APPLICATIONS OF DEEP LEARNING AND MULTI-PERSPECTIVE 2D/3D IMAGING STREAMS FOR REMOTE TERRAIN CHARACTERIZATION OF COASTAL ENVIRONMENTS

A Dissertation

by

MOHAMMAD PASHAEI

BS, Tehran Polytechnic, Iran, 2005
MS, University of Tehran, Iran, 2009

Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in

GEOSPATIAL COMPUTER SCIENCE

Texas A&M University-Corpus Christi
Corpus Christi, Texas

December 2021

December 2021

APPLICATIONS OF DEEP LEARNING AND MULTI-PERSPECTIVE 2D/3D IMAGING STREAMS FOR REMOTE TERRAIN CHARACTERIZATION OF COASTAL ENVIRONMENTS

A Dissertation

by

MOHAMMAD PASHAEI

This dissertation meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

Michael J. Starek, PhD
Chair

Philippe Tissot, PhD
Committee Member

Scott A. King, PhD
Committee Member

Craig L. Glennie, PhD
Committee Member

Kathleen Lynch-Davis, PhD
Graduate Faculty Representative

December 2021

ABSTRACT

Threats from storms, sea encroachment, and growing population demands put coastal communities at the forefront of engineering and scientific efforts to reduce vulnerabilities for their long-term prosperity. Updated and accurate geospatial information about land cover and elevation (topography) is necessary to monitor and assess the vulnerability of natural and built infrastructure within coastal zones. Advancements in remote sensing (RS) and autonomous systems extend surveying and sensing capabilities to difficult environments, enabling more geospatial data acquisition flexibility, higher spatial resolutions, and allowing humans to "see" in ways previously unattainable. Recent years have witnessed enormous growth in the application of small unmanned aircraft systems (UASs) equipped with digital cameras for hyperspatial resolution imaging and dense three-dimensional (3D) mapping using structure-from-motion (SfM) photogrammetry techniques. In contrast to photogrammetry, light detection and ranging (lidar) is an active RS technique that uses a pulsed laser mounted on a static or mobile platform (from air or land) to scan in high definition the 3D structure of a scene. Rapid proliferation in lidar technology has resulted in new scanning and imaging modalities with ever increasing capabilities such as geodetic-grade terrestrial laser scanning (TLS) with ranging distances of up to several kilometers from a static tripod. TLS enables 3D sampling of the vertical structure of occluding objects, such as vegetation, and underlying topography. Full-waveform (FW) lidar systems have led to a significant increase in the level of information extracted from a backscattered laser signal returned from a scattering object. With this technological advance and increase in remote sensing capabilities and data resolution, comes an increase in information gain at the cost of highly more complex and challenging big data sets to process and extract meaningful information. In this regard, utilizing

end-to-end analyzing techniques recently developed in artificial intelligence (AI), in particular, convolutional neural network (CNN), developed under deep learning (DL) framework, seems applicable. DL techniques have recently outperformed state-of-the-art analysis techniques in a wide range of applications including RS.

This work presents the application of DL for efficient exploitation of hyperspatial UAS-SfM photogrammetry and FW TLS data for land cover monitoring and topographic mapping in a coastal zone. Hyperspatial UAS images and TLS point cloud data with additional information about the scattering properties of illuminated target in the footprint of the laser beam encoded in returned waveform signals provide valuable geospatial data resources to uncover the accurate 3D structure of the surveyed environment.

This study presents three main contributions: 1) Evaluation of different DCNN architectures, and their efficiencies, to classify land cover within a complex wetland setting using UAS imagery is investigated; 2) DCNN-based single image super-resolution (SISR) is employed as a pre-processing technique on low-resolution UAS images to predict higher resolution images over coastal terrain with natural and built land cover, and its effectiveness for enhancing dense 3D scene reconstruction with SfM photogrammetry is tested; 3) Full waveform TLS data is employed for point cloud classification and ground surface detection in vegetation using a developed DCNN framework that works directly off of the raw, digitized echo waveforms. Results show that returned raw waveform signals carry more information about a target's spatial and radiometric properties in the footprint of the laser beam compared to waveform attributes derived from traditional waveform processing techniques. Collectively, this study demonstrates useful information retrieval from hyperspatial resolution 2D/3D RS data streams in a DL analysis framework.

DEDICATION

I would like to dedicate this dissertation to my family.

ACKNOWLEDGEMENTS

This work would not have been possible without the funding support of the National Geodetic Survey (NGS) and Conrad Blucher Institute for Surveying and Science (CBI) at Texas A&M University-Corpus Christi (TAMUCC). Each of the members of my Dissertation Committee has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general. I would especially like to thank Dr. Michael J. Starek, my supervisor and the chair of my committee, whose expertise was invaluable in formulating the research questions and methodology. As my teacher and mentor, he taught me more than I could ever give him credit for here. He has shown me, by his example, what a good researcher (and person) should be. I am especially indebted to Dr. Philippe Tissot and Dr. Scott King, my committee members, who have been supportive of my career goals and who worked actively to provide me research guidance to pursue those goals. I would also like to thank Dr. Craig Glennie, my committee member, from The University of Huston, for his insightful feedback, which pushed me to sharpen my thinking and brought my work to a higher level. I am truly thankful to Jacob Berryhill for all his efforts in collecting and preparing high quality remote sensing data used in this work.

I would also like to acknowledge the Texas Department of Transportation (TxDOT) for invaluable support of my doctoral studies. I am extremely grateful to all of those with whom I have had the pleasure to work during this and other projects. I would like to thank all students and staff of Measurement Analytics Lab (MANTIS) in Conrad Blucher Institute (CBI), CBI's staff and field crew, whose support and guidance are with me in whatever I pursue.

Most importantly, I wish to thank my loving and supportive family, who provide me endless supports and inspirations to finish this work.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I: INTRODUCTION

1.1. Importance of Coastal Zone Monitoring

Coastal zones are recognized as some of the most dynamic environments on Earth, and also some of the most pressured due to increasing population growth and anthropogenic development, impacts from episodic storms, and impacts from longer-term climate change and sea level rise. Coastal wetlands, which are the buffer zone between land and sea, are considered one of the most productive ecosystems on our planet; yet are under increasing threat by human activities, such as road construction, agriculture irrigation, and pollution, as well as by global warming and climate change-related stressors such as sea level rise, shoreline erosion, and flooding [1-4]. The continued loss of coastal ecosystems will have far-reaching ecological and economic impacts.

Scientific and engineering efforts to mitigate the loss of coastal environments, both natural landforms and built infrastructure, and improve their resiliency, requires updated and accurate geospatial data and information about land cover, topography, and how these regions are evolving. In this regard, continued development and progression of aerial and terrestrial based remote sensing (RS) techniques for efficient surveying of coastal zones seems inevitable. Information derived through the constant monitoring of coastal zones helps governments and decision makers effectively manage their protection and restoration plans [4-6]. It also aids scientists to have accurate assessment about the evolution of populated coastal zones and landform characteristics in coastal wetlands [6-11]. However, effective land cover monitoring and topographic mapping with RS technologies usually requires efficient data processing and analysis techniques to extract useful geospatial information from these datasets.

1.2. Unmanned Aircraft System (UAS) Photogrammetry

Over the past few decades, numerous developments in satellite and aerial remote sensing RS systems have been developed to acquire abundant Earth observations at regional to global scales for different applications. In recent years, rapid growth in autonomous systems technology has led to increased use of small Unmanned Aircraft Systems (also called Uncrewed Aircraft Systems, or UASs) equipped with digital cameras for accurate, local-scale aerial mapping. UAS have proven effective for this task due to their ability to acquire hyperspatial resolution imagery with ground sample distances (GSDs) on the order of a few centimeters or smaller as a result of low flying heights and high sensor resolutions [12-19]. Currently, the FAA defines small UAS (sUAS, referred to simply as UAS herein) as weighing less than 24.9 kgs (55 lbs.) including payload capacity. UAS provide flexible platforms that are easy to deploy for rapid data acquisition and to target specific events, such as post-storm reconnaissance [20-22]. Additionally, at localized geographic/areal extents, UAS survey missions are cost-effective relative to traditional piloted aircraft [19, 23-25].

These combined factors make UAS-based remote sensing an attractive technique for coastal zone monitoring and surveying. Several studies have utilized UAS imagery for some routine coastal surveying, infrastructure mapping, and landform monitoring applications [16-18, 23, 26]. Hyperspatial UAS imagery has also been employed for vegetation mapping in wetland areas [13, 15, 27] and mapping biomass evolution in coastal wetlands [27-29].

Moreover, in contrast to traditional aerial photogrammetry which requires manned aircraft, expensive metric-grade cameras and equipment, and time-consuming procedures for precise data processing, UAS equipped with small-format digital cameras in combination with Structure-from-Motion (SfM) photogrammetry can provide highly precise and detailed two-dimensional (2D) and three-dimensional (3D) geospatial data about the underlying topography and land cover of an

imaged scene [30]. Nowadays, commercial-grade UAS mapping platforms and SfM software tools for processing of UAS imagery are readily available for use by coastal engineers, surveyors, managers, and scientist, where it is increasingly being used in coastal research and surveying [18, 23, 26, 31-33]. Overlapping UAS image sequences processed by SfM photogrammetry, can generate very dense, geolocated 3D point clouds from the coastal environment at a level of spatial detail previously unattainable nor practical using traditional aerial photogrammetry techniques with piloted aircraft (see Figure 1.1). Furthermore, UAS-SfM surveys can generate other geographic information systems (GIS) data products including hyperspatial resolution orthomosaic images, digital surface models (DSMs), and 3D textured meshes [22, 34].

Although UAS-SfM photogrammetry can enable generation of dense 3D point cloud data from overlapping aerial imagers collected at low flying heights above ground, there still might exist large gaps in 3D structure of the surveyed environment, represented by the dense point cloud, as well as variability in the positional accuracy of the generated point cloud. This issue may arise from three major sources including the limitations of SfM photogrammetry including the camera, land cover/terrain complexities of the surveyed area, and environmental conditions [19, 25, 35]. For example, UAS-SfM photogrammetry may fail to provide detailed information about the vertical structure of vegetation cover within a coastal wetland area due to high wind causing movement of the vegetation and false matching errors, and it may fail to accurately resolve the underlying ground surface for digital terrain model (DTM) generation. Furthermore, SfM can suffer from lack of image texture or areas with monotonous surface patterns resulting in data gaps or spurious point cloud measurements. However, in spite of these challenges, UAS-SfM photogrammetry represents an inexpensive and efficient technique for consistent and fast airborne surveying of coastal land cover and topography at localized geographic scales.

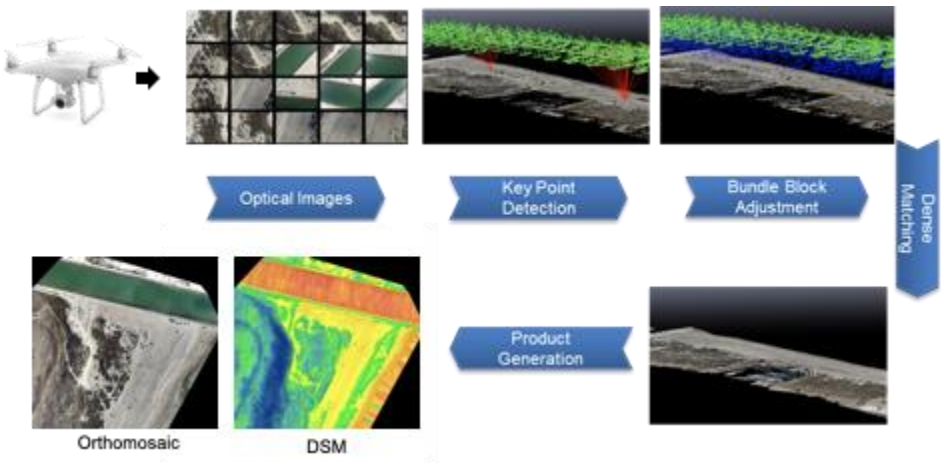


Figure 1.1. UAS-SfM photogrammetry for coastal zone mapping [23].

In addition to the 3D information (point cloud data), UAS-SfM inherently provides high resolution imagery and orthomosaics, which can subsequently be used for land cover classification and mapping tasks. However, the lower spectral resolution of the typical Red-Green-Blue (RGB) digital cameras onboard the UAS used to capture SfM imagery presents challenges for standard automated classification methods.

1.3. Light Detection and Ranging (Lidar)

Over the past few decades, airborne light detection and ranging (lidar) has evolved from a developmental technology to a proven to state-of-the-art active RS method for acquisition of accurate, high resolution land cover and topographic data by direct representation of the Earth's surface through generation of 3D point cloud data [36]. Laser ranger or range finder in a lidar system, such as an airborne laser scanning (ALS) system (Figure 1.2), consecutively transmits laser signals toward the surface of the earth and provides accurate range information between the lidar sensor and points related to different land targets.

There are two main ranging technologies and methodologies that are in widespread use for topographic applications: (1) the Time-of-Flight (TOF) or timed pulse or pulse echo method, where the travel time of a very short but intense pulse of laser radiation from the laser ranger to the object and then to the instrument, after having been reflected from the object, is accurately measured; and (2) the multiple frequency phase comparison or phase shift method for continuous wave (CW) operation using amplitude modulation (AM), where the laser rangefinder transmits a continuous beam of laser radiation instead of a discrete pulse. Phase difference between transmitted sinusoidal signal produced by the laser rangefinder and the received signal is converted to travel time [37, 38].

The TOF of the reflected pulses or phase shift of the CW is used along with the lidar sensor's geolocation data, representing the location and orientation of the lidar sensor in a predefined 3D datum, to build up a 3D point cloud representation of the surveyed area [39]. In addition, more information about the scattering properties of the illuminated target in the wavelength of the laser beam can be achieved by radiometric calibration of the lidar system. Due to its capabilities in direct and highly dense sampling from the surface of the earth and accurate representation of 3D vertical structures such as buildings, trees, and other vegetated areas, lidar systems have long been used for accurate topographic surveying and mapping, especially in forestry [40-46], 3D city and urban modeling [47-52], coastal mapping [53-59].

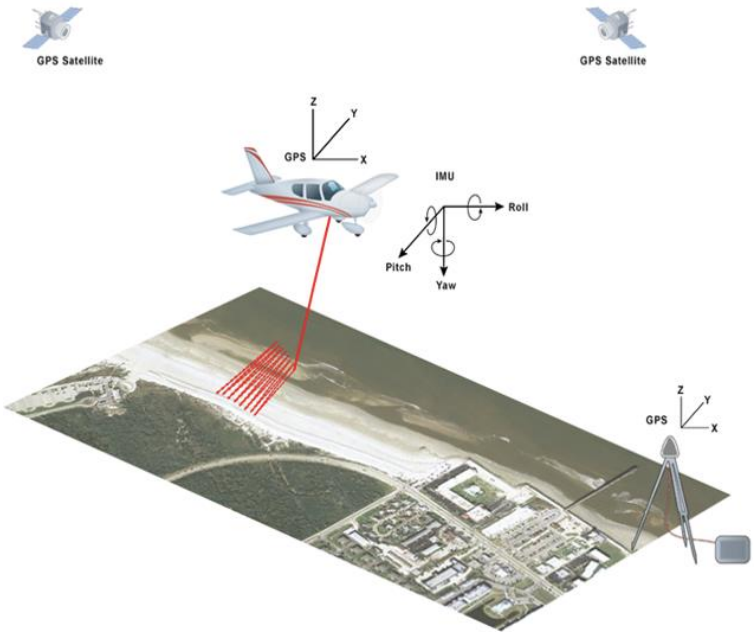


Figure 1.2. Airborne laser scanning for coastal zone mapping [60].

In the past few decades, a wide range of commercial and experimental lidar systems have been developed for different RS applications [40, 61-67]. Due to their much longer dynamic range, TOF lidar systems are the clear favorite when the ranges to be measured are long, such as lidar systems mounted in piloted aircraft. These systems can provide useful data from as close in as a meter, out to several kilometers capturing up to hundreds of thousands of points per second. Conversely, phase-based lidar systems can easily achieve a very high acquisition speed up to hundreds of thousands of points per second in short distances. However, phase-based Lidar is barely used for long range measurements since the continuous signal would have to be unacceptably powerful. Moreover, the measurement accuracy would suffer due to much higher signal-to-noise ratio and modulation waveform over long distances [38, 39]. TOF-based lidar systems are typically characterized as the analogue discrete-return measuring systems [68, 69]. It makes it possible to acquire 3D point clouds by recording accurate information concerning the

range and reflectance (amplitude at peak) of a single point with respect to each single backscattered pulse. For each emitted pulse, target detection and time-of-arrival (TOA) estimation of the returned pulse are performed in real time through analogue devices. The most useful characteristic of lidar might be that the laser energy can penetrate through small canopy gaps and measure 3D structure of tree canopy and the underlying structure on the ground surface along the transmit path of the laser beam.

Significant technological innovations have already made it possible to acquire additional information from a single pulse in lidar systems [69-72]. This is possible because a single laser signal can theoretically detect multiple objects along its path. Multi-target detection helps in separating points belonging to the soil or continuous surfaces from points belonging to different layers of vegetation [69, 73] with some limitations regarding the minimum distance between two nearby targets that can be distinguished directly from a single pulse, usually referred as Multi-Target Resolution (MTR). Echo pulses separated by shorter distances within the same laser shot cannot be physically distinguished, so that the measured range can be only estimated or even totally failed [69, 74, 75]. This multi-echo detection (also called multi-return) capability of traditional discrete-return airborne lidar systems can enable more accurate generation of DTMs under canopy, DSMs, and structure-based landcover maps.

In contrast to discrete systems, Full-Waveform (FW) lidar systems digitize and record the full energy trace of the backscattered laser signal (see Figure 1.3) [48, 67, 75-79]. In comparison to the data collected by discrete lidar systems, FW data contain additional information about the object(s) in the transmit path of the laser pulse [49, 76, 80, 81]. Waveform data collected by FW ALS systems (i.e., FW airborne lidar) has already been shown valuable for point cloud classification in both natural and built environments [48, 49, 71, 80, 82-84]. In the past few years,

a number of techniques, commonly called FW analysis (FWA) techniques, have been developed for extracting useful information from FW lidar data. The main goal in almost all FWA techniques is to precisely locate the position of each individual echo in the waveform. Moreover, by reconstructing the digitized echo pulse, through exploiting some parametric functions, such as generalized Gaussian function, more information about the physical and radiometric characteristics of the target can be extracted, which are usually used for point cloud classification or segmentation [48, 85-89].

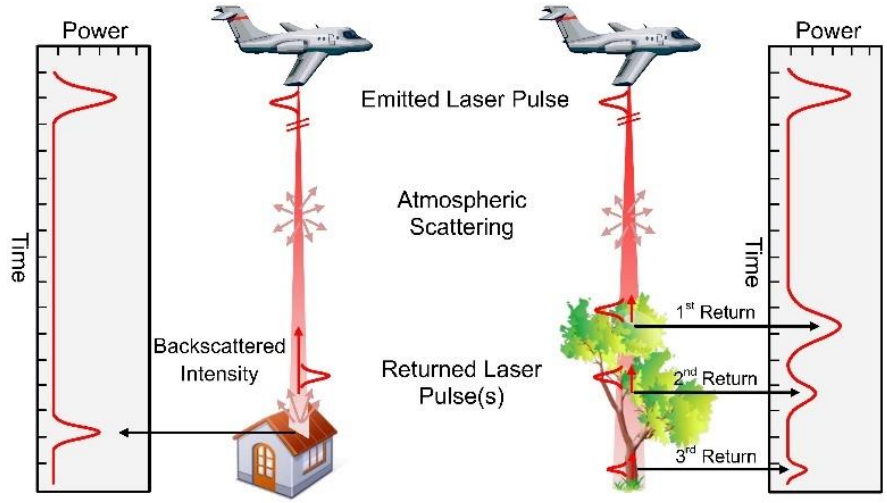


Figure 1.3. Schematic illustration of a full-waveform (FW) airborne laser scanning (ALS) system [62].

1.3.1. Terrestrial laser scanning (TLS)

Terrestrial Laser Scanning (TLS) employs a lidar scanner mounted on a tripod with a rotating sensor head to provide a 360° horizontal field-of-view and vertical field-of-view from an oblique perspective (see Figure 1.4 and Figure 1.5).

Figure 1.4. Illustration of a RIEGL VZ-2000i TLS mounted on a static tripod with an integrated RGB digital camera.

By introducing the first commercial FW Terrestrial Laser Scanning (TLS) systems in 2008, collecting FW lidar data from terrestrial platforms in a local-scale study area, is now practically possible (see Figure 1.5) [69, 74, 79, 89-92]. Despite the availability of FW TLS systems, unlike FW ALS systems, very few studies have focused on the capability of this data for TLS point cloud classification. Rogers et al. utilized discrete-return lidar data in combination with some FW features derived from a FW TLS for the assessment of the elevation uncertainty in salt marsh environment [66]. Guarnieri et al. used FW features derived from a FW TLS with a built-in FWA unit for dense vegetation filtering and create a more accurate DTM in the study area [69]. Danson et al. developed a dual wavelength FW TLS to characterize forest canopy structure [90].

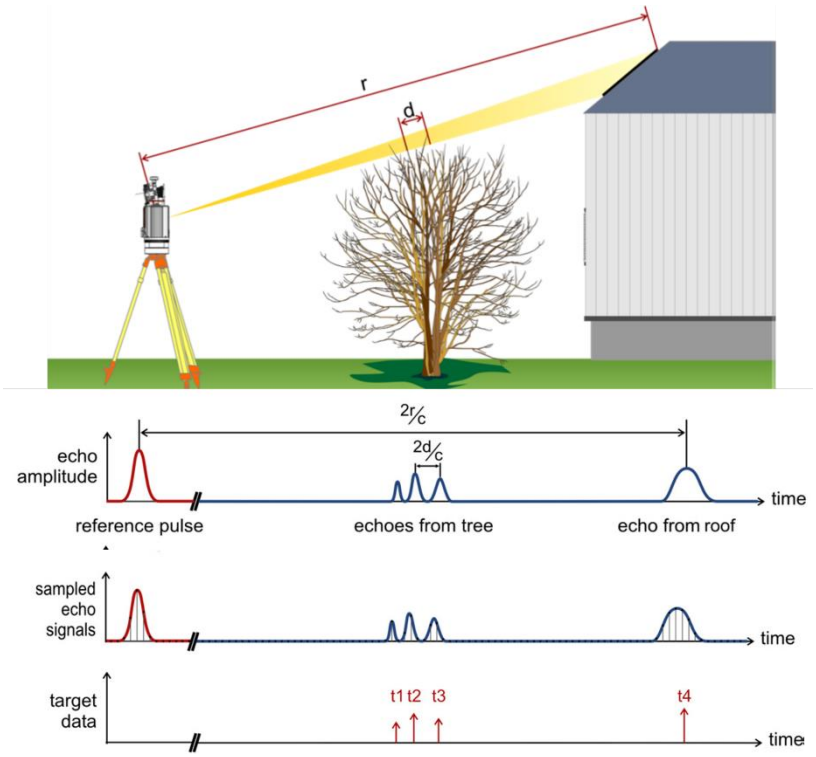


Figure 1.5. Schematic illustration of a full-waveform (FW) terrestrial laser scanning (TLS) system (image source: RIEGL).

Land cover classification and determination of ground and above-ground targets using point cloud data acquired by TLS systems in structurally complex wetland aeras, is a difficult task [93-95]. That is partly due to the complexity of the laser pulse interaction to the variety of vegetation structures, underlying topography including moist and dry land, water bodies, and other structures. Specifically, identifying ground from above-ground targets, e.g., vegetation, is essential for accurate assessment of above-ground biomass in a structurally complex environment. It can, also, help to better quantify the microtopography of the coastal environments, such as a coastal wetland by generating detailed and accurate DTM representing the terrain extension under vegetation canopy [93, 95-97].

Despite very precise and dense point cloud data collected by TLS systems, due to occlusion and the laser pulse penetrating into existing gaps in canopy structure, discriminating ground points versus above-ground points based solely on 3D point cloud data is challenging for terrain modeling applications [95, 96, 98-100]. However, FW TLS systems provide additional information about the spatial distribution and scattering properties of illuminated target(s) in the path of the laser pulse [48, 56, 69, 81, 101]. Unlike the FW ALS systems, the potential of waveform data collected by FW TLS systems for land cover monitoring and topographic mapping has not yet been fully explored (Figure 1.6).

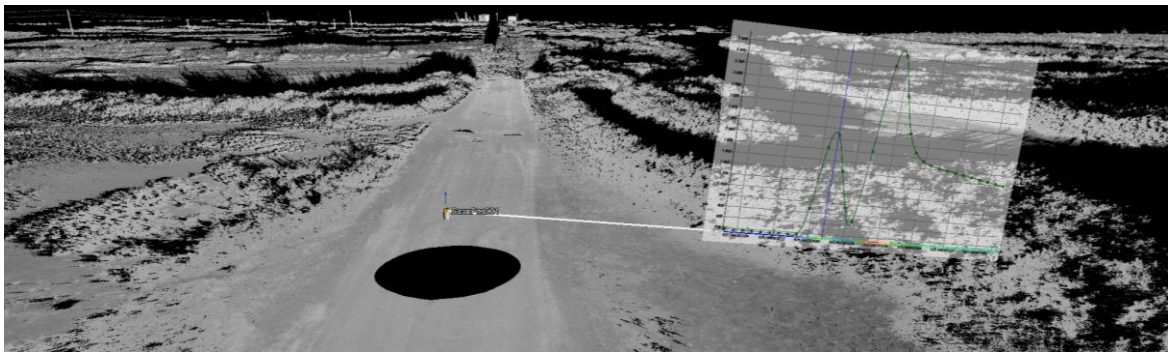


Figure 1.6. Illustration of a dense point cloud collected by a full-waveform (FW) terrestrial laser scanning (TLS) system in a complex coastal environment. Waveform data recorded for illuminated targets in the path of a single transmitted laser pulse has also been shown.

1.4. Deep Learning (DL) for Remote Sensing (RS)

With advancements in RS technology comes an exponential increase in the volume and information content of collected geospatial data. This necessitates the need for more efficient and automatic classification algorithms for fast and accurate information retrieval from raw input data.

Recent advances in Machine Learning (ML), specifically, the emerging field of Deep Learning (DL), have changed the traditional way of processing, interpreting, and manipulating

geospatial data. As a new frontier of Artificial Intelligence (AI), where feature representation and learning are carried out in an end-to-end fashion hierarchically, DL techniques have achieved huge success not only in classical computer vision tasks, but also in many other practical applications including RS [102-106]. DL methods have made significant improvements beyond the state-of-the-art techniques for data analysis in almost all domains and have attracted great interest in both academia and industrial communities.

Representation learning or feature learning procedure, which is the core concept of DL techniques, aims to explore and learn the most discriminative and representative features in an end-to-end manner within a hierarchical and relatively deep structure of feature exploration and learning [107]. Unlike almost all ML-based counterparts, such as support vector machine or multi-layer perceptron, which rely on the prior knowledge about the most informative features to achieve satisfactory results, DL models automatically explore and discover those features through their special architecture. This unique characteristic of DL architectures usually leads to the enhancement of the generalization capabilities of the DL models in problem solving for different tasks [103, 107-112].

Over the past few years, DL, in particular, Deep Convolutional Neural Networks (DCNNs), have gained significant attention in almost all analysis tasks [110, 113-115], including information retrieval from complex and huge RS data [109, 116-121]. They extract varying level of abstraction for the data in different convolutional layers. The application of DCNN techniques has been studied in a large number of land cover and land use classification tasks using hyperspectral imagery over different environments [110, 122, 123]. However, its efficiency has not been fully explored for complex land cover classification tasks such as a coastal wetland land cover monitoring, where unlike hyperspectral imagery with wide range of spectral bands, UAS-based

RGB imagery with limited spectral bands, are usually employed for mapping and monitoring of the local-scale environments [124, 125]. Limited spectral resolution along with detailed spatial information content, makes accurate land cover prediction through semantic segmentation of hyperspatial UAS images in complex environments, such as a coastal area, a challenging task. That is partly due to the high interclass similarity and intraclass variability between different object classes without a clear-cut boarder between them [126, 127].

Furthermore, DCNN-based single image super-resolution (SISR) have recently been employed as an image preprocessing technique to enhance image resolution and its information content for different applications. However, the applicability of SISR techniques in UAS-SfM photogrammetry and their capabilities in enhancing the generated geospatial data have not been fully explored. If applicable, it can optimize the efficiency of UAS data collection and quality of generated geospatial data through SfM photogrammetry in complex areas.

In addition, analyzing the recorded raw waveform TLS signal returned from the illuminated target(s) in the path of laser pulse in a DCNN framework for extracting useful information about target properties has not yet been explored. If applicable, this approach can be employed where traditional FWA techniques are not applicable or intensive calibration procedures are required to analyze the backscattered waveform for useful information retrieval.

1.5. Study Purpose and Contributions

Threats from storms, sea encroachment, and growing population demands put coastal communities at the forefront of engineering and scientific efforts to reduce vulnerabilities for their long-term prosperity. Developing techniques for continuous, fast, and accurate RS-based monitoring and mapping aids scientific understanding of the dynamic nature of coastal environments in both natural, such as wetlands, and built scenarios. This information can also aid

decision making and engineering design to better manage coastal environments and improve their resiliency.

With this motivation, this study focuses on exploration and development of DL-based techniques for exploiting dense, high resolution 2D and 3D imaging streams collected from UAS, SfM, and FW TLS for topographic mapping and efficient land cover classification of natural and built coastal environments.

First, hyperspatial UAS RGB images over a wetland area are processed within some of the most popular DCNN architectures, which have originally been developed for other image analyses and image understanding tasks in computer vision, medical imaging, and others. The main goal is to investigate the capability of advanced DL architectures in semantic image segmentation, for land cover mapping, where unlike regular images, the DCNN model is trained and evaluated on RS images acquired by the UAS flight over a complex coastal wetland environment.

Furthermore, the application of DCNN-based SISR technique, as the most recent technique in computer vision to enhance the spatial resolution and information content of typical images is explored for efficient UAS-SfM photogrammetry procedure where very high resolution (HR) images with high level of information content about the surveyed area are predicted in a DCNN model from low resolution (LR) images rather than flying at lower altitude. The main goal is to evaluate how effective SISR performs for a dense 3D reconstruction task with UAS-SfM. In return, if effective, this method could help optimize UAS-SfM data acquisition over coastal terrain by enabling UAS flights to be conducted at higher altitude/lower resolution due to time, cost, or environmental constraints.

Lastly, FW TLS data acquired within a built environment and coastal wetland is processed in a proposed DCNN model for accurate multi-class classification and enhanced representation of

the 3D structure of the surveyed environment. In this research experiment, raw waveform information related to the measured points rather than their typical spatial, radiometric, and calibrated waveform attributes from parametrically fitted waveform models or approximated in a calibration procedure, are used directly as waveform attributes for each individual point in a multi-class TLS point cloud classification task.

The contributions of this research are summarized as follows:

1. Evaluation of different DCNN architectures, and their efficiencies, to classify land cover within a complex wetland setting using UAS imagery. Research questions which are answered in this study include: (a) Can a DCNN model successfully discriminate different land cover classes in a complex wetland, where there are high inter-class similarity and intra-class variability among different classes represented by hyperspatial UAS image pixels with very limited spectral bands? (b) Is transfer learning, as a technique to reduce the number of instances required for training a DCNN model, applicable for efficiently training DCNN models on UAS images with a limited number of training instances provided for accurate wetland land cover classification? (c) Which model represents the most appropriate model among others for fast and accurate land cover prediction?
2. Investigation of DCNN-based SISR techniques for enhancing dense 3D scene reconstruction with UAS-SfM photogrammetry. Main questions which are answered in this study include: (a) Can a pretrained DCNN model for SISR efficiently generalize the transition from LR to HR image space with limited hyperspatial resolution UAS images, as training instances, through transfer learning? (b) If HR images are predicted from LR UAS images using the

underlying SISR technique, what would be the impact of this artificial transition from LR to HR image space on the information content of the predicted HR images, geometry of 2D image space, and quality of the reconstructed 3D scene using SfM?

3. Exploitation of full-waveform TLS data with DCNN framework for point cloud classification and ground surface detection within vegetation. This work develops a novel technique to classify dense point clouds acquired by a FW TLS system, equipped with a waveform digitizer and built-in online waveform processing unit where samples of the digitized single-peak echo waveform are used to populate the feature vector of the corresponding point in the point cloud for classification. A DCNN model is proposed and implemented for feature exploration and learning in multi-class classification tasks. Main questions which are answered in this study include: (a) Are the raw waveform samples, as point attributes (features), informative enough for accurate TLS point cloud classification over different environments? (b) Can the proposed classification approach outperform TLS point cloud classification based on calibrated waveform attributes provided by the TLS' built-in online waveform processing unit and calibrated look-up table (LUT) through an intense calibration procedure performed by the manufacturer? (c) If yes, what is the impact of such classification enhancement in generating an accurate 3D land cover map and DTM to represent the complex 3D structure of a natural wetland, or built environment? (d) How stable are waveform samples (attributes), temporally, versus the calibrated online waveform attributes for similar targets?

Collectively, this study demonstrates useful information retrieval from hyperspatial resolution 2D/3D RS data streams in a DL analysis framework.

1.6. Organization of the Manuscript

The remainder of the dissertation is organized into a series of three self-contained journal publications (Chapters II-IV) representing each contribution and a concluding chapter. Chapter II discusses the application and implementation of DL models for coastal wetland land cover classification, where hyperspatial resolution RGB images acquired using a UAS over a coastal wetland are processed in different DCNN models for pixel-wise classification and results evaluated in both accuracy and efficiency. Chapter III introduces DL-based SISR as a practical technique to enhance the spatial resolution of UAS imagery over a coastal environment for enhancing dense point generation with SfM photogrammetry. This section provides and discusses the results derived from applying a DL-based SISR model on LR UAS images to predict corresponding HR images input into an SfM processing workflow. Chapter IV discusses a novel technique to directly employ the raw waveform signals collected by a FW TLS system for land cover and bare-earth ground point classification for coastal environments. In this section a DCNN architecture is proposed for direct classification of raw waveform signals tested in a natural wetland and built coastal environment. Finally, Chapter V summarizes findings from Chapters II-IV as well as proposes future research directions based on this work. The citations for the journal publications reflected in Chapters II-IV are listed below.

Chapter II. Pashaei, M., Kamangir, H., Starek, M. J., & Tissot, P. (2020). Review and evaluation of deep learning architectures for efficient land cover mapping with UAS hyper-spatial imagery: A case study over a wetland. *Remote Sensing*, *12*(6), 959.

Chapter III. Pashaei, M., Starek, M. J., Kamangir, H., & Berryhill, J. (2020). Deep learning-based single image super-resolution: An investigation for dense scene reconstruction with UAS photogrammetry. *Remote Sensing*, *12*(11), 1757.

Chapter IV. Pashaei, M., Starek, M. J., Glennie, C. L., & Berryhill, J. (2021). Terrestrial Lidar Data Classification Based on Raw Waveform Samples Versus Online Waveform Attributes. Submitted to *IEEE Transactions on Geoscience and Remote Sensing*, (under revision at the time of writing this dissertation).

1.7. References

1 Cahoon, D.R., and Guntenspergen, G.R.: ‘Climate change, sea-level rise, and coastal wetlands’, National Wetlands Newsletter, 2010, 32, (1), pp. 8-12

2 Camacho-Valdez, V., Ruiz-Luna, A., Ghermandi, A., Berlanga-Robles, C.A., and Nunes, P.A.: ‘Effects of land use changes on the ecosystem service values of coastal wetlands’, Environmental management, 2014, 54, (4), pp. 852-864

3 Stedman, S.-M., and Dahl, T.E.: ‘Status and trends of wetlands in the coastal watersheds of the eastern United States, 1998 to 2004’, 2008

4 Zhang, L., Wang, M.H., Hu, J., and Ho, Y.S.: ‘A review of published wetland research, 1991-2008: Ecological engineering and ecosystem restoration’, Ecol Eng, 2010, 36, (8), pp. 973-980

5 Klemas, V.V., Dobson, J.E., Ferguson, R.L., and Haddad, K.D.: ‘A coastal land cover classification system for the NOAA Coastwatch Change Analysis Project’, Journal of Coastal Research, 1993, pp. 862-872

6 Boesch, D.F., Josselyn, M.N., Mehta, A.J., Morris, J.T., Nuttle, W.K., Simenstad, C.A., and Swift, D.J.: ‘Scientific assessment of coastal wetland loss, restoration and management in Louisiana’, Journal of Coastal Research, 1994, pp. i-103

7 Malthus, T.J., and Mumby, P.J.: ‘Remote sensing of the coastal zone: an overview and priorities for future research’, 2003

8 Ghosh, M.K., Kumar, L., and Roy, C.: ‘Monitoring the coastline change of Hatiya Island in Bangladesh using remote sensing techniques’, ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 101, pp. 137-144

9 Gens, R.: ‘Remote sensing of coastlines: detection, extraction and monitoring’, Int J Remote Sens, 2010, 31, (7), pp. 1819-1836

10 Guo, M., Li, J., Sheng, C.L., Xu, J.W., and Wu, L.: ‘A Review of Wetland Remote Sensing’, Sensors-Basel, 2017, 17, (4)

11 Tiner, R.W., Lang, M.W., and Klemas, V.V.: ‘Remote sensing of wetlands: applications and advances’ (CRC press, 2015. 2015)

12 Strecha, C., Fletcher, A., Lechner, A., Erskine, P., and Fua, P.: ‘Developing species specific vegetation maps using multi-spectral hyperspatial imagery from unmanned aerial vehicles’, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012, 3, pp. 311-316

13 Zweig, C.L., Burgess, M.A., Percival, H.F., and Kitchens, W.M.: ‘Use of unmanned aircraft systems to delineate fine-scale wetland vegetation communities’, Wetlands, 2015, 35, (2), pp. 303-309

14 Lu, B., and He, Y.: ‘Species classification using Unmanned Aerial Vehicle (UAV)-acquired high spatial resolution imagery in a heterogeneous grassland’, ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 128, pp. 73-85

15 Pande-Chhetri, R., Abd-Elrahman, A., Liu, T., Morton, J., and Wilhelm, V.L.: ‘Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery’, European Journal of Remote Sensing, 2017, 50, (1), pp. 564-576

16 Drummond, C.D., Harley, M.D., Turner, I.L., A Matheen, A.N., and Glamore, W.C.: ‘UAV applications to coastal engineering’, in Editor (Ed.)^(Eds.): ‘Book UAV applications to coastal engineering’ (Engineers Australia and IPENZ, 2015, edn.), pp. 267

17 Nikolakopoulos, K.G., Kozarski, D., and Kogkas, S.: ‘Coastal areas mapping using UAV photogrammetry’, in Editor (Ed.)^(Eds.): ‘Book Coastal areas mapping using UAV photogrammetry’ (International Society for Optics and Photonics, 2017, edn.), pp. 104280O

18 Turner, I.L., Harley, M.D., and Drummond, C.D.: ‘UAVs for coastal surveying’, Coastal Engineering, 2016, 114, pp. 19-24

19 Whitehead, K., and Hugenholtz, C.H.: ‘Remote sensing of the environment with small unmanned aircraft systems (UASs), part 1: A review of progress and challenges’, Journal of Unmanned Vehicle Systems, 2014, 2, (3), pp. 69-85

20 Colomina, I., and Molina, P.: ‘Unmanned aerial systems for photogrammetry and remote sensing: A review’, ISPRS Journal of photogrammetry and remote sensing, 2014, 92, pp. 79-97

21 Nex, F., and Remondino, F.: ‘UAV for 3D mapping applications: a review’, Applied geomatics, 2014, 6, (1), pp. 1-15

22 Klemas, V.V.: ‘Coastal and environmental remote sensing from unmanned aerial vehicles: An overview’, Journal of coastal research, 2015, 31, (5), pp. 1260-1267

23 Starek, M.J., Gingras, M., and Jeffress, G.: ‘Application of Unmanned Aircraft Systems for Coastal Mapping and Resiliency’, Sustainable Development Goals Connectivity Dilemma: Land and Geospatial Information for Urban and Rural Resilience, 2019, pp. 109

24 Starek, M.J., Davis, T., Prouty, D., and Berryhill, J.: ‘Small-scale UAS for geoinformatics applications on an island campus’, in Editor (Ed.)^(Eds.): ‘Book Small-scale UAS for geoinformatics applications on an island campus’ (IEEE, 2014, edn.), pp. 120-127

25 Hardin, P.J., Lulla, V., Jensen, R.R., and Jensen, J.R.: ‘Small Unmanned Aerial Systems (sUAS) for environmental remote sensing: Challenges and opportunities revisited’, GIScience & Remote Sensing, 2019, 56, (2), pp. 309-322

26 Green, D.R., Hagon, J.J., Gómez, C., and Gregory, B.J.: 'Using low-cost UAVs for environmental monitoring, mapping, and modelling: Examples from the coastal zone': 'Coastal Management' (Elsevier, 2019), pp. 465-501

27 Durgan, S.D., Zhang, C., Duecaster, A., Fourney, F., and Su, H.: 'Unmanned Aircraft System Photogrammetry for Mapping Diverse Vegetation Species in a Heterogeneous Coastal Wetland', Wetlands, 2020, 40, (6), pp. 2621-2633

28 Zhou, Z., Yang, Y., and Chen, B.: 'Estimating Spartina alterniflora fractional vegetation cover and aboveground biomass in a coastal wetland using SPOT6 satellite and UAV data', Aquatic Botany, 2018, 144, pp. 38-45

29 Otero, V., Van De Kerchove, R., Satyanarayana, B., Martínez-Espinosa, C., Fisol, M.A.B., Ibrahim, M.R.B., Sulong, I., Mohd-Lokman, H., Lucas, R., and Dahdouh-Guebas, F.: 'Managing mangrove forests from the sky: Forest inventory using field data and Unmanned Aerial Vehicle (UAV) imagery in the Matang Mangrove Forest Reserve, peninsular Malaysia', Forest ecology and management, 2018, 411, pp. 35-45

30 Starek, M.J., and Giessel, J.: 'Fusion of UAS-based structure-from-motion and optical inversion for seamless topo-bathymetric mapping', in Editor (Ed.)^(Eds.): 'Book Fusion of UAS-based structure-from-motion and optical inversion for seamless topo-bathymetric mapping' (IEEE, 2017, edn.), pp. 2999-3002

31 Sturdivant, E.J., Lentz, E.E., Thieler, E.R., Farris, A.S., Weber, K.M., Remsen, D.P., Miner, S., and Henderson, R.E.: 'UAS-SfM for coastal research: Geomorphic feature extraction and land cover classification from high-resolution elevation and optical imagery', Remote Sensing, 2017, 9, (10), pp. 1020

32 Gonçalves, J., and Henriques, R.: ‘UAV photogrammetry for topographic monitoring of coastal areas’, ISPRS Journal of Photogrammetry and Remote Sensing, 2015, 104, pp. 101-111

33 Scarelli, F.M., Sistilli, F., Fabbri, S., Cantelli, L., Barboza, E.G., and Gabbianelli, G.: ‘Seasonal dune and beach monitoring using photogrammetry from UAV surveys to apply in the ICZM on the Ravenna coast (Emilia-Romagna, Italy)’, Remote Sensing Applications: Society and Environment, 2017, 7, pp. 27-39

34 Papakonstantinou, A., Topouzelis, K., and Pavlogeorgatos, G.: ‘Coastline zones identification and 3D coastal mapping using UAV spatial data’, ISPRS International Journal of Geo-Information, 2016, 5, (6), pp. 75

35 Linchant, J., Lisein, J., Semeki, J., Lejeune, P., and Vermeulen, C.: ‘Are unmanned aircraft systems (UAS s) the future of wildlife monitoring? A review of accomplishments and challenges’, Mammal Review, 2015, 45, (4), pp. 239-252

36 Vosselman, G., and Maas, H.-G.: ‘Airborne and terrestrial laser scanning’ (CRC press, 2010. 2010)

37 Baltsavias, E.P.: ‘Airborne laser scanning: basic relations and formulas’, ISPRS Journal of photogrammetry and remote sensing, 1999, 54, (2-3), pp. 199-214

38 Wehr, A., and Lohr, U.: ‘Airborne laser scanning—an introduction and overview’, ISPRS Journal of photogrammetry and remote sensing, 1999, 54, (2-3), pp. 68-82

39 Shan, J., and Toth, C.K.: ‘Topographic laser ranging and scanning: principles and processing’ (CRC press, 2018. 2018)

40 Lefsky, M.A., Harding, D., Cohen, W., Parker, G., and Shugart, H.: ‘Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA’, Remote Sens Environ, 1999, 67, (1), pp. 83-98

41 Dubayah, R.O., and Drake, J.B.: 'Lidar remote sensing for forestry', Journal of forestry, 2000, 98, (6), pp. 44-46

42 Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., and Hendrix, C.J.: 'Predicting forest stand characteristics with airborne scanning lidar', Photogrammetric Engineering and Remote Sensing, 2000, 66, (11), pp. 1367-1372

43 Lim, K., Treitz, P., Wulder, M., St-Onge, B., and Flood, M.: 'LiDAR remote sensing of forest structure', Progress in physical geography, 2003, 27, (1), pp. 88-106

44 Mielcarek, M., Stereńczak, K., and Khosravipour, A.: 'Testing and evaluating different LiDAR-derived canopy height model generation methods for tree height estimation', International journal of applied earth observation and geoinformation, 2018, 71, pp. 132-143

45 Liu, K., Shen, X., Cao, L., Wang, G., and Cao, F.: 'Estimating forest structural attributes using UAV-LiDAR data in Ginkgo plantations', ISPRS journal of photogrammetry and remote sensing, 2018, 146, pp. 465-482

46 Sun, C., Cao, S., and Sanchez-Azofeifa, G.A.: 'Mapping tropical dry forest age using airborne waveform LiDAR and hyperspectral metrics', International Journal of Applied Earth Observation and Geoinformation, 2019, 83, pp. 101908

47 Park, Y., and Guldmann, J.-M.: 'Creating 3D city models with building footprints and LIDAR point cloud classification: A machine learning approach', Computers, environment and urban systems, 2019, 75, pp. 76-89

48 Mallet, C., Bretar, F., Roux, M., Soergel, U., and Heipke, C.: 'Relevance assessment of full-waveform lidar data for urban area classification', ISPRS journal of photogrammetry and remote sensing, 2011, 66, (6), pp. S71-S84

49 Höfle, B., Hollaus, M., and Hagenauer, J.: 'Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data', ISPRS Journal of Photogrammetry and Remote Sensing, 2012, 67, pp. 134-147

50 Shan, J., and Aparajithan, S.: 'Urban DEM generation from raw LiDAR data', Photogrammetric Engineering & Remote Sensing, 2005, 71, (2), pp. 217-226

51 Priestnall, G., Jaafar, J., and Duncan, A.: 'Extracting urban features from LiDAR digital surface models', Computers, Environment and Urban Systems, 2000, 24, (2), pp. 65-78

52 Yi, C., Zhang, Y., Wu, Q., Xu, Y., Remil, O., Wei, M., and Wang, J.: 'Urban building reconstruction from raw LiDAR point data', Computer-Aided Design, 2017, 93, pp. 1-14

53 Wozencraft, J., and Millar, D.: 'Airborne lidar and integrated technologies for coastal mapping and nautical charting', Marine Technology Society Journal, 2005, 39, (3), pp. 27-35

54 Lee, D.S., and Shan, J.: 'Combining LIDAR elevation data and IKONOS multispectral imagery for coastal classification mapping', Marine Geodesy, 2003, 26, (1-2), pp. 117-127

55 White, S.A., Parrish, C.E., Calder, B.R., Pe'eri, S., and Rzhanov, Y.: 'Lidar-derived national shoreline: empirical and stochastic uncertainty analyses', Journal of Coastal Research, 2011, (62), pp. 62-74

56 Parrish, C.E., Rogers, J.N., and Calder, B.R.: 'Assessment of waveform features for lidar uncertainty modeling in a coastal salt marsh environment', IEEE Geoscience and Remote Sensing Letters, 2013, 11, (2), pp. 569-573

57 Launeau, P., Giraud, M., Ba, A., Moussaoui, S., Robin, M., Debaine, F., Lague, D., and Le Menn, E.: 'Full-waveform LiDAR pixel analysis for low-growing vegetation mapping of coastal foredunes in Western France', Remote Sensing, 2018, 10, (5), pp. 669

58 Le Mauff, B., Juigner, M., Ba, A., Robin, M., Launeau, P., and Fattal, P.: 'Coastal monitoring solutions of the geomorphological response of beach-dune systems using multi-temporal LiDAR datasets (Vendée coast, France)', Geomorphology, 2018, 304, pp. 121-140

59 Starek, M.J., Vemula, R., and Slatton, K.C.: 'Probabilistic detection of morphologic indicators for beach segmentation with multitemporal LiDAR measurements', IEEE transactions on geoscience and remote sensing, 2012, 50, (11), pp. 4759-4770

60 Heritage, G., and Large, A.: 'Laser scanning for the environmental sciences' (John Wiley & Sons, 2009. 2009)

61 Lemmens, M.: 'Airborne lidar sensors', GIM international, 2007, 21, (2), pp. 24-27

62 Yan, W.Y., Shaker, A., and El-Ashmawy, N.: 'Urban land cover classification using airborne LiDAR data: A review', Remote Sens Environ, 2015, 158, pp. 295-310

63 Behera, M., and Roy, P.: 'Lidar remote sensing for forestry applications: The Indian context', Current Science, 2002, 83, (11), pp. 1320-1328

64 Wang, X., Cheng, X., Gong, P., Huang, H., Li, Z., and Li, X.: 'Earth science applications of ICESat/GLAS: A review', International journal of remote sensing, 2011, 32, (23), pp. 8837-8864

65 Hakala, T., Suomalainen, J., Kaasalainen, S., and Chen, Y.: 'Full waveform hyperspectral LiDAR for terrestrial laser scanning', Optics express, 2012, 20, (7), pp. 7119-7127

66 Rogers, J.N., Parrish, C.E., Ward, L.G., and Burdick, D.M.: 'Assessment of elevation uncertainty in salt marsh environments using discrete-return and full-waveform lidar', Journal of Coastal Research, 2016, (76), pp. 107-122

67 Mallet, C., and Bretar, F.: 'Full-waveform topographic lidar: State-of-the-art', ISPRS Journal of photogrammetry and remote sensing, 2009, 64, (1), pp. 1-16

68 Pfennigbauer, M., and Ullrich, A.: ‘Multi-wavelength airborne laser scanning’, in Editor (Ed.)^(Eds.): ‘Book Multi-wavelength airborne laser scanning’ (2011, edn.), pp.

69 Guarnieri, A., Pirotti, F., and Vettore, A.: ‘Comparison of discrete return and waveform terrestrial laser scanning for dense vegetation filtering’, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012, 39, (B7), pp. 511-516

70 Wagner, W., Ullrich, A., Melzer, T., Briese, C., and Kraus, K.: ‘From single-pulse to full-waveform airborne laser scanners: potential and practical challenges’ (na, 2004. 2004)

71 Pirotti, F.: ‘Analysis of full-waveform LiDAR data for forestry applications: a review of investigations and methods’, iForest-Biogeosciences and Forestry, 2011, 4, (3), pp. 100

72 Renslow, M., Greenfield, P., and Guay, T.: ‘Evaluation of multi-return LIDAR for forestry applications’, US Department of Agriculture Forest Service-Engineering, Remote Sensing Applications. http://www. ndep. gov/USDAFS/LIDAR. pdf [Consulta: 12 de marzo de 2009], 2000

73 Pfennigbauer, M., and Ullrich, A.: ‘Three-dimensional laser scanners with echo digitization’, in Editor (Ed.)^(Eds.): ‘Book Three-dimensional laser scanners with echo digitization’ (International Society for Optics and Photonics, 2008, edn.), pp. 69500U

74 Pirotti, F., Guarnieri, A., and Vettore, A.: ‘Vegetation filtering of waveform terrestrial laser scanner data for DTM production’, Applied Geomatics, 2013, 5, (4), pp. 311-322

75 Hartzell, P.J., Glennie, C.L., and Finnegan, D.C.: ‘Empirical waveform decomposition and radiometric calibration of a terrestrial full-waveform laser scanner’, IEEE Transactions on Geoscience and Remote Sensing, 2014, 53, (1), pp. 162-172

76 Mallet, C., Soergel, U., and Bretar, F.: ‘Analysis of full-waveform lidar data for classification of urban areas’, in Editor (Ed.)^(Eds.): ‘Book Analysis of full-waveform lidar data for classification of urban areas’ (2008, edn.), pp.

77 Parrish, C.E., Jeong, I., Nowak, R.D., and Smith, R.B.: ‘Empirical comparison of full-waveform lidar algorithms’, Photogrammetric Engineering & Remote Sensing, 2011, 77, (8), pp. 825-838

78 Persson, Å., Söderman, U., Töpel, J., and Ahlberg, S.: ‘Visualization and analysis of full-waveform airborne laser scanner data’, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2005, 36, (3/W19), pp. 103-108

79 Ullrich, A., and Pfennigbauer, M.: ‘Echo digitization and waveform analysis in airborne and terrestrial laser scanning’, in Editor (Ed.)^(Eds.): ‘Book Echo digitization and waveform analysis in airborne and terrestrial laser scanning’ (2011, edn.), pp. 217-228

80 Alexander, C., Tansey, K., Kaduk, J., Holland, D., and Tate, N.J.: ‘Backscatter coefficient as an attribute for the classification of full-waveform airborne laser scanning data in urban areas’, ISPRS Journal of Photogrammetry and Remote Sensing, 2010, 65, (5), pp. 423-432

81 Höfle, B., and Hollaus, M.: ‘Urban vegetation detection using high density full-waveform airborne lidar data-combination of object-based image and point cloud analysis’ (na, 2010. 2010)

82 Koenig, K., and Höfle, B.: ‘Full-waveform airborne laser scanning in vegetation studies—a review of point cloud and waveform features for tree species classification’, Forests, 2016, 7, (9), pp. 198

83 Lasaponara, R., Coluzzi, R., and Masini, N.: ‘Flights into the past: full-waveform airborne laser scanning data for archaeological investigation’, Journal of Archaeological Science, 2011, 38, (9), pp. 2061-2070

84 Adams, T., Beets, P., and Parrish, C.: ‘Extracting more data from LiDAR in forested areas by analyzing waveform shape’, Remote Sensing, 2012, 4, (3), pp. 682-702

85 Wagner, W., Hollaus, M., Briese, C., and Ducic, V.: ‘3D vegetation mapping using small-footprint full-waveform airborne laser scanners’, International Journal of Remote Sensing, 2008, 29, (5), pp. 1433-1452

86 Guo, L., Chehata, N., Mallet, C., and Boukir, S.: ‘Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests’, ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66, (1), pp. 56-66

87 Gross, H., Jutzi, B., and Thoennessen, U.: ‘Segmentation of tree regions using data of a full-waveform laser’, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2007, 36, (part 3), pp. W49A

88 Reitberger, J., Schnörr, C., Krzystek, P., and Stilla, U.: ‘3D segmentation of single trees exploiting full waveform LIDAR data’, ISPRS Journal of Photogrammetry and Remote Sensing, 2009, 64, (6), pp. 561-574

89 Hartzell, P., Glennie, C., Biber, K., and Khan, S.: ‘Application of multispectral LiDAR to automated virtual outcrop geology’, ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 88, pp. 147-155

90 Danson, F.M., Gaulton, R., Armitage, R.P., Disney, M., Gunawan, O., Lewis, P., Pearson, G., and Ramirez, A.F.: ‘Developing a dual-wavelength full-waveform terrestrial laser scanner to characterize forest canopy structure’, Agricultural and Forest Meteorology, 2014, 198, pp. 7-14

91 Mark Danson, F., Sasse, F., and Schofield, L.A.: ‘Spectral and spatial information from a novel dual-wavelength full-waveform terrestrial laser scanner for forest ecology’, Interface Focus, 2018, 8, (2), pp. 20170049

92 Di Salvo, F., and Brutto, M.L.: ‘Full-waveform terrestrial laser scanning for extracting a high-resolution 3D topographic model: A case study on an area of archaeological significance’, European Journal of Remote Sensing, 2014, 47, (1), pp. 307-327

93 Stovall, A.E., Diamond, J.S., Slesak, R.A., McLaughlin, D.L., and Shugart, H.: ‘Quantifying wetland microtopography with terrestrial laser scanning’, Remote Sens Environ, 2019, 232, pp. 111271

94 Nguyen, C., Starek, M.J., Tissot, P., and Gibeaut, J.: ‘Unsupervised clustering method for complexity reduction of terrestrial lidar data in marshes’, Remote Sensing, 2018, 10, (1), pp. 133

95 Owers, C.J., Rogers, K., and Woodroffe, C.D.: ‘Terrestrial laser scanning to quantify above-ground biomass of structurally complex coastal wetland vegetation’, Estuarine, Coastal and Shelf Science, 2018, 204, pp. 164-176

96 Feliciano, E.A., Wdowinski, S., and Potts, M.D.: ‘Assessing mangrove above-ground biomass and structure using terrestrial laser scanning: A case study in the Everglades National Park’, Wetlands, 2014, 34, (5), pp. 955-968

97 Xie, W., Guo, L., Wang, X., He, Q., Dou, S., and Yu, X.: ‘Detection of seasonal changes in vegetation and morphology on coastal salt marshes using terrestrial laser scanning’, Geomorphology, 2021, 380, pp. 107621

98 Coveney, S., and Stewart Fotheringham, A.: ‘Terrestrial laser scan error in the presence of dense ground vegetation’, The Photogrammetric Record, 2011, 26, (135), pp. 307-324

99 Resop, J.P., and Hession, W.C.: ‘Terrestrial laser scanning for monitoring streambank retreat: Comparison with traditional surveying techniques’, Journal of Hydraulic Engineering, 2010, 136, (10), pp. 794-798

100 Lague, D., Brodu, N., and Leroux, J.: ‘Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (NZ)’, ISPRS journal of photogrammetry and remote sensing, 2013, 82, pp. 10-26

101 Doneus, M., Pfennigbauer, M., Studnicka, N., and Ullrich, A.: ‘Terrestrial waveform laser scanning for documentation of cultural heritage’, in Editor (Ed.)^(Eds.): ‘Book Terrestrial waveform laser scanning for documentation of cultural heritage’ (2009, edn.), pp.

102 Arel, I., Rose, D.C., and Karnowski, T.P.: ‘Deep machine learning-a new frontier in artificial intelligence research [research frontier]’, IEEE computational intelligence magazine, 2010, 5, (4), pp. 13-18

103 Nogueira, K., Penatti, O.A., and Dos Santos, J.A.: ‘Towards better exploiting convolutional neural networks for remote sensing scene classification’, Pattern Recognition, 2017, 61, pp. 539-556

104 Hu, F., Xia, G.-S., Wang, Z., Huang, X., Zhang, L., and Sun, H.: ‘Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification’, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8, (5)

105 Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., and Fraundorfer, F.: ‘Deep learning in remote sensing: A comprehensive review and list of resources’, IEEE Geoscience and Remote Sensing Magazine, 2017, 5, (4), pp. 8-36

106 Zhang, L., Zhang, L., and Du, B.: ‘Deep learning for remote sensing data: A technical tutorial on the state of the art’, IEEE Geoscience and Remote Sensing Magazine, 2016, 4, (2), pp. 22-40

107 LeCun, Y., Bengio, Y., and Hinton, G.: ‘Deep learning’, nature, 2015, 521, (7553), pp. 436-444

108 He, K., Zhang, X., Ren, S., and Sun, J.: ‘Deep residual learning for image recognition’, in Editor (Ed.)^(Eds.): ‘Book Deep residual learning for image recognition’ (2016, edn.), pp. 770-778

109 Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., and Wang, J.: ‘Deep learning in environmental remote sensing: Achievements and challenges’, Remote Sens Environ, 2020, 241, pp. 111716

110 Chen, Y., Lin, Z., Zhao, X., Wang, G., and Gu, Y.: ‘Deep learning-based classification of hyperspectral data’, IEEE Journal of Selected topics in applied earth observations and remote sensing, 2014, 7, (6), pp. 2094-2107

111 Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M.P., Shyu, M.-L., Chen, S.-C., and Iyengar, S.S.: ‘A survey on deep learning: Algorithms, techniques, and applications’, ACM Computing Surveys (CSUR), 2018, 51, (5), pp. 1-36

112 Deng, L.: ‘A tutorial survey of architectures, algorithms, and applications for deep learning’, APSIPA Transactions on Signal and Information Processing, 2014, 3

113 Krizhevsky, A., Sutskever, I., and Hinton, G.E.: ‘Imagenet classification with deep convolutional neural networks’, in Editor (Ed.)^(Eds.): ‘Book Imagenet classification with deep convolutional neural networks’ (2012, edn.), pp. 1097-1105

114 Dong, C., Loy, C.C., He, K., and Tang, X.: ‘Image super-resolution using deep convolutional networks’, IEEE transactions on pattern analysis and machine intelligence, 2015, 38, (2), pp. 295-307

115 Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F.E.: 'A survey of deep neural network architectures and their applications', Neurocomputing, 2017, 234, pp. 11-26

116 Li, Y., Zhang, H., Xue, X., Jiang, Y., and Shen, Q.: 'Deep learning for remote sensing image classification: A survey', Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8, (6), pp. e1264

117 Yu, X., Wu, X., Luo, C., and Ren, P.: 'Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework', GIScience & Remote Sensing, 2017, 54, (5), pp. 741-758

118 Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., and Benediktsson, J.A.: 'Deep learning for hyperspectral image classification: An overview', IEEE Transactions on Geoscience and Remote Sensing, 2019, 57, (9), pp. 6690-6709

119 Zhao, W., and Du, S.: 'Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach', IEEE Transactions on Geoscience and Remote Sensing, 2016, 54, (8), pp. 4544-4554

120 Guan, H., Yu, Y., Ji, Z., Li, J., and Zhang, Q.: 'Deep learning-based tree classification using mobile LiDAR data', Remote Sensing Letters, 2015, 6, (11), pp. 864-873

121 Zhang, L., Shao, Z., Liu, J., and Cheng, Q.: 'Deep learning based retrieval of forest aboveground biomass from combined LiDAR and landsat 8 data', Remote Sensing, 2019, 11, (12), pp. 1459

122 Makantasis, K., Karantzalos, K., Doulamis, A., and Doulamis, N.: 'Deep supervised learning for hyperspectral data classification through convolutional neural networks', in Editor (Ed.)^(Eds.): 'Book Deep supervised learning for hyperspectral data classification through convolutional neural networks' (IEEE, 2015, edn.), pp. 4959-4962

123 Audebert, N., Le Saux, B., and Lefèvre, S.: ‘Deep learning for classification of hyperspectral data: A comparative review’, IEEE geoscience and remote sensing magazine, 2019, 7, (2), pp. 159-173

124 Papakonstantinou, A., Batsaris, M., Spondylidis, S., and Topouzelis, K.: ‘A Citizen Science Unmanned Aerial System Data Acquisition Protocol and Deep Learning Techniques for the Automatic Detection and Mapping of Marine Litter Concentrations in the Coastal Zone’, Drones, 2021, 5, (1), pp. 6

125 Liu, T., Abd-Elrahman, A., Morton, J., and Wilhelm, V.L.: ‘Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system’, GIScience & remote sensing, 2018, 55, (2), pp. 243-264

126 Rezaee, M., Mahdianpari, M., Zhang, Y., and Salehi, B.: ‘Deep convolutional neural network for complex wetland classification using optical remote sensing imagery’, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11, (9), pp. 3030-3039

127 Doughty, C.L., and Cavanaugh, K.C.: ‘Mapping coastal wetland biomass from high resolution unmanned aerial vehicle (UAV) imagery’, Remote Sensing, 2019, 11, (5), pp. 540

CHAPTER II: REVIEW AND EVALUATION OF DEEP LEARNING ARCHITECTURES FOR EFFICIENT LAND COVER MAPPING WITH UAS HYPERSPATIAL IMAGERY: A CASE STUDY ON WETLAND

Abstract

Deep learning has already been proved as a powerful state-of-the-art technique for many image understanding tasks in computer vision and other applications including remote sensing (RS) image analysis. Unmanned aircraft systems (UASs) offer a viable and economical alternative to a conventional sensor and platform for acquiring high spatial and high temporal resolution data with high operational flexibility. Coastal wetlands are among some of the most challenging and complex ecosystems for land cover prediction and mapping tasks because land cover targets often show high intra-class and low inter-class variances. In recent years, several deep convolutional neural network (CNN) architectures have been proposed for pixel-wise image labeling, commonly called semantic image segmentation. In this work, some of the more recent deep CNN architectures proposed for semantic image segmentation are reviewed, and each model's training efficiency and classification performance are evaluated by training it on a limited labeled image set. Training samples are provided using the hyper-spatial resolution UAS imagery over a wetland area and the required ground truth images are prepared by manual image labeling. Experimental results demonstrate that deep CNNs have a great potential for accurate land cover prediction task using UAS hyper-spatial resolution images. Some simple deep learning architectures perform comparable or even better than complex and very deep architectures with remarkably fewer training epochs. This performance is especially valuable when limited training samples are available, which is a common case in most RS applications.

2.1. Introduction

Remote sensing (RS) is the major source of spatial information related to the earth's surface, offering a wide range of sensors and platforms to monitor land cover and its spatial distribution. Recently, Unmanned Aircraft Systems (UASs) are widely employed in numerous RS applications including natural resource management [1-3]. In comparison with traditional RS, UAS technology stands out for its low-cost operation and ability to acquire image data with high spatial and temporal resolution in a flexible fashion at local scales. UAS usually flies at low altitudes and captures high spatial resolution (few cm to sub-cm) images. In combination with the recent advancement in image analysis algorithms, those high-quality images may significantly improve the overall accuracy of image-derived products in many different RS tasks. For instance, pixel-level labeling, which is frequently used in computer vision tasks such as semantic image segmentation and instance segmentation, is eminently applicable to UAS hyper-spatial resolution imagery. Semantic image segmentation refers to the process of associating each individual pixel of an image with a predefined class label [4]. On the other hand, instance segmentation refers to the task that treats multiple objects of the same class as distinct individual objects (instances) [5].

Wetlands are known as one of the most important ecosystems on our planet. They can be characterized as transitional areas between permanently flooded deep water environments and well-drained highlands, where the water table is usually at or near the surface and the land is inundated by shallow water [6]. Coastal wetlands are important as highly dynamic natural ecosystems offering remarkable services essential to people and the environment including, wildlife habitat for myriad species of marine and terrestrial plants and animals, storm protection, erosion control, nutrient filtering, and recreation as tourist stops. These services are estimated to value at billions of dollars [7]. Authors in [8] highlight the need for monitoring wetland vegetation and its distribution to detect changes in the terrestrial-aquatic transition. Studies show that world

wetland loss and degradation has been accelerated for the last three decades mostly due to both anthropogenic and natural factors. According to a report published by the US Fish and Wildlife Service (FWS) and the National Oceanic and Atmospheric Administration's (NOAA) National Marine Fisheries Service (NMFS), a net loss of about 361,000 acres of coastal wetlands in the eastern United States occurred between 1998 and 2004 -an average net reduction of 59,000 acres per year [6]. Sustainable management of any dynamic ecosystem requires, among other parameters, a thorough understanding of its different types of land cover.

Coastal wetland classification is challenging because vegetation and other land cover objects modulate with water level fluctuation and other environmental stressors, leading to sometimes rapid and frequent changes in the type and spatial distribution of a certain land cover [9]. The ability to accurately and quickly monitor and predict land cover undergoing rapid and seasonal variations in response to changing environmental factors, including seasonal and climate changes, topography, sea-level rise, water temperature, altered flooding and salinity patterns, etc. [10-12], is crucial for updated and/or continuous land cover monitoring systems. Wetland land cover processes as well as other dynamic landscapes are further complicated by the need for frequent data collection methods, and the subsequent demands for faster and automatic algorithms analyzing very high spatial, temporal, and spectral resolution imagery by the monitoring system with the lowest level of human intervention. In particular, achieving such continuous or near-real time land cover monitoring systems becomes more challenging where expert knowledge is required for designing and extracting the most efficient and discriminative features for different states of the land cover due to the change in participating factors. Furthermore, pixel-wise labeling using mere spectral information in natural environments usually gives rise to unsatisfactory results due to higher inter-class spectral similarity and intra-class spectral variability among natural

targets [13]. This issue is highlighted especially where high-spatial resolution imagery from a lower spectral resolution sensor is employed for classification, such as consumer-grade digital RGB cameras commonly employed on small UAS for mapping purposes [14]. Moreover, natural targets such as vegetation or water bodies are not usually enclosed by well-defined boundaries in airborne images resulting in more uncertainties in the pixel-wise labeling for the land cover prediction. In addition, due to high spatial autocorrelation among natural targets, the relationship between the target pixel and its neighboring pixels need to be incorporated into subsequent analyses [15]. Thus, to take full advantage of the UAS-based high-spatial resolution imagery, image analysis algorithms exploiting spatial, spectral, contextual, and textural information, collectively, are highly recommended for precise land cover prediction [16-19].

Exploiting sophisticated techniques and algorithms along with some level of field operations for ground truthing and results validation are often a few required components for accurate monitoring of the wetland or other natural environments through remote sensing image classification [11, 12, 17]. In traditional RS classification techniques, pixel-wise classification methods assume each pixel is pure and typically labeled to the most likely land cover category. Object-based image analysis (OBIA) techniques, on the other hand, provided a new paradigm to classify RS images, where, by utilizing both spectral and contextual image features, it can outperform the pixel-based techniques [13, 15]. By exploiting OBIA techniques, geographical objects, instead of individual pixels, form the basic unit for image analysis [16]. Unlike pixel-based analysis, in OBIA, a certain image is segmented into relatively homogeneous and semantically coherent objects based on a predefined homogeneity criteria at different scales [16]. In other words, spectral information is aggregated per object, where other textural and contextual information become available for conducting image classification on objects rather than pixels

[20]. Several studies have already shown the higher performance of object-based image classification techniques than pixel-based methods, especially when high-spatial resolution images are employed [13, 20, 21]. In general, both pixel-wise and OBIA strategies for land cover or land use classification, take advantage of a wide variety of supervised or unsupervised machine learning (ML) classification algorithms [22-26].

In recent years, however, due to the striking achievement of deep learning models in outperforming almost all state-of-the-art techniques in a wide range of applications, the RS community is shifting its attention to deep learning models. The large number of publications exploiting these models in different RS image analyses and the reported accuracies demonstrate the potential of deep learning in this field of study [27-30]. The recent success of deep convolutional neural networks (CNNs) has enabled substantial progress in many image understanding tasks including pixel-wise semantic image segmentation due to a rich hierarchical feature learning process. Hierarchical features are learned through an end-to-end trainable framework in which higher levels of the feature hierarchy are formed by the precise composition of the lower-level features [31-34]. Learned features, at multiple levels of abstraction, provide a unified, highly complex mapping function from input to output taking only as input the raw data. Such complex mapping not only considers the spectral information of each individual pixel in the image, but also takes all textural, contextual, and spatial information related to each individual pixel into account. Thanks to the recent rise of transfer learning techniques, it is possible to take a pre-trained deep CNN model, trained over a large dataset in a supervised or unsupervised manner, and leverage high complex mappings learned by very deep CNN models to perform effectively on downstream tasks [35]. In addition, due to exploiting end-to-end trainable models within the deep learning framework, efficient feature engineering, which is the biggest concern for almost all

traditional classification techniques, is entirely eliminated. This paves the path for developing fully autonomous and online land cover prediction systems. All these characteristics are extremely important in many image analyses in different RS tasks. Specifically, deep CNN models have been successfully used for RGB, multispectral, and hyperspectral RS image analyses in various applications [36-39]. Very recently, deep CNNs have been specifically applied to wetland studies, including land cover classification. Results and findings confirm where adequate labeled training samples are available, deep CNN models usually outperform the traditional and machine learning classification techniques [3, 40-43].

The objectives of this work include: (1) employing some of the most popular deep CNN architectures extensively used in computer vision community for semantic image segmentation on hyper-spatial resolution UAS images acquired over a coastal wetland for land cover prediction; (2) investigating the feasibility of deep learning architectures and evaluating the performance of different deep CNN models in pixel-wise image labeling where labeled training samples are limited and natural targets that appear in UAS images with high spatial resolution exhibit high complexity in their spectral and textural information without clear borders to distinguish other neighboring targets; (3) identifying a deep learning architecture representing, among others, a high performance CNN model from speed and accuracy points of view which can be effectively used in many RS applications where complex pixel-level analyses on high-spatial resolution imagery are required.

The author should emphasize that a comprehensive study on coastal wetland classification to perform detailed analyses of vegetation or other land cover properties is not the objective of this work. Furthermore, the study of land cover changes over time in the coastal wetland setting due to changes in participating environmental factors is not a goal at this stage. Nonetheless, due to the

complexity of the coastal wetland setting relative to many other natural environments, in terms of providing higher inter-class spectral similarity and higher intra-class spectral variability, variable target boundaries and spatial distributions, and mixed pixels, this environment has been chosen as a suitable and challenging case study. For evaluating the efficiency of the employed deep CNN models, performance metrics commonly employed for evaluating model performance of semantic image segmentation tasks in computer vision are utilized. These metrics usually take the ground truth images as the existing reality and compare the predicted images with the corresponding ground truth images based on manual labeling of the image data.

2.2. Deep Learning for Semantic Image Segmentation

Advancing deep learning architectures to tackle pixel-wise image labeling is a natural step in the progress from coarse to fine inference [4]. The origin of convolutional neural networks could be located at handling classification tasks where a certain category was predicted for the entire image [44]. Target localization and detection in computer vision tasks was the next necessary step towards fine-grained inference providing further information, other than classes. Instance segmentation which joins detection and segmentation is an additional improvement towards fine-grained inference [45].

Fully Convolutional Network (FCN) [4] is considered a milestone in transforming classification-purposed CNNs for semantic image segmentation by replacing fully connected layers with convolutional ones to output spatial maps instead of classification scores. Moreover, to compensate for low resolution prediction maps due to several down-sampling steps within pooling layers, FCN includes several fractionally-strided convolutions, also known as deconvolutions or transposed convolution [46, 47], combined with a simple bilinear or any learnable interpolation allowing per-pixel labeled output. FCN can be trained end-to-end to

efficiently learn to predict pixels' categories for an image of arbitrary size. This approach achieved significant improvement over traditional methods on the PASCAL Visual Object Classes (VOC) [48] standardized image dataset with high efficiency at inference time. Despite its simplicity and flexibility, FCN architecture suffers from some critical limitations when it is applied for certain applications. FCN has a fixed receptive field which makes the network unable to capture contextual information appropriate for pixel-wise labeling for objects that are substantially smaller or larger than the predefined fixed receptive field [34]. As a result, predictions are more uncertain for local ambiguous regions. Feature maps that are used for prediction in several layers of the CNN architecture have contextual information appropriate for the classification task, not the pixel-wise labeling. Additionally, the entire network is usually trained to be spatially invariant, which does not let the network take useful global context information into account. Furthermore, the network suffers from lack of instance-awareness which is very important in some image understanding tasks [34].

Since the introduction of FCN in 2015, a wide range of research has focused on how to provide dense segmentation maps with pixel-level accuracy from arbitrary sized images. Recently introduced deep learning architectures owe their high performances in precise semantic segmentation to several factors including:

(a) introduction of more advanced and deeper CNN feature encoders that are efficiently trained using recently developed advanced optimization algorithms.

(b) utilizing a more advanced decoding strategy to the final low-resolution encoded feature maps in an encoder--decoder architecture using deconvolution or dilated convolution to efficiently increase their resolution for pixel-wise prediction.

(c) using the skip connection to introduce low-level abstract information to the high-level abstract information to build highly accurate feature maps representing pixel-level feature information.

2.2.1. Feature encoders

Feature encoders are simply described as a stack of convolution layers in combination with activation functions, usually (*ReLU*) [49], and pooling layers, usually Max-Pooling, which construct a hierarchical representation of the input data containing low-level to high-level abstract information [47]. LeNet [50] is considered as the first CNN-based feature encoder introduced by LeCun et al. in 1998. However, AlexNet [51], the first deep CNN architecture, introduced by Alex Krizhevsky in 2012 is a landmark in deep learning history. Several key factors are contributing to this progress: (1) the efficient training procedure implemented on the modern GPUs [51], (2) the proposal of the *ReLU* activation function, which had significant contribution in boosting training and made convergence much faster, and (3) the availability of a huge dataset, e.g., ImageNet [52] to train models with high capacity which include millions of trainable parameters. VGG-Net [53], GoogLeNet [54], Residual Network (ResNet) [55], and Densely Connected Network (DensNet) [56] are a few examples of popular architectures that are frequently employed for feature extraction in very deep CNN models.

2.2.1.1.VGG-Net

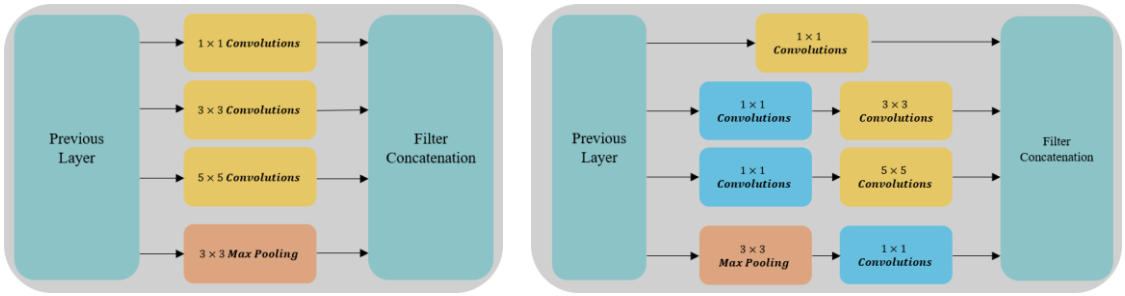
VGG-Net [53] was invented in 2014 by Oxford's Visual Geometry Group as a successful effort to build and train a very deep CNN. VGG-Net showed that the depth of a network is a critical component in CNNs to achieve high performance in recognition or classification. By shrinking the convolution kernels to 3×3 yet increasing the number of sequences of convolutional layers and

feature maps in each convolution layer, VGG isable to train deeper architecture with appropriate receptive field comparable with AlexNet for recognition tasks.

2.2.1.2. GoogleNet

GoogLeNet (a.k.a. Inception Net) from Google in 2015 was proposed by Szegedy et al. [54] with the objective of reducing computation complexity compared to the traditional CNNs. Inception module, which makes building block for the network, is a combination of 1×1, 3×3, and 5×5 convolutional kernels and a pooling layer. The motivation behind inception module is to increase the receptive field without losing fine information. By learning and combining features with different scales in parallel in each inception module, GoogLeNet is able to learn feature hierarchy in a multi-scale manner while its innovative architecture reduces the number of trainable parameters in a really deep framework (22 layers) to less than 5 million parameters in comparison to 62 million and 138 million parameters in AlexNet and VGG-Net, respectively. To train a deep stack of inception modules in an efficient way, bottleneck approach is exploited in which extra 1×1 convolutions reduce the dimensionality of feature maps that enter the inception module from the previous layer. This helps to avoid parameter explosion in inception modules and the overfitting problem in the whole network. Figure 2.1 illustrates the architecture of the inception module. Other versions of inception modules including BN-Inception [57], Inception V2, and Inception V3 [58] were later proposed. In order to increase the efficiency and performance of inception modules, in 2017, Szegedyetal et al. proposed a combined version of inception modules and residual network (ResNet) modules known as Inception-ResNet [59]. Xception [60], which stands for extreme version of inception, was proposed by Chollet et al. in 2017. The motivation behind it is to disjointly map cross-channels and spatial information in feature maps as their correlation is sufficiently decoupled. As a result, the depth-wise separable convolutions from

inception modules are modified in Xception modules as separable pointwise convolutions follow by depth-wise convolutions.



(a) (b)

Figure 2.1. Inception modules. (a) Naïve Inception and (b) Inception V1.

2.2.1.3. ResNet

As mentioned above, deeper networks can improve the performance of deep learning approach to solve complex visual tasks, but they are more prone to the notorious problem of vanishing/exploding gradients during training as well. It may lead to not only saturated accuracy, but also degradation of training accuracy. ResNet designed by He et al. in 2015 exploits residual blocks to overcome the vanishing gradient problem in very deep CNNs by introducing identity shortcut connections to successive convolution layers as shown in Figure 2.2. The shortcut connections in residual blocks help gradients flow easily in back propagation step which leads to gaining accuracy during the training phase in a very deep network. Referring to Figure 2.2, each unit calculates a residual function $F(x) = H(x) - x$, in which x is the output of the previous residual unit and $H(x)$ denotes the desired underlying mapping. More precisely, if y_l is the output of the l-th residual unit with weights w_l, then,

$$y_l = x_l + F(x_l, w_l) \quad (1)$$

$$x_{(l+1)} = f(y_l) \tag{2}$$

where $f(\,)$ is the activation function.

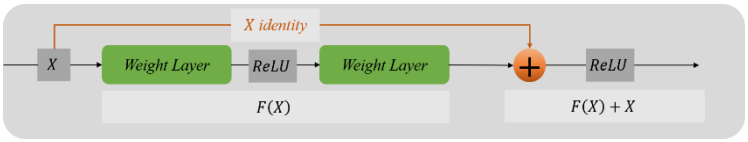


Figure 2.2. Basic diagram of residual unit.

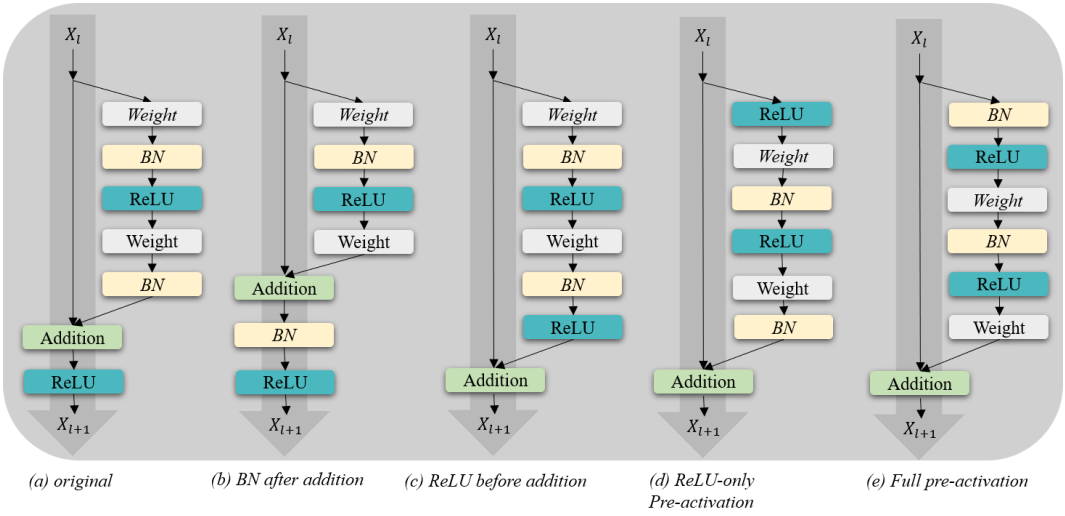


Figure 2.3. Different variants of residual units.

According to Figure 2.3, different variants of residual unit were proposed, which consists of different combinations of convolutional layers, batch normalization (BN) [57], and rectified linear unit [61] activation function [55, 62]. In our experiment, we use the full pre-activation variant of residual unit proposed by He et al. [55, 62] to build our architectures, which use ResNet as their feature encoder. ResNeXt [63] proposed by Saining Xie in 2017 is a highly modularized version of ResNet architecture based on split transform aggregate strategy as an inception module for image classification. Its innovative, simple design results in homogeneous, multi-branch architecture that has only a few hyper-parameters to set. This approach exposes a new dimension

called cardinality, the size of the set of transformations, as an essential factor in addition to other critical factors such as depth and width. The network is constructed by stacking repeating building blocks that aggregate a set of transformations with the same topology. Inspired by a residual network, several modifications, new designs, and architectures were proposed for different image understanding tasks [55, 64-66]. For instance, Figure 2.4 illustrates an inception-ResNet block called Inception ResNet-A module of the Inception ResNet-v2 network [59]. Other variants of inception-ResNet blocks including Inception ResNet-B and Inception ResNet-C modules were also proposed by Szegedy et al. in 2017 [59].

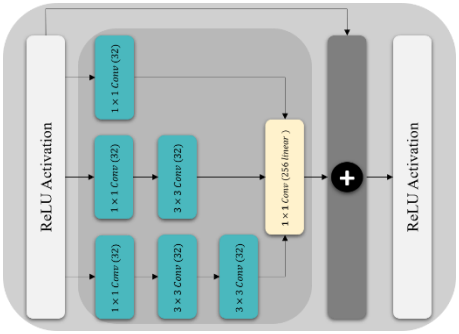


Figure 2.4. Inception-ResNet block.

2.2.1.4. DenseNet

Inspired by ResNet and the idea that shorter connections between layers close to the input and those close to the output can help to train substantially deeper CNNs more accurately and efficiently, Huang et al. proposed DenseNet [56] in 2017. The architecture consists of densely connected CNN blocks in which the output feature maps of each layer are concatenated with the output feature maps of all successor layers in a dense block as shown in Figure 2.5. If l-th layer receives all the feature maps from all preceding layers, x_0, x_1, ..., x_{l-1}, as input then:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (3)$$

where $[x_0, x_1, \dots, x_{l-1}]$ represents simple concatenation of feature maps produced in layers $0, 1, \dots, l-1$ and H_l is defined as a composite function of three consecutive operations including BN, followed by a *ReLU* and a 3×3 convolution. A transition layer composed of a a batch normalization layer and a 1×1 convolution followed by a 2×2 pooling operation is introduced between two consecutive dense blocks to reduce the dimensionality and spatial resolution of derived feature maps. DenseNet architecture consists of several densely connected blocks and transitional blocks, which are placed between two adjacent densely connected blocks. DenseNet concept alleviates the vanishing gradient problem, encourages feature propagation and feature reuse while substantially reducing network parameters.

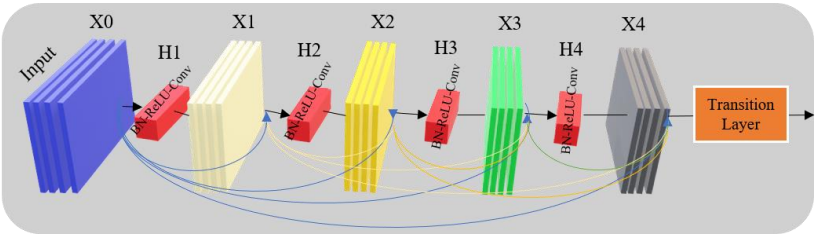


Figure 2.5. Illustration of a 5-layer sense block with a growth rate of $k = 4$.

2.2.1.5. MobileNet

Since the advancement of deep learning, the general trend has been to make deeper and more complicated networks to improve model performance [53, 58, 59]. However, these advances to improve accuracy are not necessarily making networks more efficient with respect to size and speed. In many real-world applications such as self-driving car, robotics, and augmented reality, the timely-fashioned or almost real-time prediction and recognition tasks need to be carried out on a computationally limited platform.

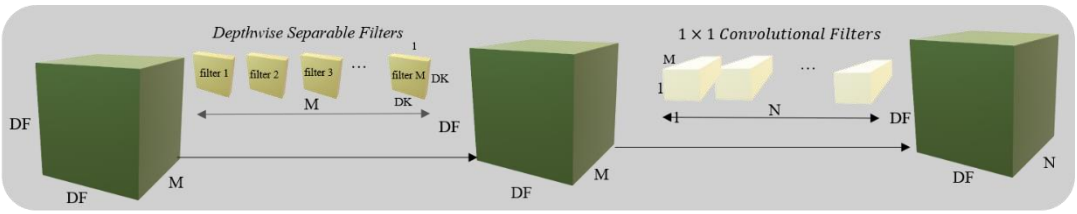
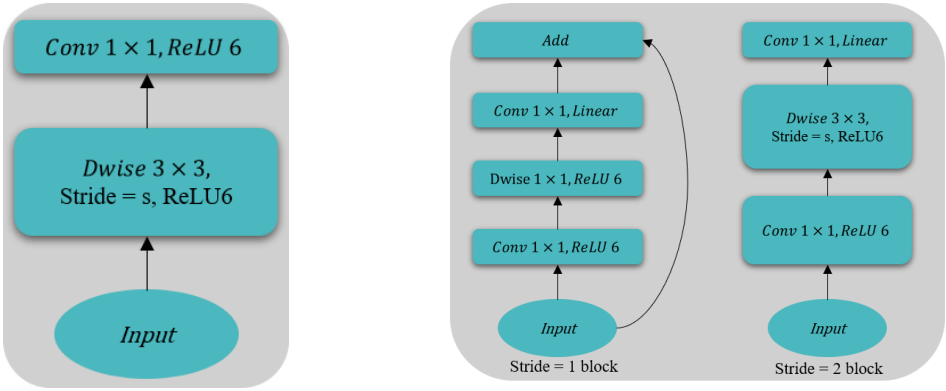


Figure 2.6. Depthwise separable convolution concept.



(a) (b)

Figure 2.7. MobileNet architecture modules.

Inspired by depth-wise separable convolutions to reduce the computation in the first few layers, a class of efficient models, called MobileNets [67, 68], for mobile and embedded vision applications was introduced by Howard et al. in 2017. This class of models presents a streamlined-base architecture that uses depth-wise separable convolutions to build lightweight deep neural networks. According to Figure 2.6, the depth-wise separable convolution is a form of factorized convolutions factorizing a standard convolution into a depth-wise convolution, which applies a single filter to each input channel, and a 1×1 convolution called a pointwise convolution to change the dimensions and linearly combine the output feature maps from depth-wise convolutions. The depth-wise separable convolution technique results in a drastic reduction in computation complexity and model size. Figure 2.7 illustrates two variants of MobileNet

architectures. According to Figure 2.7, in MobileNetV1 [67], there are two layers including depth-wise and pointwise convolutions. M and N are the number of input and output channels, respectively, and D_F and D_K are the sizes of feature maps and filter size, respectively. BN and *ReLU* activation function are both applied after convolutional layers. MobileNet introduces two hyper-parameters to the network including width multiplier, $\alpha \in (0,1]$, to control the input width of a convolutional layer and resolution multiplier $\rho \in (0,1]$, to control the input image resolution of the network. $\alpha = 1$ and $\rho = 1$ are hyper-parameters for the baseline MobileNets and $\alpha < 1$ and $\rho < 1$ are considered for any reduced computation MobileNets. Computational cost and the number of parameters are reduced by roughly α^2. However, the accuracy drops off as α and ρ decrease.

MobileNetV2 [68] is a significant improvement over MobileNetV1 with high potential of reaching the state-of-the-art performance for mobile visual recognition tasks. It was also built upon the idea of depth-wise separable convolution already applied in MobileNetV1 as efficient building blocks. In MobileNetV2, there are two types of blocks. One block is a residual block with stride of 1 and a second block with stride of 2 for downsampling. Both blocks include three layers. The first layer of each block in MobileNetV2 includes a 1×1 convolution with *ReLU* activation function. The second layer is a depth-wise convolution, and the third layer is another 1×1 convolution but without any activation function.

2.2.2. Decoding approaches

As explained earlier, an encoder is simply a deep learning architecture such as VGG-Net, GoogLeNet, ResNet, etc., making a hierarchical representation of input data. The final feature maps derived from encoders are usually coarse representations of the input image which needs to be upsampled to higher resolution feature maps. Decoding, on the other hand, is a strategy that

aims to efficiently exploit encoded feature maps provided by the encoder to form an output that is the closest match to the intended output, usually corresponding ground truth.

Deconvolution or transposed convolution [46, 69] is conceptually required in deep CNN architectures for pixel-wise predictions as feature maps are continuously down-scaled within several convolution and pooling layers. As mentioned earlier, FCN architecture enables upsampled feature maps with resolution comparable to the input image through a fractionally-strided convolution step in combination with a simple bilinear interpolation. However, due to the lack of an efficient trainable deep deconvolution network, FCN fails to achieve the high accuracy in pixel-wise labeling, especially when it is required to reconstruct highly nonlinear structures of object boundaries [70].

The deconvolution network was first discussed for image reconstruction from its feature representation by Zeiler et al. [47]. To resolve ambiguity induced by Max-pooling layers, the network stored the pooled locations, which need to be retrieved in an unpooling operation. To predict pixel-wise segmentation map, in 2015, Noh et al. proposed a trainable deep deconvolution network composed of deconvolution and unpooling layers [70]. SegNet [71] designed by Badrinarayanan et al. in 2015 consists of a deep encoder network and a hierarchy of decoders---one corresponding to each encoder followed by a pixel-wise classification layer. Appropriate decoders are fed by Max-pooling indices computed in the pooling steps of the corresponding encoder to perform deconvolution with nonlinear upsampling of their input feature maps. To produce dense feature maps in the decoder, the resulting sparse upsampled feature maps are, then, convolved with trainable filters. U-Net [72] developed in 2015 is an innovative deep learning architecture first developed for biomedical image segmentation by Ronneberger et al. and was then extensively used for image segmentation in many other fields with different encoders such as

ResNet, DenseNet, and Inception modules. The network has a symmetrical architecture characterized by an encoder with a series of convolution and Max-pooling layers in the contracting path and a decoder containing a mirrored sequence of convolution and upsampling layers in the expanding path of the network. U-Net is able to concatenate low level abstract information, extracted from the first convolutional layers of the encoder (contracting path) and high-level semantic abstraction information, extracted from the final layers of encoder, in the decoder (expanding path), resulting in a finer and more accurate prediction map. This strategy resulted in high performance especially when only a limited training dataset is available [72]. Motivated by a Laplacian pyramid developed for compact image coding [73], in 2016, Ghiasi et al. proposed a network called Laplacian Pyramid Reconstruction (LRR) in which low-resolution feature maps are used to reconstruct a low-frequency segmentation map. Feature maps are, then, refined by adding high-frequency details. Refinement network (RefineNet) [74], proposed by Lin et al. in 2017, is a generic multi-path network which explicitly exploits all available information along the downsampling path to enable high-resolution image labeling using long-range residual connections. This network consists of three main components: Residual convolution unit (RCU), which exploits features at multiple scales, multi-resolution fusion, which merge multi-resolution features, and chained residual pooling, which aims to capture background context from a large image region by fusing the output feature maps of all pooling blocks together with the input feature map.

Inspired by DenseNet, in 2017, Jegou et al. proposed a One Hundred Layers Tiramisu network, commonly called Fully Convolutional DenseNet (FC-DenseNet) [75]. The architecture extends the DenseNet to a fully convolutional network for a semantic segmentation task. The upsampling path includes convolution, upsampling operations called transition up, and skip

connections. Transition up modules consist of a transposed convolution to upsample the previous feature maps. Upsampled feature maps are then concatenated with corresponding feature maps in the downsampling path using skip connections to prepare the input for the next upsampling dense block. To mitigate the parameter explosion problem, the input of a dense block is not concatenated with its output in the upsampling path e.g., transposed convolution is applied only on feature maps derived by the last dense block instead of the concatenation of all derived feature maps so far.

Other innovative techniques were also proposed for dense semantic segmentation, which, unlike the convolution/deconvolution design, do not introduce new parameters to upsample feature maps. Atrous convolution [76, 77], usually called dilated convolution, originally developed for computing undecimated wavelet transform (UWT) [78] is employed to effectively enlarge the field of view of feature maps without increasing the number of parameters or computation complexity. Atrous or dilated convolution in the context of CNNs aims for expanding the receptive field of the network. They generate high-resolution feature maps capturing multi-scale contextual information from the input data. Dilated convolution introduces a new hyper-parameter called dilation rate to the convolution layers, which specifies the expansion rate of receptive field enabling the network to exploit a larger receptive field without losing spatial information.

In 2014, DeepLab [76], introduced by Chen et al. from Google, proposes atrous convolution instead of deconvolution for feature upsampling. Atrous convolution offers an efficient mechanism to control the receptive field of the network and finds the best trade-off between precise localization, with the small receptive field, and context assimilation, with the large receptive field. The output of the network is interpolated, with bilinear interpolation, and goes through the fully connected conditional random fields (CRF), which fine-tune the result for a more accurate and detailed segmentation map. Different variants of DeepLab architecture were later

proposed with some modification on the original network. Atrous Spatial Pyramid Pooling (ASPP) was proposed in DeepLabV2 [31] to robustly segment objects at multiple scales. ASPP probes incoming feature maps at multiple sampling rates and field-of-views capturing objects and image context in multiple scales. In DeepLabV3 [79], to handle the problem of multi-scale object segmentation, a cascade or parallel atrous convolution design is employed to capture multi-scale context by adopting multiple dilation rates. DeepLabV3 outperformed its predecessors without dense CRF post-processing and attained comparable performance with other state-of-the-art models. Authors in DeepLabV3+ [79] decided to add a decoder module to the former variant in which the encoded features are first upsampled by a factor of 4, instead of 16 as in [80], and then the resulting feature maps were concatenated with corresponding mid-level features from the network backbone. Moreover, to reduce computational complexity, they adopted the Xception module [60] and applied depth-wise separable convolution to both the ASPP and decoder.

Yu et al. [77] developed a deep learning architecture in 2015 specifically designed for dense prediction based on dilated convolution concept. This convolutional network module combines multi-scale contextual information without losing spatial resolution. Pyramid scene parsing network (PSPNet) [81] introduced in 2017 exploits the capability of global context information by different region-based context aggregation methods by employing a pyramid pooling module in combination with the proposed pyramid scene parsing network. To do pixel-wise prediction, PSPNet extends pixel-level feature to a specially designed global pyramid pooling one. Then, the local and global clues jointly form the final prediction.

2.2.3. Transfer learning

The idea of transfer learning was motivated by the fact that people can intelligently apply knowledge previously learned to solve a task in one domain to solve a new problem in the same

or different domain [82]. In the deep learning context, features learned by a CNN architecture to solve a problem in a certain domain are reusable for solving problems in some other domains, as the first layers of the network in related domains usually tend to learn the same sorts of features. Transfer learning is a highly practical approach to tackle the issue of training a very deep architecture where a limited supply of target training data is available. This could be due to data scarcity, or methods to collect and label the data may be time consuming and expensive requiring expert knowledge. In contrast to many computer vision tasks that can take advantage of thousands of freely available images related to the underlying task, in most RS applications, e.g., land cover mapping, satellite or aerial imagery missions can be very expensive or time consuming. Data collections are a function of many participating factors including flight height, ground sampling distance (GSD), environmental conditions at the time of observation, and camera/sensor settings. Furthermore, a limited number of aerial images are acquired in every flight mission and the acquired images are not always available to the public to enable generation of large, labeled data repositories for a specific type of environment or land cover. UAS provides a cost effective and flexible means to collect high-resolution aerial imagery over localized geographic extents; however, dense repositories of UAS imagery acquired over a specific type of natural environment that is expertly labeled for training deep CNNs to perform land cover prediction are presently non-existent.

Common practice in transfer learning is to copy the whole or just the first n layers of a pre-trained network, already trained on a huge dataset, to exploit them in a new task and then back-propagate the errors from the new task into the copied features to fine-tune them to the new task. In another approach, especially where the training sample size is significantly limited or the new task is closely related to the task from which a transferred feature is derived, the first n feature

layers can be left frozen, meaning that they do not change during training on the new task. The choice of whether or not to fine-tune the copied first n feature layers depends on the size of the available dataset for the new target task. In a case where the target dataset is small, fine-tuning may lead to overfitting, especially when the network contains a large number of parameters. On the other hand, if the target dataset is rich enough or the number of network's parameters is small, where overfitting does not seem to be a problem, then fine-tuning copied features to the new task can highly improve the performance [35]. In such case, training the network from scratch may also be taken as an option.

2.2.4. Performance metrics

This section describes the most common performance or evaluation metrics used in the context of semantic image segmentation. Usually, overall performance of a deep learning architecture in semantic image segmentation task is described in terms of overall accuracy of pixel-wise labelling, time, and memory usage. Overall accuracy of a network is a measure which usually describes the correctness of labelling as a simple ratio representing the number of correctly classified pixels over the total number of manually classified pixels in the ground truth. Pixel-wise or per-class accuracy is another measure that usually aims to report the percent of correctly classified pixels for each individual class. Pixel-wise accuracy is closely related to overall accuracy. In fact, binary mask employed in pixel-wise accuracy assessment may return quantities more than just true positive (TP), which represents the number of correctly labeled pixels, and true negative (TN), which represents the number of pixels that are correctly identified as not belonging to a certain class. False positive (FP) represents the number of pixels belonging to other classes misclassified as the target class, and false negative (FN) represents the number of pixels that belong to the target class but are misclassified as belonging to other classes. They are two of the most

important quantities for which the binary mask may be designed to account. Accordingly, the overall accuracy per-class can be formulated as [83]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Pixel-wise accuracy metric is not reliable and may provide misleading results when a certain class representation is small within the whole dataset. Precision and recall are two metrics that can help to interpret the overall accuracy of each class more accurately even in the case of unbalanced classes. Precision or positive predictive value (PPV) describes the purity of positive detection procedure relative to all pixels that have already been truly classified in the ground truth [83]:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Recall, or true positive value (TPV), on the other hand, effectively describes the completeness of the positive predictions relative to all pixels that have already been truly classified in the ground truth [83]:

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

The F-score is a widely used performance metric for classification and segmentation tasks, which consists of the harmonic mean of precision and recall metrics [83]:

$$F - \beta = \frac{(\beta^2 + 1)TP}{(\beta^2 + 1)\,TP + \beta^2 FN + FP} \tag{7}$$

where β is a scaling factor between the precision and recall. F1-score, one of the more widely used F-measure metrics is formulated by setting $\beta = 1$ [83]:

$$F1 = \frac{2 \times TP}{2TP + FN + FP} \tag{8}$$

Intersection over Union (IoU), also known as Jaccard index, is a standard performance measure for the object category segmentation. IoU measure represents the similarity ratio between the predicted region and the corresponding ground truth region for an object presented in the dataset [84]:

$$IoU = \frac{TP}{TP + TP + FN} \tag{9}$$

Mean Intersection over Union (mIoU) is a common performance metric for semantic segmentation that is calculated by averaging over all IoU values computed for all existing semantic classes. Other performance metrics, such as time, memory, and power, are highly dependent on the available hardware, software, and the specific deep learning architecture chosen for solving a classification task. Providing such metrics becomes more crucial when a deep learning framework is employed in online applications such as autonomous driving and mobile systems where the memory and power is more limited.

2.3. Materials and Methods

2.3.1. Study site

The study site is a coastal marsh located on a barrier island along the southern portion of the Texas Gulf Coast, USA, bounded by Corpus Christi Bay, the Laguna Madre, and the Gulf of Mexico called the Mustang Island Wetland Observatory as shown in Figure 2.8. The study area as imaged by the UAS is 11 hectares. Elevation within the wetland slopes gradually and is nearly flat, with the highest elevation in the study area at about $0.8\ m$ (NAVD88). The wetland is located on the bay side of the island Figure 2.8 and is oriented in a northeast to southwest trend, with the Gulf of Mexico located to the east and Corpus Christi Bay to the west. The dominant vegetation species are *Schizachyrium littorale* (*Nash*) (*coastal bluestem*) and *Spartina patens* (*Aiton*) (*gulf*

cordgrass) commonly found growing in mats. The second most prevalent environment of this study area is tidal flat; it ranges in elevation from $-0.05\,m$ to $0.5\,m$ (NAVD88) [85]. Low regularly flooded tidal/algal flats are significantly less abundant than high flats in this area. These local tidal flats are designated as wind-tidal flats because flooding occurs mainly due to wind-driven tides [86, 87]. Blue-green algae can be prevalent in the lower portion of the tidal flats after long periods of inundation. Furthermore, salt marsh vegetation can be found sparingly in portions of the tidal flat areas. Low marsh areas are very high in biologic productivity usually ranging in elevation from $-0.1-0.3\,m$ (NAVD88). More frequently inundated areas near tidal creeks are dominated by taller vegetation, primarily *Avicennia germinans* (*black mangrove*). High marsh environment is the least abundant in the study area imaged by the UAS. It varies in range from approximately $0.2-0.8\,m$ (NAVD88) well above the mean high tide; therefore, it is rarely inundated. These characteristics briefly illustrate the highly dynamic and complex nature of the coastal wetland and the need for applying accurate algorithms for detailed land cover mapping through analyzing UAS hyper-spatial imagery.

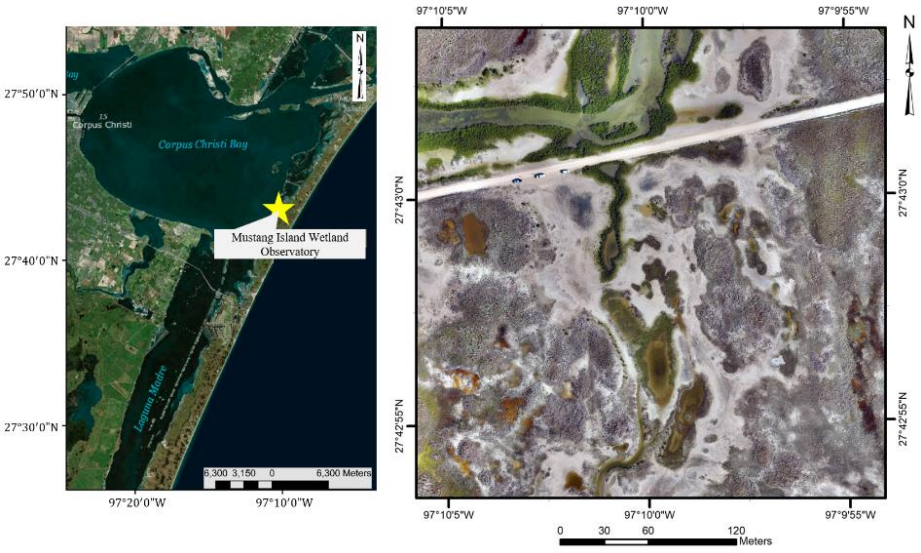


Figure 2.8. Mustang Island Wetland Observatory study site location (Left); UAS orthoimage of the study area showing the dirt road, exposed tidal flats, water bodies, and surrounding vegetated land cover (Right).

2.3.2. Data collection and preparation

Phantom 3 multi-rotor UAS, manufactured by Shenzhen DJI Sciences and Technologies Ltd (SZ DJI Technology Co., Ltd.) headquartered in Shenzhen, Guangdong province, China, was employed to collect required images for this study. This platform is equipped with a CMOS RGB sensor to capture 12 megapixel images with a resolution of 4000×3000 pixels. The flight was designed at an altitude of 90 m above the ground resulting in an average GSD of around 3 cm. Imagery was collected at 80% sidelap and endlap flown in a grid pattern with parallel flight lines and a 90° (nadir) camera orientation. This high amount of overlap was used to perform Structure-from-Motion (SfM) photogrammetry processing and orthorectify the imagery to remove perspective and relief distortion and generate a large orthoimage that covers the study area. The performance and visual quality of land cover prediction using different deep CNN models is evaluated on a certain part of the study area that most original images belonging to that area are kept for validation purpose. Because in RS applications, land cover is usually predicted on orthorectified images, the visual quality of land cover prediction is illustrated on an orthoimage mosaic of validation images. The reader is referred to [88] for more details on SfM photogrammetry.

In this work, 300 images were manually selected from the total set of acquired UAS images (about 500 images) that cover the whole study area to reduce repetitive information from image overlap. Due to the high resolution of the original imagery, the image set can rapidly exhaust the

whole GPU's memory when directly fed to any deep convolutional network. Therefore, we randomly extract 10,000 image patches of resolution 512×512 pixels from the set of 300 raw images. From those image patches, 1000 image patches are held as a validation data set for evaluating the model performance after each training epoch. Every image, at most, represents four classes: vegetation, water bodies, tidal flat, and road. In our experiment, tidal flat is assigned to surfaces exposed within intertidal areas. All temporarily flooded areas or permanently submerged lands are considered water bodies. Areas covered by any type of vegetation is called vegetated area. Finally, road represents the artificially elevated dirt surface of exposed ground that has not been affected by tides. The different land cover classes can be observed in the orthoimage mosaic displayed in Figure 2.8, which was generated from all UAS images acquired over the study area using the SfM photogrammetry software. All needed ground truth data for training and validation were manually prepared through supervised labeling by interpretation and delineation of land cover boundaries in the image patches. This was done by color labeling of all existing pixels in each original image patch to a representative class using a labeling app in MATLAB software for pixel-level image labeling. According to our predefined color for each target, pixels belonging to vegetation, tidal flat, water, and road are represented by green, orange, blue, and brown, respectively. It should also be mentioned that a set of 64 raw UAS images from a portion of the study site that had a representative distribution of the land cover classes were set aside for independent evaluation of model performance results as presented below in Section 4. These images, or patches extracted from them, were not used as part of the training set described above.

2.3.3. Deep learning architectures

This subsection introduces the deep learning architectures evaluated in this study for performing pixel-wise image segmentation task (i.e., land cover mapping) with UAS hyper-spatial

imagery acquired over a complex coastal wetland environment. The chosen architectures are extensively used in a wide range of applications beyond RS including computer vision and medical image processing.

2.3.3.1. Encoder-Decoder (SegNet)

SegNet architecture, displayed in Figure 2.9, is examined in this study, which is a relatively old deep learning network for semantic image segmentation task. It uses VGG network as its encoder to hierarchically extract features from input images. The encoder network consists of 13 convolutional layers corresponding to the first 13 convolutional layers in the VGG-16 network. In our experiment, we use weights from pre-trained VGG-16 network to initialize the training process. Each encoder layer has a corresponding decoder layer that upsamples the feature maps by using the stored pooled indices.

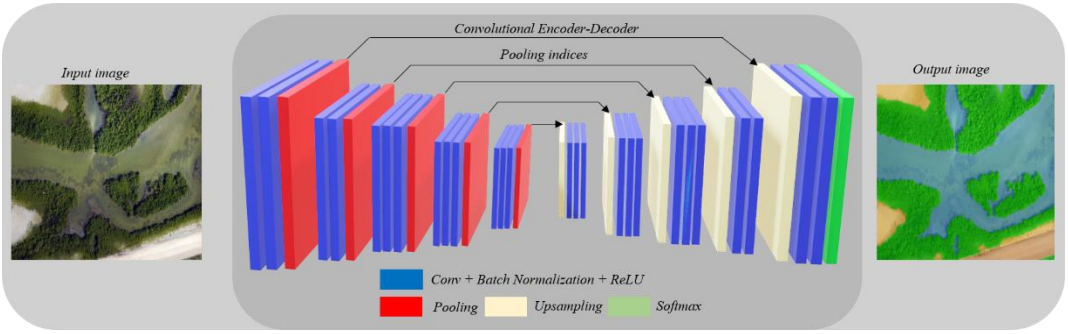


Figure 2.9. An illustration of the Encode-Decoder (SegNet) architecture.

2.3.3.2. U-Net

U-Net is a famous deep architecture based on an encoder--decoder principle that instead of using pooling indices, it transfers and exploits the entire feature maps from encoder to decoder. Upsampling strategy can have a great impact on the final accuracy of pixel-wise image classification. Comparing the performance of SegNet and U-Net architecture can tell us more

about the effectiveness of those two upsampling strategies. Figure 2.10 illustrates U-Net architecture with ResNet-34 network for feature extraction in this study.

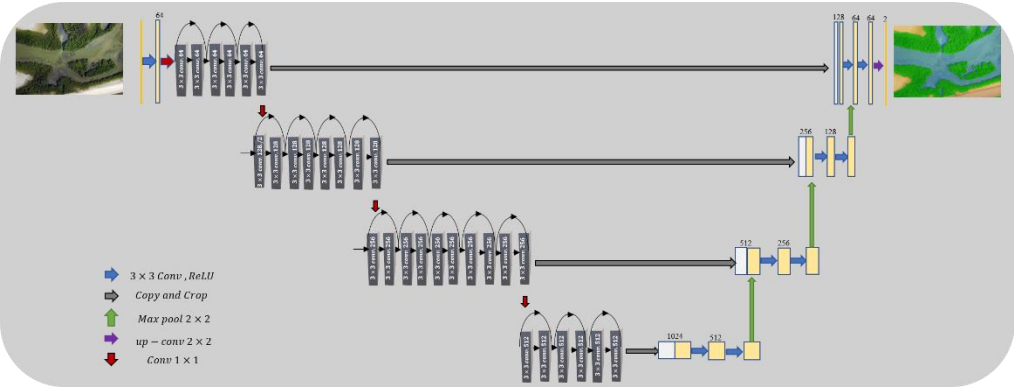


Figure 2.10. An illustration of U-Net architecture with ResNet34 as encoder.

2.3.3.3. FC.DenseNet

To explore the efficiency of DensNet architecture in feature learning for pixel-wise classification of coastal wetland images, the one hundred layer tiramisu model (FC-DenseNet), as shown in Figure 2.11, is employed which uses 56 convolutional layers, with four layers per dense block and a growth rate of 12. Similar to U-Net architecture, FC-DenseNet exploits U-shape encoder-decoder structure with skip connections between the downsampling and the upsampling paths to add higher resolution information to the final feature map. Unique characteristics of feature reuse, compactness, and substantially reduced number of parameters in FC-DenseNet architecture is evaluated in our experiment based on its performance when training the network from scratch using a limited dataset, which is the case here.

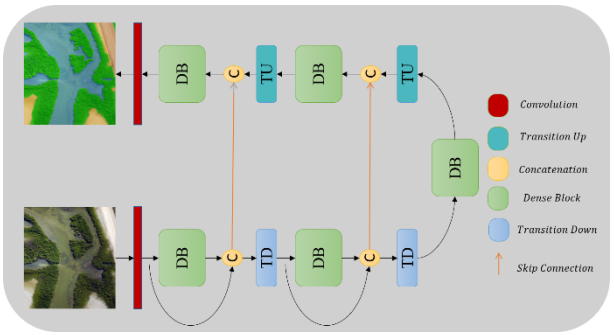


Figure 2.11. An illustration of FC-DenseNet architecture.

2.3.3.4. DeepLabV3+

Effectiveness of ASPP to encode multi-scale contextual information in images acquired over complex coastal wetland is investigated by examining DeepLabV3+ architecture illustrated in Figure 2.12. This architecture is able to perform several parallel atrus convolution with different rates.

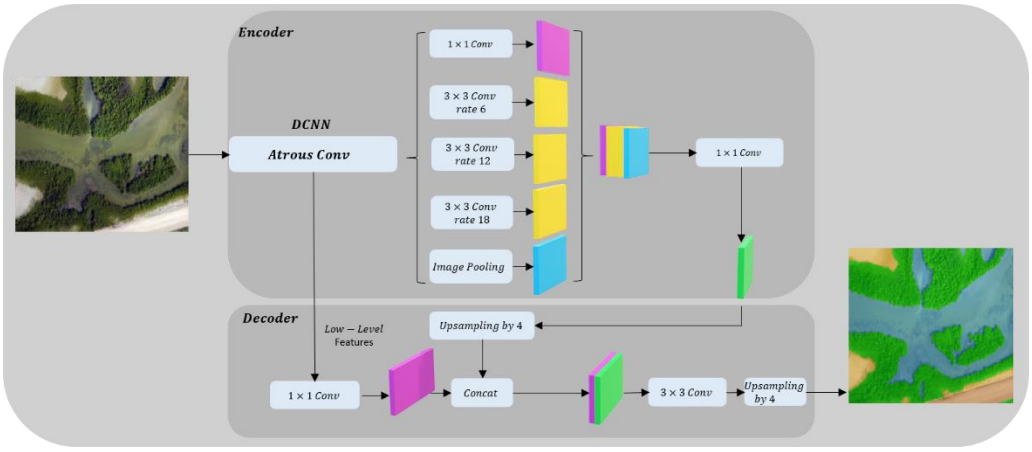


Figure 2.12. An illustration of DeepLab V3+ architecture.

2.3.3.4. PSPNet

As illustrated in Figure 2.13, PSPNet, which uses pyramid pooling module for more reliable prediction, is also investigated for this study. Specifically, this module is able to extract global context information through aggregating different regional context information.

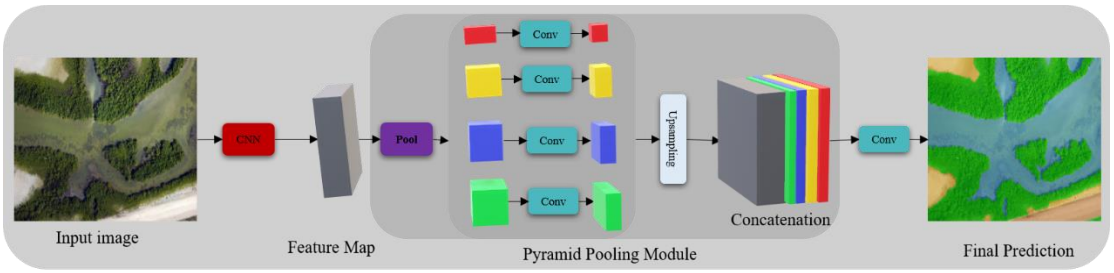


Figure 2.13. An illustration of PSPNet architecture.

2.3.3.6. MobileU-Net

Considering the idea of depth-wise separable convolution in MobileNet and feature map upsampling in U-Net architecture, MobileU-Net architecture, illustrated in Figure 2.14, is implemented in this study. The performance of this architecture in pixel-wise image labeling of hyper-spatial UAS images may give us an estimation of the accuracy achievement in real-time land cover mapping.

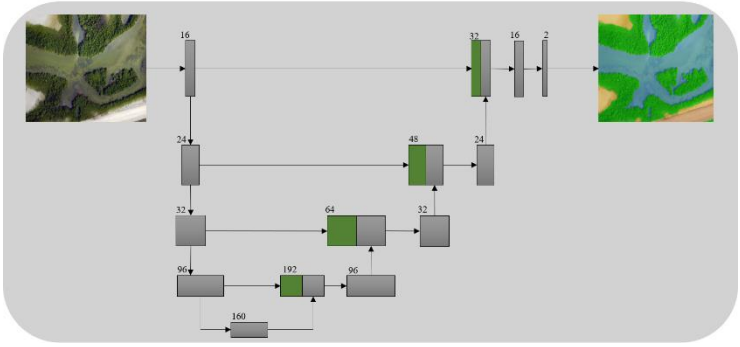


Figure 2.14. An illustration of MobileU-Net architecture.

In our experiment, we use a pre-trained ResNet-34 network as a feature encoder in all employed architectures excluding Encoder-Decoder (SegNet) and FC-DenseNet architectures. To predict each image pixel's category, all employed deep architectures include a multi-class softmax classifier on top, which is fed by the output upsampled feature map from the final layer of the network to produce pixel-wise class probabilities. Cross-entropy and Adam optimizer [89] are selected as the loss function and optimization algorithm, respectively. Adam optimizer computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients [89] and realizes the benefits of both AdaGrad [90] and RMSProp [91]. It includes several parameters that need to be carefully set. Popular deep learning libraries generally use the default parameters recommended by the paper including learning rate parameter $\alpha = 0.001$, two exponential decay rate parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and $\epsilon = 1 \times 10^{-8}$, which prevents any division by zero in the implementation. In our experiment, we set all optimization parameters according to those recommended values.

Weight initialization is carried out for all employed networks. Except for FC-DenseNet, weight parameters in other networks are initialized by transfer learning. For FC-DenseNet, we decided to train the network from scratch since we did not find a pre-trained FC-DenseNet on large datasets such as ImageNet. FC-DenseNet has very few parameters, about 10 times less than recent state-of-the-art models; thus, it is worth it to train this network from scratch and compare its performance over our limited dataset with the performance of pre-trained encoders. All deep CNN models in this experiment were trained using the same training samples under the same conditions for 200 epochs.

All experiments were carried out on Amazon Web Service (AWS) with one high-performance NVIDIA K80 GPU, with 2496 parallel processing cores and 12 GB of GPU memory and high frequency Intel Xeon E5-2686 v4 processors under CUDA version 10.0.

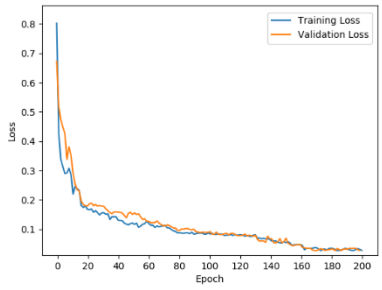
2.4. Results

Figure 2.15 illustrates the training and validation losses for all employed deep CNN architectures trained under the same training dataset. The validation loss curves closely follow corresponding training loss curves showing the ability of the deep CNN models in generalization. Normalized confusion matrices in Figure 2.16 display the performance of the deep CNN models on each individual land cover target.

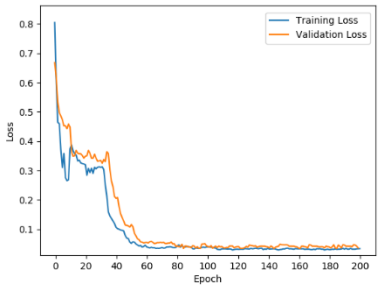
Table 2.1 illustrates the land cover prediction results achieved for the different deep CNN architectures employed in the image segmentation experiment. The first two columns represent overall accuracy for training (OA-Tr.) and validation (OA-Val.). Precision (Prec.), Recall (Rec.), F1-score (F1), and mIoU are included for evaluating the performance of each architecture as these are some of the most widely used metrics.

Figure 2.17 displays a cropped orthoimage from the upper portion of the study area and its corresponding ground truth labels. This area was selected for model validation purposes because it provides a nice distribution of the different land cover classes. Images from this area were not included in the training samples. Figure 2.17 stems from a set of 64 overlapping UAS images that were orthorectified and mosaicked together as part of the SfM photogrammetric processing used to create the full study area orthoimage (Figure 2.8). To classify the orthoimage, the full image is not fed directly into the model due to its large size. Small image patches (512×512 pixels) are extracted and then fed into the model to undergo pixel-wise labeling. After the land cover class(es) contained in each individual image patch are predicted by the model, they are then reassembled to

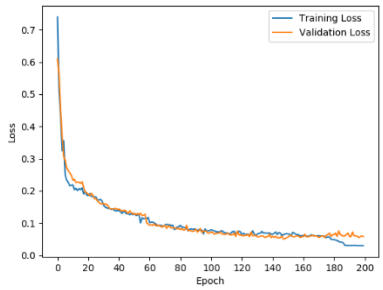
generate the full resolution image. The land cover maps predicted for this orthoimage, using all employed deep CNN models in this study, are displayed in Figure 2.18 (a-f). Interestingly, land cover classes predicted by all employed CNN models closely resemble the ground truth image in Figure 17. However, FC-DenseNet, UNet, and DeepLabV3+ are the most accurate representations of the ground targets in this complex wetland environment.



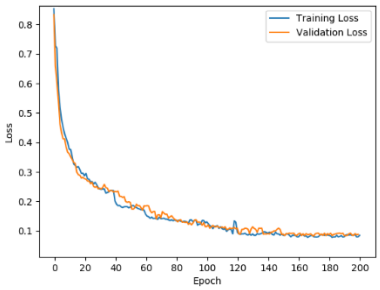
(a) FC-DenseNet.



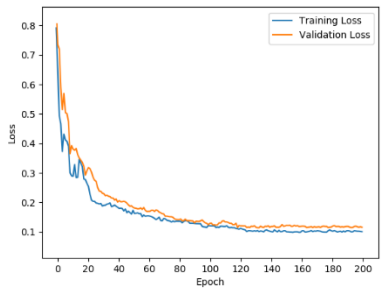
(b) U-Net.



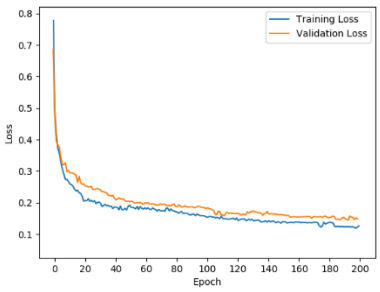
(c) DeepLab V3+.



(d) PSPNet.

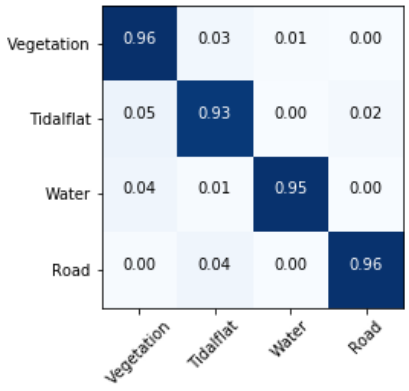


(e) MobileU-Net.

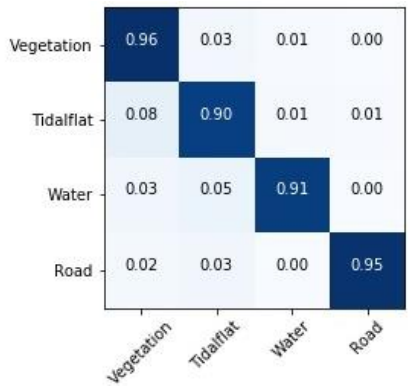


(f) SegNet.

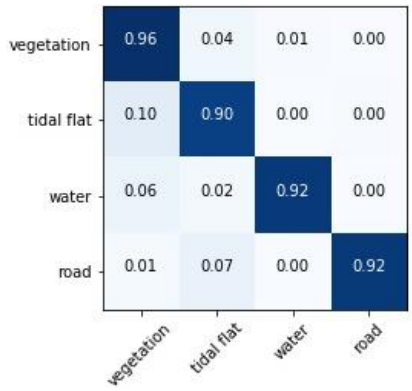
Figure 2.15. Average loss per epoch for training and validation steps.



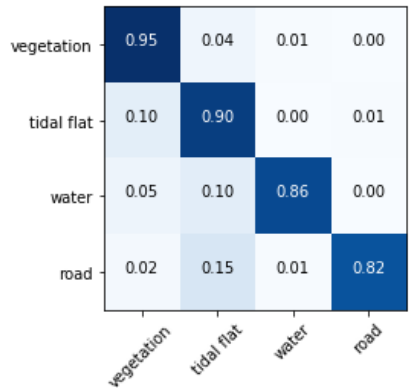
(a) FC-DenseNet.



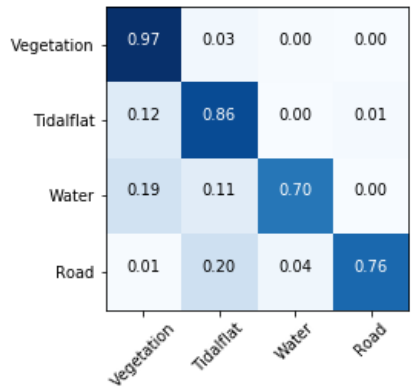
(b) U-Net.



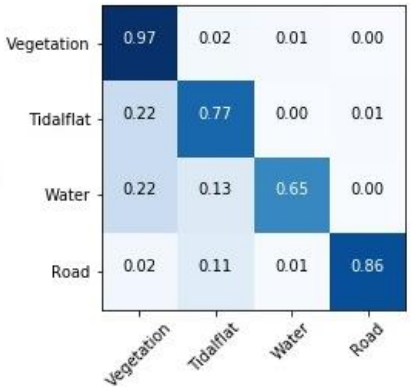
(c) DeepLab V3+.



(d) PSPNet.



(e) MobileU-Net.

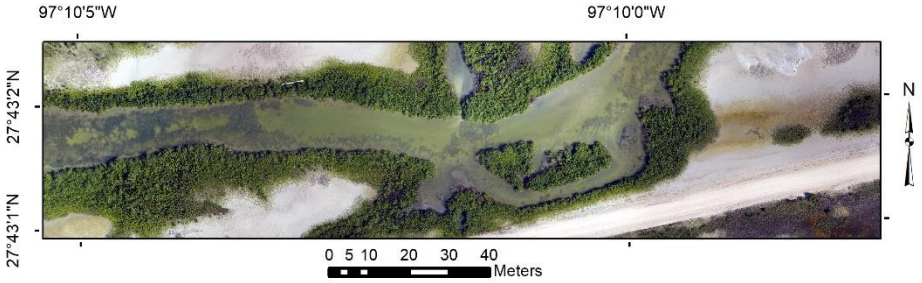


(f) SegNet.

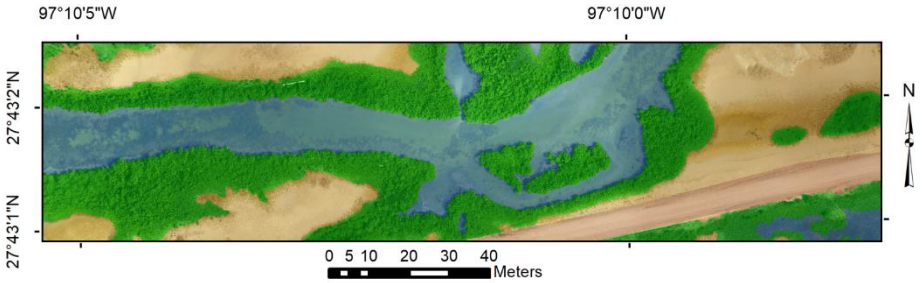
Figure 2.16. Normalized confusion matrices for the coastal wetland land cover prediction task using different deep CNN architectures.

Table 2.1. Coastal wetland land cover classification results.

Model	OA-Tr.	OA-Val.	Prec.	Rec.	F1	mIOU	Vegetation	Tidal flat	Water	Road
FC-DenseNet	0.97	0.95	0.95	0.95	0.95	0.90	0.96	0.93	0.94	0.96
U-Net	0.96	0.95	0.95	0.94	0.94	0.91	0.95	0.93	0.95	0.95
DeepLab V3+	0.94	0.93	0.91	0.90	0.90	0.89	0.94	0.88	0.87	0.89
PSPNet	0.91	0.89	0.89	0.88	0.88	0.83	0.96	0.89	0.87	0.83
MobileU-Net	0.89	0.85	0.88	0.79	0.84	0.75	0.97	0.85	0.69	0.76
SegNet	0.88	0.82	0.91	0.81	0.82	0.69	0.97	0.77	0.65	0.85

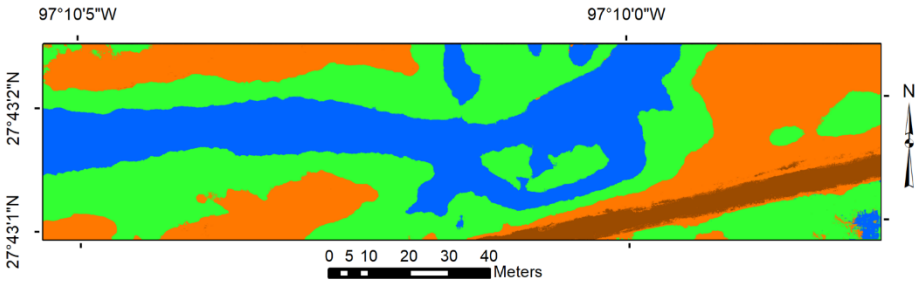


(a) Orthoimage

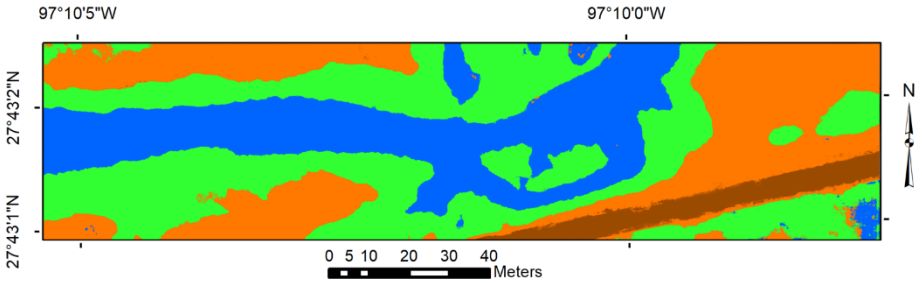


(b) ground truth

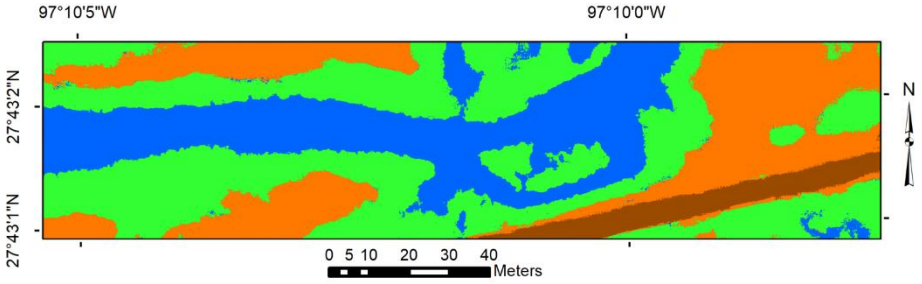
Figure 2.17. Original orthoimage generated by mosaicking 64 ortho-rectified UAS images over the wetland study site and related ground truth image.



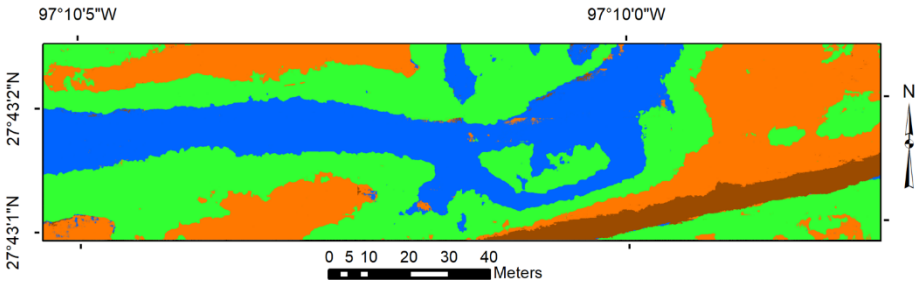
(a) FC-DenseNet.



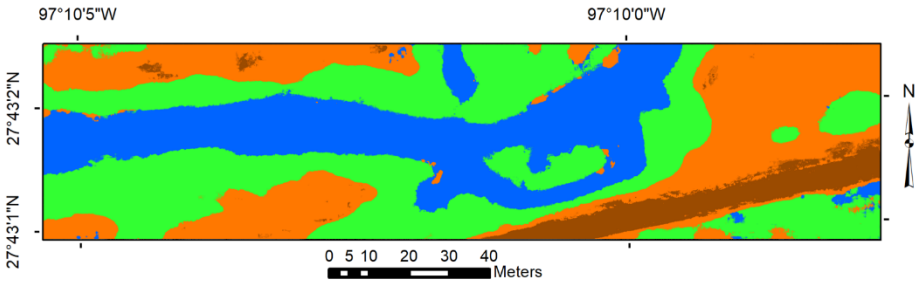
(b) U-Net.



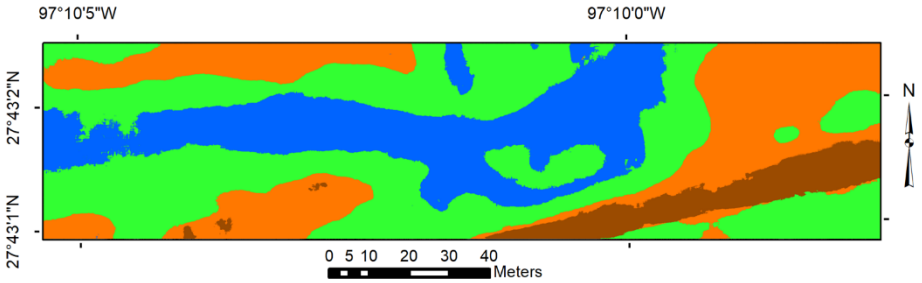
(c) DeepLabV3+.



(d) PSPNet.



(e) MobileU-Net.



(f) SegNet.

Figure 2.18. Land cover map prediction over prepared orthoimage for part of the coastal wetland test area.

2.5. Discussion

Referring to Figure 2.15, FC-DenseNet, U-Net, and DeepLabV3+ show lower loss values for both training and validation losses w.r.t MobileU-Net, PSPNet, and SegNet models, resulting in higher training and validation accuracies according to Table 2.1. For the SegNet model, the validation losses keep a certain distance above the training losses explaining the larger difference between validation and training accuracies reported for this model. Furthermore, still referring to Figure 2.15, U-Net is showing a higher speed of convergence during the training phase. This suggests that the skip connections from encoder to decoder have a high contribution in smoothing the gradient descent's path towards the global minimum in the high-dimensional weight space. Additionally, in comparison to FC-DenseNet, the fine-tuning strategy of the transfer learning

technique employed by U-Net yielded reduced training epochs. This approach helps to exploit the advantages of deeper CNNs with a larger number of trainable parameters where the available training resources are limited (as is the case here due to manual labeling). FC-DenseNet also takes advantage of skip connections in its encoders, and between encoders and decoders, which helps with the flow and convergence of gradient descent through reuse of features. However, due to training the network from scratch, more training steps to converge is necessary.

The fine-tuning strategy of transfer learning yielded very good results in all models with pre-trained VGG-16 and ResNet-34 architectures as their encoder for feature learning. This is a promising result given that the structure of low-level and high-level natural/wetland terrain features in our dataset are noticeably different from those that appear in the ImageNet dataset used for training the deep CNN architectures. Furthermore, the overall accuracy achieved by training FC-DenseNet from scratch confirms that the dramatic reduction in the number of parameters of this architecture with respect to. other state-of-the-art deep learning architectures enables it to learn optimum features when presented with relatively limited training samples.

Regarding the F1 score and mIoU values depicted in Table 2.1, the first three CNN models exhibit the highest performance among the others. According to the confusion matrices displayed in Figure 2.16, three of the employed networks, FC-DenseNet, U-Net, and DeepLabV3+, were successful in predicting labels for pixels belonging to all existing classes with accuracy above 90%. Almost all deep networks were successful in predicting pixels belonging to the vegetation class with an accuracy greater than 95%. Compared to the other classes of targets, vegetation represents the least confused class. Referring to Figure 2.16, especially when SegNet, PSPNet, and MobileU-Net models were employed, road pixels are mostly confused with tidal flat pixels, and pixels belonging to water bodies are more likely to be misclassified as tidal flat and vegetation. It

should be noted that discriminating pixels belonging to the tidal flat class from those belonging to the road class at this study site is a difficult task. These two classes exhibit very high inter-class similarities due to the road being a dirt road comprised of similar sand material to that of the exposed ground areas within tidal flat areas but with some mixed gravel.

The comparable overall accuracy of the FC-DenseNet architecture trained from scratch to that of the U-Net and DeepLab V3+ architectures, which use fine-tuned encoders, illustrates that the compactness in the number of parameters of FC-DenseNet makes it a good choice among many recently developed CNN architectures for pixel-wise labeling for training from scratch under limited training samples. The high performance of the U-Net architecture, trained based on the transfer learning technique, provided the most accurate and efficient choice among the others for pixel-wise labeling. Its performance justifies that the employed transfer learning technique does very well when it is employed to learn hierarchical features in high-spatial resolution UAS or RS images over natural terrain like wetlands. Such image sets and features are significantly different from the features of standard image datasets, such as ImageNet [52]. High performance of the DeepLabV3+ architecture demonstrates the effectiveness of ASPP in this network, which is able to properly encode multi-scale contextual information of the coastal wetland land cover captured in the images. However, this network needs more training steps to reach comparable performance with respect to U-Net. Our experiment with PSPNet at the wetland study site shows that the pyramid pooling module together with the pyramid scene parsing network is more effective in predicting vegetation and tidal flat areas than water and road areas. MobileU-Net and SegNet achieved less accuracy among all employed architectures for semantic image segmentation. Results achieved by our MobileNet architecture is based on baseline settings for its hyper-parameters, which include $\alpha = 1$ for width multiplier and $\rho = 1$ for resolution multiplier.

Decreasing those two hyper-parameters can dramatically decrease the performance of the network. However, MobileNets have the potential to be employed effectively in some real-time RS applications. As mentioned earlier, MobileNets were built as small, low-latency, low-power models parameterized to meet the resource constraints of a variety of mobile and embedded vision applications. These type models require less computational power and capacity for near real-time applications compared to very deep architectures with a higher capacity for learning due to their larger number of parameters. SegNet, like other employed architectures in this experiment, performed very well in vegetation areas but was much less accurate in classifying other targets. It is suspected that SegNet's inefficiency for pixel-wise labeling of the other targets, which are more challenging, stems from the network's inefficiency for exploiting low-level and high-level abstract features throughout the network and in its inefficient upsampling method.

It is worth mentioning that the information needed for training any of the evaluated classification architectures was obtained through supervised labeling by interpretation and delineation of land cover boundaries in the UAS images. This interpretation includes labeling a relatively large number of images by a human operator. This may result in different types of errors in the labelling of land cover types, and most notably in those circumstances in which categories are very heterogeneous and the landscape is complex. This is especially worrisome for non-domain experts or practitioners of deep learning who may not be familiar with the key characteristics that differentiate one land cover type or boundary from another. In this case, training was limited to four relatively distinct classes of importance to our wetland monitoring efforts, as opposed to more refined classes, to try and reduce those issues. Although different types of vegetation and land cover exist in the study area, this grouping aided our ability to efficiently label the data and serve the study purpose. However, the high level of classification accuracy reported here, to some

degree, may be a function of this class structure. Efforts to classify the land cover into more distinct categories and capture more biodiversity will be posed with greater labeling and training challenges and require more domain expertise. Classification accuracy may be lower in such cases than those reported here, especially if relying on low spectral resolution RGB imagery alone as evaluated in this study. Inevitably, some mixing of classes will occur during the labeling process, regardless of expertise or attention to detail, and these challenges will grow over heterogeneous and complex natural landscapes like coastal wetlands. This problem can be exacerbated when attempting to perform pixel-level labeling using very high-resolution imagery, such as created from a low-altitude UAS flight. This is due to a large amount of within class spectral variability when viewing land cover at zoomed in geographic scales (here cm-level). The errors in labeling are specifically maximized when pixels belonging to the borderlines are going to be labeled because natural targets do not usually express clear borders. In some landscapes, two or more different targets can be so mixed together that the operator cannot decide which label should be given to that specific pixel or area. Inevitably, it becomes highly subjective. Such areas can be seen in the lower right part of Figure 2.17 where a vegetation area has been submerged in shallow water. In this work, it was classified as a water body/submerged landscape. Additionally, at this specific study site, discriminating pixels belonging to edges of the tidal flat class from those belonging to the dirt road was a difficult task because those two classes exhibit very high inter-class similarities at their boundaries. As a result, the uncertainty for labeling road pixels close to the boundaries increased.

Lastly, coastal wetlands are among some of the most dynamic and complex ecosystems on the planet. Many different factors, such as seasonal and climate changes, water temperature, altered flooding and salinity patterns, sea-level rise, topography, etc. [10, 12], contribute to the current

state of the land cover and its physical properties at the time of recording the remote sensing observations. Thus, the authors emphasize that the classification results shown here, based on the classes chosen to be examined, are valid for the specific data set acquired at a certain time over the study site. The results cannot be necessarily generalized to the same coastal wetland area imaged at a different time, or at a different land cover state, without further analyses. Ambient environmental conditions, such as lighting or wind, can impact data captured in an UAS image. Similarly, flight design including altitude above ground and camera perspective (e.g., oblique versus nadir) will impact the GSD and appearance of land cover features. As a result, the visual representation of the same target may deviate from one exposure to another in a single UAS flight mission and across repeat data acquisitions. For this study, UAS data acquisition targeted calm winds and a bright, sunny day. The flight was conducted during the middle part of the day to reduce shadowing and enhance scene brightness. Furthermore, the entire scene was mapped in under thirty minutes so variation of ambient lighting during flight was minimal. Camera angles were kept at nadir to provide a top-down view for orthoimage generation and reduce shadowing of terrain from oblique perspectives.

Future efforts will need to examine the generalizability and stability of these models to perform repetitive classification using a time series of images captured from repeat UAS flights under varying conditions. However, we believe that the high capacity of deep CNN models to efficiently extract informative and discriminative features from the raw UAS images in an end-to-end manner have the potential to be extended further by training deep CNN models using a time-series of UAS images acquired over the same area. An efficient deep network trained using appropriate training samples acquired at different times and labeled by expert knowledge will be able to capture more properties about a certain land cover target at a different state of the wetland

or other environment. Such models could provide a powerful framework for designing any automatic or online land cover prediction system aiming to offer high performance regardless of the conditions at the time of data acquisition.

2.6. Conclusion

Wetlands provide a challenging natural environment for performing high accuracy land cover prediction with hyper-spatial resolution UAS imagery due to high intra-class variability and low inter-class disparity often observed between classes. For decades, semantic image segmentation for land cover mapping tasks in the RS field has relied heavily on the tedious procedure of manually designing and extracting the most informative hand-crafted features from the available data, which are then fed into different machine learning techniques for classification or segmentation. On the other hand, the accuracy of any prediction technique is highly dependent on the contribution of those features for discriminating different targets that are captured in high-spatial to hyper-spatial resolution images, such as those acquired by UAS flying at low altitude.

In this research, we exploited state-of-the-art deep learning frameworks, commonly called deep CNNs, to automatically explore high-dimensional hierarchical feature spaces and find the most informative and discriminative features for performing a pixel-wise image labeling task for land cover mapping. Among the many available deep CNN architectures, this study investigated the performance of some of the most recent very deep CNN architectures that are heavily employed for pixel-level labeling in many different applications. Six different networks were evaluated, FC-DenseNet, U-Net, DeepLabV3+, MobileU-Net, PSPNet, and Encoder-Decoder (SegNet), for performing a pixel-wise classification task using UAS hyper-spatial resolution images acquired over a coastal wetland area. Results of the study revealed that hierarchical features learned by the deep learning frameworks are highly efficient for discriminating different targets in a complex

wetland environment and providing accurate pixel-level land cover predictions for the target classes investigated (vegetation, tidal flat, water, road). Specifically, fine-tuning of deep architectures with tens of millions of parameters is the best strategy when there is a limited labeled dataset as was the case in this study. This is also the case for most current RS land cover mapping applications where large repositories of relevant labeled datasets are not available. In this study, FC-DenseNet trained from scratch outperformed the other architectures regarding the overall accuracy performances (Table 2.1) based on the validation dataset. However, U-Net architecture with ResNet34 encoder outperformed the other architectures based on training speed while achieving comparable accuracy to FC-DenseNet. These results suggest that U-Net is the most efficient architecture for the UAS hyper-spatial pixel-wise classification task explored here. Skip connections in FC-DenseNet and U-Net architecture play a significant role in these networks' ability for faster training and/or achieving higher overall accuracies. DeepLabV3+, which uses the ASPP technique to account for objects at multiple scales, was also very successful at pixel-level prediction in our study case. Furthermore, results from per-class accuracy revealed that almost all networks were able to successfully predict pixels belonging to the vegetation area with high accuracy.

The experiment with the U-Net architecture employing a ResNet34 encoder revealed that fine-tuning using the transfer learning technique works well for hyper-spatial UAS image analyses. Furthermore, the transfer learning technique in combination with skip connections applied to the architecture of CNNs significantly reduced the need for a large number of training epochs, and large labeled data resources, typically required for training deep CNNs without sacrificing their high classification performance. In this study, FC-DenseNet, with 56 convolutional layers, trained from scratch performed comparably well with the U-Net architecture regarding the overall

classification accuracy evaluated on the training dataset. This suggests that the parameter-based compactness of FC-DenseNet makes it a good choice among other deep CNN architectures for accurate pixel-wise labeling in RS applications where transfer learning may not be efficiently applicable and/or higher level of generalization with a limited training sample is required. However, as long as training from scratch is applied to FC-DenseNet, it would need more training epochs to reach an overall accuracy comparable to U-Net using a pre-trained encoder with the same number of training samples.

In conclusion, the results of this study demonstrate the high potential for exploiting recent deep CNN architectures to perform pixel-wise land cover mapping with hyper-spatial resolution imagery acquired from a small UAS equipped with an RGB camera or other RS method. Transfer learning is highly applicable for training deep CNNs in RS applications to help achieve state-of-the-art performances when faced with limited labeled data resources. Finally, coastal wetlands are highly diverse natural environments providing a range of complexities if attempting to identify more refined land covers, such as vegetation types. Such efforts will likely demand more advanced sensors to capture finer spectral information from the different targets. Future work will explore deep CNN architectures for pixel-wise labeling of multispectral and hyperspectral images to predict land cover in a coastal wetland setting. Furthermore, this study did not evaluate the uncertainty involved in training each individual CNN architecture. The classification performance, reported on training and validation datasets, is based on a single training and validation process on each CNN architecture. Evaluating CNN model's uncertainty during training phase may be considered in future work.

2.7. References

1 Boon, M.A., Greenfield, R., and Tesfamichael, S.: ‘Wetland assessment using unmanned aerial vehicle (UAV) photogrammetry’, 2016

2 Laliberte, A.S., Rango, A., and Herrick, J.: ‘Unmanned aerial vehicles for rangeland mapping and monitoring: a comparison of two systems’, in Editor (Ed.)^(Eds.): ‘Book Unmanned aerial vehicles for rangeland mapping and monitoring: a comparison of two systems’ (2007, edn.), pp.

3 Pashaei, M., and Starek, M.J.: ‘Fully Convolutional Neural Network for Land Cover Mapping In A Coastal Wetland with Hyperspatial UAS Imagery’, in Editor (Ed.)^(Eds.): ‘Book Fully Convolutional Neural Network for Land Cover Mapping In A Coastal Wetland with Hyperspatial UAS Imagery’ (IEEE, 2019, edn.), pp. 6106-6109

4 Long, J., Shelhamer, E., and Darrell, T.: ‘Fully convolutional networks for semantic segmentation’, in Editor (Ed.)^(Eds.): ‘Book Fully convolutional networks for semantic segmentation’ (2015, edn.), pp. 3431-3440

5 Hariharan, B., Arbeláez, P., Girshick, R., and Malik, J.: ‘Simultaneous detection and segmentation’, in Editor (Ed.)^(Eds.): ‘Book Simultaneous detection and segmentation’ (Springer, 2014, edn.), pp. 297-312

6 Stedman, S.-M., and Dahl, T.E.: ‘Status and trends of wetlands in the coastal watersheds of the eastern United States, 1998 to 2004’, 2008

7 Pendleton, L.H.: ‘The economic and market value of coasts and estuaries: what's at stake?’, The economic and market value of coasts and estuaries: what's at stake?, 2010

8 Ross, M.S., Reed, D.L., Sah, J.P., Ruiz, P.L., and Lewin, M.: ‘Vegetation: environment relationships and water management in Shark Slough, Everglades National Park’, Wetl Ecol Manag, 2003, 11, (5), pp. 291-303

9 Belluco, E., Camuffo, M., Ferrari, S., Modenese, L., Silvestri, S., Marani, A., and Marani, M.: ‘Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing’, Remote Sens Environ, 2006, 105, (1), pp. 54-67

10 Cahoon, D.R., and Guntenspergen, G.R.: ‘Climate change, sea-level rise, and coastal wetlands’, National Wetlands Newsletter, 2010, 32, (1), pp. 8-12

11 Silvestri, S., Marani, M., and Marani, A.: ‘Hyperspectral remote sensing of salt marsh vegetation, morphology and soil topography’, Physics and Chemistry of the Earth, Parts a/B/C, 2003, 28, (1-3), pp. 15-25

12 Taramelli, A., Valentini, E., Cornacchia, L., Monbaliu, J., and Sabbe, K.: ‘Indications of dynamic effects on scaling relationships between channel sinuosity and vegetation patch size across a salt marsh platform’, Journal of Geophysical Research: Earth Surface, 2018, 123, (10), pp. 2714-2731

13 Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., and Weng, Q.: ‘Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery’, Remote Sens Environ, 2011, 115, (5), pp. 1145-1161

14 Hsieh, P.-F., Lee, L.C., and Chen, N.-Y.: ‘Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing’, IEEE Transactions on Geoscience and Remote Sensing, 2001, 39, (12), pp. 2657-2663

15 Blaschke, T.: ‘Object based image analysis for remote sensing’, ISPRS journal of photogrammetry and remote sensing, 2010, 65, (1), pp. 2-16

16 Dronova, I., Gong, P., and Wang, L.: 'Object-based analysis and change detection of major wetland cover types and their classification uncertainty during the low water period at Poyang Lake, China', Remote Sens Environ, 2011, 115, (12), pp. 3220-3236

17 Small, C., and Milesi, C.: 'Multi-scale standardized spectral mixture models', Remote Sens Environ, 2013, 136, pp. 442-454

18 Pande-Chhetri, R., Abd-Elrahman, A., Liu, T., Morton, J., and Wilhelm, V.L.: 'Object-based classification of wetland vegetation using very high-resolution unmanned air system imagery', European Journal of Remote Sensing, 2017, 50, (1), pp. 564-576

19 Li, M., Zang, S., Zhang, B., Li, S., and Wu, C.: 'A review of remote sensing image classification techniques: The role of spatio-contextual information', European Journal of Remote Sensing, 2014, 47, (1), pp. 389-411

20 Whiteside, T.G., Boggs, G.S., and Maier, S.W.: 'Comparing object-based and pixel-based classifications for mapping savannas', International Journal of Applied Earth Observation and Geoinformation, 2011, 13, (6), pp. 884-893

21 Gao, Y., and Mas, J.F.: 'A Comparison of the Performance of Pixel Based and Object Based Classifications over Images with Various Spatial Resolutions', 2008

22 Blanzieri, E., and Melgani, F.: 'Nearest neighbor classification of remote sensing images with the maximal margin principle', IEEE Transactions on geoscience and remote sensing, 2008, 46, (6), pp. 1804-1811

23 Goncalves, M., Netto, M., Costa, J., and Zullo Junior, J.: 'An unsupervised method of classifying remotely sensed images using Kohonen self-organizing maps and agglomerative hierarchical clustering methods', Int J Remote Sens, 2008, 29, (11), pp. 3171-3207

24 Kavzoglu, T., and Mather, P.: ‘The use of backpropagating artificial neural networks in land cover classification’, Int J Remote Sens, 2003, 24, (23), pp. 4907-4938

25 Vapnik, V.: ‘The nature of statistical learning theory’ (Springer science & business media, 2013. 2013)

26 Breiman, L.: ‘Random forests’, Machine learning, 2001, 45, (1), pp. 5-32

27 Chen, Y., Lin, Z., Zhao, X., Wang, G., and Gu, Y.: ‘Deep learning-based classification of hyperspectral data’, IEEE Journal of Selected topics in applied earth observations and remote sensing, 2014, 7, (6), pp. 2094-2107

28 Zou, Q., Ni, L., Zhang, T., and Wang, Q.: ‘Deep learning based feature selection for remote sensing scene classification’, IEEE Geoscience and Remote Sensing Letters, 2015, 12, (11), pp. 2321-2325

29 Chen, Y., Zhao, X., and Jia, X.: ‘Spectral–spatial classification of hyperspectral data based on deep belief network’, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8, (6), pp. 2381-2392

30 Cheng, G., Zhou, P., and Han, J.: ‘Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images’, IEEE Transactions on Geoscience and Remote Sensing, 2016, 54, (12), pp. 7405-7415

31 Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A.L.: ‘Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs’, IEEE transactions on pattern analysis and machine intelligence, 2017, 40, (4), pp. 834-848

32 Lin, G., Shen, C., Van Den Hengel, A., and Reid, I.: 'Efficient piecewise training of deep structured models for semantic segmentation', in Editor (Ed.)^(Eds.): 'Book Efficient piecewise training of deep structured models for semantic segmentation' (2016, edn.), pp. 3194-3203

33 Dai, J., He, K., and Sun, J.: 'Instance-aware semantic segmentation via multi-task network cascades', in Editor (Ed.)^(Eds.): 'Book Instance-aware semantic segmentation via multi-task network cascades' (2016, edn.), pp. 3150-3158

34 Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., and Garcia-Rodriguez, J.: 'A review on deep learning techniques applied to semantic segmentation', arXiv preprint arXiv:1704.06857, 2017

35 Yosinski, J., Clune, J., Bengio, Y., and Lipson, H.: 'How transferable are features in deep neural networks?', arXiv preprint arXiv:1411.1792, 2014

36 Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P.: 'Convolutional neural networks for large-scale remote-sensing image classification', IEEE Transactions on Geoscience and Remote Sensing, 2016, 55, (2), pp. 645-657

37 Romero, A., Gatta, C., and Camps-Valls, G.: 'Unsupervised deep feature extraction for remote sensing image classification', IEEE Transactions on Geoscience and Remote Sensing, 2015, 54, (3), pp. 1349-1362

38 Paoletti, M.E., Haut, J.M., Plaza, J., and Plaza, A.: 'A new deep convolutional neural network for fast hyperspectral image classification', ISPRS journal of photogrammetry and remote sensing, 2018, 145, pp. 120-147

39 Palsson, F., Sveinsson, J.R., and Ulfarsson, M.O.: 'Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network', IEEE Geoscience and Remote Sensing Letters, 2017, 14, (5), pp. 639-643

40 Liu, T., Abd-Elrahman, A., Morton, J., and Wilhelm, V.L.: ‘Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system’, GIScience & remote sensing, 2018, 55, (2), pp. 243-264

41 Liu, T., and Abd-Elrahman, A.: ‘Deep convolutional neural network training enrichment using multi-view object-based analysis of Unmanned Aerial systems imagery for wetlands classification’, ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 139, pp. 154-170

42 Pouliot, D., Latifovic, R., Pasher, J., and Duffe, J.: ‘Assessment of convolution neural networks for wetland mapping with landsat in the central Canadian boreal forest region’, Remote Sensing, 2019, 11, (7), pp. 772

43 Hu, Y., Zhang, J., Ma, Y., An, J., Ren, G., and Li, X.: ‘Hyperspectral coastal wetland classification based on a multiobject convolutional neural network model and decision fusion’, IEEE Geoscience and Remote Sensing Letters, 2019, 16, (7), pp. 1110-1114

44 Bengio, Y.: ‘Gradient-based optimization of hyperparameters’, Neural computation, 2000, 12, (8), pp. 1889-1900

45 Romera-Paredes, B., and Torr, P.H.S.: ‘Recurrent instance segmentation’, in Editor (Ed.)^(Eds.): ‘Book Recurrent instance segmentation’ (Springer, 2016, edn.), pp. 312-329

46 Zeiler, M.D., and Fergus, R.: ‘Visualizing and understanding convolutional networks’, in Editor (Ed.)^(Eds.): ‘Book Visualizing and understanding convolutional networks’ (Springer, 2014, edn.), pp. 818-833

47 Zeiler, M.D., Taylor, G.W., and Fergus, R.: ‘Adaptive deconvolutional networks for mid and high level feature learning’, in Editor (Ed.)^(Eds.): ‘Book Adaptive deconvolutional networks for mid and high level feature learning’ (IEEE, 2011, edn.), pp. 2018-2025

48 Everingham, M., Van Gool, L., Williams, C.K., Winn, J., and Zisserman, A.: 'The pascal visual object classes (voc) challenge', International journal of computer vision, 2010, 88, (2), pp. 303-338

49 Nair, V., and Hinton, G.E.: 'Rectified linear units improve restricted boltzmann machines', in Editor (Ed.)^(Eds.): 'Book Rectified linear units improve restricted boltzmann machines' (2010, edn.), pp.

50 LeCun, Y.: 'LeNet-5, convolutional neural networks', URL: http://yann. lecun. com/exdb/lenet, 2015, 20, (5), pp. 14

51 Krizhevsky, A., Sutskever, I., and Hinton, G.E.: 'Imagenet classification with deep convolutional neural networks', Advances in neural information processing systems, 2012, 25, pp. 1097-1105

52 Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: 'Imagenet: A large-scale hierarchical image database', in Editor (Ed.)^(Eds.): 'Book Imagenet: A large-scale hierarchical image database' (Ieee, 2009, edn.), pp. 248-255

53 Simonyan, K., and Zisserman, A.: 'Very deep convolutional networks for large-scale image recognition', arXiv preprint arXiv:1409.1556, 2014

54 Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: 'Going deeper with convolutions', in Editor (Ed.)^(Eds.): 'Book Going deeper with convolutions' (2015, edn.), pp. 1-9

55 He, K., Zhang, X., Ren, S., and Sun, J.: 'Deep residual learning for image recognition', in Editor (Ed.)^(Eds.): 'Book Deep residual learning for image recognition' (2016, edn.), pp. 770-778

56 Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q.: ‘Densely connected convolutional networks’, in Editor (Ed.)^(Eds.): ‘Book Densely connected convolutional networks’ (2017, edn.), pp. 4700-4708

57 Ioffe, S., and Szegedy, C.: ‘Batch normalization: Accelerating deep network training by reducing internal covariate shift’, in Editor (Ed.)^(Eds.): ‘Book Batch normalization: Accelerating deep network training by reducing internal covariate shift’ (PMLR, 2015, edn.), pp. 448-456

58 Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z.: ‘Rethinking the inception architecture for computer vision’, in Editor (Ed.)^(Eds.): ‘Book Rethinking the inception architecture for computer vision’ (2016, edn.), pp. 2818-2826

59 Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A.: ‘Inception-v4, inception-resnet and the impact of residual connections on learning’, in Editor (Ed.)^(Eds.): ‘Book Inception-v4, inception-resnet and the impact of residual connections on learning’ (2017, edn.), pp.

60 Chollet, F.: ‘Xception: Deep learning with depthwise separable convolutions’, in Editor (Ed.)^(Eds.): ‘Book Xception: Deep learning with depthwise separable convolutions’ (2017, edn.), pp. 1251-1258

61 Glorot, X., Bordes, A., and Bengio, Y.: ‘Deep sparse rectifier neural networks’, in Editor (Ed.)^(Eds.): ‘Book Deep sparse rectifier neural networks’ (JMLR Workshop and Conference Proceedings, 2011, edn.), pp. 315-323

62 He, K., Zhang, X., Ren, S., and Sun, J.: ‘Identity mappings in deep residual networks’, in Editor (Ed.)^(Eds.): ‘Book Identity mappings in deep residual networks’ (Springer, 2016, edn.), pp. 630-645

63 Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K.: ‘Aggregated residual transformations for deep neural networks’, in Editor (Ed.)^(Eds.): ‘Book Aggregated residual transformations for deep neural networks’ (2017, edn.), pp. 1492-1500

64 Huang, G., Sun, Y., Liu, Z., Sedra, D., and Weinberger, K.Q.: ‘Deep networks with stochastic depth’, in Editor (Ed.)^(Eds.): ‘Book Deep networks with stochastic depth’ (Springer, 2016, edn.), pp. 646-661

65 Veit, A., Wilber, M., and Belongie, S.: ‘Residual networks behave like ensembles of relatively shallow networks’, arXiv preprint arXiv:1605.06431, 2016

66 Wu, Z., Shen, C., and Van Den Hengel, A.: ‘Wider or deeper: Revisiting the resnet model for visual recognition’, Pattern Recognition, 2019, 90, pp. 119-133

67 Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H.: ‘Mobilenets: Efficient convolutional neural networks for mobile vision applications’, arXiv preprint arXiv:1704.04861, 2017

68 Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C.: ‘Mobilenetv2: Inverted residuals and linear bottlenecks’, in Editor (Ed.)^(Eds.): ‘Book Mobilenetv2: Inverted residuals and linear bottlenecks’ (2018, edn.), pp. 4510-4520

69 Zeiler, M.D., Krishnan, D., Taylor, G.W., and Fergus, R.: ‘Deconvolutional networks’, in Editor (Ed.)^(Eds.): ‘Book Deconvolutional networks’ (IEEE, 2010, edn.), pp. 2528-2535

70 Noh, H., Hong, S., and Han, B.: ‘Learning deconvolution network for semantic segmentation’, in Editor (Ed.)^(Eds.): ‘Book Learning deconvolution network for semantic segmentation’ (2015, edn.), pp. 1520-1528

71 Badrinarayanan, V., Kendall, A., and Cipolla, R.: ‘Segnet: A deep convolutional encoder-decoder architecture for image segmentation’, IEEE transactions on pattern analysis and machine intelligence, 2017, 39, (12), pp. 2481-2495

72 Ronneberger, O., Fischer, P., and Brox, T.: ‘U-net: Convolutional networks for biomedical image segmentation’, in Editor (Ed.)^(Eds.): ‘Book U-net: Convolutional networks for biomedical image segmentation’ (Springer, 2015, edn.), pp. 234-241

73 Burt, P.J., and Adelson, E.H.: ‘The Laplacian pyramid as a compact image code’: ‘Readings in computer vision’ (Elsevier, 1987), pp. 671-679

74 Lin, G., Milan, A., Shen, C., and Reid, I.: ‘Refinenet: Multi-path refinement networks for high-resolution semantic segmentation’, in Editor (Ed.)^(Eds.): ‘Book Refinenet: Multi-path refinement networks for high-resolution semantic segmentation’ (2017, edn.), pp. 1925-1934

75 Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., and Bengio, Y.: ‘The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation’, in Editor (Ed.)^(Eds.): ‘Book The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation’ (2017, edn.), pp. 11-19

76 Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A.L.: ‘Semantic image segmentation with deep convolutional nets and fully connected crfs’, arXiv preprint arXiv:1412.7062, 2014

77 Yu, F., and Koltun, V.: ‘Multi-scale context aggregation by dilated convolutions’, arXiv preprint arXiv:1511.07122, 2015

78 Holschneider, M., Kronland-Martinet, R., Morlet, J., and Tchamitchian, P.: ‘A real-time algorithm for signal analysis with the help of the wavelet transform’: ‘Wavelets’ (Springer, 1990), pp. 286-297

79 Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H.: 'Encoder-decoder with atrous separable convolution for semantic image segmentation', in Editor (Ed.)^(Eds.): 'Book Encoder-decoder with atrous separable convolution for semantic image segmentation' (2018, edn.), pp. 801-818

80 Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H.: 'Rethinking atrous convolution for semantic image segmentation', arXiv preprint arXiv:1706.05587, 2017

81 Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J.: 'Pyramid scene parsing network', in Editor (Ed.)^(Eds.): 'Book Pyramid scene parsing network' (2017, edn.), pp. 2881-2890

82 Pan, S.J., and Yang, Q.: 'A survey on transfer learning', IEEE Transactions on knowledge and data engineering, 2009, 22, (10), pp. 1345-1359

83 Goutte, C., and Gaussier, E.: 'A probabilistic interpretation of precision, recall and F-score, with implication for evaluation', in Editor (Ed.)^(Eds.): 'Book A probabilistic interpretation of precision, recall and F-score, with implication for evaluation' (Springer, 2005, edn.), pp. 345-359

84 Rahman, M.A., and Wang, Y.: 'Optimizing intersection-over-union in deep neural networks for image segmentation', in Editor (Ed.)^(Eds.): 'Book Optimizing intersection-over-union in deep neural networks for image segmentation' (Springer, 2016, edn.), pp. 234-244

85 Paine, J.G., White, W.A., Smyth, R.C., Andrews, J.R., and Gibeaut, J.C.: 'Mapping coastal environments with lidar and EM on Mustang Island, Texas, US', The Leading Edge, 2004, 23, (9), pp. 894-898

86 Nguyen, C., Starek, M.J., Tissot, P., and Gibeaut, J.: 'Unsupervised clustering method for complexity reduction of terrestrial lidar data in marshes', Remote Sensing, 2018, 10, (1), pp. 133

87 Nguyen, C., Starek, M.J., Tissot, P., and Gibeaut, J.: ‘Unsupervised Clustering of Multi-Perspective 3D Point Cloud Data in Marshes: A Case Study’, Remote Sensing, 2019, 11, (22), pp. 2715

88 Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J., and Reynolds, J.M.: ‘‘Structure-from-Motion’photogrammetry: A low-cost, effective tool for geoscience applications’, Geomorphology, 2012, 179, pp. 300-314

89 Kingma, D.P., and Ba, J.: ‘Adam: A method for stochastic optimization’, arXiv preprint arXiv:1412.6980, 2014

90 Duchi, J., Hazan, E., and Singer, Y.: ‘Adaptive subgradient methods for online learning and stochastic optimization’, Journal of machine learning research, 2011, 12, (7)

91 Tieleman, T., and Hinton, G.: ‘Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude’, COURSERA: Neural networks for machine learning, 2012, 4, (2), pp. 26-31

CHAPTER III: DEEP LEARNING-BASED SINGLE IMAGE SUPER-RESOLUTION: AN INVESTIGATION FOR DENSE RECONSTRUCTUON WITH UAS PHOTOGRAMMETRY

Abstract

The deep convolutional neural network (DCNN) has recently been applied to the highly challenging and ill-posed problem of single image super-resolution (SISR), which aims to predict high-resolution (HR) images from their corresponding low-resolution (LR) images. In many remote sensing (RS) applications, spatial resolution of the aerial or satellite imagery has a great impact on the accuracy and reliability of information extracted from the images. In this study, the potential of a DCNN-based SISR model, called enhanced super-resolution generative adversarial network (ESRGAN), to predict the spatial information degraded or lost in a hyper-spatial resolution unmanned aircraft system (UAS) RGB image set is investigated. ESRGAN model is trained over a limited number of original HR (50 out of 450 total images) and virtually generated LR UAS images, by downsampling the original HR images using the bicubic kernel by factor 4. Quantitative and qualitative assessments of super-resolved images using standard image quality measures (IQMs) confirm that the DCNN-based SISR approach can be successfully applied on LR UAS imagery for spatial resolution enhancement. The performance of DCNN-based SISR approach for the UAS image set closely approximates performances reported on standard SISR image sets with mean peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index values of around 28 dB and 0.85, respectively. Furthermore, by exploiting the rigorous Structure-from-Motion (SfM) photogrammetry procedure, an accurate task-based IQM for evaluating the quality of the super-resolved images is carried out. Results verify that the interior and exterior imaging geometry, which are extremely important for extracting highly accurate spatial information from UAS imagery in photogrammetric applications, can be accurately retrieved from

a super-resolved image set. The number of corresponding keypoints and dense points generated from the SfM photogrammetry process are about 6 and 17 times more than those extracted from the corresponding LR image set, respectively.

3.1. Introduction

Remote sensing (RS) In most remote sensing (RS) applications, high-resolution (HR) images are usually more demanding in a wide range of image analysis tasks leading to more precise and accurate RS-derived products [1-3]. HR imagery is usually more desirable in all applications, including RS imagery, because improved pictorial information makes visual interpretation easier for a human and helps to purify representation for automatic machine perception [4]. In RS applications, the resolution of a digital imaging system can be classified in four different ways: *spatial resolution*, *spectral resolution*, *radiometric resolution*, and *temporal resolution*. In the context of accurate feature mapping and positioning in RS, spatial resolution is of the greatest challenge.

Spatial resolution of a digital imaging system is primarily defined by the pixel density in the image space, which is measured in pixels per unit area. Spatial resolution in the object space represents the level of spatial detail that can be discerned in an image, the higher the resolution, the more image details. Limited spatial resolution in a certain image is primarily a function of the imaging sensor or acquisition device [4]. The spatial resolution of imagery, usually referred to as ground sample distance (GSD) in RS applications, is determined by the sensor size or the dimension of charge-coupled device (CCD) or charge-coupled device (CCD) chip, the number of sensor elements, the focal length of the imaging device, and its distance from the imaging target. Regardless of the other factors contributing to the spatial resolution of imagery, such as focal length and the distance from sensor to the target, GSD of an image and the quality of its high-

frequency contents deteriorate mainly due to some manufacturing limitations and imperfections of an imaging sensor.

One straightforward way to improve the spatial resolution or GSD of imagery is to build a more compact sensor in which the sensor's pixel density is increased by reducing the sensor element size. However, this reduction in sensor element size may dramatically reduce the amount of light incident on each sensor element, causing the so called *shot noise* [5]. Furthermore, capture of high frequency image detail is also limited or degraded by the sensor optics, such as lens blur, lens aberration, and aperture diffraction, or any external sources of image degradation including image motion due to moving objects [4]. Constructing high-quality imaging sensors with perfect optical components, capturing very high spatial resolution images with high-quality image content, is restrictively expensive and not practical in most real scenarios. This is especially true when referring to the rapid rise in the use of small, unmanned aircraft systems (UASs) for RS and photogrammetry applications [4]. Such small UASs are typically equipped with low-cost, consumer-grade digital RGB cameras. Besides the cost, the resolution of these typical UAS cameras is also limited by the camera speed and hardware storage. Physical constraints of the sensing platform or environment, such as with satellite imagery, can put additional constraints on the use of very high-resolution sensors. Furthermore, in some imaging systems, HR image content may not be always achievable due to inherent restrictions within the system itself including built-in downsampling procedures to handle bandwidth limitations, different types of noise related to the sensor electronics and atmosphere, compression techniques, etc. [6].

An alternative approach to hardware-based solutions for spatial resolution enhancement is to accept the image degradation and apply signal processing techniques to attempt to recover fine image details degraded or almost lost during image capture. These approaches are often referred

to as *Super-Resolution* (SR) image reconstruction techniques. SR techniques attempt to recover HR images from LR images, and this task remains an important yet challenging topic in image processing that has a wide range of applications in computer vision and image understanding tasks [7-10]. SR techniques not only improve image perceptual quality, but also help to improve the final accuracy of many computer vision tasks [11-13]. Application of SR techniques on highly detailed and complex RS data introduces more challenges to the SR problem [14, 15]. Most traditional image SR techniques use highly sophisticated signal processing algorithms with a very high computational complexity [15, 16]. Considering the size and the volume of required super-resolved images for some RS applications, such as generating a precise digital surface model (DSM) using aerial or satellite photogrammetry, traditional SR techniques are highly inefficient for such applications. Furthermore, some techniques require multiple LR images from the same scene with high temporal resolution to resolve the SR problem [15, 17, 18]. However, due to costs or limitations for acquiring the necessary imagery, complexity of natural and built terrain, scarcity of multi-view sensors, and need for accurate image registration algorithms, acquiring and processing such images for SR is a difficult task [15]. In addition, complicated and versatile interaction of most RS sensors with atmosphere and objects, image displacements due to topographic anomalies, land cover characteristics, and participation of shaded areas due to the Sun-sensor-object geometry in RS images, make the SR problem a highly challenging task for almost all developed techniques in this field [15].

Deep learning (DL), specifically deep convolutional neural network (DCNN), has recently been applied to a wide range of image analysis tasks [19, 20] including the highly challenging and ill-posed problem of predicting HR images from LR images in an end-to-end manner. These methods have already shown their superiority over almost all traditional techniques by achieving

state-of-the-art performance on various SR benchmarks [21-23]. Currently, DCNN-based single image super-resolution (SISR) techniques have been employed to increase the geometrical and interpretation quality of RS imagery [24-26]. However, few studies have focused on applying DCNN-based SISR on UAS-based imagery, typically acquired at low altitudes with high resolution, where the accuracy of the spatial information captured by the images is critical for the reliability of results drawn from subsequent analyses [27, 28]. Recently, super-resolution generative adversarial network (SRGAN) [21], is considered as one of the most efficient DCNN-based SISR models for recovering very fine details in predicted HR images from corresponding LR images . Offering finer image content is always one of the most important characteristics of HR images in different RS applications, which can lead to higher accuracy and reliability in almost all spatial and non-spatial RS products. SRGAN has already proved its superiority over many other DCNN-based SISR models for recovering very fine details in predicted HR images, which are highly valuable for improving human image perception. However, the quality of the recovered image details and their potential for enhancement of hyper-spatial resolution UAS imagery for photogrammetric applications, such as dense 3D reconstruction of a scene, has not yet been fully explored. With this motivation, this paper focuses on the application of DCNN to SISR for UAS image enhancement. The contributions of the paper are as follows:

(a) An overview of the SR problem and DCNN approaches for SISR is provided with emphasis on generative adversarial network (GAN) architecture. GAN-based models are fully reviewed including their specific loss functions. Additionally, different learning strategies and image quality measures (IQMs) typically employed for SISR tasks are reviewed.

(b) A high performance DCNN-based SISR model based on GAN architecture [29], known as enhanced SRGAN (ESRGAN) [30], is adopted and trained on a set of LR UAS images virtually generated by downsampling the original HR image set by factor 4. Additive white Gaussian noise is applied to the LR imagery to make the SISR task more challenging. Such noise can always appear in any digital imaging and image transmission systems due to the electronics, imaging sensor quality, and the interaction of the digital imaging system with the natural environment, such as the level of illumination, temperature, etc. [31]. Model performance in recovering the degraded or lost image details and noise reduction in the predicted super-resolved images is then carried out using standard IQMs. In this experiment, IQMs include peak signal-to-noise ratio (PSNR), structure similarity (SSIM) index, and a qualitative analysis through visually inspecting resulting SR images.

(c) A task-based IQM using Structure-from-Motion (SfM) photogrammetry is carried out on the predicted SR image set.

(d) A comprehensive comparative analysis of SfM derived photogrammetric data products, resulting from processing of the LR, HR, and SR UAS image sets, is carried out. Those products include: the camera calibration and camera pose information, densified 3D point clouds, and digital surface models (DSMs).

Regarding the UAS-SfM task-based evaluation for SR described above, the primary objectives of the experiment are summarized as follows:

(1) The performance of the adopted DCNN-based SISR model on retrieving both the interior and exterior geometry of the UAS imagery is investigated. In the SfM photogrammetry, the accuracy and reliability of all derived parameters, within the

robust *bundle adjustment* (BA) computations, are extremely related to the accuracy and reliability of extracted *keypoint* features from raw images. Any image distortions and artefacts introduced by adding noise or upsampling images can dramatically affect the reliability of derived parameters within BA computations.

(2) The potential of the employed DCNN-based SISR model to downgrade the level of inherent and additional noise introduced to the original HR images is investigated. In most image-based 3D reconstruction algorithms, including SfM photogrammetry, lower level of noise in the underlying image set results in estimating the imaging and scene geometry with higher accuracy. That is due to the fact that the feature detection operators, using sophisticated image processing algorithms, extract *keypoint* features with higher accuracy and lower uncertainty across multiple images in an UAS image set. To do this, the naive pre-trained ESRGAN model, with upscaling factor 1, is taken as an image restoration network. The idea is to explore the effectiveness of ESRGAN model, trained on a large number of images within several standard image sets, to downgrade the inherent noise and restore the original UAS HR images.

The remainder of this chapter is organized as follows. Section 2 briefly describes image SR as an image upscaling technique to recover the degraded or lost image details in LR images. Section 3 introduces some of the pioneering DCNN-based SISR architectures. GAN-based architecture and its certain cost function for SISR task is later described in section 3. Learning strategies in Section 4 introduce different cost functions that are usually used in DCNN-based SISR models. Different metrics developed for evaluating the quality of resulting SR images are explained in Section 5. Section 6 explains the experiment including the employed DCNN-based SISR model. Section 7 reports the qualitative and quantitative results showing the performance of

ESRGAN model on virtually generated LR UAS images based on standard IQMs and a task-based IQM using SfM photogrammetry. Section 8 discusses the results in detail. Lastly, Section 9 provides a conclusion and future perspective.

3.2. Image Super-Resolution

Image SR refers to techniques which aim to restore a HR image from its LR counterpart(s). Their main goal is to recover the high frequency details lost in LR images and remove the degradation caused by the imaging device and/or environment [32, 33]. SR is a topic of great interest in digital image processing and many computer vision related applications including, HDTV [34], medical imaging [35, 36], satellite imaging [37], face recognition [38], security and surveillance [39]. The basic idea in most SR techniques is to extract the non-redundant image content in multiple LR images and combine them to generate a HR image [5]. Single image interpolation is an easy approach within many available SR techniques, which can be used to increase the image size [4]. However, several works showed that it does not provide any additional information and would dramatically decimate details of the image [4, 22, 40].

Generally, the SR problem assumes the LR image represents a downsampled, noisy, and blurred (by an unknown low-pass filter) version of HR data. Due to the non-invertibility of the degradation process, SR problem is inherently ill-posed [41]. In other words, it is an under-determined inverse problem, of which the solution is not unique. In the typical SR framework, as depicted in Figure 3.1, the LR image I_x is modeled as follows [42]:

$$I_x = \mathcal{D}(I_y; \delta) \tag{10}$$

where I_y is the corresponding HR image, $\mathcal{D}$ represents a degradation function, and δ is a set of parameters, e.g., the parameters of the unknown convolutional kernel, the scaling factor, and some noise related factors, contributing to the degradation process. Under general conditions, the

degradation process from $\mathcal{D}$ is unknown and only LR image, I_x, is provided. Thus, the SR operation, the reverse path in Figure 3.1, is an extremely challenging task, which effectively results in a one-to-many mapping from LR to HR image space [23].

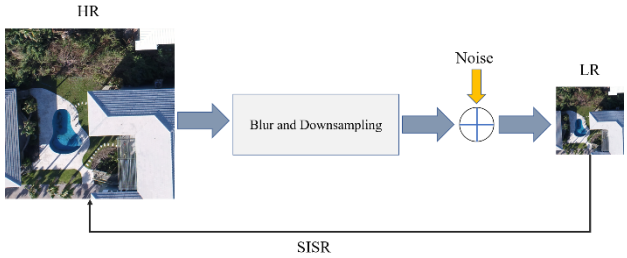


Figure 3.1. The overall framework for SISR.

Researchers are required to recover the corresponding HR image $\hat{I}_y$ from the LR image I_x, so that $\hat{I}_y$ is identical to the ground truth HR image I_y, as follows [42]:

$$\hat{I}_y = \mathcal{F}(I_x; \theta) \tag{2}$$

where $\mathcal{F}$ is the super-resolution model and θ represents the parameters of $\mathcal{F}$. Generally, degradation models combine several operations as follows [42]:

$$(I_y; \delta) = (I_y \otimes \kappa) \downarrow_s + \eta_\xi, \quad \{\kappa, s, \xi\} \subset \delta \tag{3}$$

where $(I_y \otimes \kappa)$ represents the convolution between a blur kernel κ and the HR image I_y, $\downarrow_s$ represents a downsampling process with factor s, and η_ξ is some additive white Gaussian noise with standard deviation ξ.

SR techniques typically assume that high-frequency image contents are redundant and can be reconstructed from low-frequency contents making the SR technique an inference problem [41]. Some SR techniques assume that for reconstructing a HR image of a certain scene, multiple LR instances of the same scene with different perspectives are available. These techniques are categorized as multi-image SR (MISR) approaches [16]. Such methods attempt to invert the

downsampling process by exploiting the explicit redundancy and constraining the ill-posed problem with additional information. However, MISR methods are usually computationally expensive because they require complex image registration and fusion in LR image space, where the accuracy of those processes directly affects the quality of the resulting super-resolved images [41]. An alternative approach is single image super-resolution (SISR) [43]. These techniques attempt to exploit the implicit redundancy available in the LR images, in the form of local spatial correlation in an image or additional temporal correlations in a video and recover lost or deteriorated high-frequency content from a single LR instance. In SISR techniques, prior information is usually required to constrain the solution space [44].

3.3. Deep Learning for SISR

Learning-based methods, also known as example-based methods [4, 45-47], aim at estimating an effective mapping from LR to HR image pairs due to their fast computation and superior performance relative to many other traditional techniques [23]. These methods usually exploit machine learning (ML) algorithms to learn the statistical relationships between the HR and corresponding LR images from a substantial number of training samples [23]. Traditional methods for SISR suffer from a few drawbacks [23, 41]: 1) unclear and potentially very complex definition of the mapping between the LR and HR image spaces; 2) established sub-optimal high-dimensional mapping; 3) most traditional methods rely upon handcrafted features with expert domain knowledge. Recently, deep learning-based SISR methods have achieved remarkable improvements over all traditional and ML approaches [21-23]. These methods take advantage of the huge capacity of DL models to be able to provide an extremely nonlinear mapping in a very high-dimensional space from the input space to the solution space, and efficiently explore that

space to find the best solution. These methods usually take a DCNN architecture for low to high-level feature encoding and nonlinear feature mapping.

3.3.1. DCNN architectures for SISR

A variety of super-resolution models based on DCNN architectures have been proposed so far. Most of those models focus on supervised super-resolution, requiring both LR images and corresponding HR images, usually as ground truth (GT). These approaches are mostly composed of a set of major components and processing strategies including the model's main framework, upsampling method, network architecture, and learning strategy.

Super-resolution convolutional neural network (SRCNN) by Dong et al. [22, 48] in Figure 3.2 is a pioneering work in DCNN-based SISR approach. Despite its striking success, SRCNN model suffers from the following issues [23]. 1) Inputs to SRCNN are LR images upsampled to coarse HR images at a desired size using traditional methods (e.g., bicubic interpolation). Introducing interpolated images as inputs to the network have three main drawbacks: (a) severe over-smoothing and noise amplification effects introduced to interpolated inputs can result in further inaccurate estimations of the image content; (b) employing interpolated versions of images, instead of the original LR image, as input is very time-consuming and increases computational complexity almost quadratically [49]; and (c) assuming an unknown kernel in the downsampling process makes adopting a specific interpolated input, as an estimation of the output, unjustified. 2) As mentioned previously, most SR techniques undertake the assumption that the high-frequency content is redundant and can be accurately predicted from the low-frequency data [50]. Thus, exploring more contextual information within large regions of LR images to capture sufficient information for retrieving high-frequency details in predicted HR images seems inevitable. Theoretical work in DL show more contextual information can be achieved by designing very deep

architectures with larger receptive fields, which can result in expanding the final solution space [19, 51-54]. In some situations, effectively attaining more hierarchical representations can be achieved by increasing the DL network depth [51]. In recent years, many different CNN-based architectures have been developed, which exploit a very deep and sophisticated architecture, including residual and/or dense feature mapping [19, 54], to solve complex problems more efficiently [23, 42].

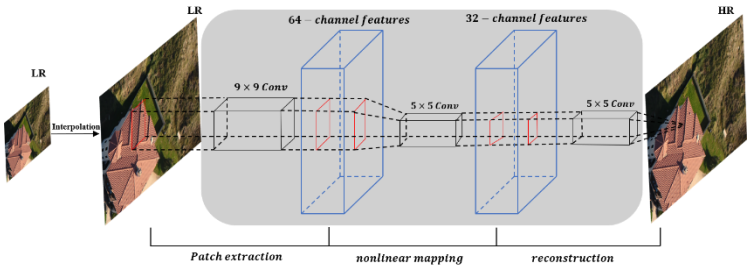


Figure 3.2. Sketch of the SRCNN architecture.

3.3.2. Generative Adversarial Network (GAN) for SISR

Introduction of recent innovative and deeper CNN-based architectures for SISR has already led to breakthroughs in accuracy and speed. Photo-realistic SISR GAN (SRGAN) [21], illustrated in Figure 3.3, was introduced for recovering the finer texture details when resolving at large upscaling factors. Those recovered fine details in SR images not only make predicted HR images more appealing to human, but also have a great impact on the accuracy and reliability of imaging geometry and scene details when they are retrieved by the SfM photogrammetry process.

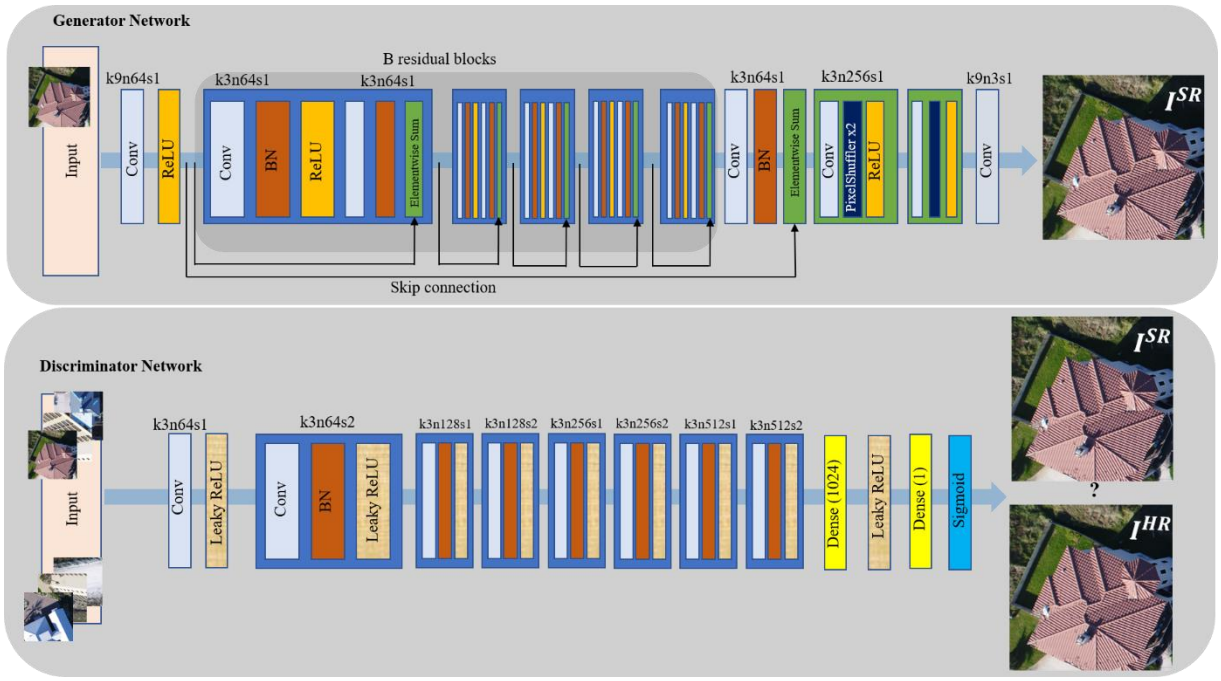


Figure 3.3. Architecture of Generator and Discriminator Network for SISR task with corresponding kernel size (k), number of feature maps (n), and stride (s) indicated for each convolutional layer.

The basic SRGAN model is built upon the residual blocks [19] and trained under the perceptual loss in a GAN framework, which makes it capable of predicting photo-realistic images for the upscaling factor of 4 [21]. The SRGAN model has shown significant improvement on overall visual quality of SR images over all previously introduced PSNR-oriented methods [21, 30].

GAN [29] introduced by Goodfellow et al. tries to solve the adversarial min-max problem [21]:

$$min_{\theta_G} max_{\theta_D} \quad \mathbb{E}_{I^{HR} \sim p_{train}(I^{HR})}\left[\log D_{\theta_D}(I^{HR})\right] \tag{4}$$

$$+ \mathbb{E}_{I^{LR} \sim p_G(I^{LR})}\left[\log\left(1 - D_{\theta_D}\left(G_{\theta_G}(I^{LR})\right)\right)\right]$$

where it allows the network to train a generative model G with the purpose of fooling a discriminator D that is simultaneously trained to discriminate the SR images from the original HR images.

The formulated perceptual loss consists of a weighted sum of a *content loss* ($\mathcal{L}_X^{SR}$) and an *adversarial loss* component ($\mathcal{L}_{Gen}^{SR}$) as follows [21]:

$$\mathcal{L}^{SR} = \mathcal{L}_X^{SR} + 10^{-3}\,\mathcal{L}_{Gen}^{SR} \tag{5}$$

Content loss motivated by perceptual similarity chooses the solution based on the perceptual similarity from the high dimensional solution space [21]. Instead of relying on pixel-wise losses, Ledig et al. define *VGG loss* based on *ReLU* activation layers and 19 layers VGG network [51], where *VGG loss* is computed as the Euclidean distance between the feature representations of a reconstructed image $G_{\theta_G}(I^{LR})$ and the ground truth image I^{HR} as follows [21]:

$$\mathcal{L}_{VGG/i,j}^{SR} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}\left(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}\left(G_{\theta_G}(I^{LR})\right)_{x,y}\right)^2 \tag{6}$$

where $\phi_{i,j}$ represents the feature map obtained by the *j-th* convolution (after activation) before the i-th maxpooling layer within the VGG-19 network. $W_{i,j}$ and $H_{i,j}$ describe the dimensions of the respetive feature maps within the VGG network.

Adversarial loss, which is the generative component of SRGAN to the perceptual loss, encourages the network to favor solutions residing on the natural image manifold [21]. The generative loss ($\mathcal{L}_{Gen}^{SR}$) is evaluated, in a probabilistic framework, based on the performance of the discriminator $D_{\theta_D}(.)$ over a training sample set as

$$\mathcal{L}_{Gen}^{SR} = \sum_{n=1}^{N} -\log D_{\theta_D}\left(G_{\theta_G}(I^{LR})\right) \tag{7}$$

where, $D_{\theta_D}\left(G_{\theta_G}(I^{LR})\right)$ represents the probability that the generated image $G_{\theta_G}(I^{LR})$ is a natural HR image. Because of exploiting adversarial loss, the discriminator network is trained to push SISR solutions to the natural image manifold.

3.4. Learning Strategies

Learning the end-to-end mapping function $\mathcal{F}$ to map a LR imageI^{LR} to the corresponding reconstructed SR image $I^{SR} = I^{HR}$, which is an approximation of the real HR image I^{HR}, requires the estimation of network parameters θ. This is attained via minimizing the loss between the super-resolved images $I^{SR} = \mathcal{F}(I^{LR}; \theta)$ and the corresponding HR images I^{HR}. In this section, different loss functions that are widely used in SISR techniques are introduced. For the sake of brevity, the subscript y is dropped from the ground truth (target) HR image I_y and the reconstructed HR image I_y in the rest of this section.

3.4.1. Pixel loss

Pixel loss evaluates the pixel-wise difference between two images, mainly in the form of L_1 distance, i.e., mean absolute error (*MAE*), or L_2 distance, i.e., mean square error (*MSE*). In so doing, it attempts to capture and solve the inherent uncertainty in retrieving lost high-frequency components by minimizing related loss functions as follows [42]:

$$\mathcal{L}_{pixel-L_1}(I^{HR}, I^{SR}) = \frac{1}{hwc}\sum_{i,j,k}\left|I^{HR}_{i,j,k} - I^{SR}_{i,j,k}\right| \tag{8}$$

$$\mathcal{L}_{pixel-L_2}(I^{HR}, I^{SR}) = \frac{1}{hwc}\sum_{i,j,k}\left(I^{HR}_{i,j,k} - I^{SR}_{i,j,k}\right)^2 \tag{9}$$

where h, w, and c are the height, width, and number of channels of the reconstructed images, respectively. Charbonnier loss [55, 56], is a variant of L_1 loss, given by [42]:

$$\mathcal{L}_{pixel-Cha}(I^{HR}, I^{SR}) = \frac{1}{hwc}\sum_{i,j,k}\sqrt{\left(I^{HR}_{i,j,k} - I^{SR}_{i,j,k}\right)^2 + \epsilon^2} \quad (10)$$

where ϵ is a small constant (e.g., 1×10^{-3}) for numerical stability.

The pixel loss constraint results in a super-resolved image I^{SR}, which is close to the ground truth HR image $\hat{I}^{HR}$ in the pixel values. In comparison with L_2 loss, the L_1 loss shows higher performance and better convergence [42, 57]. Using pixel loss as the loss function favors a high peak signal-to-noise ratio (PSNR). According to its definition, PSNR is heavily correlated with pixel-wise deviation, where minimizing pixel loss directly maximizes PSNR [21]. Moreover, it is partially related to the image perceptual quality. Thus, pixel loss has become the most widely used loss function in SR field.

Minimizing the pixel loss encourages finding plausible solutions, based on pixel-wise average, in the high dimensional solution space. In return, such solutions can be overly-smooth with poor perceptual quality [21, 58, 59]. Thus, in order to capture the reconstruction error and image quality more efficiently, a variety of other loss functions, such as content loss [59] and adversarial loss [21], were introduced to the SR field.

3.4.2. Perceptual/Content loss

To evaluate image quality based on perceptual similarity, perceptual-driven approaches have also been proposed [60, 61]. More convincing results from the image perceptual point of view, for both SR and artistic style-transfer tasks, are offered in this category [21, 61, 62]. By minimizing the error in the feature space instead of the pixel space, perceptual loss or content loss, attempts to improve the image visual quality. Denoting feature maps computed within the *l-th* layer of the network as $\phi^{(l)}(.)$, the content loss is evaluated using the Euclidean distance between corresponding feature maps from the original and super-resolved images as follows [42]:

$$\mathcal{L}_{content}(I^{HR}, I^{SR}; \phi, l) = \frac{1}{h_l w_l c_l} \sum_{i,j,k} \sqrt{\left(\phi_{i,j,k}^{(l)}(I^{HR}) - \phi_{i,j,k}^{(l)}(I^{SR})\right)^2} \tag{11}$$

where h_l, w_l, and c_l represent the height, width, and number of channels of the extracted feature maps in layer l, respectively.

Content loss encourages transferring the learned knowledge of hierarchical image features from a pre-trained classification network, usually VGG or ResNet, to the SR task [12, 21, 30, 63].

3.4.3. Adversarial loss

Adversarial learning [29] is adopted for SR task in a straightforward way, in which SR model is considered as a generator, and a discriminator network is added to the model to discriminate the generated image I^{SR} from the real image I^{HR}. Adversarial loss for SRGAN [21] is as follows [42]:

$$\mathcal{L}_{gan-G}\left(I^{LR}; D_{\theta_G}\right) = -\log D_{\theta_D}\left(G_{\theta_G}(I^{LR})\right) \tag{12}$$

$$\mathcal{L}_{gan-D}\left(I^{HR}, I^{SR}; D_{\theta_D}\right) = -\log D_{\theta_D}(I^{HR}) - \log D_{\theta_D}(I^{SR}) \tag{13}$$

where $\mathcal{L}_{gan-G}$ and $\mathcal{L}_{gan-D}$ denote the adversarial loss of the generator G_{θ_G}, which is the SR model, and the discriminator D_{θ_D}, which is a deep CNN model for binary classification, respectively. θ_G and θ_D are the parameters of the generator and discriminator, and $I^{SR} = G_{\theta_G}(I^{LR})$ is the generated image approximating the corresponding ground truth HR image.

In practice, some researchers employ a combination of multiple loss functions in their DCNN-based SISR architectures for more efficient learning and to better constrain different aspects of SR image reconstruction [12, 21, 55, 64, 65]. However, how to efficiently combine multiple loss functions with effective weights emphasizing their contribution in the learning process, remains an active area of SR research.

3.5. Image Quality Metrics

Image quality metrics, usually referred to as image quality measures (IQMs), are measures focusing on significant visual attributes of images where they attempt to quantify the perceptual assessments of an image when it is evaluated in a certain image quality assessment (IQA) approach [58]. IQA approaches are categorized into subjective methods, which focus on quantifying human perception, and objective methods, which are based on some computational models [58]. The subjective methods can be more accurate, but they are usually inconvenient, time-consuming, and expensive to implement [58]. As a result, objective methods are currently considered the mainstream among IQMs. Since the objective methods cannot efficiently capture the human visual perception, the metrics evaluated under those methods may show some inconsistency with those from subjective methods [58].

Objective IQA methods are divided into three types [58] including: (1) full-reference methods requiring corresponding images with perfect or high quality image content; (2) reduced-reference methods, which apply IQMs on the extracted features from both images and their corresponding high quality counterparts; (3) no-reference methods, which try to evaluate image quality in a blind way without any reference images. In supervised SISR, high quality HR images are usually available for evaluating different IQMs. This section introduces some of the most commonly used IQMs, covering both subjective IQA methods and objective IQA methods. **3.5.1. Peak Signal-to-Noise Ratio (PSNR)**

PSNR measure refers to ratio between a signal's maximum power and the power of the signal's noise, which affects the quality of signal's representation. Due to the very wide dynamic range (i.e., ratio of highest and lowest values) of most signals, the PSNR is usually expressed in the logarithmic decibel scale. PSNR is used to measure the reconstruction quality of lossy

transformations including image compression and inpainting. For image SR task, PSNR is defined using the maximum possible pixel value in the underlying image, and the mean squared error (MSE) between two corresponding images. Given the high-quality image I and the corresponding reconstructed (super-resolved) image $\hat{I}$, both of which include N pixels, the MSE and the PSNR measures are defined as follows [23]:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(I_i - \hat{I}_i\right)^2 \tag{14}$$

$$PSNR = 10\ log_{10}\left(\frac{L^2}{MSE}\right) \tag{15}$$

L denotes the maximum possible pixel value in the image. For $8 - bit$ image representations, for example, L equals to 255 and the typical values for the PSNR may vary from $20\ dB$ to $40\ dB$, where the higher the PSNR value, the better the quality of the reconstructed image as it tries to minimize MSE between the images with respect to the maximum pixel value of the input image. When L is fixed, the PSNR is only related to the pixel-wise distances between two images represented by MSE. The ability of MSE, and consequently PSNR, to capture perceptually relevant differences, such as high texture detail, is very limited meaning that the PSNR does not care about human visual perception and photo-realistic characteristics of the image. This often leads to poor performance of PSNR when used to assess the quality of super-resolved images in natural scenes. However, due to the lack of an efficient and comprehensive IQM that considers image quality from all perspectives, PSNR remains the most widely used metric for evaluating image quality in SR tasks.

3.5.2. Structural Similarity (SSIM) index

Similar to the human visual system, which is highly adapted for extracting structural information from the viewing scene, SSIM index provides a perceptual metric that quantifies

image quality degradation based on perceived image quality [66]. Made up of three relatively independent terms, luminance, contrast, and structure, SSIM index estimates the visual impact of those factors when they are modified in the reconstructed image. Those modifications may comprise shifts in image luminance, alterations in image contrast, and any other remaining deviations collectively identified as structural changes [58]. For an original high-quality image I and its reconstructed counterpart $\hat{I}$, the SSIM index is defined as follows [67]:

$$SSIM(I,\hat{I}) = [C_l(I,\hat{I})]^{\alpha}[C_c(I,\hat{I})]^{\beta}[C_s(I,\hat{I})]^{\gamma} \tag{16}$$

where α, β, and γ control the relative significance of each of the three terms of the index. In some implementations, $\alpha = \beta = \gamma = 1$ [58]. The luminance, C_l, contrast, C_c, and structural, C_s, components of the SSIM index are defined as follows [67]:

$$C_l(I,\hat{I}) = \frac{2\mu_I\mu_{\hat{I}} + C_1}{{\mu_I}^2 + {\mu_{\hat{I}}}^2 + C_1} \tag{17}$$

$$C_c(I,\hat{I}) = \frac{2\sigma_I\sigma_{\hat{I}} + C_2}{{\sigma_I}^2 + {\sigma_{\hat{I}}}^2 + C_2} \tag{18}$$

$$C_s(I,\hat{I}) = \frac{\sigma_{I\hat{I}} + C_3}{\sigma_I\sigma_{\hat{I}} + C_3} \tag{19}$$

where μ_I, σ_I and $\mu_{\hat{I}}$, $\sigma_{\hat{I}}$ represent the means and standard deviations of the original high-quality image and the corresponding reconstructed image, respectively, and $\sigma_I\sigma_{\hat{I}}$ is the covariance of the two images. The constants C_1, C_2, and C_3 in Eq.17-19 help to avoid instability when the denominators are close to zero. The formulation given in Eq.16 guarantees *symmetry*, where $SSIM(I,\hat{I}) = SSIM(\hat{I},I)$. Moreover, the index ensures a *bounded* $SSIM(I,\hat{I}) \leq 1$. Furthermore, there is a *unique maximum*, where $SSIM(I,\hat{I}) = 1$ if and only if $I = \hat{I}$. For an $8-bit$ grayscale image containing $L = 2^8 = 256$ gray levels, $C_1 = (k_1.L)^2$, $C_2 = (k_2.L)^2$, and $C_3 = C_2/2$, where

$k_1 \ll 1$ and $k_2 \ll 1$ are very small constants for avoiding instability. According to the above formulas, SSIM can be represented as follows [67]:

$$SSIM(I, \hat{I}) = \frac{(2\mu_I\mu_{\hat{I}} + C_1)(\sigma_I\sigma_{\hat{I}} + C_2)}{(\mu_I{}^2 + \mu_{\hat{I}}{}^2 + C_1) + (\sigma_I{}^2 + \sigma_{\hat{I}}{}^2 + C_2)} \quad (20)$$

In addition, to deal with uneven distribution of image statistical features or distortions, it is more reliable to perform image quality assessment locally rather than globally. Thus, mean structural similarity (mSSIM) [58] is proposed for locally assessing SSIM. This technique splits the images into multiple windows in which the SSIM of each window is evaluated, and finally averages it over all windows across the image. Because it evaluates the image reconstruction quality from the perspective of the human visual system, SSIM index better meets the requirements of perceptual assessment. The efficiency of SSIM-based IQM outperforms those based on MSE and the related PSNR over natural images including a wide variety of image distortions [67]. Those properties make SSIM index a widely used IQM among others in most SR tasks [68, 69]. However, in some cases, SSIM index may lead to similar results in evaluation of image performance with PSNR metric [58].

3.5.3. Task-based evaluation

Evaluating image reconstruction performance via other image analysis tasks is also an effective IQM [11-13, 70]. Specifically, this technique feeds the original high-quality image and the corresponding reconstructed image into a trained model for a specific vision task and evaluates the reconstruction quality by comparing the relative impact of reconstructed images on the prediction performance with respect to that from high quality original HR images. The vision tasks used for this evaluation technique include face recognition [71, 72], face alignment and parsing [63, 73], and object recognition [12, 74]. However, certain vision tasks may focus on some specific image attributes that are more favorable to the task and may not be aware or care about the visual

perceptual quality of the image. For example, most object recognition models mainly focus on the high-level semantics while ignoring the image contrast and noise. But on the other hand, in some domain-specific applications, such as super-resolving surveillance video for face recognition, task-based IQM may reflect the performance of the SR models.

3.6. Methods and Materials

In this SISR experiment, enhanced SRGAN (ESRGAN) [30] model is employed which improves the original SRGAN model in three aspects. First, ESRGAN improves the network by designing a Residual-in-Residual Dense Block (RRDB), illustrated in Figure 3.4, which offers higher capacity and easier training. Second, the Relativistic average GAN (RaGAN) [75], which learns to distinguish a more realistic image from a corresponding less realistic image, replaces the original discriminator in SRGAN, which simply judges whether an image is real or fake. According to [75], this improvement allows the ESRGAN generator to recover more realistic texture details. Third, ESRGAN adjusts the perceptual loss in the original SRGAN model by using VGG features before activation, rather than features after activation. This empirically leads to sharper edges and more visually pleasing results. Some properties of ESRGAN model are discussed below in more details.

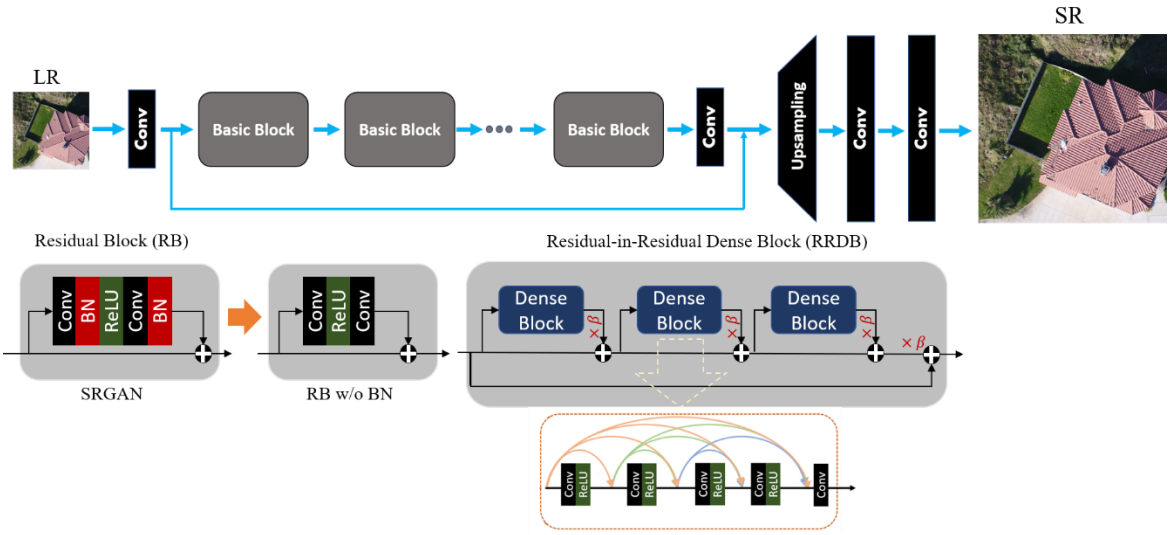


Figure 3.4. Basic architecture of SRResNet with different possible residual blocks.

3.6.1. Network architecture

Network Architecture: ESRGAN employs the basic architecture of SRResNet [21] for feature learning in the LR feature space. ESRGAN introduces two modifications to the generator architecture of SRGAN to improve the quality of the super-resolved images, G: (1) it removes all batch normalization (BN) layers; (2) it replaces the original basic residual block (RB) in SRGAN with a more compact RRDB architecture. According to Figure 3.4, by optimally combining multi-level residual blocks, the RRDB design improves the perceptual quality of super-resolved images [30]. When the statistics of image batches for training and testing are significantly high, BN layers tend to introduce unpleasant artefacts limiting the generalization ability [30]. Removing BN layers, especially under the GAN framework which is more prone to artefact generation, leads to consistent higher performance, lower computational complexity, and better generalization in the network [30, 57].

In addition to the architectural improvement, to facilitate training a very deep network, ESRGAN exploits residual scaling technique [53, 57] to prevent instability in training by scaling

down the residuals using a scaling factor between 0 and 1 before adding them to the main path. Moreover, ESRGAN employs a smarter initialization technique, which has empirically been shown to provide easier training when the initial parameter variance becomes smaller [30].

3.6.1.1. Relativistic discriminator

discriminator expressed as $D(I) = \sigma\big(C(I)\big)$, where σ is the sigmoid function and $C(I)$ is the discriminator output. This definition estimates the probability that the input image I is the original HR (real) image or the super-resolved (fake) image. In contrast, a relativistic discriminator predicts the probability that the original HR image I^{HR} is relatively more realistic than the super-resolved image I^{LR} as shown in Figure 3.5. The Relativistic average Discriminator (RaD) [75] is formulated as: $D_{Ra}(x_r, x_f) = \sigma\left(C(x_r) - \mathbb{E}_{x_f}\big[C(x_f)\big]\right)$, where D_{Ra} is RaD function and x_r and x_f are the real (original HR) and fake (super-resolved) images, respectively. $\mathbb{E}_{x_f}[.]$ represents average over all generated or fake images in each individual mini-batch. The discriminator loss, $\mathcal{L}_D^{SR}$, is defined as follows [30]:

$$\mathcal{L}_D^{Ra} = -\mathbb{E}_{I^{HR}}\big[log\big(D_{Ra}(I^{HR}, I^{SR})\big)\big] - \mathbb{E}_{I^{SR}}\big[log\big(1 - D_{Ra}(I^{SR}, I^{HR})\big)\big] \quad (21)$$

The adversarial loss for generator, $\mathcal{L}_G^{SR}$, is in a symmetrical form as

$$\mathcal{L}_G^{Ra} = -\mathbb{E}_{I^{HR}}\big[log\big(1 - D_{Ra}(I^{HR}, I^{SR})\big)\big] - \mathbb{E}_{I^{SR}}\big[log\big(D_{Ra}(I^{SR}, I^{HR})\big)\big] \quad (22)$$

where I^{LR} and $I^{SR} = G(I^{LR})$ stand for the input LR image and the predicted super-resolved image, respectively. In contrast to the adversarial loss for the generator in the original SRGAN model, $\mathcal{L}_{Gen}^{Ra}$ in Eq.7, that only gradients from the generated images take part in adversarial training, the adversarial loss for the generator in ESRGAN, $\mathcal{L}_G^{Ra}$ in Eq.22, contains both I^{SR} and I^{HR}. This property causes the gradients from both real images and generated images to participate in adversarial training [30].

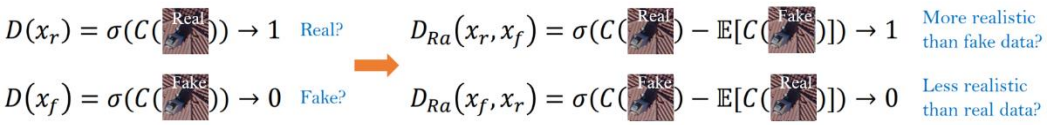


Figure 3.5. The standard (left) and relativistic (right) discriminators employed in the standard and relativistic GAN architectures, respectively.

3.6.1.2. Perceptual loss

ESRGAN suggests a more effective perceptual loss $\mathcal{L}_{percep}$ by computing distances between corresponding feature maps before activation rather than after activation, as practiced in the original SRGAN model. Employing features before the activation layers overcomes two drawbacks in the original design including extreme sparsity in the activated feature maps, and inconsistent brightness reconstruction compared with the original HR image. Specially within a very deep network, sparsity within feature maps leads to weak supervision and inferior performance. The loss function for the generator in ESRGAN model is as follows [30]:

$$\mathcal{L}_G = \mathcal{L}_{percep} + \lambda \mathcal{L}_G^{Ra} + \eta \mathcal{L}_1 \tag{23}$$

where $\mathcal{L}_1 = \mathbb{E}_{I^{LR}} \| G(I^{LR}) - I^{HR} \|_1$ is the content loss that evaluates the L_1 distance between super-resolved image $G(I^{LR})$ and the original HR image I^{HR}, and λ and η are coefficients to balance different loss terms.

3.6.2. IQMs for SR images

In this experiment, a comprehensive quantitative and qualitative assessment is performed on the resulting SR images by exploiting some standard IQMs that are frequently used for assessing the performance of different SISR models. Furthermore, a task-based IQM based on the SfM photogrammetry [76] procedure is carried out. Applying any type of image processing algorithm on a raw aerial image set can dramatically affect the precision and accuracy of retrieving

the interior and exterior geometry of camera at image acquisition time. That, consequently, may lead to a significant decrease in the quality and final accuracy of main SfM photogrammetry products, such as point clouds, DSMs, and orthoimages. The authors believe that the chosen task-based IQM can more accurately exhibit the effectiveness and performance of DCNN-based SISR to enhance the spatial resolution of LR imagery in RS applications. More specifically, where highly accurate spatial products from processing RS images are required.

3.6.2.1. Standard IQM methods

PSNR and SSIM index are evaluated as standard IQMs for quantitative assessment of predicted SR images. Choosing those two IQMs enables performance comparison in DCNN-based SISR applications when it is applied on two different categories of images (general images and aerial RS images).

3.6.2.2. SfM photogrammetry for task-based IQM

SfM photogrammetry procedure, as illustrated in Figure 3.6, is employed on all available image sets including HR ground truth, LR, and predicted SR image sets. SfM photogrammetry is a low-cost method, based on stereoscopic photogrammetry, for highly accurate topographic reconstruction using a series of overlapping images acquired from multiple viewpoints [76]. In contrast to traditional photogrammetry, in SfM photogrammetry, interior geometry of the camera, usually referred to as interior orientation (IO) parameters, positions and orientation of each camera station with respect to the scene's global coordinate system, commonly called exterior orientation (EO) parameters, and the geometry of the scene, i.e., the 3D coordinate of each point of the 3D scene, are resolved automatically. All required parameters are calculated simultaneously based on the highly redundant and iterative bundle adjustment (BA) computations using a rich database of corresponding image features automatically extracted from a set of multiple overlapping images

[77]. SfM photogrammetry addresses the key problem of determining the 3D locations of a large number of corresponding features extracted from multiple overlapping images, taken from different positions and angles with respect to the 3D scene.

Most image-based 3D reconstruction software that work based on the SfM photogrammetry principle, first solve for camera IO and EO parameters followed by a multi-view stereo (MVS) algorithm to escalate the density of the sparse point cloud generated by the SfM algorithm [76]. In the first step, several overlapping images are imported into the software, and a *keypoints* detection algorithm, usually the popular scale invariant feature transform (SIFT) algorithm [78], is applied to detect *keypoints* and *keypoint* correspondences across and between all images using a *keypoint descriptor*. In the SIFT algorithm, for example, the *keypoint descriptor* is determined by computing local image gradients and transforming them into a representation substantially insensitive to some image feature variations, including illumination, orientation, and scale [78]. These descriptors are unique enough to allow features to be matched in large image datasets. The BA technique is performed to minimize the errors in the phase of finding point correspondences [76].

In addition to solving for IO and EO parameters, which indicate camera calibration and pose parameters, respectively, the SfM algorithm generates a sparse point cloud using the image coordinates of all corresponding *keypoints*, IO, and EO parameters of the camera in all imaging station. The coordinate system related to the generated point cloud is arbitrary. In order to transform the point cloud coordinate system to any local or global coordinate system, a georeferencing phase should be adopted. In that phase, a few ground control points (GCPs) with known 3D coordinates in a local or global coordinate reference frame using land surveying or initial camera positions, e.g. using global navigation satellite system (GNSS), is required}. In this

experiment, it is not necessary to perform the georeferencing step since all images are processed in the same reference frame. The IO and EO parameters for each camera are used as the input to the MVS algorithm. Leveraging the known IO and EO parameters for each individual camera, MVS initiates an intense search algorithm to find more correspondences along all existing epipolar lines in all overlapping images. The accuracy of the MVS algorithm and the quality of the dense point cloud generated by the MVS algorithm is highly dependent on the reliability of the IO and EO parameters calculated from the initial BA computations [79].

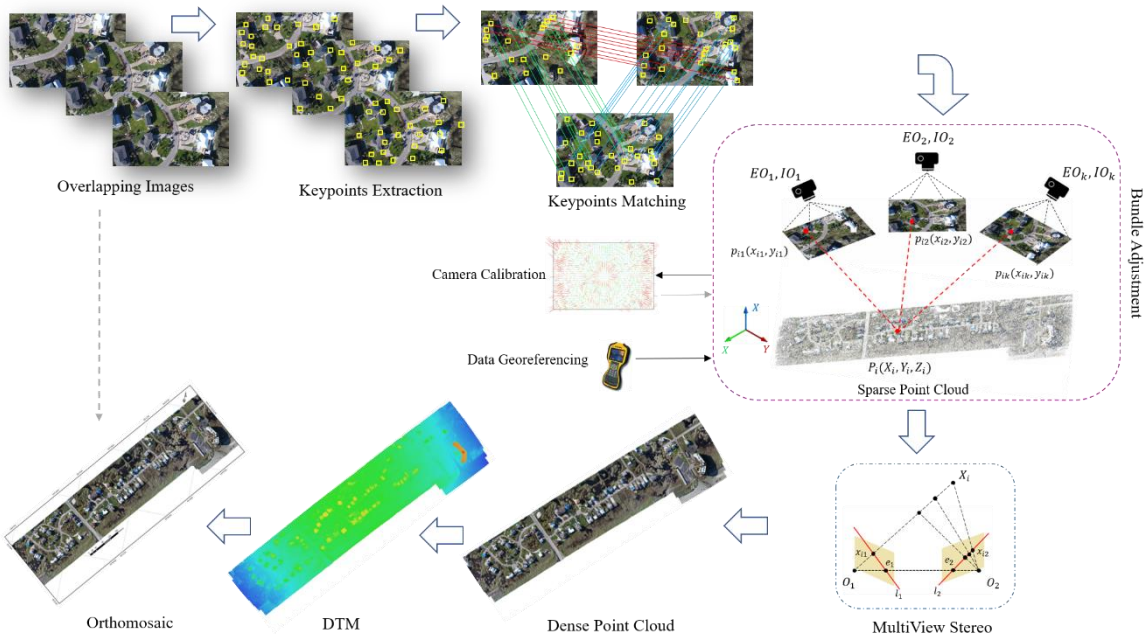


Figure 3.6. Steps of the SfM photogrammetry.

Images captured at high spatial resolutions, in general, return the most *keypoints* and *keypoints* correspondences in overlapping images. In addition to the major contribution of the natural texture in the 3D scene, the quality of the generated point cloud highly depends on several other factors including the density, sharpness, contrast, and resolution of the image content within the image set [76]. Moreover, decreasing the image acquisition distance, or flight height above ground, leads to an increase in the image spatial resolution or a finer GSD. This will further

enhance the spatial density and spatial resolution of the resulting point cloud [76]. However, the uncertainty in *keypoints* extraction and matching, which is a typical issue in all low quality LR images, may result in poor estimation of a camera's IO and EO parameters leading to a very inaccurate and erroneous 3D point cloud.

3.6.3. Study site and dataset

Port Aransas is a town located on Mustang Island along the southern Texas Gulf of Mexico coastline, USA Figure 3.7. In 2017, Hurricane Harvey, a category 4 hurricane, made landfall to the north of Port Aransas along San Jose Island on the night of August 25, 2017. The southern portion of the eye wall passed within close proximity to Port Aransas causing extensive damage, primarily due to extreme winds but also surge coming from the bay side of the island.

Figure 3.7. Port Aransas study site located along the southern Texas Gulf of Mexico coastline. The square box (top) shows the UAS flight area, which has been illustrated with more details in the UAS-derived ortho-image (bottom).

A few days after the landfall of Harvey, a small UAS photogrammetric survey was conducted over a section of the town directly bordering the Gulf-facing shoreline Figure 3.7. The purpose was to inspect and evaluate structural damages to residential and commercial properties caused by the catastrophic storm. The flight mission covers almost $0.275\ km^2$ of Port Aransas. Phantom 4 Pro multi-rotor UAS (SZ DJI Technology C.o., Ltd) was employed to conduct the survey. The platform was equipped with a 1 inch CMOS RGB sensor to capture 20 megapixel imagery at a resolution of 5472×3648 pixels. The flight altitude was designed to achieve a GSD of 2.5 cm, resulting in a flying height above ground level of about 90 m with forward lap and side lap around 80% and 70%, respectively. A total of 450 HR images were acquired over the study site. These images are used for the purposes of this study.

3.6.4. Data preparation and model training

In order to fine-tune pre-trained ESRGAN parameters with the existing dataset, 50 non-overlapping images were chosen from the original HR dataset as ground truth for fine-tuning ESRGAN during training phase. Scaling factor of 4 was set between LR and HR images. LR training images were obtained by down-sampling corresponding HR images. MATLAB bicubic kernel function was employed for image down-sampling, where its scale factor was set to 0.25. To make the SISR problem more complicated and realistic, additive white Gaussian noise with mean 0 and standard deviation of one-tenth of the standard deviation of each channel in RGB image was later added to the LR image set. Due to the high resolution of the original imagery, feeding the full-size images into the DCNN model rapidly exhausts the whole GPU's memory. However, in training phase, large image patches help very deep convolutional networks with wider receptive fields to capture more semantic information from the training samples. Therefore, this experiment was performed by extracting 1500 random image patches of resolution 1000×1000

pixels from the original HR images. Figure 3.8 illustrates HR image and corresponding ground truth HR image for a training sample. The model is trained in the RGB channels, and data augmentation with random horizontal flips and 90° rotations is employed on the training image set. Testing and evaluation of model performance is then done on 1000 image patches randomly extracted from the remaining 400 images in the original HR and corresponding LR image sets. It should be emphasized here that due to the large overlap between the employed UAS images, objects are sometimes captured by multiple images resulting in the appearance of the same object in the training and testing image sets. However, it should also be noted that such objects are captured from different viewing angles, causing different perspective and radiometric distortions for each specific object, or portion of the object, appearing in multiple images. Furthermore, the presence of such similar scenes within the training image set is necessary for performing transfer learning effectively, in which the weight parameters from a pre-trained DCNN model trained over a large dataset is applied to leverage complex mappings learned by very deep CNN models for performing a downstream task [80]. The weight parameters taken from the pre-trained model are, then, fine-tuned by training the model using a new dataset specific to the prediction task. In fact, one of the main reasons behind the transfer learning technique is to help the DCNN model to effectively capture a priori information related to the new task by fine-tuning the parameters of the underlying model using a new dataset for a different but related task. In the SISR technique, such a priori information can be provided to the SISR model by introducing information related to objects that are present in the acquired scene. Furthermore, the main goal of this study is to show the effectiveness of the SISR technique for recovering degraded or lost image details in the LR UAS images by fine-tuning a DCNN-based SISR model on a very limited set of HR UAS images.

The original ESRGAN model, before fine-tuning, is also employed to investigate the capability of the pre-trained ESRGAN, to enhance the image content and downgrade the inherent noise in the original HR images. The idea is that such a pre-trained model, trained on some standard datasets, may be capable of capturing the behavior of some types of noise that might be common in many imaging systems. To do this experiment, the original HR image set is fed to the original pre-trained ESRGAN with scaling factor of 1.

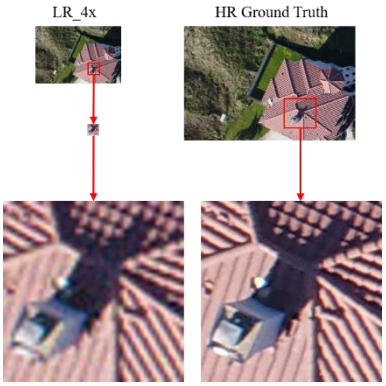


Figure 3.8. LR and corresponding HR image patches.

The Pytorch [81] implementation of ESRGAN model was chosen for training over the UAS dataset. The training process starts by initializing the ESRGAN model with weights from the pre-trained network trained on some of the well-known benchmarks in SISR such as the DIV2K dataset [82], the Flickr2K dataset [83] and the OutdoorSceneTraining (OST) dataset [64], which include thousands of high-quality HR images with a broad diversity in texture and contextual information. The performance of the trained model has already been tested on widely used SR benchmarks such as Set5 [45], Set14 [47], BSD100 [84], Urban100 [85], and the PIRM self-validation dataset [86]. Table 3.1 summarizes the information related to the ESRGAN model setup and optimization settings for training the model on the UAS image set. According to the table, dense block architecture for generator was set to $64 \times 5 \times 5$, which includes 64 kernels of size

5×5. The generator is comprised of 23 residual-in-residual dense blocks (RRDBs). The learning rate α was set to 0.0001, and Adam optimizer was chosen for updating weights during training. Two exponential decay rate parameters in Adam optimizer β_1 and β_2, were set to 0.9, and 0.999, respectively. ε parameter in the optimization algorithm was set to 1×10^{-7} to avoid any division by zero. The experiment was carried out with 100 epochs on Google Colab, *Google's free cloud service*, with one Intel(R) Xeon(R) CPU $2.30\ GHz$ and one high-performance Tesla $K80$ GPU, having 2496 CUDA cores and 12 GB GDDR5 VRAM. Fine-tuning the network took around 48 hours and inference time for predicting the super-resolved image was 10 seconds/image.

Table 3.1. ESRGAN model and training parameters setup.

Dense block	RRDB	Learning rate	Adam optimization parameters
$64 \times 5 \times 5$	23	$\alpha = 0.0001$	$\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 \times 10^{-7}$

3.7. Results

This section provides comprehensive qualitative and quantitative experimental results on predicted super-resolved, SR_{pre}, images from LR images, virtually downsampled form original (ground truth) HR, HR_{gt}, UAS image set with additive white Gaussian noise. Also, the result of applying ESRGAN model on HR_{gt} with scale factor 1, as an image enhancement network, to generate enhanced HR images, HR_{enh}, is investigated. Furthermore, the results of the task-based IQM using the SfM photogrammetry procedure implemented with the original and super-resolved imagery is reported.

3.7.1. Qualitative results

Figure 3.9 illustrates the qualitative assessment of the SISR performance using ESRGAN model on two different test samples. According to the visual inspection, and as observed in Figure 3.9, the ESRGAN model is able to upscale the LR images by factor 4 and predict SR images with

high similarity in perceptual and visual quality when they are compared with the corresponding HR counterparts. A closer look at the qualitative results in this experiment reveals some noise removal properties learned within the SISR model trained on a sufficient number of LR and corresponding HR images.

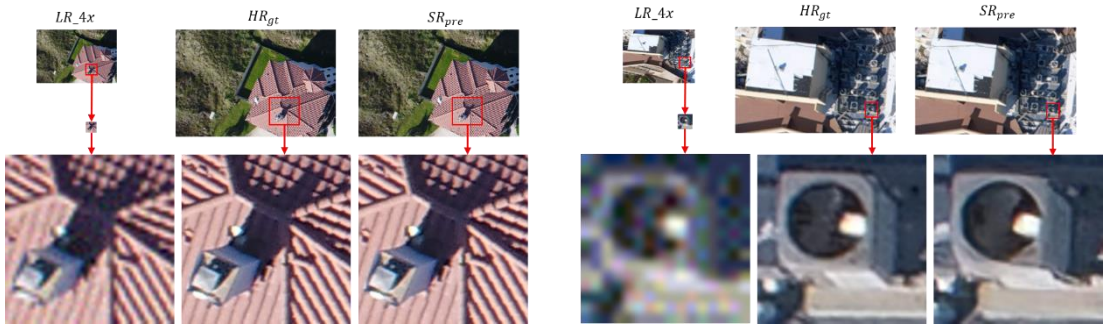


Figure 3.9. Illustration of the qualitative comparison between the predicted SR image and corresponding LR and ground truth HR images for two test images.

3.7.2. Quantitative results

For quantitative evaluation of the SISR performance, in this experiment with ESRGAN model, PSNR value and SSIM index were calculated for test image set and the enhanced HR (HR_{enh}) image set. Table 3.2 illustrate the lowest, highest, and average PSNR values and SSIM indices for both image sets. The range of values for both PSNR and SSIM index in Table 3.2, resulting from evaluating ESRGAN performance on SR_{pre} image set, is comparable in values reported for those IQMs when ESRGAN, or any other high-performance DCNN-based SISR model, is applied on standard SISR image sets [21, 23, 30]. The values of the standard IQMs represented in Table 3.2 confirm that SISR can be effectively applied for recovering lost or degraded details in LR UAS imagery, and hopefully on a wide range of imagery in RS applications, including aerial and satellite imagery, with a comparable performance.

Table 3.2. PSNR and SSIM index calculated on image sets.

Image set	Minimum PSNR/SSIM	Maximum PSNR/SSIM	Mean PSNR/SSIM
SR_{pre}	25/0.67	32/0.90	28/0.85
HR_{enh}	43/0.91	49/0.99	82/0.96

3.7.3. Task-based IQM and related results

Further investigation of ESRGAN model performance in a task-based image quality evaluation using SfM photogrammetry reveals more about the impact of image super-resolving on the internal and external geometry of imagery and the geometry of the reconstructed $3D$ scene. All available UAS image sets including the downsampled noisy LR image set (LR), the original ground truth HR image set (HR_{gt}), the predicted super-resolved image set (SR_{pre}), and enhanced HR image set (HR_{enh}) were separately imported to Agisoft Metashape software [87] for SfM photogrammetric processing. Each image set was processed using the exact same settings and workflow procedure to ensure a fair comparative evaluation could be made on the impact of SR imagery to the BA computations and $3D$ reconstruction (i.e., point cloud).

BA computations, using *keypoints* extracted from each individual image in each image set, also result in an accurate estimation of camera calibration (IO) parameters in a self-calibration procedure using a pre-defined camera calibration model. Camera parameters evaluated within BA computations include the focal distance f, principal pint coordinates (c_x, c_y), radial distortion coefficients (k_1, k_2, k_3, k_4), decentering distortion coefficients (p_1, p_2, p_3, p_4), and affinity and Skew transformation coefficients (B_1, B_2), which represent a specific distortion in digital imaging sensors accounting for scale distortion and non-orthogonality of pixel elements in the x, and y directions of the digital sensor [88]. Table 3.3 illustrates the camera calibration results for *LR*, HR_{gt}, SR_{pre}, and HR_{enh} UAS image sets. It should be noted that the reported results are based on running the BA computations once on each individual image set using the full camera calibration

model. Results for camera parameters evaluated with relatively high uncertainty, including k_2, k_3, k_4, p_3, p_4, B_1, and B_2 have not been reported in the table. According to Table 3.3, the evaluated values of IO parameters for SR_{pre} image set, especially, the sensor element (or pixel) size, focal distance, f, principal point offset c_x, c_y, and the first coefficient of radial lens distortion, K_1, which are among the most critical camera calibration parameters, closely approximate the real values derived from HR_{gt} image set. Referring to Table 3.3, the calibrated IO parameters for LR image set are different from IO parameters for HR_{gt}, SR_{pre}, and HR_{enh}, meaning that the parameters defining the internal imaging geometry in LR UAS image set is different than those in the other HR UAS image sets. It should be emphasized here that the number of selected *keypoints* and the level of certainty in finding their correspondences in multiple images within an image set can have a significant impact on the stability of BA computations and the accuracy of the estimated IO and EO parameters.

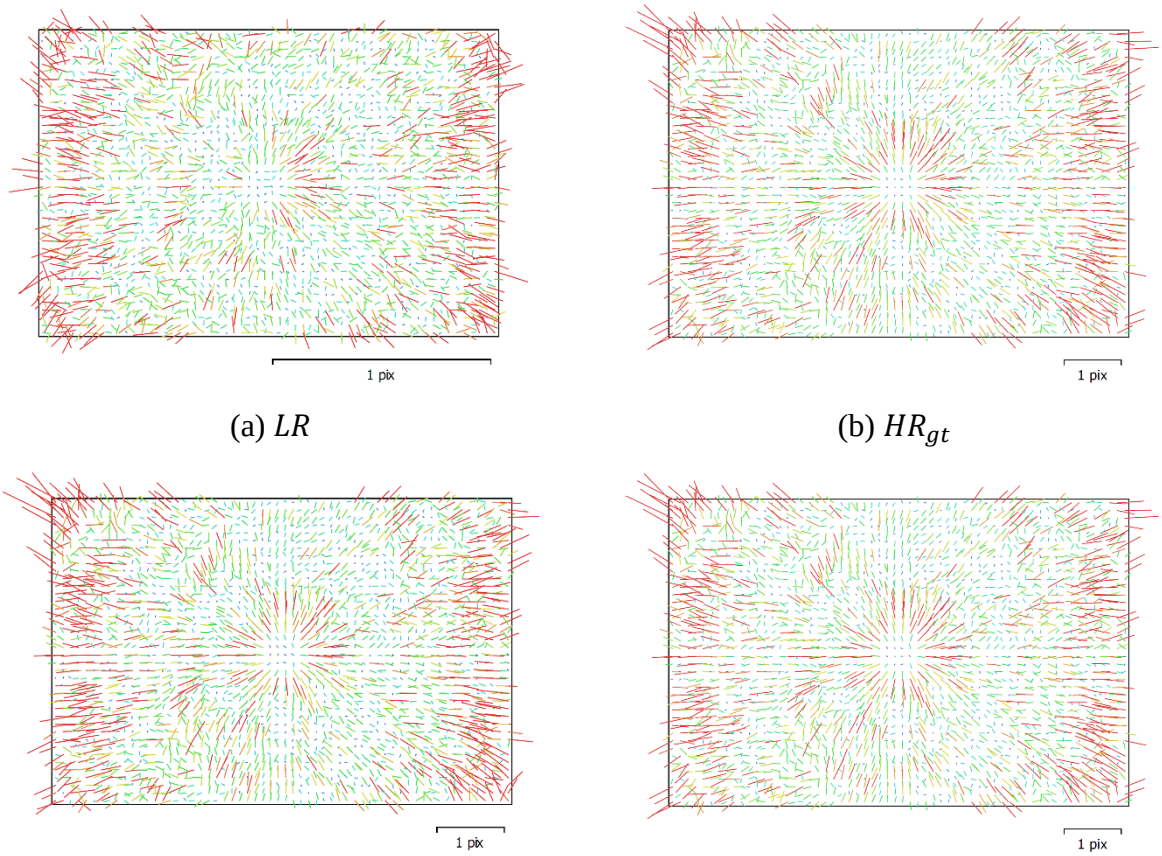
Table 3.3. Camera calibration results.

Parameters	LR	HR_{gt}	SR_{pre}	HR_{enh}
Pixel size (mm)	0.0096	0.0024	0.0024	0.0024
F $(pixel)$	911.785	3689.370	3701.798	3681.261
c_x $(pixel)$	-0.988	-49.869	-57.713	-40.43
c_y $(pixel)$	0.727	-13.880	-16.251	-15.321
k_1	0.007	0.005	0.007	0.004
p_1	0.002	-1.707×10^{-5}	-2.815×10^{-5}	-1.603×10^{-5}
p_2	0.001	-1.022×10^{-5}	-1.478×10^{-5}	-1.020×10^{-5}

Figure 3.10 displays plots representing the average reprojection error vectors from BA computations across the image space for LR, SR_{pre}, HR_{enh}, and HR_{gt} UAS image sets. This error quantifies the distance between a certain *keypoint* location on an image and the location of the

corresponding $3D$ point reprojected on that image. The magnitude of reprojection error in the image space depends on the quality of estimated camera calibration parameters and pose parameters, as well as on the quality of the extracted *keypoints* on each individual image [87]. Maximum and RMS of reprojection errors across the image space, and the average camera location errors with respect to the $3D$ scene have been depicted in Table 3.4 for LR, HR_{gt}, SR_{pre}, and HR_{enh} image sets. According to the table, both the maximum and RMS of the reprojection errors in SR_{pre} image space are closely comparable with those derived from HR_{gt} image set. The errors related to the quality of the $3D$ space, reconstructed by SR_{pre} image set, confirm the same quality in scene reconstruction when HR_{gt} image set is employed. In addition, Figure 3.11 illustrates a graphical view of the camera locations and their errors represented by the error ellipsoids for all UAS image sets.

The process of point cloud densification was carried out on each individual UAS image set after BA computations and digital surface models (DSMs) were later generated from the $3D$ point cloud data by the post-processing within SfM photogrammetry software. Figure 3.12 displays the dense point cloud over a small area of the study site for all UAS image sets. Moreover, Table 3.5 summarizes the processing report from SfM photogrammetry for each individual image set. According to Figure 3.12 and Table 3.5, visual and quantitative inspections on the density of the resulting dense point cloud, which is the average number of points per square meter, demonstrate that the dense point cloud generated from HR_{gt}, SR_{pre}, and HR_{enh} are about 17 times denser than the dense point cloud generated from the LR image set.



(a) LR (b) HR_{gt}

(c) SR_{pre} (d) HR_{enh}

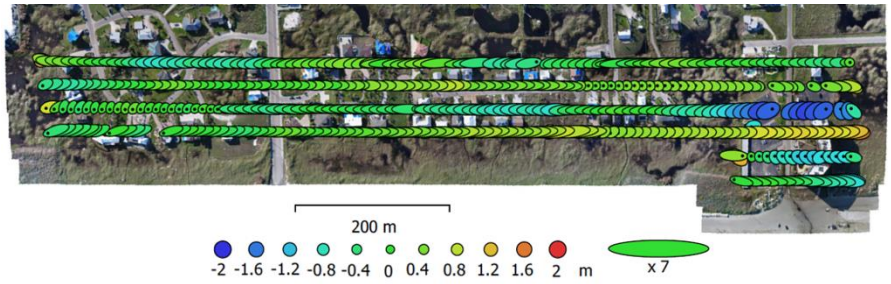
Figure 3.10. Average reprojection error vectors plotted on image space for LR, HR_{gt}, SR_{pre}, and HR_{enh} image sets. Colors of the error vectors represent increasing magnitudes of the reprojection error progressing from blue to red, respectively. The scale bar at bottom shows the magnitude of the error vector in pixel units.

Table 3.4. Bundle adjustment results for reprojection and camera location errors.

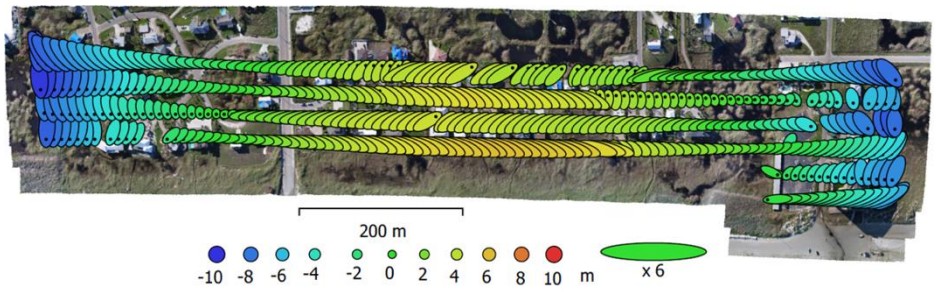
Image set	LR	HR_{gt}	SR_{pre}	HR_{enh}
Max. reprojection error ($pixel$)	15.901	56.960	57.210	55.050
Reprojection error ($pixel$)	0.498	0.787	0.993	0.635
X-error (m)	1.770	2.400	2.417	2.324
Y-error (m)	2.322	2.663	2.669	2.399
Z-error (m)	0.550	4.341	4.183	3.990
XY-error (m)	2.920	3.586	3.601	3.503
Total error (m)	2.972	5.631	5.520	5.420

Table 3.5. SFM photogrammetry report summary for different image set.

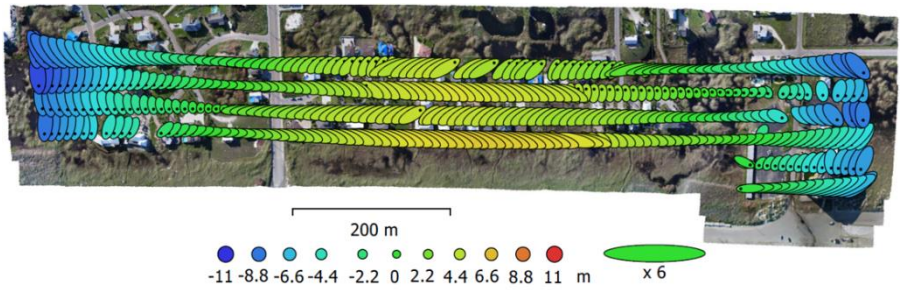
Parameters	LR	HR_{gt}	SR_{pre}	HR_{enh}	LR to SR_{pre}	HR_{gt} to HR_{enh}
Num. of images	440	440	440	440	0.0%	0.0%
Flying altitude (m)	106	106	107	106	0.9%	0.0%
Tie points (pts.)	1,398,877	11,051,665	8,268,475	11,630,227	490.0%	5.2%
Dense cloud (pts.)	1,805,966	31,041,604	31,052,606	31,940,817	1619.4%	2.8%
Point density (pts./m^2)	5.82	94.5	94.4	94.9	1521.9%	0.4%
DSM resolution (cm/pix.)	41.40	10.30	10.30	10.30	75.1%	0.0%



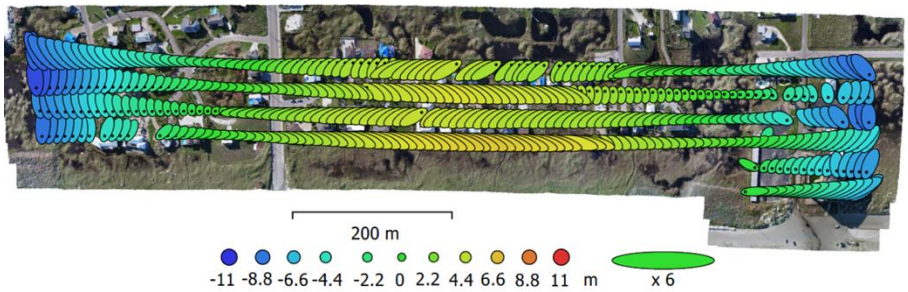
(a) LR



(b) SR_{pre}

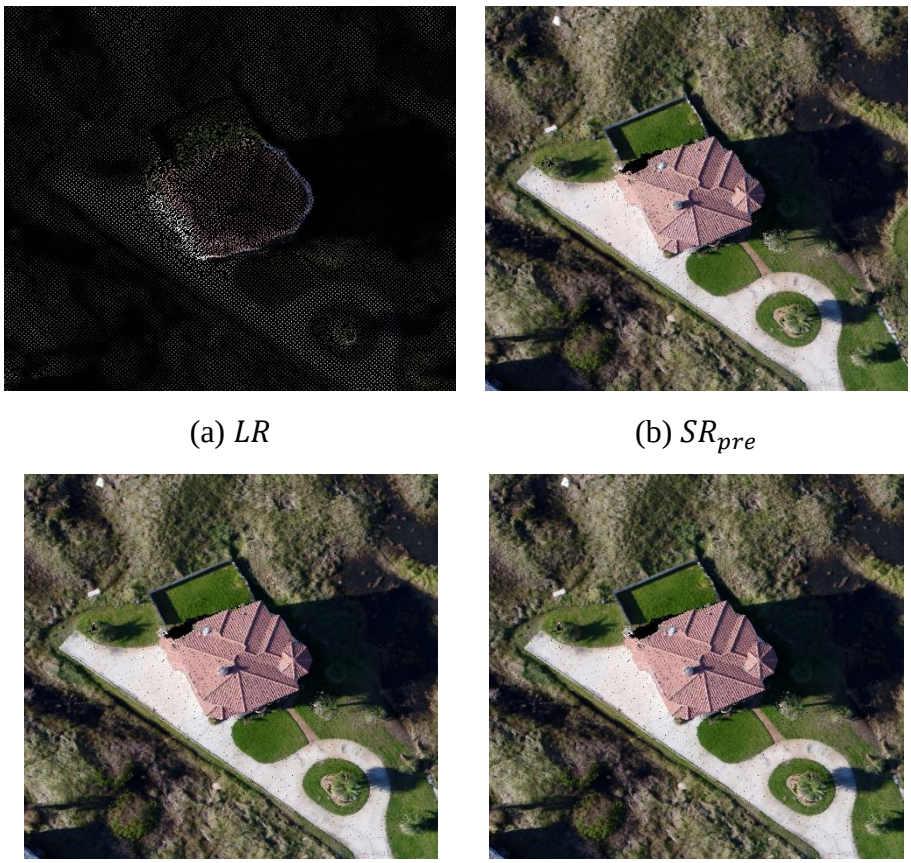


(c) HR_{enh}



(d) HR_{gt}

Figure 3.11. Camera locations and related uncertainties for LR, SR_{pre}, HR_{enh}, and HR_{gt} image sets. Ellipse color represents Z error. Errors in X and Y directions are represented by ellipse shape. Black dot within each individual ellipse represents estimated camera locations.



(a) LR (b) SR_{pre}

(c) HR_{enh} (d) HR_{gt}

Figure 3.12. Resulting dense RGB point cloud computed within the SfM photogrammetry process using LR, SR_{pre}, HR_{enh}, and HR_{gt} image sets.

To investigate how closely the DSM generated based on the SR_{pre} image set approximates the corresponding DSM generated from HR_{gt} image set, DSM from SR_{pre} was subtracted from the DSM generated from HR_{gt} image set. Figure 3.13 displays the resulting differential surface.

Referring to Figure 3.13, the average height difference between the two DSMs is about -0.5 cm. However, there are some areas showing large height differences. These areas are mostly related to the edges of tall man-made and natural objects. Areas with lack of texture, such as water bodies, also contribute to the large height differences observed in Figure 3.13. The histogram in Figure 3.14 displays a statistical representation of the pixel-wise height differences based on the frequency of occurrence for pixel values in differential DSMs after filtering blunders.

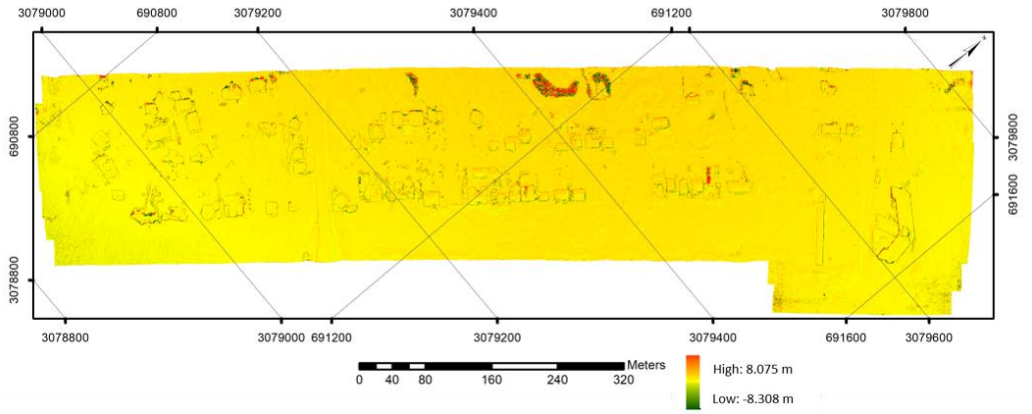


Figure 3.13. Illustration of DSM difference between HR_{gt} and SR_{pre} image sets.

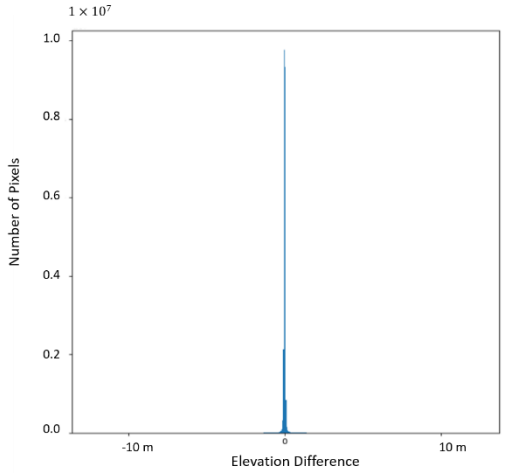


Figure 3.14. Illustration of height-difference histogram derived by subtracting DSM for SR_{pre} image set from DSM for HR_{gt} image set.

3.8. Discussion

Visual inspection of mage samples in SR_{pre} and corresponding HR_{gt} image sets confirms that the ESRGAN model performs much better over man-made objects and natural objects with definite boundaries than other targets, as shown in Figure 3.9. One reason may be due to the fact that natural objects usually comprise extremely intricate structures and severely random patterns with very fine details. In addition, natural objects, such as vegetation, may be moving due to the wind during image acquisition in an outdoor environment, inducing dynamic image motions in the recorded images. More accurate visual inspection on SR_{pre} images demonstrates that the model is able to predict super-resolved images with lower level of noise and blur when they are visually compared with the corresponding HR_{gt} images. This noise reduction property of the model, however, in some natural targets, such as vegetated areas may result in removing unpleasing pseudo-noise patterns within those areas. This noise reduction capability in ESRGAN model is more evident over man-made structures and surfaces as illustrated in the right example of Figure 3.9.

Such image enhancement and noise removal characteristics can also be observed on both natural and man-made objects that appear in HR_{enh} image set, where the HR_{gt} images were used as input and the naive pre-trained SISR model, with scale factor 1×1, was used as an image restoration network. This observation demonstrates that pre-trained ESRGAN, on several standard image sets for SISR, has been able to capture, to some extent, the behavior of some types of noise that are common in almost all digital imaging systems. Considering the fact that this model has already been trained to predict SR images with scale factor 2 and 4, the observations with scale factor 1 divulges that there might be some types of noise that may commonly appear in different image scales where the pre-trained network has been able to differentiate them from the real signal.

The high IQM values reported for the HR_{enh} image set in Table 3.2 is due to the high degree of similarity in image content and quality between corresponding images in HR_{enh} and HR_{gt} image sets. This observation demonstrates that pre-trained ESRGAN can be used as an image restoration network when it is employed with scale factor 1.

It is worth mentioning that employing pre-trained ESRGAN, without fine-tuning the parameters using LR and corresponding HR_{gt} UAS image sets for predicting the super-resolved images (SR_{pre}), decreases the model performance around 15% for both PSNR and SSIM index in this experiment. The relatively high values for those standard image quality metrics on SR_{pre} UAS image set, whose contents are intrinsically different from those on which the vanilla ESRGAN model has been trained, verifies that the transfer learning technique and fine-tuning of the pre-trained parameters significantly helps the DL-SISR model to extract more related semantic information from the UAS images. This information is optimally encoded as abstract information within multiple layers of a DCNN-SISR model. Interestingly, according to Table 3.2, the vanilla ESRGAN model trained on standard image sets, resulted in high values for PSNR and SSIM index when it was employed on the HR_{gt} image set as an image restoration network. This is regardless of the fact that the model did not previously see the UAS images for which it has been employed to predict on in this experiment.

Results of the task-based IQM using SfM photogrammetry adds more to the previous findings. Referring to Table 3.3, calibrated sensor element size, or image pixel size, for *LR* images is about 4 times bigger than that for images in other image sets, which is compatible with our experiment. The calibrated focal lengths in SR_{pre} and HR_{enh} image sets closely approximate the real focal length evaluated in HR_{gt} ground truth image set. The difference in calibrated focal length for LR, SR_{pre}, and HR_{enh} image sets from the calibrated focal length for HR_{gt} image set are

-0.010 mm, -0.030 mm, and 0.020 mm, respectively. Furthermore, calibrated c_x, and c_y values show an accurate estimation of the principal point location in SR_{pre} images with respect to the HR_{gt} images. For LR images, however, those calibrated parameters show a very different location for the principal point in LR image space.

Referring again to Table 3.3, the remaining calibration parameters, including radial and decentering lens distortion coefficients, affinity, and skew transformation parameters in SR_{pre} and HR_{enh} image sets show a high degree of compatibility with HR_{gt} parameters confirming that lens distortion parameters and other sensor related distortions can be accurately estimated in both super-resolved SR_{pre} images and restored HR_{enh} images. However, interpreting the values of those coefficients, especially between LR and HR_{gt} images, is not very meaningful because some of them are usually highly correlated with other parameters, especially the focal length, principal point location, and the first coefficient of radial lens distortion [88, 89].

Referring to Figure 3.10, the behavior of the average reprojection error in SR_{pre} image space accurately approximates that in the original HR_{gt} image space. This finding can be supported further by our above findings when referring to the calibrated camera parameters, where results showed that the internal geometry of the sensor can be accurately recovered in the SR_{pre} images. The plot related to the average reprojection error in LR image space represents less similarity with the error behavior in HR_{gt} and SR_{pre} image space, especially in the center of the image space. On the other hand, the average reprojection error plot for HR_{enh} image space (Figure 3.10-d) is very similar to the reprojection error plot for the HR_{gt} image space (Figure 3.10-b). This observation demonstrates that image restoration processing carried out on the HR_{gt} images within the pre-

trained ESRGAN has not meaningfully changed the IO parameters of the camera derived from the SfM analytical self-calibration procedure.

According to Table 3.4, investigation on maximum reprojection error and its RMS in the SR_{pre} and HR_{enh} image spaces shows that they closely approximate those values in the HR_{gt} image space with sub-pixel magnitudes. However, RMS of reprojection error in HR_{enh} image space is about 20% less than it is in HR_{gt} image space. Part of this decrease in reprojection error might be due to the noise reduction process in HR_{enh} image space with respect to the original HR_{gt} image space. Referring to the average camera location errors in Table 3.4, SR_{pre} and HR_{enh} image sets closely approximate those in the original HR_{gt} image set. This suggests that the SISR process employed with factor 4 on the LR image set and employed with the image restoration process on HR_{gt}, preserves the external imaging geometry with respect to the $3D$ scene. As depicted in Table 3.4, pre-trained ESRGAN model with scaling factor 1, as image restoration network, resulted in 3% improvement on total error in camera positions for HR_{enh} image set. There is also 2% improvement in that error for SR_{pre} dataset. Figure 3.11 shows that camera locations and their positional errors in the HR UAS imagery can be accurately retrieved in the predicted SR image set. Furthermore, it shows that image enhancement performed with the employed pre-trained ESRGAN model does not dramatically change the external imaging geometry.

Carefully exploring the differential DSM in Figure 3.13 reveals that such areas include natural and man-made water bodies with lack of texture and the edges of tall natural and man-made structures. Filtering out those areas from the original differential DSM and calculating some statistics over them shows that the minimum, maximum, and standard deviation (SD) of height difference in those areas are -8.308 m, 8.075 m, and 30 cm respectively. The height-difference

histogram in Figure 3.14, for filtered differential DSM, confirms that the geometry of the reconstructed 3D scene, as reflected by the DSM, can be accurately retrieved with a SD around 2.50 cm. The minimum, maximum, and mean of height-differences within the filtered differential DSM are about -4.85 cm, 5.73 cm, and -0.02 cm, respectively.

It is worth mentioning that there are numerous environmental and sensor-related factors as well as flight design parameters which contribute to the quality and the spatial resolution of images captured by the UAS. Texture quality, related to each individual object in the scene, can highly affect the training and inference phases of the DCNN-based SISR model, which subsequently affects the results of the SfM process. Ambient environmental conditions, such as lighting or any instability of the platform during image capturing, such as due to the wind, can impact the above results. Similarly, flight design including altitude above ground and camera perspective (e.g., oblique versus nadir) will impact the GSD and appearance of land cover features. As a result, the visual representation of the same target may deviate from one exposure to another in a single UAS flight mission and across repeat data acquisitions. Thus, the authors emphasize that the results shown here, are valid for the specific data set acquired at a certain time over the certain study site. The results presented here, in terms of reconstruction accuracy, cannot be necessarily generalized to other sites with very different targets and textures, or the same area imaged at a different time and during different environmental conditions, without further experimentation. However, we believe that the high capacity of deep CNN models to efficiently extract informative contextual features from the raw UAS images in an end-to-end manner have the potential to be extended further by training DCNN-based SISR models using a time-series of UAS images acquired over the same area, or UAS images captured from the same area under different weather conditions. Also, training and evaluating the performance of a certain DCNN-based SISR model on multiple

UAS image sets including images from different areas with a wider range of targets and varying textures may be considered for further analyses.

3.9. Conclusion

SISR seeks to obtain HR images from corresponding LR images, which is a notoriously arduous and ill-posed problem. Investigating different IQMs evaluated on SR images predicted from corresponding LR images in a DCNN-based SISR network revealed two important findings with respect to this study's experiment on UAS imagery. First, the quantitative measures of image quality, including PSNR and SSIM index, applied to the super-resolved UAS imagery, confirm that the DCNN-based super-resolution technique employed here (ERSGAN architecture) can achieve the same level of performance for spatial-resolution and pictorial information enhancement relative to the original HR ground truth image set. Both quantitative and qualitative assessment of SR images showed that the level of additive white noise to the LR image remarkably decreases in the SR image. Furthermore, visual comparison of SR images with corresponding HR images in some areas showed that the SR image may exhibit less amount of noise.

The second important finding relates to the task-based IQM performed using SfM photogrammetry. Results confirmed that the geometry of UAS image acquisition can be recovered in SR images with high accuracy. Camera interior and exterior parameters, evaluated by processing SR images in auto-calibration module within the SfM photogrammetry procedure, closely approximate the original results derived from the same procedure on the ground truth HR images. Preserving the geometry of imagery can significantly increase the reliability of using super-resolution techniques in many different RS applications, specifically where extracting spatial information from RS images is required. The densified point cloud generated by SfM photogrammetry on the SR UAS images is about 15 times richer than the point cloud generated

from the artificially degraded LR UAS images, which provides more details about the underlying terrain. Furthermore, the differential DSM and related height-difference histogram show the STD around 2.5 cm, which confirm the closeness of two reconstructed surfaces from SR and HR image sets.

Overall, results from this study's experiment on UAS imagery show that DCNN-based SISR enhancement techniques can exploit spatial and non-spatial information in LR and HR imagery for effectively discriminating the signal from noise in image space resulting in high performance in recovering image details and more visually appealing images for different RS applications. For example, one practical application of the SR technique for UAS mapping is that it can potentially enable flights at higher altitudes and lower GSDs to cover more area in a certain time duration, thereby leading to more flight efficiency. Then, a DCNN-based SISR technique, such as presented in this study, could be applied to super-resolve the imagery to a specific resolution and generate a dense point cloud from SfM photogrammetry, and subsequently DSM or orthoimage, as though the data were acquired from a UAS flight conducted at a lower altitude and with similar quality.

Future work will seek to investigate the real scenario of employing SISR to reduce UAS image acquisition flight time for aerial surveying operations when mapping of a relatively large area at high resolution is demanded. This will be investigated by employing two UAS image sets acquired at two different altitudes over the same area. Performance of the DCNN-based SISR model to super-resolve the LR (high altitude) images can then be assessed by comparing SfM processing results with the super-resolved LR images and original HR (low altitude) images in terms of 3D reconstruction fidelity and image quality. Furthermore, for more accurate comparison between the geometry of the original HR image space and predicted HR (super-resolved) image

space, interior camera parameters and uncertainties involved in recovering them need to be carefully analyzed. Accurate estimation of camera parameters, uncertainty assessment, and sensitivity analysis of those parameters play an important role in high-accuracy measurement and object localization in 3D scene. The effect of different lighting and environmental conditions, and the impact of different study sites with different objects of varying textures, on model performance may also be explored. Another possibility for further exploration related to the above experiment} is to employ hyper-spatial resolution UAS imagery for training a DCNN-based SISR model in order to improve the spatial resolution of satellite imagery. Finally, examining the most optimized DCNN-based SISR techniques, with the lowest time-complexity in training and inference phases, might be a topic of great interest where it can help pave the path for integration of SISR into real-time remote sensing application scenarios.

3.10. References

1 Stumpf, R.P., Holderied, K., and Sinclair, M.: 'Determination of water depth with high-resolution satellite imagery over variable bottom types', Limnology and Oceanography, 2003, 48, (1part2), pp. 547-556

2 Aizawa, K., Komatsu, T., and Saito, T.: 'Acquisition of very high resolution images using stereo cameras', in Editor (Ed.)^(Eds.): 'Book Acquisition of very high resolution images using stereo cameras' (International Society for Optics and Photonics, 1991, edn.), pp. 318-328

3 Dare, P.M.: 'Shadow analysis in high-resolution satellite imagery of urban areas', Photogrammetric Engineering & Remote Sensing, 2005, 71, (2), pp. 169-177

4 Yang, J., Wright, J., Huang, T.S., and Ma, Y.: 'Image super-resolution via sparse representation', IEEE transactions on image processing, 2010, 19, (11), pp. 2861-2873

5 Chaudhuri, S.: 'Super-resolution imaging' (Springer Science & Business Media, 2001. 2001)

6 Al-falluji, R.A.A., Youssif, A.A.-H., and Guirguis, S.K.: 'Single image super resolution algorithms: A survey and evaluation', Int. J. Adv. Res. Comput. Eng. Technol, 2017, 6, pp. 1445-1451

7 Vega, M., Mateos, J., Molina, R., and Katsaggelos, A.K.: 'Super-resolution of multispectral images', The Computer Journal, 2009, 52, (1), pp. 153-167

8 Zhang, H., Zhang, L., and Shen, H.: 'A super-resolution reconstruction algorithm for hyperspectral images', Signal Processing, 2012, 92, (9), pp. 2082-2096

9 Zhang, H., Yang, Z., Zhang, L., and Shen, H.: 'Super-resolution reconstruction for multi-angle remote sensing images considering resolution differences', Remote Sensing, 2014, 6, (1), pp. 637-657

10 Greenspan, H.: ‘Super-resolution in medical imaging’, The computer journal, 2009, 52, (1), pp. 43-63

11 Haris, M., Shakhnarovich, G., and Ukita, N.: ‘Task-driven super resolution: Object detection in low-resolution images’, arXiv preprint arXiv:1803.11316, 2018

12 Sajjadi, M.S., Scholkopf, B., and Hirsch, M.: ‘Enhancenet: Single image super-resolution through automated texture synthesis’, in Editor (Ed.)^(Eds.): ‘Book Enhancenet: Single image super-resolution through automated texture synthesis’ (2017, edn.), pp. 4491-4500

13 Bai, Y., Zhang, Y., Ding, M., and Ghanem, B.: ‘Sod-mtgan: Small object detection via multi-task generative adversarial network’, in Editor (Ed.)^(Eds.): ‘Book Sod-mtgan: Small object detection via multi-task generative adversarial network’ (2018, edn.), pp. 206-221

14 Tuna, C., Unal, G., and Sertel, E.: ‘Single-frame super resolution of remote-sensing images by convolutional neural networks’, Int J Remote Sens, 2018, 39, (8), pp. 2463-2479

15 Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H., and Zhang, L.: ‘Image super-resolution: The techniques, applications, and future’, Signal Processing, 2016, 128, pp. 389-408

16 Borman, S., and Stevenson, R.L.: ‘Super-resolution from image sequences-a review’, in Editor (Ed.)^(Eds.): ‘Book Super-resolution from image sequences-a review’ (IEEE, 1998, edn.), pp. 374-378

17 Hardie, R.C., Barnard, K.J., and Armstrong, E.E.: ‘Joint MAP registration and high-resolution image estimation using a sequence of undersampled images’, IEEE transactions on Image Processing, 1997, 6, (12), pp. 1621-1633

18 Tipping, M.E., and Bishop, C.M.: ‘Bayesian image super-resolution’, Advances in neural information processing systems, 2003, pp. 1303-1310

19 He, K., Zhang, X., Ren, S., and Sun, J.: ‘Deep residual learning for image recognition’, in Editor (Ed.)^(Eds.): ‘Book Deep residual learning for image recognition’ (2016, edn.), pp. 770-778

20 Pashaei, M., Kamangir, H., Starek, M.J., and Tissot, P.: ‘Review and evaluation of deep learning architectures for efficient land cover mapping with UAS hyper-spatial imagery: A case study over a wetland’, Remote Sensing, 2020, 12, (6), pp. 959

21 Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., and Wang, Z.: ‘Photo-realistic single image super-resolution using a generative adversarial network’, in Editor (Ed.)^(Eds.): ‘Book Photo-realistic single image super-resolution using a generative adversarial network’ (2017, edn.), pp. 4681-4690

22 Dong, C., Loy, C.C., He, K., and Tang, X.: ‘Image super-resolution using deep convolutional networks’, IEEE transactions on pattern analysis and machine intelligence, 2015, 38, (2), pp. 295-307

23 Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., and Liao, Q.: ‘Deep learning for single image super-resolution: A brief review’, IEEE Transactions on Multimedia, 2019, 21, (12), pp. 3106-3121

24 Liebel, L., and Körner, M.: ‘Single-image super resolution for multispectral remote sensing data using convolutional neural networks’, ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016, 41, pp. 883-890

25 Wang, C., Liu, Y., Bai, X., Tang, W., Lei, P., and Zhou, J.: ‘Deep residual convolutional neural network for hyperspectral image super-resolution’, in Editor (Ed.)^(Eds.): ‘Book Deep residual convolutional neural network for hyperspectral image super-resolution’ (Springer, 2017, edn.), pp. 370-380

26 Haut, J.M., Fernandez-Beltran, R., Paoletti, M.E., Plaza, J., Plaza, A., and Pla, F.: 'A new deep generative network for unsupervised remote sensing single-image super-resolution', IEEE Transactions on Geoscience and Remote sensing, 2018, 56, (11), pp. 6792-6810

27 Arun, P.V., Herrmann, I., Budhiraju, K.M., and Karnieli, A.: 'Convolutional network architectures for super-resolution/sub-pixel mapping of drone-derived images', Pattern recognition, 2019, 88, pp. 431-446

28 Burdziakowski, P.: 'Increasing the geometrical and interpretation quality of unmanned aerial vehicle photogrammetry products using super-resolution algorithms', Remote Sensing, 2020, 12, (5), pp. 810

29 Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: 'Generative adversarial networks', arXiv preprint arXiv:1406.2661, 2014

30 Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C.: 'Esrgan: Enhanced super-resolution generative adversarial networks', in Editor (Ed.)^(Eds.): 'Book Esrgan: Enhanced super-resolution generative adversarial networks' (2018, edn.), pp. 0-0

31 Dougherty, G.: 'Digital image processing for medical applications' (Cambridge University Press, 2009. 2009)

32 Park, S.C., Park, M.K., and Kang, M.G.: 'Super-resolution image reconstruction: a technical overview', IEEE signal processing magazine, 2003, 20, (3), pp. 21-36

33 Chang, H., Yeung, D.-Y., and Xiong, Y.: 'Super-resolution through neighbor embedding', in Editor (Ed.)^(Eds.): 'Book Super-resolution through neighbor embedding' (IEEE, 2004, edn.), pp. I-I

34 Goto, T., Fukuoka, T., Nagashima, F., Hirano, S., and Sakurai, M.: ‘Super-resolution System for 4K-HDTV’, in Editor (Ed.)^(Eds.): ‘Book Super-resolution System for 4K-HDTV’ (IEEE, 2014, edn.), pp. 4453-4458

35 Peled, S., and Yeshurun, Y.: ‘Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging’, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 2001, 45, (1), pp. 29-35

36 Shi, W., Caballero, J., Ledig, C., Zhuang, X., Bai, W., Bhatia, K., de Marvao, A.M.S.M., Dawes, T., O’Regan, D., and Rueckert, D.: ‘Cardiac image super-resolution with global correspondence using multi-atlas patchmatch’, in Editor (Ed.)^(Eds.): ‘Book Cardiac image super-resolution with global correspondence using multi-atlas patchmatch’ (Springer, 2013, edn.), pp. 9-16

37 Thornton, M.W., Atkinson, P.M., and Holland, D.: ‘Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping’, Int J Remote Sens, 2006, 27, (3), pp. 473-491

38 Gunturk, B.K., Batur, A.U., Altunbasak, Y., Hayes, M.H., and Mersereau, R.M.: ‘Eigenface-domain super-resolution for face recognition’, IEEE transactions on image processing, 2003, 12, (5), pp. 597-606

39 Zhang, L., Zhang, H., Shen, H., and Li, P.: ‘A super-resolution reconstruction algorithm for surveillance images’, Signal Processing, 2010, 90, (3), pp. 848-859

40 Yang, C.-Y., Huang, J.-B., and Yang, M.-H.: ‘Exploiting self-similarities for single frame super-resolution’, in Editor (Ed.)^(Eds.): ‘Book Exploiting self-similarities for single frame super-resolution’ (Springer, 2010, edn.), pp. 497-510

41 Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., and Wang, Z.: 'Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network', in Editor (Ed.)^(Eds.): 'Book Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network' (2016, edn.), pp. 1874-1883

42 Wang, Z., Chen, J., and Hoi, S.C.: 'Deep learning for image super-resolution: A survey', IEEE transactions on pattern analysis and machine intelligence, 2020

43 Yang, C.-Y., Ma, C., and Yang, M.-H.: 'Single-image super-resolution: A benchmark', in Editor (Ed.)^(Eds.): 'Book Single-image super-resolution: A benchmark' (Springer, 2014, edn.), pp. 372-386

44 Baker, S., and Kanade, T.: 'Limits on super-resolution and how to break them', IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24, (9), pp. 1167-1183

45 Bevilacqua, M., Roumy, A., Guillemot, C., and Alberi-Morel, M.L.: 'Low-complexity single-image super-resolution based on nonnegative neighbor embedding', 2012

46 Yang, J., Wang, Z., Lin, Z., Cohen, S., and Huang, T.: 'Coupled dictionary training for image super-resolution', IEEE transactions on image processing, 2012, 21, (8), pp. 3467-3478

47 Zeyde, R., Elad, M., and Protter, M.: 'On single image scale-up using sparse-representations', in Editor (Ed.)^(Eds.): 'Book On single image scale-up using sparse-representations' (Springer, 2010, edn.), pp. 711-730

48 Dong, C., Loy, C.C., He, K., and Tang, X.: 'Learning a deep convolutional network for image super-resolution', in Editor (Ed.)^(Eds.): 'Book Learning a deep convolutional network for image super-resolution' (Springer, 2014, edn.), pp. 184-199

49 Dong, C., Loy, C.C., and Tang, X.: ‘Accelerating the super-resolution convolutional neural network’, in Editor (Ed.)^(Eds.): ‘Book Accelerating the super-resolution convolutional neural network’ (Springer, 2016, edn.), pp. 391-407

50 Tong, T., Li, G., Liu, X., and Gao, Q.: ‘Image super-resolution using dense skip connections’, in Editor (Ed.)^(Eds.): ‘Book Image super-resolution using dense skip connections’ (2017, edn.), pp. 4799-4807

51 Simonyan, K., and Zisserman, A.: ‘Very deep convolutional networks for large-scale image recognition’, arXiv preprint arXiv:1409.1556, 2014

52 Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: ‘Going deeper with convolutions’, in Editor (Ed.)^(Eds.): ‘Book Going deeper with convolutions’ (2015, edn.), pp. 1-9

53 Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A.: ‘Inception-v4, inception-resnet and the impact of residual connections on learning’, in Editor (Ed.)^(Eds.): ‘Book Inception-v4, inception-resnet and the impact of residual connections on learning’ (2017, edn.), pp.

54 Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q.: ‘Densely connected convolutional networks’, in Editor (Ed.)^(Eds.): ‘Book Densely connected convolutional networks’ (2017, edn.), pp. 4700-4708

55 Lai, W.-S., Huang, J.-B., Ahuja, N., and Yang, M.-H.: ‘Deep laplacian pyramid networks for fast and accurate super-resolution’, in Editor (Ed.)^(Eds.): ‘Book Deep laplacian pyramid networks for fast and accurate super-resolution’ (2017, edn.), pp. 624-632

56 Bruhn, A., Weickert, J., and Schnörr, C.: ‘Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods’, International journal of computer vision, 2005, 61, (3), pp. 211-231

57 Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K.: 'Enhanced deep residual networks for single image super-resolution', in Editor (Ed.)^(Eds.): 'Book Enhanced deep residual networks for single image super-resolution' (2017, edn.), pp. 136-144

58 Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P.: 'Image quality assessment: from error visibility to structural similarity', IEEE transactions on image processing, 2004, 13, (4), pp. 600-612

59 Johnson, J., Alahi, A., and Fei-Fei, L.: 'Perceptual losses for real-time style transfer and super-resolution', in Editor (Ed.)^(Eds.): 'Book Perceptual losses for real-time style transfer and super-resolution' (Springer, 2016, edn.), pp. 694-711

60 Bruna, J., Sprechmann, P., and LeCun, Y.: 'Super-resolution with deep convolutional sufficient statistics', arXiv preprint arXiv:1511.05666, 2015

61 Gatys, L.A., Ecker, A.S., and Bethge, M.: 'Texture synthesis using convolutional neural networks', arXiv preprint arXiv:1505.07376, 2015

62 Gatys, L.A., Ecker, A.S., and Bethge, M.: 'Image style transfer using convolutional neural networks', in Editor (Ed.)^(Eds.): 'Book Image style transfer using convolutional neural networks' (2016, edn.), pp. 2414-2423

63 Bulat, A., and Tzimiropoulos, G.: 'Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans', in Editor (Ed.)^(Eds.): 'Book Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans' (2018, edn.), pp. 109-117

64 Wang, X., Yu, K., Dong, C., and Loy, C.C.: 'Recovering realistic texture in image super-resolution by deep spatial feature transform', in Editor (Ed.)^(Eds.): 'Book Recovering realistic texture in image super-resolution by deep spatial feature transform' (2018, edn.), pp. 606-615

65 Yuan, Y., Liu, S., Zhang, J., Zhang, Y., Dong, C., and Lin, L.: 'Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks', in Editor (Ed.)^(Eds.): 'Book Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks' (2018, edn.), pp. 701-710

66 Wang, Z., Simoncelli, E.P., and Bovik, A.C.: 'Multiscale structural similarity for image quality assessment', in Editor (Ed.)^(Eds.): 'Book Multiscale structural similarity for image quality assessment' (Ieee, 2003, edn.), pp. 1398-1402

67 Channappayya, S.S., Bovik, A.C., Caramanis, C., and Heath, R.W.: 'SSIM-optimal linear image restoration', in Editor (Ed.)^(Eds.): 'Book SSIM-optimal linear image restoration' (IEEE, 2008, edn.), pp. 765-768

68 Sheikh, H.R., Sabir, M.F., and Bovik, A.C.: 'A statistical evaluation of recent full reference image quality assessment algorithms', IEEE Transactions on image processing, 2006, 15, (11), pp. 3440-3451

69 Wang, Z., and Bovik, A.C.: 'Mean squared error: Love it or leave it? A new look at signal fidelity measures', IEEE signal processing magazine, 2009, 26, (1), pp. 98-117

70 Dai, D., Wang, Y., Chen, Y., and Van Gool, L.: 'Is image super-resolution helpful for other vision tasks?', in Editor (Ed.)^(Eds.): 'Book Is image super-resolution helpful for other vision tasks?' (IEEE, 2016, edn.), pp. 1-9

71 Fookes, C., Lin, F., Chandran, V., and Sridharan, S.: 'Evaluation of image resolution and super-resolution on face recognition performance', Journal of Visual Communication and Image Representation, 2012, 23, (1), pp. 75-93

72 Zhang, K., Zhang, Z., Cheng, C.-W., Hsu, W.H., Qiao, Y., Liu, W., and Zhang, T.: 'Super-identity convolutional neural network for face hallucination', in Editor (Ed.)^(Eds.): 'Book Super-identity convolutional neural network for face hallucination' (2018, edn.), pp. 183-198

73 Chen, Y., Tai, Y., Liu, X., Shen, C., and Yang, J.: 'Fsrnet: End-to-end learning face super-resolution with facial priors', in Editor (Ed.)^(Eds.): 'Book Fsrnet: End-to-end learning face super-resolution with facial priors' (2018, edn.), pp. 2492-2501

74 Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y.: 'Image super-resolution using very deep residual channel attention networks', in Editor (Ed.)^(Eds.): 'Book Image super-resolution using very deep residual channel attention networks' (2018, edn.), pp. 286-301

75 Jolicoeur-Martineau, A.: 'The relativistic discriminator: a key element missing from standard GAN', arXiv preprint arXiv:1807.00734, 2018

76 Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J., and Reynolds, J.M.: ''Structure-from-Motion'photogrammetry: A low-cost, effective tool for geoscience applications', Geomorphology, 2012, 179, pp. 300-314

77 Snavely, N.: 'Scene reconstruction and visualization from internet photo collections: A survey', IPSJ Transactions on Computer Vision and Applications, 2011, 3, pp. 44-66

78 Lowe, D.G.: 'Distinctive image features from scale-invariant keypoints', International journal of computer vision, 2004, 60, (2), pp. 91-110

79 Furukawa, Y., and Hernández, C.: 'Multi-view stereo: A tutorial', Foundations and Trends® in Computer Graphics and Vision, 2015, 9, (1-2), pp. 1-148

80 Yosinski, J., Clune, J., Bengio, Y., and Lipson, H.: 'How transferable are features in deep neural networks?', arXiv preprint arXiv:1411.1792, 2014

81 Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A.: 'Automatic differentiation in pytorch', 2017

82 Agustsson, E., and Timofte, R.: 'Ntire 2017 challenge on single image super-resolution: Dataset and study', in Editor (Ed.)^(Eds.): 'Book Ntire 2017 challenge on single image super-resolution: Dataset and study' (2017, edn.), pp. 126-135

83 Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H., and Zhang, L.: 'Ntire 2017 challenge on single image super-resolution: Methods and results', in Editor (Ed.)^(Eds.): 'Book Ntire 2017 challenge on single image super-resolution: Methods and results' (2017, edn.), pp. 114-125

84 Martin, D., Fowlkes, C., Tal, D., and Malik, J.: 'A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics', in Editor (Ed.)^(Eds.): 'Book A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics' (IEEE, 2001, edn.), pp. 416-423

85 Huang, J.-B., Singh, A., and Ahuja, N.: 'Single image super-resolution from transformed self-exemplars', in Editor (Ed.)^(Eds.): 'Book Single image super-resolution from transformed self-exemplars' (2015, edn.), pp. 5197-5206

86 Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., and Zelnik-Manor, L.: 'The 2018 pirm challenge on perceptual image super-resolution', in Editor (Ed.)^(Eds.): 'Book The 2018 pirm challenge on perceptual image super-resolution' (2018, edn.), pp. 0-0

87 Agisoft, L.: 'Metashape–photogrammetric processing of digital images and 3D spatial data generation', in Editor (Ed.)^(Eds.): 'Book Metashape–photogrammetric processing of digital images and 3D spatial data generation' (2019, edn.), pp.

88 Fraser, C.S.: ‘Digital camera self-calibration’, ISPRS Journal of Photogrammetry and Remote sensing, 1997, 52, (4), pp. 149-159

89 Remondino, F., and Fraser, C.: ‘Digital camera calibration methods: considerations and comparisons’, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2006, 36, (5), pp. 266-272

CHAPTER IV: TERRESTRIAL LIDAR DATA CLASSIFICATION BASED ON RAW WAVEFORM SAMPLES VERSUS ONLINE WAVEFORM ATTRIBUTES

Abstract

In this study, the potential of raw samples of digitized echo waveforms collected by full-waveform (FW) terrestrial laser scanning (TLS) for point cloud classification is investigated. Two different FW TLS systems are employed, both equipped with a waveform digitizer for access to the raw waveform and online waveform processing which assigns calibrated waveform attributes to each point measurement. Point cloud classification based on samples of the raw single-peak echo waveform is compared with point cloud classification based on the calibrated online waveform attributes. A deep convolutional neural network (DCNN) is designed and implemented for the supervised classification. Random forest classifier is used as a benchmark to evaluate the performance of the proposed DCNN model. In addition, feature importance and temporal stability of the raw waveform samples versus the calibrated waveform attributes for point cloud classification are reported. Classification results are evaluated at two study sites, a built environment on a university campus and a natural coastal wetland environment. Results show that direct classification of the raw waveform samples outperforms classification based on the set of calibrated waveform attributes at both study sites. Results also show that the contribution of the range, as the only geometric attribute in the raw waveform feature vector, significantly increases the classification performance, while it has a relatively negligible impact on the classification performance based on the calibrated waveform features. Finally, performance of the DCNN for filtering ground points to generate a digital terrain model (DTM) based on classification of the raw waveform samples is assessed. Results are evaluated at the wetland site and compared to a DTM generated from a progressive morphological filter and real-time kinematic (RTK) GNSS survey.

4.1. Introduction

Conventional terrestrial and airborne laser scanning systems based on the Time-of-Flight (ToF) measurement principle, which are characterized as analogue discrete return light detection and ranging (lidar) systems, have long been used for topographic mapping and other remote sensing (RS) applications. For each emitted laser pulse, echo detection and time-of-arrival (TOA) estimation of the backscattered laser pulse are performed in real-time by analogue devices. In discrete return systems, the estimation of the TOA is highly affected by range walk, i.e., the amplitude of echo pulse detected by the receiver frontend [1]. The analogue estimators may yield significant range errors or completely fail in accurately detecting multiple targets along the laser transmit path, depending on the temporal separation between consecutive targets with respect to the emitted laser pulse width [2].

In contrast to discrete return systems, in echo-digitizing lidar systems, the complete return signal from the reflecting target is sampled at high rate and recorded in a digital form prior to performing the target detection [3]. Small footprint full-waveform (FW) airborne laser scanning (ALS) systems have been developed in the past few decades [4]. More recently, terrestrial, mobile, and unmanned airborne lidar systems with the capability of recording FW data are also becoming more readily available. Echo pulse attributes, such as amplitude and width, derived from the waveform signal backscattered from a reflecting object are shown to be useful for classification of lidar data collected over natural and built environments [4-6].

However, extracting the fundamental properties of the returned waveform, such as the number of relevant peaks and parameters describing the shape of each detected echo in the waveform signal is a challenging task in signal processing [4]. Moreover, the echo pulse attributes need to be discriminative enough to be exploited as relevant features in the feature vector of the

target for efficient classification. Depending on the employed FW lidar system for collecting waveform data, and the required accuracy to extract waveform attributes, different techniques have been developed for waveform decomposition and modelling [4, 7-10], By carrying out a radiometric calibration procedure on waveform data, more relevant features can also be introduced to the feature vector of the target to improve the overall accuracy of the classification task [3, 5].

In some lidar systems, especially terrestrial laser scanning (TLS) systems, the system response model is usually unknown or too complex for modeling and decomposing the waveform using typical parametric functions, such as the well-known generalized Gaussian function [2]. To take the advantage of the capability of FW TLS systems in digitizing and recording the return signal for classification tasks, an intensive calibration procedure for approximating the actual system response model seems inevitable [2, 11]. This approximated model will later be used as the basic template for waveform decomposition and modeling which can be accomplished, almost in real-time, by an internal FW processing unit in some lidar systems [11, 12].

Due to the fact that digitized waveform samples are the fundamental source of data for modeling of the waveform shape and extracting echo parameters, samples of the raw echo waveform may have the potential to be directly employed as waveform features representing physical characteristics of the illuminated target. One advantage of this approach is that conventional FW analysis (FWA) techniques are not required for extracting common waveform attributes. Therefore, uncertainties in evaluating the echo parameters due to the low capacity of the parametric functions for fitting to the echoes are eliminated. In addition, due to the lower sampling rate of the digitizer (usually 2 ns) and higher non-linear system response in TLS systems, with respect to FW ALS systems, common FWA techniques are not usually applicable in a FW TLS system for the full dynamic range of the lidar system [2, 11].

4.2. Background

Conventional terrestrial A FW lidar system is capable of digitizing and recording the complete temporal energy profile of the backscattered laser signal from the reflecting target, where the "waveform" is the term that refers to the shape of the echo signal [3]. In comparison to discrete-return lidar systems, the data collected by FW lidar systems contains additional information about the physical and spatial properties of the illuminated target in the footprint of the laser beam [4]. Specifically, in critical target situations where the target location with respect to the nearby targets or its spatial distribution along the travel path of the laser pulse causes uncertainty in range determination or target identification, analyzing the additional information may help to partly resolve those ambiguities [13]. This additional information is typically derived from detection and modeling of each individual echo within the digitized waveform signal. Two of the most important echo attributes include echo pulse amplitude, which is related to the radiometric characteristics of the target, and pulse width, which is a measure of the target-laser beam configuration and/or surface roughness at scales comparable to the laser wavelength [14]. These echo attributes and their derivatives, including the reflectance and geometry of the target with respect to the laser beam, such as the backscatter cross-section and backscattering coefficient, have been widely used as relevant waveform features for target classification or segmentation [15, 16].

Analyzing the waveform, which encapsulates both radiometric and geometric properties of the illuminated target, is usually accomplished in an offline (or post-processing) mode using a pre-defined FWA technique [4]. Some lidar systems, however, such as Riegl VZ-Line TLS systems, offer an online waveform processing approach [2]. In this approach, the analysis of the returned waveform is carried out in real-time, where the actual system response, derived from an intensive system calibration procedure accomplished by the lidar manufacturer, is exploited for waveform

decomposition and modeling [2]. This is due to the fact that the tremendous dynamic range of the TLS system usually leads to a large degree of nonlinearity in the characteristics (e.g., scale) of the system response. However, by providing the raw digitized waveform data, Riegl VZ-Line scanners offer the end-user the flexibility to apply more in-depth and advanced analysis on the raw waveform data to achieve satisfactory results.

4.2.1. Waveform features for classification

The development of FW lidar systems and the advancement in FWA algorithms have brought interest to explore the suitability of features derived from digitized returned waveform signals for lidar data classification [16-19]. FW lidar data classification has been met through different approaches. Some work focus only on the geometric properties of targets and explores the most relevant geometric features describing different targets for classification[1]. This approach emphasizes the improvement in geometric representation of the measured target for classification purposes. Improvements in the geometric representation of the target will help to better discriminate different targets in the classification procedure [20-23]. Other studies explore the combination of basic waveform features, such as the echo width and (uncalibrated) amplitude, with some calibrated attributes, such as the backscatter cross-section or backscattering coefficient, derived from the echo waveform in the feature vector of the target [15-17, 24]. Some geometric features related to the targets in lidar data, such as elevation differences and surface normal, and features related to the position of a detected echo in the waveform or coefficients describing the deviation of the echo pulse from the ideal transmitted pulse may also be considered to enhance the classification performance [16, 25, 26]. The potential of parameters related to the structure of the waveform, such as the rise time to the first echo, ratio between tree canopy and ground energy,

total waveform energy, and height of median energy have also been explored for airborne lidar data classification [27, 28].

Although the benefits of FW lidar data are particularly profound for forestry and vegetation segmentation due to its ability to provide accurate and detailed information about the vertical structure of the vegetated area and the terrain elevation underneath [29, 30], FWA has also been found advantageous for the challenging task of classification of natural and built objects in developed environments. Additional information about the reflecting properties of natural and built structures and their spatial distribution, encoded in the waveform data, have been shown to be relevant for land cover mapping and target classification in urban and natural areas [18, 25, 29, 31, 32]. However, in complex areas, the basic waveform features become less discriminative for a multi-class classification task [5, 16]. Nevertheless, some research studies have developed FWA techniques to derive more accurate basic features as well as some advanced features from the backscattered signal [15]. Studies have shown that employing basic and advanced waveform features along with careful radiometric calibration of the data improves multi-class classification in urban areas [15, 16, 33]. Furthermore, adding features such as the total number of echoes within each waveform and the position of the echo in the waveform together with some geometric and/or spectral features derived from the lidar system or integrated sensors such as multispectral or hyperspectral cameras can significantly increase the accuracy of the multi-class classification of lidar data over complex environments [5, 26, 34, 35].

4.2.2. Objectives of this work

This paper seeks to investigate the utility of raw digitized waveform samples for multi-class classification of targets within built and natural environments. Classification performance is evaluated at two study sites: a university campus and a coastal wetland. Natural and made-made

targets found in the selected study sites are classified through direct classification of the corresponding backscattered waveforms. The FW TLS systems employed in this study are equipped with online waveform processing capability and a waveform digitizer, which provides the end-user with digitized waveform samples for advanced post-acquisition analysis.

The hypothesis in this study is that the relevant waveform features for classification, either derived from the online waveform processing or in an offline mode using the post-processing FWA, are not always available. Moreover, in either case, samples of the digitized waveform are the primary source for waveform modeling and feature extraction. Thus, this study aims to investigate the potential of the raw samples of digitized single-peak echo waveforms for target classification. In this work, feature vectors containing samples of the raw digitized waveform are referred to as offline waveform feature vectors (attributes). Spatial information related neighboring targets are not included in the proposed classification approach. Due to the high correlation between the target's distance and the received optical energy by the lidar system, the range to the target is the only geometric feature which is included in the feature vector of the illuminated target. However, for the sake of completeness, the discriminative capability of samples of the waveform for multi-class classification in the absence of the range attribute will also be reported.

The potential of digitized raw waveform samples is compared with the capability of calibrated waveform features, derived from the online waveform processing, for target classification. The calibrated waveform features are referred to as the online waveform features throughout this paper. Due to the importance of single-peak echo waveforms in representing geophysical characteristics of targets, where the calibrated parameters directly relate to the spatial and radiometric properties of the illuminated target, only waveform samples related to single-peak echo waveforms are considered for waveform classification and feature analysis. In addition, as it

has been explained in Section 5, the majority of waveform data collected at each study site are single-peak echo waveforms. For the sake of fairness in the comparative analysis, only the online waveform attributes related to single-return measurements are considered for classification.

At the built environment (campus) study site, online and offline waveform attributes in the feature vectors are analyzed for classification using FW TLS datasets collected at two different points in time. Each individual dataset includes multiple scan positions within the same environment. Employing these datasets is crucial for investigating the robustness of the suggested waveform classification approach. First, by having multiple scan positions within one dataset, the training and testing of the classifier can be carried out on two separately collected data sets. Second, collecting FW TLS data from multiple scan positions significantly decreases the correlations between a certain target category and some properties of the measured waveform that are highly correlated to the TLS-target geometric configuration, such as the range to the target. Finally, two waveform datasets collected at two different points in time from the same study area makes it possible to perform a temporal stability analysis on both online and offline waveform feature vectors.

At the natural environment (wetland) study site, classification is performed using a different FW TLS system than that used in the built environment but manufactured by the same company. This TLS is also equipped with online waveform processing and a waveform digitizer. This study site provides an additional evaluation of the robustness of the raw waveform classification approach within natural terrain and based on a different TLS system.

The main contributions of this paper are as follows: 1) classification performance on raw waveform data is compared with classification based on online waveform attributes (features) derived from a calibrated lookup table (LUT) provided by the TLS system manufacturer; 2) a deep

convolutional neural network (DCNN) architecture is developed and employed for the multi-class classification task, where the offline (raw waveform data) and online waveform feature vectors related to each target are used as input to the DCNN model. Furthermore, the classification performance of the DCNN model is compared with that from a random forest (RF) model and important features in both online and offline waveform feature vectors are reported; 3) temporal stability, and performance in different environments, of the offline versus online waveform features for classification is investigated; 4) task evaluation of the DCNN for filtering ground points to generate a digital terrain model (DTM) of the wetland surface based on online and offline waveform feature vectors is performed.

The remainder of this paper is organized as follows: Section 3 describes online and offline FWA. Section 4 introduces the study sites and collected FW TLS data. The methodology employed for FW TLS data classification and evaluation based on raw and online waveform attributes is presented in Section 5. Results are presented and discussed in Section 6. Finally, Section 7 concludes with some future work considerations.

4.3. Full-waveform Analysis Approaches

4.3.1. Offline full-waveform analysis

Offline FWA, which is usually employed in FW ALS systems is a post-processing approach to detect pulses and related attributes, e.g., amplitude, and width from the digitized echo signal. Those echo pulse attributes can later be used to derive some information about the scattering characteristics of the illuminated targets. Different approaches proposed to extract the target backscattering properties encoded in the digitized waveform can be broadly categorized into two main approaches: (1) deconvolution-based methods [9] and (2) methods based on fitting the digitized echo waveform with basic parametric functions [3]. With the assumption that the system

response can be described and modeled with an ideal Gaussian function, which is usually true in FW ALS systems with limited dynamic range, Gaussian decomposition and modeling has become a widely accepted FWA approach in FW ALS systems [3, 25, 29]. However, applying Gaussian decomposition and modeling on FW data collected from FW TLS systems, with a large dynamic range, usually leads to unsatisfactory results [11].

4.3.2. Online full-waveform processing

Since 2008, Riegl Laser Measurement Systems, GmbH, Horn, Austria, has developed a line of lidar systems, commonly called V-Line, based on pulsed ToF technology with real-time echo digitization and online waveform processing capabilities [12]. Indeed, Riegl V-Line lidar systems combine the advantages of analogue detection systems, in which lidar survey results are provided without the need for post-processing, with those of airborne echo digitizing lidar systems [1]. In contrast to FW ALS systems, in which digitized returned waveforms are stored during flight for FWA in an offline or post-processing mode, the lack of computational power in TLS systems for real-time processing has led Riegl to implement online waveform processing for V-Line scanners including VZ-Line TLS products [1].

As opposed to FW ALS systems in which an ideal Gaussian pulse, usually, closely approximates the sensor's system response, the Riegl VZ-Line systems exploit the actual sensor's system response derived from an intensive calibration procedure performed by the manufacturer [2]. This actual system waveform is employed for the waveform decomposition and derivation of physical observables describing the scattering properties of the target, such as the target's laser cross-section or calibrated relative reflectance, within an automatic procedure called online waveform processing [2, 11]. In this approach, the nonlinear scale characteristics of the system response are perfectly captured by the calibrated sensor's system response, resulting in the utmost

accuracy and precision in the echo decomposition and reconstruction [2, 11]. In addition to calibrated relative reflectance, the online waveform processing in VZ-Line TLS systems provides calibrated amplitude and pulse deviation for each detected echo signal.

Calibrated amplitude: The amplitude of the optical echo signal detected in the receiver depends on a number of factors including system-related factors, such as emitted laser pulse and the receiver aperture, and target-related factors, such as target's laser radar cross-section which is a function of the target's reflectance in the laser's wavelength, target's size, and the directivity of the target's reflection [12]. By means of a precise calibration procedure during manufacturing, the amplitude of every detected pulse is given relative to the amplitude of an echo signal at the detection threshold of the LiDAR system as [12]:

$$A_{db} = 10 \times log\left(\frac{P_{echo}}{P_{DL}}\right) \quad (1)$$

where A_{db} is amplitude in decibel, P_{echo} is the optical input power, and P_{DL} is the minimum detectable input power. The logarithmic measure, given above, covers the wide dynamic range of the employed TLS system. However, the amplitude of the echo signal, detected by the receiver, is highly correlated with the range value making it difficult to discriminate different targets based purely on their amplitude readings for target classification.

Calibrated relative reflectance: The target reflectance is a physical target property, which refers to the fraction of the incident laser's optical power that is reflected by the target at the laser wavelength. However, the reflected optical power measured at the receiver is highly correlated with the target range. The calibrated relative reflectance is defined as the ratio of the absolute amplitude of the target to the amplitude of a target of known reflectance at the same range, orthonormal to the laser beam and with a size larger than the laser footprint [12]. The target of the known reflectance is usually a white diffuse target with the reflectance of about 100%. This

quantity relates the echo intensity to the target reflectance independent of the range to the target. The Riegl VZ-Line TLS systems also take the directivity of the target reflectance into account, making the relative reflectance comparable with the normalized laser radar cross-section. The relative reflectance in dB, ρ_{rel}, measured by the LiDAR system is evaluated as [20]:

$$\rho_{rel} = A_{dB} - A_{dB,Ref}(R) \tag{2}$$

where A_{dB} is the calibrated amplitude, $A_{dB,Ref}(R)$ is the calibrated amplitude of the reference target at range R.

Pulse deviation: Online waveform processing in VZ-Line systems provides information about the pulse shape figure, where the shape of the echo pulse is compared with the expected (and undistorted) pulse shape for each individual echo pulse. In fact, pulse deviation represents the deviation of the returned pulse from the actual system response already evaluated and stored in the instrument [1]. The stored calibrated system response model, which encompasses the entire dynamic range of the TLS system, captures a large portion of systematic changes in the echo shape as a function of range, enabling accurate fits to the waveform [11, 12]. However, even the ideal echo signal, backscattered from an extended flat target orthonormal to the laser beam, still shows discrepancies with the stored system response [12]. This deviation especially increases for overlapping echoes, returning from targets located at a distance smaller than the multi-target resolution (MTR) distance (for example, $0.8\,m$ in the Riegl VZ-400 TLS systems), and for broadened echoes returning from slanted targets [12]. The pulse deviation is given as [12]:

$$\delta = \sum_{i=1}^{N} |s_i - p_i| \tag{3}$$

where N is the number of samples in the digitized echo signal with digital number (DN) value s_i, and p_i is the digital number corresponding to the sample from the equivalent system response.

4.4. Study Sites and Data

4.4.1. Study sites

To perform this experiment on a built environment, part of the campus of Texas A&M University-Corpus Christi, TX, USA was selected. The campus area includes natural and man-made structures such as palm trees, grass fields, asphalt roads and buildings, which are used as target categories for classification. To evaluate the potential of the raw waveforms to discriminate tree canopy from grass fields, the tree category is divided into two separate sub-classes (trunk and canopy) in the classification process, described further below. Figure 4.1 illustrates the campus study site with the area of $94{,}600\ m^2$ displayed in two different views of the co-registered point cloud dataset collected by the Riegl VZ-400 FW TLS system at two different points in time.

Figure 4.1. Co-registered TLS point cloud of the campus. Side view is colored gray by reflectance. Top view (left) is color-coded by height. Circles show six TLS positions in the October survey. White circles represent two TLS positions for the July survey.

The second study site, Mustang Island Wetland Observatory (MUI), is part of a coastal wetland located on a barrier island along the southern portion of the Texas Gulf Coast, USA, bounded by Corpus Christi Bay to the west and the Gulf of Mexico to the east. Two prominent target categories in the selected coastal wetland are vegetated land cover and tidal flat areas. Tidal flat areas are bare-earth sediment surfaces, usually devoid of vegetation, and alternately submerged

and exposed to the air by changing tide and water levels. In this study, tidal flat areas include exposed, lower lying and tidal inundated wetland surface areas as well as exposed, upland and periodically inundated wetland surface areas within zones of vegetation cover. The vegetated areas include densely vegetated areas and areas with sparse vegetation cover. Some other target categories found in this study site include a dirt road and power lines. Thus, tidal flat, vegetation, road, and power lines are used for multi-class classification of this natural environment. Figure 4.2 illustrates the coastal wetland study site with the area of $185{,}430\,m^2$ visualized with a georeferenced point cloud collected at 8 different scan positions with a Riegl VZ-2000i FW TLS.

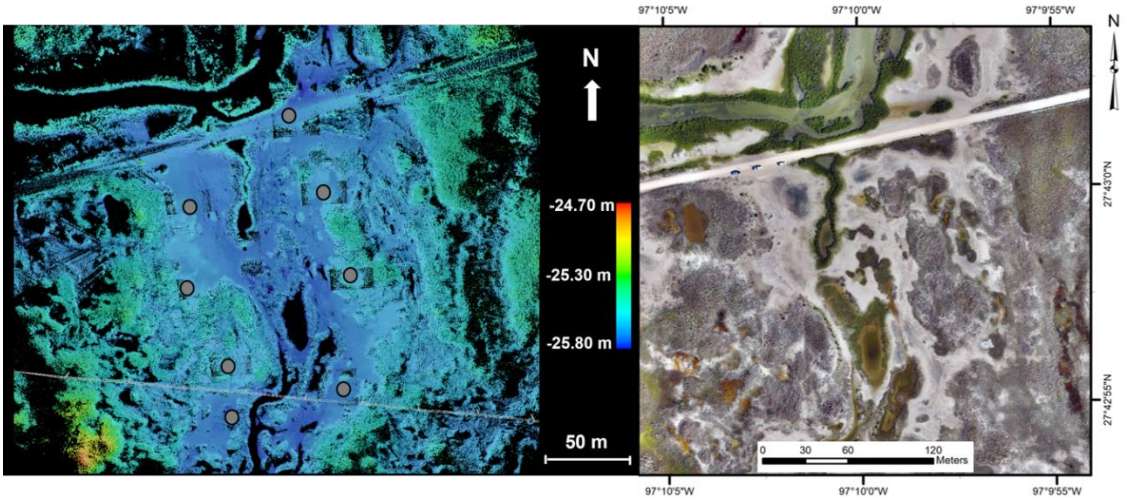


Figure 4.2. Georeferenced point cloud collected from the coastal wetland, color-coded based on ellipsoidal height. The gray circles demarcate the TLS positions. The orthoimage on the right shows the land cover of the study site.

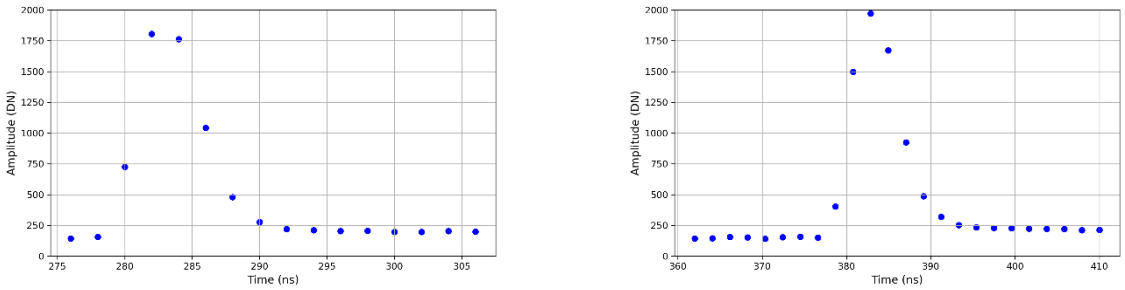
4.4.2. Full-waveform TLS data

The Riegl VZ-400 and VZ-2000i FW TLS systems were employed for collecting data from the campus and wetland study sites, respectively. Specifications of both lidar systems are given in Table 4.1. These TLS systems not only perform online waveform processing, but also digitize and

record the entire echo waveform at a sampling rate of $500\,\mathrm{MHz}$ or one sample per two nanoseconds. Figure 4.3 illustrates single-peak digitized echo waveforms recorded by the Riegl VZ-400 and VZ-2000i from extended targets with the same reflectance values of about -3.2 dB derived from the online waveform processing. According to the figure, the Riegl VZ-400 TLS system records 16 samples related to the echo waveform while the Riegl VZ-2000i TLS system records 24 samples. Both cases are for an extended target perpendicular to the path of the emitted laser beam. However, both systems use the same digitization rate; the time separation between two consecutive samples in each individual waveform is $2\ ns$.

Table 4.1. Technical specifications of the Riegl VZ-400 and VZ-2000i.

Technical Specifications	Riegl VZ-400	Riegl VZ-2000i
FOV	$360° \times 100°$	$360° \times 100°$
Max. Measurement range (Long range mode)	For natural targets: $\rho \geq 20\%$ up to 280m $\rho \geq 80\%$ up to $600m$	For natural targets: $\rho \geq 20\%$ up to $1300m$ $\rho \geq 90\%$ up to $2500m$
Measurement rate	42,000meas./sec (Long range mode) 122,000meas./sec (High speed mode)	21,000meas./sec (Long range mode 500,000meas./sec (High speed mode)
Beam divergence	0.3mrad	0.3mrad
Laser wavelength	1550nm	1550nm
Angular resolution	0.0005°	0.0005°
Range accuracy	5mm	5mm
Range precision	3mm	3mm



(a) VZ-400 (b) VZ-2000i

Figure 4.3. Single-peak digitized echo waveforms with 2 ns spacing measured by Riegl VZ-400 and VZ-2000i TLS generated from laser pulse returns from extended targets with similar reflectance values.

4.4.2.1. Campus study site

Two separate FW TLS surveys, performed at two different times using a Riegl VZ-400 TLS system, over the selected campus area have been considered for FW data classification. The first TLS survey was carried out on July 14th, 2020, at two different scan positions. The second survey was carried out on October 31st, 2020, at six scan positions, where two of those scan positions were located at the same TLS positions and heights used to acquire data on July 14th. The average temperature and humidity during data collection on July 14th are 37°C and 78% and on October 31st are $20°C$ and 55%, respectively.

In Figure 4.1, the side view shows a TLS point cloud colored by calibrated relative reflectance values, while the top view represents the same point cloud data color-coded according to height. Circles in Figure 4.1 show the scan positions in the collected datasets on October 31st, where the two white circles show the TLS positions in common with the TLS survey conducted on July 14th. The dataset collected on October 31st is used for training and testing the classifier. Having several scan positions for the October dataset is crucial for the robustness of the suggested

classification approach. First, by having multiple scan positions within the study area, instances for training and testing the underlying classifier can be chosen from separate scan positions or a combination of them. Second, collecting FW TLS data from multiple scan positions significantly decreases the correlations between the shape of the return waveform and the geometric configuration of the target with respect to the TLS system. Furthermore, the trained classifier on the October dataset is also used to classify the dataset collected on July 14th, which enables analysis of the temporal stability of online and offline waveform feature vectors for classification.

For both TLS surveys, point cloud/waveform data were collected at each scan position in panoramic mode with a $360°$ horizontal field-of-view (FOV) and $100°$ (from $-40°$ to $+60°$) vertical FOV using the scanner's high-speed acquisition mode with FW recording turned on. The pulse repetition rate (PRR)was set to 300 KHz, corresponding to 122,000 measurements per second, and the minimum step angle was set to $0.0024°$ equivalent to 4 mm point spacing at 100 m.

Registration and fine alignment of individual scan positions into a cohesive point cloud was performed with Riegl RiSCAN PRO, version 2.12.1, software package, using the Multi-Station Adjustment (MSA) plugin. MSA results reported by the RiSCAN PRO software show the final horizontal and vertical accuracy of TLS scan co-registration are 0.006 m and 0.004 m, respectively, with angular precision better than $0.004°$ for all angular parameters. Registered point cloud data from both TLS surveys was locally referenced within a project-oriented coordinate system.

4.4.2.2. Wetland study site

The TLS survey of the wetland study area was conducted on February 23rd, 2021, at 8 different scan positions using the Riegl VZ-2000i FW TLS with an integrated RTK GNSS receiver.

The average temperature and humidity during data acquisition in the coastal wetland study area are 24℃ and 56%, respectively. Figure 4.2 illustrates the TLS locations on wetland study site using gray circles on the georeferenced point cloud color-coded based on ellipsoid height. The orthoimage given in Figure 4.2 shows the land cover within the wetland study site. Point cloud and digitized waveform data were collected at each scan position in panoramic mode with a 360° horizontal FOV and 100° (from −40° to +60°) vertical FOV using the scanner's high-speed acquisition mode. The PRR was set to 600 KHz, corresponding to 250,000 measurements per second, and the minimum stepping angle was set to 0.0024°.

The scan positions were co-registered using the same procedure described above for the campus study site and georeferenced based on the VZ-2000i's integrated RTK GNSS receiver, which received corrections from the Texas Department of Transportation (TxDOT) real time network (RTN) during data acquisition. This approach provided absolute positional accuracy down to a few centimeters. Spatial referencing was set to the North American Datum of 1983 (NAD83), National Adjustment 2011, State Plane Coordinate System, Texas South Zone for the horizontal point cloud coordinates. Vertical coordinates were referenced to the NAD83 ellipsoid. Georeferencing of the TLS data at the wetland site was necessary for assessment of waveform classified ground point data for DTM generation and comparison to RTK survey data, described further below.

4.4.3. RTK GNSS control points on coastal wetland

A network of 132 RTK GNSS points was collected on the bare-earth surface in the coastal wetland study site using an Altus NR3 (Septentrio) RTK GNSS rover with cellular-based corrections provided by the TxDOT RTN. Coordinates at each sample point were computed from a 10 second observation average at 1 Hz sample rate. Ellipsoidal height (vertical) accuracy using

this procedure is estimated to be within 2.7 cm (1 sigma). All RTK data were collected in the same reference frame as the TLS survey.

These RTK points serve as ground truth (i.e., vertical control points) for evaluating DTMs generated from the offline and online waveform classified point cloud data explained later in the paper. RTK points were distributed throughout the study area, collected on surfaces within vegetated and exposed land cover. From the total set of collected control points, 30 points represent hard surfaces, including tidal flats and dirt road areas, and the remaining 102 points characterize vegetated areas, including both densely vegetated areas and areas with sparse vegetation. RTK GNSS points collected on hard surfaces are used to evaluate the vertical accuracy of the TLS data.

4.5. Methodology

For the purposes of this work, a filtering procedure must first be implemented on all collected TLS datasets to extract single return points derived from online waveform processing and corresponding single-peak echo waveforms. In addition, for a multi-class supervised classification task, an appropriate number of ground truth instances need to be generated for both study sites.

4.5.1. Single-peak echo waveforms

The Riegl RiSCAN PRO, version 2.12.1 software package was used for visualizing, filtering, and exporting the point cloud derived from the online waveform processing with selected attributes. Riegl also provides a software toolkit called RiWaveLib library for advanced research and analysis purposes on the raw waveform data acquired by the Riegl VZ-Line scanners. The digitized echo waveform corresponding to a selected point in the point cloud is accessible through the timestamp attribute assigned to that point, derived from the online waveform processing.

The point cloud data collected from each scan position is filtered to include points related to the single-peak echo waveforms. Such filtering is necessary because the radiometric calibration of the lidar instrument and the resulting relative reflectance are valid for extended targets. In other words, radiometric calibration in a lidar system assumes that the received intensity values are from a single target with a size larger than the footprint of the laser beam. Moreover, due to the higher variations in the echo shape caused by the influence of central obscuration for targets measured at a close range to the scanner, especially up to $9\,m$ [11], the single-peak echo waveforms and corresponding points in the point cloud data, at both study sites, are also filtered to exclude data points collected at distances less than 9 m from the TLS instrument. The net result of the filtering procedure is the exclusion of about 25% and 30% of the waveforms collected from the campus and coastal wetland study sites.

The point attributes exported from the Riegl RiSCAN PRO software include the 3D coordinates, range to the scanner, calibrated amplitude, calibrated relative reflectance, and pulse deviation. For the sake of simplicity, amplitude and reflectance are used rather than calibrated amplitude and calibrated relative reflectance, respectively, in the remaining sections of this paper. The 3D coordinates are not included in the classification process but they are employed for visualization purposes.

To explore the raw digitized waveform data, computer programs were developed using the software library RiWAVELib, and compiled with Microsoft Visual C++ on Windows platform. Examining the single-peak echo waveforms acquired by the Riegl VZ-400 TLS system shows that more than 98% of waveforms contain 16 or 24 samples or digital numbers (DNs). The same examination on the single-peak echo waveforms collected by the Riegl VZ-2000i TLS system shows that more than 97% of the single-echo echo waveforms contains 24 or 32 samples. Echo

waveforms with more than 24 samples recorded by the Riegl VZ-400 or with more than 32 samples measured by the the Riegl VZ-2000i usually belong to highly inclined surfaces in the path of the laser beam or are a consequence of merged echoes from targets spaced closer than the target separation resolution of the lidar sensor. Thus, to avoid including waveforms resulting from a cluster of nearby targets and to save computation time, a second filtering procedure is applied on the waveform data and related points in the point cloud, for both study areas. As a result of this procedure, the echo waveforms with more than 24 samples and 32 samples collected by the Riegl VZ-400 TLS and VZ-2000i TLS systems, respectively, are removed from the classification procedure and the blank elements in the waveform vectors related to the shorter waveforms are padded with the DN of the last sample.

4.5.2. Ground truth preparation

Ground truth instances were generated from the data collected at all scan positions in each study site by manual inspection of the acquired point cloud.

4.5.2.1. Campus study site

The total number of ground truth points and corresponding waveform instances generated from the filtered dataset collected by the Riegl VZ-400 TLS over the campus study site in October and July are given in Table 4.2. From the total number of points in the filtered October dataset, 1,000,000 ground truth points with online waveform features and corresponding waveforms were randomly selected for training the classifier, where each target category participates in training the classifier with 200,000 random point and waveform instances. The number of randomly selected points and related waveforms from the October dataset, used for validation from each target category is given in Table 4.2. It is worth noting that the training and validation sets do not include

shared instances. Table 4.2 also reports the number of points and corresponding waveform instances for testing the classifier on the July dataset after training using the October dataset.

Table 4.2. Total number of ground truth instances generated from the two collected datasets over the campus study area. For each dataset, the number of ground truth instances randomly sampled for training and testing is given.

Dataset	Point	waveform	Asphalt	Building	Grass	Tree trunk	Tree canopy
October	59,470,887	59,470,887	26,639,548	20,179,562	7,391,730	1,711,501	4,548,546
July	12,112,869	12,112,869	3,530,778	3,009,509	2,094,158	1,419,761	2,058,663
Train (Oct)	1,000,000	1,000,000	200,000	200,000	200,000	200,000	200,000
Test (Oct)	20,000,000	20,000,000	6,500,000	6,500,000	3,600,000	1,000,000	2,400,000
Test (Jul)	5,500,000	5,500,000	1,700,000	1,700,000	655,000	385,000	860,000

Table 4.3. Total number of ground truth instances generated from the collected dataset over the coastal wetland study area. For each dataset, the number of ground truth instances randomly sampled for training and testing is given.

Dataset	Point cloud	Raw waveform	Tidal flat	vegetation	Road	Power line
Coastal wetland	18,181,280	18,181,280	10,862,557	5,796,947	1,085,289	436,487
Train	800,000	800,000	200,000	200,000	200,000	200,000
Test	17,381,280	17,381,280	10,662,000	5,596,947	885,289	236,487

4.5.2.2. Wetland study site

Ground truth instances generated from the filtered dataset collected by the Riegl VZ-2000i TLS system over the coastal wetland area are given in Table 4.3. It summarizes the total number of generated ground truth points and corresponding waveforms, and the number of instances that belong to each target category, for training and testing of the underlying classifier. According to

the table, the same number of ground truth instances were randomly selected from the total number of generated ground truth instances for each target category.

4.5.3. Online vs. offline waveform feature vectors

For both TLS systems used in this study, the online waveform feature vector related to each individual single-return point instance includes the range to the scanner, amplitude, reflectance and pulse deviation. The corresponding single-peak offline waveform feature vectors, however, have different lengths for each TLS system. The offline waveform feature vectors include the range value and a series of 24 DNs for the waveforms measured by the Riegl VZ-400 TLS system, while for the VZ-2000i TLS systems the feature vector includes 32 DNs and the range value. As mentioned earlier, the range value is the only geometric attribute which is included in both online and offline feature vectors. It is assumed that the dependency of the intensity to the range from the target, which is resolved during radiometric calibration of the echo waveform signals, can be partially captured by the classification algorithm when the classifier is trained on feature vectors including the range attribute. However, for the sake of completeness, the same feature vectors excluding the range attribute are also used for training and validating the same classifiers.

4.5.4. DCNN architecture for FW TLS data classification

DCNN architectures have significantly outperformed almost all traditional ML approaches for classification and segmentation tasks in an end-to-end manner [36]. While a large number of DCNN architectures have been developed for image and 3D point cloud classification and segmentation, the potential of a DCNN architecture has not been fully explored for FW classification [32, 37].

The proposed DCNN architecture for FW TLS data classification, developed as part of this study, is shown in Figure 4.4. The input to the network is a matrix of data of size $N \times M$, where N

is the number of input instances that are simultaneously fed to the network for classification and M is the number of elements in the input vector. For example, for the offline waveform feature vector classification, the input vector includes $M = 25$ elements, in which 24 elements represent 24 samples of the digitized waveform measured by the Riegl VZ-400 FW TLS system and the remaining element represents the recorded range to the target. In the case that the online waveform feature vectors are fed to the network for classification, M is equal to 4, where the first three elements represent three online waveform attributes (e.g., amplitude, reflectance, and pulse deviation) related to the measured target and the range value.

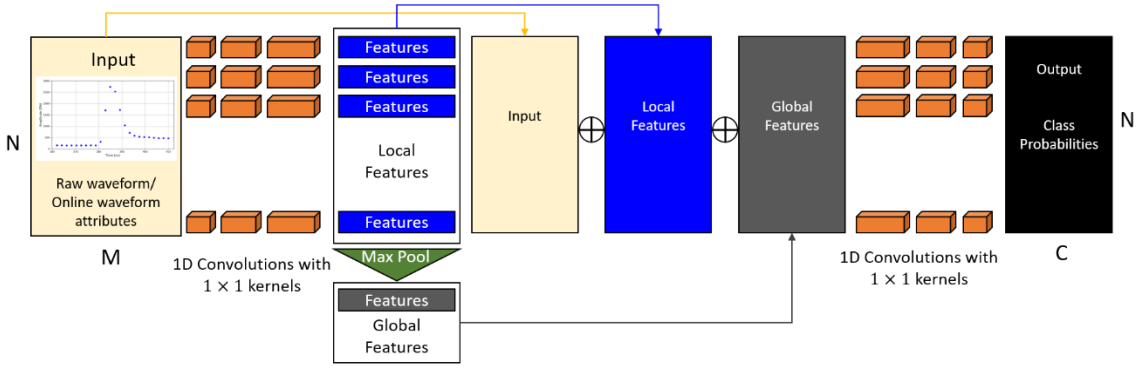


Figure 4.4. Proposed DCNN architecture for FW TLS data classification.

According to Figure 4.4, the first block of the proposed DCNN architecture takes the input data and computes the local features for each input vector by using three 1D convolutional kernels of size 1×1 with batch normalization. Each convolutional layer is then followed by a nonlinear activation function, such as ReLU:

$$f(x) \approx ReLU(\boldsymbol{W}x + \boldsymbol{b}) \tag{24}$$

where x is the input vector or the feature vector computed in an earlier convolutional layer, $\boldsymbol{W}$ is the learnable weight parameters, and $\boldsymbol{b}$ is the bias parameter.

Local features derived in the first convolutional block are fed into a max pooling layer to extract global features from the input feature vectors. As a symmetric function, max pooling layer produces the same output feature vector without any dependence on the order of the input data. The second part of the network concatenates the input vector with both the local and global feature vectors and the resulting vector is fed to the second set of convolutional layers, where three 1D kernels of size 1×1 with batch normalization and the ReLU activation function is applied on each individual input feature vector. To solve the classification of the input data, the feature vector resulting from the last convolutional layer is fed into the classifier defined on top of the DCNN architecture, where the class probability is calculated for each individual input vector by the softmax layer as:

$$p_i = \frac{e^{y^i}}{\sum_{j=1}^{C} e^{y^j}} \tag{25}$$

where p_i is the class probability of the class i with output value of y^i and C is the total number of classes.

Furthermore, due to the fact that collected FW TLS data may include severe imbalanced instances in different classes, the DCNN model uses the weighted categorical cross-entropy loss for training. The loss function can be formulated as:

$$\mathcal{L}_{CE} = \sum_{n=1}^{N} \sum_{c=1}^{C} W_c t_{n,c} \, log\left(y_{n,c}\right) \tag{26}$$

where, $\mathcal{L}_{CE}$ is the categorical cross-entropy loss, $t_{n,c}$ is the ground truth value in one-hot vector representation, and $y_{n,c}$ is the value showing the predicted probability of class c for the input vector n. W_c is the weight for class c, which can be defined as:

$$W_c = \frac{1}{ln\left(1.2 + \frac{a}{b}\right)} \tag{27}$$

where a is the number of the instances of the same target category and b is the total number of instances in all target categories.

In this classification experiment on FW TLS data, the first set of convolutional layers include 256, 512, and 1024 filters, ending with a bottleneck layer of dimension 1024. Also, the second set of the convolutional layers include three sets of 1024, 512, and 256 filters, ending with a bottleneck layer of dimension 256.

To train the DCNN model, the learning rate α was set to 0.001, and Adam optimizer [38] was chosen for updating weights during training. Two exponential decay rate parameters in the Adam optimizer β_1 and β_2, were set to 0.9, and 0.999, respectively. ϵ parameter in the optimization algorithm was set to 1×10^{-7} to avoid any division by zero. The experiment was carried out with 300 epochs on Google Colab, Google's free cloud service, with one Intel(R) Xeon(R) CPU 2.30 GHz and one high-performance Tesla K80 GPU, having 2496 CUDA cores and 12 GB GDDR5 VRAM.

4.5.5. Random forest for FW TLS data classification

In order to compare the performance of the proposed DCNN architecture for FW lidar data classification with a traditional ML-based classification approach, a random forest (RF) classifier is employed. The RF algorithm is an ensemble ML technique which uses a large number of tree-like classifiers in the ensemble and achieves a classification accuracy comparable to boosting technique [39]. RF is a very robust classifier against overfitting the training data and does not require any assumptions about the distribution of the data [40]. Furthermore, due to its ability to handle big, unbalanced, and high-dimensional data, it is one of the most popular machine learning

(ML) techniques for supervised classification of RS data, including hyperspectral imagery and lidar data [26, 40, 41]. In addition, the RF classifier estimates the importance of each feature in the feature vector of the training instances. This capability can be exploited to find the most discriminative features in both online and offline feature vectors.

Although RF classifier is not sensitive to the user-defined values for hyperparameters, in this study, the grid search method along with the 5-fold cross validation (CV) technique was employed to find the best settings for the hyper-parameters. The trained classifier is then used to evaluate the classification performance on two separate test sets given in Table 4.2. The best hyperparameter settings found for efficiently training the RF classifier over the campus dataset include 500 trees with a maximum depth of 20 for online waveform feature vectors and 1000 trees with a maximum depth of 50 for offline waveform feature vectors. The maximum number of features for splitting a node, minimum number of samples required for splitting a node, and minimum number of samples required in a leaf are 2, 2, and 2, respectively, for training the RF classifier with both online and offline waveform feature vectors. In addition, both RF classifiers use the bootstrap technique for sampling data points during training and validation.

4.5.6. Point cloud filtering for DTM generation

Discrimination between the ground and above-ground targets is one of the most interesting, yet challenging topics in the applications of lidar data, including TLS, for generating accurate an DTM in natural environments. In this work, classified TLS point cloud data collected at the coastal wetland study site are used to filter ground points from above-ground objects and subsequently generate a DTM of the wetland ground surface. To do so, the point cloud data classified by the online and offline waveform features using the proposed DCNN classifier are simply filtered according to their predicted label. The resulting filtered datasets include points related to the tidal

flat and dirt road areas within the study site, which are collectively called hard surface areas for the purposes herein. Those data sets are later used to generate the DTM model.

To evaluate the accuracy of the DTM generated from DCNN-based classification, a baseline ground point set is generated using the well-known progressive morphological filter (PMF) proposed by Zhang et al. [42]. The accuracy of the PMF filtering result is evaluated by computing the vertical distance from a triangulated irregular network (TIN) model generated from the PMF classified ground point set to the RTK GNSS points. The PMF DTM is then used to evaluate the accuracy of the DCNN-based classification of hard surface points, and subsequent DTM, based on online and offline waveform features.

4.6. Point Cloud Filtering for DTM Generation

4.6.1. Built environment classification using online waveform features from the October test set

Figure 4.5 visualizes the distribution of online waveform features for each target category from the online waveform feature vectors of the October dataset and Table 4.4 summarizes some statistics related to these features. It is worth noting that for better visualization of the feature distributions given in Figure 4.5, the upper bound of the x-axis in each plot is limited to the feature value that covers the distribution of 99.5% of the data.

For almost all target categories shown in Figure 4.5, the distribution of each feature has overlap with features in other target categories. This usually leads to high inter-class similarity for underlying target categories and consequently a decrease in classification performance. Referring to the figure, the range distribution plot shows a large overlap for all target categories. This plot simply shows that no specific target can be correctly classified based solely on its range from the scanner. The distributions for calibrated amplitude also show high overlap for different target

categories. However, the asphalt and grass classes show narrower amplitude distributions than the other targets. Building and tree trunk classes show the largest overlap in their amplitude distribution. In addition, asphalt and tree canopy show the largest overlap in amplitude distributions.

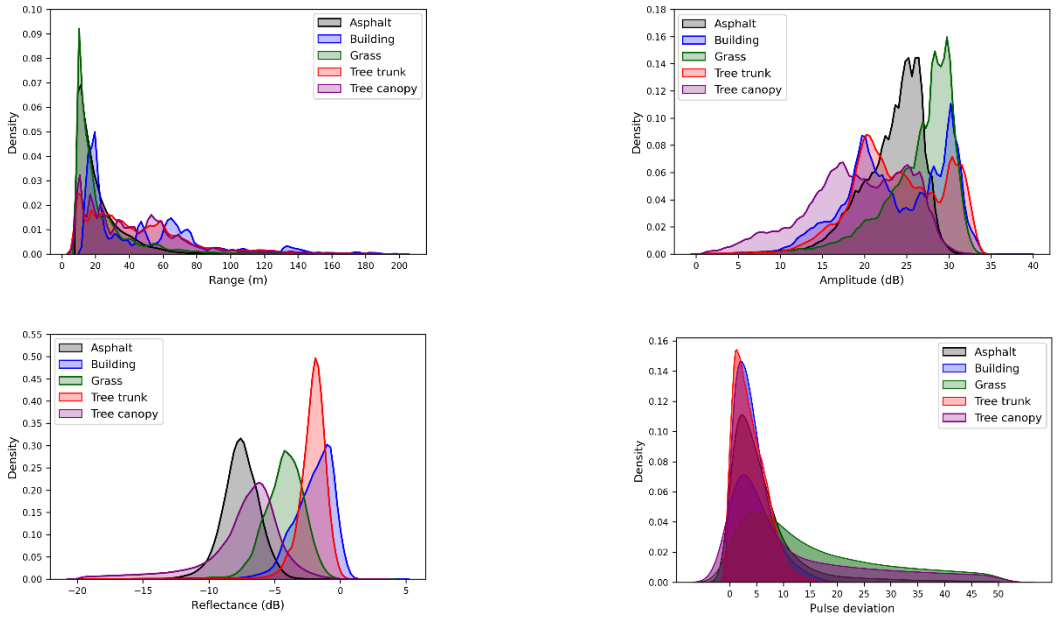


Figure 4.5. Distribution of the online waveform features for different targets in the training dataset.

It should be noted that the amplitude feature given by the online waveform processing is not calibrated with respect to the range in comparison to relative reflectance, and as such, the amplitude feature shows wider distributions and higher overlaps for almost all target categories. The plot representing the distribution of pulse shape, also, shows large overlap areas for different targets. Distributions of relative reflectance show the highest separability among different target categories with respect to the other online waveform features. That is expected due to the careful radiometric calibration of the TLS system by the manufacturer. However, different targets still show considerable overlap for reflectance values. The most noticeable overlaps are between asphalt and tree canopy classes and also between tree trunk and building classes. Furthermore,

except for the relative reflectance, the other online waveform features represent multimodal distributions.

Table 4.4. Summary of statistics for online waveform features in the training dataset. Each column gives the minimum, maximum, mean, and standard deviation of the related feature for the underlying target.

Target	Range (m)	Amplitude (m)	Reflectance (dB)	Pulse deviation
Asphalt	8.00, 199.70, 22.41, 15.10	0.34, 45.06, 23.77, 3.76	−20.07, 3.44, −7.64, 1.41	−1.00, 344.00, 7.43, 10.20
Building	13.61, 200.00, 48.56, 39.49	0.43, 43.73, 23.79, 6.01	−20.16, 2.49, −2.10, 1.68	−1.00, 311.00, 4.65, 4.13
Grass	8.35, 200.00, 23.93, 21.04	0.63, 35.59, 27.06, 4.33	−20.14, 5.79, −4.23, 1.46	−1.00, 400.00, 32.45, 37.56
Tree trunk	9.10, 200.00, 44.75, 31.93	0.40, 34.56, 26.27, 5.75	−20.06, 2.14, −2.17, 1.40	−1.00, 381.00, 3.70, 3.82
Tree canopy	8.56, 200.00, 42.83, 30.69	0.48, 35.36, 19.56, 6.28	−20.35, 3.33, −7.39, 2.98	−1.00, 422.00, 30.64, 43.75

Distributions of the reflectance attribute and its mean values for each individual target category, given in Figure 4.5 and Table 4.4, respectively, show its important role in separating asphalt and tree canopy from the building and tree trunk classes. It also helps to discriminate grass from all other target categories. Referring to the statistics reported in Table 4.4, mean reflectance shows comparable values between asphalt and tree canopy classes and also between building and tree trunk classes, making this feature less discriminative for instances in those target categories. Furthermore, pulse deviation is a more discriminative feature than calibrated amplitude. Referring to Figure 4.5 and Table 4.4, pulse deviation shows higher mean and standard deviation for grass

and tree canopy than other classes due to the spatial distribution of those targets in the path of the laser beam, making it a relatively strong feature for discriminating those classes from the others.

Fig. 6 illustrates the importance of each feature in the online waveform feature vectors, from the October training set, reported by the RF classifier. Using a boxplot to show feature importance also gives visual information about the distribution of features in the feature vector. As predicted earlier, the relative reflectance has the highest importance for classification based on the online waveform feature vector. The lower importance of the amplitude with respect to the range is partly due to the fact that the pulse amplitude is not compensated for range.

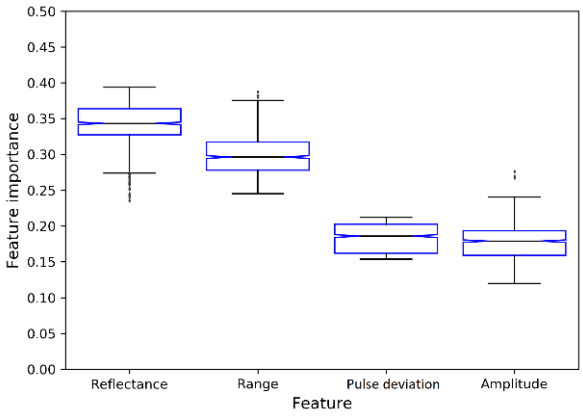


Figure 4.6. Feature importance from RF classifier trained on online waveform feature vectors.

Amplitude has the lowest discriminative capability in this classification experiment. However, the amplitude mean and standard deviation given in Table 4.4 and its density distribution plot shown in Figure 4.5, shows a degree of power for separating asphalt and tree canopy from grass instances.

The classification results for the online waveform feature vectors from the October test set, including and excluding the range attribute, using the RF classifier and the proposed DCNN-based classifier are given in Table 4.5 and Table 4.6, respectively. Each table summarizes the

performance of the underlying classifier using the confusion matrix, precision, recall, and F1-score for each individual target category. Furthermore, the weighted average of those metrics has also been reported, where the average metrics take into account the imbalance of the test set.

Table 4.5. RF-based Classification performance for online waveform features from the October test set. The values above and below the horizontal lines show the results for online feature vectors including and excluding the range values, respectively.

Reference points	Asphalt	Building	Grass	Tree trunk	Tree canopy	Precision	Recall	F1-score
Asphalt	0.80 0.78	0.01 0.01	0.07 0.07	0.00 0.00	0.13 0.15	0.72 0.71	0.80 0.78	0.75 0.74
Building	0.01 0.01	0.80 0.75	0.03 0.04	0.13 0.17	0.03 0.04	0.82 0.76	0.80 0.75	0.81 0.75
Grass	0.08 0.08	0.06 0.06	0.76 0.75	0.03 0.04	0.07 0.07	0.77 0.76	0.76 0.75	0.77 0.75
Tree trunk	0.00 0.00	0.11 0.18	0.04 0.04	0.84 0.76	0.01 0.01	0.82 0.76	0.84 0.76	0.83 0.76
Tree canopy	0.22 0.22	0.02 0.02	0.09 0.10	0.01 0.01	0.66 0.65	0.73 0.71	0.66 0.65	0.69 0.68
Weighted average						0.77 0.74	0.77 0.74	0.77 0.74
Overall accuracy					77% 74%			

According to Table 4.5 and Table 4.6, the overall classification accuracy reported from the RF classifier is comparable with that from DCNN model. In addition, excluding the range attribute from the online waveform feature vectors caused a decrease in the overall accuracy of about 3% for both classifiers. Both classifiers show similar performance in discriminating different target categories based on their online waveform features.

Table 4.6. DCNN-based Classification performance for online waveform features from the October test set. The values above and below the horizontal lines show the results for online feature vectors including and excluding the range values, respectively.

Reference points	Asphalt	Building	Grass	Tree trunk	Tree canopy	Precision	Recall	F1-score
Asphalt	0.85 0.85	0.00 0.00	0.06 0.06	0.00 0.00	0.09 0.08	0.70 0.70	0.85 0.85	0.77 0.77
Building	0.01 0.01	0.79 0.76	0.02 0.03	0.14 0.17	0.04 0.04	0.81 0.68	0.79 0.76	0.80 0.72
Grass	0.09 0.08	0.06 0.06	0.76 0.77	0.03 0.03	0.06 0.05	0.79 0.76	0.76 0.77	0.77 0.77
Tree trunk	0.00 0.00	0.12 0.30	0.04 0.05	0.83 0.63	0.01 0.01	0.81 0.73	0.83 0.63	0.82 0.68
Tree canopy	0.26 0.26	0.02 0.02	0.09 0.10	0.01 0.01	0.63 0.61	0.76 0.76	0.63 0.61	0.69 0.68
Weighted average						0.76 0.73	0.77 0.73	0.77 0.73
Overall accuracy					77% 73%			

Misclassified instances resulting from the classification of online waveform feature vectors, including and excluding the range attribute, follow the same pattern in Table 4.5 and Table 4.6. According to the F1-score values, both classifiers show the highest performance on building and tree trunk categories. However, they show a lower skill in detecting tree canopy instances. According to the confusion matrices given in Table 4.5 and Table 4.6, tree canopy has the highest rate of misclassified instances with asphalt. This observation was predictable by referring to the reflectance distribution plot given in Figure 4.5, where reflectance distributions for the asphalt and tree canopy classes shows the largest overlap. Buildings, on the other hand, shows the highest misclassified instances with tree trunk, which was, also, predictable by examining the plots in

Figure 4.5. In addition, referring to the confusion matrices, the grass category has about 25% misclassified instances which are distributed among other target categories.

It is worth noting that the significant difference between precision and recall for asphalt in both classifiers, given in Table 4.5 and Table 4.6, shows that despite the relatively large number of misclassified instances of asphalt in other classes, notably tree canopy and grass, both classifiers are still able to correctly detect a large portion of asphalt returns. Conversely, the relatively higher precision than recall for tree canopy class derived from both classification methods shows that the underlying classifier is more skillful in detecting instances that do not belong to the tree canopy than detecting instances that do actually belong to that class. The above classification results are consistent with the information retrieved from the feature distribution and feature importance plots given in Figure 4.5 and Figure 4.6, respectively.

4.6.2. Built environment classification using offline waveform features from the October test set

Figure 4.7 illustrates the feature importance plot reported by the RF classifier for training based on the offline waveform feature vectors related to the targets measured for the campus study site. S_1 to S_{24} in horizontal axis of the plot show sample indices for the measured waveforms. According to the plot, the range to the target has the highest importance for classification. This was predictable due to the high correlation between the intensity (amplitude) of the echo signal and range to the target. Referring to Figure 4.7, it is interesting to note that waveform samples related to the rise-time and fall-time of the return waveform, which usually happen around samples S_2 and S_6, respectively, in the single-peak echo waveforms measured by the Riegl VZ-400 TLS system, are more important than other samples. Furthermore, according to the plot, samples representing the rise-time and fall-time of the signal are almost equally important, with samples

closely representing the amplitude of the echo waveform. Analyzing the DNs for all recorded waveforms indicates that the peak of the echo signal usually occurs somewhere between the S_4 and S_6 samples. This observation confirms the importance of rise-time and fall-time of the echo waveform for classification in [28], where the authors highlight the importance of those features in the waveform feature vector for discriminating different tree types.

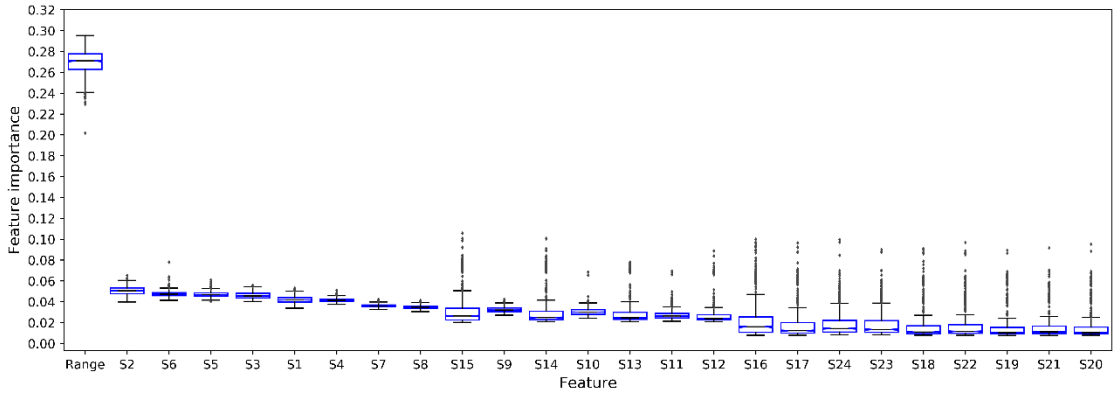


Figure 4.7. Feature importance from the RF classifier trained using offline waveform feature vectors.

According to Figure 4.7, waveform samples S_1-S_8 follow a symmetric distribution with a limited range of outliers, whereas the majority of samples related to the falling tail of the waveform, S_9-S_{24} follow asymmetric, positively skewed distributions with a larger range of outliers, which makes them less important features for efficiently training the classifier. In other words, samples related to the falling tail of the waveform carry less discriminative information for target classification.

Table 4.7. RF-based Classification performance for offline waveform features from the October test set. The values above and below the horizontal lines show the results for online feature vectors including and excluding the range values, respectively.

Reference points	Asphalt	Building	Grass	Tree trunk	Tree canopy	Precision	Recall	F1-score
Asphalt	0.89 0.81	0.00 0.05	0.04 0.03	0.00 0.05	0.07 0.06	0.82 0.64	0.88 0.80	0.85 0.72
Building	0.01 0.12	0.80 0.67	0.02 0.01	0.13 0.16	0.04 0.04	0.76 0.59	0.80 0.67	0.78 0.63
Grass	0.05 0.09	0.04 0.03	0.82 0.75	0.03 0.04	0.06 0.09	0.83 0.82	0.82 0.75	0.82 0.79
Tree trunk	0.00 0.14	0.22 0.33	0.03 0.03	0.73 0.44	0.02 0.06	0.79 0.55	0.73 0.43	0.76 0.78
Tree canopy	0.12 0.11	0.01 0.07	0.07 0.10	0.01 0.07	0.79 0.65	0.81 0.72	0.79 0.65	0.80 0.68
Weighted average						0.80 0.67	0.80 0.66	0.80 0.66
Overall accuracy					80% 67%			

The classification results for the offline waveform feature vectors from the October test set, including and excluding range, using the RF classified and the proposed DCNN-based classifier are given in Table 4.7 and Table 4.8, respectively. According to those tables and considering Table 4.5 and Table 4.6, which shows classification results of the online waveform feature vectors, the overall classification accuracy on the offline waveform feature vectors is 3% higher than that for the online waveform feature vectors, when the RF classifier is used for classification. However, the classification results on the offline waveform feature vectors using the DCNN-based classifier show a noticeable improvement of 10% in overall accuracy relative to the RF and DCNN-based classification performance using online waveform feature vectors. In addition, referring to Table 4.7, excluding range from the offline waveform feature vector reduces the performance of the RF by 13%, while this reduction, according to Table 4.8, is about 7% for classification based on the DCNN model.

Table 4.8. DCNN-based Classification performance for offline waveform features from the October test set. The values above and below the horizontal lines show the results for online feature vectors including and excluding the range values, respectively.

Reference points	Asphalt	Building	Grass	Tree trunk	Tree canopy	Precision	Recall	F1-score
Asphalt	0.94 0.87	0.00 0.05	0.02 0.03	0.00 0.02	0.04 0.03	0.89 0.73	0.93 0.87	0.91 0.80
Building	0.01 0.07	0.89 0.82	0.01 0.01	0.07 0.05	0.02 0.05	0.80 0.60	0.88 0.82	0.84 0.69
Grass	0.04 0.08	0.02 0.02	0.88 0.83	0.02 0.02	0.04 0.06	0.89 0.85	0.88 0.83	0.88 0.84
Tree trunk	0.00 0.09	0.17 0.40	0.02 0.03	0.80 0.42	0.01 0.06	0.83 0.77	0.80 0.42	0.82 0.54
Tree canopy	0.06 0.08	0.02 0.08	0.06 0.07	0.00 0.04	0.86 0.74	0.87 0.78	0.86 0.74	0.87 0.76
Weighted average						0.86 0.75	0.87 0.74	0.87 0.74
Overall accuracy					87% 74%			

Comparing Table 4.7 and Table 4.8 (offline waveform features) with Table 4.5 and Table 4.6 (online waveform features), shows a relatively similar pattern for misclassified class instances. Moreover, exploring the classification performance of the DCNN for each individual target category using online and offline waveform feature vectors as reported in Table 4.6 and Table 4.8, respectively, shows that classification based on the samples of the raw waveform significantly improves classification performance across almost all classes.

Interestingly, the RF and DCNN-based classification results for the tree trunk category shown in Table 4.5 and Table 4.6 (online waveform features) when compared to results shown in Table 4.7 and Table 4.8 (offline waveform features) reveals that higher classification performance can be achieved for this target category using the online waveform features rather than the raw

waveform samples. Referring to the reflectance distribution plot given in Figure 4.5, this may be due to the relatively narrow distribution of the calibrated reflectance attribute related to tree trunks that may help the underlying classifier more effectively detect instances in that category.

4.6.3. Built environment classification using online/offline waveform features from the July test set

The classification results for both online and offline waveform feature vectors from the July test set using the DCNN-based classifier is given in Table 4.9. According to the table the overall accuracy of the classification on offline waveform features is 15% higher than that for the online waveform features. Discrepancies between the mean and standard deviation of each individual feature in the October versus July test set, for the online and offline waveform feature vectors, are given in Figure 4.8 and Figure 4.9, respectively. Because the equivalent waveform data and 3D points with online waveform features are used in both October and July datasets, for simplicity, the mean and standard deviation of the range value for each target category have only been displayed in Figure 4.8.

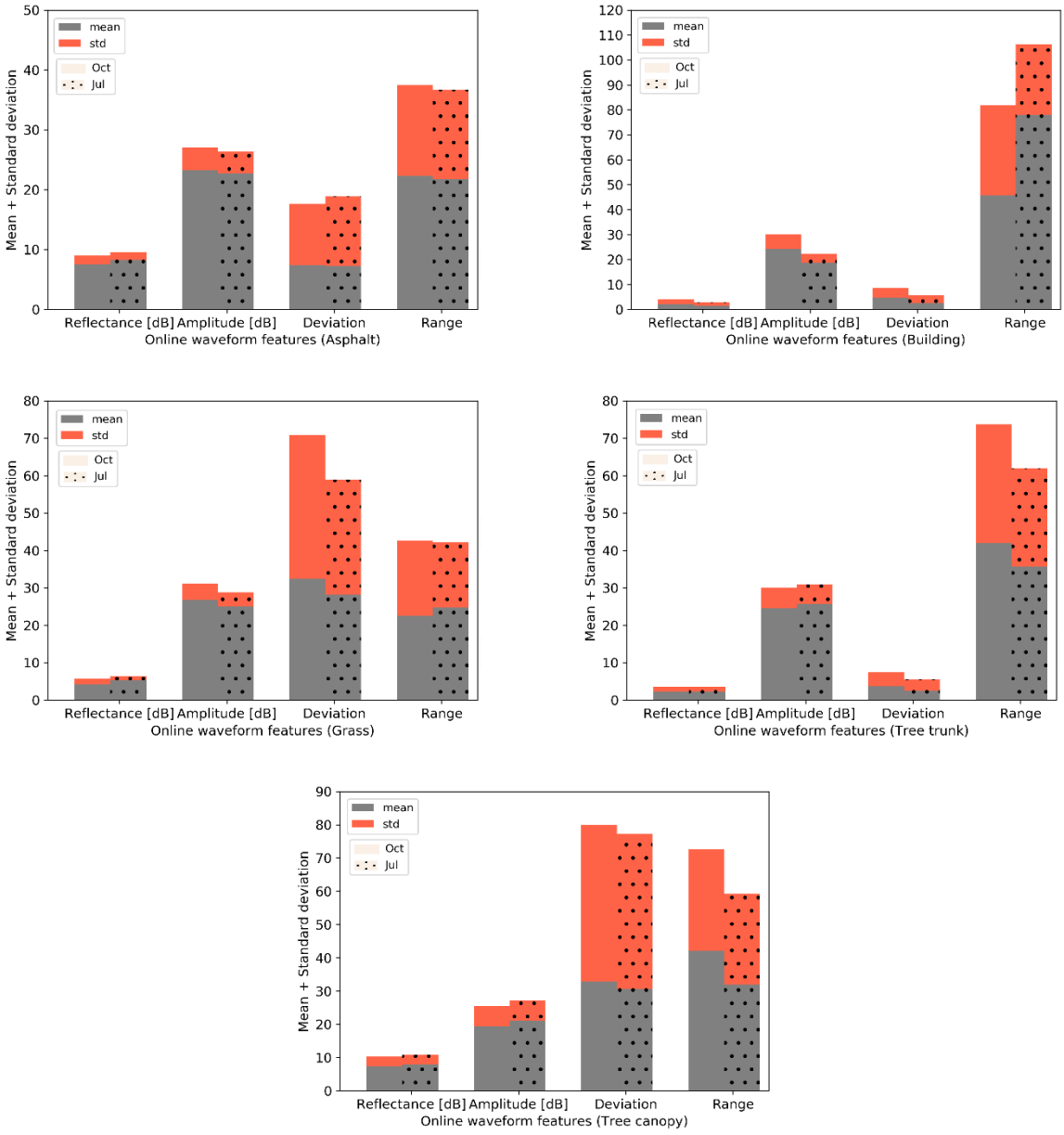


Figure 4.8. Discrepancies in the mean and standard deviation of online waveform attributes for different target categories measured at two different points in time.

Referring to Figure 4.8, the mean and standard deviation of the range value in the October and July datasets are more similar for the asphalt category and less comparable for other target categories. The larger the separation of the range statistics between the two datasets for a certain

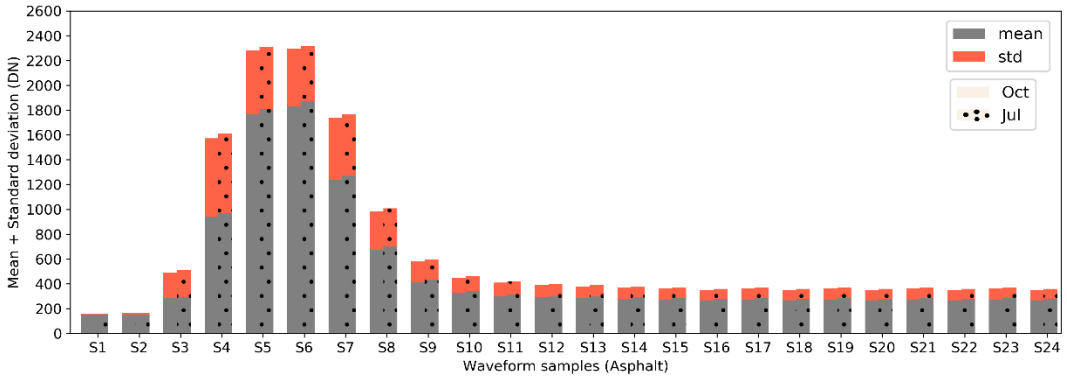
target, the larger the differences observed in reflectance values for that target. Consequently, the differences in the classification performance for that target category across the two survey dates is also larger.

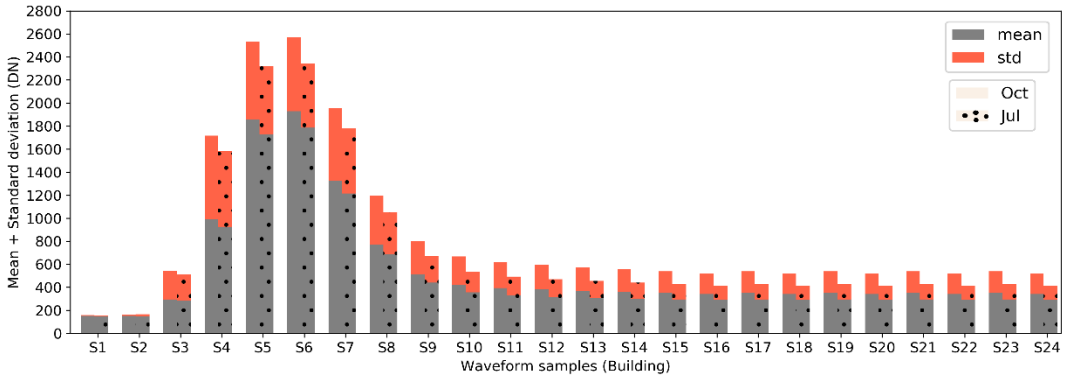
Table 4.9. DCNN-based classification performance for the July test set. The values above and below the horizontal lines show the results for offline and online feature vector classification, respectively.

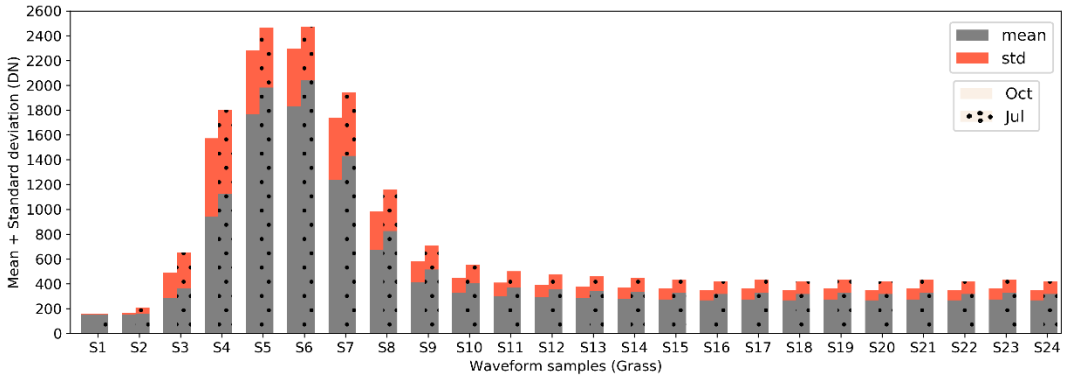
Reference points	Asphalt	Building	Grass	Tree trunk	Tree canopy	Precision	Recall	F1-score
Asphalt	0.81 0.89	0.07 0.00	0.01 0.01	0.00 0.00	0.11 0.10	0.86 0.64	0.93 0.89	0.89 0.74
Building	0.00 0.00	0.81 0.59	0.01 0.00	0.15 0.35	0.04 0.05	0.69 0.69	0.89 0.59	0.77 0.64
Grass	0.09 0.16	0.07 0.04	0.78 0.64	0.00 0.01	0.06 0.16	0.87 0.83	0.85 0.67	0.86 0.72
Tree trunk	0.00 0.01	0.36 0.41	0.04 0.05	0.58 0.51	0.02 0.01	0.87 0.42	0.59 0.51	0.70 0.46
Tree canopy	0.14 0.35	0.02 0.02	0.08 0.09	0.00 0.00	0.76 0.53	0.85 0.63	0.82 0.53	0.83 0.58
Weighted average						0.82 0.64	0.81 0.63	0.81 0.63
Overall accuracy					81% 65%			

Considering the F1-scores given in Table 4.9 and comparing those values with equivalent values given in Table 4.6 and Table 4.8 confirm that discrepancies in the mean and standard deviation for each individual sample of the raw waveform collected at two different points in time have relatively less impact on the classification performance than differences in the statistics related to the online waveform features. The drop in the classification performance over natural targets can be partly due to the impact of seasonal changes on some properties of those targets, where, for example, green and dry grass or tree canopy represent changing backscattering

properties. Moreover, atmospheric attenuation factors on the laser energy, such as the humidity index, air pressure, and temperature, related to each collected dataset contribute to the related echo waveforms and subsequently derived online waveform attributes, resulting in different misclassification rates in one dataset relative to the other.







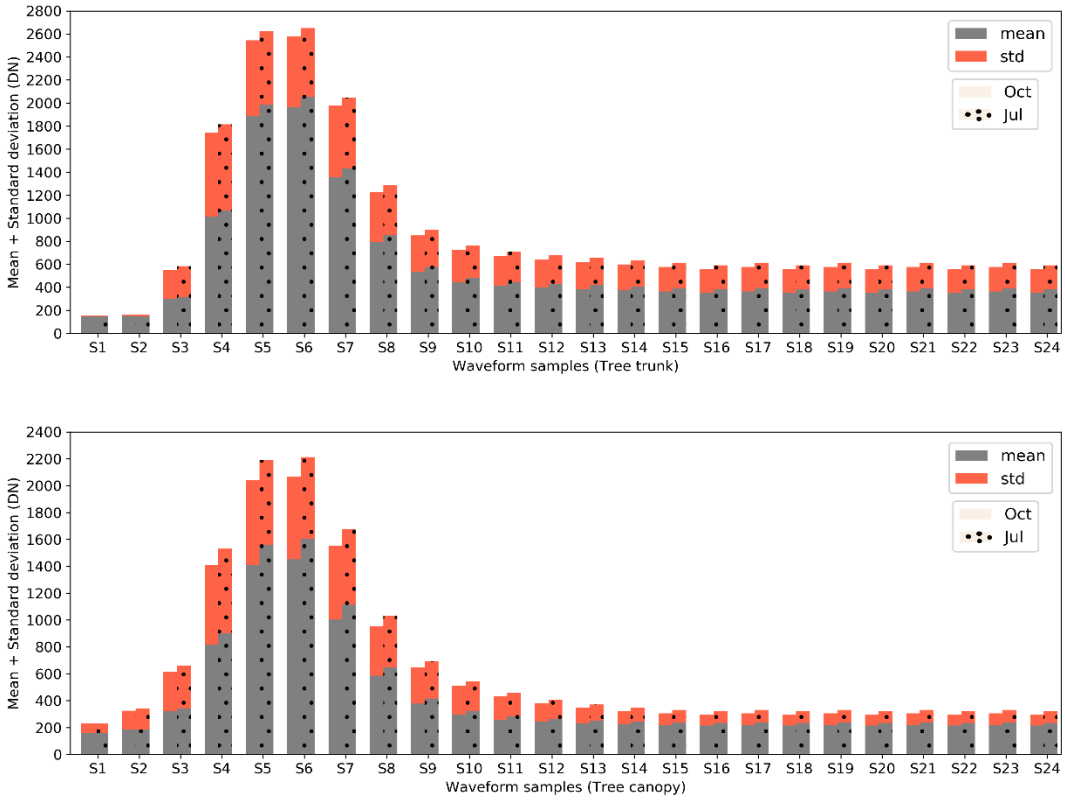
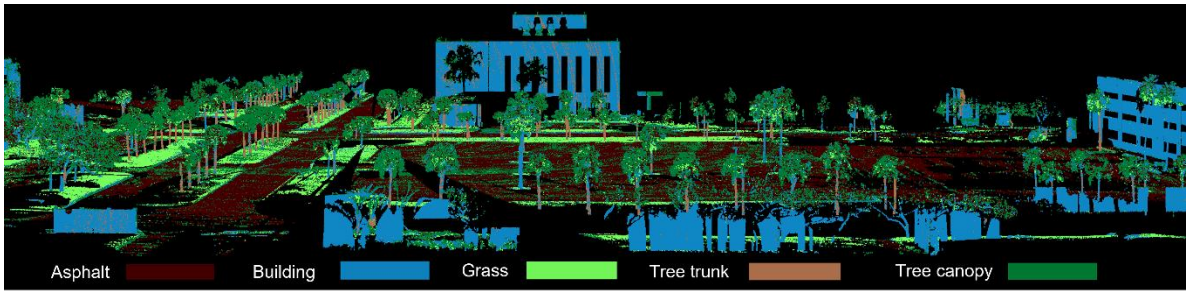
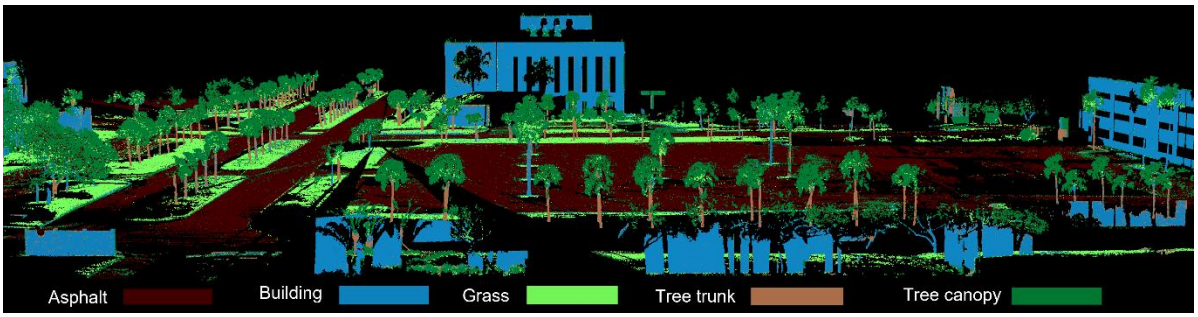


Figure 4.9. Discrepancies in the mean and standard deviation of waveform samples for different target categories measured at two different points in time.

Finally, Figure 4.10 illustrates the qualification of the classification performance for both the online and offline waveform data from the October test set. According to the figure, it is clear that the misclassified building and tree trunk instances in the online waveform classification is higher than the offline waveform classification. Also, the higher rate of misclassified instances for asphalt and tree canopy in the online waveform classification can be seen in Figure 4.10.



a) Online waveform



b) Offline waveform

Figure 4.10. Qualification of classification over the campus study area using online and offline waveform feature vectors from October test set.

4.6.4. Natural environment classification

The performance of the proposed DCNN-based classifier on both online and offline waveform feature vectors derived from the TLS survey over the coastal wetland study site is given in Table 4.10. According to the table, the overall accuracy of the multi-class classification for the coastal wetland using the proposed DCNN model on offline waveform feature vectors is 13% higher than that for the online waveform feature vectors. The F1-score reported in Table 4.10, shows that the classification performance on online and offline feature vectors related to both tidal flat and vegetation are more comparable than the performance for the road and power line classes. The calibrated reflectance feature in the tidal flat areas and vegetation areas shows that this online

waveform attribute can easily discriminate a large number of instances belonging to those categories.

Table 4.10. DCNN-based Classification performance on the coastal wetland area. The values above and below the horizontal lines show the results for offline and online feature vector classification, respectively.

Reference points	Tidal flat	Vegetation	Road	Power line	Precision	Recall	F1-score
Tidal flat	0.98 0.97	0.01 0.02	0.01 0.01	0.00 0.00	0.92 0.81	0.98 0.97	0.95 0.88
Vegetation	0.01 0.04	0.98 0.95	0.00 0.01	0.01 0.01	0.95 0.80	0.98 0.95	0.97 0.87
Road	0.12 0.36	0.00 0.40	0.87 0.24	0.00 0.00	0.97 0.89	0.87 0.24	0.92 0.38
Power line	0.09 0.19	0.15 0.20	0.00 0.01	0.76 0.60	0.96 0.95	0.76 0.60	0.85 0.74
Weighted average					0.95 0.83	0.95 0.82	0.94 0.79
Overall accuracy				95% 82%			

The higher performance of the classification based on the raw waveform samples relative to the online waveform attributes is more noticeable for the road and power line categories. The road class at this study site is comprised of dirt and sediment, similar in composition to the upland less submerged parts of the tidal flat area, making these two areas challenging for classification. As observed in Table 4.10, classification of the road based on the offline waveform features had a significantly higher classification accuracy compared to classification of the road based on the online waveform features (92% F1-score versus 36%, respectively). According to the confusion matrix results there are large number of misclassified instances with other target categories for online waveform features. Although instances related to the tidal flat and road show very close

calibrated reflectance values in their online waveform feature vectors, results suggest that samples of the raw waveform significantly improved discrimination of those two target categories, perhaps due to differences in surface roughness. The qualitative results of the DCNN-based classification for both the online and offline waveform feature vectors are given in Figure 4.11.

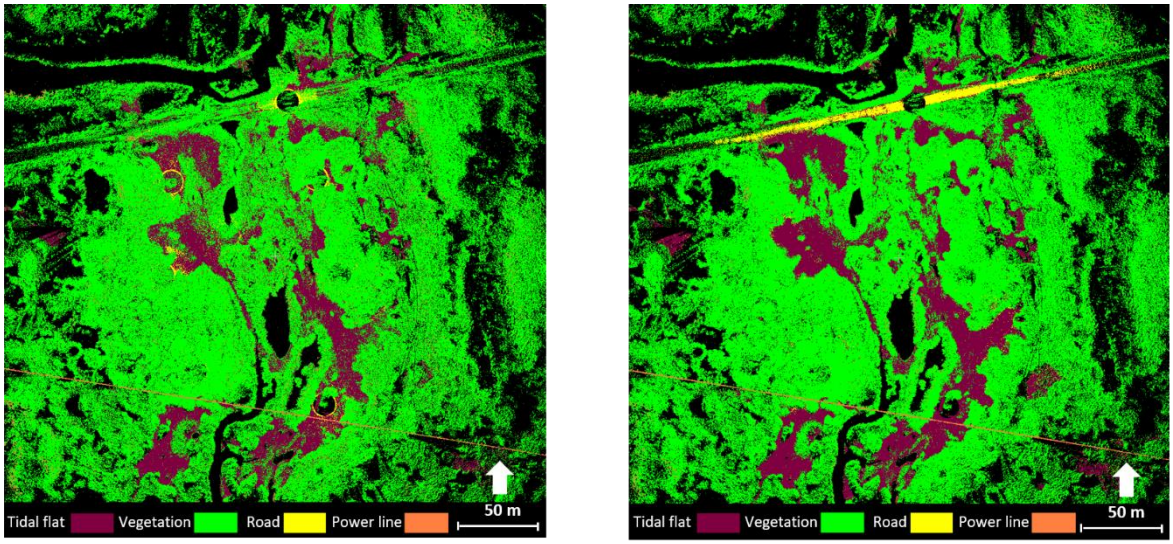
Table 4.11. Statistics of vertical error (m) between Riegl VZ-2000i TLS measurements and RTK GNSS points collected on hard surfaces and vegetated surfaces before and after applying PMF.

		Before PMF	After PMF
Statistics (m)	$\Delta Z^{TLS-GNSS}_{\text{hard surfaces}}$	$\Delta Z^{TLS-GNSS}_{\text{vegetated surfaces}}$	$\Delta Z^{TLS-GNSS}_{\text{vegetated surfaces}}$
Mean	0.009	0.116	0.028
Min	-0.017	-0.029	-0.020
Max	0.049	0.711	0.130
St. Dev	0.019	0.159	0.062
$RMSE_Z$	0.021	0.197	0.068

4.6.5. Terrain surface modeling for the coastal wetland site

The classified points based on both online and offline waveform features are used to approximate terrain models (DTMs) for the coastal wetland study site. Classified points are filtered based on their predicted labels from the proposed DCNN classifier, where the ground points refer to the set of points predicted as road or tidal flat areas. Recall the tidal flat class includes exposed, lower lying and tidal inundated wetland surface areas and upland, periodically inundated wetland surface areas in proximity to sparse or dense vegetation. To evaluate the fidelity of DTMs generated from the DCNN-based filtering result with online and offline waveform features, a

classified set of ground points output from the PMF filter applied to the original TLS point cloud is used.



(a) Online waveform (b) Offline waveform

Figure 4.11. Qualification of the classification over the study area using online and offline waveform feature vectors.

The vertical differences between the RTK GNSS points collected on the exposed wetland/tidal flat surfaces and road surfaces, here called hard surfaces for brevity, and a local TIN model constructed from the original TLS points shows a bias of $+0.009\ m$, which is in the range of the vertical accuracy of the RTK GNSS survey method employed. Table 4.11 reports vertical accuracy statistics for the TLS data relative to the RTK GNSS points collected over hard surfaces and vegetated surfaces, both dense and sparsely vegetated, before and after applying PMF filtering. In addition, the scatterplot for the RTK GNSS ellipsoid heights compared to the TLS measured ellipsoid heights on hard surfaces and vegetated surfaces before and after applying PMF filtering are shown in Figure 4.12. The goodness-of-fit or coefficient of determination, r^2, is 0.95 for regression lines related to the hard surfaces in both plots. Whereas, the r^2 coefficient is 0.81 and

0.92 for the regression lines representing the height difference over vegetated surfaces before and after applying PMF, respectively.

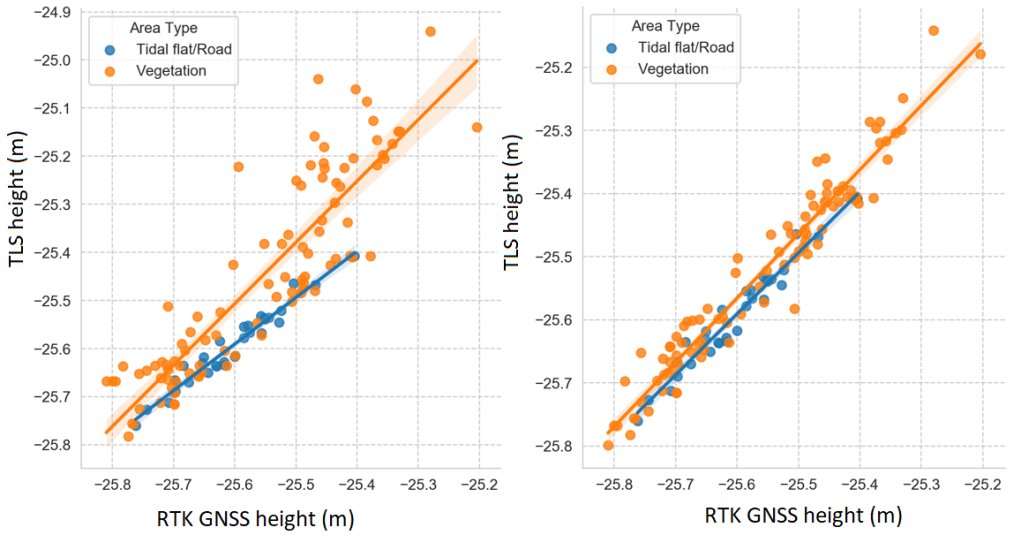


Figure 4.12. Scatterplot of RTK GNSS ellipsoid heights versus TLS ellipsoid heights on hard surfaces and vegetated surfaces before (left) and after (right) applying PMF.

The statistics given in Table 4.11 and plot in Figure 4.12, clearly show that the vegetated areas can cause a significant bias in the process of modeling the terrain surface, as expected. Moreover, statistics and regression plots given in Table 4.11 and Figure 4.12, respectively, show the high performance of the PMF algorithm in filtering the above-ground targets and identifying the ground points in the vegetated areas. This justifies the use of the PMF filtering solution as a representative ground point set for evaluating performance of the DCNN-based classification of hard surfaces (tidal flat areas and road areas) using online and offline waveform feature vectors.

Table 4.12 summarizes statistics related to the vertical distance between a TIN surface model constructed from the PMF ground point set, using LAStools (rapidlasso GmbH) point cloud processing software, and the classified point set on hard surfaces resulting from the DCNN-based classification of both the online and offline waveform feature vectors. According to the table,

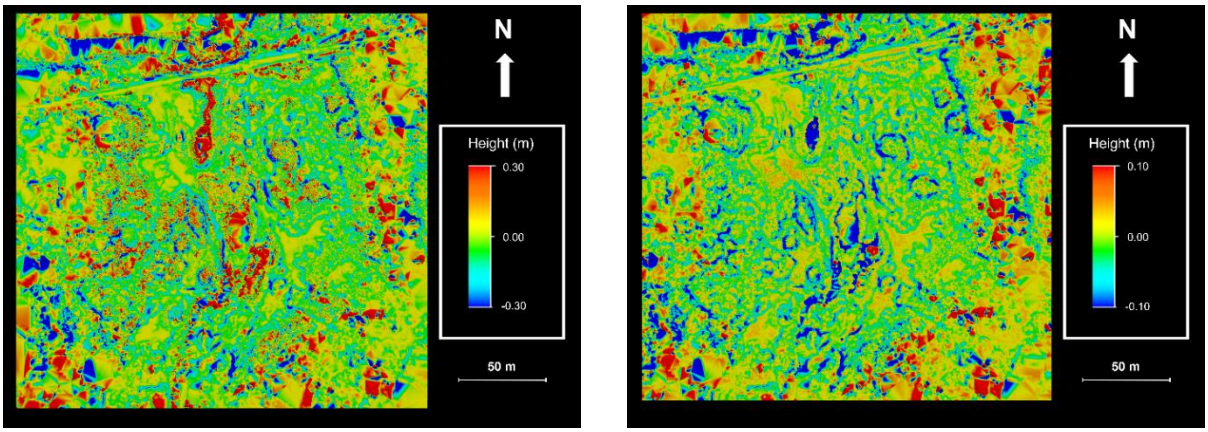
predicted tidal flat and road points from offline waveform features can model the terrain surface with the uncertainty of about one order of magnitude lower than that from predicted points based on the online waveform features.

Table 4.12. Statistics of vertical distance between TIN surface constructed on terrain points derived from PMF and classified terrain points, including tidal flat and road, derived from DCNN-based classification on offline and online waveform feature vectors.

Statistics (m)	$\mathrm{DTM}^{\mathrm{Offline}} - \mathrm{DTM}^{\mathrm{PMF}}$	$\mathrm{DTM}^{\mathrm{Online}} - \mathrm{DTM}^{\mathrm{PMF}}$
Mean	0.000	0.004
Min	−0.020	−0.090
Max	0.212	0.430
St. Dev	0.005	0.040
RMSE_Z	0.005	0.040

Figure 4.13 shows the differences between a DTM generated from the PMF ground point set and DTMs generated from the DCNN-based classified points on hard surfaces, including tidal flat and road areas, using online and offline waveform feature vectors. All DTMs have been generated using LAStools software with a given step size (resolution) of $0.1\,m$ in both X and Y directions. The DTM generated from the offline waveform classification result more closely approximates the DTM generated from the PMF ground point set for both densely vegetated areas, in the middle and upper part of the figure, and sparse vegetation areas on the left and right side of the figure. According to Figure 4.13, the range of uncertainty in terrain height on classified tidal flat and road resulted from the online waveform features is significantly higher than that from the offline waveform features. The main reason for the higher vertical uncertainty for classified points from the online waveform features relative to the offline waveform features is the higher rate of

misclassified instances of vegetation with tidal flat in the classification based on the online waveform features which results in lower precision value for the tidal flat category. Also, the larger rate of misclassification between tidal flat and vegetation instances in online waveform classification caused a higher vertical uncertainty over the vegetated areas.



(a) $DTM^{Online} - DTM^{PMF}$ (b) $DTM^{Offline} - DTM^{PMF}$

Figure 4.13. Differential DTMs computed by subtracting the DCNN-based DTM, computed from online and offline waveform features, from the PMF-based DTM.

4.7. Conclusion

In this study, the potential of the raw samples of TLS single-peak echo waveforms versus calibrated waveform features from online waveform processing, were explored for point cloud classification within built and natural environments. FW data were collected by the Riegl VZ-Line FW TLS systems in multiple scan positions in each study area, where in addition to the 3D coordinates for each measurement, the calibrated waveform features, from the online waveform processing, and equivalent digitized waveform data were recorded. Also, a DCNN-based classifier was proposed for both online and offline waveform feature vector classification, where its performance was compared with the performance achieved based on RF classification on the same

datasets, and feature importance in each feature vector (online versus offline) was reported. This experiment showed that the samples of the digitized waveform can be more discriminative for certain target classes than the limited number of calibrated waveform features from online waveform processing, which resulted in higher overall classification performance. Furthermore, exploring feature statistics related to both the online and offline features for each individual target showed that they undergo some discrepancies over time. These might be due to the variation in some environmental factors during data collection, including weather conditions, or the internal fluctuations in the transmitted laser pulse over time. However, according to the results, the offline waveform feature vector shows more temporal stability than the equivalent online waveform feature vector.

Results for the selected wetland environment, showed that the classification based on samples of the raw waveform outperforms that on the calibrated waveform features. In addition, a filtering procedure to discriminate terrain points based on the predicted label for TLS measurements is more accurate when the classified dataset derived from a raw waveform classification rather than classifying using calibrated waveform features.

The approach for FW TLS data classification based on the raw waveform samples proposed in this work is adaptable to FW airborne lidar and other modalities. The approach is especially useful when the lidar system response for modeling the waveform is complicated or unknown. It is also advantageous where, due to a low sampling rate of the digitizer such as is common in FW TLS systems, accurate modeling of the waveform signal may not be practically feasible.

Some limitations related to the proposed point cloud classification approach should be kept in mind. This approach uses only single-peak echo waveforms for classification. Moreover, the proposed DCNN model has a relatively simple architecture for feature encoding. In addition, to

have a more accurate assessment on the potential of the proposed classification approach, it should be evaluated on more complex built and natural environments with more target categories. Substantial variations of the returned waveforms and consequently the derived cross-section (calibrated reflectance) values over distances shorter than range resolution, as an inherent limitation of any lidar system (TLS or ALS), should also be considered when interpreting the lidar data or classification performance over complex environments.

As future work, more advanced DCNN-based architectures may be developed to more effectively explore waveform feature space for classification. Classification of multi-echo waveforms may also be considered as future work. The capability of the proposed classification approach will also be assessed on more complex built and natural environments with more sophisticated target categories. In addition, the proposed raw waveform classification approach can be employed for advanced target identification and filtering procedures in complex environments where the inclusion of geometric information to the feature vector of each individual measurement can boost the performance of the FW lidar data analysis.

4.8. References

1 Guarnieri, A., Pirotti, F., and Vettore, A.: 'Comparison of discrete return and waveform terrestrial laser scanning for dense vegetation filtering', International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012, 39, (B7), pp. 511-516

2 Ullrich, A., and Pfennigbauer, M.: 'Echo digitization and waveform analysis in airborne and terrestrial laser scanning', in Editor (Ed.)^(Eds.): 'Book Echo digitization and waveform analysis in airborne and terrestrial laser scanning' (2011, edn.), pp. 217-228

3 Wagner, W.: 'Radiometric calibration of small-footprint full-waveform airborne laser scanner measurements: Basic physical concepts', ISPRS Journal of Photogrammetry and Remote Sensing, 2010, 65, (6), pp. 505-513

4 Mallet, C., and Bretar, F.: 'Full-waveform topographic lidar: State-of-the-art', ISPRS Journal of photogrammetry and remote sensing, 2009, 64, (1), pp. 1-16

5 Höfle, B., Hollaus, M., and Hagenauer, J.: 'Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data', ISPRS Journal of Photogrammetry and Remote Sensing, 2012, 67, pp. 134-147

6 Reitberger, J., Krzystek, P., and Stilla, U.: 'Analysis of full waveform LIDAR data for the classification of deciduous and coniferous trees', Int J Remote Sens, 2008, 29, (5), pp. 1407-1431

7 Chauve, A., Mallet, C., Bretar, F., Durrieu, S., Pierrot-Deseilligny, M., and Puech, W.: 'Processing full-waveform lidar data: modelling raw signals', in Editor (Ed.)^(Eds.): 'Book Processing full-waveform lidar data: modelling raw signals' (2007, edn.), pp. 102-107

8 Stilla, U., Yao, W., and Jutzi, B.: 'Detection of weak laser pulses by full waveform stacking', International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2007, 36, (Part 3), pp. W49A

9 Jutzi, B., and Stilla, U.: ‘Range determination with waveform recording laser systems using a Wiener Filter’, ISPRS Journal of Photogrammetry and Remote sensing, 2006, 61, (2), pp. 95-107

10 Persson, Å., Söderman, U., Töpel, J., and Ahlberg, S.: ‘Visualization and analysis of full-waveform airborne laser scanner data’, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2005, 36, (3/W19), pp. 103-108

11 Hartzell, P.J., Glennie, C.L., and Finnegan, D.C.: ‘Empirical waveform decomposition and radiometric calibration of a terrestrial full-waveform laser scanner’, IEEE Transactions on Geoscience and Remote Sensing, 2014, 53, (1), pp. 162-172

12 Pfennigbauer, M., and Ullrich, A.: ‘Improving quality of laser scanning data acquisition through calibrated amplitude and pulse deviation measurement’, in Editor (Ed.)^(Eds.): ‘Book Improving quality of laser scanning data acquisition through calibrated amplitude and pulse deviation measurement’ (International Society for Optics and Photonics, 2010, edn.), pp. 76841F

13 Parrish, C.E., Jeong, I., Nowak, R.D., and Smith, R.B.: ‘Empirical comparison of full-waveform lidar algorithms’, Photogrammetric Engineering & Remote Sensing, 2011, 77, (8), pp. 825-838

14 Mallet, C., Soergel, U., and Bretar, F.: ‘Analysis of full-waveform lidar data for classification of urban areas’, in Editor (Ed.)^(Eds.): ‘Book Analysis of full-waveform lidar data for classification of urban areas’ (2008, edn.), pp.

15 Alexander, C., Tansey, K., Kaduk, J., Holland, D., and Tate, N.J.: ‘Backscatter coefficient as an attribute for the classification of full-waveform airborne laser scanning data in urban areas’, ISPRS Journal of Photogrammetry and Remote Sensing, 2010, 65, (5), pp. 423-432

16 Mallet, C., Bretar, F., Roux, M., Soergel, U., and Heipke, C.: 'Relevance assessment of full-waveform lidar data for urban area classification', ISPRS journal of photogrammetry and remote sensing, 2011, 66, (6), pp. S71-S84

17 Fieber, K.D., Davenport, I.J., Ferryman, J.M., Gurney, R.J., Walker, J.P., and Hacker, J.M.: 'Analysis of full-waveform LiDAR data for classification of an orange orchard scene', ISPRS Journal of Photogrammetry and Remote Sensing, 2013, 82, pp. 63-82

18 Azadbakht, M., Fraser, C.S., and Khoshelham, K.: 'Synergy of sampling techniques and ensemble classifiers for classification of urban environments using full-waveform LiDAR data', International journal of applied earth observation and geoinformation, 2018, 73, pp. 277-291

19 Lai, X., Yuan, Y., Li, Y., and Wang, M.: 'Full-waveform LiDAR point clouds classification based on wavelet support vector machine and ensemble learning', Sensors-Basel, 2019, 19, (14), pp. 3191

20 Reitberger, J., Krzystek, P., and Heurich, M.: 'Full-waveform analysis of small footprint airborne laser scanning data in the Bavarian forest national park for tree species classification', 3D Remote Sensing in Forestry, 2006, 218, pp. 227

21 Ma, L., Zhou, M., and Li, C.: 'Land covers classification based on Random Forest method using features from full-waveform LiDAR data', International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017, 42, (2/W7)

22 Bruggisser, M., Roncat, A., Schaepman, M.E., and Morsdorf, F.: 'Retrieval of higher order statistical moments from full-waveform LiDAR data for tree species classification', Remote Sens Environ, 2017, 196, pp. 28-41

23 Crespo-Peremarch, P., Fournier, R.A., Nguyen, V.-T., van Lier, O.R., and Ruiz, L.Á.: 'A comparative assessment of the vertical distribution of forest components using full-waveform

airborne, discrete airborne and discrete terrestrial laser scanning data', Forest Ecology and Management, 2020, 473, pp. 118268

24 Chen, D., Peethambaran, J., and Zhang, Z.: 'A supervoxel-based vegetation classification via decomposition and modelling of full-waveform airborne laser scanning data', Int J Remote Sens, 2018, 39, (9), pp. 2937-2968

25 Höfle, B., and Hollaus, M.: 'Urban vegetation detection using high density full-waveform airborne lidar data-combination of object-based image and point cloud analysis' (na, 2010. 2010)

26 Guo, L., Chehata, N., Mallet, C., and Boukir, S.: 'Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests', ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66, (1), pp. 56-66

27 Neuenschwander, A., Magruder, L., and Gutierrez, R.: 'Signal processing techniques for feature extraction and classification using small-footprint full-waveform airborne LIDAR', in Editor (Ed.)^(Eds.): 'Book Signal processing techniques for feature extraction and classification using small-footprint full-waveform airborne LIDAR' (IEEE, 2008, edn.), pp. III-676-III-679

28 Neuenschwander, A.L., Magruder, L.A., and Tyler, M.: 'Landcover classification of small-footprint, full-waveform lidar data', Journal of applied remote sensing, 2009, 3, (1), pp. 033544

29 Crespo-Peremarch, P., Tompalski, P., Coops, N.C., and Ruiz, L.Á.: 'Characterizing understory vegetation in Mediterranean forests using full-waveform airborne laser scanning data', Remote Sens Environ, 2018, 217, pp. 400-413

30 Luo, S., Wang, C., Xi, X., Nie, S., Fan, X., Chen, H., Ma, D., Liu, J., Zou, J., and Lin, Y.: 'Estimating forest aboveground biomass using small-footprint full-waveform airborne LiDAR data', International Journal of Applied Earth Observation and Geoinformation, 2019, 83, pp. 101922

31 Doneus, M., Pfennigbauer, M., Studnicka, N., and Ullrich, A.: 'Terrestrial waveform laser scanning for documentation of cultural heritage', in Editor (Ed.)^(Eds.): 'Book Terrestrial waveform laser scanning for documentation of cultural heritage' (2009, edn.), pp.

32 Zorzi, S., Maset, E., Fusiello, A., and Crosilla, F.: 'Full-waveform airborne LiDAR data classification using convolutional neural networks', IEEE transactions on geoscience and remote sensing, 2019, 57, (10), pp. 8255-8261

33 Azadbakht, M., Fraser, C., and Khoshelham, K.: 'Improved urban scene classification using full-waveform lidar', Photogrammetric Engineering & Remote Sensing, 2016, 82, (12), pp. 973-980

34 Shen, X., Cao, L., Chen, D., Sun, Y., Wang, G., and Ruan, H.: 'Prediction of forest structural parameters using airborne full-waveform LiDAR and hyperspectral data in subtropical forests', Remote Sensing, 2018, 10, (11), pp. 1729

35 Sun, C., Cao, S., and Sanchez-Azofeifa, G.A.: 'Mapping tropical dry forest age using airborne waveform LiDAR and hyperspectral metrics', International Journal of Applied Earth Observation and Geoinformation, 2019, 83, pp. 101908

36 Rawat, W., and Wang, Z.: 'Deep convolutional neural networks for image classification: A comprehensive review', Neural computation, 2017, 29, (9), pp. 2352-2449

37 Liao, W., Van Coillie, F., Gao, L., Li, L., Zhang, B., and Chanussot, J.: 'Deep learning for fusion of APEX hyperspectral and full-waveform LiDAR remote sensing data for tree species mapping', IEEE Access, 2018, 6, pp. 68716-68729

38 Kingma, D.P., and Ba, J.: 'Adam: A method for stochastic optimization', arXiv preprint arXiv:1412.6980, 2014

39 Breiman, L.: 'Random forests', Machine learning, 2001, 45, (1), pp. 5-32

40 Gislason, P.O., Benediktsson, J.A., and Sveinsson, J.R.: ‘Random forests for land cover classification’, Pattern recognition letters, 2006, 27, (4), pp. 294-300

41 Jin, Y., Liu, X., Chen, Y., and Liang, X.: ‘Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: A case study of central Shandong’, Int J Remote Sens, 2018, 39, (23), pp. 8703-8723

42 Zhang, K., Chen, S.-C., Whitman, D., Shyu, M.-L., Yan, J., and Zhang, C.: ‘A progressive morphological filter for removing nonground measurements from airborne LIDAR data’, IEEE transactions on geoscience and remote sensing, 2003, 41, (4), pp. 872-882

CHAPTER V: CONCLUSION

5.1. Summary

Application of RS is important for efficient and precise monitoring and modeling of land cover and topography within dynamic coastal environments. The focus of this dissertation was on the application of DL-based techniques, in particular DCNN architectures, to retrieve useful information from hyperspatial RS data collected over a coastal environment by advanced geodetic imaging technologies including UAS-SfM and FW TLS. Retrieved geospatial information can be utilized to better understand, monitor, and model the topography and spatial distribution of different land cover targets in coastal environments. The developed methods and techniques enable processing of large and complex 2D/3D data streams collected over coastal environments in DL framework for the highest possible information gain from raw data.

In Chapter II of this work, as the first contribution, some of the most advanced DCNN architectures developed for different applications were evaluated for land cover prediction using UAS imagery collected over a complex coastal wetland study area. The main objective of this study was to investigate the generalization capacity of advanced DCNN architectures, originally developed for different image-based analyses in other applications, for land cover prediction using hyperspatial UAS RGB images acquired over the wetland environment. It also explored transfer learning, due to UAS data scarcity, to train some of the most popular DCNN architectures developed for pixel-wise image segmentation and find the most efficient architectures for land cover classification in the study area.

Results showed that employed DCNN models for semantic image segmentation can be practically fine-tuned for land cover prediction in a complex environment, such as a coastal wetland, which is, inherently, a challenging task in RS applications. DCNN-based hyperspatial

UAS image segmentation is able to exploit transfer learning to effectively train the most advanced and efficient DCNN models, including FC-DenseNet and U-Net, for accurate land cover prediction with limited training data. It is important to emphasize that by exploiting transfer learning to fine-tune weights in the underlying DCNN model, the overall accuracy of pixel-wise UAS image segmentation is comparable for almost all DCNN models. One reason for this observation is the effectiveness of DCNN architectures in feature exploration and learning the most discriminative features from input data in a hierarchical fashion. According to the findings in this study, FC-DenseNet and U-Net showed higher overall accuracy for land cover classification in the coastal wetland area. Also, U-Net model represented the fastest training phase among other DCNN models.

It is worth noting that different coastal wetlands and other landscapes may introduce different levels of complexity in the spatial distributions and radiometric characteristics of targets. This can lead to the ineffectiveness of the DCNN model, trained on data collected on one study site, to predict the land cover for a different study site. However, due to the fact that DL models, including DCNN architectures, are trained in an end-to-end manner, the proposed DCNN-based technique for is considered efficient for classification and mapping of costal wetlands, where constant monitoring of the land cover and its evolution are demanded.

Chapter III described the second contribution of this work, where it investigated the application of DCNN-based SISR technique, developed in computer vision, to predict higher spatial resolution UAS images from lower resolution images acquired over a built coastal environment. The main objective of this work was to examine the possibility of optimizing the UAS-SfM photogrammetry procedure for topographic mapping and generating accurate geospatial products, including dense 3D point clouds and detailed DSM models.

To reach the above objective, a pretrained DCNN architecture developed for SISR was implemented and fine-tuned to enhance the spatial resolution and information content of virtually down-sampled UAS LR images, which approximately simulate UAS flight at a higher altitude. The trained DCNN-based SISR model predicts corresponding HR image for each individual input LR image. The investigation showed that super-resolved UAS images, i.e., predicted HR images which closely approximate original HR images, can be successfully predicted from input LR images. The implemented DCNN-based SISR can effectively enhance the spatial resolution of the predicted SR image by factor 4, which is equivalent to improvement in the GSD of collected UAS images due to decreasing flying height by factor 0.5. Moreover, examining the SfM-photogrammetry products generated from LR, original HR and predicted SR image sets, including the retrieved camera's interior and exterior orientation parameters and the qualitative and quantitative properties of derived geospatial products related to the reconstructed 3D environment (e.g., point cloud and DSM), confirm the effectiveness of the proposed approach for optimizing UAS-SfM photogrammetry.

Finally, Chapter IV of this manuscript introduced a novel technique to extract useful information about the illuminated targets, encoded in backscattered raw TLS waveforms. This information was later used for point cloud labeling in both built and natural environments for land cover and topographic mapping.

The main objective was to develop a novel technique to classify 3D points, representing major targets found in the study site, through the direct classification of the backscattered laser energy represented by the received waveform signal in a FW TLS system. In this study, raw waveform information returned from each illuminated target was used to populate the feature vector of corresponding TLS points. Furthermore, a DCNN architecture was proposed to explore

waveform feature space and directly classify each individual point in the dense point cloud based on a hierarchy of discriminative features learned in an end-to-end manner by the DCNN model. This study evaluated the discriminative capability of raw waveform samples (attributes) versus the calibrated online waveform attributes derived from online waveform processing unit in the FW TLS system for target classification.

Point cloud classification based on raw waveform attributes in the feature vectors of measured points versus calibrated online waveform attributes showed more than 10% improvement in overall accuracy of the classification for both natural (coastal wetland) and built (campus) study sites. This observation confirms that the raw waveform data contains more information about the spatial and radiometric properties of the target. Although online waveform attributes for each measurement are derived from a calibrated TLS's built-in look-up-table provided by the manufacturer through an intense calibration procedure, those attributes cannot fully capture the geospatial and radiometric properties of the measured target. Furthermore, the study showed that the waveform samples in the feature vector of measured targets, which represent the scattering properties of the target, are temporally more stable than the calibrated online waveform attributes for the same targets. Results showed that exploiting samples of raw waveform rather than calibrated waveform attributes leads to 15% improvement in classification accuracy of TLS point clouds collected at different points in time from the same scene using a pre-trained DCNN model.

In addition, the classified point cloud in the coastal wetland environment using the raw waveform samples and calibrated waveform attributes showed that the waveform information can help to better discriminate ground points from above ground targets, such as vegetation, in a point

cloud filtering procedure according to predicted class labels. This led to generating a more accurate DTM in the coastal wetland area based on the waveform data.

Collectively, this study showed the potential of high information content in hyperspatial resolution 2D and 3D RS data collected by advanced geodetic imaging technologies (UAS-SfM, FW TLS) along with the applicability and potential of DL techniques, specifically DCNN architectures, to extract useful information from these data streams for land cover monitoring and mapping applications. The findings of this research enhance exploitation of RS data as well as quality and reliability of generated geospatial products in support of coastal zone monitoring and surveying. Additionally, the developed computational techniques in this study are generalizable to a wide range of terrain characterization problems using 2D and 3D RS data streams, such as change detection and real-time post-disaster mapping.

5.2. Future Directions

Regarding this work, further enhancements in RS-based land cover monitoring and topographic mapping can be considered in future work. More accurate land cover prediction in natural and built environments by exploiting different RS sensors and combining different remotely sensed data, such as UAS multispectral imagery along with airborne FW lidar data, should be targeted in future work. In addition, the impacts of virtual transition from LR to HR image space on land cover prediction can be examined. Furthermore, the employed DCNN-based SISR needs to be further investigated in a real-world UAS-SfM photogrammetry scenario in which HR images are predicted from truly collected LR images through the UAS flight at a relatively high altitude. Furthermore, analyzing single-echo raw waveform data to classify the illuminated target needs to be extended to multi-echo waveforms to take full advantage of the multi-target detection capability of lidar instruments. This should exponentially improve the quantity and

quality of geospatial information (i.e., land cover) derived from returned raw waveforms, which can lead to a more accurate and thorough understanding of the 3D structure of the surveyed area.