AUTOMATED EARTHWORK PROGRESS MONITORING FOR CONSTRUCTION PROJECTS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

AHMET BAHADDİN ERSÖZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
CIVIL ENGINEERING

AUGUST 2023

Approval of the thesis:

AUTOMATED EARTHWORK PROGRESS MONITORING FOR CONSTRUCTION PROJECTS

submitted by **AHMET BAHADDİN ERSÖZ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** i**n Civil Engineering, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Erdem Canbay
Head of the Department, **Civil Engineering** _____

Assoc. Prof. Dr. Onur Pekcan
Supervisor, **Civil Engineering, METU** _____

Examining Committee Members:

Prof. Dr. Berna Unutmaz
Civil Engineering, Hacettepe University _____

Assoc. Prof. Dr. Onur Pekcan
Civil Engineering, METU _____

Assoc. Prof. Dr. Emre Akbaş
Computer Engineering, METU _____

Assoc. Prof. Dr. Güzide Atasoy Özcan
Civil Engineering, METU _____

Asst. Prof. Dr. Seda Selçuk
Civil Engineering, Çankaya University _____

Date: 10.08.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : Ahmet Bahaddin Ersöz

Signature :

ABSTRACT

AUTOMATED EARTHWORK PROGRESS MONITORING FOR CONSTRUCTION PROJECTS

Ersöz, Ahmet Bahaddin
Doctor of Philosophy, Civil Engineering
Supervisor: Assoc. Prof. Dr. Onur Pekcan

August 2023, 172 pages

This thesis presents a novel approach to earthwork progress monitoring on construction sites using Unmanned Aerial Vehicles (UAVs), machine learning, and computer vision. Traditional methods, often time-consuming and prone to error, are addressed through the developed methodology, enhancing efficiency and accuracy. A fully autonomous pipeline is established, from image acquisition to volume calculation. The utilization of UAVs for imaging during operational hours improves safety measures by eliminating the risks associated with manual inspections. A real-world aerial image dataset with pixel-level annotation from a wide variety of construction environments annotation is presented. A machine learning model is trained using this dataset that isolates construction machinery from images, achieving an average precision score exceeding 90% for typical machinery categories. An image inpainting algorithm is presented to remove these machines from images, enhancing the visual representation of construction terrains for volume calculation. Furthermore, a novel volume calculation algorithm is proposed, providing an accurate and efficient method for estimating earthwork volumes. This study performs a comparative experiment of point clouds generated by UAVs and Terrestrial Laser Scanning, demonstrating the UAV-based approach's viability as a

reliable and cost-effective alternative. Despite these advancements, the research acknowledges challenges in UAV performance due to weather variations, difficulties detecting diverse construction machinery, and the complexity of dealing with closely positioned machines. In conclusion, this research paves the way for more efficient construction management practices. Future research can focus on expanding the machine learning training datasets to improve the process further and integration with 4D Building Information Modeling (BIM).

Keywords: Earthwork Progress Tracking, Construction Image Dataset, Unmanned Aerial Vehicles, Deep Learning, Image inpainting

ÖZ

İNŞAAT PROJELERİ İÇİN OTOMATİK HAFRİYAT İLERLEME TAKİBİ

Ersöz, Ahmet Bahaddin
Doktora, İnşaat Mühendisliği
Tez Yöneticisi: Doç. Dr. Onur Pekcan

Ağustos 2023, 172 sayfa

Bu tez, İnsansız Hava Araçları (İHA), makine öğrenimi ve bilgisayar görüşü kullanarak inşaat sahalarındaki hafriyat ilerlemesini izlemek için yeni bir yaklaşım sunmaktadır. Genellikle zaman alıcı ve hataya açık olan geleneksel yöntemler, geliştirilen metodoloji ile ele alınarak verimlilik ve doğruluk artırılmaktadır. Görüntü alımından hacim hesaplamasına kadar tamamen otonom bir veri akışı oluşturulmuştur. Çalışma saatlerinde görüntüleme için İHA'ların kullanılması, manuel yöntemlerle ilişkili riskleri ortadan kaldırarak güvenlik önlemlerini iyileştirir. Çok çeşitli inşaat sahalarından piksel düzeyinde etiket içeren havadan çekilmiş görüntü kümesi sunulmuştur. Bu veri kümesi kullanılarak, inşaat makinelerini görüntülerden izole eden ve tipik makine kategorileri için %90'ı aşan ortalama bir hassasiyet puanı elde eden bir makine öğrenimi modeli eğitilmiştir. Bu inşaat makinelerini görüntülerden kaldırmak ve hacim hesaplaması için inşaat arazilerinin görsel temsilini geliştirmek için bir görüntü boyama algoritması sunulmuştur. Ayrıca, hafriyat hacimlerini tahmin etmek için doğru ve verimli bir yöntem sağlayan yeni bir hacim hesaplama algoritması önerilmiştir. Bu çalışma, İHA'lar ve yersel lazer tarama tarafından üretilen nokta bulutlarının karşılaştırmalı bir testini gerçekleştirerek, İHA tabanlı yaklaşımın güvenilir ve uygun maliyetli bir alternatif olarak uygulanabilirliğini göstermektedir. Bu gelişmelere rağmen

araştırma, hava koşullarındaki değişimler, çeşitli inşaat makinelerini tespit etmedeki zorluklar ve birbirine yakın konumdaki makinelerin ayrıştırılmasının karmaşıklığı nedeniyle İHA performansındaki zorlukları kabul etmektedir. Sonuç olarak, bu araştırma daha verimli inşaat yönetimi uygulamalarının önünü açmaktadır. Gelecekteki araştırmalar, süreci daha da iyileştirmek için makine öğrenimi eğitim veri setlerini genişletmeye ve 4 Boyutlu Yapı Bilgi Modellemesi (BIM) ile entegrasyona odaklanabilir.

Anahtar Kelimeler: Hafriyat İlerleme Takibi, İnşaat Görüntü Kümesi, İnsansız Hava Araçları, Derin Öğrenme, Görüntü Boyama

To my beloved mom

ACKNOWLEDGMENTS

First and foremost, I am deeply grateful to my supervisor, Dr. Onur Pekcan, whose guidance, support, and mentorship have been invaluable throughout this journey. His commitment to my academic and personal growth has been helpful in shaping this thesis. More than a supervisor, he has been a source of inspiration and encouragement, and I am genuinely thankful for that.

I extend my heartfelt appreciation to the members of the thesis monitoring committee, Dr. Güzide Atasoy Özcan and Dr. Emre Akbas. Their constructive feedback and insightful suggestions played a pivotal role in refining and advancing this study. I am also thankful to the esteemed jury members for their reviews contributing to its enhancement.

Special thanks go to Serhat Erinmez and Ali Özdemir, whose unwavering assistance in conducting field studies has been indispensable to the successful completion of this research. Their dedication and collaboration have enriched my work beyond measure. I also extend my gratitude to the members of the AI2LAB research team for their companionship and collaborative spirit.

I am fortunate to have the invaluable support of my fiancée, Beste Ozyurt, who has been by my side through every challenge. Her presence, encouragement, and belief in me have been a constant source of strength, and I consider myself incredibly lucky to have her in my life.

In closing, this thesis is dedicated to in loving memory of my dear mother, Meryem Ersöz, whose enduring love inspires me every day. Though she is no longer with me, her spirit lives on. Her influence on my life will forever remain a driving force behind my aspirations. I would also like to extend my heartfelt thanks to my father and sisters for their constant encouragement and strength.

I am also grateful to all those who have supported me in ways seen and unseen, shaping my academic path and contributing to completing this thesis.

This research is supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under grant number 122M055.

TABLE OF CONTENTS

ABSTRACT v
ÖZ vii
ACKNOWLEDGMENTS x
TABLE OF CONTENTS xii
LIST OF TABLES xv
LIST OF FIGURES xvi
CHAPTERS
1 INTRODUCTION 1
1.1 Overview and Problem Statement 1
1.2 Objectives of the Research 6
1.3 Scope of the Thesis 8
1.4 Thesis Organization 9
2 LITERATURE REVIEW 11
2.1 Object Detection using Machine Learning 12
2.1.1 General Purpose Datasets 12
2.1.2 Construction-Specific Datasets 16
2.1.3 Applications in Construction 22
2.2 Object Removal using Image Inpainting 24
2.2.1 Inpainting Datasets 26
2.2.2 Inpainting Applications 27
2.3 Earthwork Progress Monitoring Approaches 31
2.3.1 Laser-based Approaches 31
2.3.2 Photogrammetry-based Approaches 36
2.4 Volume Calculation Methods 41
2.5 Current Gap in the Literature 43

3 DEVELOPMENT OF MACHINE LEARNING MODEL.....47
3.1 Introduction..... 47
3.2 Image Collection..... 48
3.3 Image Annotation..... 56
3.4 Dataset Statistics..... 61
3.5 Performance Evaluation..... 67
3.6 Results and Discussion..... 73
4 METHODOLOGY FOR EARTHWORK VOLUME CALCULATION.....79
4.1 Introduction..... 79
4.2 Data Collection..... 80
4.3 Construction Machine Segmentation and Removal..... 82
4.4 3D Surface Reconstruction..... 86
4.4.1 Feature Extraction and Matching..... 87
4.4.2 Sparse Point Cloud Construction..... 90
4.4.3 Dense Point Cloud Reconstruction..... 91
4.4.4 Surface Reconstruction..... 92
4.5 Volume Calculation..... 96
5 FIELD EXPERIMENTS.....101
5.1 Study Area..... 101
5.2 Data Collection..... 107
5.2.1 UAV Imaging..... 109
5.2.2 Terrestrial Laser Scanning..... 114
5.3 Construction Machines Segmentation..... 119
5.4 Construction Machines Removal..... 124
5.5 3D Surface Reconstruction..... 128

5.6 Volume Calculation 134
5.7 Results and Discussion 138
6 CONCLUSIONS 143
6.1 Summary 143
6.2 Contributions 146
6.3 Limitations and Future Work 149
REFERENCES 153
CURRICULUM VITAE 171

LIST OF TABLES

TABLES

Table 3.1 Results of the Clustered Data 72
Table 3.2 Results of the Unclustered Data 72
Table 3.3 Classwise AP Results (IoU=50%) 73
Table 5.1 GCP World Coordinates 112
Table 5.2 Point Cloud Registration and Georeferencing Results 118
Table 5.3 Number of Instances Included in the Images 120
Table 5.4 Segmentation Results 121
Table 5.5 Classwise AP Results (IoU=50%) 121
Table 5.6 Inpainting Results 127
Table 5.7 3D Reconstruction Results 129
Table 5.8 Truck Count 137
Table 5.9 Comparison of Methods 138
Table 5.10 Earthwork Progress using UAV-Based Approach 138

LIST OF FIGURES

FIGURES

Figure 1.1. Flowchart of Proposed Method 8
Figure 2.1. Annotated Images from COCO Dataset 13
Figure 2.2. Image Examples from MOCS Dataset 19
Figure 2.3. Synthetic Image of Construction Machine 21
Figure 2.4. Image Restoration and Text Removal 25
Figure 2.5. Sample Images From Places2 Dataset 27
Figure 2.6. Steps of Construction Worker Removal 30
Figure 2.7. 3D Laser Scanner 33
Figure 2.8. Sphere Target for Laser Scanning 34
Figure 3.1. Aerial Perspective View of Images Collected from Construction Sites 49
Figure 3.2. Examples of Aerial Images from the Construction Sites 51
Figure 3.3. Examples of Aerial Images from the Parking Areas 53
Figure 3.4. Negative Image Samples that Do Not Include Machinery 54
Figure 3.5. Different Weather Conditions on the Same Aerial Spot 55
Figure 3.6. Selected Object Classes for Image Annotation 58
Figure 3.7. Privacy Protections for Faces, Brand Logos, and Facets 59
Figure 3.8. CVAT Image Labeling Interfaces 61
Figure 3.9. Number of Images and Number of Objects for Each Type of Construction Machine 62
Figure 3.10. Number of Objects per Image (%) 63
Figure 3.11. Number of Categories per Image (%) 63
Figure 3.12. Distribution of the Object Size (%) 64
Figure 3.13. Similar Images from the Literature 65
Figure 3.14. Clustering Strategy Examples for Overlapping Images 66
Figure 3.15. Illustration of Intersection over Union (IoU) 70
Figure 3.16. Example Images of Segmentation Results 74
Figure 3.17. Results of Incorrect Object Segmentations 77

Figure 4.1. Demonstration of Overlapping Images 81
Figure 4.2. Comparisons of Visual Recognition Methods 84
Figure 4.3 Original and Inpainted Images using LaMa Algorithm 86
Figure 4.4 SfM-MVS Flowchart 87
Figure 4.5. SfM-MVS Pipeline Visualization 90
Figure 4.6 Preprocessing Steps to Calculate Volumes 93
Figure 4.7. 3D Demonstrations 94
Figure 4.8. Poisson Surface Reconstruction 96
Figure 4.9. Ray Length and Volume Calculation Pseudocode 97
Figure 4.10. Volumetric Difference Calculation of an Artificial Object 99
Figure 4.11. Illustration of Rays Between Two Meshes 100
Figure 5.1. Location of the Construction 102
Figure 5.2. Illustration of the Project 102
Figure 5.3. CAD Drawings of the Construction 104
Figure 5.4. Views of the Excavation Area 105
Figure 5.5. Borehole Plan of the Construction Site 106
Figure 5.6. Soil Profile of the Construction Area 106
Figure 5.7. Site Visits for the Construction Site 108
Figure 5.8. Locations of the Ground Control Point 110
Figure 5.9. Details of GCP Surveying 111
Figure 5.10 Selected Drone – DJI Mavic Pro 113
Figure 5.11 Flight Mission Plan Used for All UAV Flights 114
Figure 5.12 Laser Scanner Device – Leica BLK360 115
Figure 5.13 Polystyrene Sphere Targets 116
Figure 5.14 T1 Laser Scanning Station Locations 117
Figure 5.15 T2 Laser Scanning Station Locations 117
Figure 5.16 Laser Scanner and UAV on the site 119
Figure 5.17 Aerial View of Construction Site 120
Figure 5.18. Correct Object Segmentations 122
Figure 5.19. Incorrect Object Segmentation 123

Figure 5.20. T1 Inpainting Results 125
Figure 5.21. T2 Inpainted Results 126
Figure 5.22 Resultant Orthophotos 130
Figure 5.23 Resultant Point Clouds 131
Figure 5.24 Laser Scanner Point Clouds 132
Figure 5.25 Manual Construction Machine Removal 132
Figure 5.26 Laser Scanner Point Clouds Visualization 133
Figure 5.27 3D Reconstruction Steps 134
Figure 5.28 Volume Distribution Visualization 136
Figure 5.29 Typical Truck Model from the Construction Site 137

CHAPTER 1

INTRODUCTION

1.1 Overview and Problem Statement

Earthwork operations are critical and often among the most costly activities involved in construction projects (Hare et al., 2011). Consequently, even the most minor miscalculated volume for excavation or fills on-site can result in substantial financial losses.

Beyond accounting for a significant portion of the project budget, earthwork operations also serve as the foundation of the construction project, laying the groundwork for subsequent operations. Any delays in these foundational earthwork activities can result in significant schedule delays and cost overruns throughout the project. Therefore, accurate and regular monitoring is necessary to complete a construction project within the allocated timeframe and budget. Contractors are responsible for tracking ongoing changes throughout the construction duration.

Calculating the earthwork volume with precision is crucial. Accurate volume calculations enable project managers to provide more precise estimates for labor, equipment, and material costs associated with these operations. Topographical change data is vital as it is the primary information for reviewing earthwork design, monitoring progress, estimating earthwork volume, and documenting the project. To calculate topographical changes effectively, it is essential to accurately represent the topography of the earthwork area.

There are numerous methods available for monitoring topographical changes. Considering their primary technology, it is possible to categorize them as laser-based and photogrammetry-based techniques. Each method has its advantages and disadvantages depending on the specific situation. Laser-based methods operate by

emitting a laser beam towards a target and measuring the time it takes to reflect the beam back to a receiver on the device. This time measurement is then used to calculate the distance to the target. Total stations and laser scanning devices are the most commonly used laser-based measurement tools in construction and earthwork operations.

Total stations, although a traditional method, may be ineffective in certain circumstances. They can be labor-intensive and time-consuming to use because they measure only one point distance per measurement. As the size of the area increases, their efficiency decreases. For example, creating a representation of a 100-by-100-meter earthwork site using a total station would require approximately two days of fieldwork and several hours of CAD drawing time (Chae et al., 2011). While these methods can be cost-effective for small-scale surveying, the financial and labor costs can significantly increase for larger areas (Akgul et al., 2018). Moreover, deploying personnel under challenging terrains poses inherent safety risks.

Terrestrial laser scanning technology (TLS) offers a potential solution as it rapidly captures high-resolution 3D data over a large area. It generates a comprehensive point cloud by measuring multiple laser beams' distances. However, the effectiveness of TLS may be significantly reduced on topographically complex sites. In such cases, frequent repositioning of the scanner is necessary to avoid obstructions, increasing the survey duration and overall cost. Another option is airborne Light Detection and Ranging (LiDAR) surveys, but they require a significant financial investment, which may not be feasible for most monitoring budgets (Hugenholtz et al., 2015).

In conclusion, laser-based monitoring of the progress of earthwork operations poses significant challenges that often involve extensive manual labor and resources. There is a need for a method that can accurately and efficiently survey the topography of earthwork, enabling analysis, forecasting, supervision, and documentation.

Photogrammetry-based methods have emerged as a practical alternative to laser-based technologies and have gained popularity over the years due to their cost-

effectiveness, accessibility, and ease of use. This process involves converting images into 3D models using computer vision algorithms. However, close-range photogrammetry, which typically consists of capturing images from the ground, has limitations. One significant challenge is that the camera viewpoint from the ground may not encompass the entire excavation site, resulting in incomplete data capture and gaps in the final 3D model. Additionally, capturing images by walking in excavation sites can pose safety risks due to unstable ground or other hazards, presenting methodological challenges and safety concerns.

Recently, Unmanned Aerial Vehicles (UAVs), commonly known as drones, have emerged as a solution to address the limitations of close-range photogrammetry. UAVs offer several advantages for surveying tasks, including their portability, maneuverability, and small size, which make them easy to transport and operate. They are handy for periodic observations of large areas, enabling high operational efficiency and rapid data collection. UAVs are also cost-effective and time-efficient, capable of capturing high-resolution images (Ajayi & Ajulo, 2021). Their ability to create point clouds is highly valued, as these 3D point clouds can be utilized for accurate volume estimations in cut-and-fill calculations, providing invaluable insights into earthworks. A study by Siebert and Teizer (2014), demonstrated the successful application of UAVs in areas ranging from 100 to 100,000 m2, maintaining an acceptable accuracy margin of just a few centimeters. This promising performance indicates a bright future for UAV technology in earthwork monitoring.

In their studies, Hugenholtz et al. (2015) emphasized the value of small UAVs for volumetric analysis in various scales of earthwork projects. They found that UAVs offer a more effective and safer alternative to conventional ground-based approaches, especially in complex or hazardous survey environments. On the other hand, Tamin et al. (2019) highlighted the practical advantages of UAVs compared to Terrestrial Laser Scanning (TLS) techniques, despite the longer processing time of TLS. They noted that UAVs provide easier deployment, reduced personnel requirements, enhanced safety measures, and faster acquisition of aerial images, confirming their potential as an effective monitoring tool in earthwork projects.

Despite the promising opportunities that UAV photogrammetry presents for tracking earthwork progress, there are still challenges regarding earthwork volume calculations. One significant obstacle is the presence of moving objects, particularly construction machines, in UAV images, as they can distort actual ground-level measurements. Only a limited number of studies in the literature have addressed the issue of non-terrain objects affecting the accuracy of volumetric difference calculations in UAV-based earthwork monitoring.

In the visualization of height differences using point cloud data, Cho et al. (2021) highlighted the difficulties caused by moving objects such as cars and construction equipment. These objects had a significant impact on the generation of height differential maps. Similarly, W. J. Kim et al. (2022) emphasized that temporary facility objects captured in UAV mapping results can introduce errors in earthwork volume computations. Both studies highlight the urgent need to refine existing methodologies to mitigate inaccuracies caused by non-terrain objects in UAV-based photogrammetry. Further research and development are required to address these challenges and improve the accuracy and reliability of UAV-based earthwork volume calculations, particularly in the presence of non-terrain objects.

There are geometry-based point cloud filtering algorithms available to remove objects such as trees and buildings from point cloud data generated from photogrammetry. These methods aim to identify geometric threshold values by considering changes in the ground slope and the shape of the point cloud. The most widely used filters in research are the Cloth Simulation Filter (CSF) and the Simple Morphological Filter (SMRF) (Pingel et al., 2013; W. Zhang et al., 2016).

However, these filters have certain limitations. They require careful adjustment of several parameters for each dataset, and selecting inappropriate parameter values can result in poor filtering results. This makes them less user-friendly and less adaptable to different datasets without proper fine-tuning. Moreover, these filters face challenges when encountering sudden or sharp changes in terrain, such as slopes

commonly found in excavation areas. The algorithms may incorrectly interpret these natural variations in terrain as objects, leading to inaccuracies in the filtered results.

In conclusion, there is a need for an automated method that can accurately represent the earthwork topography without the presence of machinery and calculate volumetric operations for forecasting, supervision, and documentation purposes. Future research and development should aim to address the limitations of existing filtering algorithms and develop more robust and adaptable approaches to improve the accuracy and reliability of UAV-based earthwork monitoring.

This research introduces an autonomous pipeline for monitoring earthwork progress in construction sites, covering image acquisition to volume calculation. A dataset of 2155 annotated aerial images from diverse construction environments is presented. The validation of deep learning techniques on this dataset demonstrates precision rates exceeding 90% for standard categories, showcasing a robust methodology.

UAV imaging is favored over TLS due to its positive impact on productivity and safety. Notably, image inpainting achieves a Structural Similarity Index Measure (SSIM) score of 90%, enabling the removal of machinery from images and facilitating terrain reconstruction. The resulting reconstructed point clouds using UAV photogrammetry, with a ground sampling accuracy of 1.3 cm, contribute to precise earthwork calculations. Furthermore, a novel volume calculation algorithm aligns favorably with estimates based on truck cycle counts.

However, challenges include accurately distinguishing diverse machinery and addressing closely located machinery recognition. Weather-resistant systems and dataset augmentation are recommended for improved algorithmic performance. Future directions involve integrating 3D soil profiles for cost estimation and incorporating 4D Building Information Modeling for excavation progress monitoring. This comprehensive approach advances earthwork monitoring in construction, spanning technology, algorithms, and operations.

1.2 Objectives of the Research

The main objective of this study is to develop an automated earthwork progress tracking system by utilizing Unmanned Aerial Vehicle (UAV) imaging coupled with machine learning and computer vision algorithms. The study aims to address the drawbacks of conventional earthwork volume calculation methods and provide an efficient solution to overcome these challenges. To achieve this goal, this thesis is divided into several constituent parts:

- ***Benefiting the advantages of UAV imaging:*** The study aims to demonstrate the significant benefits of using UAVs for imaging compared to traditional ground-based methods. It seeks to show how UAVs can be effectively deployed during construction hours, unaffected by the movement of construction machines. This overcomes limitations related to safety concerns and operational complexities typically encountered by ground-based methods.
- ***Machine learning for construction machine segmentation:*** In machine learning, the study aims to develop and evaluate a deep learning model designed explicitly for detecting construction machines in construction site images. The primary goal is to create an open image dataset containing commonly encountered types of machinery such as dump trucks, excavators, and loaders.
- ***Image inpainting for construction machine removal:*** The study focuses on testing and validating the implementation of an image inpainting algorithm. This algorithm aims to remove construction machines from the images, creating a fully automated pipeline to represent the excavation topography accurately.
- ***Volume calculation methodology using UAV-based photogrammetry:*** The final goal is to develop a method for volume calculation that utilizes point clouds derived from UAV-based photogrammetry. The study will compare the earthwork volume estimations generated by this proposed technique with

those produced by traditional methods such as terrestrial laser scanning. This comparison will help to understand the accuracy and effectiveness of the newly proposed method.

The study aims to develop an automated system that can accurately track earthwork progress using UAV imaging, machine learning, and computer vision techniques by addressing these constituent parts. Figure 1.1 presents a graphical depiction of the proposed system for earthwork progress tracking following the objectives outlined above.

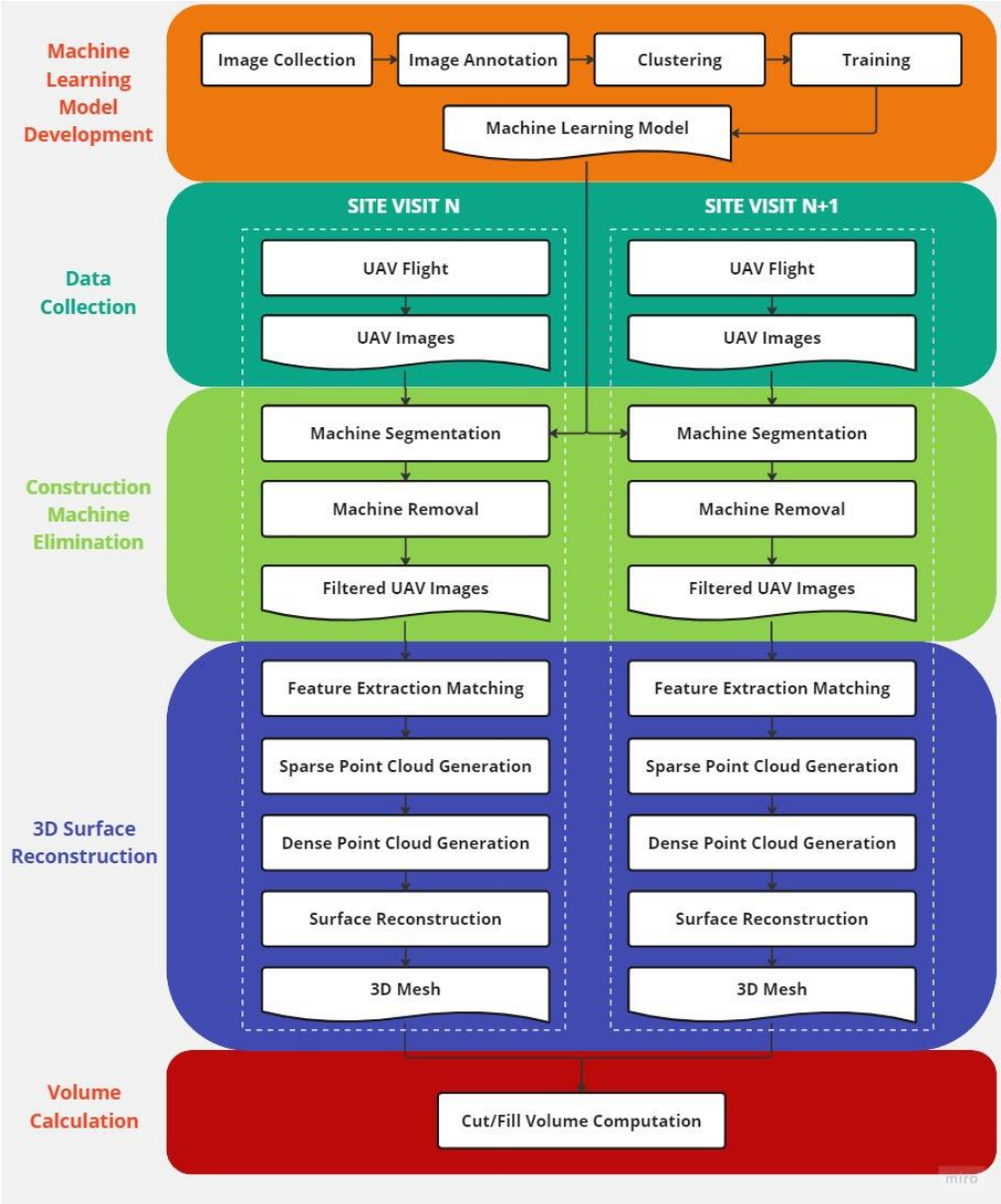


Figure 1.1. Flowchart of Proposed Method

1.3 Scope of the Thesis

This thesis addresses critical challenges in earthwork operations, primarily focusing on the accurate and efficient monitoring and calculation of earthwork volume. The specific areas of focus within this thesis are as follows:

The research scope is primarily limited to earthwork operations at construction sites, specifically concentrating on the initial stages of earthwork, starting from excavation and fill processes and continuing until the commencement of foundation construction. The inclusion of consistent UAV imaging from job sites multiple times is an essential aspect of this thesis's scope.

The investigation primarily tackles the challenge posed by the movement of construction machinery within the construction site, which can interfere with accurate topographic measurements. This study focuses on eight commonly found types of vehicles in construction job sites, including dump trucks, excavators, backhoe loaders, wheel loaders, compactors, dozers, graders, and cars. Other machines, such as asphalt paving and drilling machines, appear less frequently and are grouped under the "other" category. Construction machines not included in the "other" category are not within the scope of the construction machine segmentation and removal process.

1.4 Thesis Organization

The dissertation is divided into six major chapters that are structured to present the material in a logical and coherent manner, systematically guiding the reader through the intricacies of the study.

In the first chapter, "Introduction," the problem statement, objectives, research questions, and significance of the study are introduced.

The second chapter, titled "Literature Review," provides an exhaustive review of the relevant literature. This chapter offers a comprehensive overview of the current state of earthwork progress monitoring, the use of Unmanned Aerial Vehicles (UAVs) in construction, machine learning and datasets, and image inpainting applications.

The development and evaluation of aerial image datasets and machine learning models are described in detail in Chapter 3, "Development of Machine Learning Model."

Chapter 4, "Methodology for Earthwork Volume Calculation," provides a step-by-step description of the UAV-based photogrammetry procedure, the machine learning techniques used for the segmentation of construction machinery, and the image inpainting approach utilized.

In Chapter 5, "Field Experiments," the proposed methodology is applied to an actual construction site, demonstrating its applicability and efficacy. It evaluates the system's performance in a real-world environment and discusses any challenges or limitations that arise.

The concluding chapter is titled "Conclusions" (Chapter 6). It provides a summary of the key findings, contributions made by the research, obstacles encountered, and recommendations for future research.

CHAPTER 2

LITERATURE REVIEW

The development and integration of cutting-edge technologies have played a crucial role in earthwork calculations. Particularly, the application of deep learning techniques to object detection has fundamentally changed the identification and classification of objects within an image taken from the construction. In addition, the capability to remove unwanted objects from an image using image inpainting techniques has increased the outputs' precision and provided a clearer representation of the work area. In addition to data processing, with advancements in data collection technology over recent years, traditional methodologies for earthwork monitoring have been largely supplanted by innovative techniques employing Unmanned Aerial Vehicles (UAVs), Global Navigation Satellite Systems (GNSS), Terrestrial Laser Scanners (TLS), and photogrammetry. The combination of these techniques provides a comprehensive set of tools for processing and interpreting imagery associated with construction projects.

This literature review is organized to guide the reader sequentially through the interrelated aspects of earthwork calculations. Beginning with the foundations of image processing via object detection and image inpainting, we then transition to a broader exploration of the various methodologies used in earthwork calculations before concluding with a focused examination of the essential techniques used for volumetric computation. This systematic approach is intended to provide a comprehensive understanding of the current state of technology and methodologies in earthwork operations.

2.1 Object Detection using Machine Learning

Rapid advances in deep learning have contributed to promising developments in civil engineering. Utilizing deep learning methodologies can transform how engineers approach productivity, safety, and efficiency in the construction sector. This chapter explores the complex relationship between deep learning and construction, focusing on object detection of construction equipment. Prior to looking deeper into the specifics of its application in the construction industry, it is crucial to understand some fundamental deep-learning terminology.

Deep learning is a subset of machine learning, a subfield of artificial intelligence involving the development and application of algorithms that allow machines to learn from data and make decisions or predictions. Multiple-layered artificial neural networks are used to construct deep learning models, hence the term "deep." These models acquire knowledge directly from data. The successful application of those techniques requires comprehensive and diverse image datasets.

Through a review of the current literature, this section investigates the image datasets and the application of deep learning to object detection in the context of earthwork monitoring. The discussion encompasses a variety of techniques, their utility, their limitations, and the potential for future developments in this field.

2.1.1 General Purpose Datasets

For the advancement and evaluation of deep learning algorithms, it is crucial to have access to large datasets. These datasets provide an enormous amount of training and testing data for machine learning models. Numerous large-scale datasets have been introduced over time and have had a significant impact on the fields of computer vision and machine learning. Figure 2.1 illustrates sample annotated images from one of the most known datasets, COCO (T. Y. Lin et al., 2014).

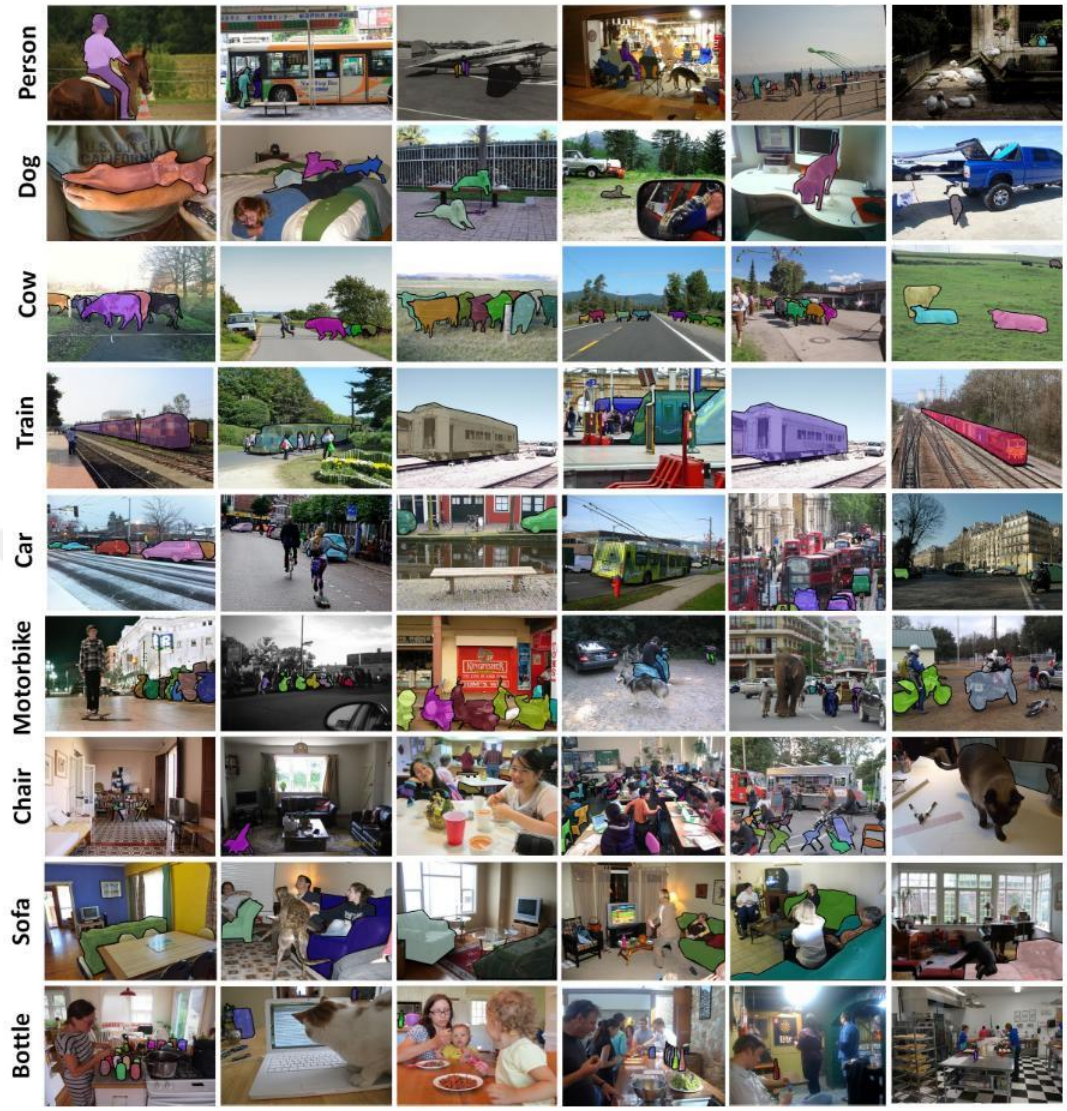


Figure 2.1. Annotated Images from COCO Dataset (T. Y. Lin et al., 2014)

Researchers from the California Institute of Technology published Caltech-101 and Caltech-256 as two of the earliest datasets. Caltech-101 consists of 101 classes (Fei-Fei et al., 2004), whereas Caltech-256 consists of 256 classes (Griffin et al., 2007), which include a wide variety of objects such as cars, motorbikes, and airplanes. Each dataset includes 9,146 and 30,607 images, respectively. Caltech-256 improves upon its predecessor by offering greater intra-class variability, more images per category, and the inclusion of cluttered backgrounds and various object orientations. As a

result, Caltech-256 presents a more realistic and challenging benchmark for computer vision algorithms. Similarly, the Canadian Institute for Advanced Research (CIFAR) introduced the CIFAR-10 and CIFAR-100 datasets, which were collected by Alex Krizhevsky (2009). CIFAR-10 contains 60,000 32x32 pixel color images divided into ten classes with 6,000 images per class, while CIFAR-100, an extension of CIFAR-10, contains 100 classes with 600 images per class, for a total of 60,000 images. Both datasets contain preprocessed, fixed-size, low-resolution images, with 50,000 training and 10,000 test images for each. The CIFAR-100 classes are subdivided into twenty "superclasses" to facilitate more complex classification tasks.

The Pascal Visual Object Classes (Pascal VOC) competition was a major event in the early days of computer vision research, propelling significant advancements in the field (Everingham et al., 2010). The competition, which was held annually from 2005 to 2012, attracted researchers from all over the world to develop innovative algorithms for tasks such as object recognition, detection, and segmentation. The datasets associated with Pascal VOC have expanded over time, from four categories and 1,578 images in 2005 to twenty categories and 11,530 images in 2012. These images included a wide variety of real-world objects, such as chairs, dogs, humans, and vehicles, providing a more inclusive and challenging benchmark for the computer vision community. The datasets also included annotations for object class labels, bounding boxes, and segmentation masks, allowing researchers to evaluate their algorithms across multiple facets of object comprehension. The Pascal VOC challenge significantly advanced and popularized deep learning techniques such as Convolutional Neural Networks (CNNs) by fostering a collaborative environment and providing a robust platform for testing and refining novel approaches in computer vision.

The Microsoft COCO (Common Objects in Context) dataset is yet another widely used image dataset with over 200,000 labels and 330,000 images released (T. Y. Lin et al., 2014). In comparison to the Pascal VOC, the COCO dataset focuses on smaller objects and encompasses 80 categories, providing practitioners of machine learning

with a more extensive range of data (Figure 2.1). Compared to its predecessors, this dataset is more challenging and realistic due to the images' depiction of complex everyday scenes containing common objects. COCO's annotations consist of object class labels, bounding boxes, segmentation masks, and image captions, allowing for the evaluation of various algorithms and techniques.

ImageNet and Google's Open Images Dataset are two extensive, large-scale datasets that have significantly impacted the field of computer vision and machine learning. ImageNet, introduced in 2009, is a colossal dataset containing approximately 14 million images (Jia Deng et al., 2009), while the Open Images Dataset, released by Google in 2017, comprises around 9 million images (Krasin et al., 2017). Both datasets encompass a wide variety of object categories, with ImageNet having over 20,000 classes and Open Images featuring around 6,000 classes, covering objects as diverse as balloons, strawberries, animals, and vehicles. The ImageNet dataset has been particularly influential, as its associated ImageNet Large Scale Visual Recognition Challenge (ILSVRC) catalyzed breakthroughs in computer vision (Berg et al., 2010), such as the success of AlexNet, which showcased the power of deep learning for image recognition tasks. The Open Images Dataset, on the other hand, has further expanded the scope of available data, providing an even more comprehensive and challenging benchmark for researchers to develop and test their algorithms. Together, these datasets have played a crucial role in shaping the field of computer vision and have been instrumental in the rapid advancements witnessed in recent years.

The last dataset should be mentioned is the LVIS (Large Vocabulary Instance Segmentation) dataset. Facebook AI Research developed the dataset to address previous datasets' category coverage limitations and encourage research in instance segmentation using a more extensive vocabulary (Gupta et al., 2019). The dataset contains over 1,000 entry-level object categories, which enables researchers to examine a greater variety of objects and scenarios. With over 2 million objects cautiously masked across 164,000 images, LVIS provides abundant data for training and evaluating machine learning models. The LVIS dataset is designed to address

the long-tail distribution of object classes, a common issue in real-world scenarios where some object classes are common while others are uncommon. This presents a unique challenge for machine learning models, as they typically perform well with frequent categories but struggle with rare or underrepresented ones. The design of the dataset encourages researchers to create algorithms that can effectively manage such imbalanced distributions, thereby pushing the limits of existing computer vision techniques. The availability of the datasets mentioned above has dramatically driven the advancement of deep learning algorithms and computer vision technology. It is predicted that as technology advances, even larger and more diverse datasets will emerge, further stimulating innovation and progress in deep learning and computer vision research.

2.1.2 Construction-Specific Datasets

While general-purpose datasets have played a pivotal role in advancing computer vision and deep learning technologies in various fields, their application in the construction industry is gaining increasing attention. With diverse images and annotations, these datasets provide valuable training grounds for machine learning models. However, construction sites' unique and complicated nature necessitates a specialized approach. This recognition has prompted researchers to develop construction-specific datasets that take into account the unique conditions, objects, and environments inherent to construction sites in order to meet the industry's specific requirements. Now we will examine these specialized datasets for developing machine learning and computer vision applications in the construction industry.

Tajeen and Zhu (2014) developed one of the earliest datasets in the field of construction. It consists of 2,000 images of five types of construction machinery (excavators, loaders, dozers, rollers, and backhoes) captured from various angles, poses, and sizes. The database is expected to evaluate the performance of existing object recognition methods for recognizing construction equipment. The standard,

impartial, and exhaustive evaluation results from the use of the image dataset can assist construction researchers and professionals in selecting a better or more appropriate recognition method to automate construction tasks, such as productivity analysis and progress monitoring. The article discusses the three primary types of object recognition techniques: geometry-based, appearance-based, and feature-based. At the time of its creation, deep learning detectors were not widely used to evaluate the performance of datasets.

After the deep learning algorithms were presented in the literature, (H. Kim, Kim, et al. (2018) presented a deep convolutional network-based construction object-detection method for recognizing construction equipment precisely. The proposed method employs transfer learning to train a model for the construction industry by transferring the training data and knowledge of models trained in other domains. To evaluate the generalization performance of the proposed construction equipment detection model, a benchmark dataset titled Advanced Infrastructure Management (AIM) dataset was developed for five classes: dump truck, excavator, loader, concrete mixer truck, and road roller. The experimental results demonstrate that the proposed method achieves a remarkable 96.33 percent mean average precision. By evaluating the performance of the dataset using Faster R-CNN and ResNet-101, they demonstrated the potential of deep learning algorithms for detecting heavy construction equipment.

Xiao and Kang (2021a) released the Alberta Construction Image Dataset (ACID), a specialized dataset comprising 10,000 images of construction equipment captured from the ground and ten object classes. To assess its suitability, they trained the dataset with YOLO-v3, Inception-SSD, R-FCN-ResNet101, and Faster-R-CNN-ResNet101. To demonstrate the reliability of ACID, authors employed four existing deep learning object detection algorithms, namely YOLO-v3, Inception-SSD, R-FCN-ResNet101, and Faster-RCNN-ResNet101, to train the data set. The mean average precision for Inception-SSD is 83.0 percent, YOLO-v3 is 87.8 percent, R-FCN-ResNet101 is 88.8 percent, and Faster-RCNN-ResNet101 is 89.2 percent. The

average detection speed of the four algorithms is 16.7 frames per second, which is sufficient for most construction automation research studies.

The Moving Objects in Construction Sites (MOCS) image dataset published by Xuehui et al. (2021) contains 41,668 images and thirteen categories from 174 different construction sites. The dataset was utilized to evaluate the performance of deep learning algorithms for object detection and instance segmentation. Contrary to ACID, the MOCS dataset contains images captured by UAVs flying at low altitudes around construction equipment. The dataset consists of 13 categories of construction site moving objects annotated with per-pixel segmentations to facilitate precise object localization (Figure 2.2). Using the MOCS dataset, 15 distinct Deep Neural Network (DNN) detectors were evaluated in a benchmark. The results indicate that all detectors trained on the dataset were able to precisely and reliably detect objects on construction sites. According to the study results, detecting workers and equipment through images and videos can aid in safety monitoring, quality control, and productivity management on construction sites. While the dataset does include aerial images, the creators of the dataset do not offer a top-down perspective as proposed by the dataset in this thesis.

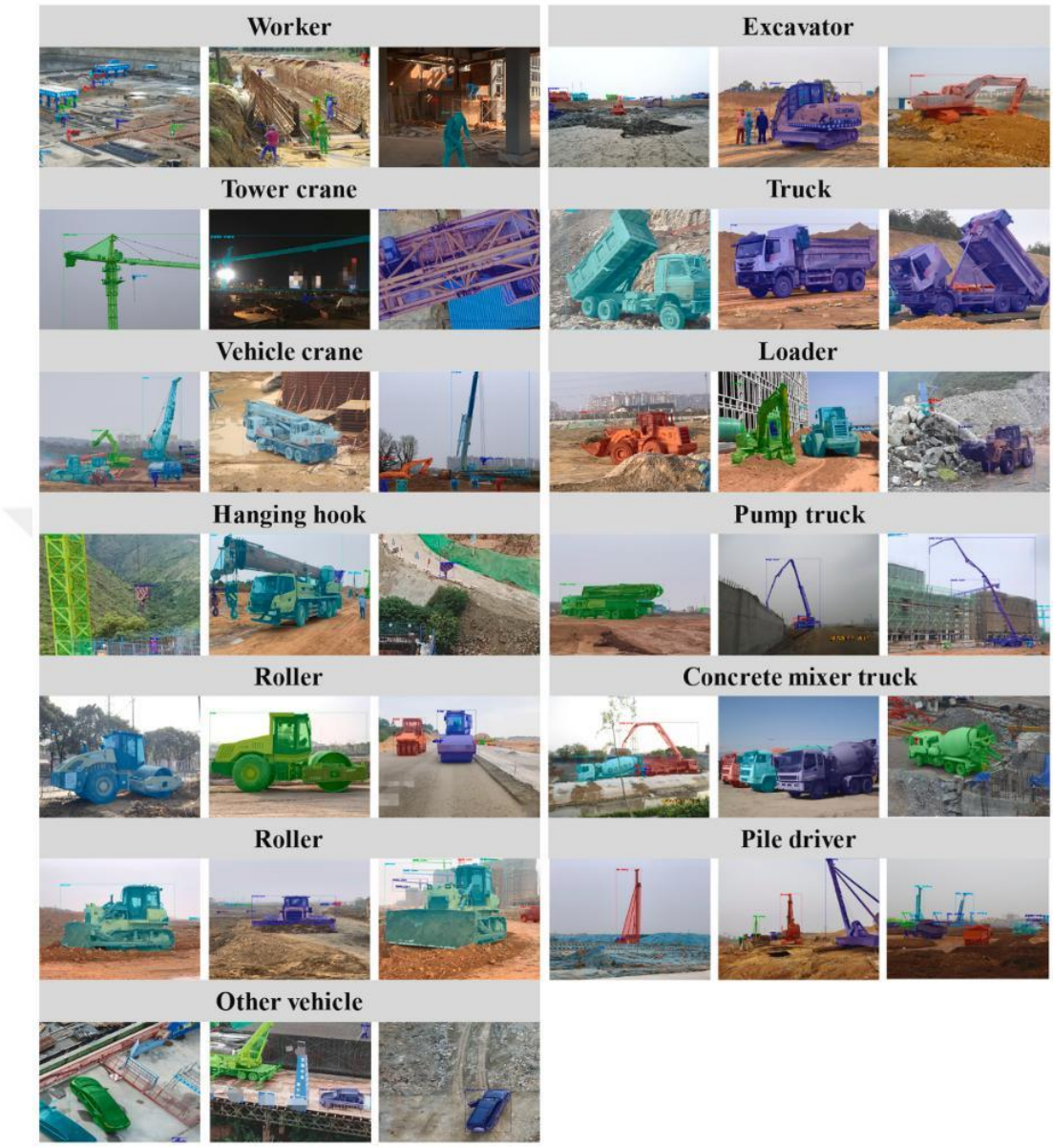


Figure 2.2. Image Examples from MOCS Dataset (Xuehui et al., 2021)

Duan et al. (2022) introduced the Site Object Detection Dataset (SODA), which includes 15 object classes, such as scaffolds, hooks, and fences, that are absent from existing construction datasets. The dataset consists of more than 20,000 images collected from multiple construction sites under varying circumstances, weather conditions, and construction phases and from a variety of angles and perspectives.

The dataset is robust in terms of diversity and quantity. Additional evaluation with two widely used deep learning-based object detection algorithms demonstrates the dataset's feasibility, achieving a maximum mAP of 81.47 percent. SODA aims to provide a valuable resource for a variety of construction-related applications.

Additionally, Del Savio et al. (2022) presented a dataset captured by four stationary cameras positioned around a construction site. The dataset comprises 1046 images extracted from video footage by a frame extraction algorithm, as well as text files containing the class and coordinates of the objects. These data can be utilized to develop computer vision techniques for use in the engineering and construction industries. The data was collected by applying a frame extraction algorithm to the videos captured by the surveillance cameras. The data is stored in a public and reputable Repositorio Institucional – Universidad de Lima repository.

Given the time-consuming nature of manually collecting and annotating large image datasets, researchers have also proposed synthetic datasets artificially mimicking the characteristics of actual data. Soltani et al. (2016) created a dataset by combining three-dimensional models of construction machines with images of construction sites' backgrounds (Figure 2.3). The location can be extracted and used for annotation using segmentation techniques. The test results demonstrate that the proposed method is able to reduce the time required to annotate images compared to conventional annotation methods while simultaneously improving detection accuracy. Bang et al. (2020) proposed an image augmentation method for constructing a large dataset to enhance the detection of construction resources. The method includes three techniques: removing-and-inpainting, cut-and-paste, and image variation. As the initial dataset for evaluating the method, 656 images of construction sites captured by unmanned aerial vehicles were utilized. The performance of a faster region-based convolutional neural network (Faster R-CNN) trained on the augmented training dataset is superior to that of a network trained on the original dataset. Hwang et al. (2022) presented an automated framework for constructing a large, high-quality image training database for construction site monitoring. The framework consists of three primary processes: (1) automated

construction image collection utilizing web crawling, (2) automated image labeling utilizing an image segmentation model, and (3) fully randomized foreground-background cross-oversampling. This study's findings can reduce the time and resources needed to develop vision-based site monitoring technologies.

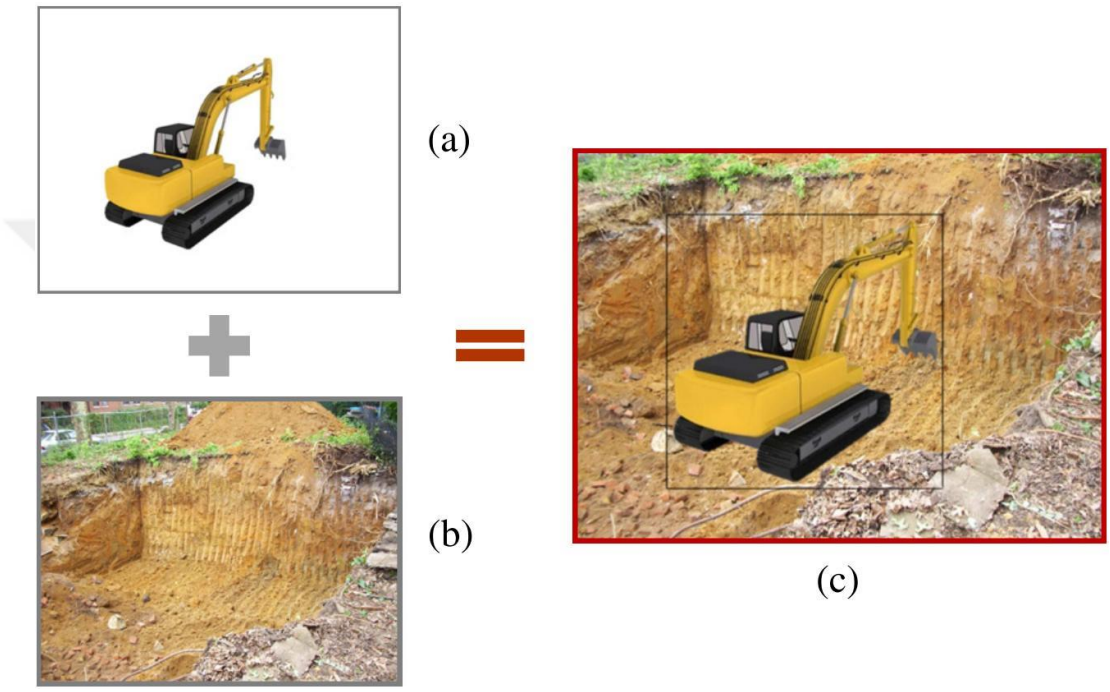


Figure 2.3. Synthetic Image of Construction Machine (Soltani et al., 2016)

These datasets, both natural and artificial, have contributed to the advancement of computer vision techniques for the engineering and construction industries. However, the absence of a publicly accessible large-scale image dataset for detecting objects on construction sites hinders research in this field. As more datasets become accessible, researchers and practitioners can utilize this wealth of data to enhance the capabilities of deep learning algorithms and computer vision technologies, thereby fostering innovations and improvements in the construction industry.

2.1.3 Applications in Construction

The availability of construction equipment-related datasets in the literature affords a variety of opportunities for diverse application areas. Identification of equipment from images is essential for progress monitoring, productivity measurements, and proactive work zone safety applications. In these applications, detecting equipment is the first step (Memarzadeh et al., 2013). In this context, deep learning algorithms have been utilized to detect workers, excavators, and dump trucks (Fang et al., 2018; Z.-H. Lin et al., 2021; Xiang et al., 2018).

Golparvar-Fard et al. (2013) proposed that a multi-class SVM classifier could be used to recognize excavator and dump truck actions from images captured by a stationary camera. Zhu et al. (2017) identified construction workers and equipment with high precision and recall. In a separate study, Luo et al. (2018) demonstrated a method for detecting 22 classes of construction-related objects and 17 types of construction activities. Using the AIM dataset mentioned in the previous chapter, Roberts and Golparvar-Fard (2019) developed a method for determining the operational activities of excavators and dump trucks. Xiao and Kang (2021b) also proposed the construction machine tracker (CMT) system for tracking multiple construction machines using image hashing features and the ACID dataset. By employing object detection and tracking algorithms, it is possible to identify and quantify critical productivity aspects of construction equipment, such as working cycles, cycle times, and delays (C. Chen et al., 2020; Gong & Caldas, 2011; J. Kim & Chi, 2019). H. Kim, Bang, et al. (2018) estimated earthwork productivity using a CNN-based technique to detect excavators and dump trucks.

Understanding the equipment's position in terms of productivity simplifies the estimation of the amount of time operators spend on each task stage (Rezazadeh Azar & McCabe, 2012b). Soltani et al. (2017) introduced a method to process images from each camera on the construction site, create a 2D skeleton of the excavator, and use the relative rotation and translation data between the cameras' coordinate systems to infer the 3D posture. Mahmood et. al (2022) created a synthetic dataset in order to

train a vision-based model to predict the 3D posture of an excavator. Safety monitoring and excavator activity analysis are facilitated by 3D pose detection. (S. Zhang & Zhang, 2022). Chi and Caldas (2011), Rezazadeh Azar and McCabe (2012a) introduced object recognition and background subtraction algorithms focused on proactive automation of work-zone safety. Rezazadeh Azar et al. (2013) and J. Kim et al. (2018) developed a framework for identifying vision-based tasks that account for the interactive characteristics of earthmoving machinery operation.

In aerial imaging, the introduction of UAVs has led to the widespread use of equipment detection and tracking (Guo et al., 2020). Meng et al. (2020) proposed a real-time detection of excavators for pipeline safety using the widely adopted YOLO v3 algorithm and 350 images captured by a UAV DJI M600 Pro. UAVs are primarily utilized for proximity monitoring applications because aerial views provide more information from more extensive areas than ground images. D. Kim et al. (2019) collected 4,512 frames using UAVs, which include excavators, wheel loaders, and workers and proposed a YOLO-v3-based automated proximity measurement technique. Similarly, Bang et al. (2021) developed a Mask R-CNN-based method for segmenting, predicting, and monitoring the proximity of workers and heavy equipment.

In summary, the literature provides a variety of datasets and methodologies related to construction equipment, allowing for a wide range of applications. Advanced techniques such as deep learning algorithms, object detection algorithms, and tracking algorithms have proven to be extremely useful for detecting equipment in images, measuring productivity, and enhancing work-zone safety. Incorporating UAVs has also increased the capabilities of aerial imaging for equipment detection, tracking, and proximity monitoring. As the construction industry evolves, these datasets and methodologies will play a crucial role in enhancing efficiency, productivity, and safety. While existing literature presents a variety of datasets and methodologies for construction equipment detection and its diverse applications, a significant gap still remains. No substantial research has been done on detecting and removing construction equipment for accurate earthwork volume calculation. This

dissertation addresses this neglected domain, highlighting the need for innovative strategies that bridge this gap in our understanding and applications.

2.2 Object Removal using Image Inpainting

Image inpainting is a method that uses accessible image information to infer and fill in missing areas. This innovative technique has numerous potential applications, including enhancing image resolution, removing obstructions from images, and repairing damaged parts. This dissertation aims to detect construction machinery in images and subsequently remove them. It is essential to investigate the capabilities of image inpainting methods for the purpose of construction machine removal.

In the proceedings of SIGGRAPH 2000, Bertalmio et al. presented "Image Inpainting," the earliest work in this field (Bertalmio et al., 2000). The researchers utilized a novel technique in which they propagated isophote lines, i.e., lines of equal pixel intensity from the exterior to the interior of the damaged or missing area (Figure 2.4). This study established a solid foundation for future image inpainting research.



Figure 2.4. Image Restoration and Text Removal (Bertalmio et al., 2000)

Classical (non-learning-based) and learning-based methods are the primary categories of image inpainting techniques. Classical techniques primarily relied on image characteristics such as continuity, smoothness, and self-similarity to extrapolate information from neighboring regions into missing sections. These methods typically involve solving optimization problems, minimizing an objective function defined on the previously stated image properties, and typically do not require extensive datasets for training. Methods such as diffusion techniques (Ballester et al., 2003; Levin et al., 2003; Tschumperlé & Deriche, 2005), exemplar-based methods (Buatik & Chaiyasarn, 2021; Engels et al., 2011; Fan et al., 2016; Witherow et al., 2019), and hybrid methods (Bertalmio et al., 2003; Starck et al., 2005) are widely employed. These methods are effective but do not scale well to larger masks in terms of both quality and execution time (Kapoor et al., 2019).

Recently, deep learning has emerged as a significant trend in computer vision. Its application to image inpainting has effectively remedied the long-standing weaknesses of classical inpainting techniques, thereby significantly enhancing the output quality. In image inpainting, the deep learning approach enables the accurate extraction of semantic information from images, thereby providing a more logical prediction of the missing content. Convolutional neural networks (CNNs) and Generative Adversarial Networks (GANs) have demonstrated proficiency in capturing abstract image data. Numerous researchers have utilized CNNs to address image inpainting issues more effectively, resulting in significant advancements (Liu et al., 2018; Yu et al., 2019). Commonly used CNN modifications include Fully Convolutional Networks (FCN) (Long et al., 2015) and U-Nets (Ronneberger et al., 2015). A popular deep-learning approach proposed by Goodfellow et al. (2014) introduced GANs, which are well-suited for image inpainting tasks due to their robust data generation capabilities.

2.2.1 Inpainting Datasets

Upon reviewing the wide variety of image inpainting methods, as object detection algorithms, their performance frequently depends on the quality and relevance of the training and evaluation datasets. The selection of an adequate dataset is critical to the development and performance of inpainting models since it offers the necessary context and learning examples for the algorithms. As a result, researchers have produced numerous image inpainting datasets, each tailored to unique image kinds, situations, or inpainting difficulties. In the following paragraphs, we will cover some of the most often-used image inpainting datasets in computer vision.

ImageNet: Although not explicitly designed for image inpainting, the previously mentioned ImageNet, a large-scale dataset containing over 14 million images across 20,000 categories is commonly used for pre-training deep learning models prior to fine-tuning on specific inpainting tasks (Russakovsky et al., 2015).

Places2: This dataset, created by B. Zhou et al. (2014), consists of over 10 million images of scenes and locations, spanning over 400 scene categories. Places2 has been extensively utilized for the training and evaluation of image inpainting algorithms, mainly when dealing with large contextual regions and various scene types (Figure 2.5).



Figure 2.5. Sample Images From Places2 Dataset (B. Zhou et al., 2014)

Paris StreetView: Paris StreetView was developed by Pathak et al. (2016) and contains street view images of Paris with artificially generated missing regions of varying sizes and shapes. The dataset has been used to train and evaluate a variety of inpainting algorithms, particularly those geared toward urban scenes.

CelebA-HQ: This high-quality dataset, introduced by Karras et al. (2018), consists of 30,000 celebrity face images with a resolution of 1024x1024. The CelebA-HQ database has been utilized for training and evaluating image inpainting algorithms centered on facial images and fine-grained details.

2.2.2 Inpainting Applications

Based on its functionality, image inpainting technology can be divided into three categories: image completion, image generation, and object removal. Image completion emphasizes filling in missing areas with probable information. On the

basis of image generation, an algorithm generates a new image. Object removal is the process of eliminating a specified image object and replacing it with feasible content.

This thesis predominantly focuses on studies related to object removal, and the applications of this specific domain are elaborated in the succeeding paragraphs. Typically, the object removal process consists of three steps: object detection, object removal, and content inpainting. Earlier classical techniques outlined the object region manually and synthesized texture using statistical inpainting methods. Despite being time-consuming and incapable of ensuring global semantic consistency, these techniques have been significantly enhanced by deep learning, which can accurately identify objects to be removed in the vast majority of scenarios.

A broad range of applications for classical image inpainting methods can be found in the literature. Exemplar-based image inpainting (Criminisi et al., 2003) is a widely applied technique in image processing, used to complete missing or corrupted parts of images by using entire areas as "exemplars" or "patches". This method identifies patches in the uncorrupted region that closely resembles the area around the missing part based on color, texture, and other visual attributes. These patches are then skillfully blended to fill the gap, resulting in a seamless appearance. One pioneering study aimed at improving 3D urban reconstruction from street-level imagery used this technique to detect and remove vegetation and vehicles automatically from the imagery (Engels et al., 2011).

In another instance, exemplar-based inpainting was utilized to fill the holes in stereo images caused by moving objects. The process proved effective in mitigating the impact of moving objects on the 3D map (Fan et al., 2016). Witherow et al. (2019) employed this method to address occlusion problems and eliminate vehicles from crowdsourced images of flooded conditions. The algorithm successfully filled in the occluded areas behind the vehicles, improving the registration accuracy and creating a more robust and dynamic flood detection method.

Recently, an exemplar-based inpainting algorithm was used to remove deformable obstacles, such as trees, people, and construction sites, from 2D images in a project aimed at reconstructing a high-quality, image-based 3D model of a heritage structure. The results demonstrated that removing these objects before creating the final 3D model led to more accurate results and enhanced visualization of heritage structures (Buatik & Chaiyasarn, 2021).

On the other hand, deep learning-based inpainting has demonstrated considerable promise across several domains, including urban modeling, shadow manipulation, construction management, and infrastructure planning. (Kapoor et al., 2019) introduced Nostalgin, a tool capable of recreating 3D city models from historical photographs. Using inpainting techniques, missing data is filled in, and the model is refined to create a more accurate representation of the past. Wei et al. (2019) demonstrated a two-stage Generative Adversarial Networks (GAN) technique that focuses on shadow inpainting and removal. It is highlighted for its ability to improve shadow removal and color retention in non-shaded areas, which represents a significant improvement over existing techniques.

The implementation of inpainting techniques has also led to significant advancements in the construction field. Bang et al. (2020) presented a method for enhancing construction resource detection in UAV-acquired images through image enhancement. The proposed image inpainting method employs GANs to remove and reconstruct specific image objects, resulting in enhanced object detection outcomes. In a similar vein, Angah and Chen (2020) proposed a context inpainting method employing the Context Encoders model and the U-Net architecture, designed to remove obstructing objects such as construction workers from images of construction sites (Figure 2.6). In addition to providing clearer site views, this method facilitates the creation of Building Information Models.



Figure 2.6. Steps of Construction Worker Removal (Angah & Chen, 2020)

W. J. Kim et al. (2022) introduced an innovative method using inpainting to improve the detection of moving objects in the field of object detection. By enhancing background detail at the pixel level, this technique outperforms conventional techniques, especially in environments with sudden changes.

Regarding infrastructure planning, J. Park et al. (2022) recommended a technique for creating vehicle-free orthomosaics from UAV imagery. Using image segmentation and inpainting algorithms based on deep learning, unwanted objects such as vehicles are effectively removed. This method is implemented before the Structure from Motion (SfM) processing step, which enables the creation of inpainted UAV image-based orthomosaics. This innovative inpainting-focused technique offers potential improvements in transportation infrastructure management, highlighting the adaptability and effectiveness of inpainting in various contexts.

In conclusion, image inpainting, specifically object removal, has significantly progressed from traditional to advanced deep learning approaches. Early research relied primarily on conventional techniques, such as diffusion techniques, exemplar-based methods, and hybrid methods. However, these techniques had scalability and quality limitations, especially with larger masks. Deep learning has revolutionized the field of image inpainting, addressing many of these challenges and significantly enhancing output quality. In this thesis, we will employ image inpainting techniques based on deep learning to detect and remove construction equipment from images in order to represent the construction field accurately.

Transitioning from the limitations and opportunities provided by image inpainting-based object removal, it is evident that this process represents a crucial step in ensuring the accuracy of earthwork measurements. Effective removal of non-terrain objects from UAV-collected images paves the way for more accurate earthwork calculations. Despite the fact that object removal aids in the preparation of our data, it is only one component of earthwork management as a whole. The ultimate application of our data is the calculation of earthwork volumes. In the following section, we will examine various methods for calculating earthwork. We will investigate the various methodologies, their feasibility, and how they contribute to earthwork management's practical tasks.

2.3 Earthwork Progress Monitoring Approaches

The estimation of earthwork is crucial to the planning and execution of construction projects, affecting project costs, schedules, and safety. The methods used for calculating the volume of earthwork have evolved significantly over time, shifting from manual measurements to sophisticated technologies capable of capturing highly accurate data. This section examines these modeling approaches, examining the industry-standard techniques. It divides these techniques into two broad categories: laser-based and photogrammetry-based techniques. Each section provides an overview of the respective techniques, providing a comprehensive understanding of their operations, advantages, and disadvantages.

2.3.1 Laser-based Approaches

This section explores laser-based methods for measuring and monitoring earthwork volume in the construction industry. Laser-based technologies, which utilize Light Detection and Ranging (LiDAR), have revolutionized the collection of high-resolution topographic data and have found widespread use in a variety of fields. LiDAR is a remote sensing technology that utilizes laser light pulses to measure

distances to a surface, allowing for the generation of three-dimensional maps and models. Laser-based methods offer numerous advantages, such as improved precision and high data density. However, they bring with them their own set of difficulties and complexities. In the following paragraphs, we will examine in depth the application, benefits, potential limitations, and most recent developments of laser-based methods for monitoring earthwork.

The National Institute of Standards and Technology (NIST) in the United States developed one of the earliest methods for monitoring the status of typical earthmoving operations on construction sites. Utilizing auto-registered LiDAR range sensing systems, the method provides useful quantities to the earthmoving contractor on the actual construction site. The system's accuracy and reliability have been demonstrated on multiple construction sites. The aim was to save time and money by eliminating manual measurements and enhancing safety by decreasing the need for workers to enter hazardous areas (Stone et al., 2000).

Building on this technological advancement, Du and Teng (2007) introduced another impactful application of modern sensor technologies. They presented a case study of the use of 3D laser scanning and GPS technology to collect data on landslides and calculate the volume of earthwork. A 3D laser scanner, Dibit Geoscanner in Figure 2.7, was used to create a detailed 3D model of the area affected by the landslide. At the same time, GPS technology was utilized to determine the coordinates of the 3D points within this model. A landslide in Taiwan was analyzed in a practical application of this method. The amount of earthwork required for landslide remediation was estimated by subtracting created digital terrain model and the 3D contour of the area before the landslide. The authors concluded that 3D laser scanning and GPS technology constitute a valuable, accurate, and efficient tools for assessing landslide remediation needs in various environmental conditions.



Figure 2.7. 3D Laser Scanner (Du & Teng, 2007)

Further expanding on the use of these technologies, Yakar et al. (2010b) evaluated the use of terrestrial laser scanning (TLS) and total station (TS) for determining the volume of excavation areas. Comparing the two, the authors evaluate each technique's precision, efficiency, and cost. By creating a 3D point cloud of the excavation area in less time, they determined that TLS provided greater accuracy and efficiency than TS, as demonstrated by their study of an actual excavation site. TLS was found to be more expensive. Although both methods were considered acceptable for volume computations, the authors highlighted the importance of selecting the method based on the project's location, area size, required accuracy, budget, and schedule. TLS was recommended when accuracy and efficiency were the most significant priority, while TS was recommended when cost was the primary concern.

Continuing the exploration of advancements in this field, Chae et al. (2011) presented the creation of a mobile 3D work environment recognition system for civil earthwork, a component of the Intelligent Excavation System project aimed at automating earthwork. A car-mounted laser scanner, named the Mobile 3D Imaging System, was developed to control a 3D laser scanner. The scanner was installed in a modified mini-van to allow mobility across the work site. The system used sphere targets for point cloud registration to generate a 3D surface model of the excavation area. Five sphere targets (Figure 2.8), each 15 cm in radius, were specifically manufactured and positioned around the earthwork site for this purpose. Having been tested on several sites, the system proved effective in enhancing the efficiency and safety of the excavation process, reducing the necessity for manual measurements, and limiting workers' exposure to hazardous areas. However, using such 3D surface modeling systems has limitations, including high costs for specialized equipment, time-consuming setup, and challenges in difficult-to-access areas.



Figure 2.8. Sphere Target for Laser Scanning (Chae et al., 2011)

Slattery et al. (2012) examined the TLS technology in the construction industry, focusing on its capacity to precisely determine earthwork quantities. The authors describe a method in which the existing terrain was meticulously scanned using available topographic data as the basis for planning these scans, thereby ensuring thorough coverage of the area of interest with the fewest possible scans. Using existing contour data of the topography, a software tool was employed to determine the most optimal locations for the scanners. Before and after the construction phase, this extensive scanning procedure was performed. Triangulated Irregular Network (TIN) files were then generated from the scanned surfaces. By comparing the TIN of the pre-construction terrain to that of the completed project, the authors were able to precisely calculate earthwork quantities, demonstrating the effectiveness of TLS technology for efficient construction planning and management.

In a more specific application of these advanced technologies, Hu et al. (2015) discussed the application of laser scanning technology for evaluating excavation quality and calculating excavation volume in hydraulic projects. The research was conducted in two phases. Initially, the authors created a 3D model of a slope using laser scanning technology. This 3D model was subsequently used to evaluate the quality of the excavation and to calculate the volume of the area excavated. Remarkably, the volume calculated using the scanned data showed only minor deviations from the measured data, demonstrating the validity of the volume calculation method proposed in this study. Utilizing laser scanning technology in this context was advantageous due to its efficiency. This technology reduced excavation quality assessment time by 70 to 90 percent compared to conventional methods, demonstrating its potential to transform practices in hydraulic projects.

Despite obstacles such as high cost, the need for skilled operation, and complexities in data processing, laser-based technologies have demonstrated significant promise in the field of earthwork monitoring. As technology advances, it has been suggested that these tools should be refined further in order to overcome these obstacles while maximizing their capabilities. As with laser-based instruments, photogrammetry-based systems offer distinct advantages for earthwork measurement. Notable

benefits of photogrammetry include cost-effectiveness and usability. In the subsequent section, the role, advancements, and implications of photogrammetry-based technologies in earthwork monitoring and management will be investigated in greater depth.

2.3.2 Photogrammetry-based Approaches

In the field of earthwork measurement and monitoring, photogrammetry-based systems, which are frequently viewed as a viable alternative to laser-based technologies, are gaining popularity. These systems utilize the process of constructing three-dimensional models or maps from two-dimensional images or video frames. Photogrammetry's affordability, accessibility, and user-friendliness have led to a dramatic increase in its use over the years, particularly by Unmanned Aerial Vehicles (UAVs). This section intends to concentrate on the aspects of photogrammetry-based methods, examining their capabilities, their applications in earthwork management, and the inherent challenges they face while highlighting the recent developments in the field.

Pioneering the incorporation of photogrammetry in earthwork management, (Yilmaz & Yakar, 2008) made significant contributions to the field of earthwork by proposing one of the first photogrammetry-based applications, namely digital close-range photogrammetry (DCRP). This technique uses photographs to create a 3D model of an inaccessible or potentially hazardous excavation area, which can then be used to calculate the volume of the area. The researchers then compared the volumes determined from these 3D photo models with those determined using total stations. For this technique, 36 test targets were randomly distributed on the object's surface before and after excavation, and their geodetic coordinates were measured using a reflectorless Topcon GPT 7PT total station. According to the results, the photogrammetric method offers a 21.89 percent time efficiency advantage, a 12.11 percent accuracy improvement, and a 33.33 percent cost efficiency advantage over the conventional method. This assertion of the DCRP method's precision and

efficacy in computing excavation volumes was supported by a similar comparative study conducted by Yakar et al. (2010a), who concluded that both methods could be used for volume calculations. However, they noted that DCRP is more precise, effective, and expensive. Therefore, they recommended DCRP for projects where accuracy and efficiency were crucial and total stations for projects where cost was the most critical factor.

In a separate study, Nassar and Jung (2012) investigated the use of photogrammetry techniques in earthworks in an effort to identify its advantages and disadvantages. The findings confirmed that photogrammetry is an effective and precise technique. Nonetheless, its application is limited by a number of factors. First, the extent of the excavation must not exceed the camera's capture range in order to achieve an acceptable resolution. Secondly, the technique's success depends on the presence of numerous and easily accessible viewpoints surrounding the site, as this is essential for taking the necessary number of photographs.

The introduction of UAV technology has revolutionized the photogrammetry field, overcoming several of its previous limitations. According to the Siebert and Teizer (2014), UAV systems can be utilized as data acquisition platforms and measurement instruments for civil engineering surveying applications. The combination of hardware, control software, and a capable program for photogrammetric flight planning makes it possible to generate highly accurate 3D point clouds from images. In addition, a performance model for estimating the position error was developed and tested in various realistic construction environments, further enhancing the data's reliability. Significantly, the authors identified UAVs as an essential element for the rapid and efficient creation of accurate and reliable 3D models for earthwork projects. However, they identified one persistent drawback: the need to improve UAV battery life to extend flight duration and increase data collection resolution.

Transitioning to a comparative study, (Hugenholtz et al., 2015) highlighted the potential value of small UAVs for volumetric analysis in small to moderate earthwork projects. By conducting a case study on stockpile volume estimation, the

authors could use softcopy photogrammetry to evaluate the accuracy of Digital Terrain Models (DTMs) derived from UAV imagery. The results demonstrated that the vertical accuracy of the UAV-derived DTMs was comparable to, if not better than, total station methods, although less precise than TLS. In addition, UAVs were highlighted as viable alternatives to ground-based methods, particularly when stockpiles were either too dangerous for physical surveys or possessed complex geometries that prevented uniform coverage. Despite the challenges, the authors concluded that the combination of UAVs and softcopy photogrammetry offers significant value in the estimation of earthwork volumes, bringing notable improvements in efficiency, accuracy, and safety.

In their respective studies, Cryderman et al. (2014) and Tamin et al. (2019) presented a compelling comparison between the accuracy and reliability of UAVs and GNSS-RTK (Global Navigation Satellite System Real-Time Kinematic) surveys in the ever-changing landscape of geospatial data acquisition. Cryderman et al. (2014) evaluated an in-house developed UAV system in a volumetric stockpile survey and determined that the results were, at a minimum, equivalent to those obtained from GNSS-RTK, a high-precision navigation method typically employed in topographic surveying. The UAV photogrammetry surveys achieved levels of accuracy suitable for 1:200 scale mapping and 0.145 m contours, resulting in more accurate and representative models of the surveyed surfaces. Concurrently, Tamin et al. (2019) compared the use of UAVs and GNSS-RTK for DEM generation, focusing on their capabilities for estimating earthwork volume in road planning and construction. Comparatively, GNSS-RTK surveying yielded a point density of 35 points/ha, whereas the UAV-based method yielded a point density of approximately 234,385 points/ha. Moreover, the volumes of excavations and embankments were found to be nearly identical when using the UAV-based DEM as a reference plane for the average excavation per unit of road length. This led to the conclusion that UAV-based surveys had more reliable resolution and point density values than GNSS-RTK-based DEM.

Delving further into the comparison between UAVs and other technologies, Tamin et al. (2019) have compared the precision of UAV-based imagery to laser scanning

techniques. The imaging was performed with a UAV, and a detailed survey was conducted with TLS to verify and compare the accuracy of both techniques. The TLS method was found to be more time-consuming than the UAV-based strategy, as indicated by the results. In addition, both methods successfully generated DEMs, and the UAV method was able to generate contours, DEMs, and volume calculations with a relative error of 0.002 percent. Notably, the use of UAVs demonstrated several practical advantages, including simple deployment, fewer personnel needs, safety during data collection, and rapid acquisition of aerial images. Comparatively, the TLS encountered limitations when surveying the upper 20-meter portion of the stockpile. These results highlight the significant benefits of using UAVs for monitoring earthwork progress.

Recent research has focused on evaluating the performance of UAV photogrammetry in monitoring earthwork progress in the construction industry, highlighting the distinct benefits and identifying the conditions required for accurate application. A study by Julge et al. (2019) demonstrated that the accuracy of UAV photogrammetry depends on a number of factors, such as the quality of the captured images, the characteristics of the surface being measured, the strategic placement of Ground Control Points (GCPs), the degree of photo overlap, and the altitude of the flight. The research provides a foundation for understanding the critical parameters that govern the efficient use of UAVs in earthworks. In conjunction with this, Wierzbicki and Nienaltowski (2019) supported the adoption of UAV systems, emphasizing their inherent utility. They enable continuous operation without the need to stop work on the excavation site, eliminate the need for personnel to enter potentially hazardous areas, and facilitate the measurement of otherwise inaccessible areas.

Expanding on these practical benefits, Ajayi and Ajulo (2021) highlighted the additional advantages of UAV photogrammetry, including its portability due to its compact size, cost-effectiveness, time efficiency, and the ability to acquire images with high spatial resolution. Importantly, Ajayi and Ajulo note that UAV photogrammetry provides a high density of points in 3D coordinates, thereby enhancing the precision of volume estimations — a crucial aspect of earthwork

operations. H. C. Park et al. (2022) presented a case in which the initial estimation of excavation volume, based on GPS data and boring test results, was significantly higher than the volume measured by UAV photogrammetry after the excavation. The resulting cost difference was a staggering USD 183,057, representing a disparity of 15.8 percent. In addition to highlighting the potential for cost savings, these findings highlight the role of UAV technology in promoting ethical business practices and transparency among stakeholders. Therefore, UAVs are becoming an indispensable tool for monitoring and managing earthworks.

Only a limited number of papers in the literature have addressed the issue of non-terrain objects affecting the accuracy of volumetric difference calculations in UAV-based earthwork monitoring. Cho et al. (2021) highlighted the difficulty of visualizing height differences due to the presence of moving objects, such as cars and construction machines, in point cloud data. While the researchers observed that the difference in net volume between models with and without moving object removal was relatively small, the presence of these objects had a significant impact on the generation of height differential maps. On the other hand, W. J. Kim et al. (2022) proposed a method for calculating earthwork volume by comparing 3D models generated by UAV mapping. The authors of the study emphasized that temporary facility objects included in UAV mapping results can cause errors in earthwork volume calculations. This indicates that future developments in UAV-generated 3D model technology should concentrate on more precise field measurements and improved methods for detecting and interpolating outlier objects. Both studies highlight the critical need for refining current methodologies to reduce inaccuracies caused by non-terrain objects in UAV-based photogrammetry.

In conclusion, photogrammetry-based systems have demonstrated tremendous potential for measuring and monitoring earthworks. They have significantly advanced the field by providing affordable, accessible, and reliable solutions. However, obstacles such as the problem of existing and temporary objects restrict the perfect accuracy of the derived measurements. Future developments should focus on improving object detection and removal techniques, paving the way for more

accurate and trustworthy earthwork volume calculations. At this point, this dissertation offers an innovative solution for non-terrain objects, which is discussed in the upcoming chapters.

2.4 Volume Calculation Methods

After reviewing the various techniques used to create a 3D representation of the earthwork area, it is clear that an accurate volume computation method is required for a complete earthwork estimation framework. While the significance of accurate volumetric computations is universally acknowledged, numerous methods have been proposed, each with its own unique advantages and disadvantages. In this section, we will concentrate on various volumetric computation methods, focusing on their applicability, benefits, and areas for improvement.

Oloufa (1991) introduced one of the first surface-based volume calculation methods that employs triangulation, specifically Delaunay triangulation, for calculating soil volumes. Traditional methods, such as the end-area method and the prismoidal method, need plenty of time and computational power, particularly for larger sites, and their accuracy can be compromised in areas with irregular boundaries or sites with multiple soil types. In contrast, this innovative triangulation method offers a variety of advantages, such as increased accuracy, adept handling of diverse soil types, faster development of elevations, and overall increased efficiency. The authors have also developed and tested a program to implement this technique across multiple sites; the results confirm the technique's ability to calculate volumes accurately and efficiently. In their study, Hintz and Vonderohe (2011) found that the average-end-area method and the surface-to-surface method produced noticeably different values, with the discrepancies reaching up to 5 percent when the cross-section interval was 30.48 meters, translating to differences in construction costs of $112,500 for fill and $95,800 for the cut in one of the tested data sets. Contreras et al. (2012) investigated the effects of various cross-section spacings on the accuracy of earthwork volume estimation, assuming that a 1-meter spacing provides the "true"

earthwork volume. Results indicate that as cross-section spacing increases, the estimation of earthwork volume becomes less accurate due to an inability to account for terrain variations. In order to improve the accuracy of earthwork volume estimation when roads are planned and located on hilly and rugged terrain, short cross-section spacing should be utilized.

In the 2010s, the average-end-area (2D) method for calculating earthwork volume in roadway engineering was widely accepted in the literature. However, CAD and modeling software developments have paved the way for a shift towards 3D or surface-to-surface methods. Cheng and Jiang (2013) have investigated the accuracy of these two methods and discovered that the 3D method has distinct advantages. Notably, while the 2D method was more accurate in flat terrain, the 3D method consistently demonstrated greater accuracy regardless of the terrain's flatness. The researchers drew the conclusion that the 3D method could be easily implemented using CAD software. In contrast, the 2D method could still provide adequate accuracy if the distance between two consecutive cross sections was maintained below a certain threshold. However, the characteristics of the terrain play a substantial role in this calculation. In support of this idea, Tanoli et al. (2018) proposed a method for developing 3D earthwork section designs using a parametric modeling approach. They calculated the volume of earthworks by superimposing and differentiating the TIN surfaces of the existing ground and final design, proving that the surface-to-surface method is more accurate than the average end area method.

Comparing the surface-to-surface method to the average-end-area method cannot definitively determine the surface-to-surface method's true accuracy because both methods have their own margin of error. Therefore, researchers have proposed novel methods to assess the accuracy of volumetric calculations. Samad et al. (2012) proposed the use of an artificial object to evaluate the accuracy of the digital close-range photogrammetry method for volume determination. The volume of the artificial object, a concrete cube block with dimensions of 145 mm for width, height, and length, was determined by hand using a ruler. Subsequently, 12-megapixel digital images of the cube were captured from various distances using a digital

camera, and the digital close-range photogrammetry method was used to calculate the cube's volume. The accuracy reported was 99 percent. A study conducted by Ajayi and Ajulo (2021) in a mining region compared the volume estimated from a UAV-based DEM to the volume obtained from a mill-machine, which is considered the standard. The results revealed a slight variance of 2.94 percent between the methods and the actual volume. This indicates that the UAV approach, which yielded a cumulative volume of 2686.252 m^3 compared to the mill machine's 2750 m^3, is accurate and time-efficient. This enhances the potential of UAVs to serve as a robust alternative to conventional, more time-consuming ground survey methods.

In conclusion, this literature review has investigated the critical domains involved in earthwork management, focusing specifically on object detection using deep learning, image inpainting-based object removal, earthwork calculation approaches, and volumetric computation methods. Each of these elements plays an essential role in earthwork estimation, offering significant advantages and posing unique challenges. Despite the progress in the literature, significant obstacles persist and must be addressed for further development. To address these issues, this thesis presents a novel framework for monitoring earthwork progress, which aims to streamline the process and improve the accuracy and robustness of earthwork estimations. Starting from the next chapter, the details and implementations of this innovative framework are elaborated, demonstrating its potential to transform the industry's current practices.

2.5 Current Gap in the Literature

The construction industry is continuously evolving due to the integration of new technologies that streamline and improve site operations. However, despite these advancements, there are notable gaps in the literature that point to areas where additional innovation could result in significant improvements. The section that follows will discuss these gaps, with a particular emphasis on the following areas:

- ***Absence of Aerial Construction Machine Image Dataset with Pixel-level Annotations and Machine Learning Model:*** The use of UAVs in construction is rising. However, there is a notable gap in the literature concerning the application of aerial image datasets and machine learning in construction. Current research frequently focuses on machine learning models trained on simplified datasets that do not adequately represent the complexity and diversity of actual construction environments. The potential of pixel-level annotations for intricate machinery recognition, which can significantly contribute to applications such as construction progress monitoring, is another overlooked aspect.

- ***Lack of Methodologies Utilizing Image Similarity and Proximity-based Clustering for Data Partitioning:*** Another critical lack in the literature is the absence of methods that employ image similarity and proximity-based clustering for data partitioning. Existing methods commonly employ random partitioning, which may result in overfitting and a lack of robustness in model performance. A novel approach employing image similarity and proximity-based clustering could result in more accurate and reliable machine learning models by ensuring a more evenly distributed, diverse training and testing dataset.

- ***Deficiency of Image Inpainting for Machine Removal and Terrain Generation:*** In the construction domain, there is also a lack of research on the application of image inpainting techniques for machine removal and terrain generation. Current methods for representing construction terrains frequently ignore the presence of machinery, which can significantly compromise the accuracy of these models. Implementing an image inpainting technique could improve visual representations and provide a more precise depiction of the terrain, thereby enhancing earthwork monitoring and volumetric analysis.

- ***Absence of a Completely Autonomous Pipeline for Earthwork Progress Monitoring:*** Despite the increasing use of automation in construction, there is a significant gap in the literature regarding the creation of a fully autonomous pipeline that manages the entire process, from data collection to volume calculation. Most existing systems require significant human intervention at various stages, which can introduce errors and inefficiency. A pipeline with complete autonomy would streamline operations, reduce the possibility of human error, and increase the process's speed and accuracy.

In conclusion, this thesis has conducted an in-depth examination of some of the most significant gaps in the existing literature on earthwork progress monitoring. The proposed method has significantly addressed these deficiencies by developing a fully autonomous pipeline for construction site monitoring using UAV photogrammetry, machine learning, and computer vision algorithms. By systematically addressing these gaps, this study has significantly advanced the understanding and practice of construction progress monitoring, laying the groundwork for future research and increasing efficiency and precision in construction site management.

CHAPTER 3

DEVELOPMENT OF MACHINE LEARNING MODEL

3.1 Introduction

The dynamic nature of construction environments and the diverse array of moving objects, such as machinery and human resources, present unique challenges for effective construction monitoring. The advent of aerial imaging technology, specifically Unmanned Aerial Vehicles (UAVs), allows a comprehensive and efficient analysis of construction sites. The aerial perspective provided by UAVs enables a thorough understanding of the site's layout, machinery performance, and overall project progress.

Despite the considerable potential of UAVs for monitoring progress, calculating earthwork volumes still presents challenges. One major obstacle is the presence of moving objects, mainly construction machinery, in UAV images, which can lead to inaccuracies in ground-level measurements. To ensure precise volumetric results, it becomes crucial to remove these objects. Nonetheless, the elimination process involves a selection procedure, which, when done manually, can be time-consuming and susceptible to human error.

Applying machine learning algorithms, specifically deep learning (DL), has shown promising results in object recognition tasks across multiple domains, including construction site monitoring, which aims to automatically identify and segment various types of machinery in an aerial image. However, developing and validating such algorithms requires the availability of large and well-annotated image datasets, which need to be improved in the context of aerial construction imagery.

To bridge this gap, this chapter presents an aerial image dataset, a comprehensive collection of aerial images from various construction sites meticulously annotated at

the pixel level. This dataset consists of 2155 images depicting 9563 construction machines across nine categories, providing an extensive resource for developing, training, and refining machine learning algorithms for construction site management.

This chapter will focus on creating datasets, exploring their unique attributes and potential applications. Additionally, we will delve into the evaluation of cutting-edge deep learning (DL) algorithms using this dataset, analyzing their performance and pinpointing areas for enhancement. We will identify the most effective machine learning model employed in the forthcoming construction machinery removal process by considering the evaluation metrics.

This chapter begins with an introduction of the methodology and processes involved in the collection and annotation of the dataset, including specifics on the image collection and annotation techniques used. In addition, the chapter describes the statistics of the dataset, performance evaluation metrics, and the outcomes of DL algorithms applied to the dataset. The chapter finally discusses the dataset's limitations and suggestions for future enhancements.

3.2 Image Collection

Over seven years, from 2015 to 2022, aerial images were collected from 25 different locations in Türkiye. As presented in Figure 3.1, these images, captured at 10 to 150 meters, provide a comprehensive view of the construction sites. The team members of Ai2LAB, the Applied Innovative and Interdisciplinary Research Laboratory, one of the research labs of the Middle East Technical University's Civil Engineering Department, predominantly obtained the earliest images in the dataset.

The primary purpose of acquiring these images was to support research endeavors focused on construction monitoring, productivity analysis, and quality control (Figure 3.1). A subset was meticulously chosen from the existing image archive to construct the dataset under discussion, comprising images showcasing various heavy machinery used in construction sites. Furthermore, additional images were procured

and integrated to ensure a comprehensive representation of diverse construction equipment, accounting for even the rare types of machinery commonly encountered in construction scenarios. This meticulous process resulted in a dataset encompassing a wide range of construction machinery for practical analysis and evaluation.

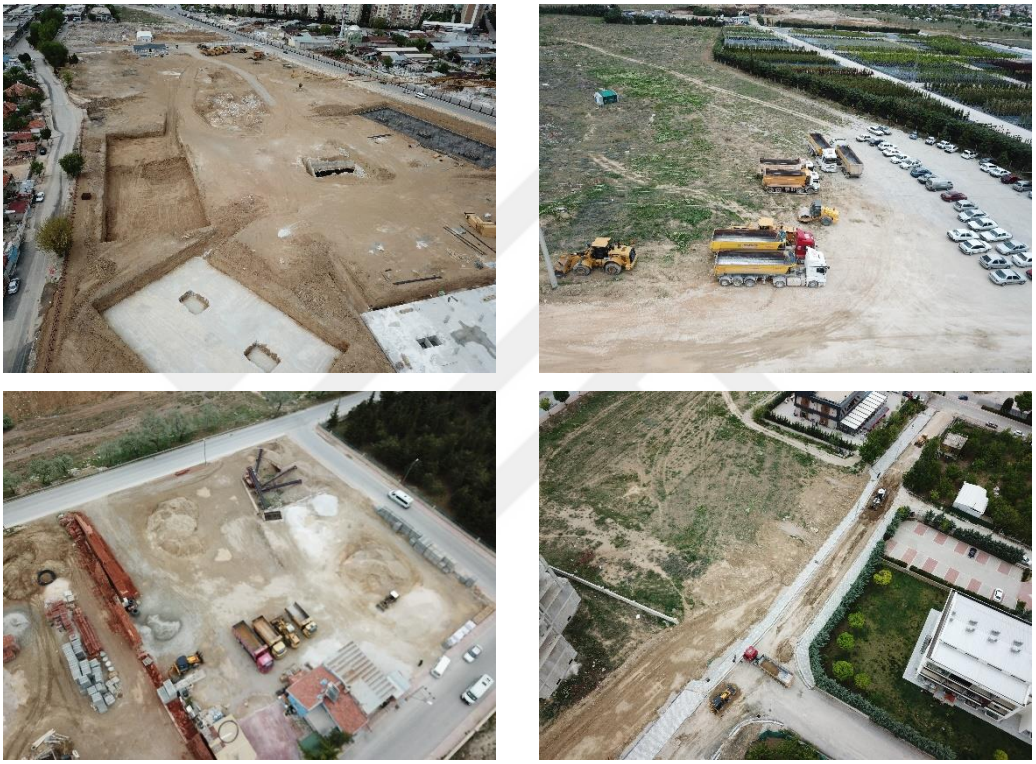


Figure 3.1. Aerial Perspective View of Images Collected from Construction Sites

Four UAVs were used to collect these images, including DJI's Phantom 4 RTK, Mavic Pro, Mavic 2 Pro, and Yuneec's H520. Each drone possessed unique specifications and capabilities, contributing to the image collection's diversity and overall quality.

Phantom 4 RTK by DJI is a drone renowned for its high-resolution camera and extensive range. These advantages allowed for capturing detailed aerial images from a considerable distance, providing a broad perspective of the construction sites. The

exceptional camera quality of the Phantom 4 RTK enabled sharper, more detailed images, ensuring the fine details of construction sites and machinery.

In contrast, DJI's Mavic Pro and Mavic 2 Pro offered distinct benefits. These drones were chosen due to their portability and simplicity of operation. Their miniature size and user-friendly controls made them incredibly convenient for rapid deployment. Additionally, they were renowned for their ability to capture stable, high-quality footage in flight. These characteristics make them ideal for capturing multiple perspectives and angles of construction sites, enriching the dataset.

The Yuneec H520, on the other hand, contributed its unique strength. This UAV has a six-rotor system that provides superior stability, especially in windy conditions. This feature was essential for capturing high-quality, clear images in less-than-ideal weather conditions, ensuring consistency throughout the dataset.

Utilizing the unique capabilities of these drones, high-quality images were captured to ensure comprehensive coverage of different types of construction machinery at multiple construction sites. It is important to note that construction image datasets in the existing literature that include UAV imagery are characterized by low-altitude and side-view images (Xuehui et al., 2021). In this dataset, however, we provide a unique perspective through UAV-captured aerial photographs. The camera angles for these images ranged from 60 to 90 degrees, providing a novel and insightful perspective from the top view for analysis.

The dataset was compiled from various construction environments, including deep excavation sites, steel structures, reinforced concrete structures, and various transportation projects (Figure 3.2). These conditions provide realistic and diverse representations of construction machinery in active operation.

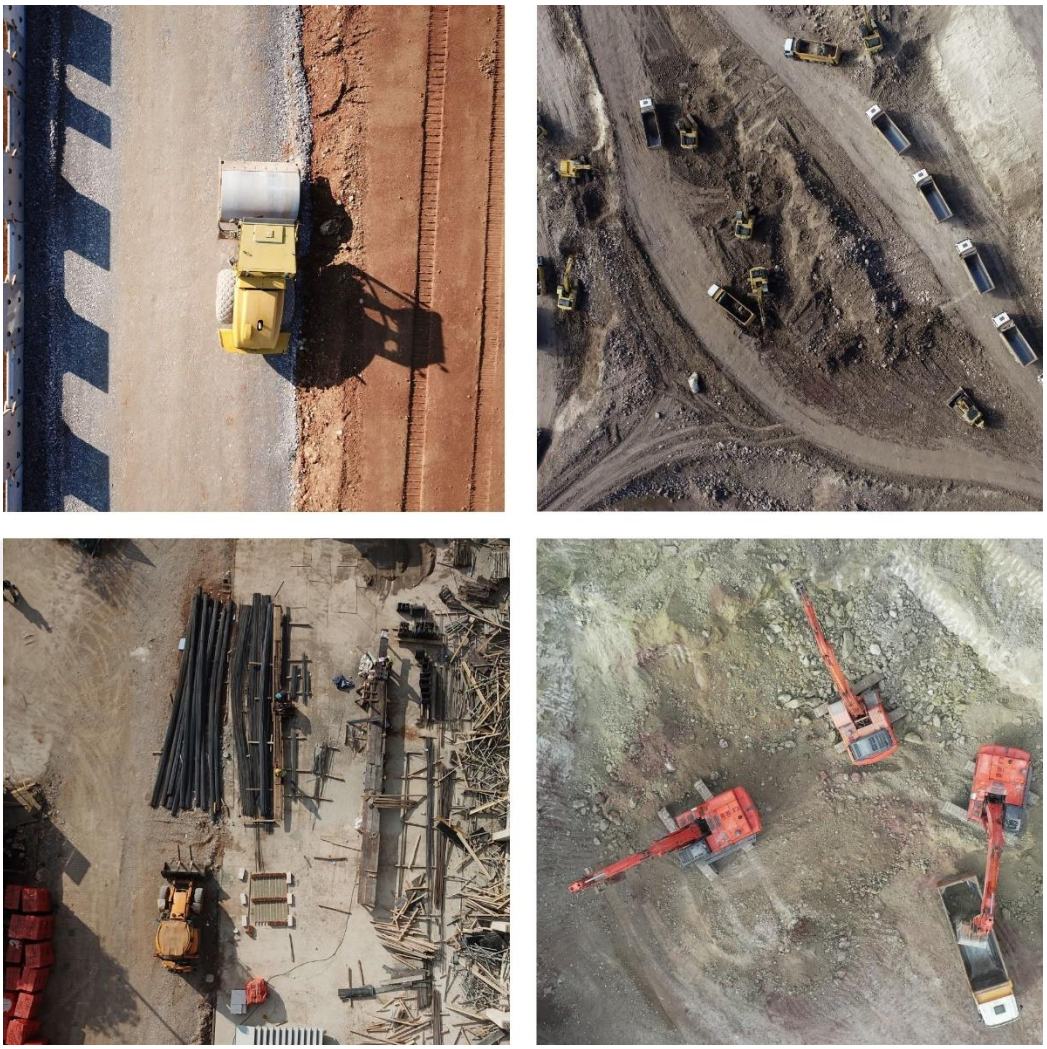


Figure 3.2. Examples of Aerial Images from the Construction Sites

In addition to active construction sites, construction machines are frequently located in designated parking zones. Construction equipment is stored in those areas when not allocated to a job site or maintained. They are characterized by the concentration of various machinery types in a single location, making them ideal for aerial imaging of a wide range of machinery. Therefore, UAV footage was also collected from these parking areas, enriching the dataset with images of parked construction machines (as shown in Figure 3.3). Adding images from these parking zones increased the variety of machinery types. It ensured diverse representations in terms of poses and sizes, both from construction sites and parking areas. This diversity aids in developing

machine learning algorithms capable of handling a wide range of real-world scenarios.

Indeed, the dataset includes images featuring various types of construction machinery and a subset of images from construction job sites where no machinery is present (Figure 3.4). Negative samples are essential to any machine learning dataset, which can improve the performance of deep learning algorithms. It also enhances the algorithm's ability to distinguish between relevant and irrelevant patterns (e.g., construction machinery). In addition, it advances the algorithm's ability to generalize, making it more robust and able to process new and unexplored data.

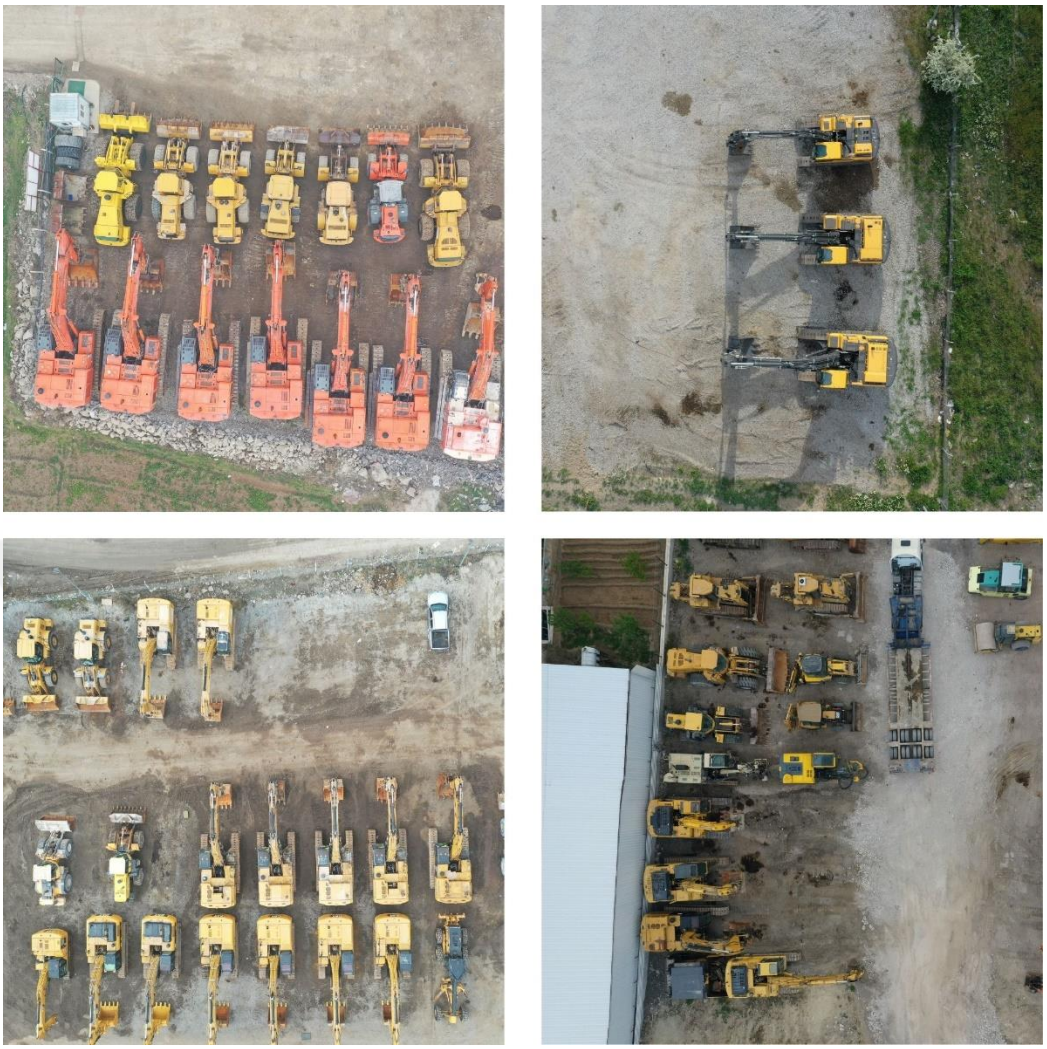


Figure 3.3. Examples of Aerial Images from the Parking Areas



Figure 3.4. Negative Image Samples that Do Not Include Machinery

This extensive dataset includes various locations, machinery types, poses, and negative samples, making it a valuable asset for developing and enhancing machine learning algorithms in construction site monitoring and management. Furthermore, it represents a significant stride towards addressing the challenge of automating the identification and removal of construction equipment from images, thus facilitating precise earthwork calculations.

Throughout the extensive data collection period, the construction sites' weather conditions experienced notable shifts, ranging from summer to winter (Figure 3.5).

Recognizing the potential influence of weather conditions on image analysis and the accurate recognition of construction vehicles, the dataset was meticulously curated to encompass a diverse collection of images captured under varying weather conditions.



Figure 3.5. Different Weather Conditions on the Same Aerial Spot

The dataset includes images captured during bright, sunny days and snowy conditions. Including diverse weather conditions is a crucial dataset component that ensures the trained deep-learning models can recognize construction machines in various environmental conditions. This weather diversity significantly improves the applicability and generalizability of the trained models in the real world. To ensure uniform lighting conditions throughout the dataset, all images must be captured during daylight hours. The consistency of lighting conditions facilitates the simplification of image analysis and the enhancement of machine learning models' precision.

Lastly, JPG was chosen as the image format for the dataset. This format was selected due to its widespread use and acceptance in image processing applications, making the dataset more accessible and user-friendly for many users. In addition, GPS coordinates were embedded in the dataset's images' metadata to facilitate location identification and categorize images from the same construction sites.

In conclusion, by incorporating variables such as various construction environments, weather conditions, daylight consistency, image format, and geographic information, the dataset is well-equipped to facilitate the development of

robust and versatile machine-learning models for construction site monitoring and management.

3.3 Image Annotation

Having explored the image collection process, we now proceed to the next pivotal stage: Image Annotation. This phase is where the raw images are transformed into a valuable repository of usable data, enabling the identification and categorization of various types of construction machinery. Image annotation plays a vital role in training effective machine learning models, as it provides the necessary ground truth labels that enable the algorithms to learn and recognize patterns accurately. This crucial step ensures that the machine learning models are equipped with the knowledge needed to monitor and analyze construction sites effectively, further advancing the field of construction technology.

By conducting a thorough analysis of construction machine datasets in academic literature and seeking input from industry experts, construction machine categories were selected for the dataset. This approach aimed to encompass the most prevalent and widely used types of construction equipment. By combining academic research and real-world expertise, the chosen categories will likely provide a comprehensive representation of the construction machinery landscape. This methodology ensures that the dataset covers various equipment commonly encountered in construction projects.

The selected categories included a variety of standard and essential construction equipment:

1. Dump trucks
2. Excavators
3. Backhoe loaders
4. Wheel loaders
5. Compactors

6. Dozers
7. Graders

Indeed, each of the selected machinery types plays a pivotal role in construction operations, and their inclusion in the dataset is paramount for developing accurate machine-learning models capable of identifying and distinguishing between them (see Figure 3.6 for visual representation). Furthermore, during the dataset creation process, it was noticed that certain types of construction machinery, such as asphalt paving machines and drilling machines, appeared less frequently on construction sites. To solve this, these less prevalent machines were grouped under "Other" rather than being given distinct categories. This approach was taken to ensure that the machine learning models would not be biased towards the more prevalent machinery types and to maintain a balanced distribution of categories.

In addition, cars frequently appeared within or close to the construction sites in the aerial images. Cars were given their category due to their frequency and potential influence on subsequent analysis. This choice helps the machine learning algorithms distinguish between construction-related machinery and other vehicles, improving their predictions' accuracy.



Figure 3.6. Selected Object Classes for Image Annotation

The decision to capture images from a top-down perspective proved highly effective in addressing occlusion. Occlusion occurs when objects in an image are partially obscured by other entities, making it challenging for machine learning algorithms to recognize them accurately. However, by adopting the top-down imaging method, this problem was largely mitigated. The advantage of the top-down perspective is that it allows for a comprehensive view of the construction equipment from above. As a result, occlusion is minimized, as most of the machinery can be seen in its entirety without being obstructed by other objects.

However, because UAVs capture images while in motion, particular challenges remain. Due to the movement of the UAV or the construction machines, vehicles may appear and disappear intermittently within the camera's field of view. This led to instances where only a portion of a machine was visible in an image. The procedure for such situations is that if more than half of a machine was visible in a snapshot, it was considered important enough to include and was annotated accordingly.

The dataset's creation was motivated by a solid commitment to protecting the privacy of people and enterprises. In the case of individuals, people's faces looking upwards at the UAV could have been included unintentionally in the dataset, given that the images were captured from an aerial perspective. To address this issue, a review procedure was implemented in which any facial features detectable in the images were blurred (Figure 3.7). This procedure was applied to prevent the identification of specific individuals and to mitigate potential privacy concerns.

Similarly, corporate entities were also subject to privacy considerations. Company names, logos, and slogans were frequently displayed on construction sites on machinery, buildings, and other infrastructure. Identifiable corporate information visible in the images was identified and obscured to prevent a potential violation of corporate privacy rights. Figure 3.7 depicts the privacy precautions that were taken to respect the privacy rights of individuals and organizations.

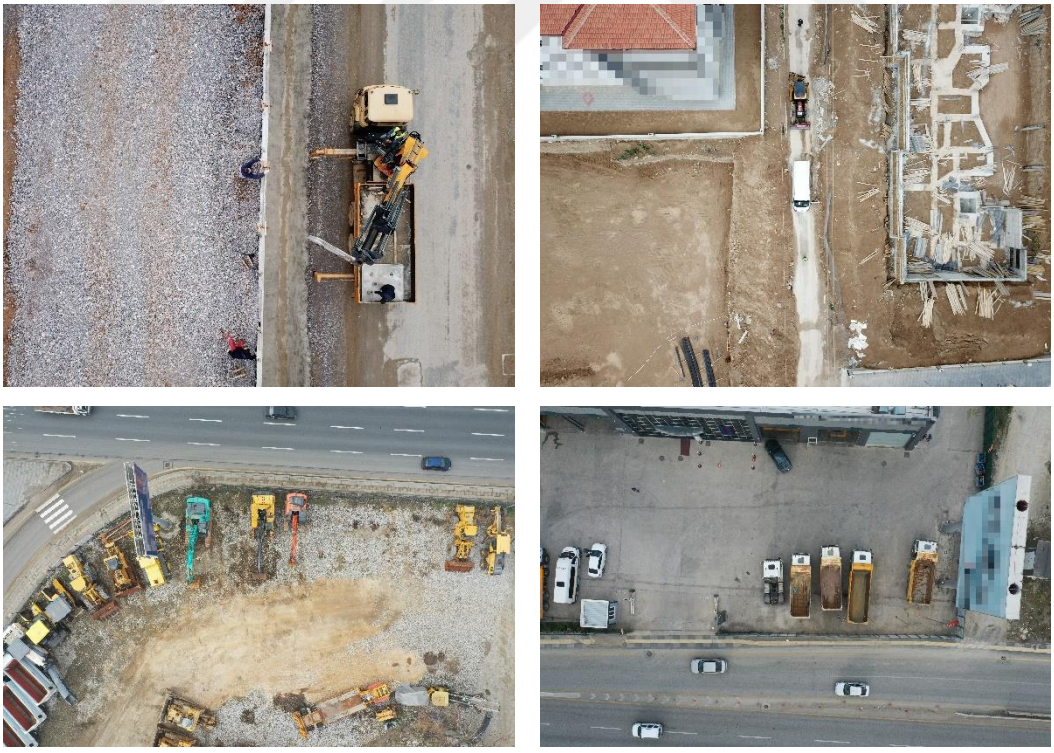


Figure 3.7. Privacy Protections for Faces, Brand Logos, and Facets

The meticulousness of image annotation played a crucial role in creating the dataset. For this purpose, a team of five specialists was assembled. Each team member received a comprehensive briefing on the annotation methodology, ensuring a uniform understanding of the process and promoting uniformity throughout the annotations. Systematically, the entire dataset was divided into 100 image subsets, and each subset was then assigned to an annotator. We used the CVAT (Computer Vision Annotation Tool) (CVAT, 2022), a robust and user-friendly tool designed specifically for image annotation. This application was installed on a web server within the infrastructure of Amazon Web Services. This configuration allowed for the parallel execution of the annotation procedure. Each annotator works independently on the subset of images assigned to them. The cloud-based annotation interfaces are presented in Figure 3.8.

The annotators spent a total of 82.38 person-hours annotating the images. Instead of simply drawing boxes around the objects of interest, the precise boundaries of each object were drawn. This method is more complicated and time-consuming than standard bounding-box annotation, but it provides significantly more information, essential for developing accurate segmentation models. After the data was annotated, it was stored in the COCO (Common Objects in Context) data format (T. Y. Lin et al., 2014), widely accepted in the machine learning community. The data was saved in the JSON (JavaScript Object Notation) file format, a lightweight and straightforward data-exchange format.

Eventually, 9563 boundaries of instances were drawn and labeled in 2155 images. The extensive and detailed process of image annotation described above lays the groundwork for an in-depth comprehension of the contents and characteristics of the dataset. With the images precisely annotated and curated, we can now overview the statistics of the dataset.

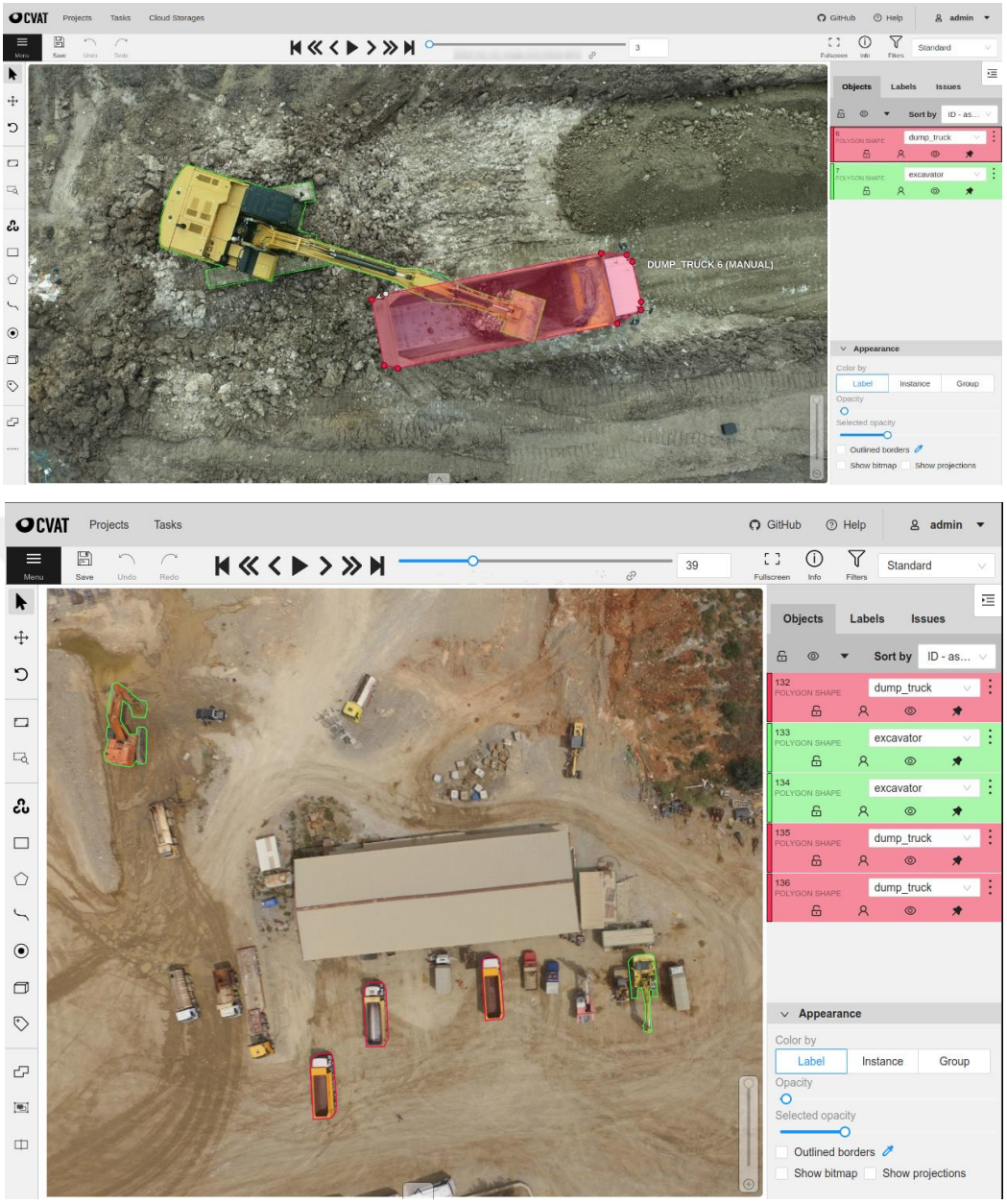


Figure 3.8. CVAT Image Labeling Interfaces (CVAT, 2022)

3.4 Dataset Statistics

This section examines the characteristics of the dataset from a statistical perspective. Each dataset attribute thoroughly explains the data's structure and composition. The

distribution of diverse construction machines across the images in the dataset is one of the most important aspects highlighted. As depicted in Figure 3.9, excavators, dump trucks, cars, and backhoe loaders occur most frequently per image.

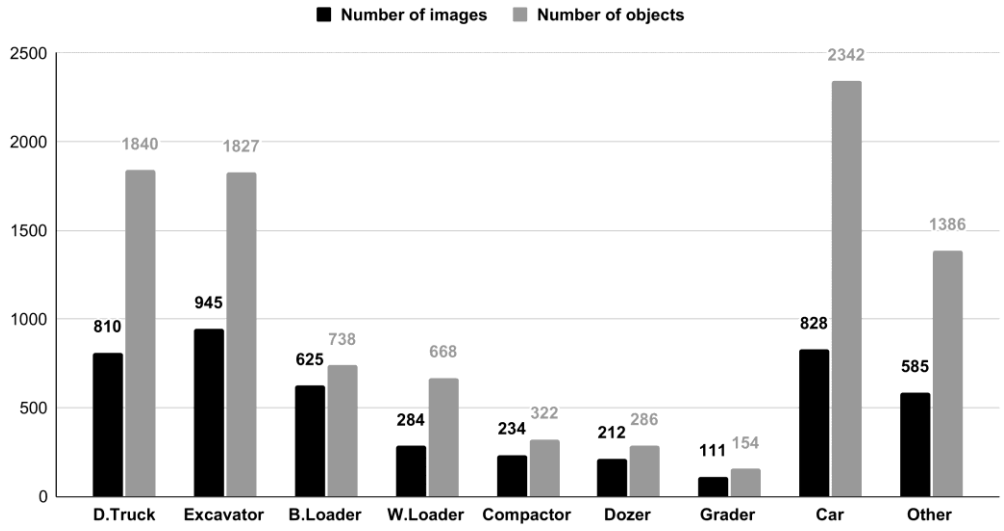


Figure 3.9. Number of Images and Number of Objects for Each Type of Construction Machine

Additionally, we investigate the distribution of objects in the images captured. Figure 3.10 details the per-image counts of various objects found in the images. Notably, 254 images, or 11.79 percent of the dataset, do not contain construction machinery. These images are considered negative samples, serving as a contrast to the images depicting construction machines. Including such negative samples in the dataset can further improve the performance of machine learning algorithms by providing a broader perspective and posing the challenge of distinguishing between relevant and irrelevant patterns.

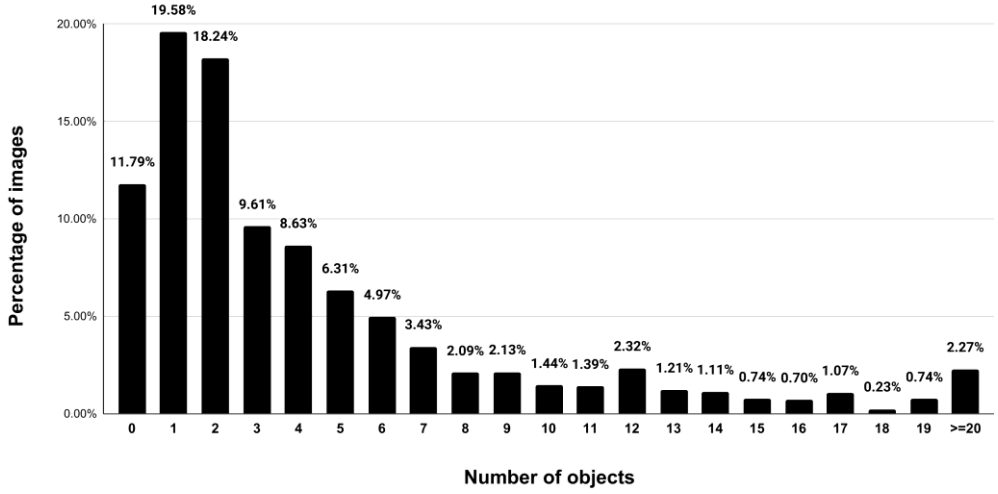


Figure 3.10. Number of Objects per Image (%)

The proposed dataset differs from past datasets produced from ground-level observations in significantly increasing the number of items seen in each image. As depicted in Figure 3.11, this is primarily attributable to the use of UAVs for image capture, which provides a broader field of view due to their altitude and camera angle.

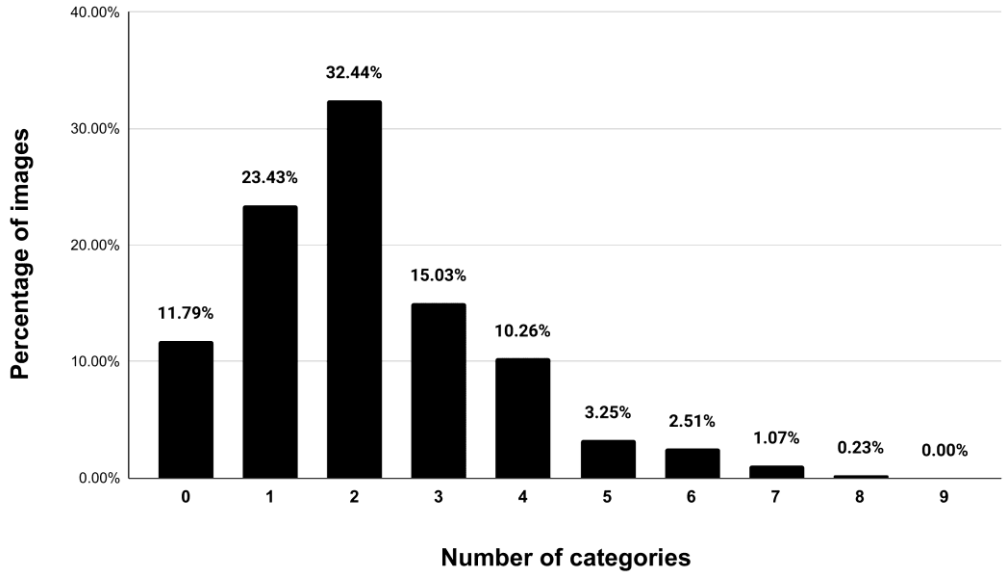


Figure 3.11. Number of Categories per Image (%)

In addition, as shown in Figure 3.12, we analyze the distribution of bounding box sizes containing objects within the images. Due to the high-altitude perspective from which the UAV footage is captured, most objects appear relatively small. This attribute presents a compelling challenge for machine learning algorithms, which are tasked with accurately detecting and classifying these objects despite their small sizes.

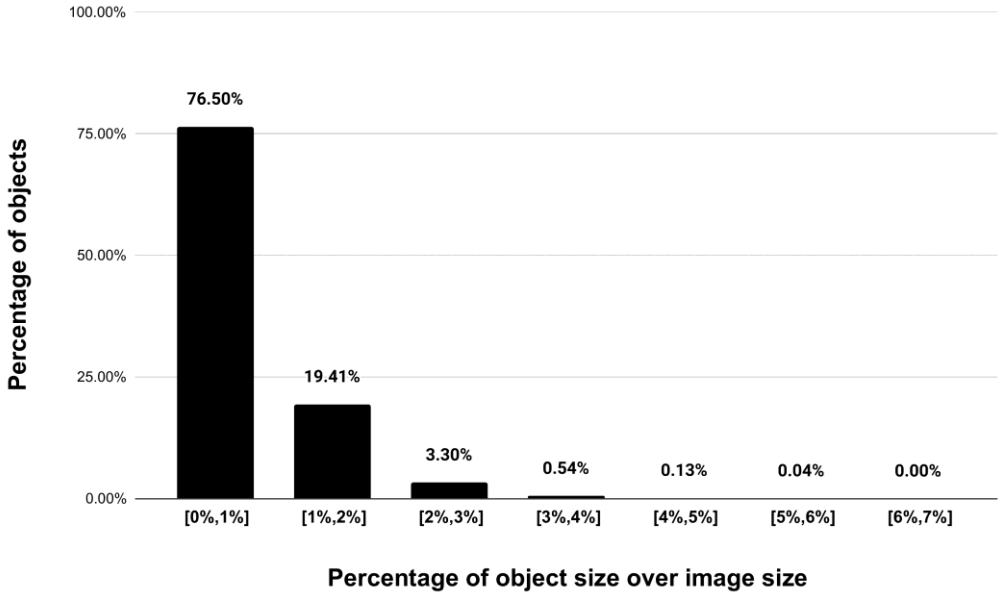


Figure 3.12. Distribution of the Object Size (%)

When examining previous studies involving similar datasets, one observation that stands out is that images depicting identical scenes were typically not separated but included in both the training and testing subsets of data. Such an arrangement could result in the inclusion of similar or even identical images in both splits. While potentially advantageous for overall model accuracy in the short term, this practice can result in overfitting, known as "memorization." The model may over-learn from the training data to the point where it memorizes particular image features and

patterns. As a result, it can perform exceptionally well with these familiar or duplicate images but struggle with novel images.



Figure 3.13. Similar Images from the Literature (Xiao & Kang, 2021a)

A more robust approach would be to ensure that the test set consists entirely of images the model has never seen, i.e., images on which it has not been trained. This strategy ensures that the performance evaluation of the model is accurate and representative of its ability to generalize to new data. However, manually removing duplicate images is a laborious and time-consuming procedure. There is a need to automate this process by employing algorithms able to evaluate image similarity. By ensuring the uniqueness of images in the training and testing sets, such an approach could streamline the process and improve the dataset's quality.

The same concern applies to the dataset we propose. As shown in Figure 3.14, specific machines appeared in multiple images with different perspectives. Therefore, it was essential to identify these overlaps and correctly group similar images to ensure an unbiased separation of the training and test datasets. To accomplish this, we utilized a method of image clustering based on similarity and proximity metrics.



Figure 3.14. Clustering Strategy Examples for Overlapping Images

A method employing Sentence Transformers, which convert text into vector representations, was utilized to calculate the percentage of image similarity (Reimers & Gurevych, 2019). This was a noteworthy choice because most of the input consists of images. Using textual descriptions of images as the basis for measuring their similarity is the method's idea. Using Sentence Transformers, this technique transforms the content of each image into textual descriptions, which are then converted into sentence embeddings of fixed-size vectors. Essentially, these embeddings contain the semantic information conveyed by the image descriptions. This method permits the calculation of similarity scores between pairs of images based on the cosine similarity or Euclidean distance of their respective sentence embeddings. It is a clever and efficient method for converting visual comparison to semantic comparison, making it ideally suited for image clustering.

Regarding the proximity metric, we use the geographical data associated with each image. The extracted GPS coordinates from each image's metadata were used to calculate the distance between images. This calculation is essential because images captured nearby typically contain similar scenes, and it is necessary to group them.

The clustering procedure was performed to assign to the same cluster images with similarity percentages exceeding 95% and within a distance of 500 meters. 1386 distinct clusters were formed through this procedure. Then, cluster identifiers, or IDs, were added to the JSON file used to label the images. These cluster IDs are essential for dividing the dataset for training and testing, with each split containing only images with unique cluster IDs. Figure 3.14 depicts an example of the image metrics utilized by this process.

After a comprehensive examination of the dataset's statistics and a discussion of the various characteristics of image data, we now transition to the study's performance evaluation phase. This crucial step allows us to evaluate the dataset's effectiveness when combined with different machine-learning models.

3.5 Performance Evaluation

Evaluation of a dataset's performance is vital in determining its value, mainly when it is developed for machine learning applications. In the following sections, we provide a comprehensive report on the outcomes of the performance analysis of the dataset using various DL algorithms. These results evaluate how effectively and precisely these models can process and interpret the dataset's information, shedding light on its overall quality and potential for practical applications.

Five distinct deep learning algorithms that are frequently employed in computer vision-based research were employed for the performance evaluation. These algorithms were selected based on their specific capabilities and performance characteristics, making them highly suitable for the proposed dataset and the tasks. According to the studies in the literature, the majority of these chosen DL

algorithms belong to the category of two-stage detectors, which have demonstrated superior accuracy rates in comparison to their one-stage counterparts (Carranza-García et al., 2021; Soviany & Ionescu, 2018). While one-stage algorithms provide faster detection speeds and the potential for real-time results in construction datasets, they typically compromise precision. As our current focus is not on developing real-time applications but rather on achieving the highest level of accuracy, the decision to use two-stage detectors was logical.

Mask R-CNN (He et al., 2018), Cascade Mask R-CNN (K. Chen et al., 2019), Mask Scoring R-CNN (Huang et al., 2019), Hybrid Task Cascade (K. Chen et al., 2019), and Pointrend (Kirillov et al., 2020) were the algorithms utilized for this study. To optimize the use of computational resources during the training process, we rescaled the dataset's images to 1300x500 pixels. We used ResNet50+FPN as the backbone for all of the chosen algorithms. ResNet50+FPN was selected due to its high performance and efficiency in image recognition tasks and its ability to capture a wide variety of features at different scales effectively.

20 epochs were utilized during the training of the algorithms. The initial learning rate of 0.02 was selected. A technique known as learning rate scheduling was utilized to improve the training process further. This involves reducing the learning rate by a particular factor at specific points in the training process to facilitate the algorithms' convergence on the optimal solution and prevent overfitting. After the 16th and 19th epochs, we reduced the learning rate by 0.1.

We conducted two training rounds to determine the impact of grouping similar images in the same splits. A clustering strategy was utilized in the first round, while this strategy was omitted in the second round. These two distinct approaches aimed to gain insight into the impact of grouping similar images together versus randomly scattering them throughout the dataset.

The dataset was split into two sections to conduct these training rounds: 80% of the images were used for training, and 20% were set aside for testing. To evaluate the performance of models trained on the dataset, we adopted metrics from the COCO

dataset. The COCO dataset is a large-scale dataset for object detection, segmentation, and captioning, and its performance evaluation metrics are highly regarded within the machine learning community (T. Y. Lin et al., 2014).

This study's primary evaluation metric is the mean average precision (mAP). This metric is a single-number measure of the quality of object detection models. To better understand the mAP used to evaluate our deep learning models, it is necessary to define certain terms:

True Positive (TP): A condition in which a model identifies the positive class correctly. In the context of object detection, this means the model correctly identifies and positions a bounding box around the object.

False Positive (FP): A condition in which a model mispredicts the positive class. This occurs in object detection when the model incorrectly places a bounding box without an object or incorrectly classifies the object contained within the bounding box.

False Negative (FN): A situation in which a model misses a positive class. This situation occurs in object detection when the model fails to place a bounding box around a correct object.

Precision: The ratio of correctly predicted positive observations to the total number of positive observations. In essence, it indicates the proportion of correctly identified objects.

$$Precision = \frac{TP}{TP + FP}$$ Equation 1

Recall: The ratio of correctly predicted positive observations to the total number of actual observations. In essence, it indicates the proportion of correctly identified actual objects.

$$Recall = \frac{TP}{TP + FN} \qquad \text{Equation 2}$$

Intersect over Union (IoU): IoU is a measure of overlap between two bounding boxes - the ground truth and the predicted bounding box (Figure 3.15). In object detection tasks, it is typically used to distinguish between true positives and false positives. The typical IoU threshold for classifying a sample as positive is 0.5, although this can vary.

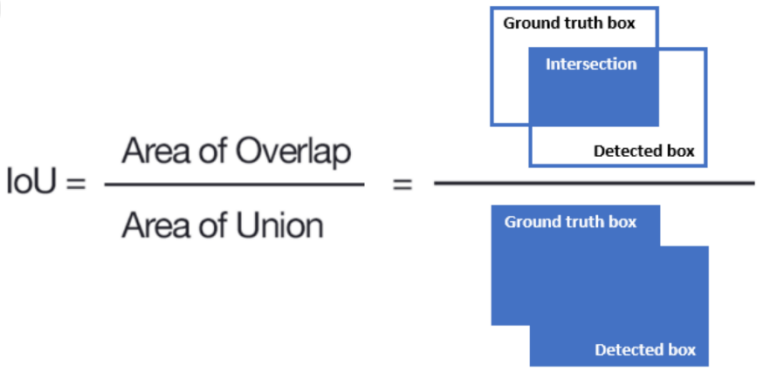


Figure 3.15. Illustration of Intersection over Union (IoU)

Mean Average Precision represents an average of precision values at different recall levels. The following steps are required for mAP calculation:

Computation of Precision-Recall Curve: Precision and recall are computed at every threshold level for each class (for all potential predictions). The Precision-Recall curve is subsequently plotted.

Calculation of Average Precision (AP): The area under a class's Precision-Recall curve represents the class's Average Precision. It is computed by averaging the precisions at a variety of recall levels (0.0, 0.1, 0.2, ... , 1.0). This essentially captures the distribution of the Precision-Recall curve and indicates the model's performance for that class.

The mAP is the mean of all APs for each class. Precisely, it is calculated using IoU thresholds ranging from 50% to 95%, with a 5% increment. This range of thresholds enables us to evaluate the robustness of our model under varying degrees of overlap between predicted and actual bounding boxes. mAP_{50} and mAP_{75} are two specific metrics derived from the mAP corresponding to the mean average precision at IoU thresholds of 50 percent and 75 percent, respectively.

In addition, we investigated the performance of our model for small, medium, and large objects in the dataset. The size classification is based on the area occupied by the objects, with small ones having an area less than 32^2 pixels, medium ones having an area between 32^2 and 96^2 pixels, and large objects having an area greater than 96^2 pixels. These categories' corresponding mAPs are designated as mAP_s, mAP_m, and mAP_l. mAPs of different IoU threshold levels of each algorithm are presented in Table 3.1 and Table 3.2. There are no objects smaller than 32^2 pixels, therefore, mAP_s was not calculated for the proposed dataset.

Table 3.1 Results of the Clustered Data

Algorithm	**mAP**	**mAP_{50}**	**mAP_{75}**	**mAP_m**	**mAP_l**
Hybrid Task Cascade	67.7	92.4	81.5	47.1	69.0
Cascade Mask R-CNN	66.2	91.0	81.4	52.4	67.6
Mask Scoring R-CNN	66.1	88.4	80.0	36.6	68.0
Pointrend	68.2	92.6	83.5	49.7	69.6
Mask R-CNN	66.6	91.6	80.4	46.8	67.8

Table 3.2 Results of the Unclustered Data

Algorithm	**mAP**	**mAP_{50}**	**mAP_{75}**	**mAP_m**	**mAP_l**
Hybrid Task Cascade	71.9	93.7	86.4	60.4	72.5
Cascade Mask R-CNN	70.7	93.4	85.0	58.9	71.4
Mask Scoring R-CNN	72.9	93.7	88.3	41.9	73.8
Pointrend	72.9	94.2	88.4	49.2	73.6
Mask R-CNN	71.4	93.9	86.6	57.1	72.3

These deep learning models were trained on a computer comprised of an Intel(R) Core(TM) i7-10700K processor running at 3.0 GHz with 12 cores and 128 GB of memory. The system also included an NVIDIA GeForce RTX 2080 Ti graphics card for faster training. Ubuntu 18.04 was the operating system that was in use. The models were trained using the MMDetection 2.18.1 toolbox, an open-source object detection software library built on the PyTorch framework (MMDetection Contributors, 2018).

As mentioned above, the IoU threshold of 50% was applied to the top-performing deep learning algorithm's class-wise Average Precision (AP). These results enable us to comprehend the per-class performance of our model (Table 3.3). Noting that all IoU values were calculated using segmentation masks instead of bounding boxes provides a more nuanced measurement of the model's prediction accuracy.

Table 3.3 Classwise AP Results (IoU=50%)

Algorithm	D.T	Exc.	B.L.	W.L.	Com.	Doz.	Gra.	Car	Other
Pointrend (clustered)	97.1	97.5	92.2	92.8	91.5	95.9	92.3	96.5	77.7
Pointrend (unclustered)	97.3	97.8	97.9	96.6	92.6	86.5	100.0	95.0	84.2

Following a comprehensive evaluation of the performance of the current machine learning models on the dataset, we transition to a comprehensive discussion and analysis of our findings. This subsequent section examines the implications of the evaluation's findings, extracting insights and discussing potential enhancements for future studies.

3.6 Results and Discussion

Currently, most construction-centric datasets rely on ground-level images captured by stationary or mobile cameras. Although UAV imagery applications have increased, comprehensive, publicly accessible aerial image datasets are still absent. The most significant contribution of this research to the existing literature is the introduction of a real-world image dataset comprising aerial perspectives of various construction machinery with pixel-level annotations. Figure 3.16 depicts some examples of the detections generated by deep learning algorithms. Subsequent sections will provide a comprehensive analysis and discussion of the results derived from the proposed dataset.

The mean average precision (mAP) metrics of five distinct object segmentation algorithms are presented in Table 3.1 and Table 3.2. These algorithms are Hybrid Task Cascade, Cascade Mask R-CNN, Mask Scoring R-CNN, Pointrend, and Mask R-CNN. Using a variety of metrics, such as mAP, mAP_{50}, mAP_{75}, mAP_s, mAP_m, and mAP_l, these algorithms were evaluated on both clustered and unclustered datasets.

Pointrend outperforms in clustered and unclustered datasets, achieving peak mAP_{50} scores of 92.6% and 94.2%, respectively. The mAP_{50} scores for the remaining four algorithms range from 92.4% to 88.4% for the clustered dataset and from 93.9% to 93.4% for the unclustered dataset.



Figure 3.16. Example Images of Segmentation Results

The use of images captured by UAVs has numerous advantages over CCTV and mobile camera images. The aerial perspective provided by high-resolution cameras mounted on UAVs and GPS technology enables precise object location on construction sites. In addition, UAVs allow for the acquisition of images in a timely and risk-free manner while minimizing the occlusion issues frequently encountered with ground-level imagery.

The proposed dataset contains images of various construction environments, including excavation sites, steel, reinforced concrete structures, and transportation projects. This diversity provides a rich resource for researchers and industry professionals, allowing for developing and refining machine learning algorithms and tackling construction management challenges across various project types.

Segmenting construction machines in images provides a more comprehensive understanding of the objects than bounding box-based detection. Generating pixel-wise masks for each piece of construction vehicle permits a more accurate representation of their boundaries, shapes, and orientations. This level of specificity facilitates comprehension of the positions and activities of the machines. Object segmentation enables the analysis of spatial relationships between machines and workers, which can evaluate and improve machinery usage, identify potential hazards or inefficiencies, and plan future activities more effectively. In addition, the segmentation of construction machines can facilitate the detection of anomalies, such as unauthorized access to restricted zones or machinery operating in unsafe conditions. Such automation can significantly enhance site safety, reduce its dependence on manual inspections, and lower the risk of accidents.

A model trained with an unclustered dataset tends to have a higher mAP than one trained with a clustered dataset. This difference results from the unclustered dataset containing similar images in both the training and testing phases, causing the model to memorize the training data. As a result, the accuracy of the testing set appears to be high but does not accurately reflect the model's performance on new, unseen images. Ensuring that the training and testing sets contain no overlapping images is essential to mitigate this issue. In this study, we clustered our dataset using an algorithm based on image similarity and proximity. Each of these clusters can be used for training and testing. This method ensures that the model is evaluated on data to which it was not exposed during the training phase. Therefore, compared to the unclustered set, the clustered set's results are more reliable and applicable to real-world applications.

The per-class AP results table demonstrates that Pointrend achieves the highest AP in most classes, including dump trucks, excavators, backhoe loaders, wheel loaders, compactors, dozers, graders, and cars (as shown in Table 3.3). However, the algorithm has difficulty accurately identifying objects in the "other" category. This category includes a wide range of categories, including mobile cranes, concrete pump trucks, cement trucks, drilling machines, asphalt producers, and forklifts. The limited number of representative images for these classes, which had to be grouped together for analysis, poses a challenge for accurate identification within this category. Due to the diverse shapes and characteristics of the objects in the "other" category, it is challenging to learn the distinctive features of this class from a small number of images. Therefore, there is a clear need for additional images to improve the performance of deep learning algorithms on this specific class of objects.

On construction sites, excavators and dump trucks frequently operate nearby while transporting materials such as soil, rock, and debris. This operational interaction can make it difficult for algorithms to accurately differentiate between the two machine types. The merging of the boundary lines between an excavator and a dump truck, as illustrated in Figure 3.17, can be a significant issue. This can lead to misclassification, incorrectly labeling two distinct machines as a single entity, thereby reducing the accuracy of identification.

In addition, accurately identifying the slimmer parts of machines, particularly those categorized as "others," is complicated. A concrete example can be seen in Figure 3.17, where the representation of the thin components of a concrete pumper machine appears incomplete. To overcome this obstacle, capturing the diverse poses of these machine classes is essential, as this enables a more comprehensive representation of the machines and improves the accuracy of identifying their slimmer components. In addition, when images are reduced to smaller dimensions (in this case, to 1300x500 pixels), the finer details of the objects may be lost or blurry. This information degradation can make it more difficult for algorithms to accurately identify and segment these components. To ensure optimal detection, it is suggested that the flight altitude and selected camera resolution should be carefully examined.



Figure 3.17. Results of Incorrect Object Segmentations

The number of images in the dataset plays a crucial role in enhancing the performance of machine learning models. The ability of the dataset to capture the diversity and complexity of construction machines improves as the number of images increases. In turn, this improves the model's ability to generalize and its performance on unseen data. To mitigate the observed limitations in the current dataset, it is recommended to increase its size with a more diverse and representative collection of images, including various poses and types of machinery. This strategy would substantially enhance the accuracy of detection and classification.

Having thoroughly investigated and discussed the application of deep learning models for construction machine detection and segmentation, we now apply these deep learning-based approaches for tracking earthwork progress. The high-precision machine detection methods established in the preceding section equip us with a machine-learning model well-suited for detecting machines in subsequent processes. The following chapter will examine a novel method for calculating earthwork volumes.

CHAPTER 4

METHODOLOGY FOR EARTHWORK VOLUME CALCULATION

4.1 Introduction

Accurate and regular monitoring is necessary to successfully complete a construction project within the allocated time and budget. The importance of earthwork operations such as excavation and filling cannot be ignored, considering the other activities involved in these projects. They are the prerequisites of the construction process, laying the foundation for subsequent operations. However, despite their significance, monitoring the progress of such earthwork operations presents a significant challenge, frequently necessitating a considerable amount of time, manual effort, and resources.

With the introduction of digital technologies, new opportunities have emerged for automating and optimizing the earthwork monitoring process. Unmanned Aerial Vehicles (UAVs) have emerged as a promising tool for capturing high-resolution images of construction sites from a bird's eye view. When these images are combined with advanced algorithms, it is possible to obtain valuable insights regarding the construction progress, such as volume calculations for earthwork.

One major challenge in using UAV images for automated earthwork progress tracking is the presence of construction machines, which can directly affect the measurements of the actual ground level. This necessitates the development of a precise and automated method for removing these machines. This chapter provides a comprehensive description of the methodology employed in this study to automate the monitoring of earthwork progress. This methodology utilizes high-resolution imagery from UAVs, deep learning, and geometry processing algorithms to precisely calculate the volume of earthworks.

The methodology is divided into four main sections: UAV imaging, construction machine segmentation and removal, 3D reconstruction, and volume calculation. Each section describes the techniques and processes used and how they contribute to the final output, accurately measuring the earthworks' volume.

The initial phase involves gathering data through UAVs obeying the principles of photogrammetric modeling. Following this, the second stage revolves around machine segmentation and removal and focuses on how the raw images captured by UAVs are processed to identify and remove heavy machinery. Through the 3D reconstruction process, in which the cleaned-up imagery is converted into a 3D model, we present a more precise and detailed representation of the construction site. Lastly, the volume calculation section describes the computation of volumetric differences to determine the volume of earthworks.

4.2 Data Collection

The first stage of the method, image acquisition, plays a crucial role in determining the quality and success of the entire process. This stage aims to capture a set of high-resolution scene images from different viewpoints, ensuring adequate overlap between adjacent images to facilitate accurate feature matching and camera pose estimation for 3D reconstruction.

UAV imaging starts with a comprehensive flight plan. Specialized flight planning software plots the flight path and defines waypoints, considering factors such as altitude, overlap between consecutive images (typically 60-80 percent overlapping), UAV speed, and battery endurance. In addition to ensuring compliance with local UAV flight regulations, the physical features of the terrain are considered when designing the flight plan. An illustration of overlapped aerial images is given in Figure 4.1.

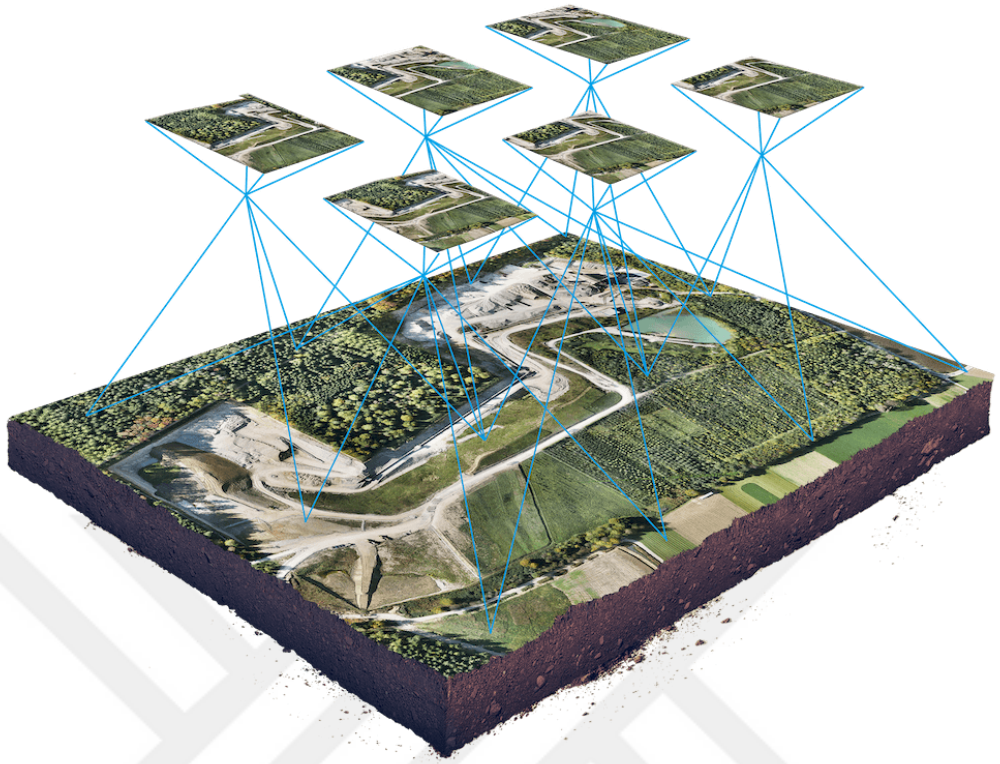


Figure 4.1. Demonstration of Overlapping Images (Wingtra, 2023)

Prior to the UAV flight, ground control points (GCPs) are set up across the designated survey area. These GCPs consist of identifiable markers strategically positioned within the survey zone, and their exact coordinates are determined using high-precision GPS equipment. During the later stages of data processing, these GCPs play a crucial role as reference points, enabling the georeferencing and enhancing spatial accuracy for the resulting 3D model. The GCP locations are carefully chosen to ensure even coverage of the area and clear visibility in the UAV images.

The UAV is launched to fly in autonomous mode to follow the predetermined path after the finalized flight plan and GCP survey. While flying, the UAV captures a series of images overlapping with the construction site. During each instance of image capture, the UAV's onboard GPS and inertial measurement unit (IMU) work in tandem to record essential data related to the UAV's position and orientation. This

collected information plays a crucial role in the subsequent image alignment and 3D reconstruction processes. By leveraging the precise position and orientation data from the GPS and IMU, the images can be accurately aligned with one another and used to construct a detailed 3D model of the surveyed area. The remote controller monitors the image acquisition process in real-time to ensure that the entire area is adequately covered and that the images meet the quality requirements.

After the image acquisition process, the captured images are stored on the UAV's onboard SD card. After landing, the SD card is removed from the UAV. The images are transferred from the SD card to a dedicated computer for further analysis.

4.3 Construction Machine Segmentation and Removal

Upon the efficient and accurate capture of aerial images, a 3D model of the construction sites can be created to facilitate earthwork calculations. The accuracy of these calculations depends on how well the construction job site is represented. The more precise 3D representation of the job site makes the volumetric calculations more accurate. As discussed in the literature Chapter 2.3.2, conventional photogrammetry-based digital terrain generation algorithms are unaware of the type of information on the point clouds. Therefore, they output all geometric features on the point cloud as the same type, including non-terrain objects such as machinery and man-made entities.

In the scope of this thesis, the goal is to detect and eliminate construction machines such as excavators, bulldozers, and trucks from these aerial images. By doing so, the digital representation of the excavation area created using machinery eliminated images accurately represents the natural terrain, allowing for precise volume calculations. This process helps to generate a digital replica of the excavation site that accurately depicts the unobstructed natural terrain. As a result, it allows for more accurate volume calculations by eliminating any distortions or inaccuracies caused by the presence of construction machinery. Therefore, we can improve the accuracy

of 3D reconstructions and subsequent volumetric analyses by carefully removing machinery from the images.

For the identification of these construction machines, we employ deep learning algorithms. As a subset of machine learning, deep learning has attracted significant attention and has shown outstanding performance in object recognition. Object detection and instance segmentation are two essential aspects of deep learning. Object detection is the process of identifying instances of a particular class, such as people, buildings, or in the context of this study, construction machines, within an image or video. It identifies the objects in the image and locates them within the image, often providing bounding boxes around detected objects (X. Wu et al., 2020).

In contrast, instance segmentation involves more detailed mapping of an image. Segmentation partitions an image into regions or pixels corresponding to individual objects, producing a more precise "mask" for each object than detection (Figure 4.2). This technique enables more precise visual recognition and comprehension, which is especially useful in detecting and removing objects.

Numerous cutting-edge algorithms have significantly improved instance segmentation. Mask R-CNN (He et al., 2018) is an innovative model that extends Faster R-CNN by adding a mask prediction branch, thereby enabling precise segmentation. Cascade Mask R-CNN (K. Chen et al., 2019) introduces a multi-stage object segmentation framework for progressive refinement. In contrast, Mask Scoring R-CNN (Huang et al., 2019) includes a mask-IoU branch to evaluate the quality of predicted instance masks. Hybrid Task Cascade (K. Chen et al., 2019) adopts a multi-task learning paradigm and incorporates the cascading concept within this framework to further improve segmentation accuracy. Last but not least, PointRend (Kirillov et al., 2020) employs a rendering-based technique to address the issue of coarse output in segmentation models, thereby enhancing the object boundary detail. These algorithms have substantially improved the accuracy and efficiency of object segmentation tasks.

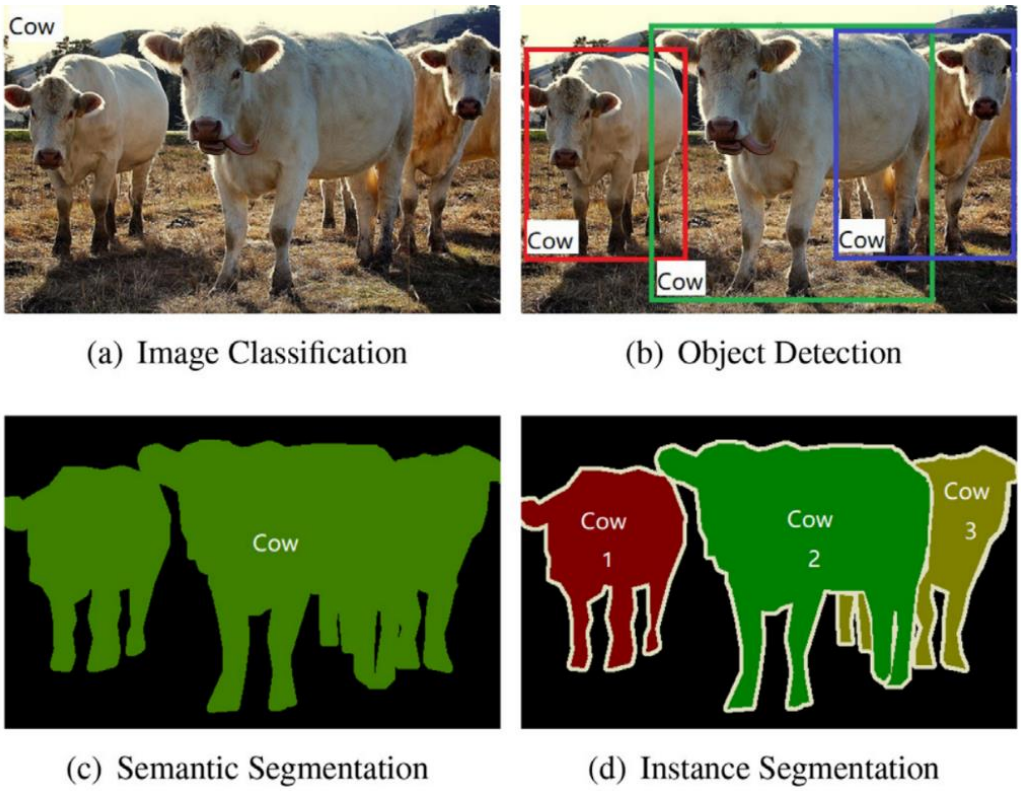


Figure 4.2. Comparisons of Visual Recognition Methods (X. Wu et al., 2020)

Implementing these algorithms for construction monitoring requires a construction-specific, robust data set. Such a dataset would ideally contain diverse construction machinery captured under various conditions and from multiple angles.

By training the machine learning model on this data set, it will be possible to detect and segment construction machinery images on the job site. In this study, we employed the machine learning model described in Chapter 3, which is intended to function as a detector for construction machinery. Following the UAV's imaging process, this model analyzes all georeferenced images and detects the construction machine objects. The identified and segmented objects are converted into image masks, which are then employed in the image inpainting process to eliminate these objects from the images.

Image inpainting is also a computer vision technique that aims to restore or reconstruct missing portions of an image by filling them based on the surrounding

context (Figure 4.3). With the aid of generated masks, the inpainting algorithm can effectively eliminate the construction machinery from the aerial images and fill those pixels with similar pixels with the construction site environment. The subsequent analysis helps to create a more accurate digital terrain model from images lacking any construction machine.

Image inpainting methods also depend on image datasets; in the context of our research, we chose to utilize a pre-trained model on the Places2 dataset (B. Zhou et al., 2014). Because it encompasses over 10 million diverse images across 400+ unique scene categories, it provides a robust and varied training environment for our model, enhancing its ability to detect and remove construction machinery in diverse scenarios. After evaluating numerous inpainting techniques mentioned in the literature, we selected the Large Mask Inpainting (LaMa) (Suvorov et al., 2022) algorithm due to its promising abilities to handle large missing regions, complex geometric structures, and high-resolution images. Using the pre-trained Places2 dataset and the LaMa algorithm, we sought robust and visually consistent inpainting results in our study. The details of the selected dataset and algorithm are given in Chapter 2.2.

In conclusion, construction machine segmentation and removal are essential for achieving high precision in earthwork volume calculations. This section has delved into the specifics of machinery segmentation and introduced the process of image inpainting and its significance in improving the 3D terrain model's accuracy. By employing these robust and cutting-edge techniques, we hope to optimize data processing quality in the earthwork monitoring process, thereby improving the precision of volumetric measurements. The following section will elaborate on the 3D reconstruction process, which converts these cleaned images into a three-dimensional digital terrain model.



Figure 4.3 Original and Inpainted Images using LaMa Algorithm (Suvorov et al., 2022)

4.4 3D Surface Reconstruction

Following the phase of construction machine segmentation and removal from aerial images, which ensures that the digital representation of the construction site accurately reflects the natural terrain, we advance to the next phase of the methodology. This next phase involves a more in-depth examination of the 3D reconstruction techniques, intending to create detailed and accurate 3D models of the terrain from a series of two-dimensional images. Utilizing photogrammetric techniques, Structure from Motion (SfM) (C. Wu, 2013) and Multi-View Stereo (MVS) (Furukawa & Ponce, 2010), the method progresses from 2D images to a detailed 3D model.

SfM-MVS can generate precise three-dimensional models of the terrain, which are essential for accurate volume calculations and excavation management. SfM

estimates the camera's position and the 3D structure of the scene simultaneously by identifying and matching features across the images, whereas MVS refines the 3D reconstruction by utilizing dense image correspondence. This results in an accurate and precise 3D representation of the terrain.

The SfM-MVS method is often combined in a pipeline to create accurate and detailed 3D models. The typical stages in a pipeline include Feature Extraction and Matching, Sparse Point Cloud Construction, Dense Point Cloud Construction, and Surface Reconstruction (Schonberger & Frahm, 2016). These are discussed in the following subsections.



Figure 4.4 SfM-MVS Flowchart

4.4.1 Feature Extraction and Matching

The initial stage, feature extraction and matching, is essential for establishing correspondence between unique points or features across multiple images. This procedure consists of two principal subtasks: feature extraction and feature matching.

Feature extraction identifies distinct points or regions within an image unaffected by various image transformations, such as rotation, scaling, and illumination changes. These specific points are known as features and serve as the basis for establishing overlapping correspondence between images. In the literature, numerous algorithms for feature extraction have been developed, each with advantages and disadvantages. The most prevalent algorithms are:

Scale-Invariant Feature Transform (SIFT) (Lowe, 1999): SIFT is a widely employed feature extraction algorithm that identifies key points and generates their descriptors

by analyzing local image gradients. SIFT features are invariant to image scale, rotation, illumination changes, and affine transformations.

Inspired by SIFT, Speeded-Up Robust Features (SURF) (Bay et al., 2008) is an alternative feature extraction method that offers faster computation while maintaining the same level of robustness. This is accomplished by SURF approximating the local image gradients with Haar wavelets, which can be efficiently computed using integral images.

ORB (Rublee et al., 2011) is a feature extraction algorithm that combines different algorithms. It is computationally efficient and provides binary feature descriptors, which enables Hamming distance-based matching to occur rapidly.

The selection of the feature extraction algorithm depends on the application's specific requirements, such as computational efficiency, robustness to image transformations, and matching precision. In our study, we have opted for the SIFT as the feature extraction algorithm of choice, driven by its robustness and the specific demands of our application.

After extracting features from each image, the next step is establishing correspondence between the features in overlapping images. This process, referred to as feature matching, compares feature descriptors to identify pairs of corresponding features. It can be completed by brute force or utilizing an algorithm.

Matching by brute force (Jakubović & Velagić, 2018): This technique involves comparing each feature descriptor in one image to each feature descriptor in the other. As matches, the descriptors with the closest Euclidean (for SIFT and SURF) or Hamming (for ORB) distances are considered. Although this comprehensive method guarantees the best matches, it is computationally costly.

The Fast Library for Nearest Neighbors (FLANN) (Muja & Lowe, 2009): FLANN is an approximate nearest neighbor search algorithm that speeds up the matching procedure by utilizing hierarchical clustering techniques. FLANN organizes feature descriptors into a tree structure, allowing for faster searches for nearby neighbors.

FLANN may not always find the optimal match, but it offers a significant speedup compared to the Brute Force matcher with a negligible loss in matching accuracy.

The feature extraction and matching stage of the SfM-MVS pipeline is essential for establishing correspondence between features across multiple images. The selection of feature extraction algorithms and matching techniques is determined by the application's specific requirements, balancing the tradeoffs between computational efficiency, robustness, and matching precision. We have chosen the FLANN for feature matching in our study due to its effective balance between speed and accuracy.

Once feature correspondence has been established between images that overlap, camera pose estimation is the next step. The objective is to calculate each image's relative camera poses, including positions and orientations. Precise camera pose estimation is essential for triangulating matched features and facilitating the subsequent 3D reconstruction process. The essential and fundamental matrices capture the geometric relationship between corresponding points in two images. Once these matrices have been computed, the camera poses can be extracted using a technique known as decomposition, which isolates the rotation and translation components of the relative camera motion. Next in the SfM-MVS pipeline is the generation of a sparse 3D reconstruction, accomplished by triangulating the matched features and simultaneously optimizing the camera poses and 3D points using bundle adjustment (Figure 4.5).

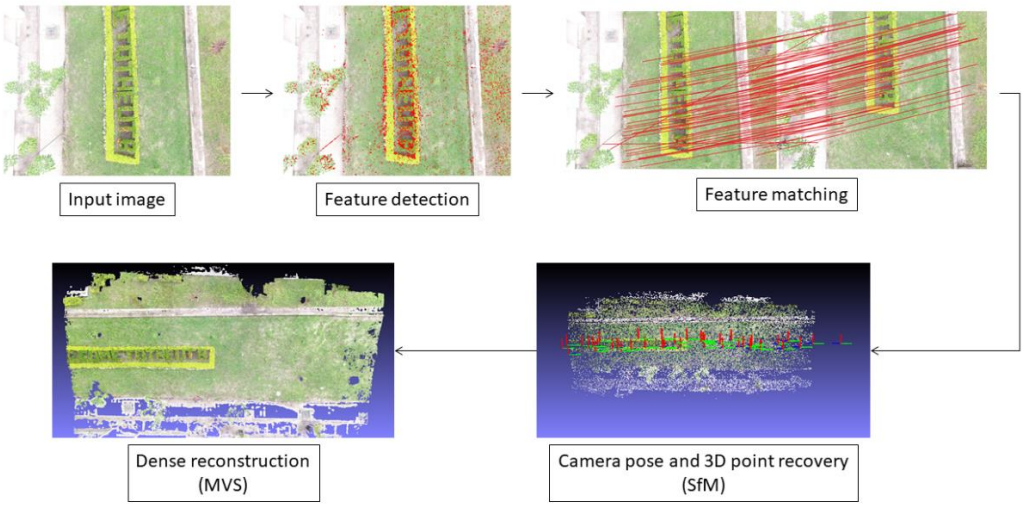


Figure 4.5. SfM-MVS Pipeline Visualization (Meza et al., 2019)

4.4.2 Sparse Point Cloud Construction

The subsequent step in the SfM-MVS pipeline is to generate a sparse point cloud construction of the scene after estimating the relative camera poses in the previous stage. This step involves triangulating the matched features to generate a sparse point cloud representing the 3D structure. Sparse point cloud construction is a crucial step in the pipeline because it provides an initial estimate of the 3D form of the scene, refined in the dense point cloud construction stage. This section gives insight into the sparse point cloud construction stage's triangulation procedure and bundle adjustment optimization.

Triangulation is the process of determining the 3D position of a scene point from its 2D projections in multiple images (Bianco et al., 2018). Given the relative camera positions and matching feature correspondences, triangulation involves locating the 3D point that best satisfies the projection constraints imposed by the cameras. In particular, this is achieved by minimizing the reprojection error, which is the difference between the observed 2D feature positions and the projected 3D point positions.

Bundle adjustment is a global optimization technique to refine camera simultaneously poses and 3D points (James et al., 2017). Bundle adjustment minimizes the overall reprojection error across all images and matching features. Typically, iterative techniques are used to solve this nonlinear optimization problem. Not only does bundle adjustment improve the accuracy of the sparse point cloud construction, but it also provides a measure of uncertainty for the estimated camera poses and 3D points. This uncertainty information helps evaluate the construction quality and guides the following stages of the SfM-MVS pipeline.

In conclusion, the sparse point cloud construction stage is crucial for producing an initial estimate of the scene's 3D structure by triangulating matched features and refining camera poses and 3D points using bundle adjustment.

4.4.3 Dense Point Cloud Reconstruction

Dense 3D reconstruction concentrates on refining and enhancing the sparse point cloud obtained in the sparse construction stage. The primary objective of dense construction is to generate a more comprehensive and accurate scene representation by incorporating a significantly larger number of 3D points. This increased density allows for the capture of finer details, resulting in a more realistic and detailed 3D model.

During this step, MVS algorithms compute depth information for each pixel in the image. MVS algorithms compare the pixel intensity values of multiple images captured from different perspectives. The depth maps obtained in the preceding step are combined to accurately and consistently represent the scene's depth information. This fusion process mitigates inconsistencies and errors in individual depth maps by utilizing information from multiple perspectives. Techniques such as volumetric depth map fusion and probabilistic depth map fusion can be utilized to achieve this objective.

Once the depth maps have been fused, the depth information can be back-projected into 3D space to generate a dense point cloud. Back-projection refers to converting 2D depth map data into a 3D representation. This densification process generates a significantly greater number of 3D points than the sparse point cloud, thereby capturing finer scene details. To ensure the accuracy and dependability of the resulting dense point cloud, it is crucial to filter and remove outliers during the densification process. They can be removed by the techniques such as smoothing, noise removal, and outlier filtering that improve the dense point cloud's overall accuracy and visual quality.

In conclusion, dense point cloud construction is a crucial step in the SfM-MVS pipeline that densifies the sparse point cloud to produce a more accurate and detailed scene representation. Multiple steps comprise the procedure, including depth map estimation, depth map fusion, point cloud densification, and dense point cloud refinement. The dense point cloud that results pave the way for surface reconstruction, resulting in a complete and aesthetically pleasing 3D model of the scene.

4.4.4 Surface Reconstruction

The dense point cloud obtained through the SfM-MVS pipeline requires a surface reconstruction for volume calculation. The surface reconstruction stage involves a series of carefully executed preprocessing steps to ensure accuracy. These steps are crucial for reducing computational demands, defining the geometry and surface structure of the data, and reconstructing the surface. This section will examine each of these crucial preprocessing steps — Region of Interest (ROI) Cropping, Voxel Downsampling, Normal Computation, and Poisson Surface Reconstruction— to comprehensively understand the methodology (Figure 4.6).

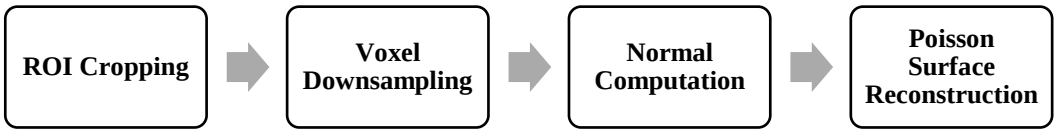
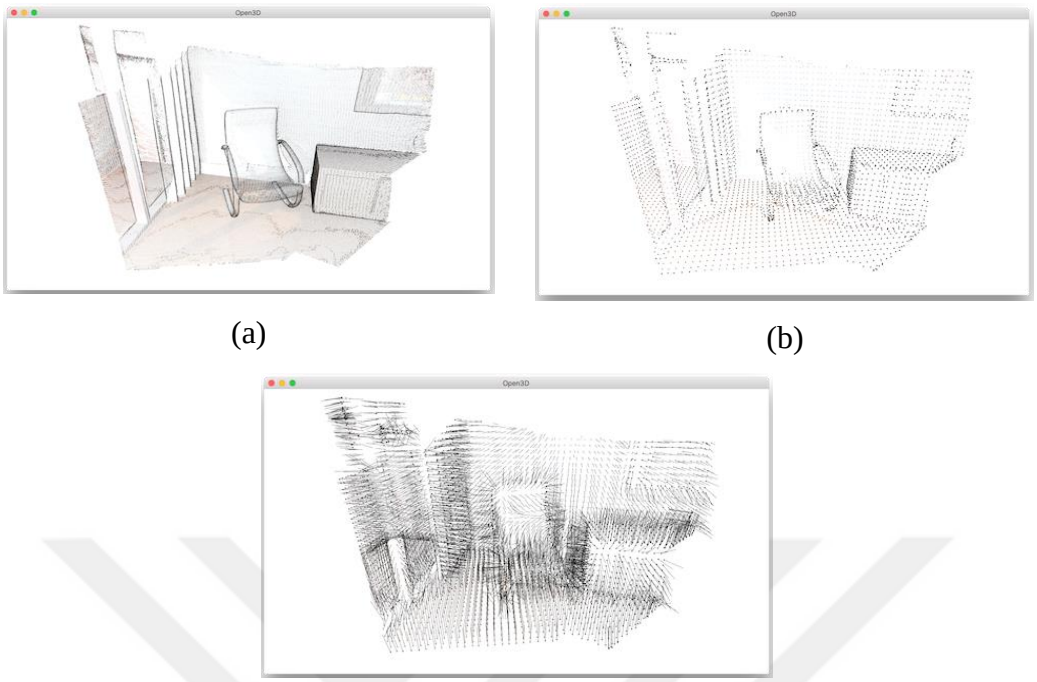


Figure 4.6 Preprocessing Steps to Calculate Volumes

In the first step of preprocessing, the ROI is extracted from the point cloud. This is crucial because aerial images, our primary data source, typically include the construction site and the surrounding neighborhood. These extra, irrelevant regions are also incorporated into the point cloud, which adds unnecessary complexity and potential inaccuracy to the data. We specify the excavation area with a polygon to isolate and remove these periphery elements, thereby effectively differentiating the construction site. Using this polygonal boundary as a reference, the point cloud is subsequently cropped. Consequently, all computations and further analysis are restricted to the data contained within this ROI, enhancing the accuracy and applicability of the results.

The next step is to downsample the point cloud using voxel downsampling. Voxel downsampling is one of the standard techniques used in processing 3D data, particularly in point cloud processing. This method segments the entire 3D space into cubical units known as voxels (the 3D equivalent of pixels). The initial high-resolution point cloud is effectively "downsampled" by replacing the set of points within each voxel with a single representative point (Figure 4.7a and Figure 4.6b). Typically, this representative point is the centroid or the mean of all points within the voxel. In addition to the point's location, other attributes, such as color and normal vectors, if present, are also averaged. The resultant downsampled point cloud is considerably smaller and, therefore, easier to process while retaining a representative overview of the original three-dimensional structure. This process is used as the initial preprocessing step to reduce the computational requirements of subsequent operations.



(a) (b)

(c)

Figure 4.7. 3D Demonstrations (a) Point Cloud (b) Voxel Downsampled Point Cloud (c) Calculated Normals (Q.-Y. Zhou et al., 2018)

Once the voxel downsampling process has successfully reduced the density of the point cloud data, the next essential step in the preprocessing pipeline is the computation of normals, which further defines the geometry and surface structure of the 3D data. Normals are critical elements in 3D data processing and analysis, as they help define the geometry and orientation of surfaces. They serve as fundamental surface representation characteristics for a 3D point cloud. At a given point on a surface, the normal is the vector that is perpendicular to the surface (Figure 4.7c). Furthermore, the orientation and consistency of these normals can significantly affect the quality of the resulting 3D mesh. Therefore, calculating and correctly orienting normals are essential processing steps. Using a covariance analysis of the points adjacent to each point of interest, normals are computed within a point cloud.

The final step, surface reconstruction, is used to generate a coherent and realistic three-dimensional representation of the scene, which is essential for volume

calculation. This phase's primary objective is to create a polygonal model, or 3D mesh, from the dense point cloud. A mesh is a geometric structure composed of connected vertices, edges, and faces, forming a continuous surface that encapsulates the scene's 3D structural layout. Several surface reconstruction algorithms, such as Poisson surface reconstruction (Kazhdan et al., 2006), Delaunay triangulation (Lee & Schachter, 1980), and Marching Cubes (Lorensen & Cline, 1987), can be used to achieve this objective. Each algorithm is designed to generate a continuous, smooth surface that accurately represents the point cloud data while minimizing the possibility of artifacts and noise.

In this study, the 3D mesh was generated using the Poisson surface reconstruction method. The Poisson method was chosen because of its ability to solve a regularized optimization problem, resulting in a smooth surface. This attribute gives the Poisson surface reconstruction an advantage over other methods, which frequently produce non-smooth results because the vertices of the resulting triangle mesh correspond directly and unmodified to the points of the point cloud. The depth, which defines the extent of the octree used for surface reconstruction, is an essential parameter of the Poisson surface reconstruction. Consequently, this parameter immediately affects the resolution of the resulting triangle mesh. A greater depth value corresponds to a mesh with greater granularity. A demonstration of the Poisson surface reconstruction algorithm is given in Figure 4.8.



(a) (b)

Figure 4.8. Poisson Surface Reconstruction (a) Input Point Cloud (b) Output 3D Mesh (Q.-Y. Zhou et al., 2018)

4.5 Volume Calculation

After completing the surface reconstruction steps, we move on to the volume calculation phase, which is the issue at the heart of the proposed method. We will discuss how processed mesh data are used to determine the volume of earthworks. The aim is to compare the output mesh from each site visit to the output mesh from the prior site visit. This comparative analysis enables us to calculate the differences in volume between the two. This is achieved by employing a volumetric difference algorithm, which determines the fill and cut volumes on the site. The resulting volume calculation is an essential set of data that informs us of the construction site's earthwork progress over time and the amount of work completed between site visits.

There are two steps involved in calculating the volumetric difference. The pseudocode of the ray length calculation and volume calculation is given in Figure 4.9. The ray length calculation involves projecting rays from each triangle vertex from the first mesh and determining the length of these rays until they intersect with the second mesh. This provides a measurement of the difference between the two 3D

meshes. After this, the volume is computed by integrating these differences over the entire mesh. The pseudocode provided below describes these algorithms in detail.

Algorithm 1 Ray Length Calculation

```
loop Over vertices of first mesh
    Draw rays in +Z and -Z direction
    Find which rays hits the surface of second mesh
    Calculate ray length
    if +Z direction ray hits then
        Multiply ray length with positive
    else
        Multiply ray length with negative
    end if
    Save ray lengths in an array
end loop
```

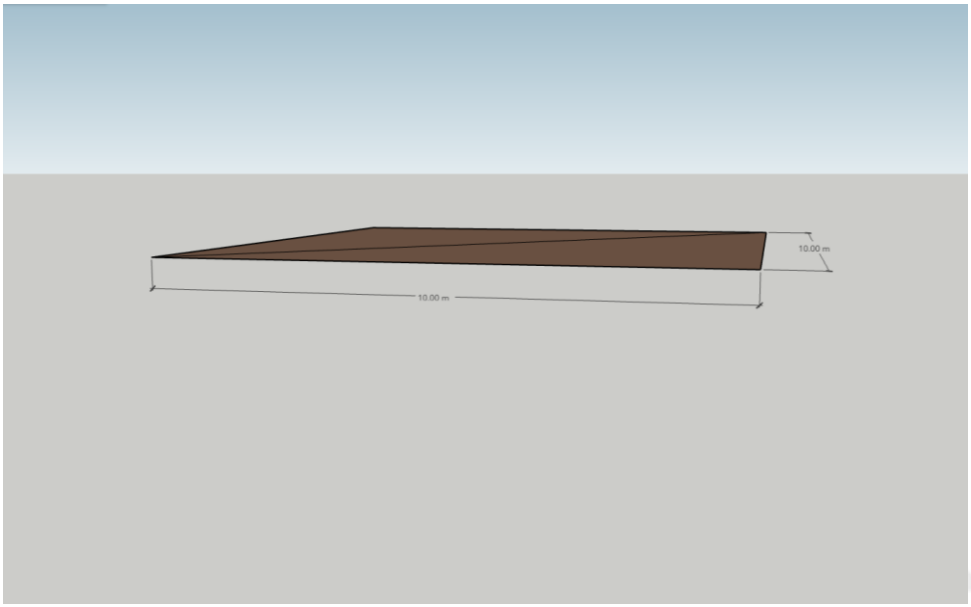
Algorithm 2 Volume Calculation

```
fillVolume = 0
cutVolume = 0
loop Over triangles of first mesh
    averageRayLength = the average ray length of three vertices (using previously calculated
ray lengths)
    crossSectionalArea = the area of the triangle, setting Z coordinates to zero
    individualVolume = averageRayLength * crossSectionalArea
    if individualVolume ≤ 0 then
        cutVolume ← cutVolume + individualVolume
    else
        fillVolume ← fillVolume + individualVolume
    end if
end loop
```

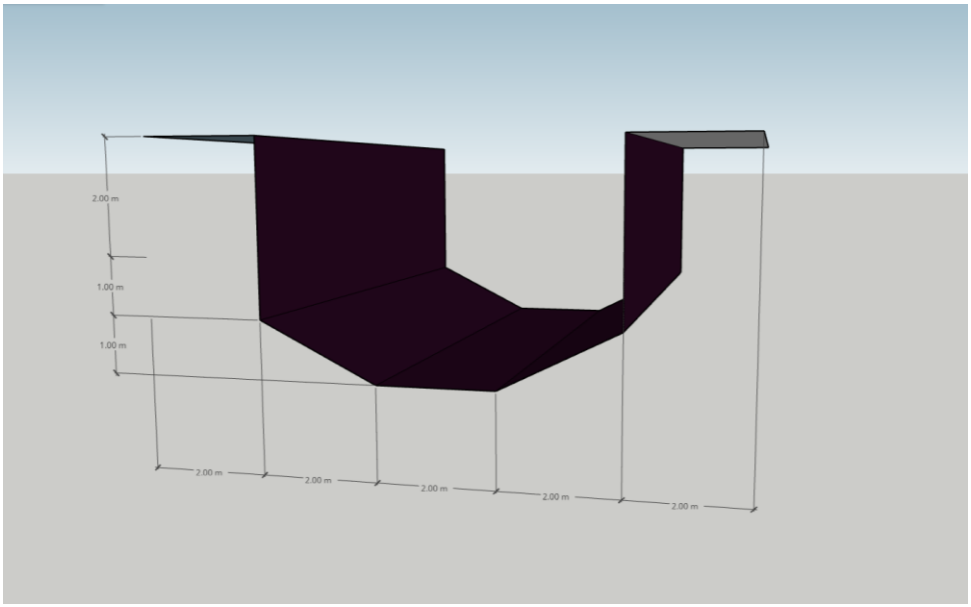
Figure 4.9. Ray Length and Volume Calculation Pseudocode

We utilize an artificial object with known geometry and volume to validate the proposed volume calculation method. The object is triangulated, and the volume calculation algorithm is applied to this triangulated form. This provides a benchmark against which we can verify the accuracy of our calculations. The first and second meshes are shown in Figure 4.10a and Figure 4.10b. The exact cut and fill volumes are 100 m^3 and 80 m^3, respectively. The cut volume is illustrated in red color, and the fill volume is illustrated in green color in Figure 4.10d. The artificial object is triangulated, and the proposed ray length and volume calculation algorithm is

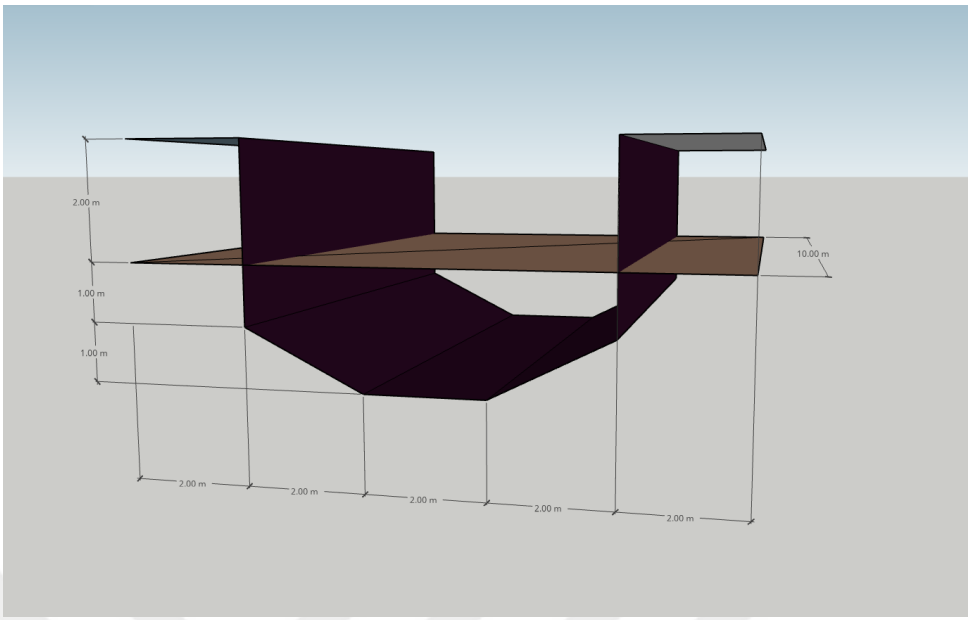
performed. This can be seen in Figure 4.11. The fill volume is calculated as 99.373 m^3 (deviation: 0.63%), cut volume is 78.1245 m^3 (deviation: 2.34%). As the number of triangles increases, the calculated results tend to converge with the exact results. Consequently, the outcomes align well with the actual volume measurements.



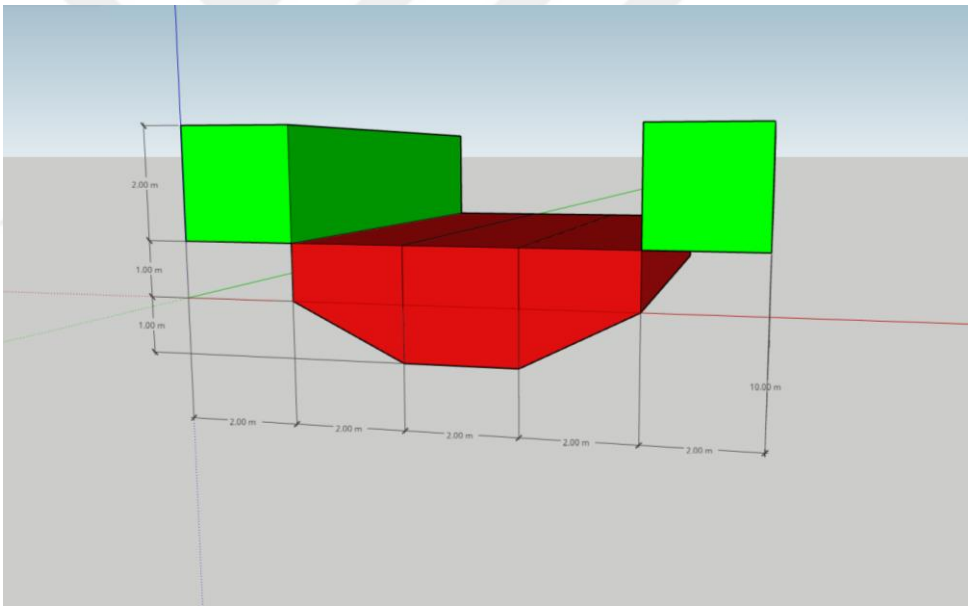
(a) First Mesh (Demonstration of the Surface Before Excavation)



(b) Second Mesh (Demonstration of Surface After Excavation)



(c) First and Second Mesh Alignment



(d) Illustration of Cut and Fill Volumes

Figure 4.10. Volumetric Difference Calculation of an Artificial Object

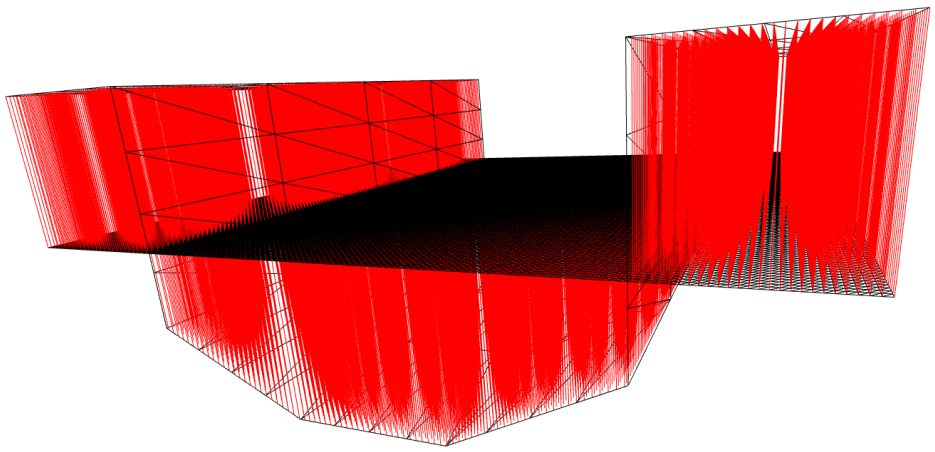


Figure 4.11. Illustration of Rays Between Two Meshes

In conclusion, the methodology described above is a comprehensive approach to automating the monitoring of earthwork progress. Utilizing UAV imaging and deep learning and geometry processing algorithms, the proposed system has the potential to significantly improve the precision, efficiency, and safety of construction site monitoring. While the methodology has been meticulously designed to address the inherent difficulties of earthwork monitoring, it must be evaluated under real-world conditions to determine its performance. This brings us to the subsequent section of the study. The following chapter describes the methodology's real-world application and evaluation in depth.

CHAPTER 5

FIELD EXPERIMENTS

After examining the method's theoretical background and specific details, it is time to assess its applicability and practicability in the field. This chapter describes the experiments designed to comprehensively evaluate the performance of the proposed system in a real-world construction setting. It will provide an in-depth analysis of the practical application of UAV imaging, deep learning, and geometry processing techniques by presenting, analyzing, and discussing the results.

5.1 Study Area

The study area is in the Koru District of Ankara, Türkiye (Figure 5.1). An empty plot of land in the region is set to undergo a significant transformation after years of inactivity. Figure 5.2 depicts plans to transform the area into a lively town square with an ornamental pool, observation terrace, seating areas, and a theater. In addition, a parking garage will be built beneath the courtyard to provide parking space for the vehicles of residents and visitors.



Figure 5.1. Location of the Construction



Figure 5.2. Illustration of the Project

The renovated square will encompass a total area of 9,211 square meters and provide various urban facilities. These facilities will consist of decorative pools, water curtains, a theater for cultural events, viewing terraces for scenic enjoyment, relaxation areas, and information kiosks (Figure 5.3a).

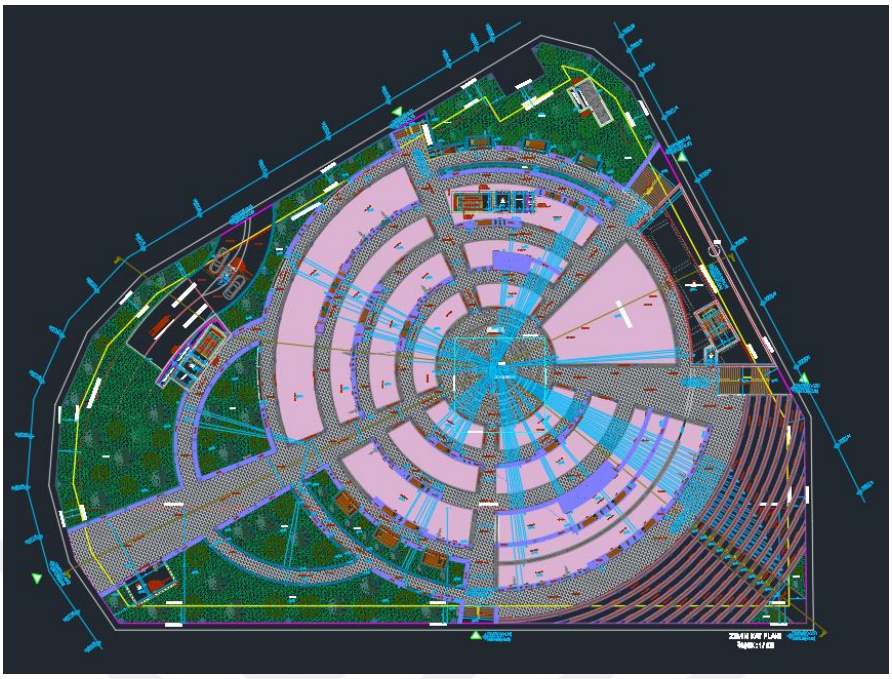
To accommodate the area's parking needs, a two-story garage covering 12,952 square meters will be constructed beneath the square. This spacious garage will provide parking for various vehicles, including 330 standard cars, 23 electric vehicles, and 22 vehicles owned by people with disabilities.

It will be necessary to excavate to a total depth of 9.20 meters as part of the construction process, and piles surround three sides of the garages. This project will ensure the underground parking garage has the proper foundation and space (Figure 5.3b).

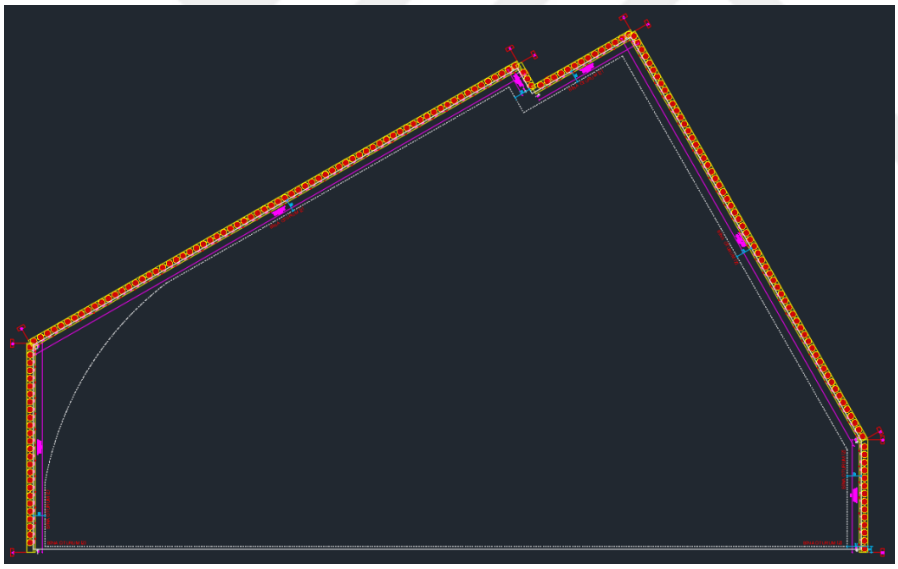
The study area has an ideal location for monitoring earthwork progress. In the course of the construction phase, the following machines were utilized:

1. Excavators
2. Backhoe loaders
3. Dump trucks
4. Concrete mixers
5. Drilling machines
6. Anchor installation machines
7. Mobile cranes

The machinery and general view of the excavation area can be seen in Figure 5.4



(a) Architectural Drawings



(b) Geotechnical Plan Drawings

Figure 5.3. CAD Drawings of the Construction



Figure 5.4. Views of the Excavation Area

Before construction began, the contractor conducted site surveys. As depicted in Figure 5.5, seven boreholes were drilled, and soil samples were gathered from the field. The field and laboratory tests revealed that the top 0.5 meters of the soil is composed of vegetable soil. Below this layer, up to a depth of 15 meters, the soil is highly plastic, slightly gravelly sandy, silty clay (CL-CH). In the examined area, groundwater was detected at an average of 8.0 and 12.5 meters depth. It is hypothesized that sand-gravel bands exist at these depths and that these regions serve as water sources. The profile of the soil is presented in Figure 5.6. Given that the excavation depth falls within this soil layer, it is reasonable to assume that the same type of soil will be encountered throughout the excavation.



Figure 5.5. Borehole Plan of the Construction Site

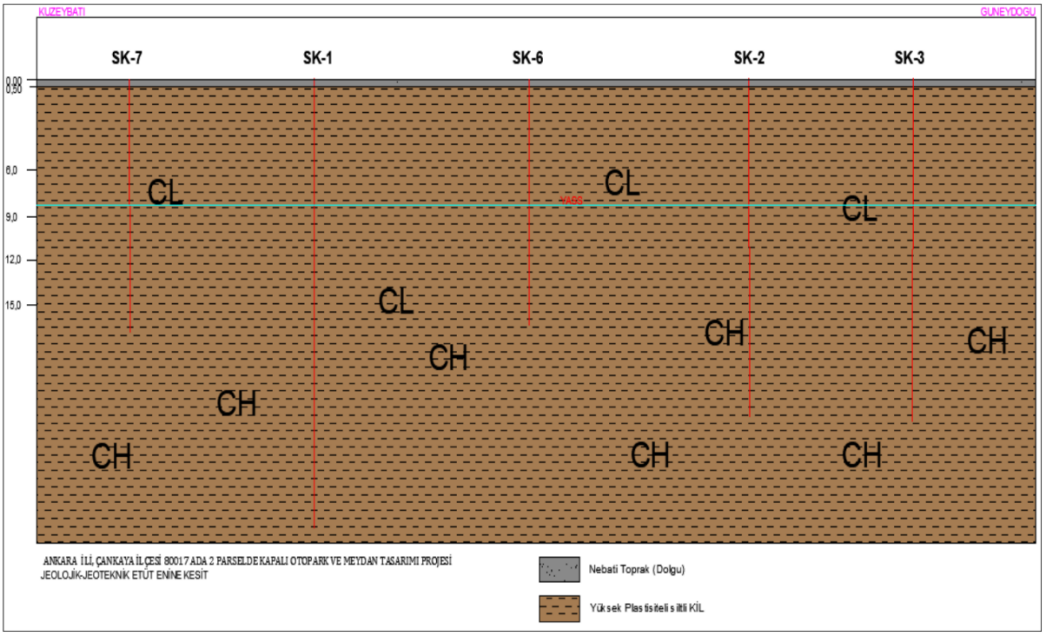


Figure 5.6. Soil Profile of the Construction Area

In conclusion, the construction of this parking garage provides an ideal study area for evaluating earthwork progress. The complexity of the project, coupled with the variety of construction machines and the unique soil profile, provides a favorable environment for the application of cutting-edge techniques in 3D data processing and volume calculation. Throughout the construction period, the ongoing analysis and monitoring of earthwork provided crucial insights that enriched this study.

5.2 Data Collection

Data acquisition is a critical step in any study involving site monitoring and evaluation, as it serves as the foundation for subsequent analysis and conclusions. In this particular instance, we investigated the use of two prominent techniques, namely UAV imaging and terrestrial laser scanning (TLS), to collect data on the progress of the earthworks. These methods enabled us to collect large amounts of data providing a thorough understanding of the project's development over 2.5 months. The UAV imaging process was carried out throughout multiple visits; T1 (05.01.2023), T2 (10.01.2023), T3 (01.02.2023), T4 (23.02.2023) and T5 (15.03.2023). These visits provide a comprehensive view of the construction site.

In contrast, TLS data were collected exclusively on T1 and T2 dates. TLS played a crucial role in comparing the data and computations derived from the UAV images. Figure 5.7 depicts the construction site's data collection timelines.



(a) T1 (05.01.2023)



(b) T2 (10.01.2023)



(c) T3 (01.02.2023)



(d) T4 (23.02.2023)



(e) T5 (15.03.2023)

Figure 5.7. Site Visits for the Construction Site

5.2.1 UAV Imaging

UAV imaging is one of the primary methods used to collect precise and high-resolution data during ongoing construction. This method uses the adaptability of UAV technology equipped with high-resolution cameras, enabling comprehensive and detailed imaging of the subject site. This section explores the various aspects of the aerial imaging procedure, including the acquisition of ground control points (GCPs), the selection of the UAV, the planning of UAV flights, and the data collection process.

Several crucial steps were taken throughout this construction project to ensure accurate data collection and effective site monitoring. To improve image alignment for 3D point cloud generation and positional accuracy within the photogrammetric process, high-precision cartographic points known as GCPs were gathered around the construction site's perimeter. Care was taken to ensure that the location of these measurements remained unchanged throughout the construction. Figure 5.8 and Figure 5.9 illustrate the places used to obtain measurements from various points surrounding the construction site.



Figure 5.8. Locations of the Ground Control Point



Figure 5.9. Details of GCP Surveying

Surveyors frequently use the real-time kinematic Global Navigation Satellite System (GNSS) to measure GPS coordinates on construction sites. In this instance, a GNSS receiver of the Hi-Target V100 model was employed, with data-link service provided by the Türkiye CORS (Continuously Operating Reference Stations) network system. TUREF / TM30 served as the coordinate system. The GCP locations are listed in the Table 5.1.

Table 5.1 GCP World Coordinates

ID	X	Y	Z
GCP1	729644.48	4420157.71	910.57
GCP2	729714.45	4420161.74	910.26
GCP3	729750.63	4420094.70	911.59
GCP4	729769.82	4420064.54	911.88
GCP5	729691.18	4420019.41	913.83
GCP6	729636.62	4420096.50	911.73

It is essential to select the proper camera and UAV to acquire high-quality images with sufficient detail for 3D surface generation. When selecting a camera, consideration must be given to resolution, sensor size, lens quality, and dynamic range. A camera with a higher resolution can capture more detailed images, resulting in more precise 3D reconstructions. Nevertheless, larger image sizes may also increase the computational complexity of subsequent 3D surface generation pipeline stages.

In this study, DJI Mavic Pro was utilized for imaging, as it possesses several advantageous features, including its miniature size, extended flight time, and precise positioning capabilities. Its foldable body makes it easier to transport the drone (Figure 5.10). The DJI Mavic Pro features a 21-minute flight time, 4000x3000 image resolution, and 3840x2160 4K video at 24/25/30 fps. Its maximum transmission distance is 2.5 miles, allowing for greater freedom and range (DJI, 2023).



Figure 5.10 Selected Drone – DJI Mavic Pro (DJI, 2023)

To ensure optimal image quality and reduce potential errors caused by inclement weather, sunny weather conditions were preferred for image capture data collection. Throughout the image capture process, construction equipment was continuously in operation. However, since the images were captured from above, this did not pose any problems for data collection and safety.

Flight planning is essential for acquiring aerial images with UAV-mounted cameras to ensure adequate image overlap and scene coverage. For successful feature matching and camera pose estimation, a 60-80 percent overlap in the flight and cross-flight direction is advised (Colomina & Molina, 2014). Flight planning should consider flight altitude, camera field of view, and speed to achieve the desired image overlap and ground resolution.

Utilizing the DroneDeploy mobile application for flight mission planning, a consistent mission plan was applied to each flight in this study (Dronedeploy, 2023). The flights were conducted at a constant altitude of 40 meters, ensuring data consistency and flight-to-flight comparability. In addition, image acquisition followed a front and side overlap of 80%, optimizing the quality of the facilitating accurate data analysis. This standardized method enabled reliable comparisons and insightful conclusions to be drawn from the collected data. All drone flights were

autonomous, ensuring consistent flight paths and minimizing the possibility of human error during data collection. An interface of the mission plan is shown in Figure 5.11.



Figure 5.11 Flight Mission Plan Used for All UAV Flights

UAV data collection was carried out for 2.5 months. After each flight, images were transferred from the UAV's SD card to a desktop computer and then uploaded to cloud storage, where they were preserved for future analysis. Aerial imaging helps us paint an accurate and detailed picture of the project's status at various stages of its life cycle.

5.2.2 Terrestrial Laser Scanning

To compare the data obtained from aerial imaging, this method was utilized primarily during the T1 and T2 construction periods. We can generate high-resolution 3D models of the construction site using cutting-edge laser scanner devices. In this section, we discuss the specifics of this data collection method,

including the selection of the laser scanner, the establishment of scanning stations, and the data collection, registration, and georeferencing processes.

We employ the Leica BLK360 laser scanner device, a 3D scanning instrument, for comparison. The BLK360 features an array of impressive specifications, including an integrated spherical imaging system and thermography panorama sensor system. These include a 1 kg lightweight design, compliance with Laser Class 1 (according to IEC 60825-1:2014), a range of up to 60 meters, and a measurement rate of up to 360,000 points per second (Leica, 2023). This sophisticated instrument ensures accurate and trustworthy results, allowing for precise data collection and evaluation.



Figure 5.12 Laser Scanner Device – Leica BLK360 (Leica, 2023)

At T1 and T2, two site visits were conducted to collect comprehensive data for the study. Due to the presence of daytime-operating construction machinery, data collection was scheduled to occur after working hours. Due to the technology's

reliance on laser points rather than visible light for precise measurements, the laser scanning process was unaffected by the waning of sunlight.

On-site sphere targets were implemented to facilitate the point cloud registration process. These lightweight and transportable polystyrene targets illustrated in Figure 5.13 were available in two different sizes, 15 cm and 20 cm diameters. There are 20 targets mounted on tripods and strategically positioned between scanning stations, with a 50-meter maximum distance between them. This placement ensured adequate coverage and made data merging during post-processing more efficient.



Figure 5.13 Polystyrene Sphere Targets

For each site visit, 27 laser scanning stations were established. At each station, the BLK360 laser scanner was utilized to collect high-density and resolution data. The scanning procedure took approximately five minutes per station, allowing for the

detailed acquisition for subsequent analysis. The locations of scanner stations are shown in Figure 5.14 and Figure 5.15. This data collection strategy allowed for comprehensive site documentation and provided a solid foundation for subsequent processing and evaluation.

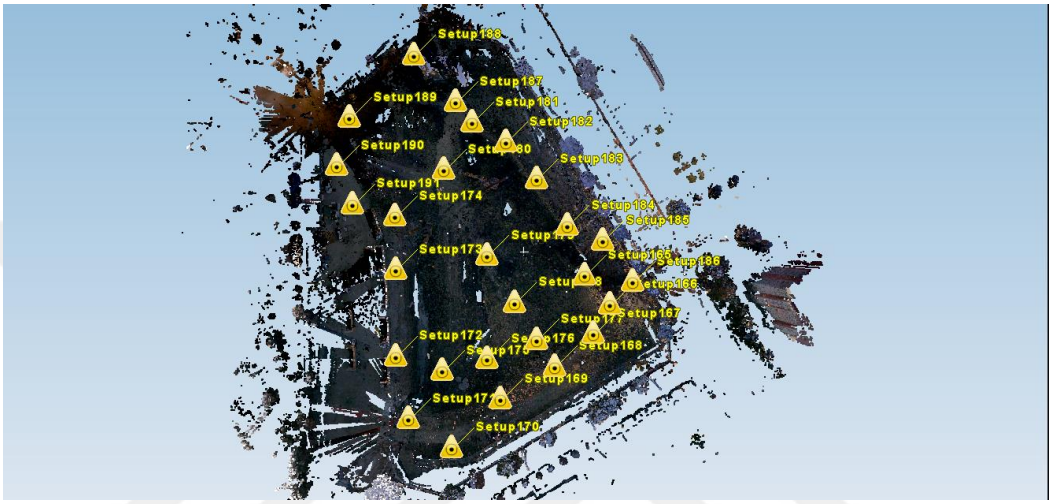


Figure 5.14 T1 Laser Scanning Station Locations

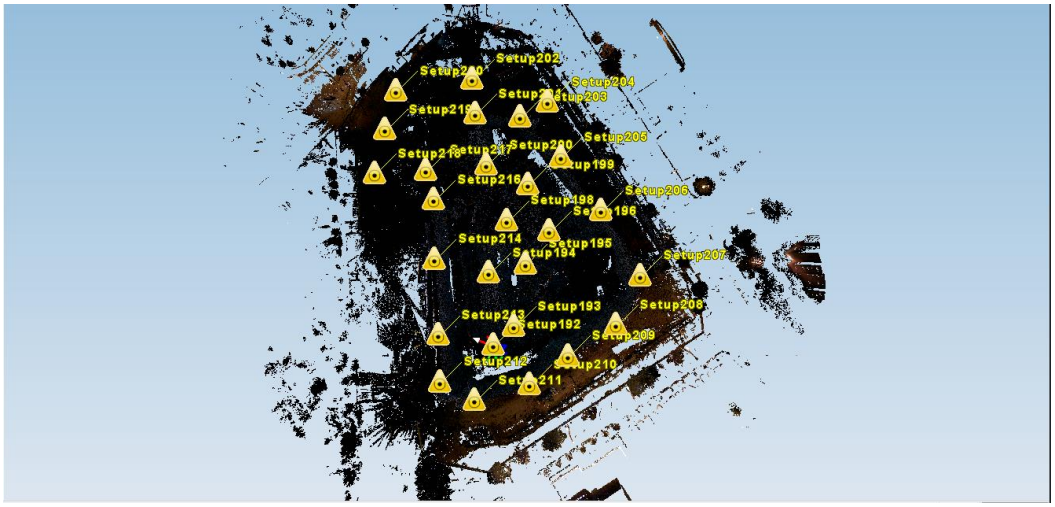


Figure 5.15 T2 Laser Scanning Station Locations

Registration of point clouds is an essential step in processing laser-scanning survey data. The point clouds in this study were registered using the Leica Geosystems Cyclone 9.3.2 software. To facilitate the registration process, sphere targets were marked on the point clouds, and a sphere mesh object was fitted to the point groups of spherical targets, producing a robust and precise representation of the sphere centers. At least three sphere targets were marked for each station's point cloud. The registration step was accomplished by minimizing the distances between the centers of spheres in each cloud, resulting in the precise alignment of the point clouds. The registration results for sphere targets are detailed in Table 5.2. Mean absolute error measures the average absolute difference between the true positions of points and their positions as predicted by the registration algorithm.

Following the registration of the point clouds, the data were georeferenced to ensure accurate positioning within a real-world coordinate system. Three GCPs were utilized to georeference the registered point clouds in this study. The data was accurately georeferenced by employing a transformation procedure that minimized the residual errors between these three GCPs and their corresponding points in the point cloud. Now positioned within a real-world coordinate system, the georeferenced point cloud data is ready for further analysis.

Table 5.2 Point Cloud Registration and Georeferencing Results

Reference	**Mean Absolute Error**
T1 Target registration	0.1 cm
T1 Georeferencing	0.7 cm
T2 Target registration	0.7 cm
T2 Georeferencing	0.6 cm

In conclusion, the data acquisition phase of the study, which included both UAV imaging and laser scanning, was carried out successfully. We have obtained accurate and detailed spatial data by utilizing sophisticated equipment types, as seen in Figure 5.16. This wealth of visual and spatial information will be the foundation for subsequent analyses.

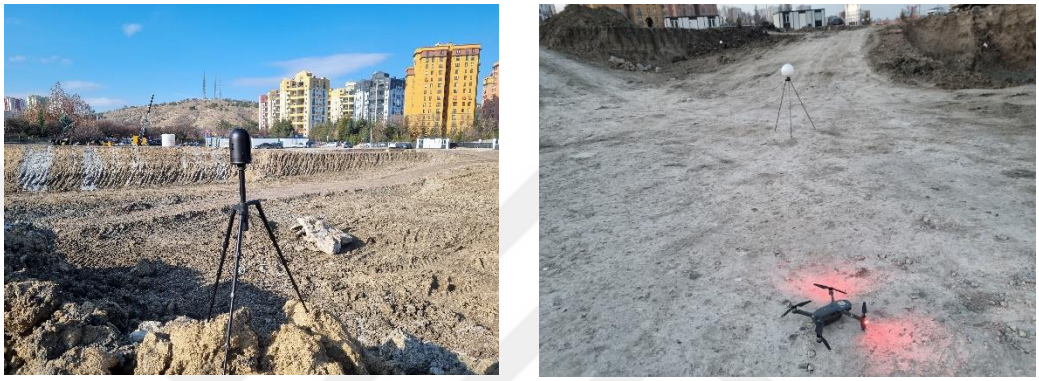
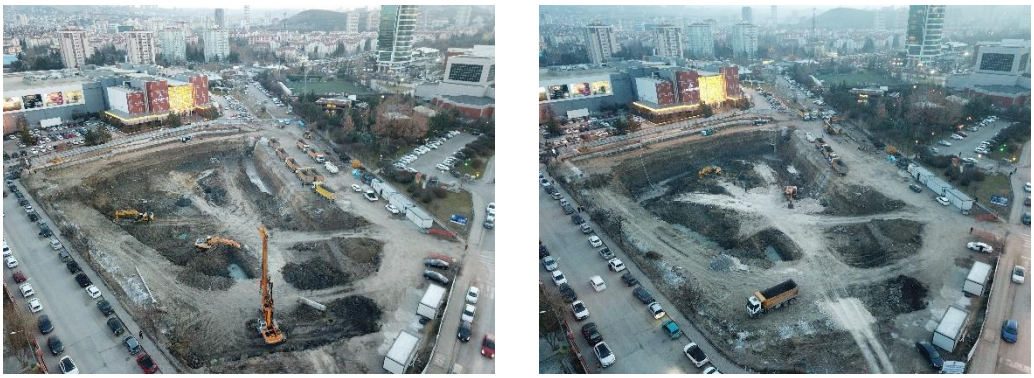


Figure 5.16 Laser Scanner and UAV on the site

5.3 Construction Machines Segmentation

This section addresses the steps of the automated construction machine segmentation procedure. The required data from the construction site was gathered from the previous steps. The UAV flight mission used for this purpose yielded 230 images for both T1 and T2 visits. All of these images are in 4000 x 3000 pixels resolution and geo-referenced.

At T1 and T2 periods, the construction site contained five distinct types of equipment: dump trucks, excavators, backhoe loaders, cars, and a category labeled "other" that included drilling machines, anchor installation machines, and concrete mixers. Given the site's layout surrounded by numerous streets, many cars were also spotted near the construction site (Figure 5.17).



(a) T1 (b) T2

Figure 5.17 Aerial View of Construction Site

The ground truth of construction machine objects boundaries are required for both T1 and T2 datasets to calculate segmentation errors. Humans usually annotate the ground truth manually for object segmentation, mark bounding boxes around objects of interest in an image, and label them with their corresponding class. In this study, since instance segmentation was performed, pixel-level masks were prepared. Table 5.4 displays statistics of the construction machine instances. There were 1708 instances, including 174 dump trucks, 57 excavators, 10 backhoe loaders, 1404 cars, and 63 other categories for T1. For T2, the total number was 1422 containing 142 dump trucks, 55 excavators, 1144 cars, and 81 other instances. Backhoe loaders were not seen in the T2 dataset images.

Table 5.3 Number of Instances Included in the Images

Dataset	D.Truck	Excavator	B.Loader	Car	Other
T1	174	57	10	1404	63
T2	142	55	0	1144	81

The trained machine learning mentioned in Chapter 3 was used to detect construction machines on the datasets T1 and T2. MMdetection library, a PyTorch-based open-

source object detection toolbox, was utilized for detection and segmentation. All steps of the proposed method was implemented in Python. The mAP outcomes for the T1 and T2 datasets are displayed in Table 5.4. The mAP for the T1 dataset was 66.6, while the mAP_{50} and mAP_{75} were 87.3 and 82.3, respectively. For the T2 dataset, the mAP was 64.5, with mAP_{50} and mAP_{75} values of 87.2 and 78.7, respectively The definition of mAPs is detailed in Chapter 3.5.

Table 5.4 Segmentation Results

Dataset	**mAP**	**mAP_{50}**	**mAP_{75}**
T1	67.6	87.3	82.3
T2	64.5	87.2	78.7

Table 5.5 provides a breakdown of the AP results by class. The AP values for T1 were as follows: 97.8 for dump trucks, 98.9 for excavators, 96.3 for backhoe loaders, 97.5 for cars, and 45.9 for other vehicles. In the T2 dataset, the AP values for dump trucks were 92.8, excavators were 96.4, backhoe loaders were not applicable, cars were 96.5, and others were 63.1. Figure 5.18 and Figure 5.19 further illustrate the precision of segmentation, depicting correct and incorrect segmentation, respectively. A discussion of these results is given in Chapter 5.7.

Table 5.5 Classwise AP Results (IoU=50%)

Dataset	**D.Truck**	**Excavator**	**B.Loader**	**Car**	**Other**
T1	97.8	98.9	96.3	97.5	45.9
T2	92.8	96.4	-	96.5	63.1

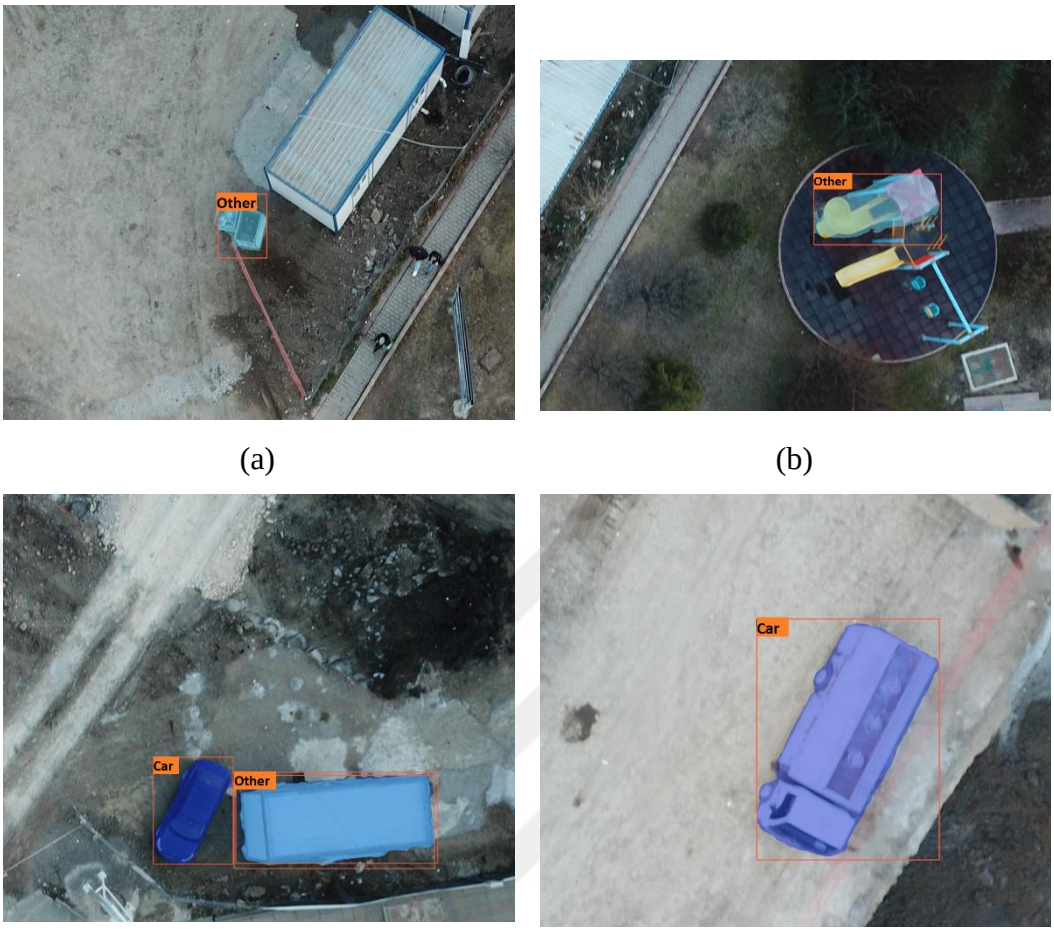


(a) (b)

(c) (d)

(e) (f)

Figure 5.18. Correct Object Segmentations



(a) (b)

(c) (d)

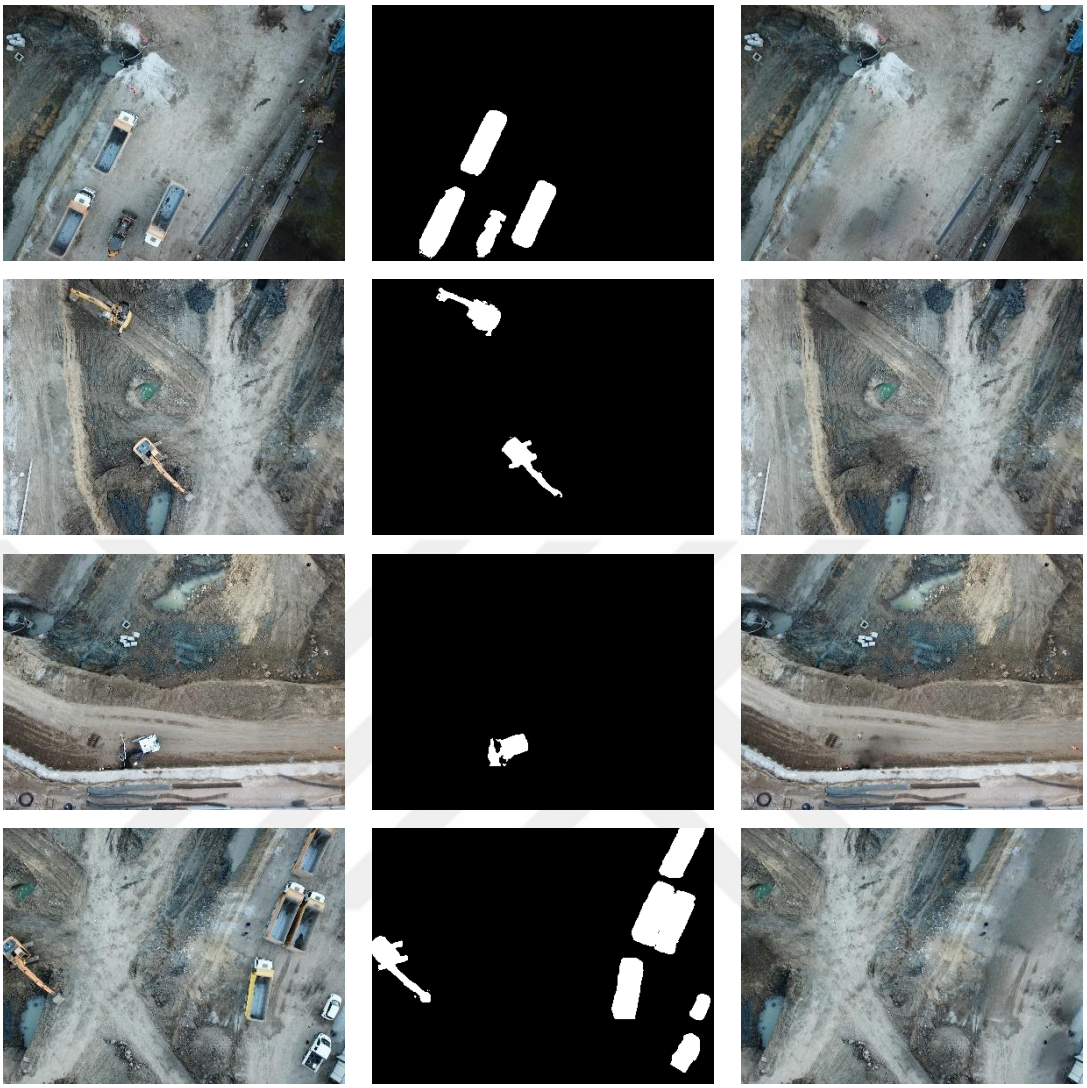
Figure 5.19. Incorrect Object Segmentation

The construction machine segmentation stage plays an essential role in the automated monitoring of earthwork progress. This section's findings provide a solid foundation for moving on to the subsequent stages of construction machine removal, 3D surface generation, and volume calculation, resulting in a comprehensive monitoring tool for construction sites.

5.4 Construction Machines Removal

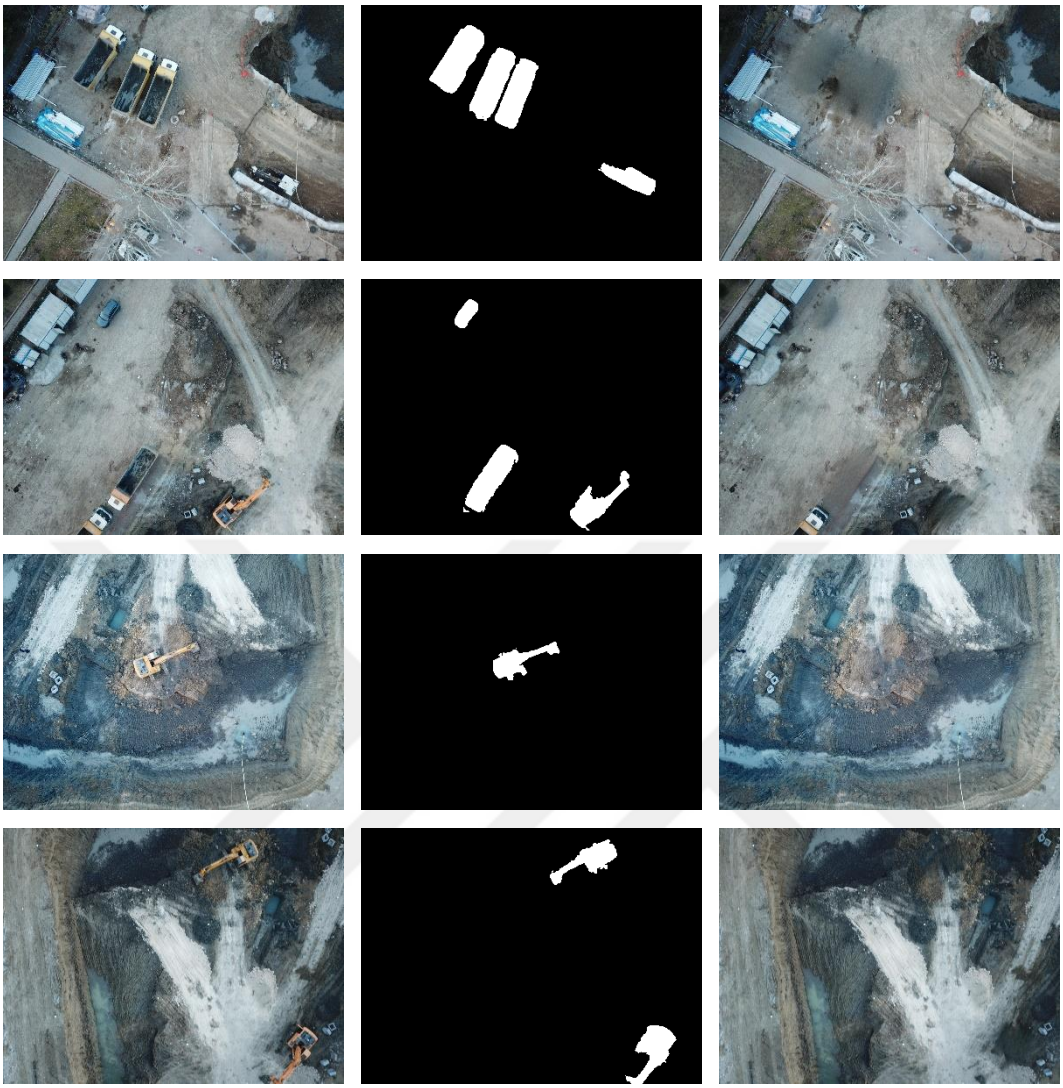
The previous section illustrated the validity of deep learning methods in detecting heavy machinery. Image masks, including object boundaries for each detected machine in the image, were generated after the segmentation. Notably, the method had difficulty identifying smaller parts of the machines when viewed from the top. To accommodate these missing elements and perimeters, image masks were dilated. Image dilation is a morphological operation that thickens objects by adding pixels to the boundaries of objects in the image. It is used for binary images in our study, which are image masks.

After dilation, the LaMa inpainting algorithm trained with Places2 dataset was applied to the images to fill the removed construction machinery pixels. LaMa is an inpainting technique based on deep learning designed for large missing regions in images (Suvorov et al., 2022). It is highly effective at restoring complex structures and generates high-quality inpainted images. More information about LaMa can be found in Chapter 4.3. Figure 5.20 and Figure 5.21 depict the results of the image inpainting procedure for datasets T1 and T2, respectively.



(a) Original images (b) Image masks (c) Inpainted images

Figure 5.20. T1 Inpainting Results



(a) Original images (b) Image masks (c) Inpainted images

Figure 5.21. T2 Inpainted Results

The efficiency of the inpainting was evaluated using a variety of standard image comparison metrics. Normalized Root Mean Square Error (NRMSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) are presented in Table 5.6 and discussed below in detail.

NRMSE is a commonly used measure of the differences between the values predicted by a model and the observed values in the real world (Eskicioglu & Fisher, 1995). It measures the average magnitude of the pixel value difference between two

images during image inpainting. A smaller NRMSE indicates that the prediction is closer to the actual data, thus indicating improved performance.

PSNR is a metric used to measure the quality of reconstruction (Eskicioglu & Fisher, 1995). It is an estimate of the maximum possible power of a signal relative to the power of the noise that corrupts it. In this context, the 'signal' is the original image, whereas the 'noise' is the inpainting error. Typically, PSNR is expressed on a logarithmic decibel scale. A greater PSNR indicates superior image quality and less distortion and noise. SSIM is a method used to compare the similarity of two images.

The SSIM index is a complete reference metric, which means that the measurement or prediction of image quality is based on an initial uncompressed or distortion-free image as the reference (Wang et al., 2004). SSIM is intended to be superior to conventional methods such as PSNR and NRMSE by considering changes in the image's structural information, luminance, and contrast. The SSIM index is a decimal value between -1 and 1, and a value of 1 is only possible when two data sets are identical.

Table 5.6 Inpainting Results

Dataset	**NRMSE**	**PSNR**	**SSIM**
T1	0.019	37.628	0.900
T2	0.018	32.595	0.901

EXIF (Exchangeable Image File Format) information was removed during the image inpainting process. EXIF data consists of information such as the date and time of image creation, GPS coordinates, and camera settings, among other things. EXIF data from the original images were copied back into the inpainted images to preserve this information for future use in the 3D reconstruction.

In conclusion, by employing image inpainting techniques, we could isolate and remove construction machines, enhancing the clarity and precision of our data. This critical preprocessing step prepares the images of terrain data for subsequent analysis, ensuring progress tracking and volume calculations accurately represent the construction site's actual state, free of temporary obstructions like construction machinery.

5.5 3D Surface Reconstruction

After construction machines were removed from the images, the process continued with the transformation of 2D images into a 3D point cloud. To reconstruct the site in 3D, the Structure from Motion (SfM) - Multi-View Stereo (MVS) technique was required. Details of these techniques are explained in Chapter 4.4. When images are collected using UAVs, the process is sometimes referred to as UAV photogrammetry, a technique used to generate 3D models. From the numerous open-source and commercial software options for managing UAV photogrammetry, we utilized SiteEye for this study. SiteEye, a comprehensive photogrammetry software developed at our university Technopolis, processes 2D and 3D data from construction job sites (SiteEye, 2023). T1 and T2 were uploaded to the SiteEye platform and analyzed in the cloud. All ground control points described in Chapter 5.2.2 on data acquisition were marked and utilized on the images via SiteEye GCP marking tool.

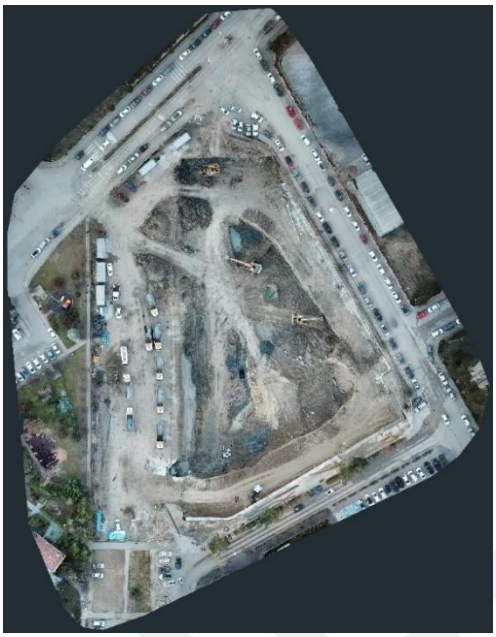
Table 5.7 displays the results of the photogrammetric process. All 230 images from each image set are being processed successfully. Creating sparse point clouds resulted in 314,299 points for T1 and 313,065 for T2. Densified point clouds grew to 32,491,933 and 32,012,565 points for T1 and T2, respectively. For both datasets, the ground sampling distance (GSD), the distance between two consecutive pixel centers measured on the ground, was calculated to be 1.3 cm. A smaller GSD indicates a higher resolution, as each pixel represents a smaller ground area.

It was determined that the GCP error, which indicates the accuracy of photogrammetric processes, was 16.7 cm for T1 and 14.4 cm for T2. This value represents the difference between the known coordinates of a point and its photogrammetrically calculated coordinates. Photogrammetric processes were conducted on original and inpainted images to demonstrate the difference. Figure 5.22 and Figure 5.23 illustrate the orthophotos, i.e., high-resolution maps and point clouds resulting from the photogrammetry procedure, respectively.

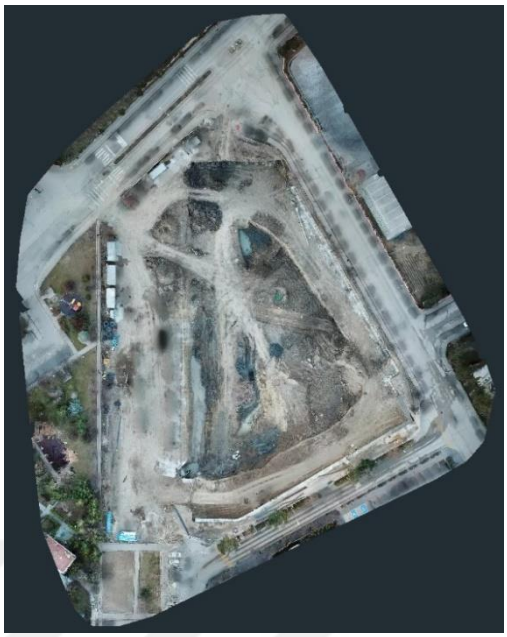
Table 5.7 3D Reconstruction Results

	T1	**T2**
Processed Images	230 of 230	230 of 230
Sparse Point Cloud	314 299 points	313 065 points
Dense Point Cloud	32 491 933 points	32 012 587 points
GSD	1.3 cm	1.3 cm
GCP Error	16.7 cm	14.4 cm

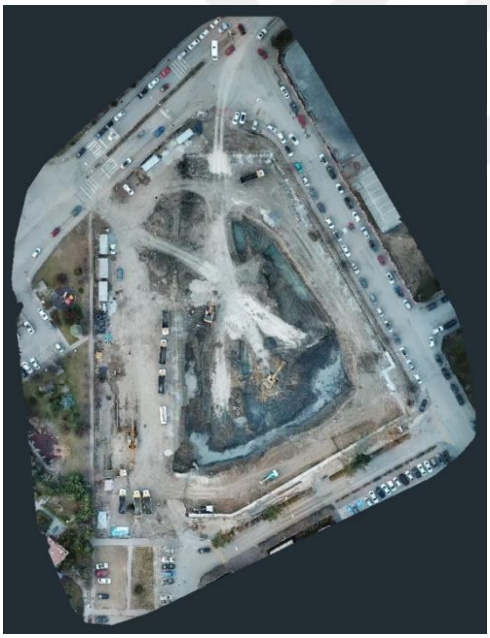
In addition to UAV photogrammetry, TLS was used to generate point clouds of the construction site. Point clouds displayed in Figure 5.24 were created by the TLS method. An imaging step was not performed since these point clouds were generated directly from the laser beams. Therefore, construction machines were manually removed from the point cloud by selecting and fencing the objects in 3D. Figure 5.25 illustrates the original and filtered point clouds after machinery removal.



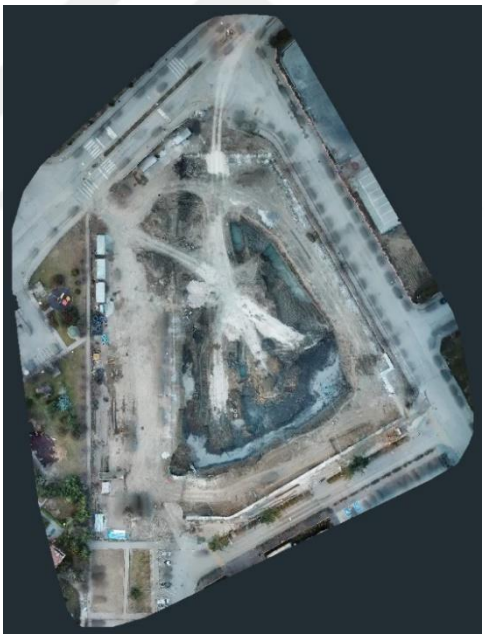
(a) T1 with Original Images



(b) T1 with Inpainted Images

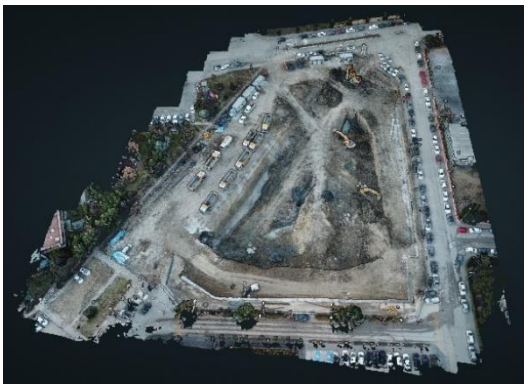


(c) T2 with Original Images



(d) T2 with Inpainted Images

Figure 5.22 Resultant Orthophotos



(a) T1 with Original Images



(b) T1 with Inpainted Images



(c) T2 with Original Images



(d) T2 with Inpainted Images

Figure 5.23 Resultant Point Clouds



(a) T1 (b) T2

Figure 5.24 Laser Scanner Point Clouds



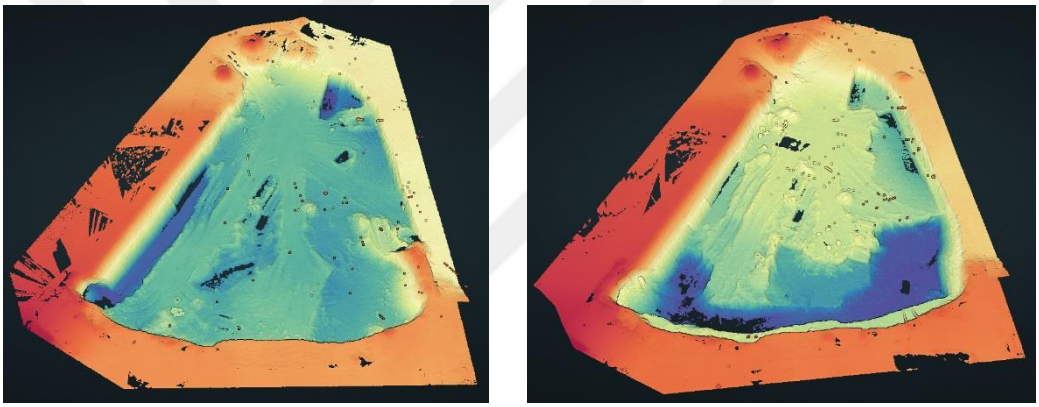
(a) Original Point Cloud (b) Filtered Point Cloud

Figure 5.25 Manual Construction Machine Removal

As the laser scanning procedure was performed in the evening, no color data could be collected. However, the excavation site's geometry was successfully obtained. Figure 5.26 also includes the colored point cloud according to Z coordinates, providing a comprehensive site visualization.



(a) T1 Point Cloud (b) T2 Point Cloud

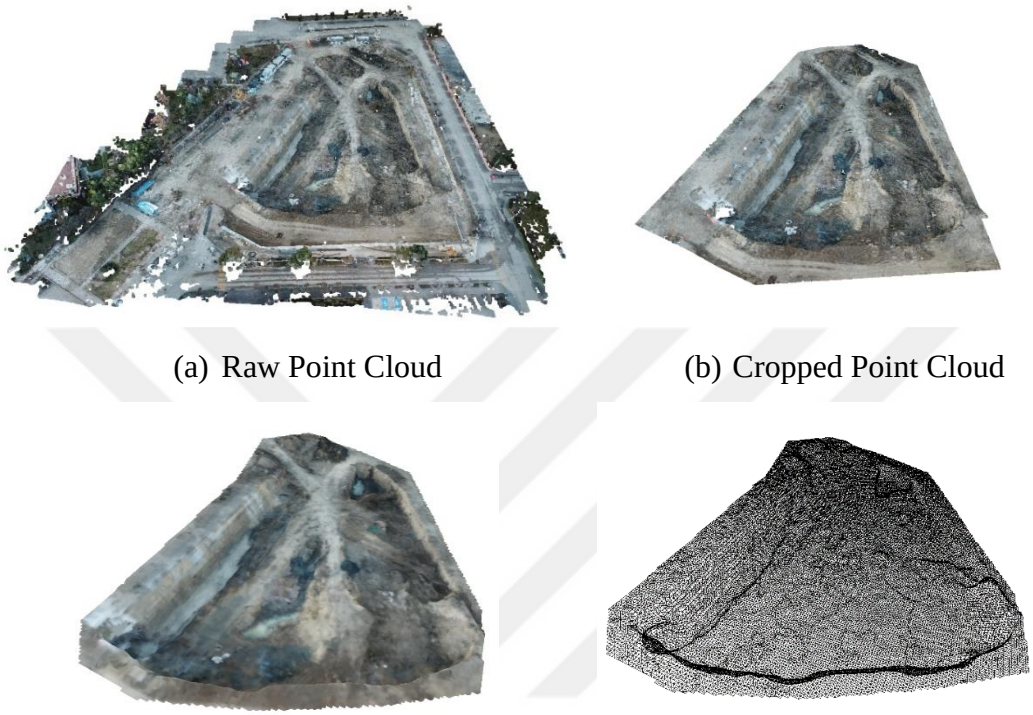


(c) T1 Colored Point Cloud (d) T2 Colored Point Cloud

Figure 5.26 Laser Scanner Point Clouds Visualization

The initial step involved the extraction of the Region of Interest (ROI) from the raw point cloud to filter out unnecessary data and improve the accuracy of calculations. A polygon defined the construction site, and the point cloud was cropped accordingly (Figure 5.27b). Subsequently, voxel downsampling was utilized to reduce the point cloud's density. Even though the resultant point cloud was less dense, it accurately represented the original 3D structure. The computation and orientation of normals, essential for defining the geometry and surface structure of the 3D data, were

performed after the downsampling process. Finally, the surface was reconstructed using the Poisson surface reconstruction method (Figure 5.27c).



(a) Raw Point Cloud (b) Cropped Point Cloud

(c) 3D Surface Generation (d) Mesh Wireframe View

Figure 5.27 3D Reconstruction Steps

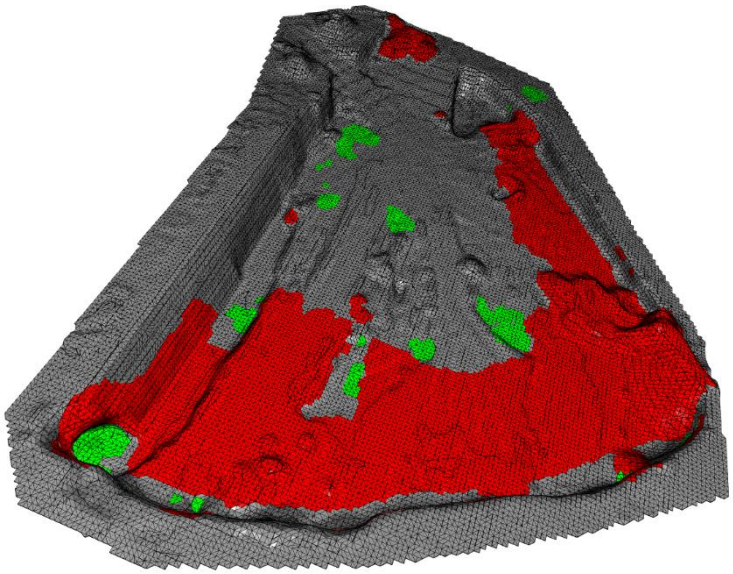
Having completed the generation of point clouds using UAV photogrammetry and TLS, we can now proceed to another essential aspect of this study: volume calculation. This component is crucial as it enables us to evaluate the differences between two successive stages of earthwork progress on construction sites.

5.6 Volume Calculation

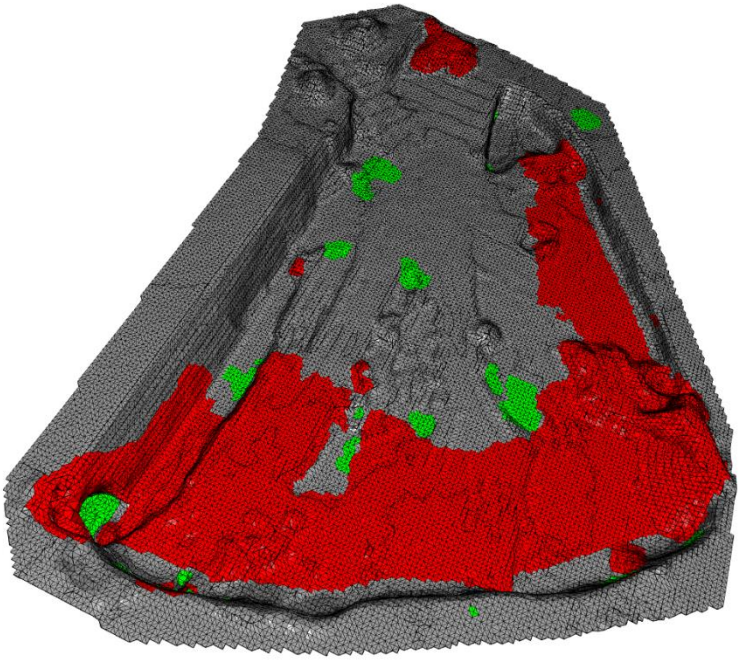
The volume calculation is vital in evaluating the differences between two successive stages of earthwork progress on construction sites. This methodology was applied to

both UAV photogrammetry and TLS generated point clouds to calculate volume changes.

After surfaces were generated, the proposed volume calculation algorithm has performed between subsequent surfaces T1 and T2. The surface reconstruction and volume calculation were performed for UAV photogrammetry and TLS resulting point clouds. For the inpainted point clouds method, which involved the use of UAV photogrammetry, the calculated volume of cut was 4432.82 cubic meters, and the volume of fill was 143.47 cubic meters. On the other hand, the terrestrial laser scanning-based point clouds method yielded a volume of cut at 4682.92 cubic meters and a volume of fill at 122.38 cubic meters. Figure 5.28 depicts the cut and fill volumes visually. The volume that is cut is represented by the color red, while the volume is filled by the color green.



(a) UAV Photogrammetry Method



(b) TLS Method

Figure 5.28 Volume Distribution Visualization

Truck counts from the construction site were also recorded to compare them with the proposed volumetric calculations. Every truck entering and exiting the construction site was meticulously recorded, complete with timestamps and license plates. Table 5.8 details the count of each truck departed from the excavation site. Figure 5.29 depicts one of the identical model trucks that comprise the fleet. Although truck bucket sizes are identical, the filling ratio may vary depending on the excavator operator. According to conversations with construction site specialists, the volume of soil loaded onto this type of truck varied between 15 and 17 m^3. This indicates that the total volume of soil removed from the construction site fell between 4,170 and 4,726 cubic meters. Three volumetric results are considered for comparison between T1 and T2 datasets: UAV photogrammetry, TLS-based point clouds, and truck count-based results tabulated in Table 5.9.

Table 5.8 Truck Count

Day	Truck Counts
6.1.2023	36
7.1.2023	29
8.1.2023	79
9.1.2023	54
10.1.2023	80
Total	278



Figure 5.29 Typical Truck Model from the Construction Site

Table 5.9 Comparison of Methods

	Cut	**Fill**
Inpainted Point Clouds	4432.82 m^3	143.47 m^3
Laser scanner Point Clouds	4682.92 m^3	122.38 m^3
Truck Count based	4,170 m^3 - 4,726 m^3	-

All the methodologies described in the previous sections were systematically applied to the other datasets collected from subsequent site visits (T3 to T5) with UAV photogrammetry with inpainted images. The cumulative volumes of cut and fill were determined through analysis. The total volume of material excavated (cut) from the site was found to be 22,386.29 cubic meters, while the total volume of material used to fill in excavated areas (fill) was recorded at 1,865.14 cubic meters across all visits. Detailed results of these calculations, including individual cut and fill volumes for each period, are summarized in Table 5.10.

Table 5.10 Earthwork Progress using UAV-Based Approach

	Cut	**Fill**
T1 & T2	4432.82	143.47
T2 & T3	7828.13	974.28
T3 & T4	7148.23	399.52
T4 & T5	2943.75	313.15
Total	22386.29	1865.14

5.7 Results and Discussion

The current research presents an innovative approach for monitoring the progress of automated earthwork on construction sites. By combining UAV imaging with

machine learning and computer vision algorithms, a solution is proposed that can mitigate a number of drawbacks associated with conventional earthwork volume calculation methods. Within the context of this study, this section will discuss the significant contributions, encountered challenges, and potential future works.

The research has made several significant contributions to the understanding and application of automated earthwork monitoring. This study provides a fully automated pipeline to calculate earthworks for the benefit of UAVs, perhaps its most significant contribution. It is possible to conduct UAV imaging during construction work hours, with the data collection process unaffected by the movement of construction machines. This is a significant advantage compared to conventional ground-based methods, which are frequently constrained by safety concerns and increased operational complexities during working hours. In addition, ground-based techniques such as laser scanning are time-consuming, requiring lengthy setup, scanning, registration, and manual point cloud editing to remove unwanted objects. Consequently, the UAV-based method proposed in this study provides a significantly more effective and adaptable alternative for earthwork monitoring.

The high performance of the construction machine segmentation model was a significant achievement of the proposed approach. This deep learning model demonstrated impressive precision, particularly for categories of construction vehicles that are commonly encountered, such as dump trucks, excavators, backhoe loaders, and cars. The Average Precision (AP) scores for these categories were consistently high, frequently exceeding the 90% threshold. This demonstrates the strength and viability of using deep learning algorithms for machinery segmentation in construction site images. This implementation is a promising step forward for using machine learning technologies to monitor construction sites.

This research achieved a significant milestone in the automation and efficiency of the construction machine removal process by applying image inpainting. The evaluation metrics NRMSE, PSNR, and SSIM produced encouraging results. For dataset T1, the NRMSE was a modest 0.019, while the PSNR, a measure of

reconstruction quality, was 37.628. The SSIM, which measures structural similarity, produced an excellent score of 0.900. For the second dataset, T2, the NRMSE decreased to 0.018, the PSNR was measured at 32,595, and the SSIM value was 0.901, nearly replicating the results for T1. These measurable outcomes validate the successful application of the image inpainting algorithm for construction machine removal, which helps to generate digital terrain models accurately.

Further reinforcing the proposed method's utility, the construction machine-removed images that were georeferenced and created independently could be processed with any photogrammetry software. The output of the proposed construction machine removal method is an image set that can serve as input for any third-party application. This interoperability highlights the adaptability of the proposed method and its potential for integration with a wide range of existing workflows or systems. This capability can increase the adoption and applicability of this research in automated earthwork monitoring for construction sites.

The Ground Sampling Distance (GSD), which describes the distance on the ground that each pixel in an image represents, was determined to be 1.3 cm for both datasets (T1 and T2). This demonstrates that the images captured sufficient detail for earthwork calculation purposes. By achieving such a small GSD, the proposed system ensures that the construction site's elaborate features are accurately represented, essential for precise earthwork volume calculation.

When it comes to the validation of earthwork volume, it is very challenging to compare it to the actual volume calculation, as each method has its flaws and attempts to represent the construction site. This allows us to compare only the most common methods. Our UAV-based method estimated a cut volume of 4432.82 m^3, a fill volume of 143.47 m^3, and a net volume of 4575.47 m^3. The difference between UAV-based and TLS-based methods is 229.83 m^3. The minor difference observed can be attributed to the manual machine removal process utilized in laser scanning, which leaves holes in the point cloud that are later filled flat during surface reconstruction.

On the other hand, the UAV-based approach inpainted and preserved the natural texture and appearance of the topography. Moreover, even if it is an approximation, truck count estimates suggested that the calculated cut volumes fall within a plausible range (4,170 - 4,726 m^3), strengthening the credibility of the volume calculations by UAV photogrammetry and TLS. This suggests the proposed deep learning and image inpainting method can provide comparably precise volume estimates, confirming its robust earthwork progress monitoring potential.

Despite these positive results, the research also encountered some challenges. Rainy and snowy conditions could potentially affect the functionality of most UAVs. Despite this, the 15-minute duration of aerial imaging allows UAV flights to be conducted during dry intervals during rainy periods. When a UAV is readily accessible at a construction site, this adaptability becomes particularly feasible. During this research, however, our base of operations was located a considerable distance from the construction site, posing particular logistical challenges. Due to the distance, it was impractical for us to take advantage of these brief weather windows on rainy days, so we could not visit the site under these conditions. This limitation emphasizes the need for a UAV on-site for more flexible imaging scheduling, particularly on construction sites in geographically challenging or remote areas.

In addition, the "Other" category, which included drilling machines, anchor installation machines, and concrete mixers, received significantly lower AP scores, 45.9 for T1 and 63.1 for T2. This suggests that the model has difficulty accurately identifying less common or more diverse types of machinery, as mentioned in Chapter 3.6. Figure 5.19 illustrates that certain non-machinery items, such as child parks, containers, and concrete blocks, are also classified under this category. Including such diverse elements further complicates the model's ability to accurately identify and classify items within this "Other" category.

In the context of image inpainting, it has been observed that proximity between construction machines presents a unique challenge. Large image masks resulting

from the inpainting procedure have the potential to produce artificial-looking regions. Due to the difficulty of recreating fine details and fine details over such expansive regions, the inpainting process may not result in seamless integration with the surrounding environment in these circumstances.

Developing a dedicated image inpainting dataset could be a promising direction for future research. A specific dataset that focuses on scenarios frequently encountered during the automated monitoring of construction sites could aid in training more specialized models. This could improve the performance of the image inpainting algorithm, which is a crucial component of the proposed method for earthwork progress monitoring. These future directions highlight the potential for additional improvements to automated earthwork progress monitoring.

In conclusion, while there are some areas for improvement and further research, the contributions of this study are significant and pave the way for future developments in automated earthwork progress monitoring for construction sites. The study provides a solid foundation for refining and expanding the proposed method, which can solve one of the main problems in geotechnics. It demonstrates the enormous potential of combining UAV imaging with advanced machine learning and computer vision algorithms in construction management.

CHAPTER 6

CONCLUSIONS

6.1 Summary

The primary objective of this thesis was to create an automated earthwork monitoring system for the construction industry. Utilizing UAV-based photogrammetry, machine learning for construction machinery segmentation, and image inpainting, was expected to solve the problem of the current lack of precise monitoring. A comprehensive literature review highlighted the methods of machine learning, image inpainting, earthwork progress monitoring, and volume calculation used in construction. This analysis resulted in the originality of this research and emphasized the need for such an integrated solution.

The research began with developing a dataset of aerial images and a machine-learning model for construction machinery. Various UAVs, including DJI's Phantom 4 RTK, Mavic Pro, Mavic 2 Pro, and Yuneec's H520, were used to collect these images in 25 regions of Türkiye. The utilization of various UAV models and camera angles contributed to the level of detail of the collected data. The images depicted a variety of settings, including excavation sites, steel and reinforced concrete structures, and transportation projects. Images devoid of construction machines were also collected to improve the algorithm's ability to distinguish between relevant and irrelevant patterns. In addition, to maintain privacy, the dataset was meticulously processed to blur personally identifiable information such as faces, company names, logos, and slogans.

The careful process of pixel-level annotations required 82.38 hours of labor. The dataset focused on eight common construction vehicle types: dump trucks, excavators, backhoe loaders, wheel loaders, compactors, dozers, graders, and cars.

Other, less common types of machinery, such as asphalt pavers and drilling machines, were classified as "other." In total, 9563 instances were labeled in 2155 images, and a clustering strategy based on image similarity and proximity was implemented. Following this, the study evaluated the performance of five cutting-edge deep learning models, namely Hybrid Task Cascade, Cascade Mask R-CNN, Mask Scoring R-CNN, Pointrend, and Mask R-CNN, for object detection and segmentation tasks. A significant milestone in the research was training the most suitable model, which followed a rigorous evaluation process. Using the trained machine learning model, this study developed an earthwork progress monitoring system that has the potential to revolutionize construction management. This system primarily aims to segment and remove construction machinery from images, producing an accurate digital terrain representation.

The application of the machine learning model is integral to this procedure. The model examines the gathered images and performs the required segmentation. In the procedure, it identifies and isolates machinery efficiently and effectively. Image inpainting is an additional essential component of the methodology. The selected image inpainting algorithm seamlessly filled in the empty portions of images left by removing construction machinery. This algorithm ensures that the digital terrain representation remains intact and accurate following the removal of machinery.

In this context, understanding photogrammetry operations is essential for digital terrain generation. The study presented the SfM-MVS pipeline, a fundamental component of photogrammetry. This pipeline consists of stages, including feature extraction and matching, sparse and dense point cloud construction, and surface reconstruction. Collectively, these steps transform the raw images into a three-dimensional representation of the construction site, enabling efficient monitoring of earthwork progress. In addition to these processes, a novel algorithm for volume calculation was implemented as part of the methodology. This algorithm calculates cut and fill volumes, which is essential for comprehending earthwork progress. The algorithm calls for generating two surface meshes representing the terrain before and after earthworks. The differences between these meshes are computed using ray-

casting to determine cut and fill earthwork volumes. Measuring an artificial object's volume demonstrated this method's precision; the calculated results deviated minimally from the actual volume, confirming the method's accuracy.

Finally, field experiments were conducted in a real-world setting to evaluate the practical applicability of the proposed system. In an excavation site in Ankara, Turkiye, the application of the created earthwork progress monitoring system was tested. The research team used UAV imaging and terrestrial laser scanning to compare the proposed monitoring procedure. Throughout five site visits, the team gathered enormous data from the excavation area. The proposed method was applied to UAV images, in which the machine learning model segmented the construction machines, and the image inpainting algorithm filled the resulting gaps. This procedure resulted in images devoid of construction machines and accurately depicting the site's natural condition. We used the processed image data to reconstruct the 3D surfaces of the excavation site. These reconstructions provided a tangible, comprehensive representation of the terrain at the site. Simultaneously, a laser scanning method was employed as a comparative measure for the first two site visits, with the construction machinery manually removed from the laser-generated point clouds.

The proposed UAV-based method and the conventional laser scanning-based method were compared. Calculated volumes for the UAV-based method were 4432.82 cubic meters for cut and 143.47 cubic meters for fill. The cut volume for the laser scanning-based method was 4682.92 cubic meters, and the fill volume was 122.38 cubic meters. The team also recorded the number of truckloads that departed the construction site as a comparison measure. Since each truck carried an average of 15 to 17 cubic meters of soil, the total volume of soil removed from the site was estimated to be between 4,170 and 4,726 cubic meters. These real-world data enhanced the comparison of the proposed system's volume calculations.

Processing all site visits, the UAV photogrammetry with inpainted images revealed that the total volume of excavated material was 22,386.29 cubic meters, while the

fill volume was 1,865.14 cubic meters. This volumetric data and the accurate digital representations of the site demonstrated the effectiveness of the proposed monitoring system for tracking earthwork progress.

In conclusion, the results were evaluated critically, with the volume calculations demonstrating the successful integration of machine learning, photogrammetry, and image inpainting in construction management. The proposed system's remarkable accuracy provides compelling proof of concept for the broader application of such technology.

6.2 Contributions

This research presents a fully autonomous pipeline for earthwork progress monitoring in construction sites using Unmanned Aerial Vehicles (UAV) photogrammetry, machine learning, and computer vision algorithms. By automating the complete chain of operations, from image acquisition to volume calculation, the potential for manual errors is reduced, and the process's speed and accuracy are enhanced. This significant development paves the way for a more efficient and reliable approach to earthwork monitoring, highlighting an important progression in construction management practices. The main contributions are as follows.

- ***Enhancing Safety and Efficiency Through UAV Imaging****:* The adoption of UAVs for imaging during operational hours has notably improved the safety and efficiency of construction sites. By removing the need for manual inspections, risks associated with human intervention are significantly minimized. Additionally, as UAVs can operate independently of ground-based activities, the continuity of construction operations is not disrupted, further enhancing overall productivity. Integrating UAVs into construction site management presents a clear advancement in ensuring safety, minimizing disruption, and improving operational efficiency.

- ***Development of Aerial Image Dataset and Machine Learning Model:*** This thesis has contributed by assembling an annotated aerial image dataset, capturing a range of real-world construction environments. This diverse collection includes images from excavation sites, steel and concrete structures, and transportation projects. With their detailed pixel-level annotations, this assortment of real-world images plays a crucial role in training machine learning models, enabling them to learn better and understand the complexities of different construction environments. This comprehensive dataset expands the opportunities for research in automated construction progress monitoring and sets a foundation for further developments in this sector. An integral component of this study is developing a machine-learning model to detect construction machinery. The model has shown considerable effectiveness in identifying standard construction machinery accurately, reflecting its robustness and precision.

- ***Image Similarity and Proximity-based Clustering for Data Partitioning:*** The research has also introduced a novel approach for data partitioning in machine learning model training and testing by employing an image similarity and proximity-based clustering strategy. This technique aims to group images with similar characteristics, enhancing the model's prediction accuracy and helping prevent overfitting. The use of this innovative data split method contributes to improving the model's reliability by ensuring a diverse and well-distributed dataset for both training and testing.

- ***Image Inpainting for Machine Removal and Terrain Generation:*** Integrating an image inpainting technique for effective machine removal is another significant achievement of this study. This method enhanced the fidelity of visual representations of construction terrains by excising machinery from the images and reconstructing the affected areas. The

efficacy of this technique played a crucial role in enhancing the overall accuracy of the terrain modeling process.

- ***Software-Independent Photogrammetry Procedure:*** Developing a software-independent photogrammetry procedure marks a significant leap in this research domain. It ensures the versatility and wide accessibility of the proposed methodology, regardless of the software environment. This software-agnostic approach, allowing the use of different photogrammetry software, has increased the flexibility and adaptability of the approach, thereby promoting its practical applicability in various construction monitoring scenarios.

- ***Novel Volume Calculation Algorithm:*** The proposed novel volume calculation algorithm is another crucial contribution of this study. It provides a new method to estimate earthwork volumes accurately and efficiently, thus enhancing the overall construction progress monitoring process. The algorithm has been designed to work with the developed machine learning model and image inpainting technique, collectively providing a more accurate and comprehensive terrain representation. This has paved the way for more precise volumetric analysis and has potential implications for cost and time efficiency in construction projects.

- ***Comparative Analysis of UAV and TLS-based Point Clouds:*** Lastly, the comparison between UAVs and Terrestrial Laser Scanning (TLS) based point clouds provides valuable insights into the relative efficacy of these methods. The findings suggest that both UAV and TLS-generated point clouds fall within a reasonable interval of difference, demonstrating the viability of using UAVs for such applications. This comparison contributes to our understanding of the strengths and limitations of these technologies in

construction site monitoring and opens the door for their practical combined usage in the future.

6.3 Limitations and Future Work

While the contributions of this research mark significant advancements in earthwork monitoring and construction management, they also present new challenges that provide avenues for further exploration and refinement. These challenges underscore the complexities inherent in applying machine learning and imaging technologies in the dynamic and diverse world of construction sites. The following paragraphs introduce these limitations:

- ***Weather Challenges in UAV-Based Construction Site Monitoring:*** Although adopting UAVs brings about significant benefits in terms of safety and efficiency, it also introduces particular challenges, particularly their susceptibility to weather conditions. Inclement weather can affect UAVs' flight performance and imaging capabilities and, in severe cases, can ground operations entirely. Therefore, future work should investigate the development of robust systems capable of withstanding diverse weather conditions or incorporating weather prediction models into the planning and scheduling of UAV-based data collection in construction site monitoring.

- ***Difficulty in Segmentation of Diverse Construction Machinery:*** Identifying the “other” class of construction machinery posed a significant challenge in this study, owing to the wide range of categories, the limited number of representative images, and the diverse shapes and characteristics of these machines. While the machine learning model demonstrated high performance in detecting regular construction machinery, the segmentation of “other” machinery types requires further enhancement. Addressing this issue may necessitate expanding the training dataset to include more varied

and representative images of diverse construction machinery types, which could boost the model's overall segmentation performance.

- ***Challenges in Detecting and Inpainting Closely Positioned Machines:*** Closely positioned machines on construction sites present another complication. For the segmentation process, these machines are often recognized as one single entity, which impacts the accuracy of machinery segmentation and identification. This issue is amplified during the image inpainting stage, where the large masks created by these closely positioned machines pose a challenge in filling them effectively. This phenomenon could potentially distort the terrain model and influence the volume estimation accuracy. Future research could focus on refining the segmentation algorithm to better distinguish individual machines even when they are closely located and enhancing the inpainting algorithm to handle large mask areas more effectively.

In addition to the challenges encountered in the research, several promising possibilities for future work can further enhance and expand the capabilities of the proposed methodology:

- ***Developing a Specialized Image Inpainting Dataset for Enhanced Construction Site Terrain Modeling:*** A future endeavor could involve the creation of a dedicated image inpainting dataset specifically for construction sites. This dataset could provide more targeted and relevant training data for the inpainting algorithm, helping it better understand and reproduce the unique characteristics of construction site terrains. The effort would, ideally, improve the construction machinery removal algorithm's performance and result in more accurate digital terrain models.

- ***Integrating 3D Soil Profiling and Cost Estimation into the Earthwork Progress Monitoring System:*** Future research could enhance the proposed system's capacity to generate a 3D soil profile from borehole logs. This would necessitate incorporating ground and subsurface layers into the model, thereby enhancing information on the soil composition and type at the excavation site. This additional functionality would provide valuable insights into the material being excavated, particularly in the case of deep excavation. Furthermore, it would also aid in calculating the tonnage and associated costs associated with soil removal. Consequently, this integration could increase the method's usefulness and contribution to effective and efficient project management in the construction industry.

- ***4D BIM Integration:*** Lastly, integrating the developed methodology with 4D Building Information Modeling (BIM) for monitoring excavation progress presents an inspiring future research direction. 4D BIM integrates geometric models with time or schedule-related information, enabling a comprehensive view of project progress. Combining the automated UAV-based earthwork monitoring system with 4D BIM could facilitate accurate excavation work tracking, providing valuable insights to project managers and enabling more informed decision-making. Such integration, however, would require substantial work in data interoperability, system integration, and possibly the development of new algorithms. Nonetheless, the potential benefits of this integration make it a worthy endeavor for future research.

REFERENCES

Ajayi, O. G., & Ajulo, J. (2021). Investigating the applicability of unmanned aerial vehicles (UAV) Photogrammetry for the estimation of the volume of stockpiles. *Quaestiones Geographicae*, *40*(1), 25–38. https://doi.org/10.2478/quageo-2021-0002

Akgul, M., Yurtseven, H., Gulci, S., & Akay, A. E. (2018). Evaluation of UAV- and GNSS-based dems for earthwork volume. *Arabian Journal for Science and Engineering*, *43*(4), 1893–1909. https://doi.org/10.1007/s13369-017-2811-9

Alex Krizhevsky. (2009). *Learning Multiple Layers of Features from Tiny Images*.

Angah, O., & Chen, A. Y. (2020). Removal of occluding construction workers in job site image data using U-Net based context encoders. *Automation in Construction*, *119*(April 2019), 103332. https://doi.org/10.1016/j.autcon.2020.103332

Ballester, C., Caselles, V., & Verdera, J. (2003). Disocclusion by joint interpolation of vector fields and gray levels. *Multiscale Modeling and Simulation*, *2*(1), 80–123. https://doi.org/10.1137/S1540345903422458

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*(September 2019), 103198. https://doi.org/10.1016/j.autcon.2020.103198

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. https://doi.org/10.1111/mice.12672

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, *110*(3), 346–359. https://doi.org/10.1016/j.cviu.2007.09.014

Berg, A., Deng, J., & Fei-Fei, L. (2010). *Large Scale Visual Recognition Challenge (ILSVRC)*.

Bertalmio, M., Sapiro, G., Caselles, V., & Ballester, C. (2000). Image inpainting. *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, 417–424. https://doi.org/10.1145/344779.344972

Bertalmio, M., Vese, L., Sapiro, G., & Osher, S. (2003). Simultaneous structure and texture image inpainting. *IEEE Transactions on Image Processing*, *12*(8), 882–889. https://doi.org/10.1109/TIP.2003.815261

Bianco, S., Ciocca, G., & Marelli, D. (2018). Evaluating the performance of structure from motion pipelines. *Journal of Imaging*, *4*(8), 1–18. https://doi.org/10.3390/jimaging4080098

Buatik, A., & Chaiyasarn, K. (2021). Deformable object removal via exemplar-based algorithm for 3D model of heritage structures. *International Journal of GEOMATE*, *21*(84), 151–158. https://doi.org/10.21660/2021.84.j2076

Carranza-García, M., Torres-Mateo, J., Lara-Benítez, P., & García-Gutiérrez, J. (2021). On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. *Remote Sensing*, *13*(89), 1–23. https://doi.org/10.3390/rs13010089

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. https://doi.org/10.1016/j.autcon.2011.02.003

Chen, C., Zhu, Z., & Hammad, A. (2020). Automated excavators activity recognition and productivity analysis from construction site surveillance videos. *Automation in Construction*, *110*(April 2019), 103045. https://doi.org/10.1016/j.autcon.2019.103045

Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Shi, J., Ouyang, W., Loy, C. C., & Lin, D. (2019). Hybrid task cascade for instance

segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 4969–4978. https://doi.org/10.1109/CVPR.2019.00511

Cheng, J., & Jiang, L. (2013). Accuracy comparison of roadway earthwork computation between 3D and 2D methods. *Procedia - Social and Behavioral Sciences*, *96*, 1277–1285. https://doi.org/10.1016/j.sbspro.2013.08.145

Chi, S., & Caldas, C. H. (2011). Automated Object Identification Using Optical Video Cameras on Construction Sites. *Computer-Aided Civil and Infrastructure Engineering*, *26*(5), 368–380. https://doi.org/10.1111/j.1467-8667.2010.00690.x

Cho, J. W., Lee, J.-K., & Park, J. (2021). Large-scale earthwork progress digitalization practices using series of 3D models generated from UAS images. *Drones*, *5*(4). https://doi.org/10.3390/drones5040147

Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *92*, 79–97. https://doi.org/10.1016/j.isprsjprs.2014.02.013

Contreras, M., Aracena, P., & Chung, W. (2012). Improving accuracy in earthwork volume estimation for proposed forest roads using a high-resolution digital elevation model. *Croatian Journal of Forest Engineering*, *33*(1), 125–142.

Criminisi, A., Pérez, P., & Toyama, K. (2003). Object removal by exemplar-based inpainting. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2*. https://doi.org/10.1109/cvpr.2003.1211538

Cryderman, C., Mah, S. B., & Shufletoski, A. (2014). Evaluation of uav photogrammetric accuracy for mapping and earthworks computations. *Geomatica*, *68*(4), 309–317. https://doi.org/10.5623/cig2014-405

CVAT. (2022). *Powerful and efficient Computer Vision Annotation Tool (CVAT)*.

https://github.com/opencv/cvat

Del Savio, A., Luna, A., Cárdenas-Salas, D., Vergara, M., & Urday, G. (2022). Dataset of manually classified images obtained from a construction site. *Data in Brief*, *42*, 108042. https://doi.org/10.1016/j.dib.2022.108042

DJI. (2023). *DJI Mavic Pro Specs*. https://www.dji.com/mavic

Dronedeploy. (2023). *Dronedeploy: Reality Capture Platform*. www.dronedeploy.com

Du, J.-C., & Teng, H.-C. (2007). 3D laser scanning and GPS technology for landslide earthwork volume estimation. *Automation in Construction*, *16*(5), 657–663. https://doi.org/10.1016/j.autcon.2006.11.002

Duan, R., Deng, H., Tian, M., Deng, Y., & Lin, J. (2022). SODA: A large-scale open site object detection dataset for deep learning in construction. *Automation in Construction*, *142*(July), 104499. https://doi.org/10.1016/j.autcon.2022.104499

Engels, C., Tingdahl, D., Vercruysse, M., Tuytelaars, T., Sahli, H., & Van Gool, L. (2011). Automatic occlusion removal from facades for 3D urban reconstruction. *Advanced Concepts for Intelligent Vision Systems: 13th International Conference, ACIVS 2011, Ghent, Belgium, August 22-25*, *6915 LNCS*, 681–692. https://doi.org/10.1007/978-3-642-23687-7_61

Eskicioglu, A. M., & Fisher, P. S. (1995). Image quality measures and their performance. *IEEE Transactions on Communications*, *43*(12), 2959–2965. https://doi.org/10.1109/26.477498

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal visual object classes (VOC) Challenge. *International Journal of Computer Vision*, *88*(2), 303–338.

Fan, L., Tan, Y., & Chen, L. (2016). 3D map repair after moving objects removing.

2016 International Conference on Advanced Robotics and Mechatronics (ICARM), *1*, 127–132. https://doi.org/10.1109/ICARM.2016.7606907

Fang, W., Ding, L., Zhong, B., Love, P. E. D., & Luo, H. (2018). Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach. *Advanced Engineering Informatics*, *37*(November 2017), 139–149. https://doi.org/10.1016/j.aei.2018.05.003

Fei-Fei, L., Fergus, R., & Perona, P. (2004). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2004-Janua*(January). https://doi.org/10.1109/CVPR.2004.383

Furukawa, Y., & Ponce, J. (2010). Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(8), 1362–1376. https://doi.org/10.1109/TPAMI.2009.161

Golparvar-Fard, M., Heydarian, A., & Niebles, J. C. (2013). Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers. *Advanced Engineering Informatics*, *27*(4), 652–663. https://doi.org/10.1016/j.aei.2013.09.001

Gong, J., & Caldas, C. H. (2011). An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations. *Automation in Construction*, *20*(8), 1211–1226. https://doi.org/10.1016/j.autcon.2011.05.005

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. *Advances in Neural Information Processing Systems*. https://doi.org/10.1145/3422622

Griffin, G., Holub, A., & Perona, P. (2007). *Caltech-256 object category dataset*. http://www.vision.caltech.edu/datasets/

Guo, Y., Xu, Y., & Li, S. (2020). Dense construction vehicle detection based on orientation-Aware feature fusion convolutional neural network. *Automation in Construction*, *112*(February), 103124. https://doi.org/10.1016/j.autcon.2020.103124

Gupta, A., Dollar, P., & Girshick, R. (2019). LVIS: A dataset for large vocabulary instance segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 5351–5359. https://doi.org/10.1109/CVPR.2019.00550

Hare, W. L., Koch, V. R., & Lucet, Y. (2011). Models and algorithms to improve earthwork operations in road design using mixed integer linear programming. *European Journal of Operational Research*, *215*(2), 470–480. https://doi.org/10.1016/j.ejor.2011.06.011

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2018.2844175

Hintz, C., & Vonderohe, A. P. (2011). Comparison of earthwork computation methods. In *Transportation Research Record* (Issue 2215, pp. 100–104). https://doi.org/10.3141/2215-11

Hu, C., Zhou, Y., Zhao, C.-J., & Pan, Z.-G. (2015). Slope excavation quality assessment and excavated volume calculation in hydraulic projects based on laser scanning technology. *Water Science and Engineering*, *8*(2), 164–173. https://doi.org/10.1016/j.wse.2015.03.001

Huang, Z., Huang, L., Gong, Y., Huang, C., & Wang, X. (2019). Mask scoring R-CNN. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 6402–6411. https://doi.org/10.1109/CVPR.2019.00657

Hugenholtz, C. H., Walker, J., Brown, O., & Myshak, S. (2015). Earthwork volumetrics with an unmanned aerial vehicle and softcopy photogrammetry. *Journal of Surveying Engineering*, *141*(1), 06014003. https://doi.org/10.1061/(asce)su.1943-5428.0000138

Hwang, J., Kim, J., Chi, S., & Seo, J. O. (2022). Development of training image database using web crawling for vision-based site monitoring. *Automation in Construction*, *135*, 104141. https://doi.org/10.1016/j.autcon.2022.104141

Jakubović, A., & Velagić, J. (2018). Image feature matching and object detection using brute-force matchers. *Proceedings Elmar - International Symposium Electronics in Marine*, *2018-Septe*(September), 83–86. https://doi.org/10.23919/ELMAR.2018.8534641

James, M. R., Robson, S., d'Oleire-Oltmanns, S., & Niethammer, U. (2017). Optimising UAV topographic surveys processed with structure-from-motion: Ground control quality, quantity and bundle adjustment. *Geomorphology*, *280*, 51–66. https://doi.org/10.1016/j.geomorph.2016.11.021

Jia Deng, Wei Dong, Socher, R., Li-Jia Li, Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/cvprw.2009.5206848

Julge, K., Ellmann, A., & Köök, R. (2019). Unmanned aerial vehicle surveying for monitoring road construction earthworks. *Baltic Journal of Road and Bridge Engineering*, *14*(1), 1–17. https://doi.org/10.7250/bjrbe.2019-14.430

Kapoor, A., Larco, H., & Kiveris, R. (2019). Nostalgin: Extracting 3D city models from historical image data. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2565–2575. https://doi.org/10.1145/3292500.3330743

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive growing of GANs for improved quality, stability, and variation. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 1–26.

Kazhdan, M., Bolitho, M., & Hoppe, H. (2006). Poisson Surface Reconstruction. *Eurographics Symposium on Geometry Processing*.

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*(December 2018), 168–182. https://doi.org/10.1016/j.autcon.2018.12.014

Kim, H., Bang, S., Jeong, H., Ham, Y., & Kim, H. (2018). Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation. *Automation in Construction*, *92*(April), 188–198. https://doi.org/10.1016/j.autcon.2018.04.002

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. https://doi.org/10.1061/(asce)cp.1943-5487.0000731

Kim, J., & Chi, S. (2019). Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles. *Automation in Construction*, *104*(May), 255–264. https://doi.org/10.1016/j.autcon.2019.03.025

Kim, J., Chi, S., & Seo, J. (2018). Interaction analysis for vision-based activity identification of earthmoving excavators and dump trucks. *Automation in Construction*, *87*(December 2017), 297–308. https://doi.org/10.1016/j.autcon.2017.12.016

Kim, W. J., Hwang, S., Lee, J., Woo, S., & Lee, S. (2022). AIBM: Accurate and instant background modeling for moving object detection. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 9021–9036. https://doi.org/10.1109/TITS.2021.3090092

Kirillov, A., Wu, Y., He, K., & Girshick, R. (2020). Pointrend: Image segmentation as rendering. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 9799–9808. https://doi.org/10.1109/CVPR42600.2020.00982

Krasin, I., Duerig, T., Alldrin, N., Ferrari, V., Abu-El-Haija, S., Kuznetsova, A., Rom, H., Uijlings, J., Popov, S., Veit, A., & others. (2017). *Openimages: A*

public dataset for large-scale multi-label and multi-class image classification.

Lee, D. T., & Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences*, *9*(3), 219–242. https://doi.org/10.1007/BF00977785

Leica. (2023). *Leica BLK360 Imaging Laser Scanner*. https://leica-geosystems.com/products/laser-scanners/scanners/blk360

Levin, A., Zomet, A., & Weiss, Y. (2003). Learning how to inpaint from global image statistics. *Proceedings of the IEEE International Conference on Computer Vision*, *1*(Iccv), 305–312. https://doi.org/10.1109/iccv.2003.1238360

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Lin, Z.-H., Chen, A. Y., & Hsieh, S.-H. (2021). Temporal image analytics for abnormal construction activity identification. *Automation in Construction*, *124*(June 2020), 103572. https://doi.org/10.1016/j.autcon.2021.103572

Liu, G., Reda, F. A., Shih, K. J., Wang, T. C., Tao, A., & Catanzaro, B. (2018). Image inpainting for irregular holes using partial convolutions. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 85–100. https://doi.org/10.1007/978-3-030-01252-6_6

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Lorensen, W. E., & Cline, H. E. (1987). Marching Cubes: a High Resolution 3D

Surface Construction Algorithm. *Computer Graphics (ACM)*, *21*(4), 163–169. https://doi.org/10.1145/37402.37422

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*. https://doi.org/10.1109/ICCV.1999.790410

Luo, X., Li, H., Cao, D., Dai, F., Seo, J., & Lee, S. (2018). Recognizing diverse construction activities in site images via relevance networks of construction-related objects detected by convolutional neural networks. *Journal of Computing in Civil Engineering*, *32*(3), 1–16. https://doi.org/10.1061/(asce)cp.1943-5487.0000756

Mahmood, B., Han, S., & Seo, J. (2022). Implementation experiments on convolutional neural network training using synthetic images for 3D pose estimation of an excavator on real images. *Automation in Construction*, *133*(September 2021), 103996. https://doi.org/10.1016/j.autcon.2021.103996

Memarzadeh, M., Golparvar-Fard, M., & Niebles, J. C. (2013). Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors. *Automation in Construction*, *32*, 24–37. https://doi.org/10.1016/j.autcon.2012.12.002

Meng, L., Peng, Z., Zhou, J., Zhang, J., Lu, Z., Baumann, A., & Du, Y. (2020). Real-time detection of ground objects based on unmanned aerial vehicle remote sensing with deep learning: Application in excavator detection for pipeline safety. *Remote Sensing*, *12*(1). https://doi.org/10.3390/RS12010182

Meza, J., Marrugo, A. G., Ospina, G., Guerrero, M., & Romero, L. A. (2019). A structure-from-motion pipeline for generating digital elevation models for surface-runoff analysis. *Journal of Physics: Conference Series*, *1247*(1). https://doi.org/10.1088/1742-6596/1247/1/012039

MMDetection Contributors. (2018). *OpenMMLab Detection Toolbox and Benchmark*. https://github.com/open-mmlab/mmdetection

Muja, M., & Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP 2009 - Proceedings of the 4th International Conference on Computer Vision Theory and Applications*, *1*, 331–340. https://doi.org/10.5220/0001787803310340

Nassar, K., & Jung, Y. (2012). Structure-from-motion approach to the reconstruction of surfaces for earthwork planning. *Journal of Construction Engineering and Project Management*, *2*(3), 1–7. https://doi.org/10.6106/jcepm.2012.2.3.001

Oloufa, A. A. (1991). Triangulation applications in volume calculation. *Journal of Computing in Civil Engineering*, *5*(1), 103–119. https://doi.org/10.1061/(asce)0887-3801(1991)5:1(103)

Park, H. C., Rachmawati, T. S. N., & Kim, S. (2022). UAV-based high-rise buildings earthwork monitoring—a case study. *Sustainability (Switzerland)*, *14*(10179). https://doi.org/10.3390/su141610179

Park, J., Cho, Y. K., & Kim, S. (2022). Deep learning-based uav image segmentation and inpainting for generating vehicle-free orthomosaic. *International Journal of Applied Earth Observation and Geoinformation*, *115*(June), 103111. https://doi.org/10.1016/j.jag.2022.103111

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context encoders: Feature learning by inpainting. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 2536–2544. https://doi.org/10.1109/CVPR.2016.278

Pingel, T. J., Clarke, K. C., & McBride, W. A. (2013). An improved simple morphological filter for the terrain classification of airborne lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *77*(October 2017), 21–30. https://doi.org/10.1016/j.isprsjprs.2012.12.002

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3982–3992. https://doi.org/10.18653/v1/d19-1410

Rezazadeh Azar, E., Dickinson, S., & McCabe, B. (2013). Server-customer interaction tracker: Computer vision–based system to estimate dirt-loading cycles. *Journal of Construction Engineering and Management*, *139*(7), 785–794. https://doi.org/10.1061/(asce)co.1943-7862.0000652

Rezazadeh Azar, E., & McCabe, B. (2012a). Automated visual recognition of dump trucks in construction videos. *Journal of Computing in Civil Engineering*, *26*(6), 769–781. https://doi.org/10.1061/(asce)cp.1943-5487.0000179

Rezazadeh Azar, E., & McCabe, B. (2012b). Part based model and spatial-temporal reasoning to recognize hydraulic excavators in construction images and videos. *Automation in Construction*, *24*, 194–202. https://doi.org/10.1016/j.autcon.2012.03.003

Roberts, D., & Golparvar-Fard, M. (2019). End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level AAA. *Automation in Construction*, *105*(August 2018), 102811. https://doi.org/10.1016/j.autcon.2019.04.006

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241. https://doi.org/10.1109/ACCESS.2021.3053408

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. *Proceedings of the IEEE International Conference on Computer Vision*, 2564–2571. https://doi.org/10.1109/ICCV.2011.6126544

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Samad, A. M., Asri, N. A., & Ahmad, A. (2012). The use of digital image for volume determination using digital close range photogrammetric method. *Proceedings - 2012 IEEE 8th International Colloquium on Signal Processing and Its Applications, CSPA 2012*, 321–324. https://doi.org/10.1109/CSPA.2012.6194742

Schonberger, J. L., & Frahm, J. M. (2016). Structure-from-Motion revisited. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 4104–4113. https://doi.org/10.1109/CVPR.2016.445

Siebert, S., & Teizer, J. (2014). Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system. *Automation in Construction*, *41*, 1–14. https://doi.org/10.1016/j.autcon.2014.01.004

SiteEye. (2023). *SiteEye: Cloud-based Photogrammetry and Project Management*. www.siteeye.co

Slattery, K. T., Slattery, D. K., & Peterson, J. P. (2012). Road construction earthwork volume calculation using three-dimensional laser scanning. *Journal of Surveying Engineering*, *138*(2), 96–99. https://doi.org/10.1061/(asce)su.1943-5428.0000073

Soltani, M. M., Zhu, Z., & Hammad, A. (2016). Automated annotation for visual recognition of construction resources using synthetic images. *Automation in Construction*, *62*, 14–23. https://doi.org/10.1016/j.autcon.2015.10.002

Soltani, M. M., Zhu, Z., & Hammad, A. (2017). Skeleton estimation of excavator by detecting its parts. *Automation in Construction*, *82*(June), 1–15. https://doi.org/10.1016/j.autcon.2017.06.023

Soviany, P., & Ionescu, R. T. (2018). Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction. *Proceedings - 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2018*, 209–214. https://doi.org/10.1109/SYNASC.2018.00041

Starck, J. L., Elad, M., & Donoho, D. L. (2005). Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing*, *14*(10), 1570–1582. https://doi.org/10.1109/TIP.2005.852206

Stone, W. C., Cheok, G., & Lipman, R. (2000). Automated earthmoving status determination. *Proceedings of the 4th International Conference and Exposition on Robotics for Challenging Situations and Environments - Robotics 2000*, *299*, 111–119. https://doi.org/10.1061/40476(299)14

Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., & Lempitsky, V. (2022). Resolution-robust large mask inpainting with Fourier convolutions. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2149–2159. https://doi.org/10.1109/WACV51458.2022.00323

Tajeen, H., & Zhu, Z. (2014). Image dataset development for measuring construction equipment recognition performance. *Automation in Construction*, *48*, 1–10. https://doi.org/10.1016/j.autcon.2014.07.006

Tamin, M. A., Darwin, N., Majid, Z., Mohd Ariff, M. F., Idris, K. M., & Manan Samad, A. (2019). Volume Estimation of Stockpile Using Unmanned Aerial Vehicle. *Proceedings - 9th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2019*, 49–54. https://doi.org/10.1109/ICCSCE47578.2019.9068543

Tanoli, W. A., Seo, J. W., Sharafat, A., & Lee, S. S. (2018). 3D design modeling application in machine guidance system for earthwork operations. *KSCE Journal of Civil Engineering*, *22*(12), 4779–4790. https://doi.org/10.1007/s12205-018-0309-y

Tschumperlé, D., & Deriche, R. (2005). Vector-valued image regularization with PDE's: A common framework for different applications. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *27*(4), 506–517. https://doi.org/10.1109/cvpr.2003.1211415

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality

assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

Wei, J., Long, C., Zou, H., & Xiao, C. (2019). Shadow inpainting and removal using generative adversarial networks with slice convolutions. *Computer Graphics Forum*, *38*(7), 381–392. https://doi.org/10.1111/cgf.13845

Wierzbicki, D., & Nienaltowski, M. (2019). Accuracy analysis of a 3D model of excavation, created from images acquired with an action camera from low altitudes. *ISPRS International Journal of Geo-Information*, *8*(2). https://doi.org/10.3390/ijgi8020083

Wingtra. (2023). *Why and how to use drones in construction and infrastructure*. https://wingtra.com/drone-mapping-applications/drones-in-construction-and-infrastructure/

Witherow, M. A., Sazara, C., Winter-Arboleda, I. M., Elbakary, M. I., Cetin, M., & Iftekharuddin, K. M. (2019). Floodwater detection on roadways from crowdsourced images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, *7*(5–6), 529–540. https://doi.org/10.1080/21681163.2018.1488223

Wu, C. (2013). Towards linear-time incremental structure from motion. *Proceedings - 2013 International Conference on 3D Vision, 3DV 2013*, 127–134. https://doi.org/10.1109/3DV.2013.25

Wu, X., Sahoo, D., & Hoi, S. C. H. (2020). Recent advances in deep learning for object detection. *Neurocomputing*, *396*, 39–64. https://doi.org/10.1016/j.neucom.2020.01.085

Xiang, X., Lv, N., Guo, X., Wang, S., & El Saddik, A. (2018). Engineering vehicles detection based on modified faster R-CNN for power grid surveillance. *Sensors (Switzerland)*, *18*(7). https://doi.org/10.3390/s18072258

Xiao, B., & Kang, S.-C. (2021a). Development of an image data set of construction

machines for deep learning object detection. *Journal of Computing in Civil Engineering*, *35*(2), 1–18. https://doi.org/10.1061/(asce)cp.1943-5487.0000945

Xiao, B., & Kang, S.-C. (2021b). Vision-based method integrating deep learning detection for tracking multiple construction machines. *Journal of Computing in Civil Engineering*, *35*(2), 04020071. https://doi.org/10.1061/(asce)cp.1943-5487.0000957

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*(October 2020), 103482. https://doi.org/10.1016/j.autcon.2020.103482

Yakar, M., Yilmaz, H. M., & Mutluoglu, O. (2010a). Close range photogrammetry and robotic total station in volume calculation. *International Journal of Physical Sciences*, *5*(2), 086–096.

Yakar, M., Yilmaz, H. M., & Mutluoglu, O. (2010b). Comparative evaluation of excavation volume by tls and total topographic station based methods. *Lasers in Engineering*, *19*(5–6), 331–345.

Yilmaz, H. M., & Yakar, M. (2008). Computing of volume of excavation areas by digital close range photogrammetry. *Arabian Journal for Science and Engineering*, *33*(1 A), 63–79.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. (2019). Free-form image inpainting with gated convolution. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, *2019-Octob*, 4470–4479. https://doi.org/10.1109/ICCV.2019.00457

Zhang, S., & Zhang, L. (2022). Construction site safety monitoring and excavator activity analysis system. *Construction Robotics*, *6*(2), 151–161. https://doi.org/10.1007/s41693-022-00077-0

Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X., & Yan, G. (2016). An easy-to-use airborne lidar data filtering method based on cloth simulation. *Remote Sensing*, *8*(6), 501. https://doi.org/10.3390/rs8060501

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning deep features for scene recognition using places database - supplementary materials. *Advances in Neural Information Processing Systems*, *27*.

Zhou, Q.-Y., Park, J., & Koltun, V. (2018). Open3D: A modern library for 3D data processing. *ArXiv Preprint*.

Zhu, Z., Ren, X., & Chen, Z. (2017). Integrated detection and tracking of workforce and equipment from construction jobsite videos. *Automation in Construction*, *81*(April), 161–171. https://doi.org/10.1016/j.autcon.2017.05.005

CURRICULUM VITAE

Surname, Name: Ersöz, Ahmet Bahaddin

Email: abersoz@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
MS	METU Civil Engineering	2016
BS	METU Civil Engineering	2013
High School	Bahçelievler Deneme Lisesi, Ankara	2008

FOREIGN LANGUAGES

Fluent English

PUBLICATIONS

1. A. B. Ersoz and O. Pekcan, “İnsansız Hava Araçları Kullanılarak İnşaat Projeleri İçin Otonom Hafriyat Takip Sistemi,” in XI. Türkiye Ulusal Fotogrametri ve Uzaktan Algılama Birliği (TUFUAB) Teknik Sempozyumu, 2022.

2. A. B. Ersoz, O. Pekcan, and O. B. Tokdemir, “Lean Project Management using Unmanned Aerial Vehicles,” Tamap Journal of Engineering, 2019. doi: 10.29371/2018.3.65.

3. M. Altun, A. B. Ersoz, T. Teke, T. Kurt, A. Akcamete-Gungor, and O. Pekcan, “Application of Artificial Neural Networks on Building Energy Estimation,” in International Conference on Engineering Technologies (ICENTE’17), 2017.

4. A. B. Ersoz, O. Pekcan, and T. Teke, “Crack Identification for Rigid Pavements using Unmanned Aerial Vehicles,” IOP Conference Series: Materials Science and

Engineering, vol. 236, no. 1, 2017, issn: 1757899X. doi: 10.1088/1757-899X/236/1/012101.

5. A. B. Ersoz, O. Pekcan, and T. Teke, “Unmanned Aerial Vehicle Based Pavement Crack Identification Method Integrated With Geographic Information Systems,” in Transportation Research Board 96th. Annual Meeting, Washington, DC. USA, 2017

6. O. Pekcan, A. B. Ersoz, and T. Teke, “Pavement Crack Detection Using Unmanned Air Vehicles,” in Transportation Research Board 95th. Annual Meeting, Washington, DC. USA, 2016.

7. O. Pekcan, A. B. Ersoz, T. Teke, G. G. Dogan, G. O. Tarakcioglu, and C. Baykal, “Damage Identification of Coastal Structures Using Image Processing Techniques,” in 35th International Conference on Coastal Engineering, Istanbul, TURKEY, 2016.