Optimizing construction site management through YOLOv5-based object detection: a comprehensive analysis of resource utilization and safety enhancement

Serhii Dolhopolov[1,†], Tetyana Honcharenko[1,†], Igor Achkasov[1,†] and Vladyslav Hots[1,†]

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

Abstract

This study introduces a YOLOv5-based object detection system for optimizing construction site management, addressing critical challenges in resource utilization and safety. We developed a custom YOLOv5 model to identify and track construction resources, equipment, and vehicles in real-time using CCTV footage. The model was trained on a dataset of 1,897 images over 30 epochs, achieving a final precision of 0.852, recall of 0.723, and mean Average Precision (mAP_0.5) of 0.792. Performance evaluation using Intersection over Union (IoU) and confusion matrix analyses demonstrated high accuracy across different object categories, with an overall precision of 88%, recall of 79%, and mAP at the 50 IoU threshold of 85% on the validation dataset. These results indicate the model's robust capability in accurately detecting and classifying various construction-related objects. The proposed system offers a comprehensive framework for integrating AI-driven object detection into construction management, potentially enhancing operational efficiency through optimized resource allocation and improving site safety via real-time monitoring. Future research will focus on refining the model's performance in diverse environmental conditions and exploring its integration with other emerging construction technologies.

Keywords

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

1. Introduction

The construction industry is undergoing a significant transformation driven by AI and ML technologies, particularly YOLOv5-based object detection, to address challenges in resource management and safety. This advanced technology offers real-time detection and classification of construction assets, enhancing operational efficiency and safety protocols [1,2].

Recent studies have demonstrated YOLOv5's efficacy in construction settings. Xue et al. developed an improved YOLOv5 algorithm for track construction safety [1], while Zhou et al. proposed a YOLOv5 model for sorting construction waste [2]. Cai et al. showcased a YOLOv4-based framework applicable to construction site management [3], and Peng et al. introduced CORY-Net, a YOLOv5 variant for power grid construction site monitoring [4,8].

YOLOv5's applications extend beyond basic object detection to analyzing equipment usage patterns, real-time monitoring of tool locations, and identifying potential safety hazards. Yang et al. demonstrated its effectiveness in monitoring safety protocol compliance [5,9], while Zeng et al.

DTESI 2024: 9th International Conference on Digital Technologies in Education, Science and Industry, October 16–17, 2024, Almaty, Kazakhstan

* Corresponding author.

† These authors contributed equally.

dolhopolov@icloud.com (S. Dolhopolov); goncharenko.ta@knuba.edu.ua (T. Honcharenko); achckasov.i@ukr.net (I. Achkasov); gots.vv@knuba.edu.ua (V. Hots)

0000-0001-9418-0943 (S. Dolhopolov); 0000-0003-2577-6916 (T. Honcharenko); 0000-0002-7049-0530 (I. Achkasov); 0000-0003-4384-4011 (V. Hots)

highlighted the importance of adapting these models to the unique challenges of construction sites [6].

The integration of YOLOv5 into construction management systems represents a paradigm shift towards data-driven decision-making and operational efficiency. Wan et al.'s work on YOLOv5 for object detection in high-resolution images underscores the model's robustness across various conditions [7], a crucial attribute for the dynamic environment of construction sites.

The purpose of this research can be summarized as follows:

- Evaluate YOLOv5's effectiveness in improving construction site efficiency.
- Assess YOLOv5's impact on construction site safety.
- Explore customization of YOLOv5 models for specific construction environments.
- Investigate integration with other technologies (e.g., drones, IoT).
- Identify challenges and limitations in deploying YOLOv5-based systems.
- Provide recommendations for future research and development.

By addressing these objectives, this study seeks to contribute to the knowledge base on advanced object detection technologies in construction site management, paving the way for 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.

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.

2.1. Dataset of the study

The success of the YOLOv5 object detection model for construction site management relies heavily on a comprehensive and representative dataset [11]. This study employed a meticulous data collection process to ensure the model's effectiveness across various construction site scenarios.

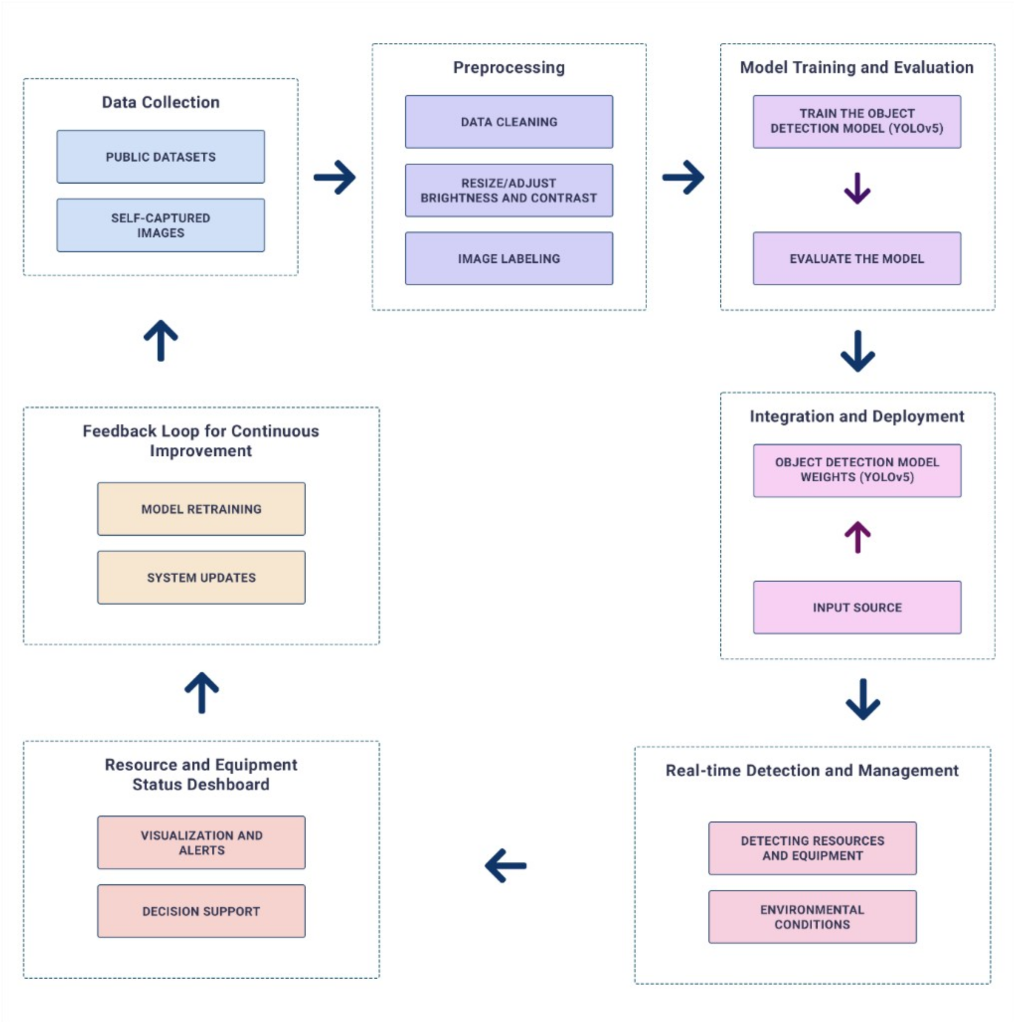


Figure 1: Proposed framework of the system.

2.2. Data collection

The dataset encompasses three main categories:

1. Equipment Utilization. Images of bulldozers, concrete mixers, and generators in both idle and active states.
2. Tool and Machinery Tracking. Images of hand drills, power saws, jackhammers, and welding machines in various usage states.
3. Vehicle Recognition. Images of cranes, dump trucks, excavators, and cement trucks in operational and idle states.

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
13	DTL	180
14	DTE	160
15	ED	210
16	EI	170
17	CTP	190
18	CTI	150

Table 1 shows the number of cases across different classes.

2.3. Activity detection methodology

The 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 relevant features indicative of the object's state.
- Temporal Analysis. Comparing features across image sequences to detect changes over time.
- State Classification. Classifying objects as active or inactive based on detected feature changes.
- Contextual Information Integration. Enhancing accuracy by incorporating knowledge of typical operational cycles.

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.

2.4. Data cleaning

Data cleaning is crucial for preparing an optimal dataset for training the YOLOv5-based object detection model [12]. The process involved:

1. Removing irrelevant images not depicting construction equipment, tools, or vehicles in specified states.
2. Eliminating duplicate images to prevent overfitting.
3. Correcting mislabelled images to ensure accurate representation of classes and states.
4. Implementing quality control measures to remove blurry, poorly lit, or obstructed images.

This meticulous process ensures a dataset optimized for training an effective and accurate YOLOv5 model, focusing on relevance, diversity, accuracy, and quality.

2.5. Image preprocessing

Image preprocessing is pivotal in enhancing the dataset's suitability for model training [13]. Key steps included:

1. Resizing all images to a uniform dimension for YOLOv5 training.
2. Adjusting brightness and contrast to simulate various lighting conditions.
3. Applying image normalization to scale pixel values.
4. Employing data augmentation techniques (rotations, translations, flipping, scaling).
5. Converting some images to different color spaces (e.g., HSV, LAB) to enhance object detection capabilities [14-15].

The preprocessed dataset was then organized into training, validation, and test sets, ensuring comprehensive model evaluation.

2.6. Splitting data

The dataset was divided into three subsets:

- Training set (70%): 1,897 images (1,610 machinery, 287 tool tracking)
- Validation set (20%): 542 images (460 equipment/vehicle, 82 tool tracking)
- Test set (10%): 271 images (230 equipment/vehicle, 41 tool tracking)

This structured approach ensures balanced representation across all classes and states.

2.7. Testing and evaluation

The model's performance was evaluated using CCTV imagery from a local construction site, focusing on accuracy and reliability in object detection and classification.

2.7.1. Intersection over Union (IoU)

IoU quantifies the accuracy of predicted bounding boxes against ground truth. The equation for IoU is given by:

$$IoU=\frac{area\,of\,overlap}{area\,of \cup ¿,¿} \tag{1}$$

where $area\,of\,overlap$ is the area where the predicted bounding box and the actual (ground truth) bounding box overlap; $area\,of \cup ¿$ 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

Precision measures the model's accuracy in predicting positive observations. The equation for Precision is given by:

$$Precision=\frac{TP}{TP+FP}=\frac{TP}{all\,detections}, \tag{2}$$

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

Recall assesses the model's sensitivity. 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) evaluates the model's accuracy across all classes. 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:

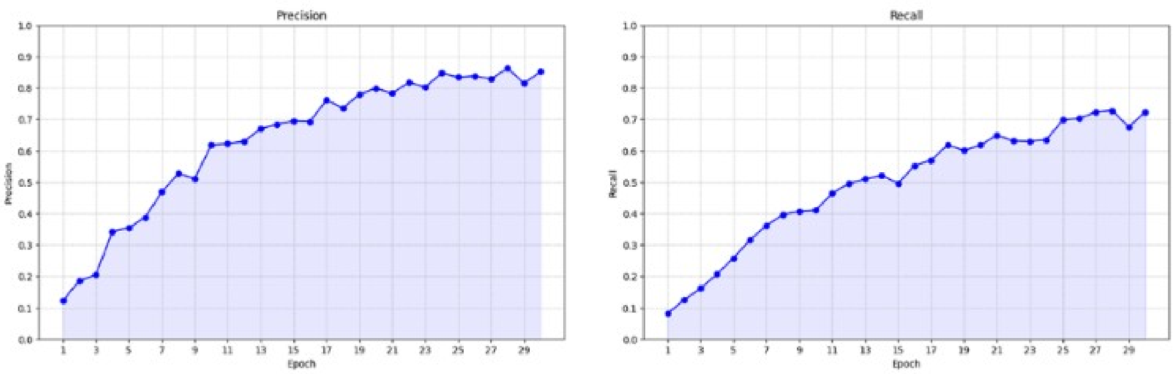
$$AP = \sum_{k=0}^{n-1} [Recall(k) - Recall(k+1)] \cdot Precision(k), \quad (5)$$

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

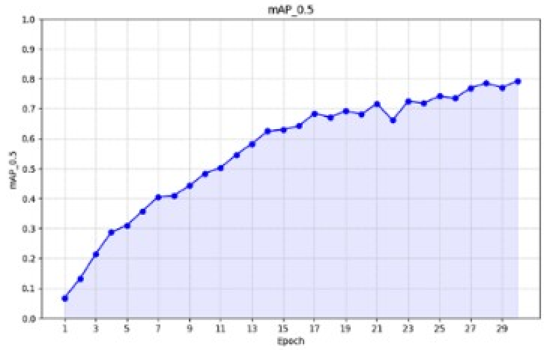
The model was evaluated using these metrics on the dataset split into training, validation, and test sets, with an IoU threshold of 0.5. This comprehensive assessment ensures the model's accuracy and reliability in real-world construction site scenarios, contributing to improved safety and efficiency.

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.



(a) (b)



(c)

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.

Table 2
Validation results on the different classes

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

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%.

4. 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.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

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] S. Dolhopolov, T. Honcharenko, O. Terentyev, K. Predun, and A. Rosynskyi, "Information system of multi-stage analysis of the building of object models on a construction site," IOP Conference Series: Earth and Environmental Science, vol. 1254, pp. 1-11, May 2023. https://doi.org/10.1088/1755-1315/1254/1/012075.

[9] T. Honcharenko, V. Mihaylenko, Y. Borodavka, E. Dolya, and 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.

[10] S. Dolhopolov, T. Honcharenko, V. Savenko, O. Balina, I. Bezklubenko, and T. Liashchenko, "Construction Site Modeling Objects Using Artificial Intelligence and BIM Technology: A Multi-Stage Approach," 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), pp. 174-179, May 2023. https://doi.org/10.1109/SIST58284.2023.10223543.

[11] 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.

[12] P. Lizunov, P., A. Biloshchytsky, A., Kuchansky, Y., Andrashko, and S., Biloshchytska, "The use of probabilistic latent semantic analysis to identify scientific subject spaces and to evaluate the completeness of covering the results of dissertation studies," Eastern-European Journal of Enterprise Technologies, vol. 4, no. 4-106, pp. 21-28, 2020. https://doi.org/10.15587/1729-4061.2020.209886.

[13] 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.

[14] 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.

[15] A. Biloshchytskyi, A., Kuchansky, Y., Andrashko, and O. Bielova, "Learning space conceptual model for computing games developers," International Scientific and Technical Conference on Computer Sciences and Information Technologies, vol. 1, art. no. 8526719, pp. 97-102, September 2018. https://doi.org/10.1109/STC-CSIT.2018.8526719.

[16] G. Ryzhakova, O. Malykhina, V. Pokolenko, O. Rubtsova, O. Homenko, and I. Nesterenko, "Construction Project Management with Digital Twin Information System", International Journal of Emerging Technology and Advanced Engineering, 12 (10), pp. 19-28, 2022. https://doi.org/10.46338/ijetae1022_03.