Doctor of Philosophy

Ground Surface Classification and Area Estimation Using UAV-Collected Images for Smart Earthmoving

August 2022

Department of Civil and Environmental Engineering

The Graduate School of Seoul National University

Daeyoun Won

Ground Surface Classification and Area Estimation Using UAV-Collected Images for Smart Earthmoving

A dissertation submitted to the Graduate School of
Seoul National University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

by
Daeyoun Won

August 2022

Approval Signatures of Dissertation Committee

Moonseo Park

Seokho Chi

Changbum Ryan Ahn

Bon-Gang Hwang

Man-Woo Park

Ground Surface Classification and Area Estimation Using UAV-Collected Images for Smart Earthmoving

지도교수 지 석 호

이 논문을 공학박사 학위논문으로 제출함

2022년 6월

서울대학교 대학원

건설환경공학부

원 대 연

원대연의 공학박사 학위논문을 인준함

2022년 8월

위 원 장 박문서 (인)

부 위 원 장 지석호 (인)

위 원 안창범 (인)

위 원 황본강 (인)

위 원 박만우 (인)

DEDICATION

To my beloved people

ACKNOWLEDGEMENTS

It is a great honor for this opportunity to give credit to those who have supported and encouraged me during my doctoral journey.

First of all, I express my sincere gratitude to my supervisor, Prof. Seokho Chi, for his invaluable support and guidance. Since beginning my graduate course, he firmly believed in my potential and made me into a mature individual. I will never forget this grace for the rest of my life. I am also thankful to my committee members, Prof. Moonseo Park, Prof. Changbum Ryan Ahn, Prof. Bon-Gang Hwang, and Prof. Man-Woo Park, for their insightful comments and guidance. This dissertation could not have been completed without them.

I would like to extend my sincere gratitude to my C!LAB colleagues. The countless memories I had with you guys mean a lot to me. Especially, I want to thank Smart Construction Team members, Dr. Seonghyeon Moon, Gitaek Lee, and Jaehyun Hwang. The numerous research ideas I shared with you became an excellent foundation for writing this dissertation. Thank you guys for decorating a big part of my doctoral journey.

Finally, and most importantly, I express my deepest gratitude to my beloved family for their unconditional love and the sacrifices they have made for me

throughout my life. I would like to thank my grandmother, Seoksoon Lee, my father, Jongho Won, my mother, Jeomae Song, and my fiancé, Suji Hwang, for their endless love and support.

Once again, thanks to my beloved people!

Gwanak, Seoul

July 2022

Daeyoun Won

ABSTRACT

Ground Surface Classification and Area Estimation Using UAV-Collected Images for Smart Earthmoving

Daeyoun Won

Department of Civil and Environmental Engineering

The Graduate School of Seoul National University

The global construction industry suffers from a shortage of workers and low productivity. To address the issues, the industry is introducing smart construction technology. Especially developed countries have tried introducing robotic and autonomous systems to automate construction. Accordingly, industry and academia are developing and commercializing automation technology focusing on earthmoving, where construction

equipment is used extensively. For the automated equipment to work on the construction site, the equipment needs to understand site information accurately. To this end, many researchers have focused on modeling the site using a laser scanner and UAV or automatically detecting and monitoring objects existing on the site.

Above all, for the complete automation of earthmoving equipment, it is necessary to analyze and provide information on the ground surface to the automated equipment, such as on which ground surface of the site is workable and on which ground surface is accessible. In the case of road construction, it is crucial to manage the ground surface on-site according to the designed clearing limit area. The national construction standards in the United States and South Korea specify that ground surface types, such as soil, rocks, trees, and puddles, existing on the site need to be treated. Moreover, when operating automated equipment, the area it can access varies depending on the ground surface types, such as trees, rocks, and puddles. As such, the equipment needs to understand the type of ground surface on the site for its effective operation.

However, in practice, the ground surface is monitored ineffectively. Site managers patrol the site and manually check for ground surface conditions. Unfortunately, since the size of road construction sites is typically enormous,

the manual approach requires significant time and workforce resources. Moreover, it is challenging to continuously monitor changes in the ground surface due to weather or work progress. Many practitioners and researchers alike have thus identified several disadvantages of this human-dependent approach: it is time-consuming, cost-ineffective, and labor-intensive. Therefore, there is a need for a method to automatically analyze ground surface that has not been addressed in the previous studies.

Hence, this research aims to develop a methodology to automatically classify and estimate the ground surface on road construction sites using UAV and computer vision techniques. First, the author proposes a super-resolution-based data augmentation technique to build ground surface datasets. Based on the datasets, this research proposes a deep learning-based multi-label classification method to classify the ground surface types. In addition, to provide area information to the automated equipment, an unsupervised segmentation-based area estimation method is proposed to segment the classified patch and estimate the area by the classified ground surface types.

As a result of the ground surface datasets development, the classification performance improved by 0.16 before utilizing the proposed approach; the average f1 score of the classification model is 0.88. Then, the ground surface

type in a patch resulted from the classification model was segmented by the unsupervised segmentation model and quantified in the form of area (average relative error of 0.15). Lastly, the ground surface area was superimposed onto point clouds to visualize in 3-D. The final result dramatically reduced the time and workforce required to acquire the ground surface information (i.e., ground surface type, location, and area) on road construction sites.

To validate the proposed methodology, this research conducted experiments using UAV images collected from road construction sites in a different environment. As a result, the proposed methodology can automatically generate key site information, such as ground surface type (average f1 score 0.81) and area (average relative error 0.21) for automated earthmoving equipment operation in a more efficient manner than the existing practices. The time required to process input UAV images and calculate the area for each type of surface was reduced to 30% or less compared to the existing manual method. Furthermore, it was confirmed that the ground surface characteristics considered in this research play a crucial role in providing site information with automated construction equipment.

In conclusion, earthmoving equipment and site managers can automatically understand the ground surface condition. The proposed methodology can

reduce the cost and time for site management by enabling the managers and the equipment to quickly and effectively detect the ground surface. Earthmoving equipment can avoid unfavorable ground surfaces to prevent accidents and identify the workable area according to the surface type. Site managers can establish a work plan considering the ground surface condition. Finally, this research facilitated an in-depth understanding of ground surface types on the site, which could further improve opportunities for smart earthmoving.

Keywords: Construction Management; Smart Earthmoving; Ground Surface Information; Unmanned Aerial Vehicle (UAV); Computer Vision

Student Number: 2018-31863

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

1.1. Research Background

The global construction industry is suffering from a shortage of construction workers (Brucker et al. 2021). According to a 2020 report by the Associated General Contractors (AGC) in the United States (U.S.), 81% of construction companies had difficulty filling full-time and temporary positions, and 72% of them expect this trend to continue (AGC 2020). Similarly, there has been a perceived shortage of construction workers in both infrastructure and building sectors in Korea (Kim 2014; Jeong et al. 2018). To address the labor shortage issue, developed countries have tried to introduce robotic and autonomous systems to automate the construction process. For example, the U.S. Federal Highway Administration (FWHA) has invested in key automation technologies such as remote sensing and automated machine guidance and machine control since they launched the intelligent construction systems and technologies program (Torres et al. 2018; FWHA 2021). The Korean government established the smart construction technology roadmap in 2018 and funded a large-scale national research and development (R&D) project, which aims to develop advanced smart construction technologies such as machine control, automated construction vehicles, and real-time data acquisition (Cho et al. 2020). The Korean

Government established the smart construction technology roadmap in 2018 and funded a large-scale national research and development (R&D) project, which aims to develop advanced smart construction technologies such as machine control, automated construction vehicles, and real-time data acquisition (Cho et al. 2020).

Accordingly, industry and academia are developing and commercializing automation technology focusing on earthworks where construction equipment is used extensively. In particular, they are developing technologies related to the control and guidance of earthmoving equipment and conducting site validations (Azar and Kamat 2017; Ha et al. 2019; Borngrund et al. 2022). In the field of construction equipment manufacturing, a remote equipment control system has been commercialized (Leonida 2020), and equipment guidance technology has been developed to position earthmoving equipment and guide work routes using GPS and 5G network technology (Nguyen and Ha 2022). In addition, equipment guidance technology using building information models was developed to provide on-site design information to equipment operators (Trimble 2022).

For automated equipment to perform tasks safely and accurately on the site, it is crucial to understand site information accurately. To this end, researchers conducted a study to model the site using a laser scanner and UAV to recognize and monitor objects automatically (e.g., equipment, workers, materials) on-site. Above all, to achieve actual earthmoving equipment

automation, the equipment needs to understand the type of ground surface at the site. For example, in the case of road construction, where the budget for earthworks is high, it is important to manage the ground surface of the site according to the designed clearing limit area. For example, when performing clearing and grubbing works, all of obstacles, such as rocks and trees must be removed, and puddles and holes created during the work must be backfilled (USDOT 2014; MOLIT 2016a). In addition, when operating automated equipment, the area accessible to the equipment varies depending on the ground surface types, such as vegetation, rocks, and puddles (Azar and Kamat 2017). As such, for the automation of earthmoving work, it is necessary to analyze and provide information about the type of ground surface to the automated equipment, such as which ground surface of the site is workable and which surfaces are accessible.

Construction site monitoring is a process of measuring, analyzing, and improving project performance, such as operational productivity and safety, and is an important task for successful project management (Kim and Chi 2021). Especially for a successful road construction project, earthwork monitoring is significant. The goal of the earthwork is to build a subsoil in a uniform and clean state, and therefore securing the quality of the subsoil plays a critical role in the entire road's design performance. Thus, it is crucial to monitor the state of the ground surface if they meet the clearing standards designated in design plans. National standards related to road construction in

the United States and South Korea specify the regulations to secure the quality of the subsoil layer (USDOT 2014; MOLIT 2016a). For example, when clearing and grubbing works are performed during earthworks, all trees, bush, downed timber, and other vegetation on the subsoil within the clearing limits need to be removed, and the holes, pits, and other settled grounds need to be backfilled, followed by site manager's visual inspection for compliance checking with the contract and industry standards. Therefore, it is necessary to check and monitor what type of ground surface exist on the subsoil layer.

Accordingly, there have been many attempts to introduce the latest equipment, such as unmanned aerial vehicles (UAVs), to construction sites to overcome the limitations of the current manual practice. Advances in UAV technology allow for more efficient image collection at construction sites compared to other devices, such as smartphone cameras, closed-circuit television, and camcorders (Ham et al. 2016; Wang et al. 2016; Bang et al. 2017a; Bang et al. 2017b; Kim et al. 2019). In particular, several companies now provide various UAV-based solutions specifically targeted at monitoring construction sites. For instance, some solutions involve conducting site modelling and surveys through on-site point-cloud generation from UAV images, comparing them with designs (Leica 2021; Meisa 2021), and sometimes creating virtual reality environments based on the collected UAV data (Angelswing 2021; DroneDeploy 2021). Meanwhile, many researchers have detected, monitored and predicted workers, equipment, and materials in

construction sites (Kim et al. 2019; Guo et al. 2020; Bang and Kim 2020; Jiang et al. 2020; Bang et al. 2020; Xuehui et al. 2021; Kim et al. 2020; Kim et al. 2021; Bang et al. 2021). Thus, many practitioners and researchers alike have proven UAVs' practicality and potential in construction site management, meaning that UAV image data have already been collected and utilized on such sites.

1.2. Problem Statement

Despite the importance of ground surface information, in practice, the information acquisition process is conducted ineffectively. Site managers patrol the site and manually check the ground surface conditions (USDOT 2014; MOLIT 2016a). Unfortunately, since the size of road construction sites is typically enormous, the manual approach requires significant time and workforce. It is challenging to continuously monitor changes in the ground surface due to weather or work progress (e.g., puddles after rain, rocks generated during earthworks in bedrock areas, and ditches and holes on the ground made by grubbing work or heavy equipment traffic.) Many practitioners and researchers alike have thus identified several disadvantages of this human-dependent approach: it is time-consuming, cost-ineffective, and labor-intensive (Kim et al. 2016; Kim et al. 2017; Kim et al. 2018; Kim and Chi 2020). Moreover, in the United States, departments of transportation are charged with managing an increasing workload with a diminishing number of inspection staffs, leading to new technology adoption to automate and improve their inspection works (Newcomer et al. 2019).

1.3. Research Objectives

To address the aforementioned challenges, this dissertation aims to propose a ground surface classification and area estimation methodology that automatically identifies and quantifies ground surface on road construction sites based on the application of a UAV and computer vision techniques, as shown in Figure 1.1. First, the author proposes a super-resolution-based data augmentation technique to build ground surface datasets. To automatically classify the ground surface types, this research proposes a deep learning-based multi-label classification method. In addition, to provide higher-level information to the automated equipment, an unsupervised segmentation-based area estimation is proposed to segment the classified patch and estimate the ground surface area.

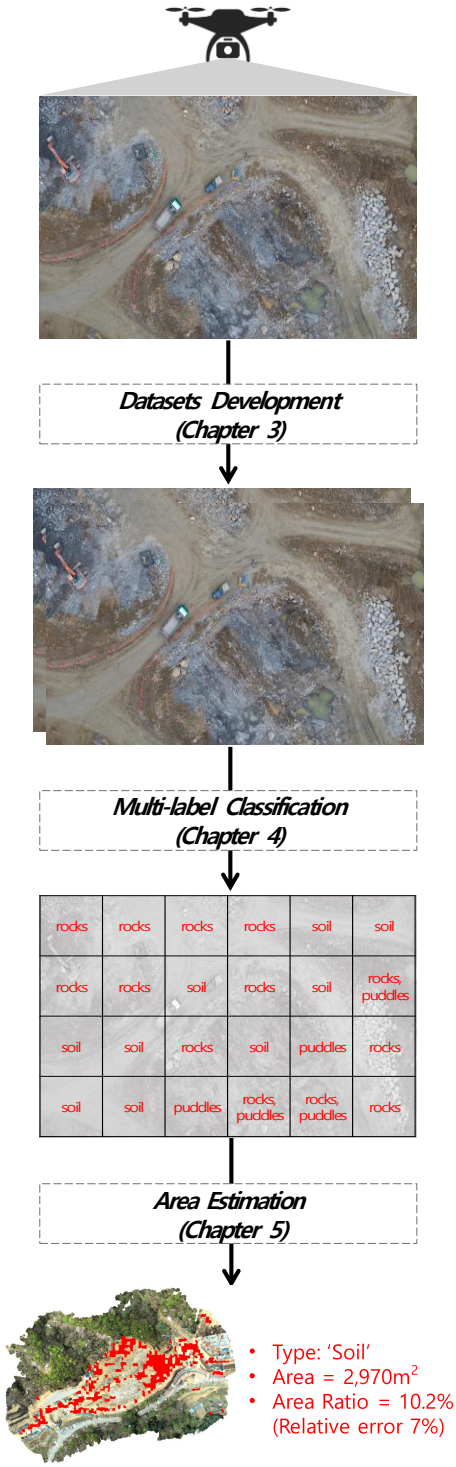


Figure 1.1 Concept of proposed methodology

The output of this research is the first attempt to automate the ground surface classification and estimation process for a road construction site. First, the output of the proposed datasets development method in Chapter 3 can enhance the quality and quantity of the dataset by upgrading the quality and quantity of UAV images, thus improving the performance of the deep learning models. Chapter 4 represents a novel multi-label image classification framework to classify automatically the ground surface from UAV images. Lastly, the area estimation method of the ground surface presented in Chapter 5 can help the automated equipment make action plans after adverse weather conditions or work progress that can affect the ground surface by automatically quantifying the identified ground surface.

1.4. Research Scope

This research defines ground surface information as ground surface types, location, and area, which is generated by classification and area estimation modules. The classification module predicts a patch-wise ground surface class in a UAV image. The area estimation module includes unsupervised segmentation which changes the patch-wise class into pixel-wise class and estimation of the area by the ground surface types. The final results are visualized in 3-D point clouds by mapping the pixel-wise class into a point-wise class.

This research focuses on ground surface types which were not addressed yet in the existing studies among other site information. Herein, this research defined the subsoil during earthwork of road construction as ground surface classes according to national construction standards as described in Table 1.1. “Soil” was defined as including only rocks under 600mm. If other objects were included, they were classified as “non-soil.” The “non-soil” was further classified into “rocks,” “trees,” “puddles,” “nets,” and “etc.” “Rocks” included rocks or stones with a diameter larger than 600mm. “Trees” included all vegetation, including stumps and bushes. “Puddles” included holes and trenches. “Nets” indicated a protection net. Any other objects (e.g., tools, devices, or anything else) corresponded to “etc.”

Table 1.1 Ground surface type definition

Type		Definition	Importance	Standards
Soil		Earthen floor	Prior to pavement works, rocks, organic matter, and other impurities need to be removed	USDOT 2014; MOLIT 2016b
Non-soil	Rocks	Includes rocks over 600mm in diameter	Rocks of a diameter larger than 600mm should be eliminated during earthworks	MOLIT 2016b
	Trees	Includes trees	All trees, such as vegetation, stumps, and bushes, need to be removed before pavement works	USDOT 2014; MOLIT 2016a
	Puddles	Includes puddles	Puddles, including holes and trenches, should be backfilled and compacted	USDOT 2014; MOLIT 2016a
	Nets	Includes protection net	A slope should be protected (e.g., with a protection net)	MOLIT 2016a; MOLIT 2016b
	Etc.	Includes any objects other than the above-defined classes	-	-

1.5. Dissertation Outline

This dissertation consists of seven chapters and the brief description of each chapter is described below.

Chapter 1 Introduction: This chapter introduces the research background, problem statement, research objectives, research scope, and the outline of this dissertation.

Chapter 2 Theoretical Background and Related Works: This chapter provides a comprehensive understanding of required site information in earthwork sites for the automated operation of construction equipment and reviews previous works for modeling the site information. Based on the review, the chapter explains the research gap to be addressed in this research.

Chapter 3 Ground Surface Datasets: This chapter covers the development process of ground surface datasets, as shown in Figure 1.2. The state-of-the-art (SOTA) algorithm for image super-resolution and its impacts on a deep learning-based classification model were described.

Chapter 4 Ground Surface Classification: In this chapter, the author investigates the characteristics of UAV image and ground surface types, and

suggests a novel classification method to automatically identify the ground surface types from UAV images, as shown in Figure 1.2. Two different multi-label classification methods and two different SOTA deep-learning-based classification networks are applied and comparatively evaluated for deriving the best-performance model.

Chapter 5 Ground Surface Area Estimation: This chapter aims to estimate the area by the ground surface types by quantifying 2-D pixel-wise ground surface class resulted from the previous chapters, as shown in Figure 1.2. The suggested unsupervised segmentation method segments the labeled 2-D patch to estimate class-wise area by the ground surface types. Then, the estimated 2-D area is superimposed onto point clouds through the 3-D mapping, making it into labeled point clouds to visualize the results in 3-D for the automated equipment operation.

Chapter 6 Experimental Design and Analysis (Validation): This chapter explains the experimental design, results, and discussion to confirm the technical feasibility and generalize the applicability of this research, as shown in Figure 1.2. By applying all the proposed methodology described from Chapter 3 to Chapter 5, the area estimation results are defined as the final output of the proposed methodology. Finally, the performances of proposed methodology are comparatively evaluated with a manual method.

Chapter 7 Conclusion: This chapter summarizes the research outcomes and discusses key findings, practical applications, and future research directions in the field of UAV-based jobsite management for the automation construction equipment.

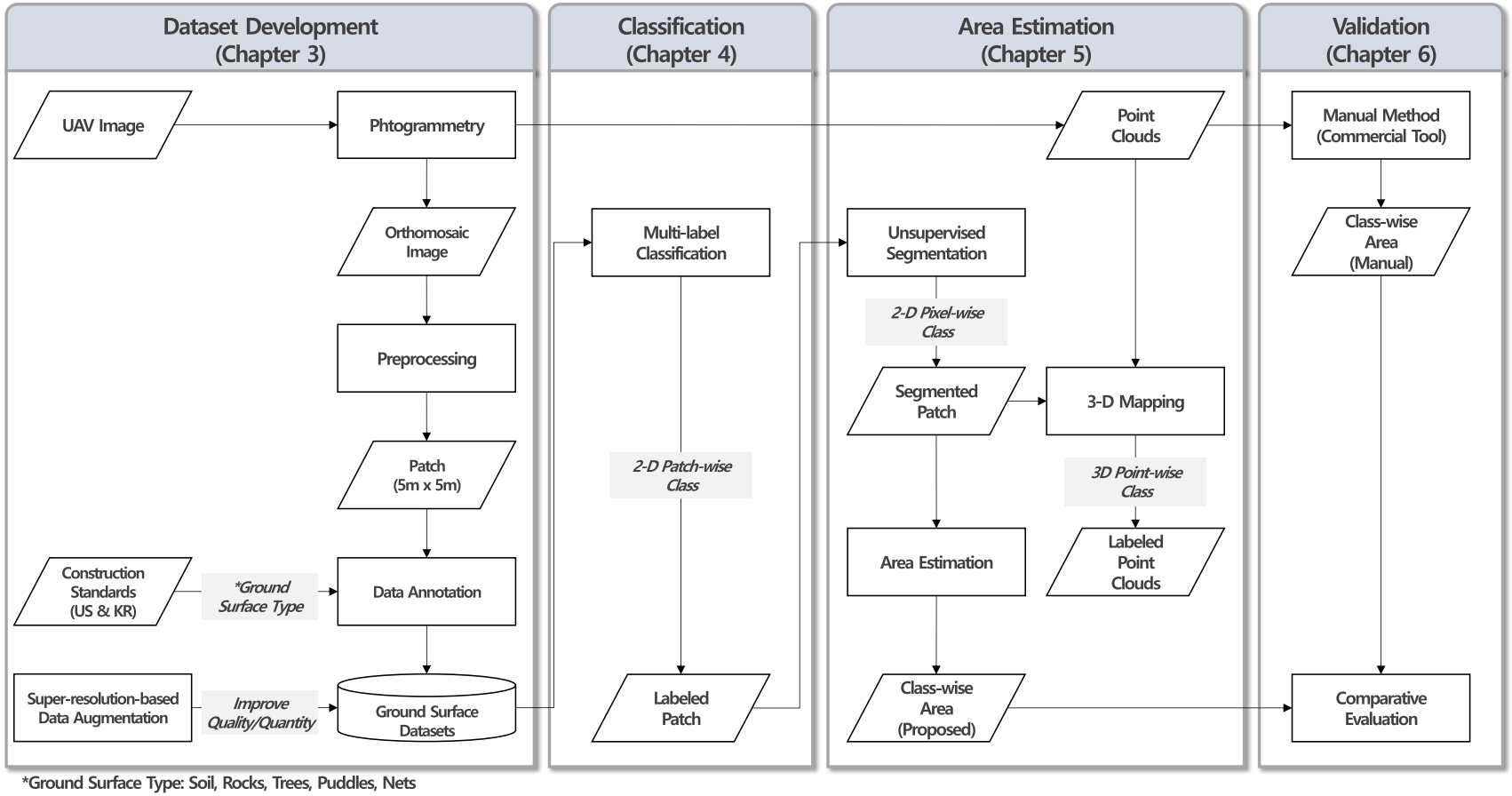


Figure 1.2 Research framework

Chapter 2. Theoretical Background and Related Works

This chapter provides a comprehensive understanding of the required site information for the automation of construction equipment in earthmoving sites and reviews previous works for modeling the site information. Many researchers focused on (1) 3-D modeling of the site using laser scanners and UAVs and (2) automatically detecting and monitoring ground surface on the site. However, to achieve the automation of earthmoving equipment, above all else, information on the ground surface type of the site is required, and research that automatically analyzes and provides this information is currently lacking. Detailed explanations of the research on (1) and (2) and the research gap are described in the following sections.

2.1. Theoretical Background

2.1.1. Earthmoving in Road Construction

Earthworks are significant for a successful road construction project. As shown in Figure 2.1, road construction can be broadly divided into earthworks and pavements according to the layers piled up. The ultimate purpose of the earthworks is to build a subsoil in a uniform state, which plays an important role in the design performance of the entire road. Thus, it is necessary to secure the quality of the subsoil layer. National standards such as USDOT's Federal Road Project Standard or Korea Highway Corporation's Standard Specifications specify that surface objects such as vegetation, rocks, and puddles be removed and managed according to the clearing limit (USDOT 2014; MOLIT 2016). Ground surface needs to be managed when clearing, and grubbing works are performed. For example, all trees, brush, downed timber, and other vegetation within the clearing limits need to be removed, and the holes, pits, and other depressions need to be backfilled, followed by evaluation through visual inspection by the site manager for compliance with the contract and prevailing industry standards (USDOT 2014; MOLIT 2016). It is necessary to monitor where and how many ground surface (e.g., trees, rocks, puddles) are on the road construction sites.

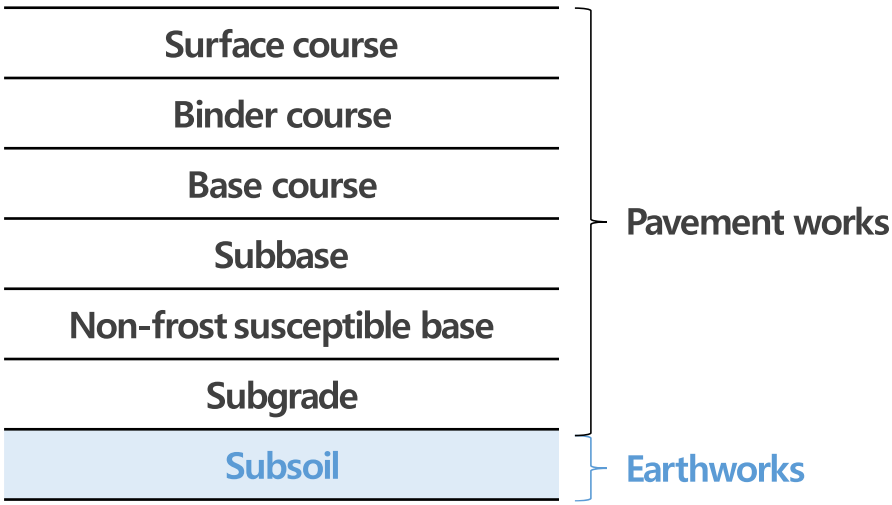


Figure 2.1 Road construction layers

In the existing practice, site managers currently patrol the site and check the ground surface conditions manually. A site manager travels in a vehicle and visually monitors where and how many objects exist on the ground surface of the site. This practice consumes a lot of time and effort for the site manager, considering the size of the road construction site that is generally distributed over a wide area. In addition, the state of objects on the ground frequently changes as construction progresses, so the manager needs to monitor the object on site continuously. For example, additional rocks during earthworks in bedrock areas or additional puddles after rainfall/snowfall require continuous monitoring over time. Accordingly, many studies addressed that human-dependent approach is time-consuming, cost-

ineffective, and labor-intensive (Kim et al. 2016; Kim et al. 2017; Kim et al. 2018; Kim and Chi 2020). Furthermore, as the number of on-site management personnel has been reduced in the US DOT recently, the government sectors are paying a lot of attention and investment in introducing new technologies to automate the above inspection tasks (Newcomer et al. 2019). Amid the limitations and interests of these practices, efforts to automate inspection work are reflected in academia, and many studies are being conducted. In particular, much research using UAVs with excellent mobility is being conducted for road construction site.

2.1.2. On-site Modeling for Smart Earthmoving

The operation of automated construction equipment requires a digitalized on-site model. A complete 3D model of the site environment can be created using multiple laser scans from different locations and angles, which are then automatically registered (Chae et al. 2011). 3D designs have been used as a target task compared with the actual environment sensed by laser scanning and stereovision to reach the designed profile (Yamamoto et al. 2009; Sung and Kim 2016). On the basis of a scanned section of the environment, intelligent navigation systems guide the equipment to a target location (Seo et al. 2011; Kim et al. 2012) and display updated job plans (Halbach and Halme 2013). However, laser scanning requires a lot of time and manpower to model the entire construction site, especially a road construction site covering a large area. In order to overcome the shortcomings of laser scanning, many researchers have used UAV applications for site modeling.

For the site modeling using UAVs, many researchers have focused on earthwork volume estimation and displacement measurement of the ground. For example, some studies have used UAVs to monitor the slope stability (Xiao et al. 2018) and have measured earthwork quantity using the photogrammetry method (Siebert and Teizer 2014). Their works show the potential of UAV and photogrammetry techniques for modeling construction sites, and these findings have encouraged other researchers to further conduct

related research and expand UAV applications in site modeling. Bang et al. (2017) have proposed an image-stitching method to generate high-resolution panoramic images suitable for representing a large construction site using UAV images, while Jiang, W., et al. (2020) have proposed a UAV-based 3D-reconstruction method for on-site layout planning. Furthermore, other research has integrated point clouds from UAV-based photogrammetry and other resources such as laser scanning for heavy equipment planning (Moon et al. 2019) and compared them with building information models (Kim et al. 2021). The results of these earlier works show the high potential of UAV applications for on-site modeling.

2.1.3. On-site Object Modeling for Smart Earthmoving

The recognition of construction resources has an important role in achieving the fully automated operation of equipment (Tajeen and Zhu 2014; Naghshbandi et al. 2021). By utilizing a variety of sensors and CV technology, the equipment can understand its surroundings and make decisions for performing tasks based on the rich description of the detected objects in the environment. Many studies have been conducted to automatically recognize workers or equipment by applying various CV-based methods to closed-circuit television images installed at construction sites (Fang et al. 2018; Kim et al. 2018; Kim and Chi 2020; Kim and Chi 2021). For more enriched geometry information (i.e., 3D), laser-scanning technology has been used to automatically produce geometry documentation of objects. Mobile Light Detection and Ranging (LiDAR) technology has been utilized to scan highways and classify obstructions (i.e., traffic signs, road surface strips, poles and bridges) on the highways based on semantic segmentation: voxel region growing (Gargoum and Karsten 2021) and supervised segmentation (Wu et al. 2019; Ma et al. 2021; Ma et al. 2022). Another study utilized point cloud datasets containing common construction vehicles to train a k-NN classifier for object recognition and classification (Chen et al. 2016). However, closed-circuit television and laser scanning incur excessive costs and time to cover a sizable construction site. Advances in UAV technology allow for more efficient image collection at construction sites compared to

other devices (Ham et al. 2016; Wang et al. 2016; Bang et al. 2017a; Bang et al. 2017b; Kim et al. 2019).

The mainstream UAV application for object detection is to monitor objects on-site, such as workers, heavy equipment, and materials. Guo et al. (2020) suggest a method to detect construction vehicles using UAV images based on an orientation-aware feature algorithm. The detection performance has been improved by using generative adversarial networks, cut-and-paste, and image-transformation techniques (Bang et al. 2020). A follow-up study introduced an image dataset including UAV images and benchmarks to detect workers and equipment at the construction site, followed by verifying whether the detectors trained on the proposed dataset could effectively detect objects using benchmark detectors (Xuehui et al. 2021). The findings of these existing studies have led researchers to apply UAVs for safety monitoring activities involving objects detected on construction sites. For example, one study monitored the proximity between mobile construction resources (i.e., worker, loader, and excavator) on UAV images using a convolutional neural network for object localization and image rectification for distance measurement (Kim et al. 2019). Another study proposed a method to provide the textual description of the site, such as construction materials, from UAV images using the image-captioning method (Bang and Kim 2020). A later study proposed a system that detects workers, manlifts, and cranes from UAV

images and monitors the proximity between the workers and the equipment working areas based on a game engine (Kim et al. 2021).

Inspired by the above studies, other researchers have even tried to predict the future proximity of objects. For example, Kim et al. (2020) suggested a framework to predict future locations of moving workers and heavy equipment, calculating their proximity by leveraging CV and deep neural networks based on UAV images. Following this study, Bang et al. (2021) suggested a proximity-monitoring method to detect moving workers, excavators, and dump trucks and their future locations, postures, directions, and speeds based on UAV-acquired video frames. As described above, existing studies have shown the exceptional detection performance of UAV vision-based construction object monitoring. Their findings have demonstrated that UAV-based visual sensing approaches are potent and effective tools for recognizing on-site objects.

2.1.4. Current Gap in The Research

For automated equipment to safely and accurately perform tasks in the field, it is vital to understand site information. Above all, to achieve effective automation, it is necessary to automatically provide information on which ground surface of the site the equipment needs to work and which ground surface it can access. Currently, this information is obtained by directly checking the site manually. Therefore, there is a need to develop an automated method to classify surface types—an issue not currently addressed in the extant literature.

To automatically classify the surface types, a detection model that can learn the visual characteristics of the ground surface types encompassing the various sizes, shapes, and colors is required. Moreover, a model is required to distinguish different surface types, such as soil, rocks, trees, and puddles. To this end, this research proposes a deep learning-based multi-label classification method. The author applied deep learning architectures—ResNet and VIT—which show excellent performances in classifying complex and diverse classes (He et al. 2016; Dosovitskiy et al. 2020). The specifics of this method are explained in section 3.

2.2. Related Works

2.2.1. Image Super-resolution

Research using aerial photography such as UAV images can obtain a wide range of areas at once. Still, it has problems with low image quality, such as low resolution and blur, which degrades the performance of the model trained from this data. To solve this problem, researchers in computer science, computer vision, and remote sensing are actively conducting research to improve image quality by applying super-resolution techniques. The super-resolution technique is divided into two types according to the number of low-resolution images. A multi-image super-resolution technique uses time series information for the same area and a single-image super-resolution technique that uses only a single image. The multi-image super-resolution technique reconstructs a single high-resolution image from multiple low-resolution images taken in the same area under the same shooting conditions (Fernandez et al. 2017). However, in general, aerial images, it is not easy to obtain images under the same shooting conditions for a given area due to uncontrollable factors such as weather conditions such as clouds or snow, moving objects, and natural disasters.

On the other hand, the single-image super-resolution technique is receiving a lot of attention from academia because of its convenience and efficiency because it generates a high-resolution image using only one input

image. Single-image super-resolution methods are largely divided into interpolation, reconstruction, and train-based methods. The traditional interpolation method utilizes only information on pixels adjacent to the target pixel to be restored, which has the disadvantage of creating an unnatural structure in the reconstructed image (Keys 1981). The reconstruction method reconstructs pixels using prior knowledge information (e.g., gradient profile sharpness) possessed by the image (Yan et al. 2015). However, as the resolution of the image increases, the calculation takes a lot of time and there is a disadvantage in that the restoration performance is deteriorated.

With the recent development of deep learning technology, academia's attention is focused on the train-based single image super-resolution method. Dong et al. proposed for the first time a deep neural network consisting of three layers, which bicubic downsamples the original image in the preprocessing step to create a low-resolution image and compares it with the original image, and characterizes the difference to build the model (Dong et al. 2014). Kim et al. proposed a model showing high image restoration performance by constructing Deep Neural Networks consisting of 20 layers to learn residual components (Kim et al. 2016). However these studies have a problem in which the outline part is blurred in the reconstructed super-resolution image due to the loss function problem, and there is a limitation that the problem becomes more pronounced as the resolution scale increases (Choi et al. 2020). To overcome this problem, several researchers proposed a

super-resolution model based on a Generative Adversarial Network (GAN) to solve the problem of outline blur and improve the image resolution to excellent performance (Ledig et al. 2017; Vassilo 2020). Click or tap here to enter text. However, the GAN-based super-resolution technique often had a problem of color distortion, where the color between the original low-resolution image and the output image was different.

In order to solve the limitations of the CNN- and GAN-based super-resolution methods described above, a super-resolution technique using Vision Transformer has recently emerged (Chen et al. 2021; Cao et al 2021). However, since these studies divide the image into patches and perform attention on each of them independently, the pixels near the boundary do not adequately utilize the information about the surrounding pixels, resulting in border artifacts of the patch. The Swin Transformer architecture has solved this problem, and it is currently exhibiting state-of-the-art performance on several benchmark datasets (Liang et al. 2021). A high-resolution image is generated by performing image restoration through shallow feature extraction and deep feature extraction with a low-resolution image as input. In this case, the deep feature extraction utilizes a network composed of Swin Transformer Block, which solves the patch border artifacts problem and achieves the best performance in several benchmark datasets.

2.2.2. Image Classification

Since 2012, image classification technology has been developed with high performance in various tasks according to the development of technology using deep learning based on Convolutional Neural Networks (CNN). The most famous task among them is the problem of automatically classifying different kinds of images with ImageNet. Before the advent of CNN-based image classification technology, combining several classifiers with Fisher Vector in technology for extracting feature points called Scale-Invariant Feature Transform (Lowe 2004) had a Top-5 error rate of 26.2% (Simonyan et al. 2013). AlexNet, the first CNN-based network, won the ImageNet challenge 2012 with a big difference Top-5 error rate of 15.4% (Krizhevsky et al. 2012). AlexNet was the first to perform well in image classification with a deep convolutional neural network (CNN) structure. Characteristic. In addition, data augmentation techniques such as image inversion, image position change, and average value subtraction were used. This technique is used in various fields other than the current image classification. Next, VGGNet is a thesis proposed by K. Simonyan et al. at Oxford University in the UK in 2014. This paper presented a method using ReLU and a convolutional neural network with a deeper structure than the existing AlexNet. The basic idea of the thesis is to increase the non-linearity by stacking many 3×3 convolutional layers and adding ReLU. Although VGGNet has a simple structure, it is easy to learn and has excellent

performance, so it has been widely used. The disadvantage of VGGNet is that the number of parameters is too large (especially, the number of parameters in the FC layer located at the last part occupies more than 20% of the total parameters), and memory usage is high.

GoogLeNet is a network announced by Google, known as Inception, as a model that won 1st place in the 2014 ILSVRC (Szegedy et al. 2015). In general, it is known that the performance of a deep learning network improves as the structure deepens and the layer becomes wider. However, when learning, as the number of parameters increases, problems such as overfitting or gradient vanishing occur, making learning difficult. Here, the core idea of GoogLeNet is to create a network that can achieve optimal performance using limited computational resources. Several convolution layers are configured as one module, and the network is constructed by concatenating the result values (i.e., concatenation). Softmax is added to the middle layer to solve the problem of gradient vanishing that may occur as the network deepens during training.

ResNet is a network that achieved a Top-5 error rate of 3.57% in the ImageNet task with an ensemble configuration by stacking up to 152 layers, which are more than eight times deeper than the existing VGG16 network (He et al. 2016). As seen in GoogLeNet, the deep learning network is known to have better performance as the structure is deep and wide, but there is a problem that the deeper the network, the more parameters, and the more

difficult it is to learn. ResNet solved this problem by proposing the concept of skip connection. ResNet has been published up to 152 layers deep, but structures as deep as 1002 have since been published (He et al. 2016). ResNet's CNN network has been used in various ways as a base CNN network in the object detection field, and it has shown high performance in the Faster-RCNN-based detector. After that, DenseNet, a network that gave a transformation to ResNet, appeared. In the case of ResNet, skip connection is taken only to the next layer, whereas DenseNet has a structure in which skip connection is taken to the entire layer. This can further reduce the gradient vanishing problem, promote feature propagation enhancement and reuse, and reduce the number of parameters. Since then, several networks based on CNN networks have been introduced, such as MobileNet (Howard et al. 2017), SENet (Hu et al. 2018), NASNet (Zoph et al. 2018), and AmoebaNet (Real et al. 2019).

However, CNN-based networks have the following disadvantages. Because it uses a fixed convolution filter size (window size), it cannot learn the relation with pixels outside the receptive field. Also, the weight values of the convolution filter do not change dynamically even if there is a slight change in the input because they use a fixed value after learning. These shortcomings can be solved by using Self-Attention and Transformer. First, Non-local Neural Networks announced in 2018 secured long-range dependency in both spatial and temporal axes through non-local blocks

(Wang et al. 2018). That is, the relation is learned while calculating the relation between a specific pixel and all remaining pixels in the form of a weighted sum in the input image (i.e., feature map). It can be seen as self-attention, and in CNN, no relation can be learned from pixels outside a given distance, but Non-local Neural Network is possible. The following is Criss-Cross Attention presented at the 2019 ICCV (Huang et al. 2019). Using the non-local block described above, full-image contextual information can be modeled, but there are limitations in that memory, and computational costs are very large. This is because a dense attention map is required to compute for the entire feature map. To overcome this, the Criss-Cross Attention method was proposed, which computes the attention map sparsely, and through this method, the accuracy may decrease slightly, but the computational complexity can be greatly reduced.

Next, Stand-Alone Self-Attention was presented at the 2019 NeurIPS (Ramachandran et al. 2019). This study proposed a method to replace all convolutional layers with a Local Self-Attention Layer. By applying this Local Self-Attention Layer to ResNet-50, higher accuracy could be achieved with fewer parameters and computational amount. Afterward, a paper presented at ICCV 2019 proposed a differentiable Local Relation Network method (Hu et al. 2019). Existing CNNs have the disadvantage of being unable to adaptively modify weights according to changes in input because the consequences are fixed after learning is finished. A method to do this has

been proposed. In addition, a computation based on Relative Position Encoding was proposed to apply the Self-Attention mechanism while maintaining the property that the position of the CNN also changes when the position of the input changes (i.e., translation equivariance) (Bello et al. 2019). Suppose all convolution operations are replaced with self-attention functions. In that case, computational efficiency can be increased, but the best performance can be achieved when used with convolution operations to achieve better performance. Next year, Vectorized Self-Attention was proposed in the 2020 CVPR (Zhao et al. 2020). In general convolution operation, it is common to process feature aggregation and transformation by activation function sequentially. In this study, feature aggregation and transformation were performed separately using self-attention, and an element-wise perceptron layer was used for transformation. Through Self-Attention Networks (SAN) constructed in this way, better performance was achieved than ResNet in ImageNet dataset with fewer parameters. In addition, the effect of being robust to adversarial perturbation was obtained, and the generalization performance was improved even when unseen transformations were applied to the test image.

Afterward, the first Transformer in Computer Vision communities, Vision Transformer (VIT), was proposed to show comparable performance to CNNs on large-scale computer vision datasets (Dosovitskiy et al. 2020). The input image is split into several patches, put into ResNet, extracted feature

maps, flattened, and put into the Transformer encoder. After that, a classifier is added for training. In this case, the transformer-based methods have a disadvantage: good performance is guaranteed only after pre-training with a myriad of data sets and then fine-tuning on downstream tasks (e.g., ImageNet). However, since the large data set used in the experiment is a data set used only by Google (i.e., JFT-300M), good performance may not be obtained even if the same method is applied in a research group other than Google. Comparing CNN and Transformer, CNN is a model with a lot of inductive bias, such as translation equivariance, so performance is guaranteed even with relatively small data. In contrast, VIT is a model with slight inductive bias, so a large amount of data is required. Performance is improved. This point can be both an advantage and a disadvantage of VIT. It is impressive that Google sublimated it to its advantage through many data. Still, it shows the weakness that it is difficult to apply in a field where it is difficult to secure a large amount of data.

2.2.3. Unsupervised Segmentation

Image segmentation is a method of assigning a label to all pixels in an image so that pixels that share a characteristic are assigned the same label. k-means clustering is a representative classical unsupervised segmentation method, and the target data is assigned to k clusters, where each data belongs to the cluster with the closest mean (MacQueen et al. 1967). The graph-based segmentation method is a simple greedy search method (Felzenszwalb and

Huttenlocher 2004). Clustering is performed through a specific region comparison method. Recently, several methods for data learning-based unsupervised image segmentation have been proposed. Liu et al., an efficient and versatile approach that can be switched to both unsupervised and supervised methods, has a limitation in that the boundaries are fixed using superpixels (Liu et al. 2014). Xia and Kulis estimate segmentation from the input image and perform unsupervised segmentation by reconstructing the input image from it, so that similar pixels are assigned to the same label (Xia and Kulis 2017). Croitoru et al. proposed an unsupervised segmentation method based on the deep neural network technique. This method performs binary segmentation (foreground or background) (Croitoru et al. 2019).

Recently, with the remarkable development of deep learning technology, CNN-based semantic image segmentation has also received a lot of attention (Chen et al. 2015; Long et al. 2015; Zheng et al. 2015; Badrinarayanan et al. 2017). Compared to other supervised learning methods, segmentation methods require much more human input because labeling must be performed at the pixel level. Accordingly, many researchers have studied object detectors (Tighe and Lazebnik 2013; Hariharan et al. 2014; Dai et al. 2016), object bounding boxes (Zhu et al. 2014; Chang et al. 2014) to minimize human input. , a weakly supervised learning approach using image-level class labels (Pathak et al. 2015; Pourian et al. 2015; Shimoda 2016; Shi et al. 2017) or

graffiti for training (Lin et al. 2016; Tang et al. 2018a; Tang et al. 2018b) are widely used.

Most weakly supervised segmentation algorithms generate training targets from weak labels and use the generated training set to train a model. This method mainly repeats two steps: gradient descent to train a CNN-based model on the generated target and generation of a training target by weak labels. For example, superpixels were used to propagate the semantic labels of graffiti to other pixels to fully annotate images, and train a neural network with the annotated images (Lin et al. 2016). Segments are proposed in pixel-level annotations using bounding box annotations or CPMC segments, and there is also a way to train a model with these proposals (Zhu et al. 2014). However, the weak supervised segmentation algorithm has errors in the generation of training objects and errors that can train the model in an undesired direction. Therefore, unsupervised deep learning approaches are receiving a lot of attention, and in this study, the author try to predict the ground surface type at the pixel level by using this method.

2.3. Summary

This chapter represented a comprehensive understanding of the required site information for the automation of construction equipment in road construction sites and reviewed the previous works for modeling the site

information. In conclusion, to achieve the automation of earthmoving equipment, above all else, the ground surface information of the site is required, and the research that automatically analyzes and provides this information is currently lacking; therefore, this research aims to fill the research gap by developing the proposed methodology in the following sections.

Chapter 3. Ground Surface Datasets

This chapter covers the development process of ground surface datasets as shown in Figure 3.1. The proposed method for image super-resolution and its impacts on a deep learning-based classification model were explored. A detailed explanation of the proposed method is described in the following sections.

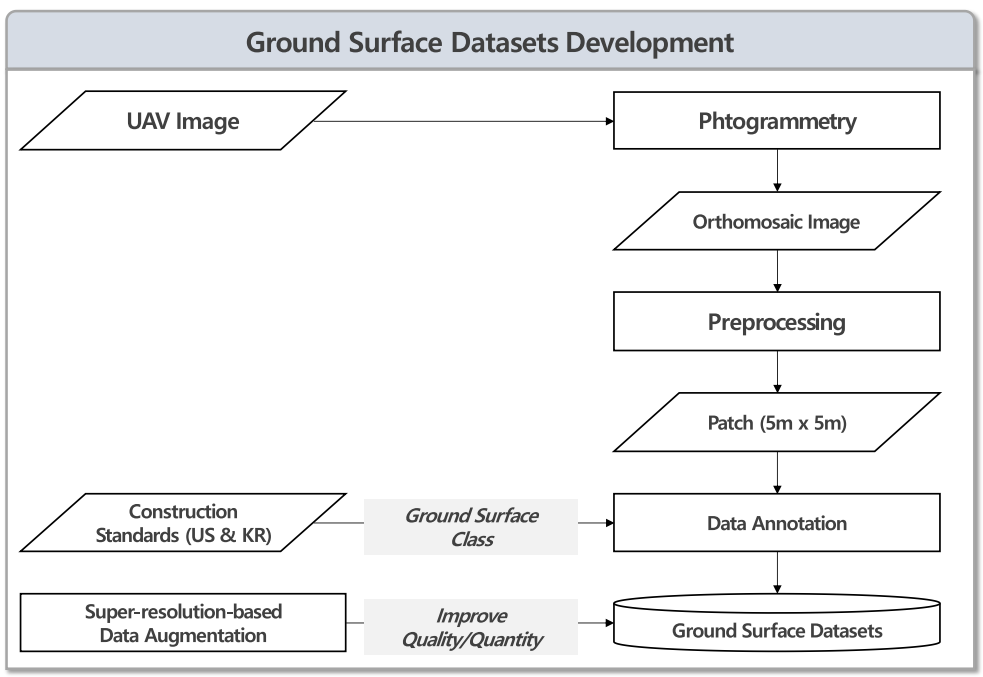


Figure 3.1 Research flow of ground surface datasets development

3.1. Data Preparation

3.1.1. UAV Image Patch

A UAV (the DJI Phantom 4 RTK) equipped with a camera (the FC6310R) was utilized, of which the field of view is 84 degrees. 199 images were collected using the UAV from a road construction site in Gyeonggi in South Korea. When photographing at 80m above the ground using the equipment, the average ground sample distance (which is the distance between two concrete pixel centers on the ground) is 0.03m. The GSD is within the minimum resolution range for humans to recognize objects in the image, which ranges from 0.25 pixel/cm to 5 pixel/cm (Cohen et al. 2009). Considering the size of one UAV image is 5,280 x 3,956 pixels, the actual size of the site contained in one UAV image was about 140 x 105 m^2, where numerous objects on the site existed simultaneously. Accordingly, there were problems with building a classification model with all of the UAV images in terms of the computing capacity required for processing the image and data annotation. To address these problems, remote sensing and geoscience communities disassembled a UAV image into several small-sized patches to fit the computing capacity of the deep-learning models—not reducing the image size to keep the spatial information on the image, which is referred to as the sliding window scheme (Han et al. 2015; Kussul et al. 2017) or patch-based scheme (Maggiori et al. 2016; Jiang et al. 2020). In this research, to

comply with the construction standard and address the computing capacity issue, the author cropped a UAV image into a patch to visually recognize and label the defined objects. As an example, this research defined “rocks” as larger than 600mm by the construction standard (USDOT 2014; MOLIT 2016b); however, it is difficult to recognize the size of “rocks” within the whole UAV image to perform data annotation.

In order to determine the optimal patch size, the patch was divided into the smallest size that can be labeled with the naked eye (3 x 3 m^2, 5 x 5 m^2, and 7 x 7 m^2), and a pilot test was conducted. An orthomosaic image (3,000pixels x 3,000pixels) was prepared for dataset development. Each dataset was built for each patch size; a total of 1,850 patches for 3 x 3 m^2, 677 patches for 5 x 5 m^2, 343 patches for 7 x 7 m^2). ResNet-based classification networks were trained with the datasets, and their performances were compared accordingly. As a result, as shown in Figure 5, the model trained on 5 x 5m^2 patch showed the highest performance. Accordingly, a UAV image was cropped the into a patch 5 x 5m^2 to comply with the construction standard and address the computing capacity issue.

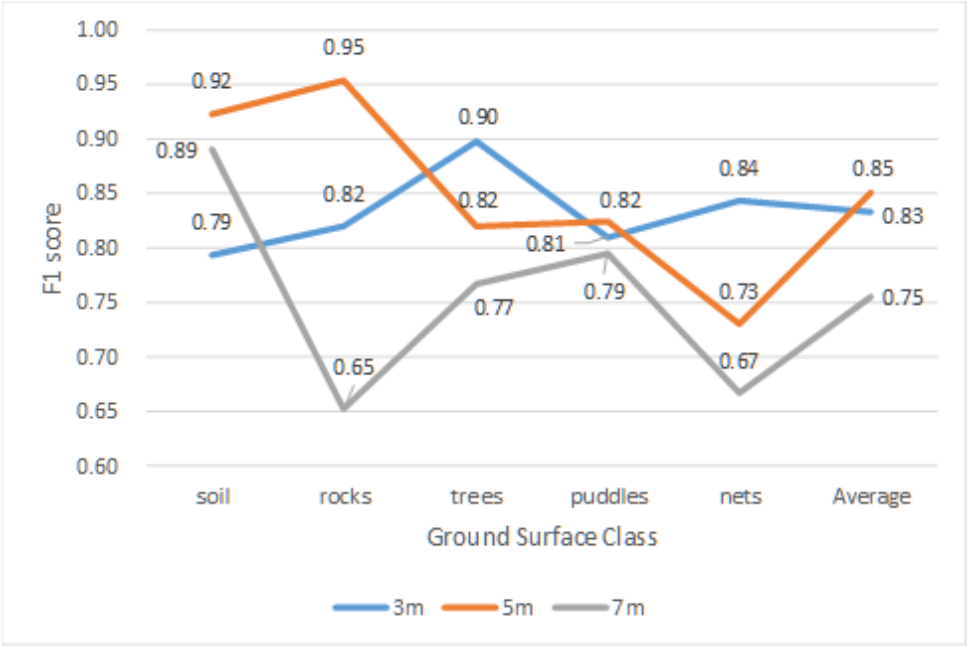


Figure 3.2 Classification model performance by patch size

3.1.2. Data Annotation

First, the collected UAV images were then processed into an orthomosaic image and cropped into 5 x 5m^2 patches. Each patch was annotated to the corresponding ground surface class (i.e., soil or non-soil). As a result, the datasets were developed with 347 “soil” patches (8.6%) and 3,649 “non-soil” patches (91.4%). Since the construction site was in the earthworks stage, “soil” was a relatively small portion. As the datasets had a class imbalance problem (i.e., “soil” 8.6% versus “non-soil” 91.4%), affecting the training performance of the classification network, this research applied data bagging as a solution. Finally, the soil patches were resampled to 3,470 patches, and a

total of 7,119 patches were prepared. The datasets were divided into 4,984 patches (70%) for training and 2,135 patches (30%) for testing.

Next, the "non-soil" (3,649 patches) were then annotated to the corresponding ground surface class (i.e., rocks, trees, puddles, nets). When multiple classes appeared in a "non-soil" patch, the author duplicated the labeling for each class that appeared in the patch. For example, if "trees" and "rocks" coexisted in a "non-soil" patch, the patch was respectively labeled "trees" and "rocks." As a result, 3,328 patches of "rocks," 3,393 patches of "non-rocks," 3,586 patches of "trees," 3,712 patches of "non-trees," 3,600 patches of "puddles," 3,641 patches of "non-puddles," 3,264 patches of "nets," and 3,377 patches of "non-nets." Each dataset was then divided into 70% for training and 30% for testing.

3.2. Super-resolution-based Data Augmentation

3.2.1. Proposed Method

In the field of computer vision, a long-standing vision problem is the poor performance of object recognition models when objects are small or difficult to see due to poor image quality. The problem is pronounced in the case of images taken from a long distance or aerial images acquired at high altitudes using UAV. In particular, when photographing an area using a UAV, it is common to take pictures while the UAV is moving (UAV velocity of 3m/sec – 10m/sec in this research), which causes quality issues such as blurring the photographed picture. In order to solve the low-quality issue of the aerial images, image restoration-related studies have been actively conducted (Lei et al. 2018; Arun et al. 2019). Furthermore, super-resolution (SR) can be a good data augmentation approach to improve data-driven models (Shorten and Khoshgoftaar 2019).

Although there are many studies that the super-resolution technique enhances the image quality and improves the model performance (Wu et al. 2015; Lei et al. 2018; Arun et al. 2019; Shim et al. 2022), a study has shown that image quality and model performance are not proportional (Shermeyer and Van 2019). In addition, a review paper suggested that using SR images as additional data for learning could be a suitable data augmentation method (Shorten and Khoshgoftaar 2019), but there was no case of actually verifying

its effectiveness. Hence, the author conducts experiment in actual construction site environments and verifies the two hypotheses: (1) applying SR improves the performance of the ground surface classification model, and (2) using additional data which SR is applied (i.e., data augmentation) improves the performance of the classification model. In the experiment for the hypothesis (1), the author builds classification models using data before and after SR application, respectively, and compares their performances. As an experiment for hypothesis (2), the author compares the performance of a model trained with data after SR application and a model trained with both data before and after SR application. The ResNet-101 network is utilized as the backbone network of the classification model.

This research applies a SOTA algorithm for super-resolution, Swin transformer-based image restoration (SwinIR), to UAV images (Liang et al. 2021). The architecture of SwinIR consists of three modules: shallow feature extraction, deep feature extraction and high-quality image reconstruction as shown in Figure 3.3. Given a low-quality input image, the shallow feature extraction module firstly extracts the features of the input image. Then, the deep feature extraction module which is composed of several residual Swin Transformer blocks (RSTB) extracts the deep features of the input image. Finally, a high-quality image is reconstructed by the shallow and deep features.

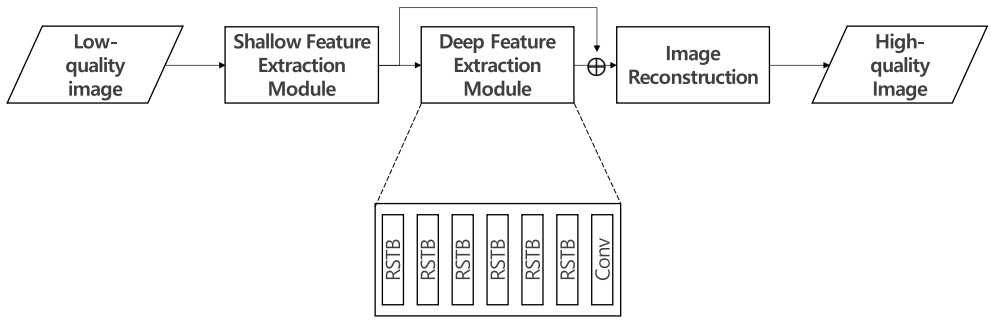


Figure 3.3 SwinIR architecture

3.2.2. Experimental Results and Analysis

Two performance metrics were determined for the experiment: blind/referenceless image spatial quality evaluator (BRISQUE) and f1 score. BRISQUE is a no-reference image quality assessment metric, ranging from 0 (high-quality) to 100 (low-quality) (Mittal et al. 2012). F1 score is the harmonic mean of the precision and recall rate. The precision and recall rate were calculated by using label-based evaluation metric (Sorower 2010) as shown in Eq. (3.1) – (3.2).

$$Precision = \frac{\sum_{i=1}^{n} Y_i \cap \hat{Y}_i}{\sum_{i=1}^{n} \hat{Y}_i} \quad (3.1)$$

$$Recall = \frac{\sum_{i=1}^{n} Y_i \cap \hat{Y}_i}{\sum_{i=1}^{n} Y_i} \quad (3.2)$$

where, n is the number of data, Y_i is true label for i^{th} data and $\hat{Y}_i$ is predicted label for i^{th} data, and $\cap$ is logical AND operator. BR model was more powerful in classifying all ground surface types.

The average BRISQUE score for the plain (i.e., before SR application) images was 30.49, and the score for the image after SR application was 27.66, which decreased by 2.83 when SR was applied. Therefore, the image quality was quantitatively improved by SR. As shown in Figure 3.4, the resolution of the image has increased; accordingly, the visibility of the object appearing in the image has increased.

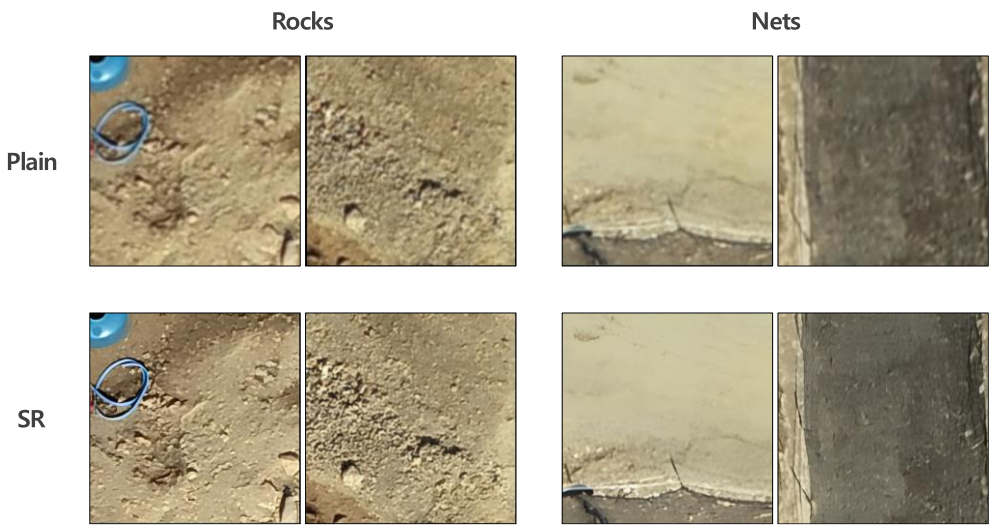


Figure 3.4 Plain and SR images: “rocks” and “nets”

The classification model performance is shown in the Figure 3.5. The model trained on SR data outperformed the model trained on plain data with

an average f1 score of 0.06 improvement: 0.02 improvement in “soil,” 0.05 improvement in “rocks,” 0.01 improvement in “trees,” 0.25 improvement in “nets.” As the sharpness of the object appearing in the image increases by improving image quality through SR, it helped the deep learning architecture extract features specialized for the object (e.g., edge of rocks), which led to the improvement of the model performance. Therefore, it can be interpreted as proof of hypothesis (1) applying SR improves the performance of the ground surface classification model.

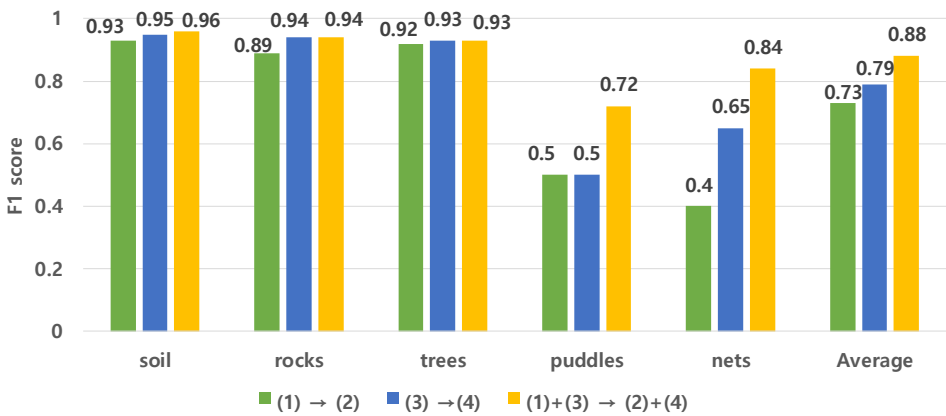


Figure 3.5 Classification model performance (Site A): Plain, SR, and Plain+SR

Moreover, the model trained on both plain and SR data outperformed the model trained on SR data only with an average f1 score of 0.09: 0.01 improvement in “soil,” 0.22 improvement in “puddles,” 0.19 improvements

in “nets.” In particular, the performance improvement in “puddles” and “nets” is noticeable, which can be interpreted as a decrease in overfitting as data increases. Figure 3.6 shows the degree to which the overfitting index of “puddles,” and “nets” change according to the training epoch. The models trained on both plain and SR data showed the lowest overfitting index over epochs. These results remark that the proposed SR-based data augmentation method reduces the overfitting and increases the model's general capability, which is consistent with the rules-of-thumb in computer science communities (Perez and Wang 2017; Aquino et al. 2017; Tellez et al. 2019).

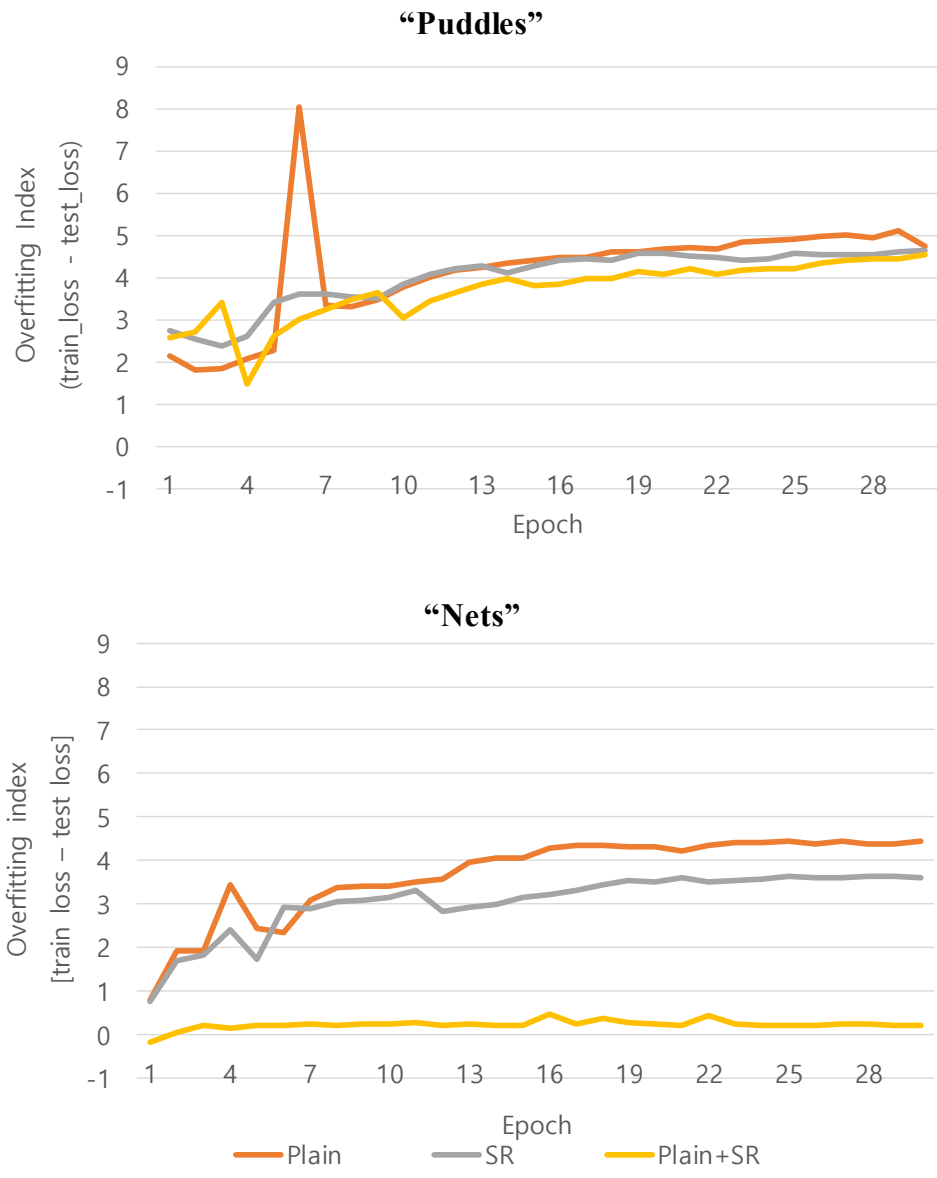


Figure 3.6 Overfitting index of "puddles," and "nets:" Plain, SR, and Plain+SR

Looking at the number of raw data in Table 3.2, the number of data in nets and puddles classes is very small due to the characteristics of the target site. Accordingly, it is interpreted that the augmentation effect was greater than that of other classes. In other words, it can be seen that the smaller the data, the greater the data augmentation effect.

Table 3.1 The number of patches: raw, bagged, augmented

Class	Raw	The number of patches	
		Bagged	SR augmented
soil	347	3,470	**6,940**
rocks	256	3,328	**6,656**
trees	1,793	3,586	**7,172**
puddles	100	3,600	**7,200**
nets	272	3,264	**6,528**

3.2.3. Generalization

In order to generalize the proposed method, this research collected additional data from other site environment (Site B), and developed a classification model and evaluated the model's performance. As a result, the model performance improved in the order of Plain, SR, and Plain+SR, as in the previous results. The model trained on SR data outperformed the model trained on plain data with an average f1 score of 0.03 improvement: 0.01 improvement in "rocks," 0.02 improvement in "trees," 0.16 improvement in

"nets." The model trained on both plain and SR data outperformed the model trained on SR data only with an average f1 score of 0.12: 0.02 improvement in "soil," 0.02 improvement in "rocks," 0.36 improvement in "puddles," 0.17 improvements in "nets." Therefore, both hypotheses presented in chapter 3.2.1 were verified.

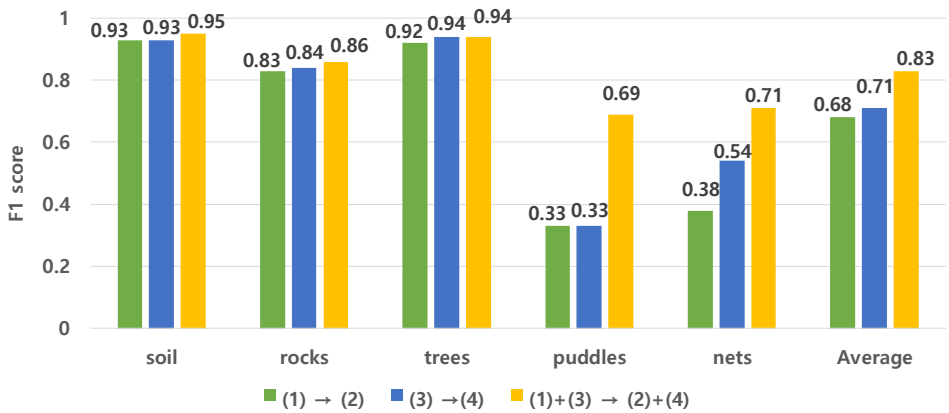


Figure 3.7 Classification model performance (Site B): Plain, SR, and Plain+SR

Furthermore, additional experiments were performed to evaluate the model's usefulness developed by the method proposed in this research. The data can be divided as shown in Figure 3.8, and the performance of the Plain+SR model trained with the data of (1) and (3) was tested using (2) and (4). The purpose is to evaluate the performance and determine whether SR is required for the model's input data when using the actual model. Site A has a

total of 17,248 patches, and Site B has a total of 23,371 patches, which were divided into Train (70%) and Test (30%), respectively, and two datasets (Plain, SR) were constructed.

	Train	Test
Plain	(1) 70% (Site A: 12,075; Site B: 16,361)	(2) 30% (Site A: 5,173; Site B: 7,010)
SR	(3) 70% (Site A: 12,075; Site B: 16,361)	(4) 30% (Site A: 5,173; Site B: 7,010)

Figure 3.8 Datasets for testing the Plain+SR model

As a result of model development, when SR is applied to both Site A and Site B, there is a slight performance improvement in some classes (soil and puddles), but the overall performance is similar, as shown in Figure 3.9. Therefore, it was found that SR can be applied only in the training stage of the model, and in actual use, acceptable model performance can be exhibited without enhancing the quality of the input image with SR.

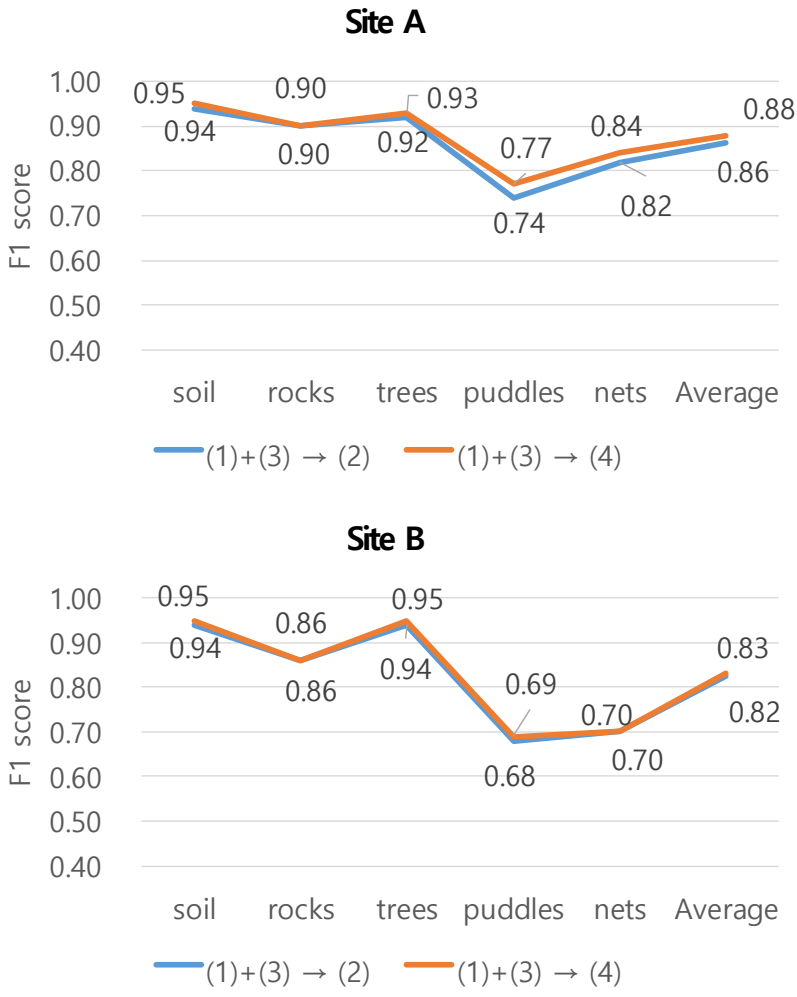


Figure 3.9 Plain+SR model performance by Plain or SR inputs

3.3. Summary

Low quality issue often occurs in aerial images acquired with UAVs, which deteriorate the performance of data-driven approach (e.g., deep learning). In order to address the issue, the author proposed a data augmentation method using the SR network. The proposed method was verified based on the experiments in actual construction site environments. As a result, the method improved the quality and quantity of UAV images, and thus improved the performance of the deep learning-based classification model.

Chapter 4. Ground Surface Classification

In this chapter, the author investigates the characteristics of UAV image and ground surface types, and suggests a novel classification method to automatically identify the ground surface types from UAV images, as shown in Figure 4.1. Two different multi-label classification methods and two different SOTA deep-learning-based classification networks are applied and comparatively evaluated for deriving the best-performance model.

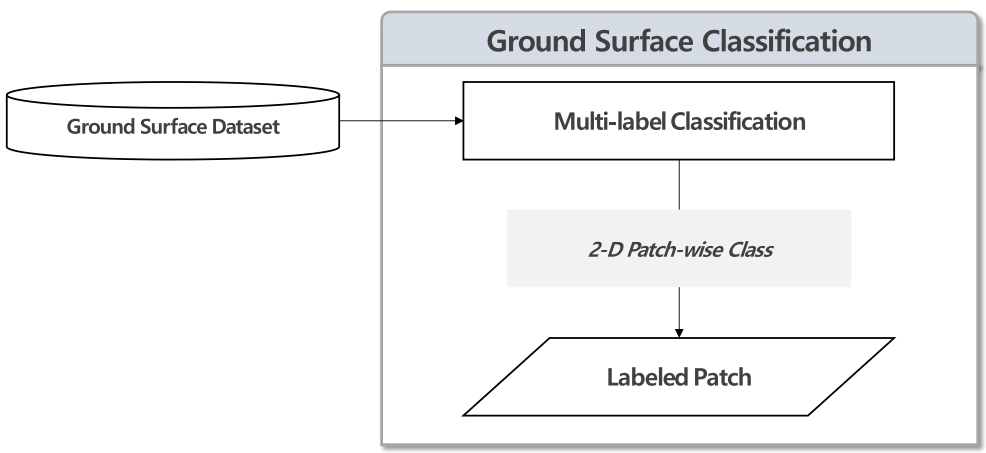


Figure 4.1 Research flow of ground surface classification

4.1. Proposed Method

The detection of the ground surface type from each patch is defined as a multi-label classification problem because several classes appear in isolation or simultaneously. Multi-label classification deals with issues where each example is represented by a single instance while simultaneously being associated with multiple class labels (Zhang et al. 2018). Both binary relevance (BR) (Read et al. 2011) and label powerset (LP) (Spolaôr et al. 2013) approaches are arguably the most intuitive solutions for learning from multi-label examples (Zhang et al. 2018) such as those described in Table 4.1. BR is a popular problem-transformation method that decomposes the multi-label task into independent binary learning tasks (Tsoumakas 2009). Accordingly, the class dependencies are ignored because each classifier is trained independently. Meanwhile, LP considers each unique set of labels in a multi-label training set as a class of a new single-label classification task (Tsoumakas 2009). Therefore, the dependencies between classes can be trained using the combinations of classes. Accordingly, the complexity of LP model can be high because the class types increased into 2^c (with c being the number of classes). This research applies these two approaches for the classification model and compared their performance.

Table 4.1. Multi-label classification methods

Method	Description	Characteristic
Binary relevance (BR)	Decompose the multi-label task into a number of independent binary learning tasks	• Independence between classes
Label powerset (LP)	Consider each unique set of labels that exists in a multi-label training set as a class of a new single-label classification task	• Dependence between classes • High complexity of the model • Vulnerable to a new combination of classes

This research built classification models for both the BR and LP using Residual Neural Network (ResNet) or Vision Transformer (VIT) as the backbone network, which are the most popular and outstanding feature-extraction architectures extensively used in both computer vision and construction management communities (Luo et al. 2019; Nath et al. 2020; Pi et al. 2020; Wortsman et al. 2022). The classification models consisted of two modules: (1) the 1st module that classified "soil;" and (2) the 2nd module, which classified "non-soil" into a detailed class. The 1st module classified an input patch into "soil" or "non-soil" through the first classifier. The classified "non-soil" patch was then passed through each classifier in the 2nd module of the BR and determined whether each class (i.e., "rocks," "trees," "puddles," "nets") existed or not in the patch (Figure 4.2a). Conversely, the 2nd module of LP was designed as shown in Figure 4.2b. The classified "non-soil" patch then passed through another classifier in the 2nd module of LP and was

classified into the detailed non-soil classes (i.e., “rocks,” “trees,” “puddles,” “nets,” “etc,” “rocks,nets” “rocks,etc” “rocks,trees” “puddles,etc”).

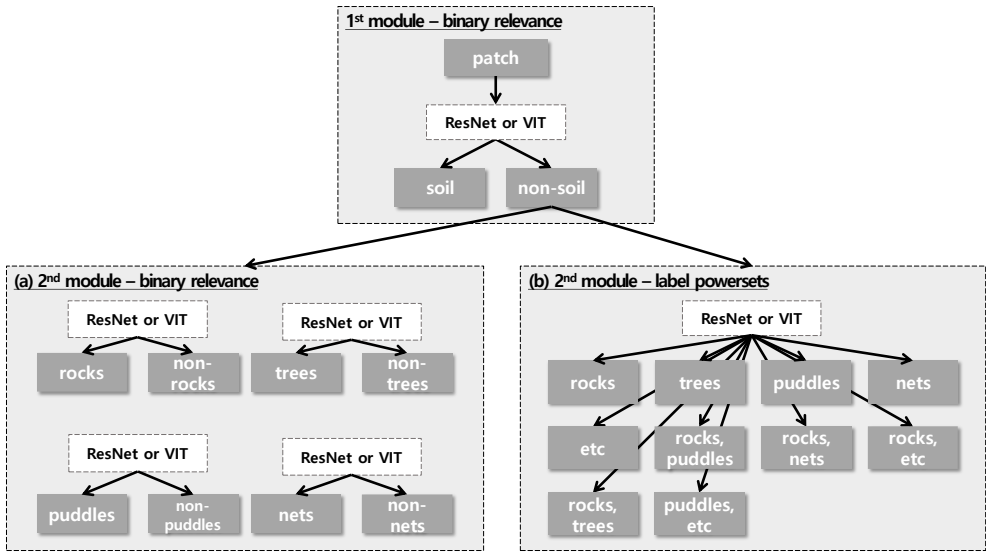


Figure 4.2. Classification networks: (a) BR model and (b) LP model

The 1st module pre-classified “soil” through a separate classifier in the proposed classification models because “soil” (i.e., subsoil layer) is essential in road construction. The ultimate goal of the earthworks stage is to make uniform subsoils that do not contain any other objects; that is, only uniform “soil” class exists (USDOT 2014; MOLIT 2016a). Accordingly, this research proposed a hierarchical classification network that first classified “soil” and subdivided “non-soil” into sub-classes (i.e., “rocks,” “trees,” “puddles,” “nets”) required to be managed according to construction standards (USDOT 2014; MOLIT 2016a).

To compile the classifier, the author tuned hyper-parameters as described in Table 4.2. The author compiled the BR modules with *sigmoid* as the output activation function, *binary cross-entropy* as the loss function, and *adam* as the optimizer. At the same time, the LP module was compiled with *softmax* as the output activation function, *categorical cross-entropy* as the loss function, and *adam* as the optimizer. *Sigmoid* is widely used for binary classification, but *softmax* is used for multi-class classification (e.g., LP) as the last layer of the model (Sharma 2019). The *adam* is used as optimizer for training the models, which is computationally efficient, requires little memory, and is well suited for problems that are large in terms of data and parameters (Kingma and Ba 2014). All of the classifiers were trained for 30 epochs.

Table 4.2 Hyper-parameters for compiling the classification network

Module	Activation function	Loss function	Optimizer	Epoch
BR	Sigmoid	Binary cross-entropy	Adam	30
LP	Softmax	Categorical cross-entropy		

Datasets are developed by the super-resolution-based data augmentation method, as described in Chapter 3, for training and validating the proposed models. Each patch is annotated to the corresponding ground surface class for the 1st module in the BR and LP models (i.e., "soil" or "non-soil"). As a result, a total of 7,119 patches are prepared for the 1st module (3,470 patches for "soil" and 3,649 patches for "non-soil." The datasets are divided into 4,984 patches (70%) for training and 2,135 patches (30%) for testing. Next, the "non-soil" patches are then annotated to the corresponding ground surface class (i.e., rocks, trees, puddles, nets, etc.) by applying the BR and LP approaches. For the BR model, the author utilizes the dataset from Chapter 3.1.3. As a result, 3,328 patches of "rocks," 3,393 patches of "non-rocks," 3,586 patches of "trees," 3,712 patches of "non-trees," 3,600 patches of "puddles," 3,641 patches of "non-puddles," 3,264 patches of "nets," and 3,377 patches of "non-nets" were prepared for the 2nd module of BR. Whereas, for the LP, when multiple classes appear in a "non-soil" patch, the author creates a new label for the multiple classes: "trees,rocks." As a result, 3560 of "rocks," 3,537 of "rocks,etc," 3,478 of "rocks,nets," 3,458 of "rocks,trees," 3,480 of "rocks,trees,etc," 3,731 of "trees," 3,563 of "trees,etc,"

3,480of “trees,nets,” 3,432 of “trees,nets,etc,” 3,816 of “etc,” 3,472 of “nets,” 3,432 of “nets,etc,” 3,475 of “puddles,” 3,480 of “puddles,etc” are prepared for the 2nd module of LP. Each dataset is then divided into 70% for training and 30% for testing.

4.2. Experimental Results and Analysis

The performance metric was determined with f1-score, which is the harmonic mean of the precision and recall rate, which were calculated by using label-based evaluation metric as described in Eq. (3.1) – (3.2). Table 4.3 shows the experimental results: the average f1 score for the BR and LP models was 0.88 and 0.68, respectively in ResNet backbone; 0.76 and 0.65, respectively in VIT backbone. As a result, the BR outperformed the LP for all ground surface classes under the same experimental conditions (i.e., the same backbone network, the same epoch values, and the same optimizer, as described in Table 4.2). Specifically, BR outperforms LP by the differences of “rocks” (0.42), “trees” (0.23), “puddles” (0.16), and “nets” (0.19) in ResNet; “rocks” (0.23), “trees” (0.13), “puddles” (0.17), and “nets” (0.07) in VIT. The differences in classification performance can be interpreted by the characteristics of the ground surface at the construction site. The classes defined in this research (i.e., “soil,” “rocks,” “trees,” “puddles,” “nets”) all exist in one population (i.e., the construction site), but the existence of an object does not affect any other object; that is, they are mutually independent. Therefore, it can be concluded that the BR approach is more suitable for ground surface classification than the LP approach by their characteristics as described earlier in Table 4.1. In particular, the two classes “rocks” and “trees” showed significant differences in performance because these two

classes often appeared together with other objects in a patch (e.g., “rocks,puddles,” “rocks,nets,” “rocks,trees,”). Since LP does not reflect the above-mentioned characteristic of the ground surface (i.e., mutually independent), the number of misclassifications increases that much in the “trees” and “rocks,” which can be interpreted as more differences in the performance than other classes.

Table 4.3 Classification model performance according to classification methods (BR and LP) and networks (ResNet and VIT)

Class	BR_ResNet	LP_ResNet	BR_VIT	LP_VIT
soil	0.96	0.96	0.98	0.98
rocks	0.94	0.52	0.71	0.44
trees	0.93	0.70	0.80	0.70
puddles	0.72	0.56	0.55	0.55
nets	0.84	0.65	0.77	0.58
Average	0.88	0.68	0.76	0.65

Otherwise, in terms of the backbone network, the average f1 score for the ResNet and VIT was 0.88 and 0.76, respectively in BR; 0.68 and 0.65, respectively in LP. Therefore, ResNet outperformed VIT in most of the ground surface classes. The differences in classification performance can be interpreted with the VIT’s characteristic. VIT can achieve the best performance only after prior learning of the model with a large amount of data and fine-tuning with a medium-level dataset (Dosovitskiy et al. 2020).

Therefore, due to the limitation of the amount of UAV image data collected from the construction site, the model performance was better in the ResNet architecture for the classification of ground surface in the construction site.

Although the experimental results showed acceptable performance for the models, there is one point to note. The models' performances were particularly poor at classifying "puddles." When only puddles appeared on the ground surface, the model classified it correctly (as shown in Figure 6a); however, when puddles appeared simultaneously with other object classes, the model had a tendency to misclassify the puddles (see Figure 6b). Due to the characteristics of water, such as color and reflection, the puddles at the construction site had similar visual attributes to the objects in them. For example, as shown in Figure 5b, when rocks were present in the puddles, they were classified as "rocks" only, or when the water was green, the puddle was classified as "trees." It was found that the object existing in the puddle had a significant effect on the model's puddle recognition capability.

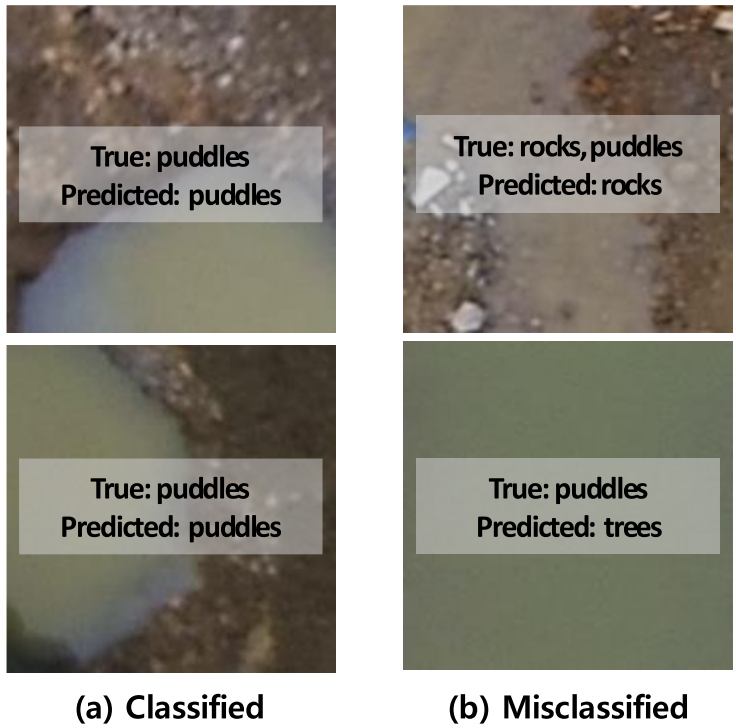


Figure 4.3 Examples of classification results: (a) classified, (b) misclassified

4.3. Summary

In this chapter, the author developed a ResNet-based multi-label classification model. The architecture of the model was constructed in that subsoil which is the final object of earthworks, was pre-classified, and then "non-soil" was further classified as "rocks," "trees," "puddles," and "nets." Two representative multi-label classification methods (i.e., binary relevance and label powersets) were applied to solve the problem of multiple classes appearing in one patch when classifying "non-soil." Moreover, two different SOTA deep-learning-based classification networks – ResNet and VIT – were applied and evaluated for deriving the best-performance model. By comparing the two methods' performances, binary relevance with ResNet (average f1 score 0.88) was the most suitable method for ground surface classification in road construction sites.

Chapter 5. Ground Surface Area Estimation

This chapter aims to estimate the area by the ground surface types by quantifying 2-D pixel-wise ground surface class resulted from the previous chapters, as shown in Figure 1.2. The suggested unsupervised segmentation method segments the labeled 2-D patch to estimate class-wise area by the ground surface types. Then, the estimated 2-D area is superimposed onto point clouds through the 3-D mapping, making it into labeled point clouds to visualize the results in 3-D for the automated equipment operation.

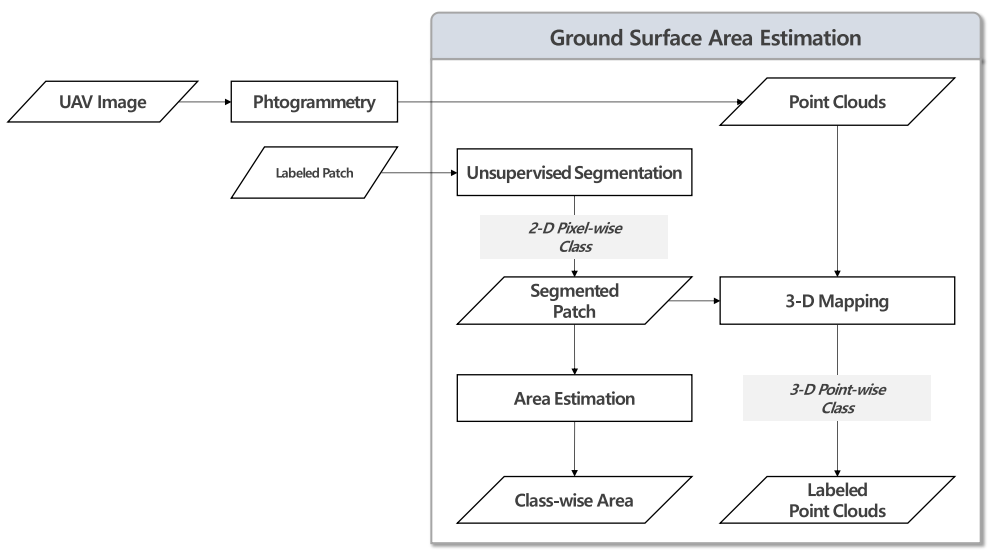


Figure 5.1 Research flow of ground surface area estimation

5.1. Proposed Method

5.1.1. Unsupervised Segmentation

The author processes the labeled patch (i.e., results from Chapter 4.1), which has a 2-D patch-wise ground surface class, through unsupervised segmentation to get a 2-D pixel-wise ground surface class information. The pixel-level information provides more accurate object shape of the ground surface than patch-level information, when applying the area estimation module in Chapter 5. This research applies convolutional neural network (CNN) for unsupervised image segmentation (Kim and Kanezaki 2020). The CNN assigns a label to a pixel that denote the cluster to which the pixel belongs. Once an image comes to the CNN, the pixel labels and feature representations are jointly optimized, and the gradient descent updates their parameters. Prediction of pixel's label and network parameters' update iterate to meet the following criteria as shown in Figure 4.5: (1) pixels of similar features should be assigned the same label (i.e., similarity loss), (2) spatially continuous pixels should be assigned the same label (i.e., spatial continuity loss), and (3) the number of unique labels should be large. Although these criteria are incompatible, the CNN minimizes the combination of similarity loss and spatial continuity loss in the back propagation with learning rate.

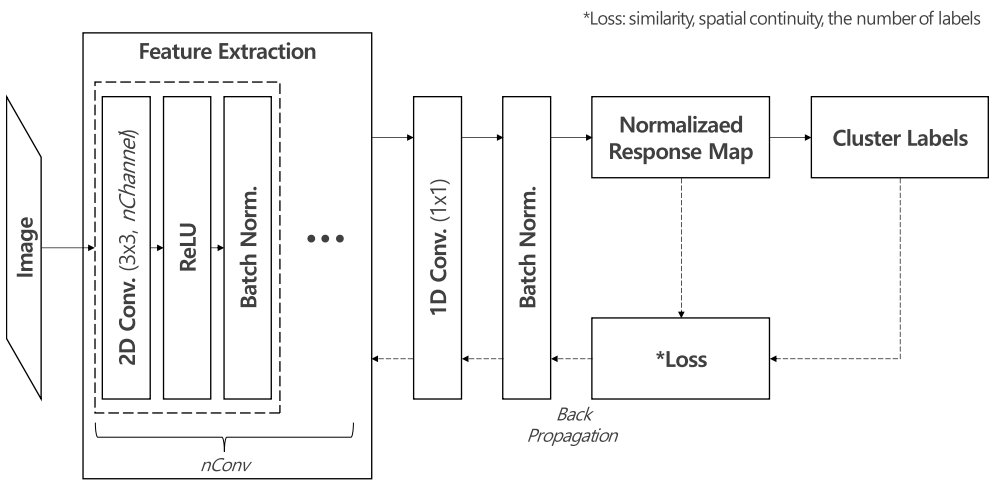


Figure 5.2 CNN-based unsupervised segmentation network

The author developed an unsupervised segmentation model by using the abovementioned CNN-based network. Hyper-parameters were tuned to build a model optimized for the ground surface. The corresponding hyper-parameters are the number of channels in a convolution layer (*nChannel*), the number of convolution layers (*nConv*), and iteration (*MaxIter*) as described in Table 4.4. Each hyper-parameter was tuned as follows: (1) configure an initial parameter combination with the above-mentioned hyper-parameters, (2) change a single parameter to be optimized within a tuning range to combine multiple parameter combinations (3) create a model for each parameter combination, (4) select the optimal parameter value by comparative evaluating of the mean Intersection over Union (mIoU) of the created models. The IoU was calculated by Jaccard index, as described in Eq. (5.1).

$$\text{Intersection Over Union (IoU)} = \frac{|True \cap Predicted|}{|True \cup Predicted|} \quad \text{Eq. (5.1)}$$

Table 5.1 Hyper-parameters for compiling the segmentation network

Hyper-parameter	Definition	Tuning range	Optimized value
nChannel	The number of channels in a convolution layer	0 – 100	30
MaxIter	The number of training iteration	100 – 1000	1,000
nConv	The number of convolution layers	1 – 10	2

An example of hyper-parameter tuning is shown in Figure 5.3 (the other hyper-parameters' tuning results are described in Appendix A). Using a segmentation annotation tool called Labelme, the author builds a set of correct answers, as shown in Figure 5.3(a), comparing them with the predicted values of the unsupervised segmentation algorithm. Finally, mIoU values for each class were derived, as shown in Figure 5.3(b), and then the average value (i.e., Average mIoU) was selected as the optimal parameter. The initial hyper-parameter setting referred from a previous study (Kim and Kanezaki 2020). In this case, the parameter to be optimized is *nChannel*, which ranges from 10 to 100. The author selected the best parameter value as 30 by comparing the average mIoU of the created models for each ground surface class. The other hyper-parameters (i.e., *nConv* and iteration) were also optimally derived in the same way as *nChannel*. As a result, optimal parameter combinations

were derived: *nChannel* 30, *MaxIter* 1,000, and *nConv* 2. Based on the model built with the optimal parameters, the labeled patch (i.e., 2-D patch-wise ground surface class) was processed into segmented patch (i.e., 2-D pixel-wise ground surface class). Then, the color occupying the largest proportion in the segmented patch was selected, and the class of the labeled patch was assigned to the pixel with having this color.

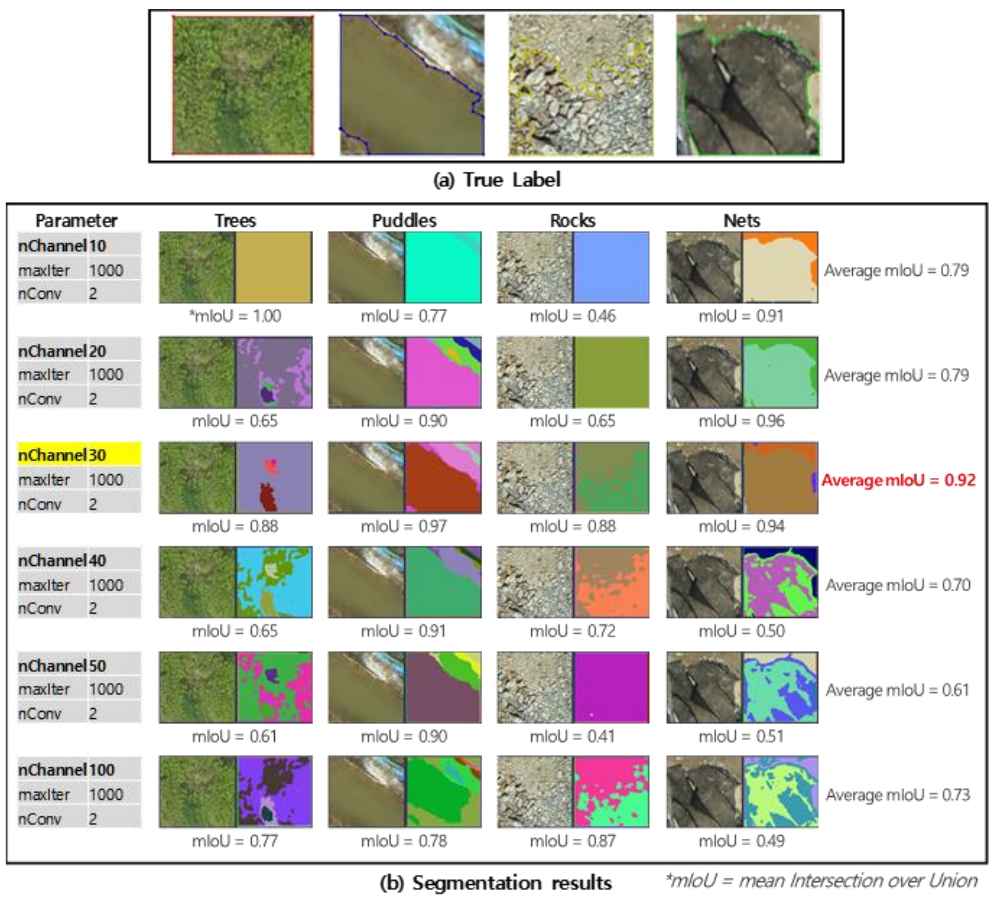


Figure 5.3 Examples of hyper-parameter tuning: *nChannel*

5.1.2. Area Estimation

Using the segmented patch resulting from the unsupervised segmentation, the ground surface pixel-wise class area is calculated as shown in the following equation.

$$\text{Class-wise Area} = Patch\ area * \alpha * \frac{1}{N} \quad \text{Eq. (5.2)}$$

where, α is segmentation ratio, and *N* is the number of classes in a patch. First, the patch area is derived as the product of the patch width and height, and then multiplied by the segmentation ratio (α), which is the proportion of pixels assigned pixel-wise class in the entire patch as shown in Figure 5.4. The input patch is pre-segmented through the unsupervised segmentation, and the index of the majority color from this pre-segmented patch is extracted through arguments of the maxima (*Argmax)*. The ratio of pixels with the index was calculated and reflected in the area calculation. Then, the calculated value (i.e., $Patch\ area * \alpha$) was divided by the number of classes (*N*). The *N* value is to prevent the overlapping calculation of the class area in the case of overlapping of several ground surface classes assigned to a patch. Finally, the class-wise area is derived by summing the area values for each patch calculated by class for the entire site.

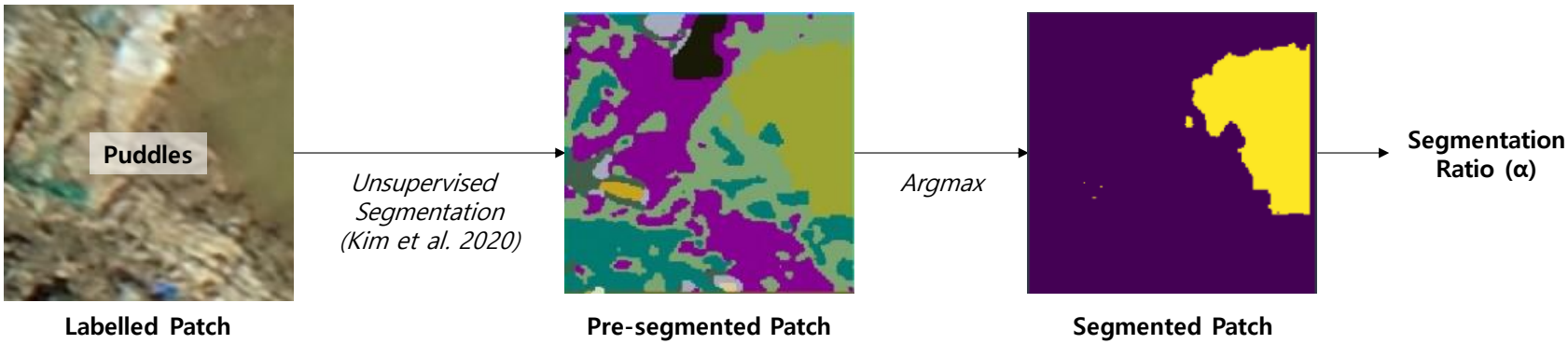


Figure 5.4 Area estimation process

5.2. Experimental Results and Analysis

This research verifies the proposed method (automated method) by comparing the area value calculated using the commercial tool (manual method).

5.2.1. Datasets preparation

First, the author prepared testing data for the experiment. Based on the UAV images collected in Chapter 3, a commercial tool, Pix4D, was used to generate the point cloud and manually quantify the area by ground surface types. The calculation method installed in Pix4D is as follows: (1) manually check the ground surface, (2) create a plane for area estimation by manually marking the boundary of the checked ground surface as shown in Figure 5.5, (3) select a plane setting method (i.e., align with lowest point, which is orthogonal projection of the point cloud on the XY plane), (4) calculate the XY plane area.

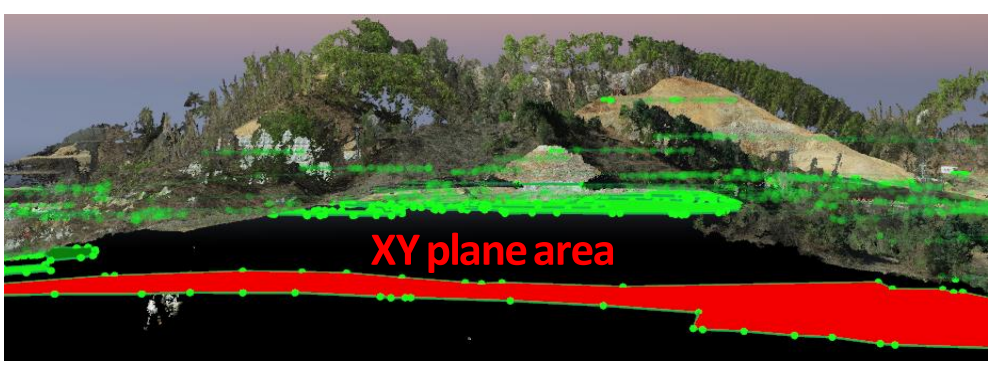


Figure 5.5 Area estimation by manual method

As shown in Figure 5.5, the area was calculated based on the Eq. 5.3.

$$A_i = L_i * W_i \quad \text{(Eq. 5.3)}$$

where, L_i is the length of the cell (i.e., ground sample distance), W_i is the width of the cell (i.e., ground sample distance). In this way, the area of all cells existing within the boundary set is calculated and these are summed. After that, the calculated values for each class are summed up and finally the area for each class is derived. It took about three days to calculate the area of all of the ground surface types on the site by the manual method when performed by one person. The calculated area for each ground surface class is shown in Table 5.1.

Table 5.2 Area calculated by manual method

Class	**Area (m^2)**
soil	3,208
rocks	2,473
trees	16,009
puddles	104
nets	2,672

5.2.2. Comparative Analysis

For comparison of the manual method and the proposed method, various conditions were set equally: (1) the density of point clouds was unified to 13,334,418 pts as shown in Figure 5.6a, and (2) the ranges of the point cloud and orthomosaic image were unified as shown in Figure 5.6b.

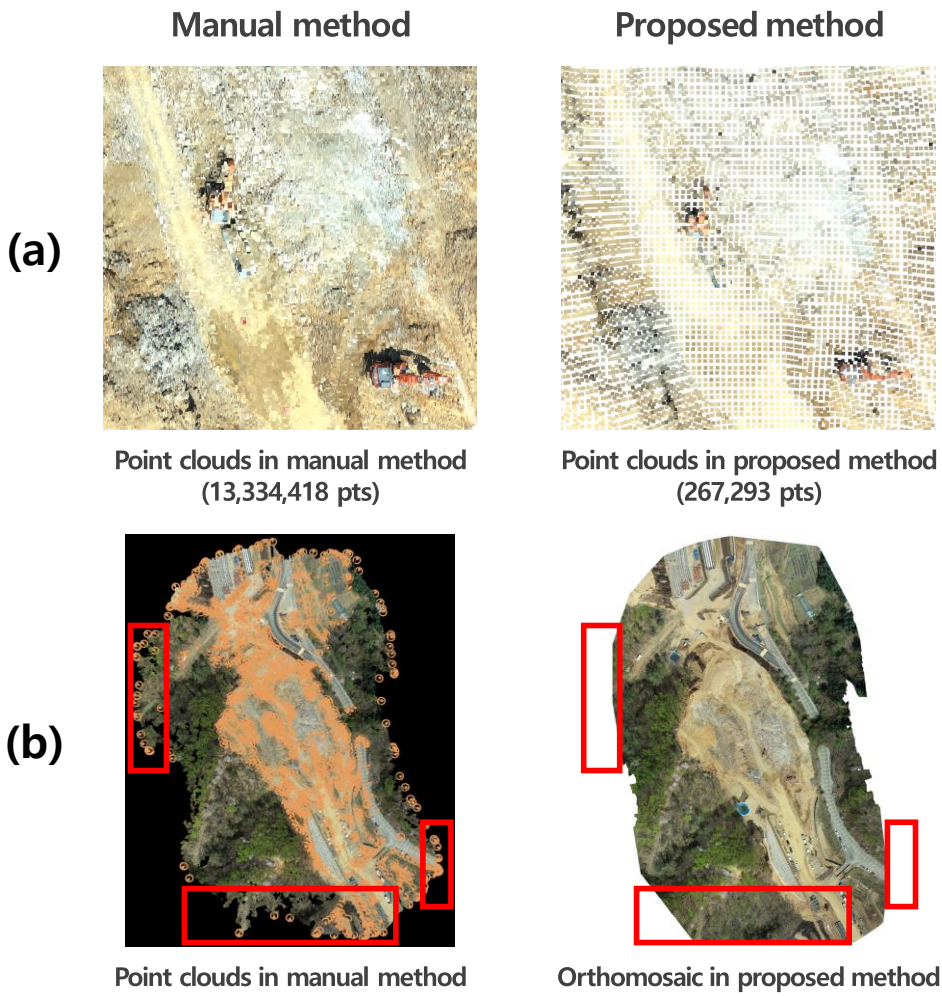


Figure 5.6 Differences between manual method and proposed method:

(a) Density difference, (b) Range difference

The error of the proposed method was analyzed by comparing the results with the results from the manual method as the correct answer. The metrics for error calculation are as described in Eq. 5.4 and Eq. 5.5.

$$\text{Absolute error} = |area\_manual - area\_automated| \quad \text{(Eq. 5.4)}$$

$$\text{Relative error} = \frac{\text{Absolute error}}{area\_manual} \quad \text{(Eq. 5.5)}$$

where, *area_manual* is a value resulted from the manual method, and the *area_automated* is a value resulted from the proposed method. As a result, the relative error was 'soil' 0.07, 'rocks' 0.11, 'trees' 0.09, 'puddles' 0.30, 'nets' 0.19, and the average relative error was 0.15, as described in Table 5.4.

Table 5.3 Experimental results: area estimation

Class	**Area_manual (m^2)**	**Estimated area (m^2)**	**Absolute error (m^2)**	**Relative error**
soil	3,208	2,970	239	0.07
rocks	2,473	2,739	266	0.11
trees	16,009	14,543	1,466	0.09
puddles	104	73	31	0.30
nets	2,672	2,164	508	0.19
Average	-	-	-	0.15

The class with the smallest error value is soil, and the class with the largest error value is puddles, which is the same result as the performance of the classification model. The comparative analysis between the area estimation results and the classification results are shown in the Figure 5.7. To compare the two models, the performance evaluation index was unified as an error. Accordingly, the evaluation index of the classification model was defined as '1 - f1 score.' The area estimation error is larger than the classification error in all classes, which can be interpreted as the classification error is accumulated in the area estimation. In most classes, the difference between the area estimation error and the classification error is 0.02 ~ 0.03, but in the case of ‘rocks’, the difference is 0.05, which means that the error difference is larger than that of other classes. The cause of this error can be interpreted as an error from the calculation method of unsupervised segmentation.

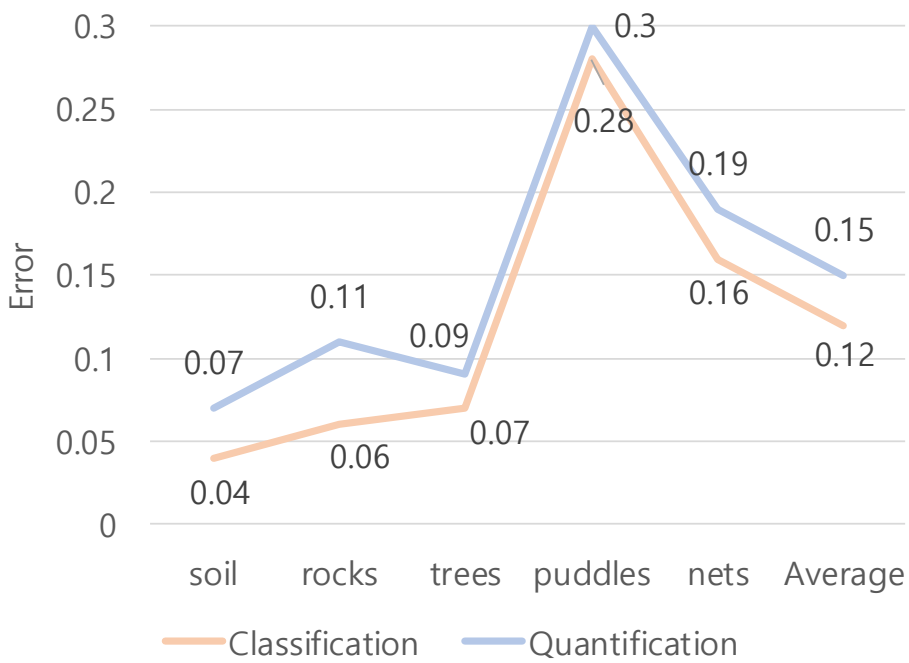


Figure 5.7 Performance comparison: classification errors vs. area estimation errors

Looking at an example in which the area estimation model has an error in 'rocks' is shown in Figure 5.8. A patch labeled with 'rocks' is generated by the classification model, which goes through the unsupervised segmentation model to be processed in a segmented patch. Then, the segment that occupies the largest area in the patch (i.e., major color) is extracted by *Argmax,* and the area is calculated for the segment. When deriving the major color with *Argmax*, in the case of 'rocks,' the area occupied by actual 'rocks' in the patch was smaller than that of other segments. Accordingly, there were cases

where the major color was extracted without segmenting the actual ‘rocks’, which led to an area calculation error.

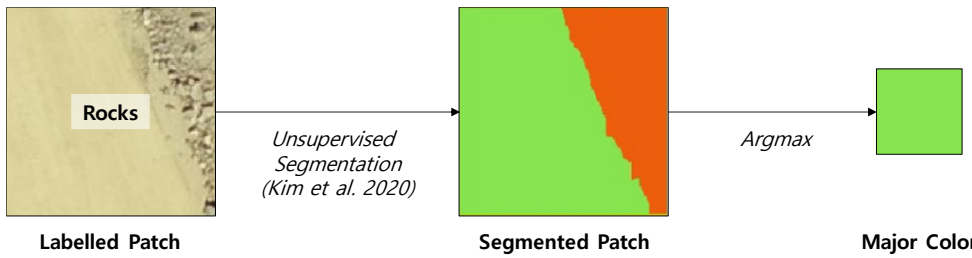


Figure 5.8 An example of major color mis-selection

5.3. 3-D Visualization

This chapter describes the visualization process of the 2-D results into 3-D, how to superimpose the segmented patch (Figure 5.9a) onto 3-D point clouds (Figure 5.9b) in the following process: (1) generates point clouds from UAV images utilizing photogrammetry techniques equipped on a commercial tool, Pix4D mapper, (2) extracts coordinates of four vertices of a pixel in the segmented patch, (3) the latitude and longitude of patch and point clouds have different coordinate systems: the coordinate system of the patch is universal transverse mercator (UTM), and the coordinate system of point clouds is world geodetic system (WGS84). The WGS84 system of the point clouds is converted into UTM, (4) find the equation of the straight-line of each side of the patch based on the latitude/longitude axis of the UTM coordinate system, and (5) if each point of the point cloud is included in the four straight-line equations, gives the patch label as the class of the point. A sample result is shown in Figure 5.2c. The other mapping results are shown in Appendix B.

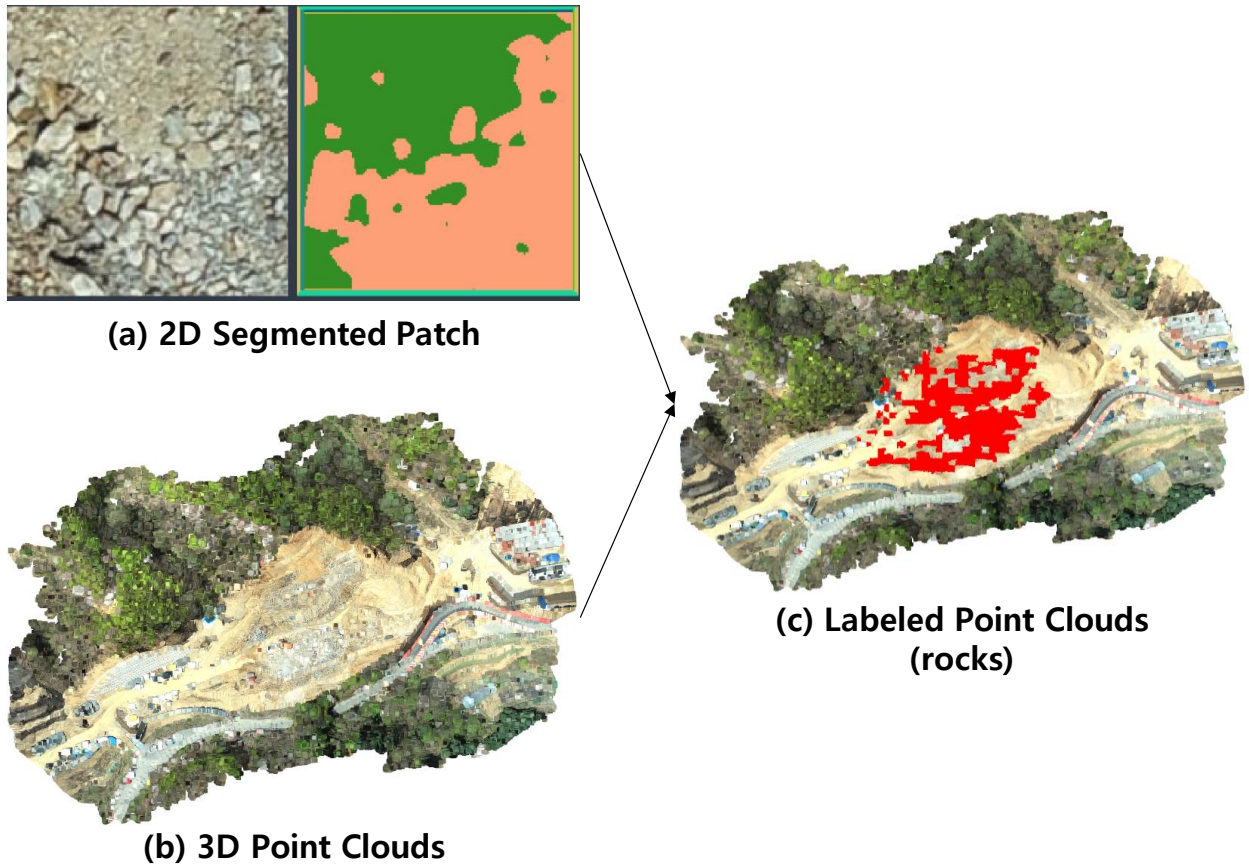


Figure 5.9 A concept of mapping: (a) 2-D segmented patch, (b) 3-D point clouds, (c) labeled point clouds (rocks)

5.4. Summary

In this chapter, the author developed a area estimation method to estimate the area by the ground surface types. This method consists of area estimation and visualization modules. The area estimation module segmented the labeled patch resulted from the previous chapter and estimated the area of the segmented patch with the average relative error 0.15. The visualization module superimposed the 2-D segmented patches onto 3-D point clouds to process the results in 3-D to provide the results with the automated equipment.

Chapter 6. Experimental Design and Analysis

This chapter presents the experimental design, results, and discussion to validate the proposed methodology and confirm the technical feasibility and the applicability of this research. The author performs experiments using UAV images collected from an actual road construction site to apply and validate the proposed methodology in real-world environments. The target construction site is in the same location as the site utilized in the previous chapters, but the data acquisition time is set one month later. As the construction period of one month passed, despite the same location, the site's characteristics were different, so it was suitable data for the validation. The area estimation result is defined as the final output of this research. The final output is generated by applying the proposed methodology (i.e., from Chapter 3 to Chapter 5 in this dissertation) and compared with the results from the manual method to evaluate the performances of the proposed methodology, as shown in Figure 6.1. The detailed process and results are described in the following sections.

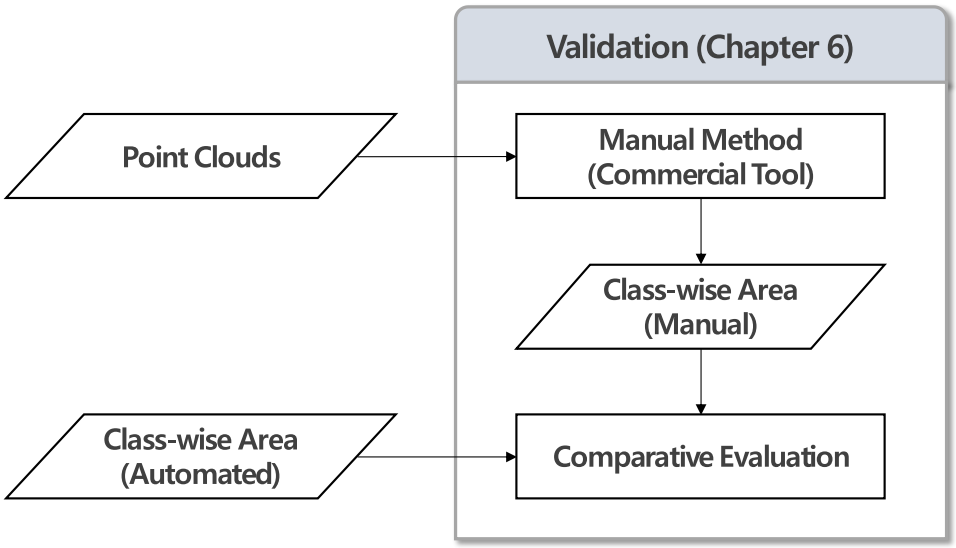


Figure 6.1 Research flow of validation of the proposed methodology

6.1. Experimental Design

The experiment is performed by the concept as shown in Figure 6.2. In practice, the practitioners need to use a commercial tool such as Pix4D to acquire the ground surface information (i.e., type, location, and area) from UAV images, referring to the manual method in Figure 6.2. As in the manual method, human input is required for class annotation, area estimation, and area summation by class to acquire the ground surface information; this research automated these processes. The collected UAV images are processed by the three modules of the proposed methodology (from chapter 3 to chapter 5 in this dissertation) to derive the final output which is the area for each ground surface class. Finally, the output is evaluated and validated by comparing the results derived by the manual method with the results from the proposed methodology. Performance criteria are f1 score for classification model, relative error for area estimation model, and processing time for the entire process.

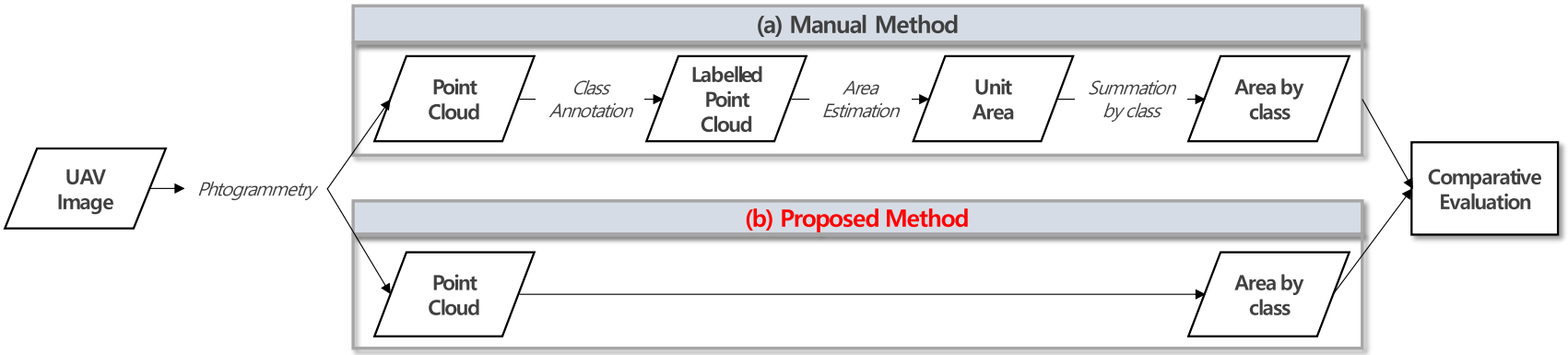


Figure 6.2 Experimental concept

The experiment was performed by the process, as described in Figure 6.3. First, the author collected the data from a road construction site in Gyeonggi in South Korea in both 2021.4.7. and 2021.5.7., using a UAV (DJI Phantom 4 RTK). A total of 524 aerial images were collected (199 images on 2021.4.7., and 325 images on 2021.5.7.) Then, the manual and proposed methods for ground surface area estimation were implemented. The results from the methods were evaluated and compared. The evaluation metrics for the performance comparison are BRISQUE for ground surface datasets; f1 score for ground surface classification; relative error and processing time for ground surface area estimation.

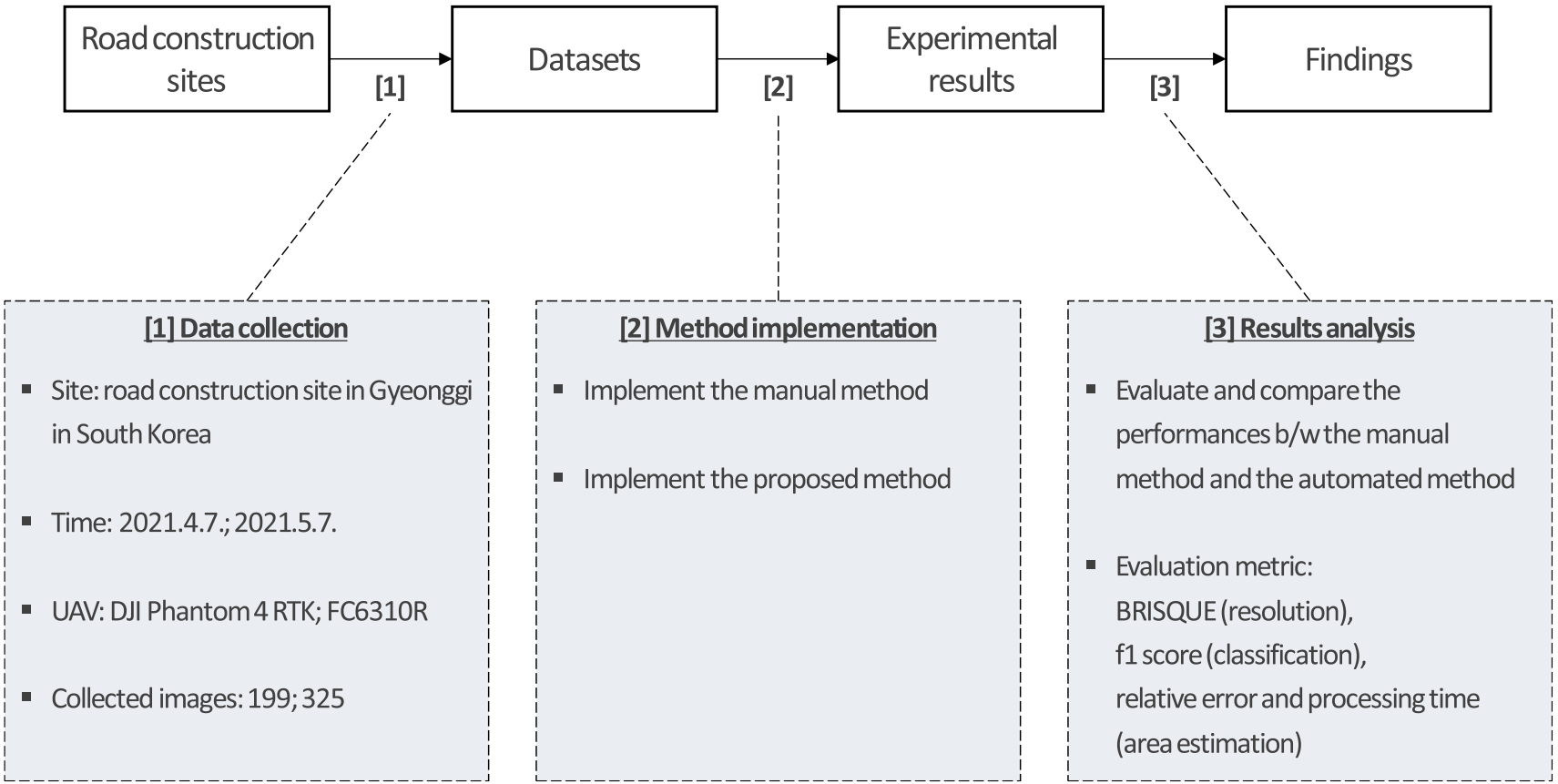


Figure 6.3 Experimental process

6.2. Experimental Results and Discussions

6.2.1. Ground Surface Datasets

The targeted jobsite was located in Gyeonggi in South Korea (courtesy of MOLIT). A UAV (the DJI Phantom 4 RTK) equipped with a camera (the FC6310R) flew to the site for data collection. As a result, 325 aerial images are collected from the site. Photographed at 80m above the ground, the image size is 5,472pixels x 3,648pixels; the average ground sample distance is 0.03m. The method presented in Chapter 3 was applied to the collected aerial images to construct the ground surface datasets. As a result, 23,371 patches were prepared for validation of the proposed methodology: 2,160 patches for "soil," 5,075 patches for "rocks," 4,934 patches for "trees," 5,658 patches for "puddles," and 5,544 patches for "nets." Super-resolution (SR) was applied to validate the model's performance and the findings from Chapter 3. As a result of SR application, the average BRISQUE score for the plain (i.e., before SR application) images was 35.12, and the score for the image after SR application was 29.27, which decreased by 5.85 after SR was applied. As shown in Figure 6.4, the resolution of the image has increased. Therefore, the image quality was quantitatively improved by SR; accordingly, the visibility of the object appearing in the image has increased.

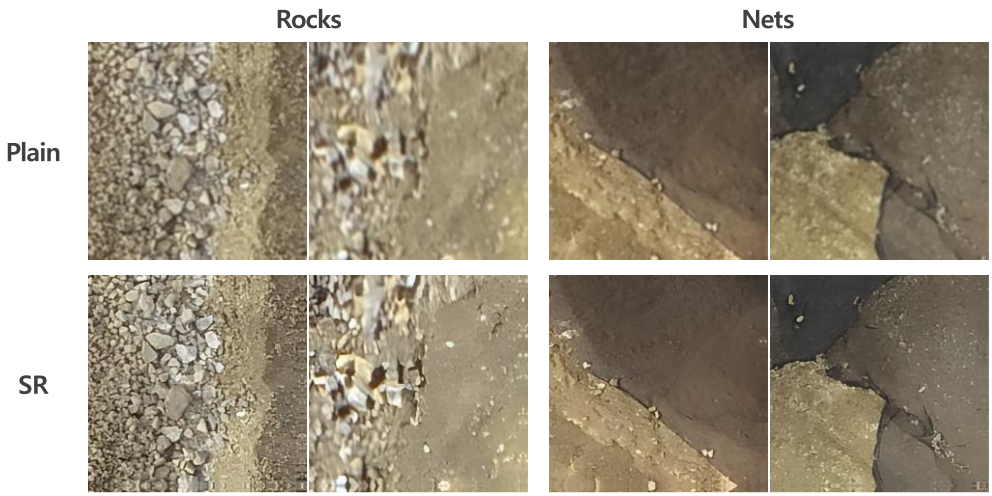


Figure 6.4 Plain and SR images: “rocks” and “nets”

6.2.2. Ground Surface Classification

The developed classification model in Chapter 4 (i.e., ResNet with BR method) was validated based on the developed datasets (i.e., plain images or SR images). When plain images or SR images were input as validation data, both showed the same performance, an average f1 score of 0.81, as described in Table 6.1. These results are in line with those in Chapter 3; the model's performance is not significantly different regardless of whether SR is applied to the input image. In other words, the proposed method suggests that SR application is required only in the model development stage (i.e., the training stage) and does not need to be applied in the actual use.

Table 6.1 Classification results by input data

Class	Input data	
	Plain	**SR**
soil	0.82	0.83
rocks	0.92	0.91
trees	0.79	0.81
puddles	0.74	0.72
nets	0.80	0.79
Average	0.81	0.81

In addition, it can be seen that the validation performance is somewhat inferior to that of the original model: the difference of f1 scores 0.14 in “soil,” 0.02 in “rocks,” 0.14 in “trees,” and 0.04 in “nets,” as shown in Figure 6.5. In particular, in the case of ‘soil’ and ‘trees,’ it can be seen that the validation f1 score is significantly lowered due to the significant difference in visual characteristics. On the other hand, in the case of 'rocks,' 'puddles,' and 'nets' at each site, the difference in visual characteristics was not significant even with the naked eye, indicating that the f1 score of the model was less degraded.

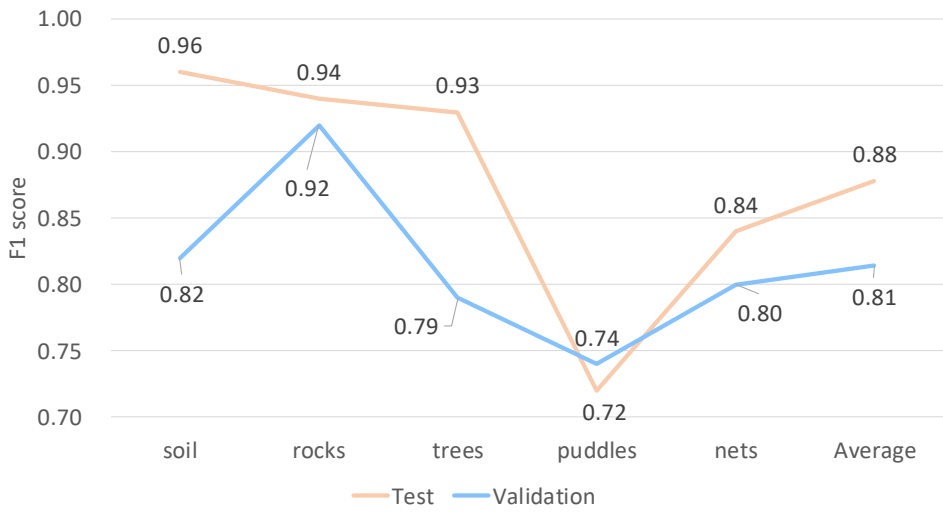


Figure 6.5 Classification results: test vs. validation

These performance deteriorations are due to the difference between the visual characteristics of ground surface types in the field environment trained by the original model and the characteristics in the new field environment in which the model was not trained. As shown in Figure 6.6, all of the ground surface types have different visual characteristics in both sites: test site (Site A) and validation site (Site B).

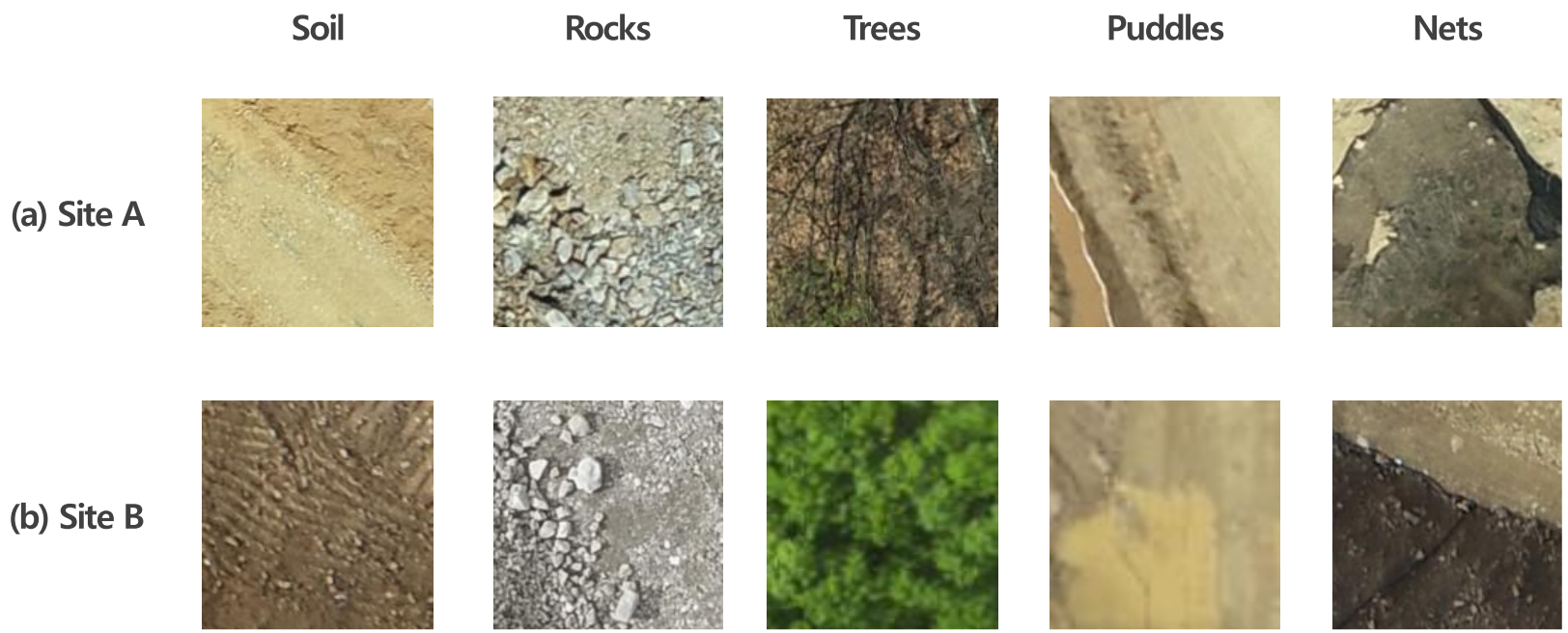


Figure 6.6 Ground surface types: (a) Site A and (b) Site B

This discrepancy in the visual characteristics by ground surface types came from the change of time and the progress of construction. In particular, since the target site is a bedrock area, ‘soil’ decreased, ‘rocks’ increased as earthworks progressed, and rainfall occurred within a month, resulting in an increase in ‘puddles,’ as shown in Figure 6.7. According to the difference in the ground surface ratio of the two sites, it can be interpreted that the different visual characteristics for each ground surface type occurred. The model could not learn these characteristics, which causes deterioration of the model’s performance.

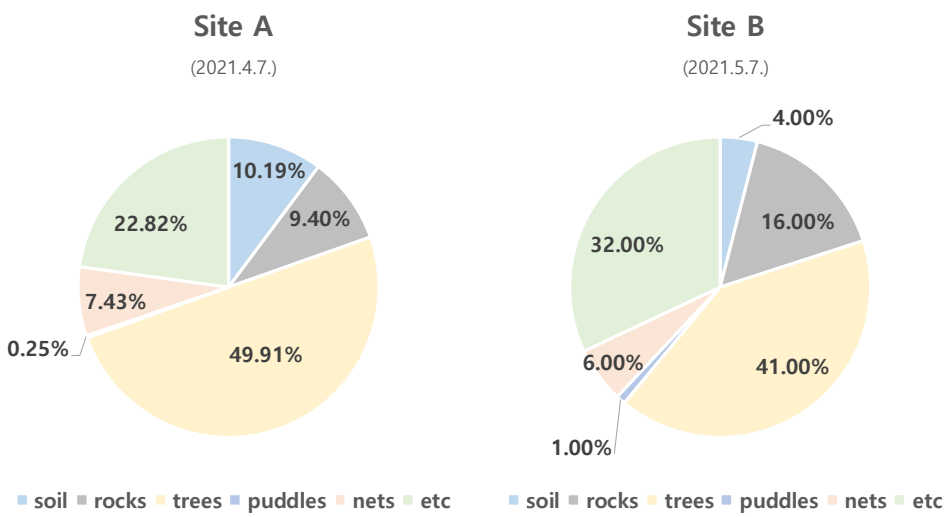


Figure 6.7 Discrepancy in visual characteristics between test site (Site A) and validation site (Site B)

6.2.3. Ground Surface Area Estimation

The ground surface area estimation was performed based on the labeled patch resulted from the classification model. First, the proposed segmentation module was applied. The examples of segmentation results are shown in Figure 6.8. In most classes, the ground surface class is successfully found and segmented inside the labeled patch.

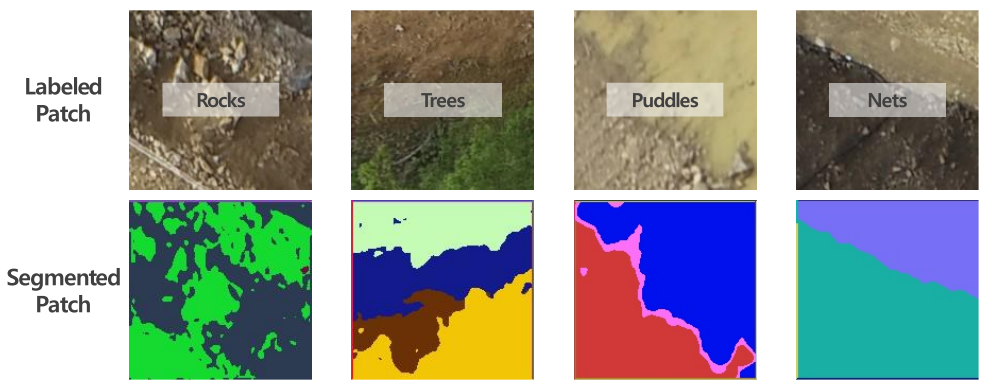


Figure 6.8 Examples of segmentation results

Next, the author prepared testing data by applying the manual method. Based on the UAV images collected in Chapter 6.2.1, a commercial tool, Pix4D, was used to generate the point cloud and manually quantify the area by ground surface types. The area estimation installed in Pix4D is as follows: (1) manually check the ground surface types in the program, (2) create a plane for estimating the area by manually marking the boundary of the checked ground surface, as shown in Figure 6.9, (3) select a plane setting method (i.e.,

align with lowest point, which is orthogonal projection of the point cloud on the XY plane), (4) calculate the XY plane area.

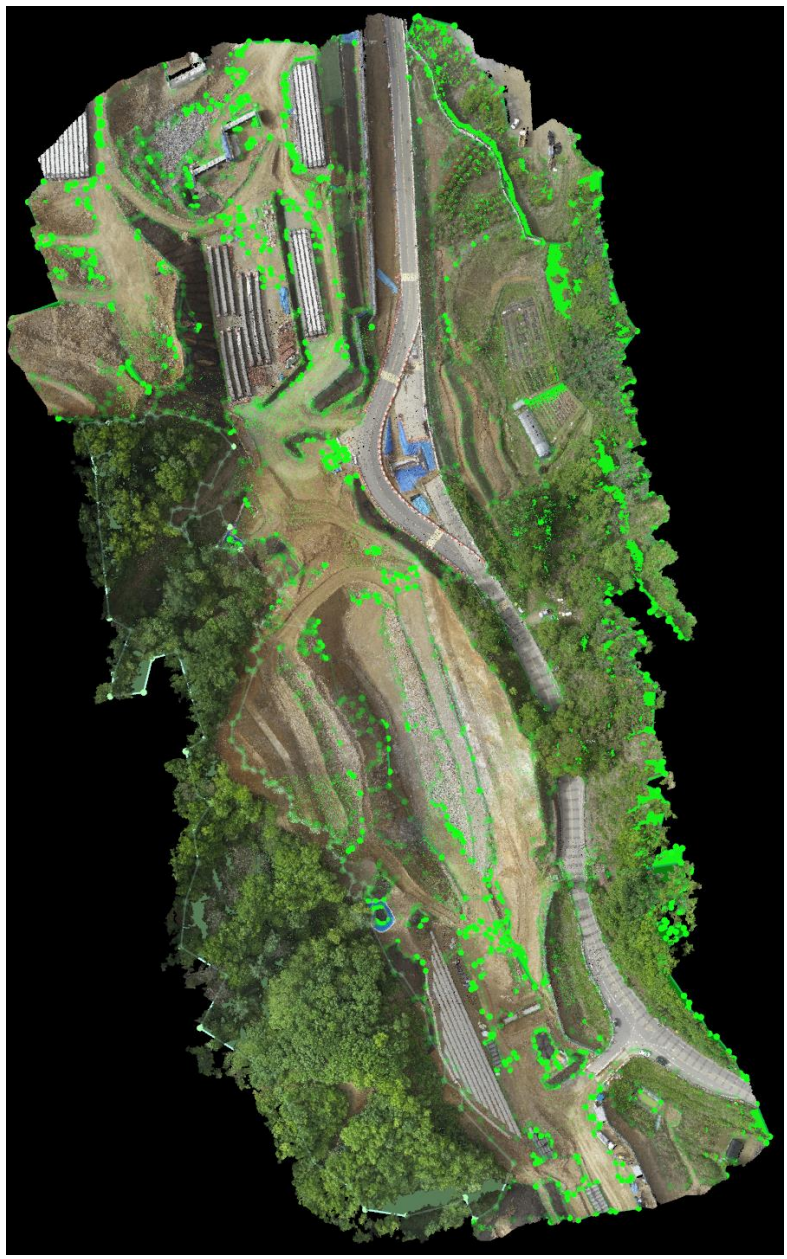


Figure 6.9 Manual boundary setup for area estimation in the manual method

Then, the area of the boundary was calculated based on the Eq. 5.3. The area of all cells existing within the boundary is calculated, and the calculated values for each ground surface type are summed up. Finally, the area for each class is derived. It took about three days to estimate the area for all of the ground surface types on the site by the manual method when performed by one person. The calculated area for each ground surface type is shown in Table 6.2.

Table 6.2 Area calculated by manual method

Class	**Area (m^2)**
soil	2,177
rocks	8,397
trees	24,615
puddles	839
nets	3,491

For comparison of the manual method and the proposed method, various conditions were set equally: (1) the density of point clouds was unified to 20,718,023 points as shown in Figure 6.10a, and (2) the ranges of the point cloud and orthomosaic image were unified as shown in Figure 6.10b.

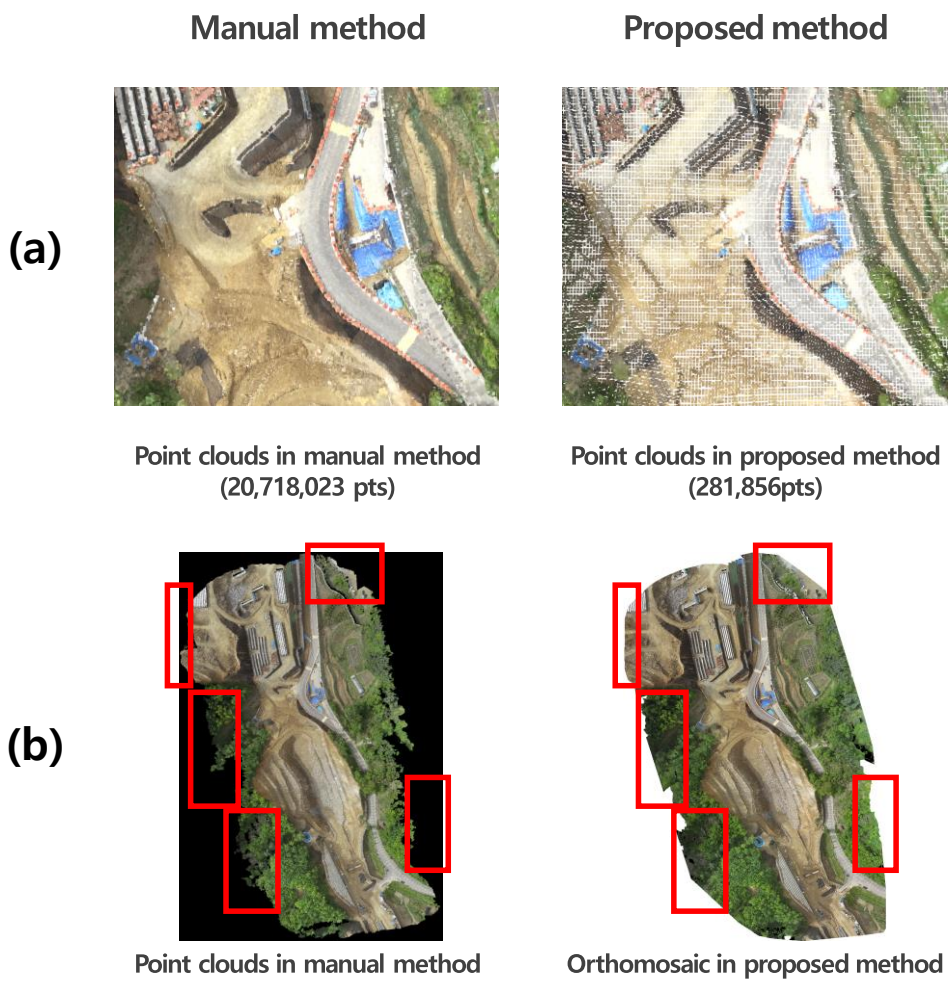


Figure 6.10 Differences between manual method and automated method: (a) Density difference, (b) Range difference

The author analyzed the error of the proposed method by comparing the results with the manual method results. The metrics for error calculation are as described in Eq. 5.4 and Eq. 5.5 (i.e., absolute error and relative error). As a result, the relative error was ‘soil’ 0.19, ‘rocks’ 0.14, ‘trees’ 0.23, ‘puddles’

0.27, ‘nets’ 0.20, and the average relative error was 0.21, as described in Table 6.3.

Table 6.3 Experimental results: area estimation

Class	Area_manual (m^2)	Estimated area (m^2)	Absolute error (m^2)	Relative error
soil	2,177	1,770	406	0.19
rocks	8,397	7,196	1,201	0.14
trees	24,615	18,953	5,661	0.23
puddles	839	615	224	0.27
nets	3,491	2,788	703	0.20
Average	-	-	-	0.21

As shown in Figure 6.11, the validation result (site B) has a slightly lower performance than the test result (Site A.) This performance gap can be interpreted as the difference between the two sites described in figure 6.6 and figure 6.7 (i.e., the difference in visual characteristics for each class and the difference in the number of data distributions for each class.) Since the difference in these characteristics is a characteristic that the model has not been trained, the model performance is lowered to that extent.

Figure 6.11 Area estimation results: test vs. validation

The class with the smallest error value is soil, and the class with the largest error value is puddles, which is the same result as the performance of the classification model. This result is inline with the result from Chapter 5.2. The comparative analysis between the area estimation results and the classification results in both sites are shown in the Figure 6.12. To compare the two models, the performance evaluation index was unified as an error. Accordingly, the evaluation index of the classification model was defined as '1 - f1 score.' The area estimation error is larger than the classification error in all classes, which can be interpreted as the classification error is accumulated in the area estimation model. In most classes, the difference between the area estimation error and the classification error is 0.00 ~ 0.02, but in the case of ‘rocks’, the difference is 0.06, which means that the error difference is larger than that of other classes. The cause of this error can be

interpreted as an error from the calculation method of unsupervised segmentation.

Figure 6.12 Performance comparison: classification errors vs. area estimation errors

Looking at an example in which the area estimation model has an error in 'rocks' is shown in Figure 6.13. A patch labeled with 'rocks' is created by the classification model, which goes through the unsupervised segmentation model to be processed in a segmented patch. Then, the segment that occupies the largest area in the patch (i.e., major color) is extracted by *Argmax,* and the area is calculated for the segment. When deriving the major color with *Argmax*, in the case of 'rocks,' the area occupied by actual 'rocks' in the patch was smaller than that of other segments. Accordingly, there were cases where the major color was extracted without segmenting the actual ‘rocks’, which led to an area calculation error.

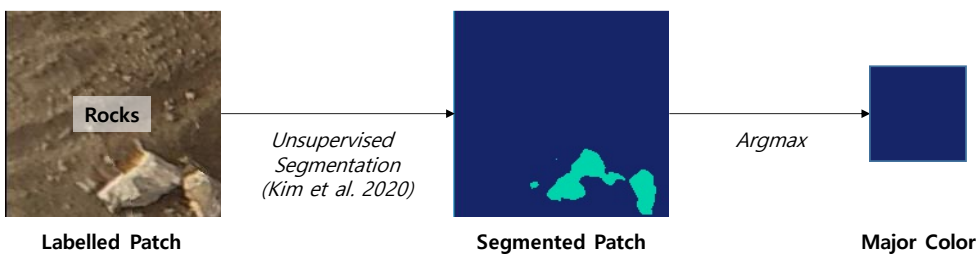


Figure 6.13 An example of major color mis-selection

The 2-D area was then mapped into 3-D by superimposing the segmented patch onto 3-D point clouds in the following process. Point clouds are generated from UAV images utilizing photogrammetry techniques equipped on a commercial tool, Pix4D mapper. Then, coordinates of four

vertices of a pixel is extracted in the segmented patch. The coordinates systems need to be unified into universal transverse mercator (UTM) because the latitude and longitude of patch and point clouds have different coordinate systems: the coordinate system of the patch is UTM, and the coordinate system of point clouds is world geodetic system (WGS84). The equation of the straight-line of each side of the patch is derived based on the latitude/longitude axis of the UTM coordinate system. If each point of the point cloud is included in the four straight-line equations, gives the patch label as the class of the point. The visualization results are shown in Appendix C.

In terms of processing time, the proposed method was overwhelming, as shown in Figure 6.14. The time required to process input UAV images and calculate the area for each type of surface was reduced to about 30% or less compared to the existing manual method. In the proposed method, it took 0.3 to 0.5 hours to generate ground surface type information from orthomosaic image through the classification model, and 6 to 7 hours were consumed to estimate the area based on the surface type information, resulting in a total of 6 ~ 8 hours were consumed. On the other hand, in the manual method, it took about 20 to 22 hours to set the boundary for each ground surface class (i.e., class annotation) from the point cloud, and it took 1 to 2 hours to calculate the unit area from the annotated point clouds. Additionally, it took 1 to 2 hours to add up the calculated unit areas, and finally, 20 to 24 hours to estimate the

area for each type of ground surface. The thing to note here is the human input. While the manual method requires continuous human input for the processing time (i.e., 20 to 24 hours,) the time required for human input among the total processing time in the proposed method is less than 10 minutes for inputting data and executing code.

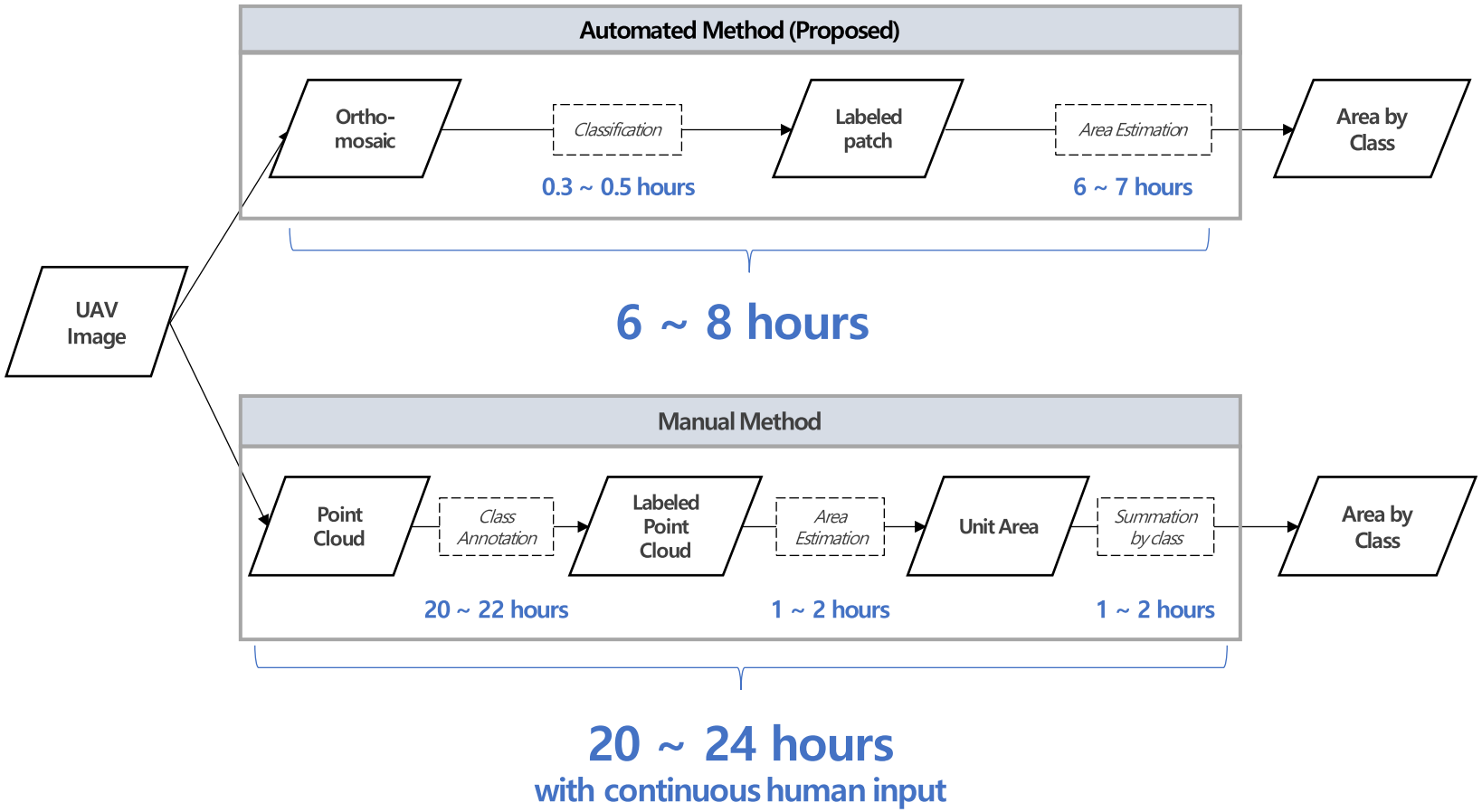


Figure 6.14 Processing time: automated method vs. manual method

6.3. Summary

This chapter validated the proposed methodology and confirmed the technical feasibility and the applicability of this research. The author performed experiments using UAV images collected from an actual road construction site to apply and validate the proposed methodology in real-world environments. The area estimation result was defined as the final output of this research, which was generated by applying the proposed methodology (i.e., from Chapter 3 to Chapter 5 in this dissertation) and compared with the results from the manual method to evaluate the performances of the proposed methodology.

The proposed data augmentation model improved the quality and quantity of UAV images. Based on the augmented data, the proposed classification model successfully classified the ground surface with the average f1 score of 0.81. The proposed area estimation model successfully segmented the labeled patch and estimated the area of the segmented patch with the average relative error 0.15 by ground surface types. The visualization model successfully visualized the results in 3-D to provide the final outputs with the automated equipment.

Chapter 7. Conclusions

This chapter summarizes the research achievements and discusses key findings, practical applications, and future research directions in the field of UAV-based jobsite management for smart earthmoving.

7.1. Summary and Contributions

For automated equipment to safely and accurately perform tasks in the field, it is vital to understand site information. Above all, to achieve effective automation, it is necessary to automatically provide information on which ground surface of the site the equipment needs to work and which ground surface it can access. Thus, it is necessary to check the ground surface (i.e., what kind of, where, and how wide ground surface exist on the jobsite.) However, in practice, the monitoring process is conducted ineffectively. Site managers patrol the site and manually check the ground surface conditions. Since the size of road construction sites is typically enormous, the manual approach requires significant time and workforce. Therefore, there is a need to develop an automated method to classify surface types—an issue not currently addressed in the extant literature. To automatically generate the ground surface information, a detection model that can learn the visual characteristics of the ground surface types encompassing the various sizes, shapes, and colors is required. Moreover, a model is required to distinguish different surface types, such as soil, rocks, trees, and puddles.

To this end, this research proposed a UAV-based ground surface classification and area estimation methodology. First, this research developed ground surface datasets, using the super-resolution-based image augmentation technique. Based on the augmented datasets, this research

utilized deep learning architectures—ResNet and VIT— as the backbone of the ground surface classification module to address the visual characteristics of the ground surface types encompassing the various sizes, shapes, and colors is required. A multi-label classification method was applied to distinguish different surface types, such as soil, rocks, trees, and puddles (average f1 score 0.88 as shown in Figure 7.1.) Then, the unsupervised segmentation model processed the labeled patch to generate a segmented patch for more accurate ground surface information by processing the patch-wise information into pixel-wise information. This integrated method (i.e., classification + unsupervised segmentation) can reduce the time required to construct a dataset by 80 to 240 times, compared to the previous supervised segmentation approaches. The identified ground surface types were further estimated by the area (average relative error 0.15 as shown in Figure 7.1) and superimposed the onto 3-D point clouds to process the results in 3-D by the visualization module, to provide the results with the automated equipment.

The technical feasibility and the applicability of the proposed methodology were validated and confirmed in the actual construction site environments. The author performed experiments using UAV images collected from an actual road construction site to apply and validate the proposed methodology in real-world environments. The proposed data augmentation model improved the quality and quantity of UAV images.

Based on the augmented data, the proposed classification model successfully classified the ground surface with the average f1 score of 0.81, as shown in Figure 7.1. The proposed area estimation model successfully segmented the labeled patch and estimated the area of the segmented patch with the average relative error 0.15 by ground surface types, as shown in Figure 7.1. The visualization model successfully visualized the results in 3-D to provide the final outputs with the automated equipment.

	Classification		Area Estimation	
Test (Site A)	Class	F1 score	Class	Relative error
	soil	0.96	soil	0.07
	rocks	0.94	rocks	0.11
	trees	0.93	trees	0.09
	puddles	0.72	puddles	0.30
	nets	0.84	nets	0.19
	Average	0.88	Average	0.15
Validation (Site B)	Class	F1 score	Class	Relative error
	soil	0.82	soil	0.19
	rocks	0.92	rocks	0.14
	trees	0.79	trees	0.23
	puddles	0.74	puddles	0.27
	nets	0.80	nets	0.20
	Average	0.81	Average	0.21

Figure 7.1 The results of test and validation

The proposed methodology can provide significant contributions to the construction industry in two aspects: academic and practical. As an academic contribution, this research fills the identified research gap by introducing a

novel framework for understanding the ground surface on road construction sites. In contrast to the existing studies that have merely emphasized the importance of managing ground surface, this research defined the ground surface types that need to be addressed, established datasets, and designed an automated classification and area estimation methodology by applying UAV and computer vision techniques. Specifically, the author customized the various SOTA deep neural networks for ground surface datasets, classification, and segmentation for the application to construction environments and provided quantitative results to demonstrate its technical feasibility. Moreover, the expandability and generalizability of the proposed approach were investigated through additional experiments involving actual construction sites. The quantitative results and theoretical findings may play a crucial role as a baseline for future research directions in the field of UAV-based construction site management and construction equipment automation.

In addition, as a practical contribution, the results of this research indicate the model's potential to reduce the cost and time needed for on-site ground surface management for earthwork of road construction by facilitating automated earthmoving equipment and field practitioners to detect any obstacles on construction sites quickly and effectively. For example, the proposed methodology can support the automated equipment and site managers in their decision-making by automatically classifying ground

surface, such as assisting managers to quickly identify the workable area of the site after bad weather, such as rainfall or snowfall, which can affect the ground surface. The equipment also can avoid unfavorable ground surfaces to prevent rollover accidents and identify the workable area according to the ground surface type. Moreover, by adjusting the classification model dataset, other classes in the field can also be addressed, which has scalability in-field application.

7.2. Improvement Opportunities and Future Research

Although the objectives of this research were achieved, there are still improvement opportunities to advance the the performance and applicability of the research findings. The future research is recommended to address the following limitations of this research.

(1) The application scope of the proposed data augmentation method in this research was limited to 2-D images. However, there is an opportunity to extend the scope into 3-D. If the photogrammetry techniques are applied to the super-resolution-applied images, the resolution of point clouds can be augmented. The augmented point clouds can be modeled into rigid objects such as building information models that can be used in various fields.

(2) In the same vein as the idea proposed in this research, methods to augment data without additional data collection are being actively developed in computer vision and computer science communities. For example, synthesizing real and virtual data is a representative method, which can be a great opportunity for improving the performance of UAV vision-based site management.

(3) The scope this research was to generate a ground surface information (i.e., type, location, and shape) and quantify the ground surface information

into the area. Considering that the recipient of this information is automated equipment or site manager, it can be further processed into advanced information. The equipment can establish a work or path planning using the ground surface information. The sites manager can link the ground surface information with the necessary resources (i.e., workers, required time) to treat the ground surface by referring to construction standards.

(4) Ground surface identified in this research were limited to ‘soil,’ ‘rocks,’ ‘trees,’ ‘puddles,’ and ‘nets.' However, other ground surface types can be managed based on UAV images, such as construction materials, equipment, and workers on the site. The proposed methodology can be applied to these object types if additional data annotation is performed for the new types in the dataset development in Chapter 3.

(5) Remarkable algorithms in computer vision communities appear one after another with a short cycle. The deep learning networks utilized in this research (i.e., ResNet, Convolution Neural Networks, Vision Transformer) can be improved with more refined algorithms to apply the various construction sites' features and expand the applicability of the proposed methodology.

Bibliography

Angelswing (2021). Drone data platform for reality capture and accurate survey. https://tinyurl.com/fsr4azem, accessed November 2021.

Aquino, N. R., Gutoski, M., Hattori, L. T., & Lopes, H. S. (2017). The effect of data augmentation on the performance of convolutional neural networks. Braz. Soc. Comput. Intell.

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

Associated General Contractors (AGC) (2020). 2020 Construction Outlook Survey Results National Results. https://tinyurl.com/3u8ew4d9

Azar, E. R., & Kamat, V. R. (2017). Earthmoving equipment automation: A review of technical advances and future outlook. Journal of Information Technology in Construction (ITcon), 22(13), 247-265.

B. N. Delaunay (1934). Sur la sphère vide. Izvestia Akademia Nauk SSSR, 7, VII, 793–800.

Bang, S., & Kim, H. (2020). Context-based information generation for managing UAV-acquired data using image captioning. Automation in Construction, 112, 103116.

Bang, S., Baek, F., Park, S., Kim, W., & Kim, H. (2020). Image augmentation to improve construction resource detection using generative adversarial

networks, cut-and-paste, and image transformation techniques. Automation in Construction, 115, 103198.

Bang, S., Hong, Y., & Kim, H. (2021). Proactive proximity monitoring with instance segmentation and unmanned aerial vehicle-acquired video-frame prediction. Computer-Aided Civil and Infrastructure Engineering, 36(6), 800-816.

Bang, S., Kim, H., & Kim, H. (2017). UAV-based automatic generation of high-resolution panorama at a construction site with a focus on preprocessing for image stitching. Automation in Construction, 84, 70-80.

Bang, S., Kim, H., & Kim, H. (2017a). Vision-based 2-D map generation for monitoring construction sites using UAV Videos. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 34). IAARC Publications.

Bang, S., Kim, H., & Kim, H. (2017b). UAV-based automatic generation of high-resolution panorama at a construction site with a focus on preprocessing for image stitching. Automation in Construction, 84, 70-80.

Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. ACM Transactions on Mathematical Software (TOMS), 22(4), 469-483.

Bello, I., Zoph, B., Vaswani, A., Shlens, J., & Le, Q. V. (2019). Attention augmented convolutional networks. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 3286-3295).

Biçici, S., & Zeybek, M. (2021). An approach for the automated extraction of road surface distress from a UAV-derived point cloud. Automation in Construction, 122, 103475.

Borngrund, C., Sandin, F., & Bodin, U. (2022). Deep-learning-based vision for earth-moving automation. Automation in Construction, 133, 104013.

Brucker, Juricic, B., Galic, M., Marenjak, S. Review of the Construction Labour Demand and Shortages in the EU. Buildings. 2021; 11(1):17. https://doi.org/10.3390/buildings11010017

Cao, J., Li, Y., Zhang, K., & Van Gool, L. (2021). Video super-resolution transformer. arXiv preprint arXiv:2106.06847.

Chae, M. J., Lee, G. W., Kim, J. Y., Park, J. W., & Cho, M. Y. (2011). A 3D surface modeling system for intelligent excavation system. Automation in Construction, 20(7), 808-817.

Chang, F. J., Lin, Y. Y., & Hsu, K. J. (2014). Multiple structured-instance learning for semantic segmentation with uncertain training data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 360-367).

Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., ... & Gao, W. (2021). Pre-trained image processing transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12299-12310).

Chen, J., Fang, Y., & Cho, Y. K. (2016). Automated equipment recognition and classification from scattered point clouds for construction management applications. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 33, p. 1). IAARC Publications.

Cho, S., Kim, D., & Yoon, W. (2020). Introduction on the National Research for Smart Construction Technology. The Magazine of the Korean Society of Civil Engineers, 68(8), 16-28.

Choi, Y., Kim, M., Kim, Y., & Han, S. (2020). A study of CNN-based super-resolution method for remote sensing image. Korean Journal of Remote Sensing, 36(3), 449-460.

D. Lin, J. Dai, J. Jia, K. He, and J. Sun, "ScribbleSup: Scribblesupervised convolutional networks for semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 3159–3167.

D. Pathak, P. Krahenbuhl, and T. Darrell, "Constrained convolutional neural networks for weakly supervised segmentation," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1796–1804.

Dai, J., He, K., & Sun, J. (2016). Instance-aware semantic segmentation via multi-task network cascades. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3150-3158).

Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In European conference on computer vision (pp. 184-199). Springer, Cham.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

DroneDeploy (2021). DroneDeploy for construction, https://tinyurl.com/3yrpsbz6, accessed November 2021.

Fang, W., Ding, L., Zhong, B., Love, P. E., & Luo, H. (2018). Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach. Advanced Engineering Informatics, 37, 139-149.

Gargoum, S. A., & Karsten, L. (2021). Virtual assessment of sight distance limitations using LiDAR technology: Automated obstruction detection and classification. Automation in Construction, 125, 103579.

Guo, Y., Xu, Y., & Li, S. (2020). Dense construction vehicle detection based on orientation-aware feature fusion convolutional neural network. Automation in Construction, 112, 103124.

Ha, Q. P., Yen, L., & Balaguer, C. (2019). Robotic autonomous systems for earthmoving in military applications. Automation in Construction, 107, 102934.

Halbach, E., & Halme, A. (2013). Job planning and supervisory control for automated earthmoving using 3D graphical tools. Automation in Construction, 32, 145-160.

Ham, Y., Han, K. K., Lin, J. J., & Golparvar-Fard, M. (2016). Visual monitoring of civil infrastructure systems via camera-equipped Unmanned Aerial Vehicles (UAVs): a review of related works. Visualization in Engineering, 4(1), 1-8.

Han, J., Zhang, D., Cheng, G., Guo, L., & Ren, J. (2015). Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. IEEE Transactions on Geoscience and Remote Sensing, 53 (6), 3325–3337.

Hariharan, B., Arbeláez, P., Girshick, R., & Malik, J. (2014, September). Simultaneous detection and segmentation. In European conference on computer vision (pp. 297-312). Springer, Cham.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. In European conference on computer vision (pp. 630-645). Springer, Cham.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Hu, H., Zhang, Z., Xie, Z., & Lin, S. (2019). Local relation networks for image recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3464-3473).

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).

Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., & Liu, W. (2019). Ccnet: Criss-cross attention for semantic segmentation. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 603-612).

I. Croitoru, S.-V. Bogolin, and M. Leordeanu, “Unsupervised learning of foreground object segmentation,” Int. J. Comput. Vis., vol. 127, no. 9, pp. 1279–1302, Sep. 2019.

Isailović, D., Stojanovic, V., Trapp, M., Richter, R., Hajdin, R., & Döllner, J. (2020). Bridge damage: Detection, IFC-based semantic enrichment and visualization. Automation in Construction, 112, 103088.

J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 3431–3440.

J. MacQueen et al., “Some methods for classification and analysis of multivariate observations,” in Proc. 5th Berkeley Symp. Math. Statist. Probab., vol. 1, Oakland, CA, USA, 1967, pp. 105–117.

Jeong, U., Bang, H., & Kim, O. (2018). Supply and Demand of Construction Skilled labor and Analysis of the Caus. Journal of the Institute of Construction Technology, 37(2), 1-5.

Jiang, W., Zhou, Y., Ding, L., Zhou, C., & Ning, X. (2020). UAV-based 3-D reconstruction for hoist site mapping and layout planning in petrochemical construction. Automation in Construction, 113, 103137.

Jiang, Y., Bai, Y., & Han, S. (2020). Determining ground elevations covered by vegetation on construction sites using drone-based orthoimage and

convolutional neural network. Journal of Computing in Civil Engineering, 34(6), 04020049.

Kim, D., Lee, S., & Kamat, V. R. (2020). Proximity prediction of mobile objects to prevent contact-driven accidents in co-robotic construction. Journal of Computing in Civil Engineering, 34(4), 04020022.

Kim, D., Liu, M., Lee, S., & Kamat, V. R. (2019). Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. Automation in Construction, 99, 168-182.

Kim, D.Y. (2014). A Basic Study on Investigation of Current Craftmen Status in Korea Construction Site. Architectural Institute of Korea, 30(11), 81-88.

Kim, H., Kim, H., Hong, Y. W., & Byun, H. (2018). Detecting construction equipment using a region-based fully convolutional network and transfer learning. Journal of Computing in Civil Engineering, 32(2), 04017082.

Kim, J., & Chi, S. (2020). Multi-camera vision-based productivity monitoring of earthmoving operations. Automation in Construction, 112, 103121.

Kim, J., & Chi, S. (2021). A few-shot learning approach for database-free vision-based monitoring on construction sites. Automation in Construction, 124, 103566.

Kim, J., Chi, S., & Hwang, B.-G. (2017). Vision-based activity analysis framework considering interactive operation of construction equipment.

ASCE International Workshop on Computing in Civil Engineering 2017. American Society of Civil Engineers, Reston, VA, 162–170. https://doi.org/10.1061/9780784480830.021.

Kim, J., Chi, S., & Kwon, T. (2016). Construction entities tracking based on functional integration and online learning with site-customized datasets. Proceedings of the CIB World Building Congress 2016, Tampere, Finland, 1118–1128.

Kim, J., Ham, Y., Chung, Y., & Chi, S. (2018). Camera placement optimization for vision-based monitoring on construction sites. 2018 Proceedings of the 35th International Symposium on Automation and Robotics in Construction, International Association for Automation and Robotics in Construction, Berlin, Germany, 748–752.

Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1646-1654).

Kim, J., Lee, S., Seo, J., Lee, D. E., & Choi, H. S. (2021). The Integration of Earthwork Design Review and Planning Using UAV-Based Point Cloud and BIM. Applied Sciences, 11(8), 3435.

Kim, K., Kim, S., & Shchur, D. (2021). A UAS-based work zone safety monitoring system by integrating internal traffic control plan (ITCP) and

automated object detection in game engine environment. Automation in Construction, 128, 103736.

Kim, S. K., Seo, J., & Russell, J. S. (2012). Intelligent navigation strategies for an automated earthwork system. Automation in Construction, 21, 132-147.

Kim, W., Kanezaki, A., & Tanaka, M. (2020). Unsupervised learning of image segmentation based on differentiable feature clustering. IEEE Transactions on Image Processing, 29, 8055-8068.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. IEEE Geoscience and Remote Sensing Letters, 14 (5), 778–782.

L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015, pp. 1–4.

Le, Q. V. (2013, May). Building high-level features using large scale unsupervised learning. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 8595-8598). IEEE.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

Lee, G. H., & Kang, M. S. (2009). Feasible plan study for pothole decrease of asphalt concrete pavement. Journal of the Korean Society of Civil Engineers, 57(12), 72-77.

Lee, G. H., Kang, M. S., & Jo, M. J. (2012). Asphalt pavement damage and reduction measures. Korean Society of Road Engineers, 14(3), 5-11.

Lee, H., Pham, P., Largman, Y., & Ng, A. (2009). Unsupervised feature learning for audio classification using convolutional deep belief networks. Advances in neural information processing systems, 22.

Lei, J., Zhang, S., Luo, L., Xiao, J., & Wang, H. (2018). Super-resolution enhancement of UAV images based on fractional calculus and POCS. Geo-spatial information science, 21(1), 56-66.

Leica (2021). Leica Aibot CX for Construction, uav.leica-geosystems.com, accessed November 2021.

Leonida, C. (2020). Giving Operators a Helping Hand. Engineering and Mining Journal, 221(2), 44-46.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 1833-1844.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1833-1844).

Lin, D., Dai, J., Jia, J., He, K., & Sun, J. (2016). Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3159-3167).

Lin, J. J., Ibrahim, A., Sarwade, S., & Golparvar-Fard, M. (2021). Bridge Inspection with Aerial Robots: Automating the Entire Pipeline of Visual Data Capture, 3-D Mapping, Defect Detection, Analysis, and Reporting. Journal of Computing in Civil Engineering, 35(2), 04020064.

Lin, J. J., Ibrahim, A., Sarwade, S., & Golparvar-Fard, M. (2021). Bridge Inspection with Aerial Robots: Automating the Entire Pipeline of Visual Data Capture, 3D Mapping, Defect Detection, Analysis, and Reporting. Journal of Computing in Civil Engineering, 35(2), 04020064.

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

Luo, X., Li, H., Wang, H., Wu, Z., Dai, F., & Cao, D. (2019). Vision-based detection and visualization of dynamic workspaces. Automation in Construction, 104, 1–13. https://doi.org/10.1016/j.autcon.2019.04.001.

Ma, Y., Easa, S., Cheng, J., & Yu, B. (2021). Automatic Framework for Detecting Obstacles Restricting 3D Highway Sight Distance Using Mobile Laser Scanning Data. Journal of Computing in Civil Engineering, 35(4), 04021008.

Ma, Y., Zheng, Y., Easa, S., Wong, Y. D., & El-Basyouny, K. (2022). Virtual analysis of urban road visibility using mobile laser scanning data and deep learning. Automation in Construction, 133, 104014.

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2016). Fully convolutional neural networks for remote sensing image classification. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 5071–5074. New York: IEEE.

Meissa (2021). Meissa smart construction platform. https://www.meissa.ai/en/, accessed November 2021.

Ministry of Land, Infrastructure and Transport (2016a). Road Construction Standard Specification.

Ministry of Land, Infrastructure and Transport (2016b). Road Design Standards.

Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain. IEEE Transactions on image processing, 21(12), 4695-4708.

Moon, D., Chung, S., Kwon, S., Seo, J., & Shin, J. (2019). Comparison and utilization of point cloud generated from photogrammetry and laser scanning: 3-D world model for smart heavy equipment planning. Automation in Construction, 98, 322-331.

N. Pourian, S. Karthikeyan, and B. S. Manjunath, "Weakly supervised graph based semantic segmentation by learning communities of imageparts," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1359–1367.

Naghshbandi, S. N., Varga, L., & Hu, Y. (2021). Technologies for safe and resilient earthmoving operations: A systematic literature review. Automation in Construction, 125, 103632.

Nath, N.D., Behzadan, A.H., & Paal, S.G. (2020). Deep learning for site safety: real-time detection of personal protective equipment. Automation in Construction, 112, 103085. https://doi.org/10.1016/j.autcon.2020.103085.

National Geographic Information Institute (NGII) (2014). General Survey Rules.

Newcomer, C., Withrow, J., Sturgill, R. E., & Dadi, G. B. (2019). Towards an automated asphalt paving construction inspection operation. In Advances in Informatics and Computing in Civil and Construction Engineering, 593-600. Springer, Cham.

Nguyen, H. A., & Ha, Q. P. (2022). Robotic autonomous systems for earthmoving equipment operating in volatile conditions and teaming capacity: a survey. Robotica, 1-25.

Omar, T., & Nehdi, M. L. (2017). Remote sensing of concrete bridge decks using unmanned aerial vehicle infrared thermography. Automation in Construction, 83, 360-371.

Oregon Department of Transportation (2015). Survey policy and procedure manual.

P. F. Felzenszwalb and D. P. Huttenlocher, “Efficient graph-based image segmentation,” Int. J. Comput. Vis., vol. 59, no. 2, pp. 167–181, Sep. 2004.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621.

Pi, Y., Nath, N.D., & Behzadan, A.H. (2020). Convolutional neural networks for object detection in aerial imagery for disaster response and recovery.

Advanced Engineering Informatics, 43, 101009. https://doi.org/10.1016/j.aei.2019.101009.

Q. Yan, Y. Xu, X. Yang, T.Q. Nguyen, IEEE Transactions on Image Processing 24 (2015) 3187–3202.

R. Fernandez-Beltran, P. Latorre-Carmona, F. Pla, International Journal of Remote Sensing 38 (2017) 314–354.

R.G. Keys, IEEE Transactions on Acoustics, Speech, and Signal Processing 29 (1981) 1153–1160.

Ramachandran, P., Parmar, N., Vaswani, A., Bello, I., Levskaya, A., & Shlens, J. (2019). Stand-alone self-attention in vision models. Advances in Neural Information Processing Systems, 32.

Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2011). Classifier chains for multi-label classification. Machine Learning, 85(3), 333-359.

Real, E., Aggarwal, A., Huang, Y., & Le, Q. V. (2019). Regularized evolution for image classifier architecture search. In Proceedings of the aaai conference on artificial intelligence (Vol. 33, No. 01, pp. 4780-4789).

S. Zheng et al., “Conditional random fields as recurrent neural networks,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1527–1537

Seo, J., Duque, L., & Wacker, J. (2018). Drone-enabled bridge inspection methodology and application. Automation in Construction, 94, 112-126.

Seo, J., Lee, S., Kim, J., & Kim, S. K. (2011). Task planner design for an automated excavation system. Automation in Construction, 20(7), 954-966.

Sharma, O. (2019, February). A new activation function for deep neural network. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 84-86. IEEE.

Shim, S., Kim, J., Lee, S. W., & Cho, G. C. (2022). Road damage detection using super-resolution and semi-supervised learning with generative adversarial network. Automation in Construction, 135, 104139.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 1-48.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep fisher networks for large-scale image classification. Advances in neural information processing systems, 26.

Sorower, M. S. (2010). A literature survey on algorithms for multi-label learning. Oregon State University, Corvallis, 18, 1-25.

Spolaôr, N., Cherman, E. A., Monard, M. C., & Lee, H. D. (2013). A comparison of multi-label feature selection methods using the problem transformation approach. Electronic Notes in Theoretical Computer Science, 292, 135-151.

Sung, C., & Kim, P. Y. (2016). 3D terrain reconstruction of construction sites using a stereo camera. Automation in Construction, 64, 65-77.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

Tajeen, H., & Zhu, Z. (2014). Image dataset development for measuring construction equipment recognition performance. Automation in Construction, 48, 1-10.

Tang, M., Djelouah, A., Perazzi, F., Boykov, Y., & Schroers, C. (2018a). Normalized cut loss for weakly-supervised cnn segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1818-1827).

Tang, M., Perazzi, F., Djelouah, A., Ben Ayed, I., Schroers, C., & Boykov, Y. (2018b). On regularized losses for weakly-supervised cnn segmentation. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 507-522).

Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J. M., Ciompi, F., & Van Der Laak, J. (2019). Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Medical image analysis, 58, 101544.

Tighe, J., & Lazebnik, S. (2013). Finding things: Image parsing with regions and per-exemplar detectors. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3001-3008).

Torres, H. N., Ruiz, J. M., Chang, G. K., Anderson, J. L., & Garber, S. I. (2018). Automation in highway construction part I: Implementation challenges at state transportation departments and success stories (No. FHWA-HRT-16-030). United States. Federal Highway Administration. Office of Infrastructure Research and Development.

Trimble Inc. (2022). Trimble WorksOS Datasheet. https://tinyurl.com/36n627xx

Tsoumakas, G., Katakis, I., & Vlahavas, I. (2009). Mining multi-label data. In Data mining and knowledge discovery handbook, 667-685. Springer, Boston, MA.

U.S. Department of Transportation Federal Highway Administration (USDOT) (2014). Standard Specifications for Construction of Roads and Bridges on Federal Highway Projects (FP-14).

United States Federal Highway Administration (FWHA) (2021). Determination of Improved Pavement Smoothness When Using 3D Modeling and Automatic Machine Guidance. https://tinyurl.com/2kbtucj2

V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

Vassilo, K. T. (2020). Single image Super Resolution with Infrared Imagery and Multi Step Reinforcement Learning. University of Dayton Dayton United States.

W. Shimoda and K. Yanai, “Distinct class-specific saliency maps for weakly supervised semantic segmentation,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2016, pp. 218–234.

Wang, L., Chen, F., & Yin, H. (2016). Detecting and tracking vehicles in traffic by unmanned aerial vehicles. Automation in Construction, 72, 294-308.

Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7794-7803).

Wortsman, M., Ilharco, G., Gadre, S. Y., Roelofs, R., Gontijo-Lopes, R., Morcos, A. S., ... & Schmidt, L. (2022). Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. arXiv preprint arXiv:2203.05482.

Wu, B., Zhou, X., Zhao, S., Yue, X., & Keutzer, K. (2019). Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In 2019 International Conference on Robotics and Automation (ICRA) (pp. 4376-4382). IEEE.

Wu, R., Yan, S., Shan, Y., Dang, Q., & Sun, G. (2015). Deep image: Scaling up image recognition. arXiv preprint arXiv:1501.02876, 7(8).

X. Liu, Q. Xu, J. Ma, H. Jin, and Y. Zhang, "MsLRR: A unified multiscale low-rank representation for image segmentation," IEEE Trans. Image Process., vol. 23, no. 5, pp. 2159–2167, May 2014.

X. Xia and B. Kulis, "W-net: A deep model for fully unsupervised image segmentation," 2017, arXiv:1711.08506. [Online]. Available: http://arxiv.org/abs/1711.08506

Xuehui, A., Li, Z., Zuguang, L., Chengzhi, W., Pengfei, L., & Zhiwei, L. (2021). Dataset and benchmark for detecting moving objects in construction sites. Automation in Construction, 122, 103482.

Yamamoto, H., Moteki, M., Shao, H., Ootuki, T., Kanazawa, H., & Tanaka, Y. (2009, June). Basic technology toward autonomous hydraulic excavator. In 26th International Symposium on Automation and Robotics in Construction (ISARC 2009) (pp. 288-295).

Z. Shi, Y. Yang, T. M. Hospedales, and T. Xiang, “Weakly-supervised image annotation and segmentation with objects and attributes,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2525–2538, Dec. 2017.

Zhang, M. L., Li, Y. K., Liu, X. Y., & Geng, X. (2018). Binary relevance for multi-label learning: an overview. Frontiers of Computer Science, 12(2), 191-202.

Zhao, H., Jia, J., & Koltun, V. (2020). Exploring self-attention for image recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10076-10085).

Zhu, J., Mao, J., & Yuille, A. L. (2014). Learning from weakly supervised data by the expectation loss svm (e-svm) algorithm. Advances in neural information processing systems, 27.

Zhu, J., Zhong, J., Ma, T., Huang, X., Zhang, W., & Zhou, Y. (2022). Pavement distress detection using convolutional neural networks with images captured via UAV. Automation in Construction, 133, 103991.

Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8697-8710).

국문 초록

스마트 토공을 위한 UAV 이미지 기반 지표면 분류 및 면적 추정

원대연

서울대학교 대학원

건설환경공학부

글로벌 건설 산업은 건설 인력 부족과 낮은 생산성으로 어려움을 겪고 있다. 이를 해결하기 위해 산업 전반에 걸쳐 건설 프로세스를 자동화하려는 움직임이 많이 나타나고 있다. 특히, 건설 선진국에서는 각종 장비 자동화 시스템을 도입하고 있고, 산·학계에서는 건설 장비가 많이 활용되는 토공사에 초점을

맞추어 자동화 기술을 개발하고 상용화하고 있다. 특히, 토공 장비의 제어와 가이던스 관련 기술을 개발하고 현장 실증을 수행하고 있다.

현장에서 자동화 장비가 작업을 안전하고 정확하게 수행하려면 현장 정보를 정확하게 이해하는 게 중요하다. 이를 위해 기존 연구자들은 레이저스캐너와 UAV 를 활용하여 현장을 3 차원 모델링하거나 현장에 존재하는 객체들을 (e.g., 장비, 작업자, 자재) 자동으로 인식하고 모니터링 하는 연구를 수행하였다. 무엇보다도 실제 토공 장비의 자동화를 이루기 위해서는 장비가 현장의 지표면 유형을 잘 파악하는 것이 필요한데, 이는 현장의 지표면 유형에 따라 장비가 수행해야하는 작업이 달라지고 장비의 안전한 운행을 위해서도 필수적이기 때문이다.

가령, 토공사의 예산 비중이 높은 도로공사의 경우, 설계된 구역에 맞추어 현장의 지표면을 관리하는 것이 중요한데 이때 지표면 유형에 따라 필요한 작업이 달라진다. 더불어, 자동화 장비 운행 시 식생, 암석, 웅덩이 등 현장의 지표면 유형에 따라 장비가 접근이 가능한 구역이 달라진다. 이처럼, 토공 작업의 자동화를

위해서는 자동화 장비가 현장의 어떤 지표면에 작업이 필요한지, 어떤 지표면에 접근이 가능한지 등에 대한 지표면 정보 (지표면 유형, 위치, 면적 등)를 자동으로 분석하여 제공하는 것이 필요한데 현재 이에 관한 연구는 부족하다.

이를 위해, 본 연구는 UAV 와 컴퓨터 비전 기술을 활용하여 도로건설 현장의 지표면을 자동으로 분류하고 면적을 산출하는 방법론을 제안한다. 먼저, 지표면 데이터셋을 구축하기 위해 초해상도 기반 데이터 증강 기법을 제안한다. 구축된 데이터셋을 바탕으로 지표면 유형을 자동으로 분류하기 위한 딥러닝 기반의 다중 레이블 분류 방법을 제안한다. 더 나아가, 자동화 장비에게 지표면 유형별 면적 정보를 제공하기 위해 분류된 지표면을 픽셀 단위로 분할하여 면적을 추정하고 그 결과를 포인트 클라우드에 맵핑하여 3 차원으로 시각화하는 방법을 제안한다.

제안한 지표면 데이터셋 구축 방법을 적용한 결과, 제안한 방법을 적용하지 않은 모델에 비해 평균 f1 score 0.16 의 분류 성능 향상을 보였다. 구축된 데이터셋으로 학습된 지표면 분류 모델은 평균 성능 f1 score 0.88 로 현장의 지표면을 토질, 암석, 식생,

웅덩이, 보호망으로 분류하였다. 더 나아가, 제안한 면적 추정 방법을 적용하여 분류된 지표면 격자를 픽셀 단위로 분할하고 면적의 형태로 정량화 하였다. 그 결과, 평균 상대오차 0.15 의 성능으로 지표면 유형별 면적을 산출하였다. 최종적으로 면적 산출 결과를 포인트 클라우드에 맵핑하여 3 차원의 형태로 정보를 시각화 하였다.

제안한 방법론을 검증하기 위해 실제 도로 건설 현장에서 수집한 UAV 이미지를 이용하여 실험을 수행하였다. 그 결과, 제안한 방법이 건설기계 자동화 운영을 위한 주요 현장 정보인 지표면 유형 (평균 f1 score 0.81) 및 지표면 면적 (평균 상대오차 0.21)을 기존 방식보다 효율적으로 생성할 수 있었다. UAV 이미지로부터 지표면 유형별 면적을 산출하는 데까지 소요되는 시간을 기존 상용 툴을 활용한 방법 보다 약 30% 이하로 단축하였다. 또한, 본 연구에서 고찰한 현장의 지표면 특성은 건설자동화장비의 운용을 위한 주요 지표면 정보를 자동으로 생성하는 데 중요한 역할을 하는 것을 확인하였다.

결론적으로 본 연구에서는 UAV 와 컴퓨터 비전 기술을 활용하여 도로 건설 현장에서 지표면 모니터링을 자동화하는 방법론을 제안하였다. 본 연구의 최종 결과물을 통해 도로건설 현장의 지표면 정보를 이해하는 데 필요한 시간과 인력을 크게 줄일 수 있고, 이는 자동화 장비가 현장에서 안전하고 효과적으로 작업하기 위한 필수적인 지표면 정보를 효율적으로 제공할 수 있다. 특히, 광활한 도로공사 현장의 지표면은 작업의 진행, 악천후 등에 따라 그 상태가 수시로 변하는 특성이 있어 이를 매번 수작업으로 파악하기 힘든데, 제안한 방법론을 통해 현장 관리자 및 자동화 장비가 건설 현장의 상황에 따라 수시로 변하는 지표면 상태를 자동으로 파악할 수 있다. 가령, 자동화 장비가 상태가 좋지 않은 지표면을 피할 수 있게 하여 전복 사고를 예방할 수 있고, 지표면 유형에 따라 작업이 필요한 구역을 효율적으로 파악할 수 있다. 본 연구는 현장의 지표면에 대한 심층적인 이해를 도모하여 스마트 건설기술 개발을 위한 지식체계에 기여할 수 있다.

주요어: 건설관리; 스마트 토공; 지표면 정보; 무인 항공기; 컴퓨터 비전

학 번: 2018-31863

Appendix

Appendix A. Hyper-parameter Tuning Results for Unsupervised Segmentation

Appendix B. 3-D Visualization Results – Chapter 5

Appendix C. 3-D Visualization Results – Chapter 6

A. Hyper-parameter Tuning Results for Unsupervised Segmentation

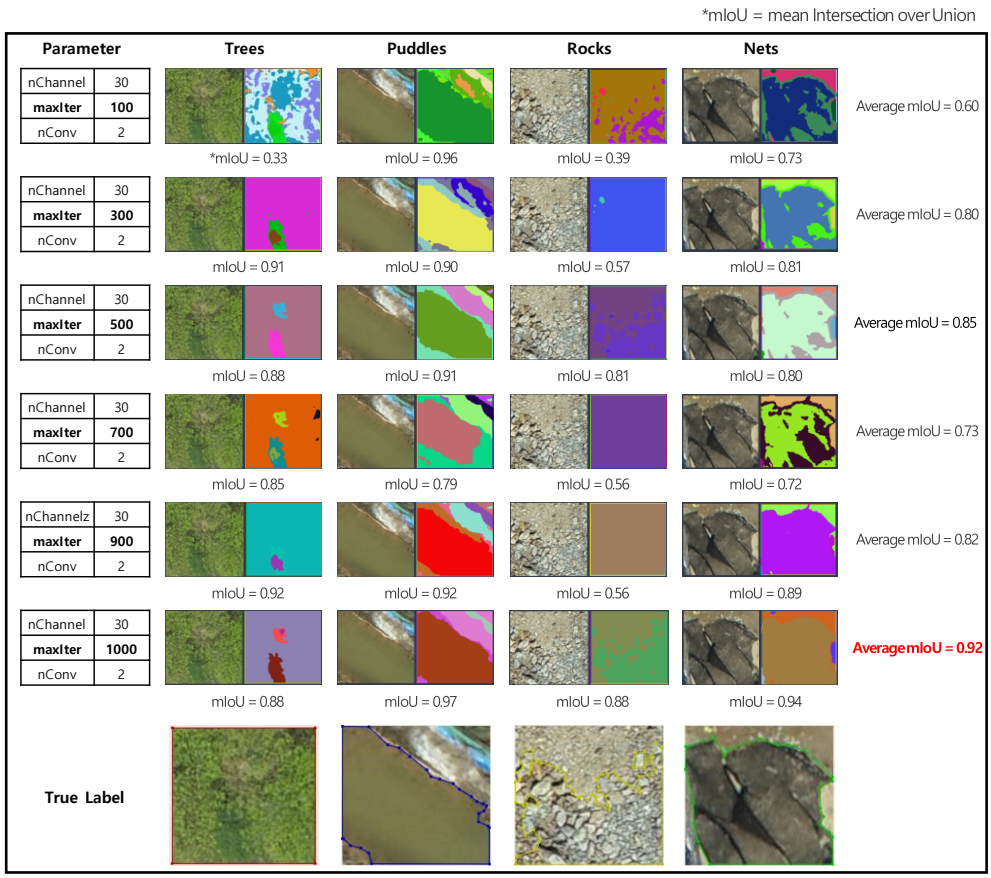


Figure A.1 Hyper-parameter tuning results for unsupervised segmentation: iteration

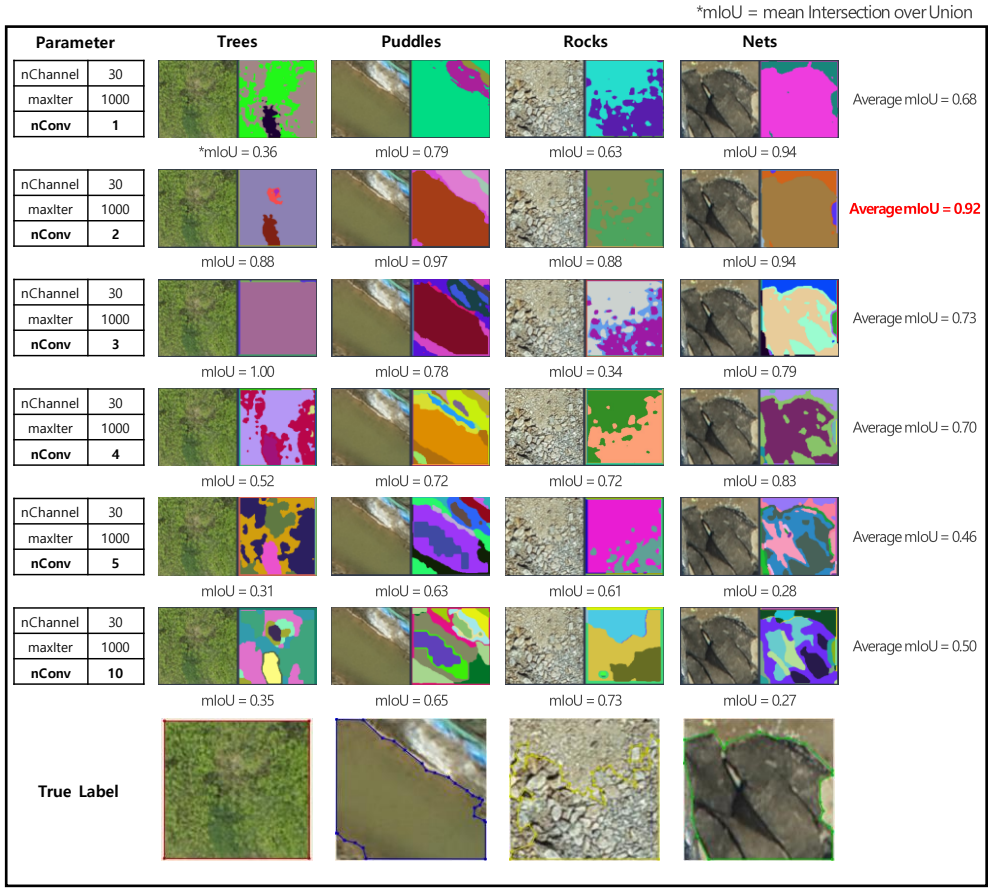


Figure A.2 Hyper-parameter tuning results for unsupervised segmentation: the number of convolution layers

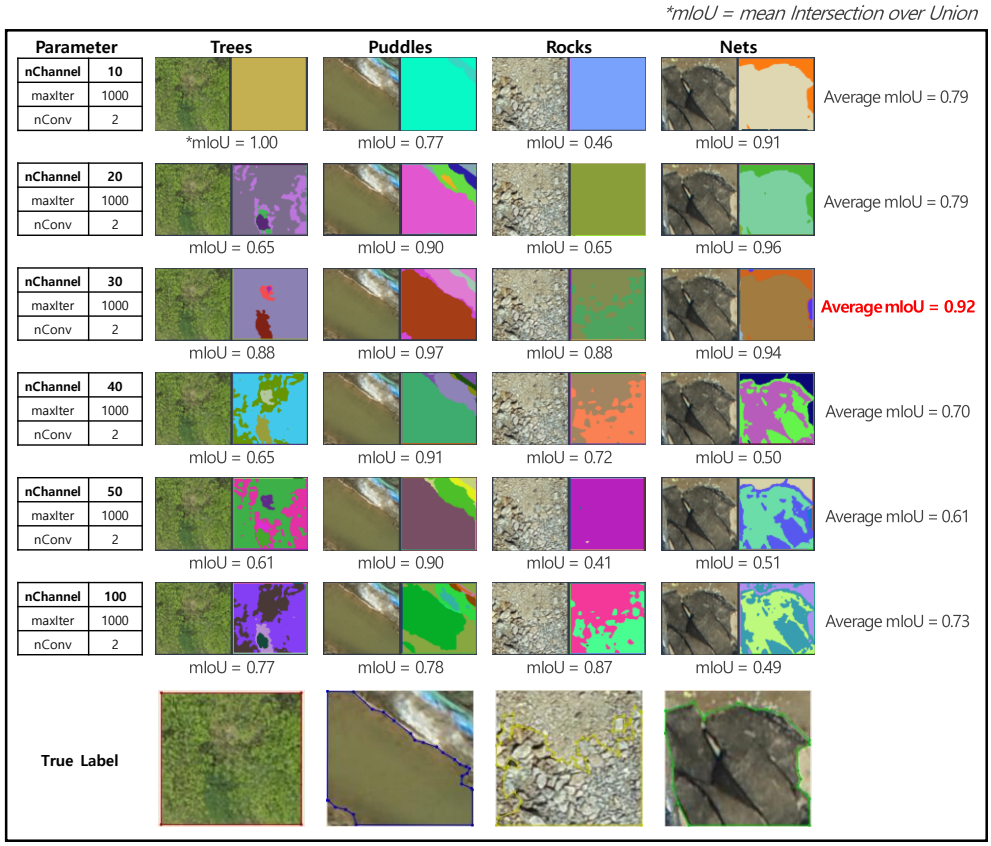
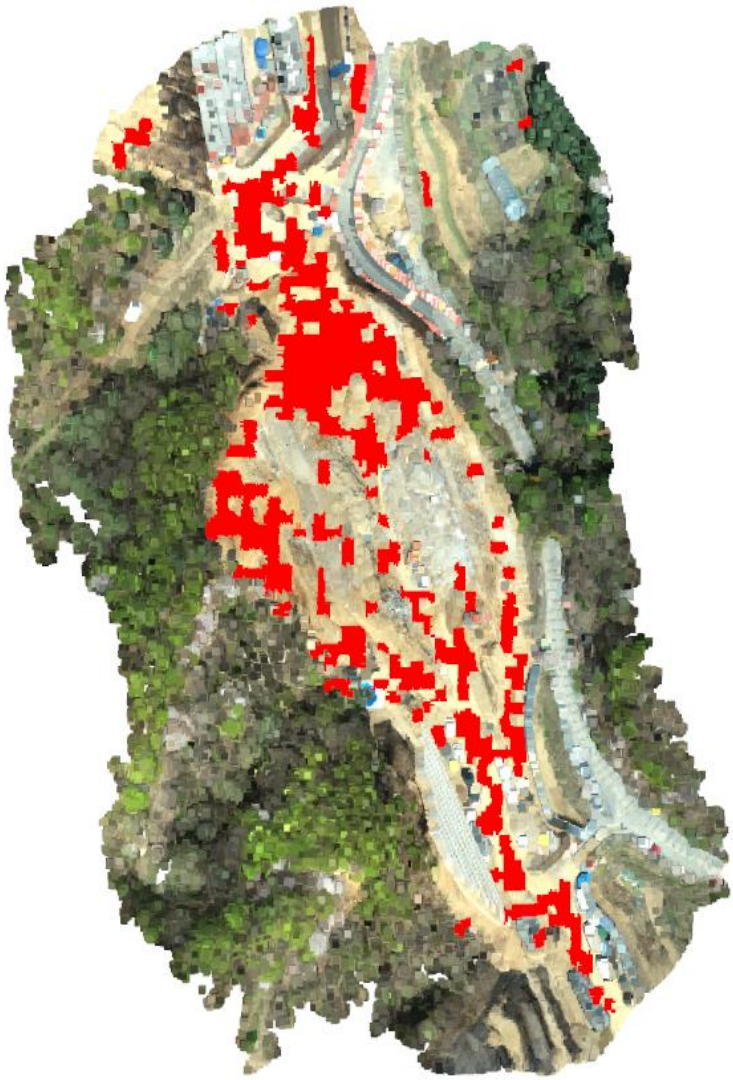


Figure A.3 Hyper-parameter tuning results for unsupervised segmentation: the number of channels

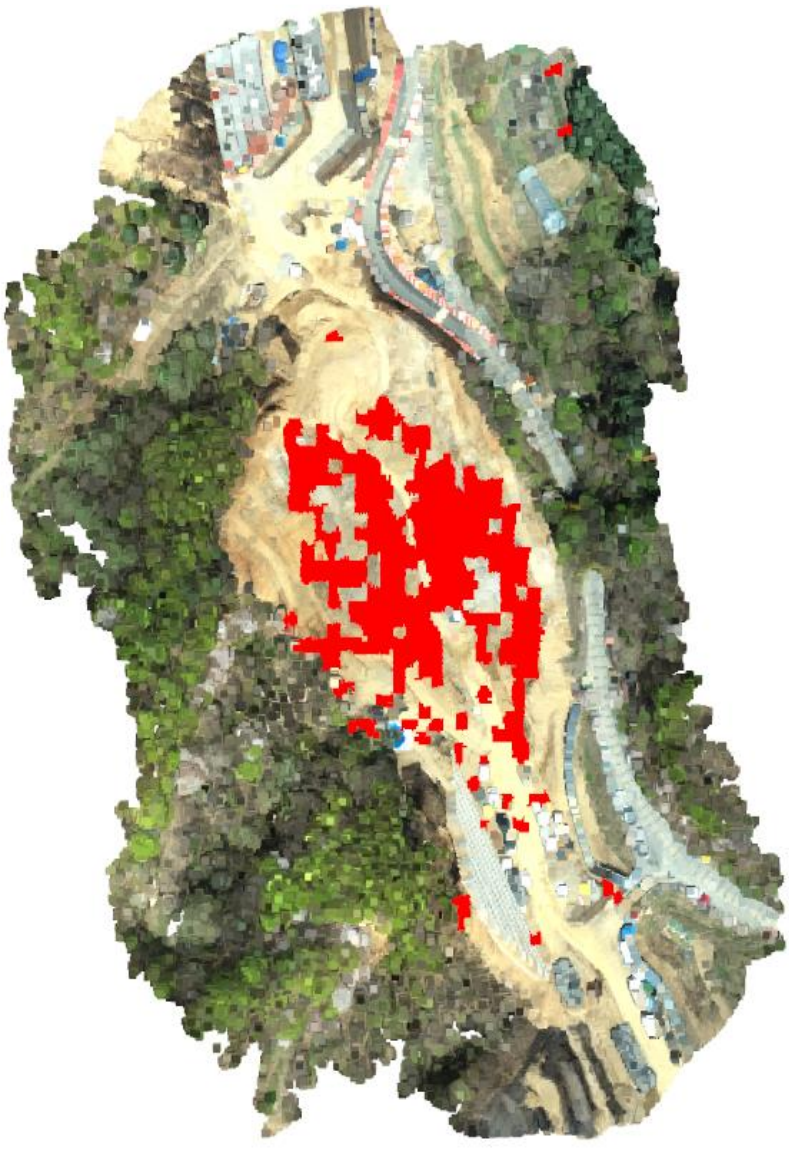
B. 3-D Visualization Results – Chapter 5



Area = 2,970m^2
*Area Ratio = 10.2%
(Relative error 7%)

*Area ratio: the ratio of the area occupied by the class to the total area of the site

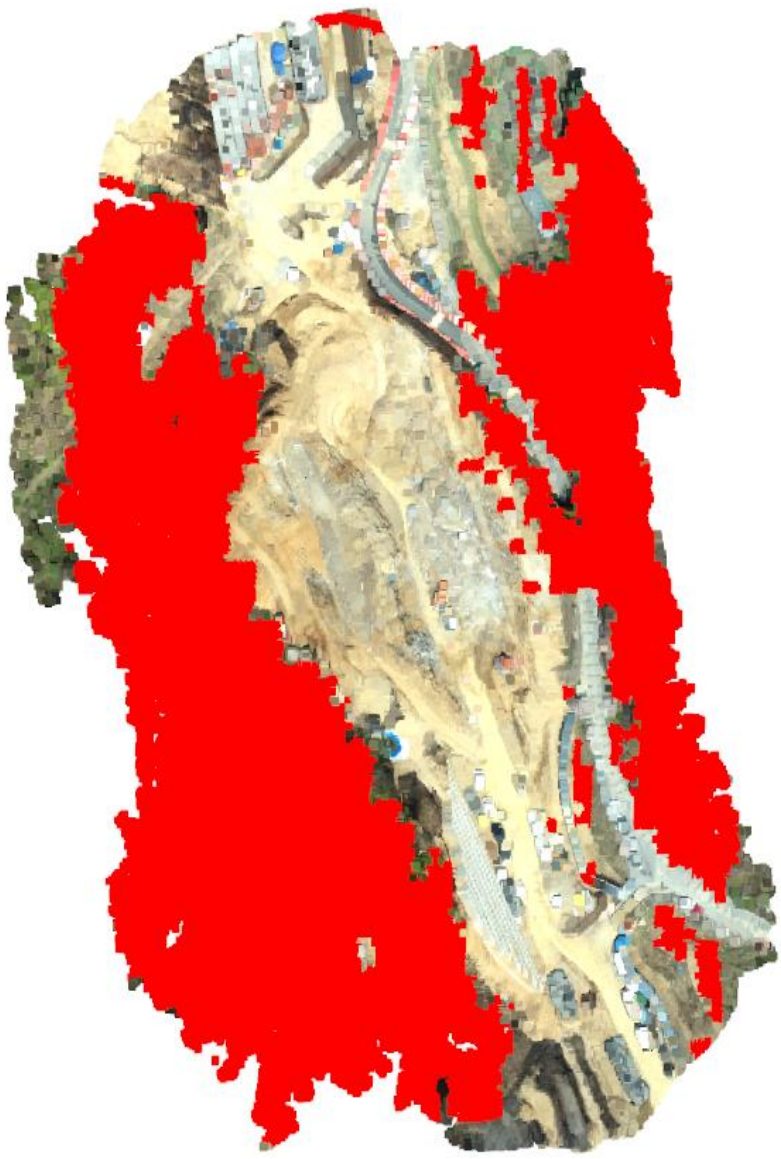
Figure B.1 Mapping and area estimation results: ‘soil’



Area = 2,739m2
*Area Ratio = 9.4%
(Relative error 11%)

*Area ratio: the ratio of the area occupied by the class to the total area of the site

Figure B.2 Mapping and area estimation results: ‘rocks’



Area = 14,543m2
*Area Ratio = 49.9%
(Relative error 9%)

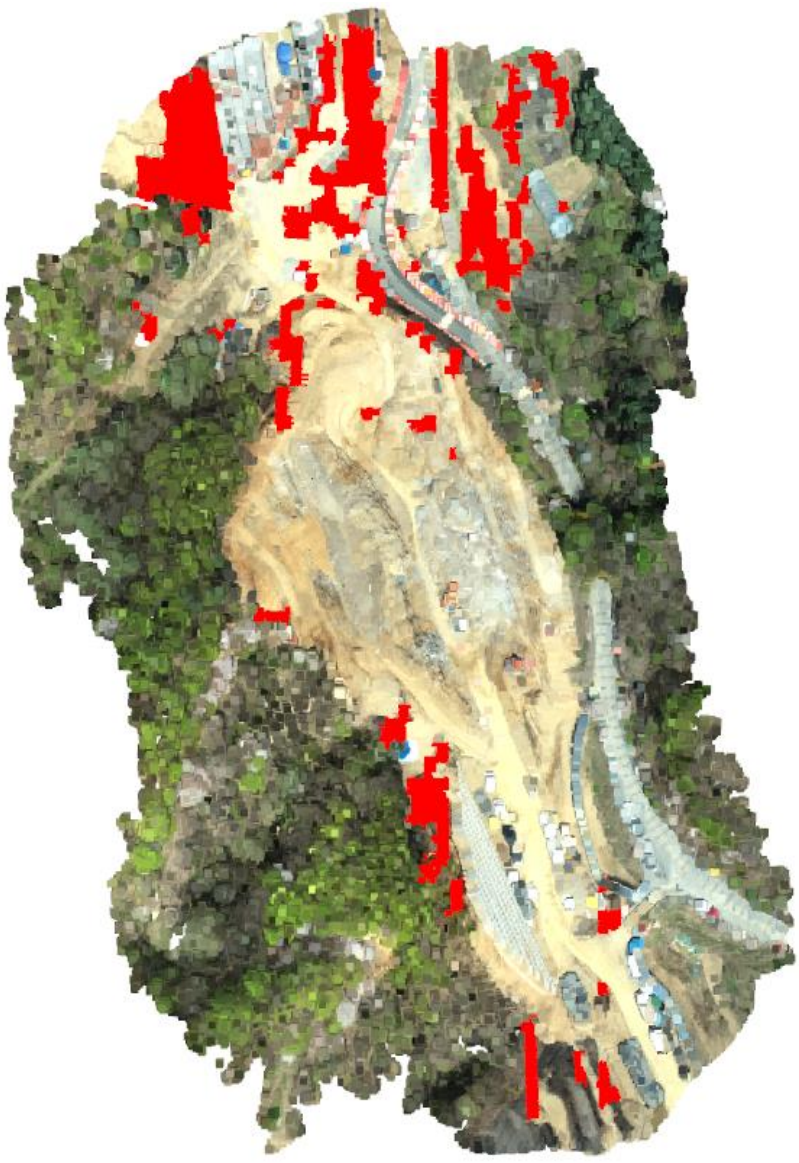
*Area ratio: the ratio of the area occupied by the class to the total area of the site

Figure B.3 Mapping and area estimation results: ‘trees’

Area = 73m2
*Area Ratio = 0.3%
(Relative error 30%)

*Area ratio: the ratio of the area occupied by the class to the total area of the site

Figure B.4 Mapping and area estimation results: ‘puddles’

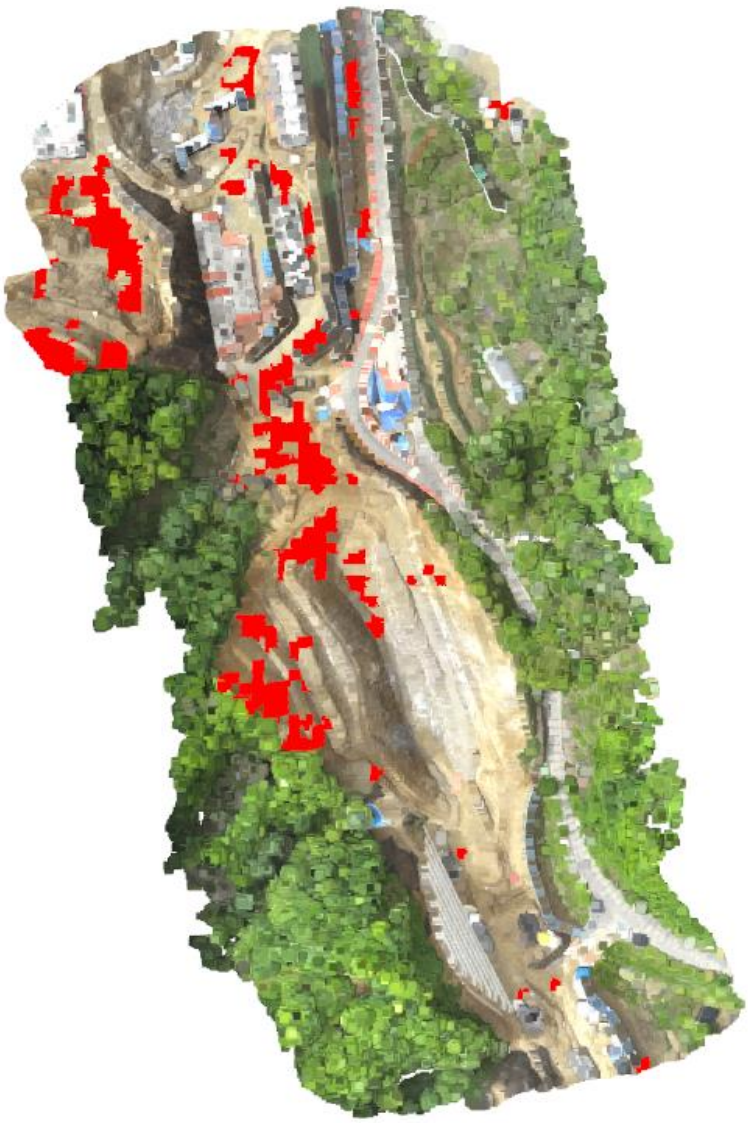


Area = 2,164m2
*Area Ratio = 7.4%
(Relative error 19%)

*Area ratio: the ratio of the area occupied by the class to the total area of the site

Figure B.5 Mapping and area estimation results: ‘nets’

C. 3-D Visualization Results – Chapter 6



Area = 1,770m^2
*Area Ratio = 3.9%
(Relative error 19%)

*Area ratio: the ratio of the area occupied by the class to the total area of the site

Figure C.1 Mapping and area estimation results: ‘soil’

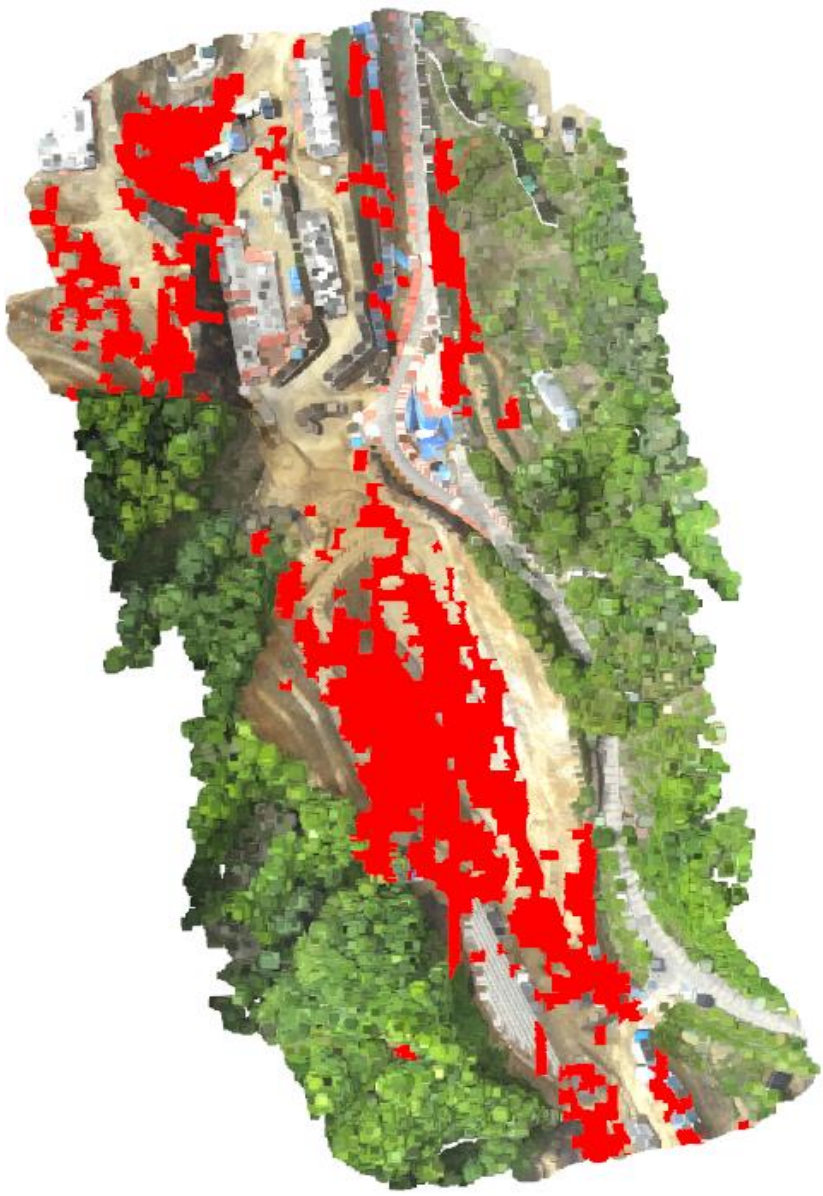
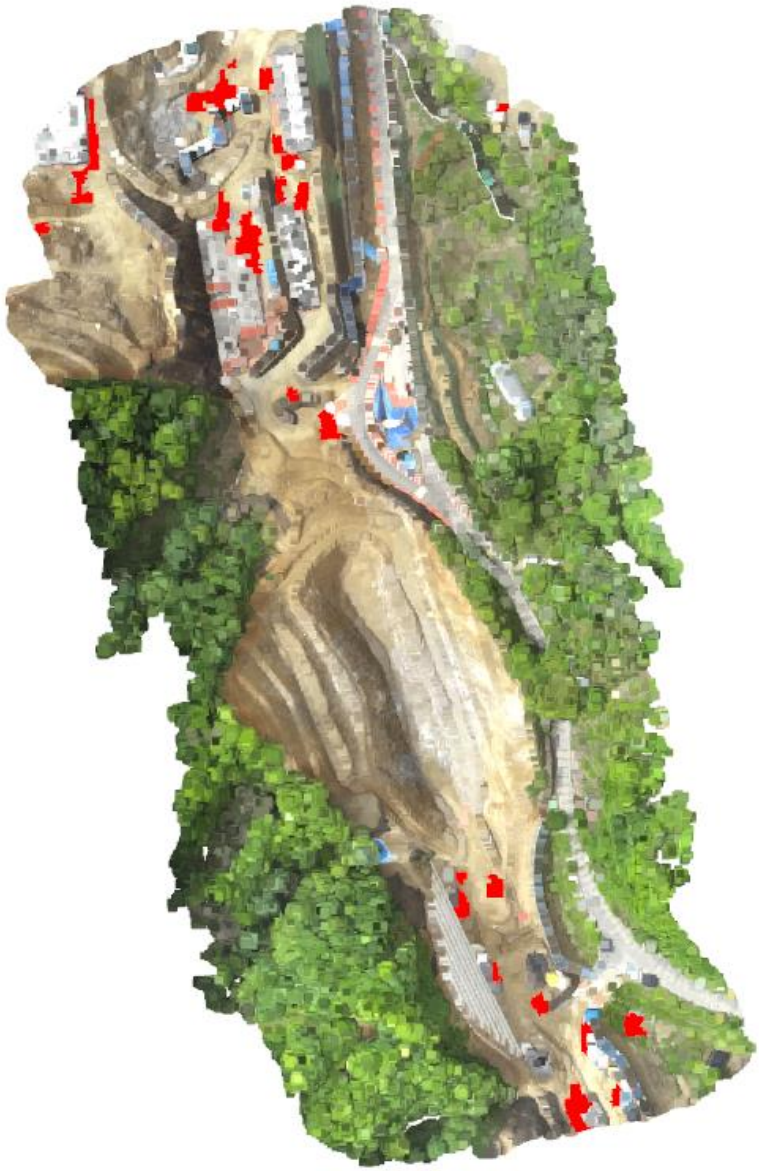


Figure C.2 Mapping and area estimation results: ‘rocks’

Area = 18,953m^2
*Area Ratio = 41.4%
(Relative error 0.23%)

*Area ratio: the ratio of the area occupied by the class to the total area of the site

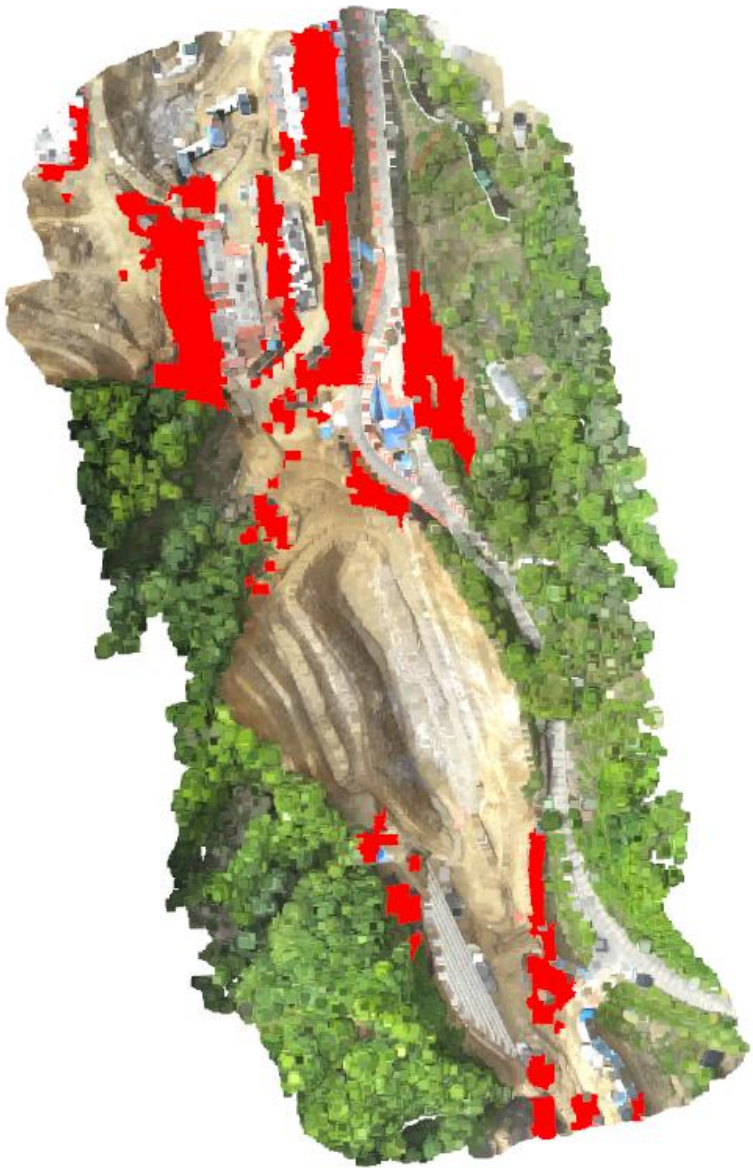
Figure C.3 Mapping and area estimation results: ‘trees’



Area = 615m^2
*Area Ratio = 1.3%
(Relative error 0.27%)

*Area ratio: the ratio of the area occupied by the class to the total area of the site

Figure C.4 Mapping and area estimation results: ‘puddles’



Area = 2,788m^2
*Area Ratio = 6.1%
(Relative error 20%)

*Area ratio: the ratio of the area occupied by the class to the total area of the site

Figure C.5 Mapping and area estimation results: ‘nets’