Joint CSCE Construction Specialty & CRC Conference 2025
Conférence conjointe spécialisée en construction de la SCGC et CRC-2025

Montreal, Quebec
July 28-31, 2025 / *28-31 juillet 2025*

Automated Heavy Machinery Recognition Using state-of-the-art deep learning methods

Ahmed Assad[1*], Yousef Albkoor[1], Awrad Abdullah[1], Shaikhah AlAteeqi[1] and Ahmed Bouferguene[2]

[1] Civil Engineering Department, Australian University, Kuwait
[2] Department of Civil & Environmental Engineering, University of Alberta, Canada
[*] a.assad@au.edu.kw

ABSTRACT: Construction projects, particularly large-scale ones, significantly depend on the effective management and utilization of heavy equipment. Nevertheless, the classification of heavy equipment has traditionally been a manual process which is labor-intensive, time-consuming and may pose some safety risks. To overcome these challenges, an efficient yet precise methodology needs to be developed to automatically identify and classify heavy equipment in construction projects. Most of the research in this domain has concentrated on the classification of a limited set of equipment classes with little focus on heavy construction equipment. Hence, this study utilizes a robust algorithm, you only look once (YOLO), to classify heavy construction machinery into 15 different types. A comprehensive dataset of equipment images was utilized to train, validate and test the develop classifier. The model achieved an overall accuracy of 94%. It was able to accurately classify most of the heavy equipment classes with an Area Under the Curve (AUC) ranging between 0.97 and 1.0. In comparison to benchmark CNN algorithms like ResNet101 and AlexNet, YOLO models enhanced accuracy by around 20%. In addition, YOLOv11 reduced the inference time by 20% compared to YOLOv8 with a less than 1% decrease in accuracy. This study highlights the potential of advanced deep learning models for real-time classification of heavy construction equipment. By offering a practical and automated solution, these models can significantly enhance efficiency, safety, and maintenance protocols on construction sites.

1. INTRODUCTION

Construction industry is one of the most important sectors of the economic development of any nation. However, it is still one of the least digitized sectors globally (Poinet 2020). In construction, effective project management shall actively seek to maximize the utilization of various available resources including labor, materials, and equipment. As construction projects are heavily equipment reliant, effective equipment management can contribute to ensuring successful completion of these projects within time and budget constraint (Nath and Behzadan 2020). Inefficient utilization patterns and frequent equipment idling may lead to higher emissions and increased project costs. In fact, effective management and utilization of heavy equipment plays a critical role in determining the efficiency, safety, and sustainability of construction projects.

The classification of heavy equipment in construction projects offers an essential first step toward achieving optimal equipment management. Traditional methods of equipment monitoring and classification often rely on manual observation and record-keeping. Such manual methods of observing and analyzing may yield misleading results due to the inherent subjectivity that stems from a heavy reliance on the skills and experience of inspectors (Sherafat et al. 2020). Moreover, conventional manual methods are labor-intensive, time-consuming, and prone to errors. In addition, those methods can be hazardous in some situations. These issues typically lead to increased project costs, operational delays, and safety risks (Sherafat et al. 2020). Innovative solutions to enhance equipment monitoring and classification within the construction industry are highly needed.

Recent advancements in artificial intelligence (AI) and computer vision can assist in developing novel equipment management techniques. Deep learning algorithms have demonstrated outstanding efficacy in object detection and classification across multiple domains. Successful implementation of such models was reported in applications related to equipment failures, optimizing maintenance schedules, and monitoring construction progress (Elghaish et al. 2022). However, utilizing deep learning models to identify heavy machinery, remains insufficiently investigated (Soltani et al. 2016). Only few research efforts attempted to employ deep learning algorithms to classify heavy construction equipment. Such efforts usually focus on a limited number of equipment types overlooking other equipment types that are commonly used mega construction projects (Soltani et al. 2016, Yamany et al. 2024). In addition, these models generally employ specific machine learning techniques that fail to address the dynamic and unstructured environment of construction sites, hence decreasing their accuracy (Yamany et al. 2024). Reliable yet efficient models to identify and classify heavy equipment are thus needed to facilitate automated progress reporting, productivity estimation, and resource allocation.

2. LITERATURE REVIEW

As previously mentioned, identifying and classifying heavy construction equipment is crucial for enabling enhanced resource allocation, improved safety protocols, and better progress monitoring (Elshaboury et al. 2024). As a result, several researchers tried to develop models for the detection and classification of heavy equipment in construction projects. Many of those efforts were based on traditional classification techniques such as artificial neural networks (ANNs), support vector machines (SVMs), and k-nearest neighbors (KNN) (Anirudh et al. 2023, Elshaboury et al. 2024). These algorithms rely heavily on manual feature extraction, requiring experts to define the features for the learning process. This dependence impacts both the model's accuracy and its generalizability. (Akinosho et al. 2020).

The new advancements in deep learning have revolutionized the domain of detection and classification. Convolutional Neural Networks (CNNs) are deep learning algorithms capable of autonomously extracting relevant features from raw image data, eliminating the need for manual feature extraction (Xiao et al. 2021). Over the last decade, the use of DL in construction has significantly increased. There are four primary uses of deep learning in construction management: automating progress monitoring and productivity estimation, automating safety alerts for workers, managing construction equipment, and automated data collection (Elghaish et al. 2022 and Al-Khiami and ElHadad 2024). For example, Wang et al. (2022) developed model for progress monitoring in construction sites, achieving an accuracy of 0.926. Xiao and Kang (2020) demonstrated the capacity of DL to enhance equipment utilization and operational efficiency.

Nonetheless, there is little emphasis in literature on the classification of heavy machinery utilized in construction projects. For example, Hou et al. (2022) introduced an improved YOLOv4 model for multi-object detection in construction machinery. The authors which combined K-means clustering and focus loss function to enhance the detection accuracy. Zeng et al. (2021) proposed construction equipment recognition and localization framework based on an improved YOLOv3 and a Grey Wolf Optimizer-enhanced Extreme Learning Machine (GWO-ELM). The authors reported a mean average precision of 0.758 in detecting six classes of construction equipment. Similarly, Arabi et al. (2020) developed a deep learning model for identifying six categories of construction equipment utilized in highway construction. Earlier, Soltani et al. (2017) investigated the effect of data sources, real or synthetic, on the classification accuracy of various classification algorithms such as SVM and AlexnNet deep learning. The authors considered three types of equipment: loaders, excavators, and trucks. AlexNet average accuracy was improved to a value of 83% when coupled with SVM. More recently, Yamany et al. (2024) exploited CNN to classify 12 classes of heavy equipment. Despite the bigger number of considered classes, the authors were only able to achieve 80%

accuracy. Some authors provided datasets for heavy equipment identification and benchmarking. For example, Xiao and Kang (2021) presented the Alberta Construction Image Dataset consisting of 10,000 images across 10 equipment types, aimed at training object detection algorithms. Similarly, Xuehui et al. (2021) introduced the MOCS dataset for moving object detection in construction, with over 41,000 images.

Consulting previous literature reveals that early models for classifying heavy construction equipment relied heavily on expert-defined inputs, which limited their accuracy and generalizability. Other models that employed DL algorithms only investigated few numbers of equipment types and/or reported low accuracy. Most of the developed models utilized specific types of CNNs with no effort to employ more recent algorithms. This study addresses these gaps by developing a novel model to classify and identify 15 distinct classes of heavy equipment in construction sites utilizing the most recent versions of YOLO algorithm. The proposed model aims to offer an automated real-time classification of a wide array of heavy construction equipment. This model reduces dependency on manual processes, minimizes errors, and supports smart construction management initiatives.

3. METHODOLOGY

Figure 1 illustrates the methodology employed in this research to classify heavy construction equipment. After conducting literature review, image data for heavy construction equipment were gathered from publicly available online sources and construction sites in Kuwait. Data was then preprocessed before being used to create the main classifiers utilizing YOLOv8 and YOLOv11 algorithms. In addition, multiple classifiers were built using different versions of ResNet and AlexNet algorithms for comparison purposes. Final classification metrics are then reported.

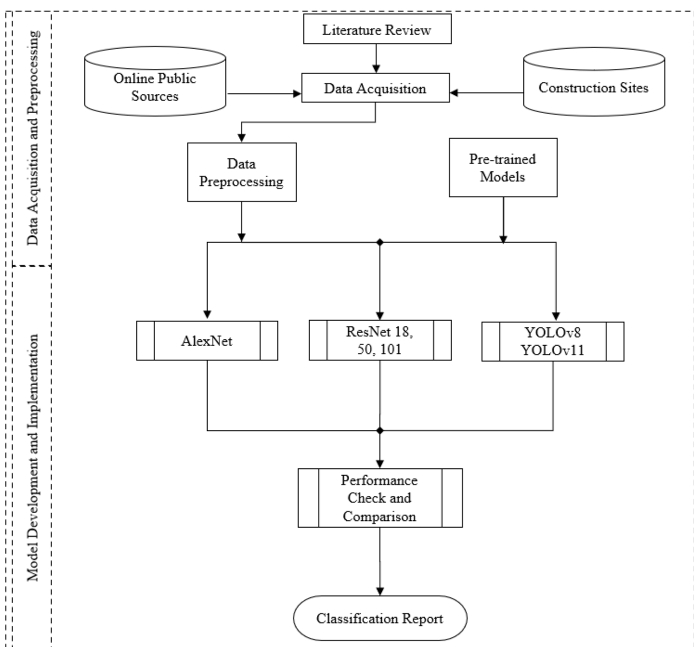

Figure 1. Research methodology

3.1 Overview of Yolo model

YOLO is a state-of-the-art artificial intelligence framework that supports a wide range of computer vision tasks across various domains (Redmon 2016). Different from conventional CNNs, YOLO algorithm processes an entire image in one pass through the network, making it highly efficient for real-time tasks (Redmon 2016). This is useful in scenarios where classification is part of a dynamic system, such as drones. YOLO provides an end-to-end deep CNN that can make predictions about bounding boxes and class probabilities at once eliminating the need for multi-stage computations (Redmon 2016). Its biggest advantage is the fast inference speed that allows real-time image processing. Since it was first introduced in 2015, multiple versions of YOLO have been developed with varying architectures and computational protocols to improve its speed, accuracy, and range of application. While the first version of YOLO was proposed for object detection tasks, more recent versions have been proposed for a wide range of tasks including classification, pose estimation, and instance segmentation (Hussain 2023). Different from earlier approaches, we utilize the most recent YOLO versions, YOLOv8 and YOLOv11, which represent a significant evolution in architecture and performance compared to older models like YOLOv3 and YOLOv4 used in previous studies. YOLOv8 introduces an anchor-free design with decoupled heads and enhanced feature fusion (e.g., C2f and SPPF modules), while YOLOv11 further incorporates the C3K2 and C2PSA modules for efficient multi-scale feature representation and improved spatial attention. These innovations support higher accuracy and faster inference, making them highly suitable for real-time applications in construction environments (Jocher et al. 2023, Jocher and Qiu 2024). A brief description of both algorithms is presented below.

In YOLOv8, the architecture is divided into two main parts: backbone and head. The Backbone is the deep learning architecture that acts as a feature extractor. Features acquired from various backbone layers are combined and fed to the head component where predictions take place. The head in YOLOv8 is decoupled so it separates the classification and regression tasks which increases the model performance (Jocher et al. 2023). Several blocks that perform the end-to-end classification in Yolov8. Frist, there is a convolutional block consisting of a 2D convolutional layer, a 2D batch normalization and a single activation function (SILU). Next is the C2f block where the resulting feature maps will be split into two portions. One portion goes to the bottleneck block, whereas the other goes directly into the concatenate block. Many bottleneck blocks can be implemented in this C2f block representing a sequence of convolutional blocks, some of which without shortcuts. The following block is spatial pyramid pooling - fast SPPF which is a modification of an earlier version of with a higher speed. The aim of the SPPF is to generate a fixed feature representation of objects of various sizes in an image without resizing or suffering spatial information loss. Finally, the detection happens in the detection block. Different from the previous versions, YOLOv8 is an anchor free model where predictions happen in the grid cell (Jocher et al. 2023).

YOLOv11 architecture is similar to this of YOLOv8 with an intermediate processing stage, neck, added between backbone and head components to aggregate and enhance feature representations across different scales. The backbone is made of numerous convolution layers that extract distinct features at various resolution levels. These layers are linked through multiple C3K2 blocks that replace the C2f block in the previous version and employ two smaller convolutions instead of one large convolution, reducing processing time. The neck consists of SPPF that performs fixed representation of objects regardless of their sizes without compromising spatial information. There is also a newly introduced cross-stage partial with spatial attention (C2PSA) block. This block helps the model learn global relationships between pixels or features at different positions to focus more effectively on important regions hence attaining better spatial representation. Next there is an upsampling layer which increases the feature map resolution of the C2PSA to match the feature map using the nearest neighbor upsampling method. The unsampled feature map will be combined with the features from this C3K2 block using concatenate block. The neck component is then connected to the head layer via three C3K2 blocks with each exclusively used to predict certain object sizes (small, medium, and large). Introducing the novel C3K2 and S2PSA blocks enhances the performance for more complex tasks with much fewer parameters (Jocher and Qiu 2024). Similar to YOLOv8, five variants of YOLOv11 are available with different depth, width, and max channel parameters to accommodate different needs of computational speed and accuracy. In this analysis, we used YOLOv8l and YOLOv11l models due to their comparable sizes in terms of the number of parameters.

3.2 Data Acquisition and Pre-processing

A comprehensive dataset of heavy construction equipment images was meticulously constructed for training and evaluating the classification model. This dataset encompasses 7,363 images categorized into 15 different classes, ensuring a diverse representation of various equipment types. After collecting image data, they were subjected to several pre-processing steps to ensure their readiness to be used in the classification model. First, watermarks and other irrelevant text were removed from the images. Next, redundant and unclear images were excluded. Images were then annotated utilizing Roboflow, an online platform for image data annotation and management (Dwyer et al. 2024). To ensure high-quality annotation, a team of five trained annotators participated in labeling the dataset images. Prior to annotation, the annotators received training on construction equipment identification using a structured guideline developed by the research team. In addition, label Assist feature was used to accelerate annotation. This feature provided bounding box suggestions based on a pre-trained object detection model (Dwyer et al. 2024). Annotators reviewed and corrected these suggestions to ensure proper and accurate annotation. Each image was subsequently quality-checked by one of the authors to maintain consistency and reduce labeling errors.

Once annotated, the dataset was divided into three subsets: training, validation, and testing with 70%, 20%, and 10%, respectively. The training dataset was used to build the models via learning complex relationships between image features and their respective classes. The validation set was used for performance enhancement while avoiding overfitting. On the other hand, the testing dataset, never seen during model calibration and enhancement, was used for the final evaluation to provide an unbiased measure of the model's accuracy.

The training dataset was enriched by some augmentation techniques to simulate variations encountered in real-world scenarios, hence, improving the robustness and generalizability of the model. In this step, images were rotated randomly within a range of -15° to +15° to account for variations in object orientation. The hue values of the images were altered randomly within a range of -9° to +9°, introducing color variations to mimic different lighting conditions and reduce sensitivity to color changes. Brightness levels of the images were adjusted randomly between -15% and +15% to account for diverse illumination conditions. A blur effect with a maximum intensity of 0.5 pixels was applied to simulate slight defocus or motion blur, improving the model's resilience to image quality inconsistencies. Finally, random noise, up to 0.1% of the total pixels, was introduced to the images to simulate distortions that may occur in real-world data acquisition.

3.3 Model Training and Validation

This study investigates the performance of some of the most used deep-learning classification models. The models included in this study are Yolov8, Yolov11, ResNet18, ResNet50, ResNet101, and AlexNet. All models were trained on an Nvidia Tesla T4 (16 GB memory) GPU. The stochastic gradient descent (SGDM) optimizer was used in all training experiments with a momentum set to 0.9. A categorical cross-entropy loss function was utilized to reduce the classification error (Liu et al. 2018). An initial learning rate of 0.01 was selected with a cosine learning rate decay. The batch size was set to 16 for all models except those with heavy backbone, ResNet50 and ResNet101, a batch size of 8 was selected. All models were trained for at least 50 epochs, or until their training loss converged. A patience parameter was defined to stop the training if the training loss plateaued with no improvements for 10 epochs. The remaining training settings followed the standard defaults of the original model implementations.

Transfer learning was used to speed up the learning process. This was achieved by exploiting pre-trained model weights as a starting point for training the new heavy equipment classification models. Specifically, pre-trained weights of models trained on the ImageNet dataset were utilized for ResNet and AlexNet models (Deng et al. 2009). On the other hand, pre-trained weights for YOLO models were retrieved from perspective GitHub repositories (Jocher et al. 2023, Jocher and Qiu 2024). In addition, Batch normalization was used in Yolo models while L2 regularization was applied for ResNet and AlexNet models to prevent overfitting, which can impair the performance of deep learning models.

3.4 Performance Evaluation

After training the deep learning models, their performance was assessed utilizing the testing dataset. The aim is to evaluate model's ability to accurately classify heavy construction equipment. The effectiveness of the developed models was assessed utilizing several performance metrics such as accuracy, precision, recall, and F1-score. Accuracy (A) is the proportion of correctly classified predictions out of the total predictions. Precision (P) is the proportion of correctly predicted positive instances out of all predicted positives. Recall (R) is the proportion of correctly predicted positive instances out of all actual positives. F1 score provides a harmonic mean of precision and recall. These metrics are given by the equations below:

$$[1]\ \mathrm{A} = \frac{\mathrm{CP}}{\mathrm{TNP}}$$

$$[2]\ \mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

$$[3]\ \mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$[3]\ \mathrm{F1\ Score} = \frac{2 \times\ \mathrm{P} \times \mathrm{R}}{\mathrm{P} + \mathrm{R}}$$

Where CP is the correct predictions, TNP is the total number of predictions, TP is true positive predictions, FP is false positive predictions, and FN is false negative predictions.

4. RESULTS AND DISCUSSION

Figure 2 displays the loss curves for the training and validation stages with the YOLOv11 model. The training loss curve exhibits a consistent and gradual decline, indicating successful learning. Correspondingly, the validation loss curve declines, indicating the model's proficiency in generalization and accurate prediction. The convergence of training loss to a low value (0.1) indicates that the model has successfully acquired the complex features of the heavy construction equipment classification task. Similar curves for YOLOv8, AlexNet, and other ResNet models are not included here due to space constraints.

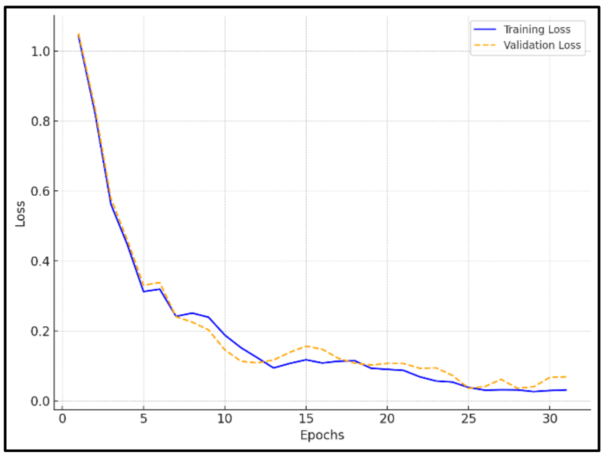

Figure 2. Training vs. validation loss for YOLOv11

Figure 3 summarizes the achieved accuracy of each deep learning classifier. It can be observed from this graph that YOLO models generally performed better than ResNet and AlexNet models. In addition, YOLOv8 model achieved slightly higher accuracy than Yolov11. However, this 0.7% accuracy increase, was accompanied by 20% increase in inference time. YOLOv11 presented a better trade-off between classification speed and accuracy.

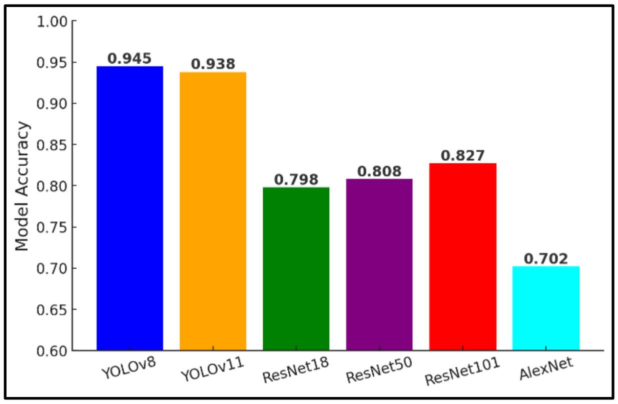

Figure 3. Accuracy of investigated models

Table 1 provides a detailed evaluation of the YOLOv11 model's performance in classifying heavy construction equipment into 15 distinct classes. The precision scores average around 0.95, with. The model shows high precision in classifying most equipment classes such as asphalt rollers, concrete mixer trucks, boom lifts, telescopic handlers, and others with 100% precision. The differences in precision scores may stem from variations in visual complexity and uniqueness within the classes, with equipment possessing more readily recognized characteristics attaining superior precision. Lower precision values could be related to visual similarities between certain equipment and other equipment types. Bulldozers, for example, were confused with scrapers and excavators.

Table 1 Classification metrics report

Class	Precision	Recall	F1 Score
Asphalt Roller	1.000	1.000	1.000
Boom Lift	1.000	1.000	1.000
Con. Mixer	1.000	0.974	0.987
Dewatering	0.923	0.923	0.923
Grader	0.950	0.950	0.950
Bulldozer	0.500	1.000	0.667
Compactor	1.000	0.979	0.989
Crane	0.941	1.000	0.970
Dump Truck	1.000	1.000	1.000
Excavator	1.000	0.950	0.974
Forklift	0.917	0.917	0.917
Pile Driving	0.967	1.000	0.983
Scraper	1.000	1.000	1.000
Skid Steer	1.000	1.000	1.000
Telehandlers	1.000	1.000	1.000

Concerning the model’s ability to correctly identify all relevant instances within a class, the recall values range from 0.92 to 1.0, demonstrating the model’s high capability to detect true positives in this category. F1-score ranges from 0.67 to 1.0. Again, bulldozer class had the lowest F1-score of 0.67, emphasizing some challenges in accurately distinguishing this class. After consulting the training set, there was relatively smaller number of bulldozer instances compared to other classes. In addition, most of bulldozer images were focused on the moldboard with few images showing other details of the equipment. Overall, the metrics in Table 1 demonstrate that the model has superior performance in classifying heavy construction equipment despite specific issues remain in bulldozer class. Resolving these challenges may necessitate increasing and diversifying the instances related to this class in the training dataset.

A Receiver Operating Characteristic (ROC) curve was then generated to thoroughly evaluate the performance of the YOLOv11 model. Figure 4 illustrates the ROC curves for the 15 heavy construction equipment classes considered in this study. It presents a visual summary of the Area Under the Curve (AUC) metrics revealing the model's efficacy in distinguishing different equipment classes. The model achieved AUC scores ranging between 0.97 and 1.0 for all classes implying superior classification accuracy.

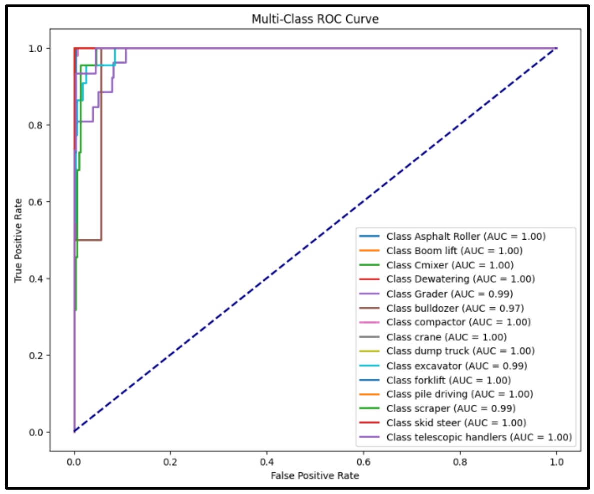

Figure 4. ROC curves of equipment

In addition to classification performance, localization accuracy was also evaluated using the mean Average Precision at Intersection over Union (IoU) threshold of 0.5 (mAP@50), a standard metric for object detection tasks. The YOLOv8 model achieved a mAP@50 score of 0.941, while the YOLOv11 model achieved a score of 0.933. These results indicate that both models are capable of accurately classifying heavy equipment and precisely localizing them. The strong localization performance further supports the models' applicability in real-time automated monitoring systems for construction environments. To further illustrate the model's performance, Figure 5 presents sample prediction outcomes from the classification model. it demonstrates the model's ability to accurately recognize various types of heavy construction equipment, including skid steer, crane, compactor, and others, across diverse environmental conditions.

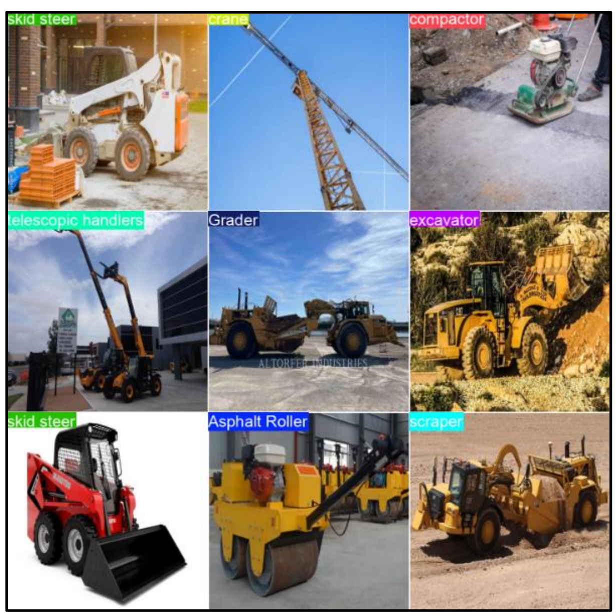

Figure 5. Sample of image prediction results

Similarly, Figure 6 demonstrates the model's prediction capabilities on drone footage. The model accurately identified and localized an excavator and a dumping truck. Such results highlight the models' suitability for automated monitoring of construction equipment in dynamic environments.

Figure 6. Prediction results for sample video frames from drone footage.

5. CONCLUSIONS

Despite the rapid expansion of deep learning applications, the construction industry remains much behind in using these techniques. This paper examined the usefulness of deep learning algorithms in accurately classifying heavy equipment in construction projects. Advanced deep learning algorithms were utilized to develop the classification models. Data augmentation and transfer learning techniques were adopted to ensure the robustness of the classifiers. Among the investigated models, YOLOv11 exhibits remarkable accuracy, with an average precision and recall scores of 0.95 and 0.98, respectively. It achieved almost a similar accuracy to YOLOv8 with 20% less inference time. Compared to some benchmark CNN algorithms such as ResNet101 and AlexNet, YOLO models improved the classification accuracy by about 11% and 23%, respectively.

Those findings demonstrate the model's efficacy in precisely distinguishing various types of construction machinery, providing a robust foundation for smart construction management. Employing advanced YOLO architectures not only achieves an outstanding overall accuracy but also enhances real-time performance. This improved efficiency translates to faster, more reliable equipment monitoring, which is critical for operational productivity and safety on construction sites. In addition, it enhances resource allocation by monitoring consumption and minimizing idle time, leading to considerable cost savings. The model can also enhance site safety by reducing accidents associated with heavy machinery. To further validate our model and improve its generalizability, future work will involve benchmarking its performance on standardized datasets.

REFERENCES

Akinosho, T., Oyedele, L., Bilal, M., Ajayi, A., Delgado, M., Akinade, O., and Ahmed, A. 2020. Deep learning in the construction industry: A review of present status and future innovations. *Journal of Building Engineering*, 32: 101827.

Al-Khiami, M., and ElHadad, M. 2024. Enhancing construction site safety using AI: The development of a custom YOLOv8 model for PPE compliance detection. *Proceedings of the EC3 Conference 2024*. European Council on Computing in Construction

Anirudh, N., Padala, S., and Reddy, H. 2023. Development of ANN-Based Risk Prediction Model in Construction Projects. In *Recent Advances in Sustainable Environment*. Springer Nature, 109–121.

Arabi, S., Haghighat, A., and Sharma, A. 2020. A deep-learning-based computer vision solution for construction vehicle detection. *Computer-Aided Civil and Infrastructure Engineering*, 35(7): 753–767.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, Florida, USA, 248–255.

Dwyer, B., Nelson, J., Hansen, T., et al. Roboflow (Version 1.0) [Software]. On-line: https://roboflow.com, Accessed: 23/12/2024.

Elghaish, F., Matarneh, S.T., and Alhusban, M. 2022. The application of deep learning in construction site management: Scientometric, thematic and critical analysis. *Construction Innovation*, 22(3): 580–603.

Elshaboury, N., Yamany, M.S., Labi, S., and Smadi, O. 2024. Enhancing local road pavement condition prediction using bayesian-optimized ensemble machine learning and adaptive synthetic sampling technique. *International Journal of Pavement Engineering*, 25(1): 2365957.

Hou, L., Chen, C., Wang, S., Wu, Y., and Chen, X. 2022. Multi-object detection method in construction machinery swarm operations based on the improved YOLOv4 model. *Sensors*, 22(19): 7294.

Hussain, M. 2023. YOLO-v1 to YOLO-v8: The rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines*, 11(7): 677.

Jocher, G., Chaurasia, A., and Qiu, J. 2023. Ultralytics YOLOv8, version 8.0.0. On-line: https://github.com/ultralytics/ultralytics, Accessed: 23/12/2024.

Jocher, G. and Qiu, J. 2024. Ultralytics YOLO11, version 11.0.0. On-line: https://github.com/ultralytics/ultralytics, Accessed: 23/12/2024.

Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. 2018. Path aggregation network for instance segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, Utah, USA, 8759–8768.

Nath, N.D. and Behzadan, A.H. 2020. Deep convolutional networks for construction object detection under different visual conditions. *Frontiers in Built Environment*, 6: 532607.

Poinet, P. 2020. *Enhancing collaborative practices in architecture, engineering and construction through multi-scalar modelling methodologies*. PhD dissertation.

Redmon, J. 2016. You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA.

Sherafat, B., Ahn, C.R., Akhavian, R., Behzadan, A.H., Golparvar-Fard, M., Kim, H., Lee, Y.C., Rashidi, A., and Azar, E.R. 2020. Automated methods for activity recognition of construction workers and equipment: State-of-the-art review. *Journal of Construction Engineering and Management*, 146(6): 03120002.

Soltani, M.M., Zhu, Z., and Hammad, A. 2016. Automated annotation for visual recognition of construction resources using synthetic images. *Automation in Construction*, 62: 14–23.

Soltani, M.M., Karandish, S.F., Ahmed, W., Zhu, Z., and Hammad, A. 2017. Evaluating the performance of convolutional neural network for classifying equipment on construction sites. *Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC)*, volume 34, Taipei, Taiwan.

Wang, Z., Zhang, Y., Mosalam, K.M., Gao, Y., and Huang, S.L. 2022. Deep semantic segmentation for visual understanding on construction sites. *Computer-Aided Civil and Infrastructure Engineering*, 37(2): 145–162.

Xiao, B., and Kang, S.-C. 2021. Development of an image data set of construction machines for deep learning object detection. *Journal of Computing in Civil Engineering*, 35(2): 04020071. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000945.

Xiao, B. and Kang, S.C. 2020. Vision-based method integrating deep learning detection for tracking multiple construction machines. *Journal of Computing in Civil Engineering*, 35(2): 04020071.

Xiao, B., Lin, Q., and Chen, Y. 2021. A vision-based method for automatic tracking of construction machines at nighttime based on deep learning illumination enhancement. *Automation in Construction*, 127: 103721.

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

Yamany, M.S., Elbaz, M.M., Abdelaty, A., and Elnabwy, M.T. 2024. Leveraging convolutional neural networks for efficient classification of heavy construction equipment. *Asian Journal of Civil Engineering*, 25(8): 6007–6019.

Zeng, T., Wang, J., Cui, B., Wang, X., Wang, D., and Zhang, Y. 2021. 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*, 291: 123268.