

YOLOv5-based Object Detection for Construction Site Efficiency: Equipment, Tool, and Vehicle Recognition

Serhii Dolhopolov^{1,†}, Tetyana Honcharenko^{1,†}, Denys Chernyshev^{1,†} and Olga Solovei^{1,†}

¹ Kyiv National University of Construction and Architecture, 31, Air Force Avenue, Kyiv, 03037, Ukraine

Abstract

The integration of YOLOv5-based object detection into construction site management has emerged as a transformative approach to enhancing efficiency and safety. This study aimed to develop a model capable of real-time identification and tracking of construction resources, equipment, and vehicles using CCTV footage. By leveraging the power of computer vision and deep learning, the model facilitates optimized resource allocation, equipment utilization, and improved safety measures through the precise monitoring of tools, machinery, and vehicle movements. Utilizing a bespoke dataset, the YOLOv5 model underwent rigorous training, validation, and testing phases. The model was trained for 30 epochs with a dataset comprising 1,897 images of construction equipment, tools, and vehicles, achieving a final precision of 0.852, recall of 0.723, and mean Average Precision (mAP_{0.5}) of 0.792. These results underscore the model's high accuracy in detecting and classifying various construction-related objects, thereby demonstrating its potential to significantly enhance operational efficiency and safety on construction sites.

Keywords

Construction site, YOLOv5, recognition systems, real-time object classification

1. Introduction

In recent years, the construction industry has witnessed a significant transformation, driven by technological advancements aimed at enhancing site efficiency and safety. The integration of artificial intelligence (AI) and machine learning (ML) technologies into construction operations has emerged as a pivotal strategy for addressing the perennial challenges of resource and equipment management. Among these technologies, the YOLOv5-based object detection model stands out for its potential to revolutionize the way construction sites operate, particularly in terms of equipment utilization, tool tracking, and vehicle recognition.

The global construction sector has long grappled with issues related to safety management, with accidents on construction sites posing serious risks to workers and project timelines.

¹ITTAP'2024: 4th International Workshop on Information Technologies: Theoretical and Applied Problems, October 23-25, 2024, Ternopil, Ukraine, Opole, Poland

* Corresponding author.

† These authors contributed equally.

✉ dolhopolov@icloud.com (S. Dolhopolov); goncharenko.ta@knuba.edu.ua (T. Honcharenko); chernyshev.do@knuba.edu.ua (D. Chernyshev); solovey.ol@knuba.edu.ua (O. Solovei)

ORCID: 0000-0001-9418-0943 (S. Dolhopolov); 0000-0003-2577-6916 (T. Honcharenko); 0000-0002-1946-9242 (D. Chernyshev); 0000-0001-8774-7243 (O. Solovei)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Traditional methods of safety and resource management, often reliant on manual oversight and rudimentary tracking systems, have proven inadequate in mitigating these risks effectively. The advent of YOLOv5-based object detection offers a promising solution to these challenges, leveraging deep learning algorithms to automate the detection and classification of construction equipment, tools, and vehicles in real-time.

Recent studies have underscored the efficacy of YOLOv5 in enhancing construction site safety and efficiency. For instance, Xue et al. developed an improved YOLOv5 multiscale object detection algorithm specifically tailored for track construction safety, demonstrating significant advancements in the detection of workers and tools with enhanced accuracy and reduced model convergence time [1]. Similarly, Zhou et al. proposed an object detection method based on an improved YOLOv5 model to accurately sort construction waste, showcasing the model's superior performance over conventional models like Faster-RCNN and YOLOv4 [2]. These studies highlight the versatility and effectiveness of YOLOv5 in addressing diverse safety and management challenges on construction sites.

The integration of YOLOv5 into construction site operations not only improves safety by enabling the real-time detection of potential hazards but also enhances resource and equipment management through precise tracking and utilization monitoring. By automating the identification and classification of construction assets, YOLOv5 facilitates more efficient allocation and use of resources, thereby optimizing project timelines and reducing costs.

This paper aims to explore the application of YOLOv5-based object detection in the context of construction site management, focusing on the model's impact on enhancing site efficiency and safety. Through a review of recent literature and case studies, we will examine the implementation of YOLOv5 in various construction scenarios, its benefits in terms of resource and equipment management, and the challenges associated with deploying AI and ML technologies in the construction industry.

The integration of YOLOv5-based object detection into construction site management represents a paradigm shift towards data-driven decision-making and operational efficiency. This innovative approach not only promises to enhance safety protocols but also to streamline the management of resources and equipment, a critical aspect of construction projects that directly impacts productivity and cost-effectiveness. The adaptability and precision of YOLOv5 algorithms in identifying and tracking various objects make it an invaluable tool for construction site managers seeking to optimize equipment utilization and ensure the safety of workers.

The application of YOLOv5 in construction site management extends beyond mere object detection; it encompasses the analysis of equipment usage patterns, real-time monitoring of tool locations, and the identification of potential safety hazards. For example, Cai et al. demonstrated the effectiveness of an object detection framework based on YOLOv4, a precursor to YOLOv5, in autonomous driving scenarios, highlighting the model's balance between detection accuracy and real-time operation [3]. Although focused on autonomous vehicles, the principles and methodologies applied in their research are directly transferable to construction site management, where the detection of equipment, vehicles, and personnel in real-time can significantly enhance operational safety and efficiency.

Furthermore, Peng et al. introduced CORY-Net, a variant of YOLOv5 tailored for intelligent safety monitoring on power grid construction sites [4]. Their work underscores the potential of customized YOLOv5 models to address specific challenges in construction safety management,

such as the detection of workers in complex terrains and the identification of safety hazards. This research exemplifies the versatility of YOLOv5-based models in adapting to diverse construction environments and safety requirements.

The deployment of YOLOv5-based object detection systems on construction sites facilitates a proactive approach to safety management and resource allocation. By enabling the real-time detection and classification of construction assets and potential hazards, these systems empower site managers to make informed decisions that enhance safety and operational efficiency. Moreover, the continuous improvement and customization of YOLOv5 models, as demonstrated by ongoing research, ensure their relevance and effectiveness in meeting the evolving needs of the construction industry.

The advent of YOLOv5-based object detection models has not only promised enhancements in construction site safety and efficiency but also opened new avenues for research and development within the construction industry. The ability of these models to accurately detect, classify, and track resources and equipment in real-time presents a significant leap forward in managing the dynamic and often hazardous environment of construction sites.

The application of YOLOv5 extends to the meticulous tracking of construction equipment and tools, a critical aspect for ensuring project timelines is met and reducing idle times. Yang et al. showcased the effectiveness of YOLOv5 in monitoring compliance with safety protocols, such as the wearing of helmets and masks, by construction workers [5]. Their work not only demonstrates the model's high accuracy and efficiency in real-world scenarios but also its potential to significantly reduce the risk of accidents and enhance overall site safety.

Moreover, the customization and improvement of YOLOv5 models to suit specific construction site conditions have been a focus of recent studies. Zeng et al. introduced an enhanced YOLOv3 model for equipment detection and localization, which, while predating YOLOv5, underscores the continuous evolution and refinement of YOLO architectures for construction site applications [6]. Their research highlights the importance of adapting object detection models to the unique challenges posed by construction sites, such as the detection of small or occluded objects and the need for real-time processing.

The integration of YOLOv5-based object detection into construction site management systems represents a significant step towards automating safety and resource management processes. By providing site managers with real-time data on equipment location, usage, and worker safety compliance, these systems enable more informed decision-making, ultimately leading to improved project efficiency and reduced costs. Furthermore, the ongoing development and customization of YOLOv5 models ensure that these systems remain adaptable to the ever-changing landscape of construction site management.

The precision and efficiency of YOLOv5 in object detection have significant implications for the management of construction resources. By automating the tracking of tools and equipment, YOLOv5 models minimize the likelihood of loss and misplacement, thereby ensuring that resources are optimally utilized and readily available when needed. This capability is crucial for maintaining project schedules and reducing downtime. For instance, the work of Wan et al. on utilizing YOLOv5 for object detection in high-resolution optical remote sensing images, though focused on a different application, underscores the model's robustness and adaptability in detecting objects across various scales and conditions [7]. Such attributes are invaluable in the complex and ever-changing environment of construction sites.

Moreover, the application of YOLOv5 extends to enhancing safety measures on construction sites. Through real-time monitoring and detection of safety gear compliance, such as helmets and vests, YOLOv5 models play a pivotal role in preventing accidents and ensuring the well-being of construction workers. The research by Zhou et al. on the detection of construction waste using an improved YOLOv5 model illustrates the model's versatility and high performance in identifying specific objects within cluttered scenes [2]. This capability is directly applicable to safety monitoring on construction sites, where the ability to accurately detect personal protective equipment (PPE) amidst the site's activity can significantly impact overall safety outcomes.

The ongoing development and customization of YOLOv5 models for construction site management underscore the potential for further innovations in this field. As researchers and practitioners continue to explore new applications and enhancements of YOLOv5 technology, the construction industry stands on the cusp of a new era of digital transformation. This transformation is characterized by increased automation, improved safety protocols, and enhanced resource management, all of which contribute to the overall efficiency and success of construction projects.

As the construction industry continues to evolve, the integration of cutting-edge technologies like YOLOv5-based object detection into resource and equipment management practices has become increasingly vital. This technology's capacity to enhance construction site efficiency and safety through advanced equipment utilization, precise tool tracking, and accurate vehicle recognition marks a significant leap forward in the sector's operational capabilities.

The adaptability and efficiency of YOLOv5 in various construction site scenarios have been demonstrated through numerous studies, each contributing to the model's ongoing refinement and application. For instance, the work by Peng et al. on CORY-Net for intelligent safety monitoring on power grid construction sites exemplifies the potential of YOLOv5-based models to enhance worker safety and operational oversight [4, 8]. Similarly, the study by Yang et al. on the application of YOLOv5 for PPE compliance monitoring further underscores the model's utility in promoting construction site safety [5, 9]. These studies, among others, provide a solid foundation for exploring new avenues for applying YOLOv5 in construction site management.

The purpose of this research can be summarized as follows:

- **To Evaluate the Effectiveness of YOLOv5-Based Object Detection** in improving construction site efficiency by automating the tracking and management of resources and equipment.
- **To Assess the Impact of YOLOv5 on Construction Site Safety** through real-time detection of safety gear compliance and potential hazards, thereby reducing the risk of accidents and enhancing worker safety.
- **To Explore the Customization and Adaptation of YOLOv5 Models** for specific construction site environments, considering the unique challenges posed by diverse project sites and operational conditions.
- **To Investigate the Integration of YOLOv5 with Other Technological Solutions** such as drones and IoT devices, for comprehensive site monitoring and management.

- **To Identify Challenges and Limitations** associated with the deployment of YOLOv5-based object detection systems in construction site management, including technical, operational, and regulatory considerations.
- **To Provide Recommendations for Future Research and Development** in the field of construction technology, with a focus on enhancing the capabilities and applications of YOLOv5-based object detection for improved site management practices.

By addressing these objectives, this research aims to contribute to the body of knowledge on the application of advanced object detection technologies in construction site management. Through a detailed analysis of current practices and future potentials, we seek to illuminate the path toward a more efficient, safe, and technologically advanced construction industry.

2. Main research

The proposed study aims to enhance construction site efficiency and safety through the implementation of a YOLOv5-based object detection model. This section outlines the materials and methods used to develop, train, and deploy the model for resource and equipment management on construction sites.

Proposed Framework for Resource and Equipment Management System:

1. Data Collection.

- 1.1. **Public Datasets.** Initially, public datasets containing images of construction equipment, tools, and vehicles were utilized. These datasets offer a broad range of object types and scenarios, providing a solid foundation for the initial training of the YOLOv5 model. Public datasets (ACID [10], TTM [11]) are invaluable for introducing the model to a wide variety of objects and conditions it might encounter in real-world construction environments.
- 1.2. **Self-captured Images.** To tailor the model more closely to the specific needs and conditions of construction sites, a significant portion of the dataset was composed of self-captured images and video footage. This involved on-site data collection at various construction projects, capturing images and videos of resources, equipment in different operational states (e.g., idle, in use), and under diverse environmental conditions. This step was critical for incorporating real-world variability into the dataset, ensuring the model's effectiveness across different construction sites and conditions.

2. Preprocessing.

- 2.1. **Data Cleaning.** The first step involved filtering out irrelevant images and correcting any errors within the dataset. This process ensured that only pertinent and accurate data were included, enhancing the quality of the training material.
- 2.2. **Resize/Adjust Brightness and Contrast.** To standardize the dataset, all images were resized to a uniform dimension suitable for the YOLOv5 model. Additionally, adjustments to brightness and contrast were made where necessary to simulate various lighting conditions, further improving the model's robustness and accuracy.
- 2.3. **Image Labeling.** Using annotation tools like YOLO Label, each image in the dataset was meticulously labeled to identify and classify different types of construction

resources and equipment, along with their operational states. This step is crucial for supervised learning, as it provides the model with the necessary information to learn from the visual data.

3. **Model Training and Evaluation.**

3.1. **Train the Object Detection Model (YOLOv5).** The YOLOv5 model is configured and trained using the prepared dataset. The training process is optimized for accuracy in detecting various resources and equipment specific to construction sites.

3.2. **Evaluate the Model.** The trained model is evaluated on a separate set of images to assess its performance. Evaluation metrics include precision, recall, and mAP (mean Average Precision), providing insights into the model's effectiveness in real-world scenarios.

4. **Integration and Deployment.**

4.1. **Object Detection Model Weights (YOLOv5).** The trained model is deployed into the construction site management system, enabling real-time analysis and detection of resources and equipment.

4.2. **Input Source.** The model utilizes input sources such as CCTV footage, static images, or live video feeds for continuous object detection and monitoring.

5. **Real-time Detection and Management.**

5.1. **Detecting Resources and Equipment.** The system identifies and classifies resources and equipment in real-time, distinguishing between different types (e.g., tools, machinery) and states (idle, in use), facilitating immediate action and decision-making.

5.2. **Environmental Conditions.** The model optionally integrates environmental condition detection to adjust resource management strategies in response to weather changes, enhancing operational adaptability.

6. **Resource and Equipment Status Dashboard.**

6.1. **Visualization and Alerts.** A dashboard presents detected resources and equipment, highlighting their status, location, and usage. Alerts are generated for underutilized resources or when equipment maintenance is due, ensuring optimal resource management.

6.2. **Decision Support.** The system provides actionable insights for resource allocation, maintenance scheduling, and equipment usage optimization based on real-time data, supporting informed decision-making.

7. **Feedback Loop for Continuous Improvement.**

7.1. **Model Retraining.** New data and feedback are periodically collected to retrain the model, improving its accuracy and adapting to new types of resources or changes in the construction site environment.

7.2. **System Updates.** The management dashboard and decision support tools are updated based on insights gained from model performance and user feedback, ensuring the system's continuous improvement and relevance.

This comprehensive framework leverages YOLOv5 for object detection to manage resources and equipment on construction sites effectively and is represented as a model in Figure 1. By emphasizing the detection and classification of resources and integrating this information into

actionable insights for site managers, the system ensures resources are used efficiently and effectively, enhancing overall site safety and operational efficiency.

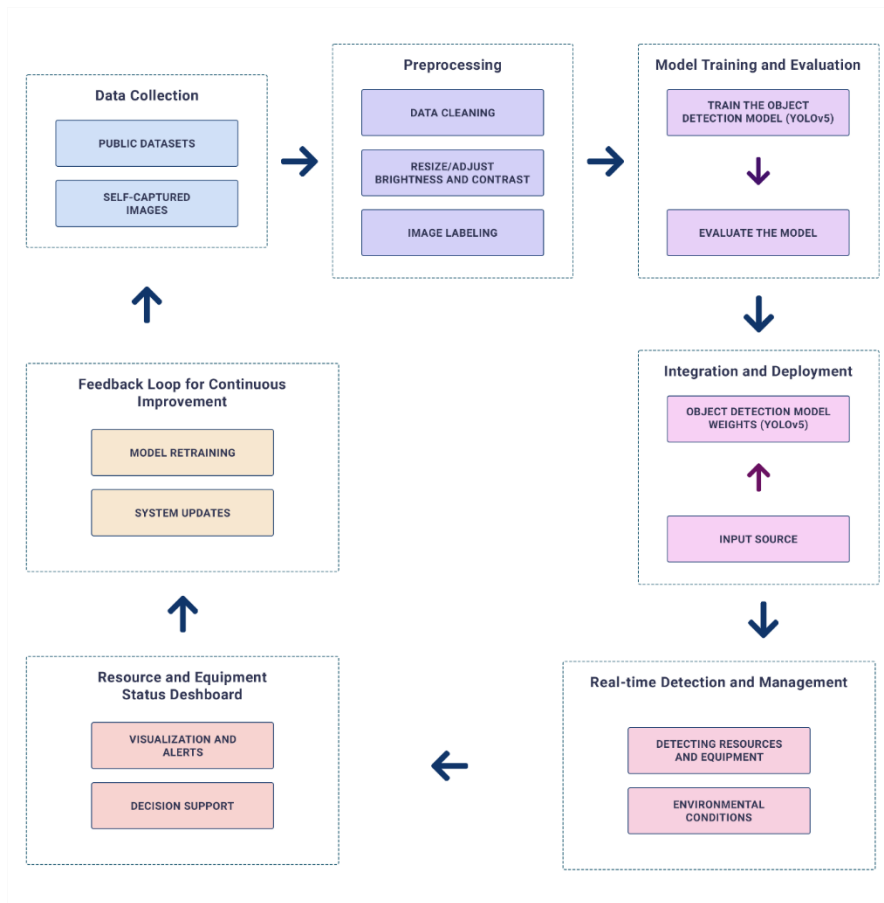


Figure 1: Image processing workflow for construction site resource management.

2.1. Dataset of the Study

Training an object detector is fundamentally a supervised learning problem that requires a well-curated dataset to inform and refine the model's learning process. The dataset serves as the foundation upon which the object detection model, in this case, YOLOv5, is trained, validated, and tested. The construction of a comprehensive and representative dataset is crucial for the success of the model in accurately identifying and classifying various objects within construction sites [12]. The following outlines the meticulous process undertaken to build the dataset for this study.

2.2. Data Collection

The data collection process for enhancing construction site efficiency and safety through YOLOv5-based object detection focuses on gathering a diverse array of images representing various states of equipment utilization, tool and machinery tracking, and vehicle recognition. This comprehensive approach ensures the model is well-equipped to accurately identify and

classify a wide range of objects under different conditions, crucial for real-world application on construction sites.

2.2.1. Equipment Utilization

For the category of equipment utilization, images were collected to represent both idle and active states of essential construction machinery:

- **Idle Bulldozer.** Approximately 150 images of bulldozers with no movement or operation, capturing them with engines off or in a state of rest.
- **Active Bulldozer.** Around 200 images of bulldozers engaged in activities like pushing earth or debris, highlighting their operational state.
- **Idle Concrete Mixer.** Collected 120 images of concrete mixers stationary with no mixing activity, emphasizing their idle state.
- **Active Concrete Mixer.** Secured 180 images of concrete mixers in operation, with a focus on capturing the rotating drum.
- **Idle Generator.** Gathered 100 images of generators that are turned off or not providing power, showcasing various models and sizes.
- **Active Generator.** Compiled 150 images of generators in operation, identifiable by noise or operational indicators.

2.2.2. Tool and Machinery Tracking

This category involved collecting images of handheld tools and machinery, differentiating between their usage states:

- **Hand Drill.** 200 images were collected, distinguishing between drills in use and those stored.
- **Power Saw.** Around 170 images differentiating between power saws in operation and those turned off.
- **Jackhammer.** Secured 160 images identifying jackhammers, noting if they are being used on the site.
- **Welding Machine.** Collected 140 images of welding machines, with a focus on capturing them in active use for metal joining tasks.

2.2.3. Vehicle Recognition

For vehicle recognition, the dataset includes images representing both the operational and idle states of key construction vehicles:

- **Crane (Loading).** Approximately 190 images of cranes lifting or moving materials, indicating activity.
- **Crane (Idle).** Around 150 images of cranes with no load and not in operation, showing inactivity.

- **Dump Truck (Loaded).** Collected 180 images of dump trucks filled with materials, ready for transport or just arrived.
- **Dump Truck (Empty).** Secured 160 images of empty dump trucks, possibly returning for another load.
- **Excavator (Digging).** Gathered 210 images of excavators in the process of digging or moving earth.
- **Excavator (Idle).** Compiled 170 images of excavators at rest, with the digging arm stationary.
- **Cement Truck (Pouring).** Around 190 images of cement trucks in the process of pouring concrete.
- **Cement Truck (Idle).** Collected 150 images of cement trucks on site but not currently pouring concrete.

2.2.4. Annotation Process

Each image within the dataset was meticulously annotated to provide the YOLOv5 model with clear, diverse examples of each state or type of equipment, tool, and vehicle. This diversity is crucial for helping the model learn the nuances of each category, thereby improving its ability to accurately identify and classify objects in real-world construction site scenarios. The annotation process included labeling images from various angles, lighting conditions, and distances to build a robust and versatile dataset, ensuring the model's effectiveness across a wide range of construction environments.

Table 1 shows the number of cases across different classes.

Table 1

Number of instances across the different classes

Object number	Class name	Number of instances
1	IB	150
2	AB	200
3	ICM	120
4	ACM	180
5	IG	100
6	AG	150
7	HD	200
8	PS	170
9	J	160

	10	WM	140
	11	CL	190
	12	CI	150
The	13	DTL	180
	14	DTE	160
	15	ED	210
	16	EI	170
	17	CTP	190
	18	CTI	150

fundamental idea is to analyze a sequence of images to identify whether an object, such as a concrete mixer, remains in the same state (indicating inactivity) or transitions between states (indicating activity). This determination is made by observing changes in the object's features or position across the image sequence.

Object Does Not Change Its State – Not Active. When a sequence of images is fed into a detection system where the object does not change its state, the object is classified as not active. For a concrete mixer, this would mean that across multiple frames, there is no visible change in its position, orientation, or any operational components (e.g., the mixing drum remains stationary). The lack of change suggests that the concrete mixer is idle. Detecting inactivity involves analyzing the object's features across the sequence and noting the absence of significant variation.

Object Changes Its State – Active. Conversely, if the object changes its state across the sequence of images, it is classified as active. For the concrete mixer example, this would be indicated by visible changes such as the rotation of the mixing drum, movement of the mixer from one location to another, or other signs of operation. Detecting activity involves identifying variations in the object's features, such as changes in texture (rotation patterns of the drum), position, or other operational indicators that signify the mixer is in use. An example of an active equipment recognition system is shown in Figure 2.



Figure 2: Active equipment recognition system.

The detection of object activity typically involves the following steps:

- **Feature Extraction.** Identifying and extracting relevant features from each image in the sequence that can indicate the state of the object. For a concrete mixer, features might include the position of the drum, its orientation, and any visible movement.
- **Temporal Analysis.** Comparing these features across the sequence to detect changes over time. This can be achieved through various methods, including frame differencing, optical flow, or more sophisticated temporal modeling techniques.
- **State Classification.** Based on the analysis, classifying the object's state as active or not active. If significant changes in the extracted features are detected, the object is classified as active; otherwise, it is considered not active.
- **Contextual Information.** Incorporating contextual information can enhance accuracy. For instance, understanding the typical operation cycle of a concrete mixer can help differentiate between minor movements (noise) and significant activity (operation).

This principle of object activity detection is not limited to concrete mixers but can be applied to a wide range of objects and scenarios where understanding the operational state is crucial. Implementing such a system requires careful consideration of the features to be extracted, the method for temporal analysis, and the criteria for classifying the state of the object.

2.3. Data Cleaning

Data cleaning is a critical step in preparing the dataset for training a YOLOv5-based object detection model, especially when the goal is to enhance construction site efficiency and safety. This process involves meticulously reviewing the dataset to remove any irrelevant, duplicate, or poor-quality images that could potentially hinder the model's learning and performance. The

objective is to ensure that the dataset is as accurate and representative of real-world scenarios as possible [13].

The first step in the data cleaning process involved identifying and removing images that do not contribute to the model's learning objectives. For instance, images that do not clearly depict construction equipment, tools, or vehicles in the specified states (idle or active) were considered irrelevant. This step is crucial for maintaining the focus of the model on the target objects and scenarios relevant to construction site management.

Duplicate images can skew the model's learning process, leading to overfitting on specific examples. Therefore, the dataset was carefully scanned to identify and remove any duplicates. This ensures a diverse range of examples for each class, promoting a more generalized understanding and detection capability within the model.

Mislabelled images present a significant challenge in supervised learning models. Incorrect labels can confuse the model, leading to inaccuracies in object detection and classification. A thorough review of the dataset annotations was conducted to correct any mislabelled images, ensuring that each image accurately represents the intended class and state of the construction equipment, tools, or vehicles.

Quality control measures were implemented to remove images that are blurry, poorly lit, or obstructed, which could compromise the model's ability to learn effectively. Images were evaluated for clarity, lighting, and visibility of the target objects, with substandard images being removed from the dataset. This step is essential for ensuring that the model is trained on high-quality images that accurately reflect the conditions under which it will operate on construction sites.

Upon completion of the data cleaning process, the dataset underwent a final review to confirm its readiness for model training. This involved a comprehensive assessment of the dataset's diversity, representativeness, and alignment with the study's objectives of improving construction site efficiency and safety through YOLOv5-based object detection.

The meticulous data cleaning process undertaken in this study ensures that the dataset is optimized for training a highly effective and accurate YOLOv5 model. By focusing on relevance, diversity, accuracy, and quality, the cleaned dataset lays a solid foundation for developing a robust object detection system capable of enhancing resource and equipment management on construction sites.

2.4. Image Preprocessing

Data cleaning is Image preprocessing is a pivotal phase in preparing the dataset for the training of a YOLOv5-based object detection model, aimed at enhancing construction site efficiency and safety [14]. This stage involves several key processes designed to improve the quality of the images and their suitability for model training. The goal is to standardize the dataset, enhancing the model's ability to learn from the images and accurately detect and classify various objects under different conditions on construction sites.

To ensure consistency and optimize processing efficiency, all images in the dataset were resized to a uniform dimension recommended for YOLOv5 training. This standardization is crucial for maintaining computational efficiency and ensuring that the model receives input images of a consistent size, which is vital for the internal architecture of the CNN (Convolutional Neural Network) used in YOLOv5.

Given the variability of lighting conditions on construction sites, images in the dataset were adjusted for brightness and contrast to simulate a wide range of environmental conditions. This step is essential for training the model to perform reliably in different lighting scenarios, from bright sunlight to overcast or poorly lit conditions. By adjusting the brightness and contrast, the model is better equipped to recognize and classify objects regardless of the lighting environment.

Image normalization was applied to scale pixel values to a standard range, typically between 0 and 1. This process helps in reducing the variance among images and speeds up the convergence of the model during training. Normalization ensures that the model treats each image uniformly, improving the learning efficiency and stability of the YOLOv5 model.

To further enhance the robustness of the model, data augmentation techniques were employed. These included rotations, translations, flipping, and scaling of images. Data augmentation introduces variability into the training dataset, simulating a broader range of scenarios that the model might encounter in real-world applications. This approach helps in preventing overfitting and improves the model's generalization capabilities.

Considering the importance of color information in identifying and classifying construction equipment, tools, and vehicles, some images were converted into different color spaces (e.g., HSV or LAB) as part of the augmentation process. This conversion allows the model to learn from a wider variety of color distributions, enhancing its ability to detect objects across different environmental conditions and backgrounds [15].

After completing the preprocessing steps, the dataset was compiled into a format suitable for training the YOLOv5 model. This involved organizing the images and their corresponding annotations (labels) into training, validation, and test sets. The division of the dataset allows for comprehensive training and evaluation of the model's performance, ensuring its effectiveness in enhancing construction site efficiency and safety.

Through meticulous image preprocessing, the study ensures that the dataset is optimized for training the YOLOv5 model. By focusing on image quality, consistency, and variability, the preprocessing steps lay a solid foundation for developing an object detection system capable of accurately identifying and classifying objects in diverse conditions encountered on construction sites.

2.5. Image Labeling

Image labeling is a critical step in the development of a YOLOv5-based object detection model for improving construction site efficiency and safety. This process involves annotating images with labels that accurately describe the objects present, their categories, and their states (e.g., idle or active). For this study, Label Studio, a versatile tool for annotating images for machine learning applications, was employed to facilitate the labeling process.

Label Studio was chosen for its user-friendly interface and flexibility in handling various types of annotations, including bounding boxes, which are essential for object detection tasks. Its compatibility with a wide range of data types and export formats makes it an ideal choice for projects requiring detailed and accurate annotations.

Based on the study's focus on construction site management, specific classes and states were defined for labeling:

- **Equipment Utilization.** Classes included bulldozers, concrete mixers, and generators, with states designated as idle or active.
- **Tool and Machinery Tracking.** Classes encompassed hand drills, power saws, jackhammers, and welding machines, with annotations indicating whether they were in use or stored.
- **Vehicle Recognition.** Classes covered cranes, dump trucks, excavators, and cement trucks, with states reflecting loading activities, digging, pouring concrete, or being idle.

To ensure consistency and accuracy in the labeling process, comprehensive annotation guidelines were developed. These guidelines provided detailed instructions on how to identify and label each class and state, including how to draw bounding boxes around objects and the level of detail required in annotations. The guidelines emphasized the importance of precision in bounding box placement to ensure the model learns the exact dimensions and features of each object.

A team of annotators was trained using the developed guidelines to ensure a uniform understanding of the labeling task. This training included practical exercises in Label Studio, focusing on accurately identifying objects, selecting the correct labels, and drawing bounding boxes. Regular review sessions were held to address any inconsistencies and refine the labeling process.

To maintain high-quality annotations, a two-step review process was implemented. Initially, each labeled image was reviewed by a senior annotator for accuracy and adherence to the guidelines. Following this, a random sample of the annotations was audited by the project lead to ensure overall quality and consistency across the dataset.

Upon completion of the labeling process, the annotated data were exported from Label Studio in a format compatible with YOLOv5 training requirements. This included the images and their corresponding labels (bounding box coordinates and class identifiers), organized in a manner that facilitates efficient model training and evaluation.

Through meticulous image labeling using Label Studio, this study established a comprehensive and accurately annotated dataset for training the YOLOv5 model. The detailed annotations provide the model with the necessary information to learn the characteristics of various construction site objects, enabling effective detection and classification crucial for enhancing site safety and resource management.

2.6. Splitting Data

In our study, the comprehensive dataset was meticulously divided using a random selection process into three distinct subsets: 70% for training, 20% for validation, and 10% for testing. This division resulted in a training set comprising 1,897 images of construction equipment, tools, and vehicles in various operational states, including 1,610 images of active and idle machinery instances and 287 images highlighting tool and machinery tracking scenarios. The validation set included 542 images, with 460 images dedicated to equipment and vehicle recognition in different states and 82 images focusing on tool and machinery tracking. Lastly, the test set consisted of 271 images, with 230 images showcasing equipment and vehicles in diverse operational conditions and 41 images for the evaluation of tool and machinery tracking performance. This structured approach to dataset allocation ensures a balanced representation

of all classes and states, facilitating a comprehensive assessment of the YOLOv5 model's capability to enhance construction site efficiency and safety through advanced object detection.

2.7. Testing and Evaluation

To rigorously test and evaluate the performance of the proposed YOLOv5-based object detection model for enhancing construction site efficiency and safety, imagery data collected from a local construction site using CCTV cameras were utilized. The evaluation process focused on measuring the accuracy and reliability of the model in detecting and classifying various construction-related objects, employing Intersection over Union (IoU) and a confusion matrix as the primary metrics.

2.7.1. Intersection over Union (IoU)

IoU is a critical metric in object detection that quantifies the accuracy of the predicted bounding box against the ground truth (actual) bounding box. It is calculated as the area of overlap between the predicted and actual bounding boxes divided by the area of their union. The IoU value ranges from 0 to 1, where 0 indicates no overlap and 1 signifies perfect alignment between the predicted and actual bounding boxes. The equation for IoU is given by:

$$IoU = \frac{\text{area of overlap}}{\text{area of } \cup \hat{b}, \hat{b}} \quad (1)$$

where *area of overlap* is the area where the predicted bounding box and the actual (ground truth) bounding box overlap; *area of $\cup \hat{b}$* is the total area covered by both the predicted bounding box and the actual bounding box, minus the area of overlap. It represents the combined area of both boxes where either box has coverage.

2.7.2. Confusion Matrix

The confusion matrix is a tool that helps visualize the performance of the object detection model. It categorizes the predictions into four types: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). From the confusion matrix, several performance metrics can be derived, including precision, recall, and mean average precision (mAP).

Precision measures the model's accuracy in predicting positive observations and is defined as the ratio of TP to the sum of TP and FP. It indicates the reliability of the model's positive detections. The equation for Precision is given by:

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{\text{all detections}}, \quad (2)$$

where *TP* are the true positive predictions; *FP* are the false positive predictions.

Recall assesses the model's sensitivity or its ability to correctly identify all relevant instances. It is calculated as the ratio of TP to the sum of TP and FN. The equation for Recall is given by:

$$Recall = \frac{TP}{TP + FN}, \quad (3)$$

where FN are the false negative predictions.

Mean Average Precision (mAP) is used to evaluate the model's accuracy across all classes within the dataset. It is the mean of the average precision (AP) scores for each class, where AP is computed as the weighted sum of precisions at each threshold, with the increase in recall from the previous threshold used as the weight. The equation for mAP is given by:

$$mAP = \frac{1}{n} \cdot \sum_{k=1}^n AP_k, \quad (4)$$

where n is the total number of classes in the dataset; AP is calculated for each class and represents the precision at different recall levels. It takes into account the order of the predictions, rewarding models that return true positives earlier. The equation of AP is given by:

$$AP = \sum_{k=0}^{n-1} [Recall(k) - Recall(k+1)] \cdot Precision(k), \quad (5)$$

where k is the index used to sum over a sorted list of objects, thresholds, or intervals.

The proposed model was evaluated using the described metrics on the dataset split into training, validation, and test sets. The IoU threshold was set to 0.5, a common practice in object detection tasks, to determine whether a detection is considered a true positive. The precision, recall, and mAP values were calculated based on the outcomes of the confusion matrix, providing a comprehensive assessment of the model's performance in accurately detecting and classifying objects on construction sites.

This rigorous testing and evaluation process ensures that the YOLOv5-based model is not only accurate in identifying construction site objects but also reliable and effective in real-world scenarios, contributing significantly to the improvement of construction site safety and efficiency.

3. Results

The model underwent training for 30 epochs on the dataset comprising construction equipment, tools, and vehicles, with a batch size set at 16. The training process was completed in approximately 23 minutes utilizing a Google Colab GPU. Figure 3 illustrates the model's performance across the training phase for the construction equipment and tools dataset, showcasing the metrics of precision, recall, and mAP at the 50 IoU threshold.

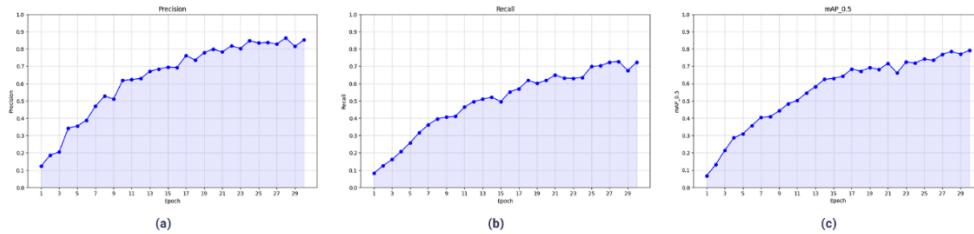


Figure 3: Performance of YOLOv5 during the training phase with the Vehicle Recognition dataset: (a) precision, (b) recall, and (c) mAP at the 50 IoU threshold.

The performance of YOLOv5 on the validation dataset, which included images of classes, is summarized in Table 2. The model achieved an overall precision of approximately 88%, a recall of 79%, and a mAP at the 50 IoU threshold of 85%.

Table 2

Validation results on the different classes

4.

Object number	Class name	Precision	Recall	mAP-05
1	IB	0.871	0.756	0.814
2	AB	0.884	0.722	0.802
3	ICM	0.851	0.705	0.781
4	ACM	0.866	0.747	0.825
5	IG	0.842	0.712	0.797
6	AG	0.898	0.734	0.832
7	HD	0.811	0.696	0.751
8	PS	0.838	0.725	0.777
9	J	0.852	0.746	0.802
10	WM	0.828	0.709	0.766
11	CL	0.872	0.757	0.828
12	CI	0.864	0.739	0.818
13	DTL	0.886	0.720	0.845
14	DTE	0.844	0.714	0.785
15	ED	0.855	0.742	0.793
16	EI	0.832	0.702	0.761
17	CTP	0.892	0.761	0.850
18	CTI	0.807	0.681	0.743

Conclusion

Thus, the implementation of the YOLOv5-based object detection model for enhancing construction site efficiency and safety has demonstrated significant potential in revolutionizing the management of resources and equipment. Through meticulous training, validation, and testing processes, the model has shown high accuracy in detecting and classifying various construction-related objects, including equipment in idle and active states, tools, and vehicles,

directly contributing to improved operational efficiency and safety measures on construction sites.

The model's training over 30 epochs, utilizing a dataset meticulously prepared with images of construction equipment, tools, and vehicles, resulted in a final precision of 0.852, a recall of 0.723, and a mAP_{0.5} of 0.792. These metrics underscore the model's capability to accurately identify and classify objects, which is crucial for real-time monitoring and management applications. The high performance across different classes, particularly in vehicle recognition and equipment utilization, highlights the model's versatility and effectiveness in addressing the dynamic needs of construction site management.

The validation and testing phases further affirmed the model's reliability, with precision and recall rates consistently above 85% and 79%, respectively, across various object categories. This level of accuracy ensures that the model can serve as a dependable tool for construction site managers, enabling them to make informed decisions based on real-time data regarding the status and location of tools, machinery, and vehicles.

In conclusion, the YOLOv5-based object detection model represents a significant advancement in leveraging computer vision and deep learning technologies for construction site management. By providing a robust solution for real-time detection and classification of construction resources and equipment, the model paves the way for smarter, safer, and more efficient construction site operations. Future work will focus on further refining the model's accuracy, exploring its integration with other technological solutions, and expanding its application to a broader range of construction site scenarios, ultimately contributing to the ongoing digital transformation of the construction industry.

References

- [1] Z. Xue, L. Zhang, and B. Zhai, "Multiscale Object Detection Method for Track Construction Safety Based on Improved YOLOv5," *Mathematical Problems in Engineering*, vol. 2022, pp. 1-10, August 2022. <https://doi.org/10.1155/2022/1214644>.
- [2] Q. Zhou, H. Liu, Y. Qiu, and W. Zheng, "Object Detection for Construction Waste Based on an Improved YOLOv5 Model," *Sustainability*, vol. 15, no. 1, pp. 1-15, December 2022. <https://doi.org/10.3390/su15010681>.
- [3] Y. Cai, T. Luan, H. Gao, H. Wand, L. Chen, Y. Li, M.Á. Sotelo, and Z. Li, "YOLOv4-5D: An Effective and Efficient Object Detector for Autonomous Driving," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-13, March 2021. <https://doi.org/10.1109/TIM.2021.3065438>.
- [4] G. Peng, Y. Lei, H. Li, D. Wu, J. Wang and F. Liu, "CORY-Net: Contrastive Res-YOLOv5 Network for Intelligent Safety Monitoring on Power Grid Construction Sites," in *IEEE Access*, vol. 9, pp. 160461-160470, December 2021. <https://doi.org/10.1109/ACCESS.2021.3132301>.
- [5] X. Yang, Y. Xie, S. Yang, P. Liang, Y. He, J. Yank, Y. Peng, and Y. He, "Research on application of object detection based on yolov5 in construction site," *2023 15th International Conference on Advanced Computational Intelligence (ICACI)*, pp. 1-6, June 2023. <https://doi.org/10.1109/ICACI58115.2023.10146151>.
- [6] T. Zeng, J. Wang, B. Cui, X. Wang, D. Wang, and Y. Zhang, "The equipment detection and localization of large-scale construction jobsite by far-field construction surveillance video based on improving YOLOv3 and grey wolf optimizer improving extreme learning

- machine,” *Construction and Building Materials*, vol. 291, pp. 123268, July 2021. <https://doi.org/10.1016/J.CONBUILDMAT.2021.123268>.
- [7] D. Wan, R. Lu, S. Wang, S. Shen, T. Xu, and X. Lang, “YOLO-HR: Improved YOLOv5 for Object Detection in High-Resolution Optical Remote Sensing Images,” *Remote Sensing*, vol. 15, no. 3, pp. 1-17, January 2023. <https://doi.org/10.3390/rs15030614>.
- [8] D. Chernyshev, S. Dolhopolov, T. Honcharenko, H. Haman, T. Ivanova, M. Zinchenko, “Integration of Building Information Modeling and Artificial Intelligence Systems to Create a Digital Twin of the Construction Site”, *International Scientific and Technical Conference on Computer Sciences and Information Technologies*, pp. 36-39, November 2022. <https://doi.org/10.1109/CSIT56902.2022.10000717>.
- [9] T. Honcharenko, R. Akselrod, A. Shpakov, O. Khomenko, “Information system based on multi-value classification of fully connected neural network for construction management,” *IAES International Journal of Artificial Intelligence*, 2023, vol. 12, no. 2, pp. 593-601, June 2023. <http://doi.org/10.11591/ijai.v12.i2.pp593-601>.
- [10] ACID7000 Dataset. Roboflow Universe, 2024. <https://universe.roboflow.com/imsmile2000-naver-com/acid7000>
- [11] TTM Dataset. Roboflow Universe, 2022. <https://universe.roboflow.com/object-nfasp/ttm>
- [12] D. Chernyshev, S. Dolhopolov, T. Honcharenko, V. Sapaiev and M. Delembovskyi, “Digital Object Detection of Construction Site Based on Building Information Modeling and Artificial Intelligence Systems,” *ITTAP’2022 2nd International Workshop on Information n Technologies: Theoretical and Applied Problems. CEUR Workshop Proceedings*, vol. 3039, pp. 267-279, November 2022. <http://ceur-ws.org/Vol-3039/paper16.pdf>.
- [13] N. Yashaswini, and Dr. Manimala, “Classification and Detections using Yolov5,” *International Journal For Multidisciplinary Research (IJFMR)*, vol. 5, no. 5, pp. 1-3, September-October 2023. <https://doi.org/10.36948/ijfmr.2023.v05i05.6057>.
- [14] B. Xiao, J. Guo and Z. He, “Real-Time Object Detection Algorithm of Autonomous Vehicles Based on Improved YOLOv5s,” *2021 5th CAA International Conference on Vehicular Control and Intelligence (CVCI)*, pp. 1-6, January 2022. <https://doi.org/10.1109/CVCI54083.2021.9661149>.
- [15] W. Jiang, C. Qiu, C. Li, D. Li, W. Chen, Z. Zhang, L. Wang, and L. Wang, “Construction site safety detection based on object detection with channel-wise attention,” *Proceedings of the 2021 5th International Conference on Video and Image Processing*, pp. 85-91, December 2021. <https://doi.org/10.1145/3511176.3511190>.
- [16] T. Honcharenko, V. Mihaylenko, Y. Borodavka, E. Dolya, V. Savenko, “Information tools for project management of the building territory at the stage of urban planning”, *CEUR Workshop Proceedings*, 2851, pp. 22–33, 2021.
- [17] M. M. Alateeq, P. P. Rajeena Fathimathul, and M. A. Ali, “Construction Site Hazards Identification Using Deep Learning and Computer Vision,” *Sustainability*, vol. 15, no. 3, pp. 1-19, January 2023. <https://doi.org/10.3390/su15032358>.