Project 2336 | March 2025

SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Smart Highway Construction Site Monitoring Using Artificial Intelligence

Mehran Mazari, PhD
Ly Jacky N Nhiayi
Yahaira Nava-Gonzalez
Mohamad H. Saleh

MINETA TRANSPORTATION INSTITUTE

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the Mineta Consortium for Emerging, Efficient, and Safe Transportation (MCEEST) funded by the U.S. Department of Transportation, the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1 and the Climate Change and Extreme Events Training and Research (CCEETR) Program funded by the Federal Railroad Administration. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce Development

To ensure the efficient movement of people and goods, we must prepare the next generation of skilled transportation professionals who can lead a thriving, forward-thinking transportation industry for a more connected world. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 25-08

Smart Highway Construction Site Monitoring Using Artificial Intelligence

Mehran Mazari, PhD

Yahaira Nava-Gonzalez

Ly Jacky N Nhiayi

Mohamad H. Saleh

March 2025

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
25-08		

4. Title and Subtitle	5. Report Date
Smart Highway Construction Site Monitoring Using Artificial Intelligence	March 2025
	6. Performing Organization Code
7. Authors Mehran Mazari: https://orcid.org/0000-0003-4988-951X Yahaira Nava-Gonzalez Ly Jacky N Nhiayi Mohamad H. Saleh	8. Performing Organization Report CA-MTI-2336
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219	10. Work Unit No. 11. Contract or Grant No. 65A0660
12. Sponsoring Agency Name and Address California Department of Transportation 1120 N Street Sacramento, CA 95814	13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplemental Notes
10.31979/mti.2025.2336

16. Abstract
Construction is a large sector of the economy and plays a significant role in creating economic growth and national development, and construction of transportation infrastructure is critical. This project developed a method to detect, classify, monitor, and track objects during the construction, maintenance, and rehabilitation of transportation infrastructure by using artificial intelligence and a deep learning approach. This study evaluated the performance of AI and deep learning algorithms to compare their performance in detecting and classifying the equipment in various construction scenes. Our goal was to find the optimized balance between the model capabilities in object detection and memory processing requirements. Due to the lack of a comprehensive image database specifically developed for transportation infrastructure construction projects, the first portion of this study focused on preparing a comprehensive database of annotated images for various classes of equipment and machinery that are commonly used in roadway construction and rehabilitation projects. The second part of the project focused on training the deep learning models and improving the accuracy of the classification and detection algorithms. The outcomes of the trained and improved deep learning classification model were promising in terms of the precision and accuracy of the model in detecting specific objects at a highway construction site. It should be noted that the scope of this project was limited to the image and video data recorded from the ground-level and cannot be extended to Uncrewed Aerial System (UAS) data. This study provides valuable insights on the potentials of AI and deep learning to improve the monitoring and thus safety and efficiency of transportation infrastructure construction.

17. Key Words	18. Distribution Statement
Infrastructure, construction, artificial intelligence, deep learning, transportation.	No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	37	

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2025.2336

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2336

ACKNOWLEDGMENTS

We would like to acknowledge the support from Caltrans staff and more specifically Abdullah Faiyaz, Aaron Chamberlin, and Balwinder Tarlok for their guidance and leadership. We also would like to thank Charles Leong for facilitating field visits, and Joseph Anderson and Shawn Silva for on-site data collection. This project could not be completed without support from Hilary Nixon and the Mineta Transportation Institute at San José State University. The author thanks Lisa Rose and Editing Press for editorial services, as well as MTI staff research data scientist Cameron Simons and graphic design assistant Katerina Earnest.

Contents

LIST OF FIGURES

Executive Summary

The construction industry's most important tasks are process monitoring and staying on schedule. However, conventional methods prove to be costly, time consuming, and inconsistent. The main goal of this project was to develop a method to detect, classify, monitor, and track equipment and other surrounding objects during the construction, maintenance, and rehabilitation of transportation infrastructure by using artificial intelligence and a deep learning approach. The main research question in this project was the following: "Can we detect and monitor various construction elements, including equipment, machinery, and workers by using videos captured during the construction, maintenance, and rehabilitation of transportation infrastructure projects and use the outcomes for process optimization, resource allocation, productivity analysis, improved work zone safety, and maximizing efficiency?" The proposed model employs Deep Learning (DL) and Computer Vision (CV) algorithms to increase the accuracy and speed of the object detection process in recorded videos.

Our research project aimed to develop and deploy a robust algorithm that can identify, detect, classify, and track different objects in the videos and images captured from the construction and rehabilitation sites, which were acquired from actual construction and rehabilitation projects in collaboration with Caltrans. The first portion of this study was focused on preparing a comprehensive database of annotated images for various classes of equipment and machinery that are commonly used in roadway construction and rehabilitation projects. The second part of the project was focused on training the deep learning models and improving the accuracy of the classification and detection algorithms. The applications of the developed algorithms in this study include, but are not limited to, improving construction efficiency, advancing the construction monitoring process, and improving work zone safety measures.

The dataset collected and processed in this project is one of the most unique and specialized datasets that has been developed for the classification of highway construction machinery. The outcomes of the trained and improved deep learning classification model are promising in terms of the precision and accuracy of the model in detecting specific objects at a highway construction site. The model achieves high confidence scores, typically above 0.8, for diverse equipment including mobile cranes, dump trucks, and excavators.

Our model also demonstrates robust performance in new scenarios, maintaining high confidence scores on unseen images, with all images meeting a floor of 0.8 and reaching highs of 0.95 for graders and compactors. Judging from the precision-recall curves, the model achieves both high precision and high recall simultaneously, suggesting excellent practical utility for construction equipment detection tasks. It should be noted that the scope of this project was limited to the image and video data recorded from the ground level and cannot be extended to Uncrewed Aerial System (UAS)-based data. Identification and detection of specific construction machinery from

UAS footage requires a separate dataset specifically curated for aerial imagery which can be pursued in the future.

1. Introduction

Construction is a large sector of the economy and plays a significant role in creating economic growth and national development (Giang et al. 2011). The global construction industry makes up around 13% of the world's Gross Domestic Product (GDP) and is projected to rise. At the same time, construction projects create a range of job opportunities for 7% of the world's working class (Barbosa et al. 2017). Salling et al. (2015) analyzed projects in Great Britain, Denmark, Sweden, and Norway and concluded that 77% of projects experienced cost overruns, with the average amount being 29% of the original contract value. The most important contributors to these overruns are traditional project management methods and poor labor productivity (Gonzalez et al. 2014).

Process monitoring and staying on the schedule are among the top priorities in the construction industry. Conventional methods prove to be costly, time consuming, and inconsistent. To overcome this issue, researchers have aimed to study innovative methods meant to address current construction monitoring flaws. The studies reviewed in this chapter focus on the use of artificial intelligence as a solution for the construction industry. Artificial intelligence has gained traction due to its limitless applications in various industries. One of those applications in the construction industry is the use of robotics and automation to collect and analyze data. It also facilitates the development of essential information when monitoring a project.

During construction, resources and time are being wasted because of the lack of productivity management practices. Moreover, projects are cyclical and require monitoring. Current transportation project progress monitoring involves paper or electronic checklists, daily reports, verbal updates, site photographs, material delivery receipts, inventory reports, and sub-contractor invoices (Vik and Brilakis 2018). Thus, most of a project manager's time is spent deciphering data while new project activities are being started. The following sections discuss the role of new technologies such as AI in the construction industry, specifically highway construction projects.

1.1 Construction Monitoring Innovations

The following section summarizes a few studies in the literature that focus on the use of process automation and other innovative technologies to improve the construction monitoring process and enhance productivity.

Vick and Brilakis (2018) aimed to automate the process of collecting information quickly and efficiently. They proposed a model-guided hierarchical space portioning data structure for detecting discrete regions of three-dimensional as-builts. This data structure was named BrickTree (Figure 2), which was used for detecting layered road design surfaces in regions of as-built cloud data. Researchers experimentally confirmed the results by achieving an F1 score of 95.2% on real-world data, supporting its feasibility.

Seo at al. (2015) analyzed computer vision techniques for construction safety and health monitoring. They divided their findings into three main categories as object detection, object tracking, and action recognition. This approach resulted in challenges for comprehensive scene understanding, varying tracking accuracy by camera position, and action recognition of various equipment and workers. The dynamic conditions of construction sites made it difficult to detect specific objects.

Rehman at al. (2022) reviewed various methods that encouraged the use of automated computer vision (CV)-based construction progress monitoring (CPM). The CV-based process originates from four sub-processes: data acquisition, information retrieval, progress estimation, and output visualization. The researchers concluded that CV-based CPM is based on resolving technical feasibility studies using image-based processing of site data, which is experimental and not connected to its applications for construction management.

Riyanto et al. (2021) used an Uncrewed Aerial Vehicle (UAV) to collect qualitative and quantitative data in real time. They applied both conventional and AI methods to measure progress of a local highway project during construction. The conventional approach required the use of a total station, where points were recorded and later translated to 2D drawings. The AI approach consisted of a UAV recording data points and images which were later translated to 3D images. The results concluded that the UAV-provided 3D photogrammetric data was faster and more productive than its counterpart.

Samsami et al. (2021) focused on illustrating photometric data from an Uncrewed Aerial System (UAS) to building information modeling (BIM) parameters and their application for automated construction monitoring. The UAS collects data and images by flying over the site of interest. The data is later processed through Structure from Motion (SfM) photogrammetric software packages to create orthophotos, point clouds, and 3D models. The data is used as an input for feature extraction to produce geometry parameters. Their findings indicated that UAS visual data and analysis is a fast and cost-effective method for tracking construction progress. The next section summarizes some of the applications of AI in construction monitoring and management.

1.2 Role of Artificial Intelligence (AI) in the Construction Industry

In this section, the integration between AI and the construction industry is explored. First, artificial intelligence must be introduced and explained. When discussing AI for construction engineering and management (CEM), five topics apply: knowledge representation and reasoning, computer vision, natural language processing, intelligence optimization, and process mining. All these topics involve AI technology and its functions.

Knowledge representation and management is an early form of AI, which consists of a symbolic representation of domain knowledge and predefined rules to create a knowledge-based system. Computers can rationally understand the available knowledge, facts, and beliefs from the real world to draw valid conclusions and communicate logic in an efficient manner. Applications in CEM include probability-based reasoning, rule-based reasoning, and the fuzzy cognitive map (FCM). Probability can be associated with risk analysis where the fault tree analysis can apply. Rule-based reasoning deploys a set of rules in the "if <conditions>, then <conclusion>" format. Other commands like AND, OR, NOT, and others would apply. A fuzzy cognitive map is developed from data or expert opinions. By combining fuzzy logic and a cognitive map, the fuzzy graph structure interprets complex relationships which provide immediate identification of root causes of a risk event under complex conditions.

CV relies on data collection equipment such as Light Detection and Ranging (LiDAR), UAV/UAS, and others to offer remote solutions when monitoring a project. After data collection, data can be converted into visual information. Computer vision technology automates the task of collecting data and will gradually replace in-person manual observation. It can also detect unsafe conditions or behavior of infrastructures or construction sites. Fang et al. (2018) concluded that their application of computer vision in detecting improper use of personal protective equipment and recognizing failure in following the safety procedure was successful. In addition, 3D point cloud is another type of data to evaluate exact conditions of facilities with spatial information. They have been applied to the entire process of construction projects to either find objects or monitor progress. Project quality is expected to improve by using 3D model reconstruction and geometry (Wang et al. 2019).

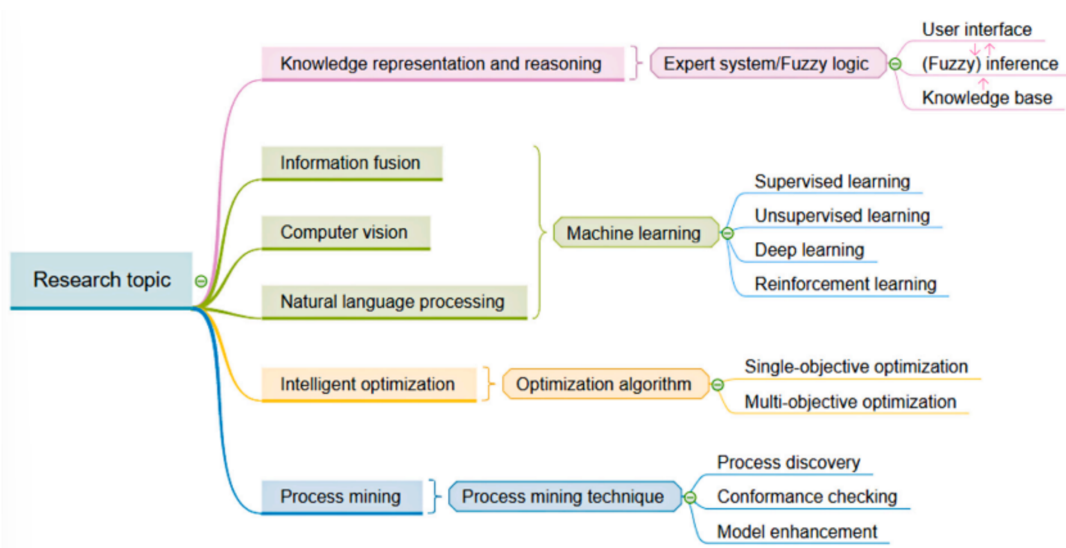
Natural language processing (NLP) and Large Language Models (LLMs) teach computers to understand language related data into the form of text and words, a human-like natural language comprehension. The traditional way of studying free-text data leaves out valuable information due to the large volumes of data; NLP and LLMs can address this shortcoming. Such technology has the potential to investigate lots of text files that can improve construction safety in the form of safety reports. These reports are typically unstructured or semi-structured with unimportant information. NLP and LLMs can extract valuable information to learn incident precursors to improve safety and lower the chance of re-occurrence. They can also convert unstructured documents with different contents into visual information such as compliance checking of BIM-based building designs recorded by Industry Foundation Classes (IFC) schema against building codes.

Intelligent optimization is a task of searching for the best solution to minimize or maximize an objective function subjected to a set of constraints. One type of this task is the simple version where the aim is to name a single best alternative. Optimization is considered helpful in prioritization, an essential task for construction management. It can maximize labor stability, minimize completion time and cost, balance workload, and analyze the changing demand of the project. Another application is for structural design problems, where it can deal with a series of design constraints to find suitable structure size and shape.

Process mining consists of exploring event logs, the connection between event logs, and the operational process. It can provide transparent and fact-based insights from real event logs to improve project monitoring. The three major types of process mining are process discovery, progress conformance, and model enhancement. Event logs can learn to automatically create process models as a reflection of the actual process and calculate metrics. Software products create a visual map to clearly describe the process and later an advanced analysis can perform diagnosis, exploration, prediction, and other functions. As construction has process deviations, bottlenecks, and hidden knowledge about productivity, the full potential of a BIM event log can be used (Pan et al. 2021).

Pan and Zhang (2021) reviewed the role of artificial intelligence in construction engineering and management (CEM). The researchers presented a systematic review using quantitative and qualitative analyses to illustrate the current and future state of AI adoption in CEM. Their study explains six trendy research topics that give an advantage to AI in CEM such as knowledge representation and reasoning, information fusion, computer vision, natural language processing, intelligent optimization, and process mining. The applications to CEM include smart robotics, cloud virtual and augmented reality (cloud VR/AR), Artificial Intelligence of Things (AIoT), digital twins, 4D printing, and blockchains. The study argues that different scenarios require different AI techniques.

Figure 1.1. Mind Map of AI Methods (Pan et al. 2021)



1.3 Benefits of AI in Construction

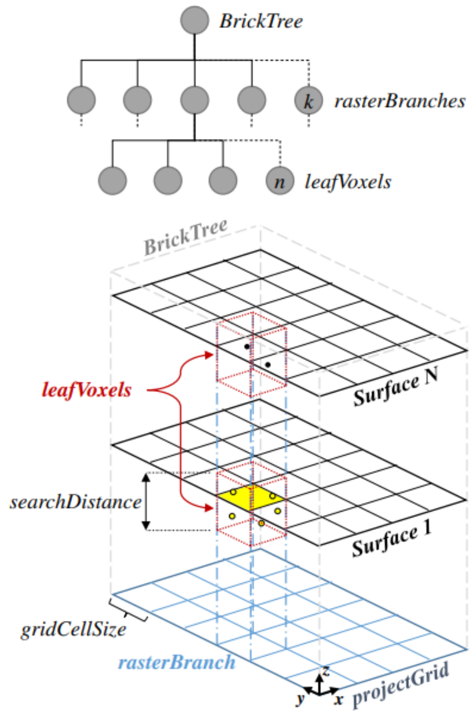
Automation processes use AI to make project management more straightforward and objective. It is known that conventional methods rely on manual observation and operation which include bias and can be confusing. For instance, on-site construction monitoring relies on cameras and sensors to automatically record data and take images and videos from the progress of a project. With AI, data can be taken without human interference and can replace traditional methods which are tedious and time-consuming. As AI models can monitor, evaluate, and predict potential risks, they can be suitable for the task of risk mitigation. Various methods are found to learn from construction site data to capture relationships from accidents and their causes. It also considers the probability and severity of the risk (Vik and Brilakis 2018).

Another important use of AI is the high efficiency in optimization. It aims to run the construction problem more smoothly and efficiently. For example, process mining is a new AI technique meant to create valuable insights into complex construction flows such as workflows. It guides the execution process to avoid any unnecessary steps and any potential future problems. AI-powered robots have been used to take over repetitive and routine construction tasks such as bricklaying and welding. For project monitoring, digitalization and computer vision play an important role in improving current trends. The following section focuses on the application of computer vision in construction monitoring.

1.4 Computer Vision Techniques

Seo at al. (2015) categorize computer vision methods into three groups: object detection, object tracking, and action tracking. Object detection can be applied to the visual data collected during the construction process and later be analyzed. They found that a 3D-based data has higher accuracy than a simple 2D approach. 3D models are less sensitive to light or color variances, hold geometrical cues, and provide better separation from the background. To make computers understand a dynamic scene like a construction site, it is necessary to label the types of project-related objects such as workers, equipment, and materials on the scene. The detection process starts with dividing the image into small spatial regions to extract features from the local windows and then classify the object of interest through supervised learning.

Figure 1.2. BrickTree Structure (Vick and Brilakis 2018)

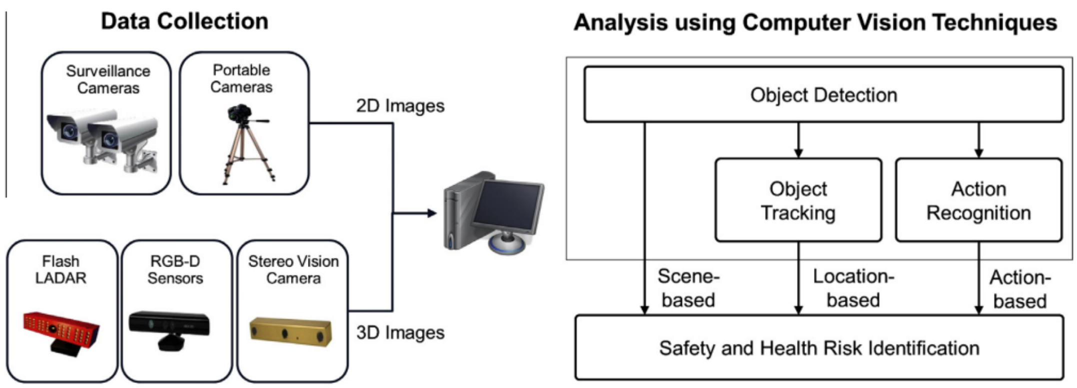


Object tracking has the advantage of not relying on sensors since cameras can cover a large field of view of the construction sites, and multiple project objects such as workers and equipment can be tracked simultaneously (Chi et al. 2020). This method can create a temporary trajectory of the detected objects as they move on site (Yilmaz et al. 2006). Once the object tracker is initialized at the first frame of video, it streams through an object detection algorithm. This algorithm tracks the 2D images of the objects by assigning labels to the tracked objects in a sequence of images.

For 2D vision tracking, three types of methods have been commonly used: point tracking, kernel tracking, and silhouette tracking. Point tracking uses feature points standing for objects and detects the object by matching those points in every frame. Kernel tracking tracks an object by computing the motion of the kernel that is the object shape and appearance in consequent frames. Silhouette tracking uses the color histogram and object edges from the object silhouette to track the object by matching said features in each frame. Applications of these techniques in construction include triangulation-based 3D positioning using various cameras (Brilakis et al. 2011) and object tracking using 3D range sensors (Shin et al. 2016).

Action recognition is beneficial when information is not sufficient to allow comprehensive understanding of the scene. This method requires image representation and action classification. Image representation types include global, local, and application-specific representations. Global representation encodes the human body, while local representation uses a collection of independent local patches. Application-specific representation uses joint locations or joint angles from human body pose and accelerations of motion. The most common approach of image representation in construction is global representation.

Figure 1.3. General Framework of Computer Vision Techniques (Seo et al. 2015)



Rehman et al. (2022) summarize the task of construction monitoring with computer vision as data acquisition, information retrieval, progress estimation, and output visualization. Data acquisition methods include the use of uncrewed aerial vehicles, handheld devices, camera systems mounted on camera stands, and surveillance cameras.

Following the review of literature related to the use of AI and computer vision in construction monitoring, the following chapter discusses the goals and scope of the current project and a summary of approach to achieve the results.

2. Project Goal and Scope

2.1 Project Goal

The main goal of this project is to develop a method to detect, classify, monitor, and track the equipment, workforce, and other surrounding objects during the construction, maintenance, and rehabilitation of transportation infrastructure by using artificial intelligence and a deep learning approach. The main research question in this project is: "Can we detect and monitor various construction elements, including equipment, machinery, and workers by using videos captured during the construction, maintenance, and rehabilitation of transportation infrastructure projects and use the outcomes for process optimization, resource allocation, productivity analysis, improved work zone safety, and maximizing efficiency?" The proposed model employs Deep Learning (DL) and Computer Vision (CV) algorithms to increase the accuracy and speed of the object detection process in recorded videos.

Our research project aims to develop and deploy a robust algorithm that can identify, detect, classify, and track different objects in the videos and images captured from the construction and rehabilitation sites, which will be crowdsourced from actual construction and rehabilitation projects in collaboration with Caltrans. Due to the lack of a comprehensive image database specifically developed for transportation infrastructure construction projects, we will focus the first portion of this study on preparing a comprehensive database of annotated images for various classes of equipment and machinery that are commonly used in roadway construction and rehabilitation projects. The second part of the project will focus on training the deep learning models and improving the accuracy of the classification and detection algorithms. The applications of the developed algorithms in this study include, but are not limited to, improving construction efficiency, advancing the construction monitoring process, and improving work zone safety measures.

2.2 Project Scope

In this study, we will evaluate the performance of AI and deep learning algorithms to compare their performance in detecting and classifying equipment in various construction scenes. Several edge cases with crowded scenes, where the target objects are occluded with other objects, will also be investigated. The detection accuracy and performance of the preliminary model will be improved once the proposed image database is developed in this study. We will provide a list of various roadway construction equipment and categorize them by activity type. Once the process of training and validation of the proposed models is complete, the algorithm will be able to detect and classify the most critical objects. Based on the availability of actual construction data, the applicability of the algorithm to both stationary and moving video sources will be evaluated. The models will be calibrated based on the properties of each image and video frame source. It should be noted that although higher video quality (i.e., higher resolution and number of frames per second) can improve the detection accuracy and tracking capabilities of the model, it will require advanced computational power and may introduce a lag in real-time tasks. Our goal is to find the optimized balance between the model capabilities in object detection and memory processing requirements.

3. Image Database Collection and Preprocessing

3.1 Introduction

In computer vision, the quality and accuracy of labeled datasets are critical for training models that perform tasks such as object detection, image classification, and semantic segmentation. The process of annotating images involves marking regions of interest and labeling objects, which provides the ground truth data required by machine learning algorithms. With the increasing demand for large, accurately labeled datasets, efficient annotation tools are essential to streamline the labeling process and ensure consistency. Roboflow and CVAT (Computer Vision Annotation Tool) are two widely adopted platforms designed to simplify and optimize dataset annotation. Both tools serve distinct but complementary purposes in the data preparation pipeline for computer vision projects.

Roboflow offers a comprehensive platform that allows users to upload images, perform image augmentations, and annotate datasets with ease. It integrates seamlessly with various machine learning frameworks, making it ideal for both beginners and experienced practitioners looking for a smooth workflow from dataset creation to model deployment. Roboflow also provides features for organizing and managing datasets, and its cloud-based nature allows users to collaborate and share datasets effortlessly.

CVAT is an open-source, web-based tool specifically tailored for detailed and precise annotation tasks. It supports a wide range of labeling options, including bounding boxes, polygons, key points, and even instance segmentation. CVAT is highly customizable and designed to handle large datasets, making it a preferred choice for projects that require high levels of precision and complex annotations. Its ability to support multiple users working simultaneously ensures that even large-scale annotation projects can be handled efficiently.

3.2 Image Data Collection

When collecting data for this project, images were carefully selected to clearly showcase the equipment of interest as the primary subject, ensuring that its features, components, and relevant details were prominent and easily identifiable. In addition to capturing the machine in high clarity, some degree of background "noise" was also desirable. This background complexity—such as varied lighting, environmental textures, and incidental objects—adds realistic variability that helps models generalize better to real-world conditions. The inclusion of controlled background elements challenges the model to differentiate between the machine and other surroundings, making it more robust and effective in diverse environments. By combining clear foreground focus with background complexity, the dataset aims to support the development of computer vision models that can reliably identify the target machine in a range of settings and contexts. A series of example images are shown below.

Figure 3.1. Example of Ideal Images for Annotation and Training

3.3 Data Labeling

Labeling images in CVAT (Computer Vision Annotation Tool) begins with setting up a new project or task and uploading the required images, which can come from local files, cloud storage, or URLs. We will then define the labels needed for this project, such as “Excavator” or “Loader” and select the correct annotation tools for each label, such as bounding boxes for object detection or polygons for segmentation tasks.

After labeling each image, we go through the dataset to review and correct any errors, ensuring consistency and high-quality annotations across the entire project. Once the quality check is complete, the annotated dataset is saved and exported in the desired format to integrate with machine learning workflows. This process ensures that every image is labeled precisely, creating a reliable dataset that enhances the accuracy and performance of computer vision models in real-world applications.

Labeling images in Roboflow begins with creating a new project and uploading the dataset. Once the images are uploaded, we can access the labeling interface to define the classes or labels relevant to this project, such as “Smooth Tandem Roller” or “Padfoot Roller.” After setting up the labels, annotation can be started by selecting the appropriate tool for labeling needs such as bounding boxes, polygons, or points, depending on the level of detail required. We will then assign the corresponding label from the list we defined earlier.

Figure 3.2. Example of CVAT Annotation and Labeling Process

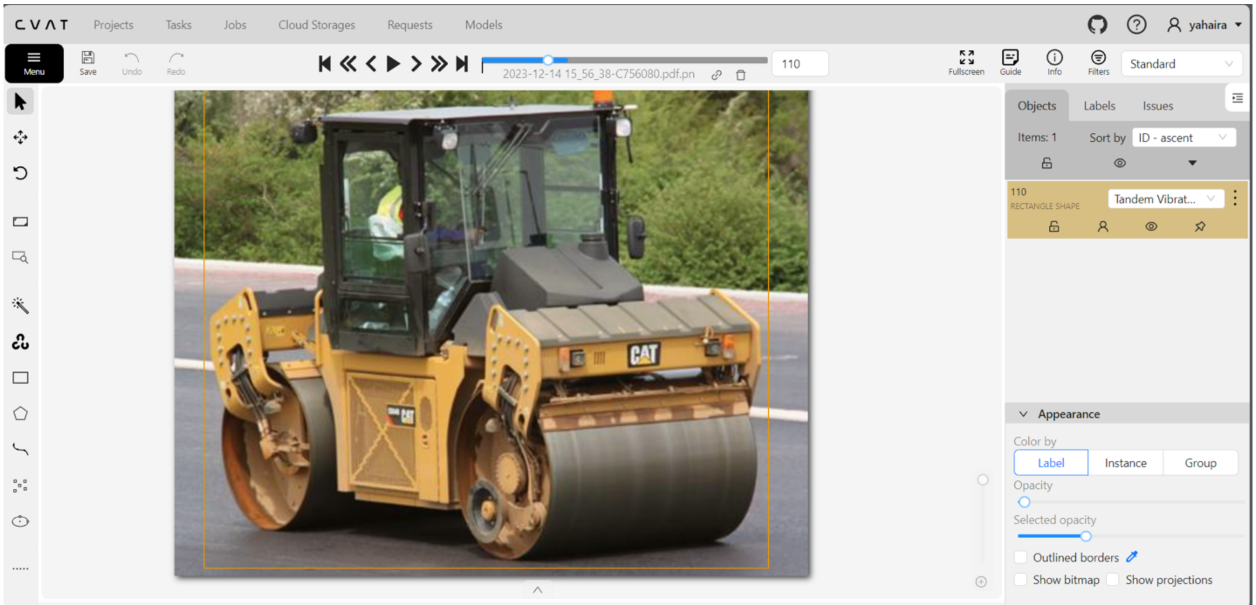
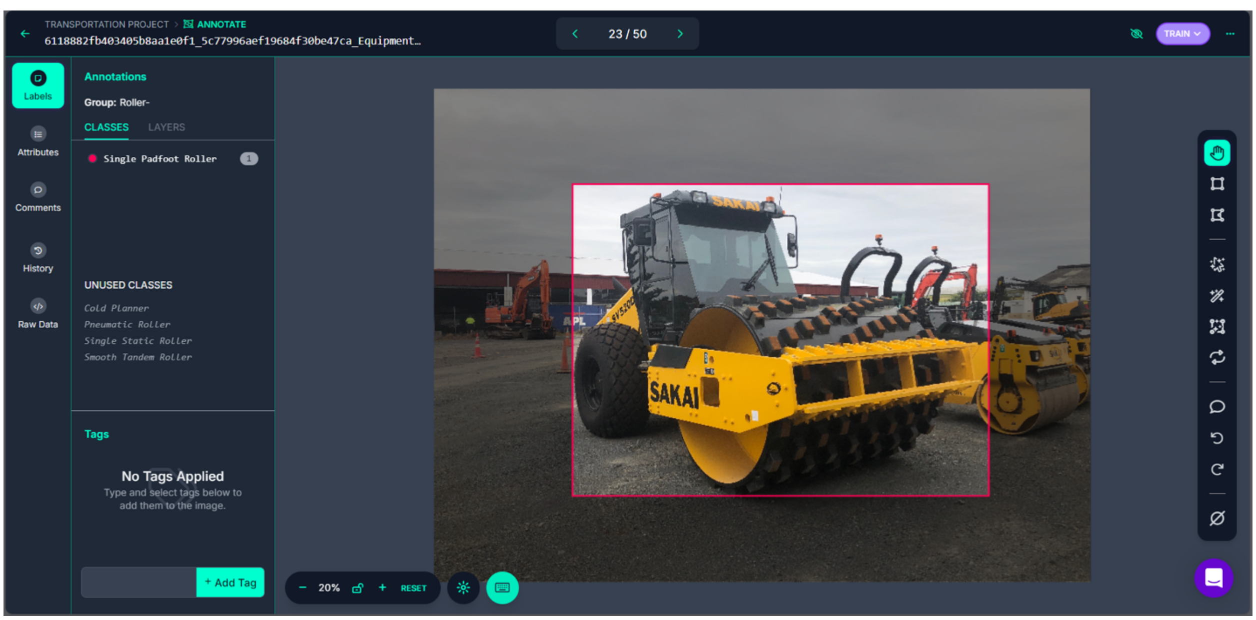


Figure 3.3. Example of Roboflow Annotation Interface



3.4 Data Augmentation

After labeling the data, data exportation is the next step. The dataset consists of regular images with no augmentations. Augmentations such as flip, rotation, shear, and noise manipulate the image so that the visibility of the object is less clear and defined. This process proves to be useful when training the model because it makes the identification task more challenging which can help considering that picture or video quality is not always optimal. Figure 3.4 shows an example of augmentations used in this project.

Figure 3.4. Example of Data Augmentation

4. Model Training, Testing, and Validation

4.1 Introduction

In computer vision, the ability to efficiently detect and segment objects in real time is important. This chapter summarizes the architecture and implementation of a deep learning model to identify and label different construction machinery on a road construction site. This section of the report outlines the comprehensive architecture of the system, detailing the data flow, as well as the integration of the You Only Look Once (YOLO) model. We describe how our custom-built model facilitates efficient communication between the user and the machine learning models, ensuring that input data are processed with precision.

A significant highlight is our approach to enhancing the YOLO model by incorporating new classes without compromising its core capabilities. By concatenating two YOLO models, we preserve existing weights while introducing custom weights for our new dataset. This innovative method allows for efficient training and accurate object recognition, even for previously untrained classes. Furthermore, we explore the state-of-the-art segmentation model, detailing its extensive training process and optimization strategies that have led to performance metrics. Our efforts in data manipulation and quality control underscore the importance of robust data handling in achieving high model accuracy.

4.2 Model Integration

The YOLO model component is designed to accept either image or video input from a ground-level point of view. Upon receiving this input, the model initiates an intermediary, enabling seamless interaction between the user interface and the trained machine learning models. The deep learning model processes the input data using the respective algorithms. The output generated by the model is then captured by the backend system.

Traditionally, when training the YOLO model on new data with pre-trained weights, the model tends to forget its old labels unless the new data includes them. To retrain the entire model on all its 80 classes, including the new classes we added for our project, one would typically need to retrain the model on the entire original dataset alongside the new dataset we collected in this project. However, this approach could be more efficient, as retraining the original model would require a significant amount of time due to the large size of the dataset. A simpler alternative would be to use a mini dataset which would have significantly less data than the original dataset for object detection. However, the drawback with this approach is that the models will not be as accurate as the YOLO model trained on the original dataset. Therefore, we needed to find a method to add new classes to the model while preserving the existing weights, without training on the original dataset. Our solution to this problem was as follows:

We essentially concatenated two YOLO models together. This approach preserves the old YOLO weights while allowing us to train custom weights for the new classes. YOLO has two outputs: one represents the class of the image and the data for the bounding box. This explains why the normal YOLO model with the original weights has an output of (N, 84, 6300) since it contains 80 classes and there are 4 coordinates to make the bounding box. The model utilized a library and employed the medium version of the YOLO model. Concatenating two models is a complex process, which will be discussed in the next section.

To concatenate the models, we first froze all the weights of the pre-trained YOLO model and froze the first 22 layers of our new model. This involved iteratively going through all the layers in the YOLO model and freezing each layer for both the pre-trained and our new model. While freezing the layers prevents their weights from being updated, the batch normalization statistics still get updated during training. This is why we needed a callback function to be added to put the frozen layers in eval mode to prevent the batch normalization values from changing. We performed this for every epoch since the model returns to training mode after the validation step, which undoes our changes. Figure 4.1 shows an example of two different kinds of machinery detected in a construction scene.

Figure 4.1. Example of Detected Construction Machines in an Image

We also utilized a state of the art deep-learning-based segmentation model that can identify and segment different construction machinery in images and videos of highway construction sites. The results of implementing this model on sample images are shown in Figure 4.2.

Figure 4.2. Performance of the Segmentation Model on Sample Images

4.2 Model Training

Our current model underwent an extensive training process, utilizing a robust dataset comprising over 10,000 diverse images. The training regimen consisted of 1000 epochs, a significant increase from our previous iterations. This extended training duration allowed the model to iterate through the entire dataset multiple times, facilitating deeper learning and pattern recognition. The large-scale dataset provided a wide array of scenarios and object variations, contributing to the model's ability to generalize effectively across different contexts.

To accelerate the training process, we implemented a dedicated GPU (Graphics Processing Unit) setup. This hardware optimization resulted in a remarkable 14-fold increase in processing speed compared to our previous training iterations. The substantial reduction in computational time allowed us to explore more complex model architectures and hyperparameter configurations. This efficiency gain was instrumental in achieving higher confidence scores and improved accuracy in object detection tasks, as it enabled us to conduct more extensive experiments and fine-tuning within a given timeframe.

The combination of an extensive dataset, increased training epochs, and accelerated processing capabilities led to significant improvements in our model's performance metrics. We observed notable enhancements in confidence scores, indicating the model's increased certainty in its predictions. Additionally, the accuracy of object detection showed marked improvement, with the model demonstrating a higher rate of correct identifications and classifications across various object categories and environmental conditions.

To streamline our data management processes, we developed several specialized scripts. These tools were designed to automate and optimize various aspects of data preprocessing and organization. One key script focuses on mapping classes, ensuring consistent and accurate labeling across the entire dataset. This is particularly crucial when dealing with large-scale datasets where manual classification can be prone to errors or inconsistencies. Another critical script in our toolkit is designed to detect images lacking proper labels. This quality control measure is invaluable when working with datasets comprising thousands of images. By automatically identifying unlabeled or mislabeled images, we can maintain the integrity of our training data. This process helps prevent the introduction of noise or inconsistencies into the model training, which could otherwise lead to reduced performance or biased outcomes.

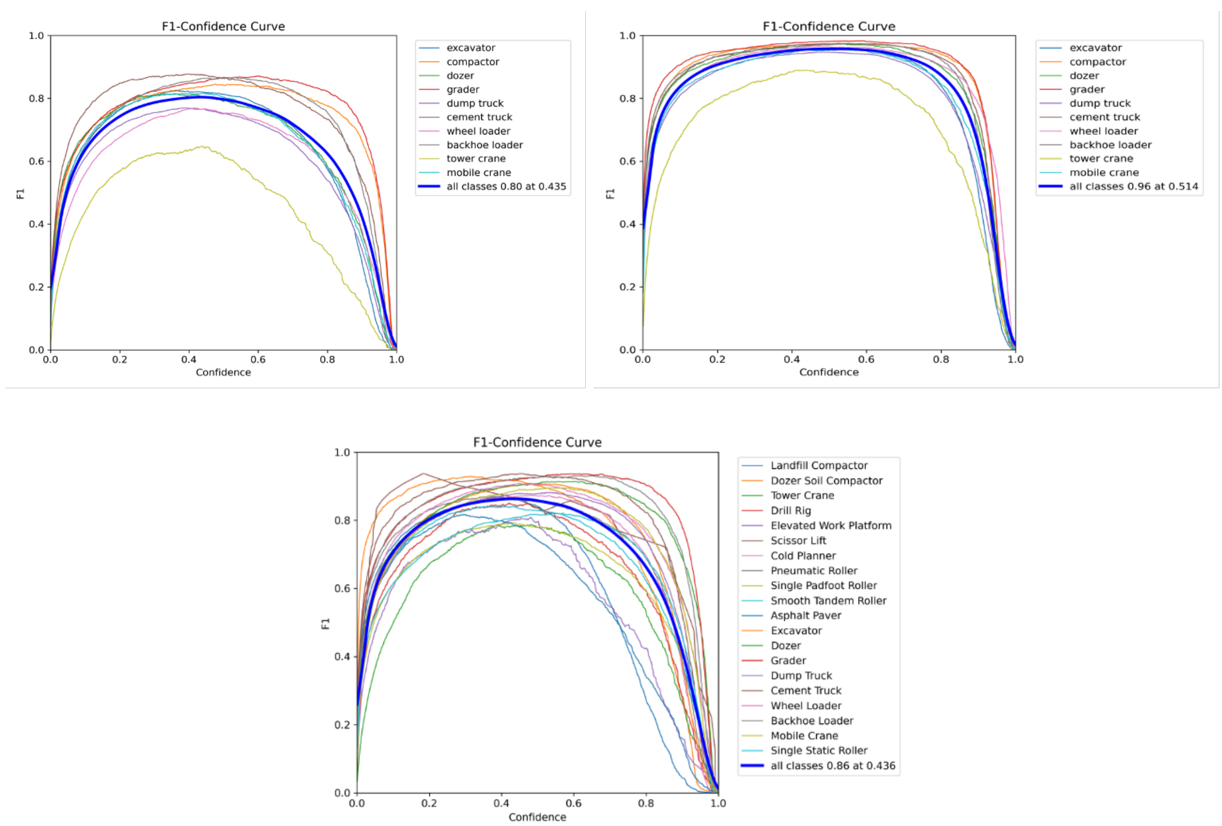
The implementation of these automated scripts significantly enhanced our ability to handle and process large volumes of data efficiently. This scalability is crucial in the context of deep learning, where the quantity and quality of training data directly impact model performance. Our approach not only improved the accuracy of our dataset but also reduced the time and resources required for data preparation, allowing us to focus more on model development and optimization.

5. Results, Discussion, and Conclusions

This chapter summarizes the results of model performance based on the dataset collected in this project. We will then provide conclusions and recommendations for future research.

The performance of a deep learning classification model is usually evaluated using F1-Confidence curves by balancing precision and recall across different confidence thresholds, making them particularly valuable for evaluating object detection systems. These curves help identify optimal operating points and reveal model behavior across varying confidence levels.

Figure 5.1. F1-Confidence Curve for All Three Models

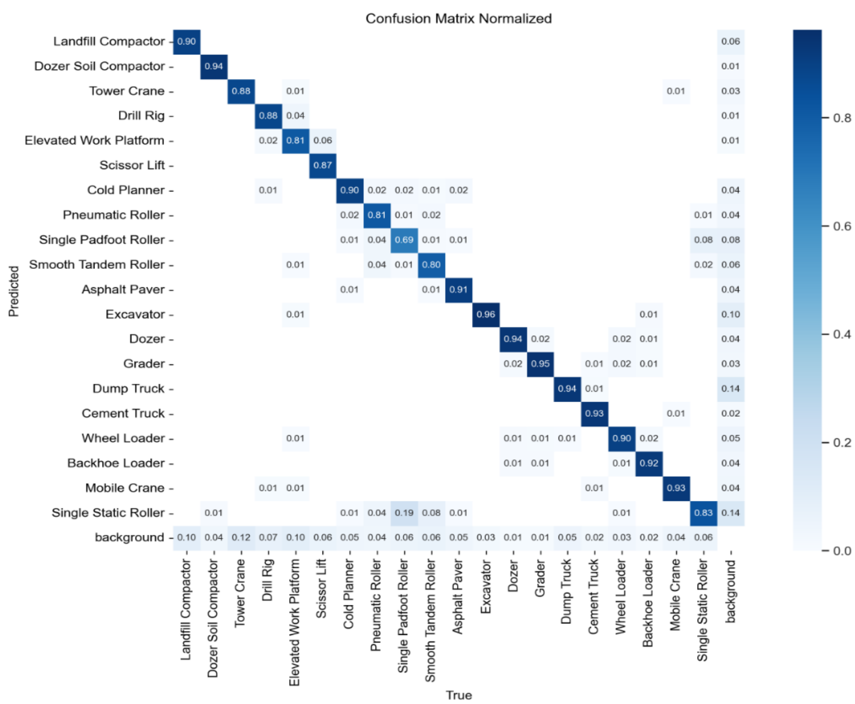


Comparison of the F1-Confidence curves between Model 1 and Model 2 reveals substantial improvements in detection performance. Model 2 achieved a higher overall F1 score of 0.96 at an optimal confidence threshold of 0.514, compared to Model 1's F1 score of 0.80 at an optimal confidence threshold of 0.435. The improved model demonstrates more consistent performance across different construction equipment classes, as evidenced by the tighter clustering of class-specific curves, suggesting enhanced generalization capabilities across various equipment types.

A normalized confusion matrix (Figure 5.2) visualizes the model's classification performance by showing the proportion of predicted classes (rows) versus true classes (columns), where perfect classification would result in values of 1.0 along the diagonal and 0.0 elsewhere. Each cell represents the fraction of predictions, with darker blues indicating higher values. The diagonal elements show strong performance with most classes achieving 0.98–0.99 classification accuracy, indicating excellent class discrimination.

The background class shows some interesting interactions, with minor misclassifications across several equipment types, notably a 0.16 rate with mobile cranes, meaning 16% of background regions were incorrectly classified as mobile cranes. Some minimal confusion (0.01–0.02) exists between functionally similar equipment pairs, such as wheel loaders with dozers and compactors, which is expected given their similar visual characteristics. The dump truck class shows slight confusion with the background (0.28), suggesting challenges in distinguishing these vehicles in complex scenes. Overall, the matrix demonstrates robust classification performance with minimal cross-class confusion, which is particularly impressive given the visual similarities between some construction equipment categories. The predominantly diagonal pattern, with very few off-diagonal elements exceeding 0.05, indicates strong class separation and reliable detection capabilities.

Figure 5.2. Confusion Matrix for Model 3



5.1 Sample Model Output

These visualizations present quantitative detection results demonstrating the model’s real-world performance on diverse construction site imagery. They aim to validate the quantitative metrics and illustrate the model's practical applicability across varying real-world scenarios and challenging conditions. Our first example shows test images with numeric class labels and colored bounding boxes, highlighting the model's ability to detect equipment in complex environments.

Figure 5.3. Sample Output for a Highway Construction Site

Tower cranes (labeled "8," green boxes) are consistently identified across multiple viewpoints, while various ground equipment such as graders and compactors (labeled "3" and "1") are accurately detected in diverse settings. The model successfully handles multiple equipment instances in single frames and maintains performance across different lighting conditions and perspectives. Figure 5.4 depicts a set of images that validate results with confidence scores and class labels, showing strong detection performance across various equipment types.

Figure 5.4. Performance of the Model in Detecting Different Construction Equipment

The model achieves high confidence scores, typically above 0.8 for diverse equipment including mobile cranes, dump trucks, and excavators. Lastly, the third set shows performance from random images concatenated together (Figure 5.5). The model demonstrates robust performance in new scenarios, maintaining high confidence scores on unseen images with all images meeting a floor of 0.8 and reaching highs of 0.95 for graders and compactors.

Figure 5.5. Construction Machinery Detection in Mixed Scenes

5.2 Conclusions

The dataset collected and processed in this project is one of the most unique and specialized datasets that has developed for classification of highway construction machinery. The outcomes of the trained and improved deep learning classification model are promising in terms of the precision and accuracy of the model in detecting specific objects at a highway construction site. It should be noted that the scope of this project was limited to the image and video data recorded from the ground level and cannot be extended to UAS-based data. Identification and detection of specific construction machinery from UAS footage requires a separate dataset specifically curated for aerial imagery which can be pursued in the future.

5.3 Recommendations for Future Research

Building upon these advancements, future research can focus on further refining the data processing pipelines and exploring more sophisticated model architectures. The focus will be to implement adaptive learning rates and advanced regularization techniques to push the boundaries of our model's performance. Additionally, future research can investigate transfer learning approaches to leverage pre-trained models, potentially accelerating the training process for specific object detection tasks while maintaining high accuracy levels.

Bibliography

Barbosa, F., Woetzel, J., & Mischke, A. (2017). *Reinventing construction: A route to higher productivity*. McKinsey Global Institute. https://www.mckinsey.com/~/media/McKinsey/Industries/Capital%20Projects%20and%20Infrastructure/Our%20Insights/Reinventing%20construction%20through%20a%20productivity%20revolution/MGI-Reinventing-Construction-Executive-summary.pdf

Brilakis, I., Park, M. W., & Jog, G. (2011). Automated vision tracking of project-related entities. *Advanced Engineering Informatics, 25*(4), 713–724. https://doi.org/10.1016/j.aei.2011.07.002

Chi, S., & Caldas, C. H. (2012). Image-based safety assessment: Automated spatial safety risk identification of earthmoving and surface mining activities. *Journal of Construction Engineering and Management, 138*(3), 341–351. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000428

Fang, W., Ding, L., Luo, H., & Love, P. E. (2018). Falls from heights: A computer vision-based approach for safety harness detection. *Automation in Construction, 91*, 53–61. https://doi.org/10.1016/j.autcon.2018.02.018

Giang, D. T., & Pheng, L. S. (2011). Role of construction in economic development: Review of key concepts in the past 40 years. *Habitat International, 35*(1), 118–125. https://doi.org/10.1016/j.habitatint.2010.06.003

González, P., González, V., Molenaar, K., & Orozco, F. (2014). Analysis of causes of delay and time performance in construction projects. *Journal of Construction Engineering and Management, 140*(1), 04013027. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000721

Pan, Y., & Zhang, L. (2021). Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Automation in Construction, 122*, 103517. https://doi.org/10.1016/j.autcon.2020.103517

Rehman, S. U., et al. (2022). Automated computer vision-based construction progress monitoring: A systematic review. *Buildings, 12*(7), 1037. https://doi.org/10.3390/buildings12071037

Riyanto, F., Juliastuti, Setyandito, O., & Pramudya, A. (2021). Realtime monitoring study for highway construction using Unmanned Aerial Vehicle (UAV) technology. *IOP Conference Series: Earth and Environmental Science, 729*(1), 012040. https://doi.org/10.1088/1755-1315/729/1/012040

Salling, K. B., & Leleur, S. (2015). Accounting for the inaccuracies in demand forecasts and construction cost estimations in transport project evaluation. *Transport Policy, 38*, 8–18. https://doi.org/10.1016/j.tranpol.2014.11.006

Samsami, R., Mukherjee, A., & Brooks, C. N. (2021). Mapping unmanned aerial system data onto building information modeling parameters for highway construction progress monitoring. *Transportation Research Record: Journal of the Transportation Research Board, 2676*(4), 669–682. https://doi.org/10.1177/03611981211064277

Seo, J., Han, S., Lee, S., & Kim, H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics, 29*(2), 239–251. https://doi.org/10.1016/j.aei.2015.02.001

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging, 35*, 1285–1298. https://doi.org/10.1109/TMI.2016.2528162

Socher, R., Huval, B., Bath, B., Manning, C. D., & Ng, A. (2012). Convolutional-recursive deep learning for 3D object classification. In *Advances in Neural Information Processing Systems* (pp. 665–673).

Stellman, J. M. (Ed.). (1998). *Encyclopaedia of occupational health and safety*. International Labour Organization.

Starbuck, R., Seo, J., Han, S., & Lee, S. (2014). A stereo vision-based approach to marker-less motion capture for on-site kinematic modeling of construction worker tasks. In *Proceedings of the 15th International Conference on Computing in Civil and Building Engineering (ICCCBE)*, Orlando, FL.

Teizer, J., Caldas, C. H., & Haas, C. T. (2007). Real-time three-dimensional occupancy grid modeling for the detection and tracking of construction resources. *Journal of Construction Engineering and Management, 133*(11), 880–888. https://doi.org/10.1061/(ASCE)0733-9364(2007)133:11(880)

Vick, S., & Brilakis, I. (2018). Road design layer detection in point cloud data for construction progress monitoring. *Journal of Computing in Civil Engineering, 32*(5), 04018029. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000772

Wang, Q., & Kim, M.-K. (2019). Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018. *Advanced Engineering Informatics, 39*, 306–319. https://doi.org/10.1016/j.aei.2019.02.007

Yilmaz, O., Javed, M., & Shah, M. (2006). Object tracking: A survey. *ACM Computing Surveys (CSUR), 38*(4), 13. https://doi.org/10.1145/1177352.1177355

About the Authors

Mehran Mazari, PhD

Mehran Mazari is an Associate Professor of Civil Engineering at California State University, Los Angeles. His research interests include Infrastructure Materials, Sustainable and Resilient Transportation Infrastructure, Applied Data Science, and Machine Learning. He has been actively involved in several federal and state-funded research projects. Mehran is the editorial board member of the CSU Journal of Sustainability and Climate Change, handling editor of the Transportation Research Record (TRR) of the Transportation Research Board, and associate editor of the International Journal of Pavement Engineering. Dr. Mazari is the founding director of the Sustainable Infrastructure Materials Research Lab (SIM-Lab) at California State University, Los Angeles. He is a member of the Highway Pavement Committee, and the Sustainable Transportation Committee of the American Society of Civil Engineers (ASCE).

Yahaira Nava-Gonzalez

Yahaira is a civil engineering undergraduate student at California State University, Los Angeles. She is the recipient of Transportation Research Board (TRB) Minority Student Fellows.

Ly Jacky N Nhiayi

Jacky is a computer science undergraduate student at California State University, Los Angeles.

Mohamad H. Saleh

Mohamad is a computer science graduate student at California State University, Los Angeles

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors

MINETA TRANSPORTATION INSTITUTE