2024 EC³
2024 European Conference on Computing in Construction
Chania, Crete, Greece
July 15-17, 2024

DETECTION AND CLASSIFICATION OF CONSTRUCTION OBJECTS BY USE OF MACHINE VISION AND DEEP LEARNING

Nicolas Nicolaou, and Symeon E. Christodoulou
Department of Civil and Environmental Engineering, University of Cyprus, Nicosia, Cyprus

Abstract

This research paper focuses on the detection and classification of objects at construction sites and analyzes the utility and potential of such detection and classification activities in the construction industry. Object detection and classification are performed by applying technologies such as machine vision (MV) and deep learning (DL) to image processing and/or in combination with object segmentation and labeling. These activities have varied applications at construction sites, including but not limited to: (1) the monitoring of workers and machinery for productivity measurement and for the prevention of accidents and collisions; and (2) the monitoring and classification of procured and installed construction materials for the evaluation of work progress. This serves as a valuable, low-cost measurement tool in the context of the management and monitoring of construction projects.

Introduction

The construction sector, even in modern times, faces numerous perennial problems and challenges related to the effective management and monitoring of construction projects.

Some of these challenges are, for example, associated with safety and health on construction sites. Despite significant progress achieved through the introduction of regulatory frameworks and legislation, a considerable number of occupational accidents are still recorded today. These accidents are not exclusively personal but are often linked to the reckless use of mechanical equipment and vehicles, as well as insufficient coordination, organization, and monitoring at construction sites. Therefore, a system using machine vision (MV) and deep learning (DL) technologies could, in this case, detect a worker in a restricted zone of the construction site, or determine whether the worker is wearing appropriate safety equipment or, even better, continuously appraise the ergonomic risks to workers (Lambrides and Christodoulou, 2023). Additionally, the system could be used to ensure the proper operation of mechanical equipment in designated areas and the maintenance of safe distances by the working personnel at the construction site.

Another critical issue commonly encountered at construction sites is resource management and construction progress monitoring. Delays, particularly in large construction projects, often occur due to material shortages and insufficient logistics in material delivery. Conflicts and construction delays are also frequent occurrences. Therefore, a mechanism based on the aforementioned technologies could be employed to monitor the transport of construction material to sites and visually document the construction process. This visual documentation can be valuable for reference and analysis.

In light of the aforementioned and other challenges faced by the modern construction industry, the use of machine vision and deep learning technologies is imperative. These technologies enable the automation of numerous technical processes on the construction site while facilitating the monitoring and resolution of various problems within it.

Literature Review

In recent years, a significant number of research studies has been conducted on the use of machine vision (MV) and deep learning (DL) technologies for detecting and classifying construction elements at construction sites. This trend began in the early 2010s, with works on the automated generation of parametric BIMs (Brilakis et al., 2010), and despite the significant improvements in relevant applications, there is still room for further development.

The article by Czerniawski and Leite (2020) introduces the automation of digital modeling of existing buildings through reality capture devices and computer vision algorithms. The goal is to facilitate the use of digital building representation technologies, promoting new forms of simulation, automation, and information provision. The research results suggest that achieving a more complete semantic coverage of building infrastructures will require a revision and intensification of relevant efforts.

Nath and Behzadan (2020) propose the validation of a genetic adversarial network (GAN) based on a deep convolutional neural network (CNN). The research involves photos taken, trained, and tested at the construction site from two internal datasets to increase image resolution when generating missing pixel information. Results demonstrate that using GAN-enhanced images can further improve the average accuracy of pre-trained models for object detection while maintaining overall processing time for real-time object detection.

In a subsequent work, Paneru and Jeelani (2021) provided an up-to-date and categorized overview of computer vision applications in construction by examining recent developments in the construction sector and the challenges that future research must address to maximize the benefits of computer vision. The authors focus on specific areas considered most likely to benefit significantly from computer vision, such as safety

management on construction sites, progress and productivity monitoring, and work quality control.

One year later, Duan et al. (2022), focused on developing a large-scale image dataset specifically collected and processed for construction sites, named SODA (Site Object Detection Dataset). This dataset includes 15 types of objects categorized into mechanical means, materials, and labor personnel. Specifically, 20,000 images were collected from various construction sites, considering different construction site conditions, weather conditions, construction phases, and shooting angles. After careful examination and processing, 19,846 images were selected, containing 286,201 objects accompanied by corresponding labels from predefined categories.

An analysis conducted indicated that the developed dataset is advantageous in terms of diversity and volume. Further evaluation using two widely accepted object detection algorithms based on deep learning (*YOLO v3* / *YOLO v4*) demonstrated the dataset's effectiveness in visualizing typical construction scenarios, achieving a maximum mean Average Precision (mAP) of 81.47%. This research contributes a large-scale dataset for the development of deep learning applications in object detection within the construction industry. It serves as a reference point for the further evaluation of corresponding algorithms in this field.

In their work, Wang et al. (2022) proposed a new semantic method aiming to extract information by integrating deep learning object detection and image captioning. This method explores important information from construction images or videos. In the proposed approach, object detection serves as an encoder to extract features of construction objects and the holistic image. By adopting this method, semantic information from construction images can be presented to project managers as a valuable tool for making crucial decisions on the construction site.

In the research work of Hou et al. (2022), a multi-object detection method based on the improved *YOLOv4* model is proposed to overcome the problem of low detection accuracy. Research results indicate that the average accuracy (mAP) of the improved *YOLOv4* model for many objects can reach 97.03%, which is 2.16% higher than that of the original *YOLOv4* detection network. At the same time, the detection speed reached 31.11 fps, a decrease of 0.59 fps, a result quite satisfactory for real-time detection data.

Zhou et al. (2022) propose an object detection method based on an improved *YOLOv5* model with high sorting accuracy of construction waste. It involves creating a dataset from images of construction waste taken in situ at construction sites. This improved model was trained, validated, and tested based on the collected images and compared with other conventional models such as *Faster-RCNN*, *YOLOv3*, *YOLOv4*, and *YOLOv7*. The *YOLOv5* model recorded an average accuracy (mAP) on the test dataset of 0.9480, indicating better performance than other conventional models in object detection.

In a recent research paper by Jog et al. (2022), full-scale validation experiments of a multi-object location tracking method for its application to resource tracking in large-scale, congested, outdoor construction sites are presented. The validation stage involved testing under harsh conditions on various large project sites. This research paper describes the process of data collection and testing, as well as the measurements and results obtained. The validation showed that the new vision tracking provides a good solution for tracking different entities in large and congested construction sites.

Research Methodology

The research work discussed herein focuses on the automated detection and classification of construction objects, and the applied research methodology was based on utilizing the Python programming language along with machine vision and deep learning technologies. In the realm of these technologies, several terms are often used interchangeably, yet they entail distinct tasks and methodologies. Classification entails assigning predefined labels or categories to input data based on their inherent features or attributes. Its primary objective is to categorize input instances into one of several predetermined classes, such as determining whether an image depicts a cat or a dog. On the other hand, the term "prediction" encompasses various interpretations, but within the domain of object classification, it involves assigning a probability score to each class to indicate the model's confidence level in its classification decision. Detection, meanwhile, pertains to the identification and localization of specific objects or phenomena within an input scene or data stream. It focuses on discerning the presence and position of objects of interest within images, videos, or sensor data, often using bounding box annotations. However, recognition, frequently conflated with detection, refers to the process of identifying and comprehending objects or patterns within an image or scene. Unlike detection, recognition entails a more profound analysis of visual content, which may include grasping the context, identifying specific object features or traits, and drawing higher-level associations or inferences based on observed patterns.

The goal was to create software, or leverage existing tools, capable of learning a series of construction objects present at a construction site. Subsequently, the software should successfully detect and classify these objects using either images from a dataset or random images. To achieve this objective, *ImageAI* (v.3.0.3) was employed. *ImageAI* (Moses, 2018) is an open-source Python library that simplifies machine vision and deep learning tasks. It is built on other libraries such as *TensorFlow* and *Keras*. From the array of tasks offered by *ImageAI*, specific codes related to image classification and object detection were utilized - activities directly aligned with the focus of this research. For each of the two tasks, a code was used for

custom model training process based on the custom classes, resulting in the creation of a model. Additional codes were employed for result extraction, verification of the resulting accuracy-performance, and broader evaluation of the respective trained models, primarily through the utilization of unseen data.

Furthermore, a dataset was created for each task, incorporating photos of all the examined objects. These data resulted from a combination of my own photos from construction sites, ready-made datasets from Kaggle, which is a platform for data science and machine learning competitions, and generally photos obtained by Google Images search service. In the context of this research, the decision was made to initially explore two distinct classes to clarify the operational mode and compatibility of *ImageAI* with the research goals. These objects were the 'column' and the 'excavator'. However, at a later stage, seven more classes (totaling 9) were added to the detection, as follows: 'beam', 'masonry', 'slab', 'window', 'person', 'safety helmet', and 'reflective jacket'. The choice of some of these object classes relates to the intent of using developed algorithms and trained models for use in health & safety applications at construction sites.

Image Classification Framework

For this task, a set of 6000+ images of the object classes to be examined was collected. Initially, a general folder was created, which contained two additional folders named 'train' and 'test,' respectively. Within each folder, a subfolder was created for each prediction object. The training photos, used to train the classification model, and the corresponding test photos, used to evaluate it, were placed in these subfolders.

In the 'train' folder/dataset, 500 photos were included for each object, while in the 'test' folder/dataset, 200 photos were included. This dataset was then utilized in the training code, where various tasks were performed, including the selection of the algorithm. *ImageAI* offers the option to use four different algorithms for training custom image classification models (*MobileNetV2*, *ResNet50*, *InceptionV3*, and *DenseNet*), each with different speed and prediction accuracy characteristics.

Additionally, other parameters such as 'batch_size' (the number of images the network will process simultaneously) and 'num_experiments' (the number of network training iterations on all training images) were set in this code. For the purposes of this work, the MobileNetV2 algorithm was chosen due to its fastest prediction speed in compare with other algorithms.

Upon each execution of training code, the model with the highest accuracy was generated and stored in the dataset folder. Additionally, the other parameters mentioned above were systematically varied during each run to elucidate their impact on the accuracy of the respective model. This measure was undertaken to facilitate the incremental enhancement of the model, which became evident with each successive iteration. In this context, accuracy represents the percentage probability that a detected object belongs to a specific class. The accuracy is calculated using the following formula:

$$\text{Accuracy} = \frac{\text{Number of Correctly Classified Images}}{\text{Total Number of Images}} * 100$$

This percentage reflects the model's confidence in the correctness of its prediction. Higher percentage probabilities generally indicate that the model is more confident in recognizing a particular class of object in the image.

At a later stage, this model was employed in another code, where its effectiveness in predicting the examined and subsequently trained objects was evaluated using both trained and random photos. A schematic overview of this methodology is depicted in Figure 1.

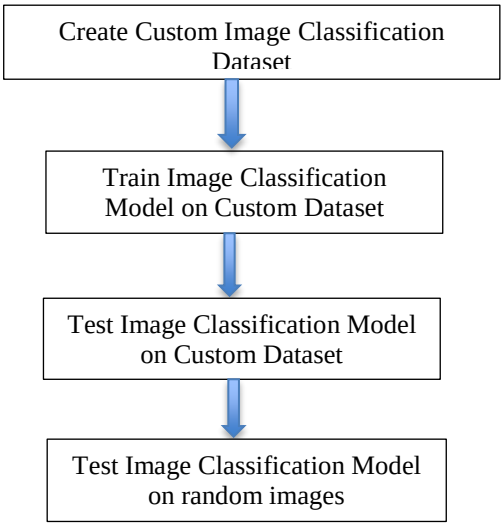

Figure 1: Image Classification methodology flowchart

Object Detection Framework

For this task, a set of 2700+ images was collected for the examined classes. Initially, a general folder was created, which included two subfolders named 'train' and 'validation,' respectively. Within each of these folders, two additional subfolders were created. The first, named 'images,' contained photos of the examined objects without separating them based on the object they depict. The second, named 'annotations,' contained the corresponding assignments for these classes, in txt format.

To create the assignments, an open-source graphic annotation tool for images, Labelimg, was employed. The process involved creating bounding boxes and labels in each photo and assigning them to each of the examined objects for the purpose of learning object detection. As part of this activity, 300+ photos were collected for each object, with 70-80% stored in the 'train' folder for training the detection model and the remainder in the 'validation' folder for evaluating the model's performance during training.

This dataset was then input into the training code, where, among other tasks, algorithm selection was performed. *ImageAI* provides the option to use two different algorithms to train custom image object detection models, namely *YOLOv3* and *TinyYOLOv3*, each with varying speed and accuracy characteristics for prediction. In this code, additional parameters such as 'batch_size' and 'num_experiments' were set, as previously explained.

During the training process for object detection, the initially used model did not include specific objects such as those found at construction sites. The model training with additional, construction site objects enriches the utilized pre-trained model and facilitates its use on construction-related image detection applications. Additionally, the option for training using a pre-trained *YOLOv3* model was specified. For the purposes of this work, both algorithms were employed. Future work shall aim the incorporation of newer releases of YOLO models (e.g., *YOLOv8*) and training datasets (e.g., *SODA*).

Each time the code was executed, the model with the highest accuracy in terms of mAP50 (mean Average Precision at 50%) was generated and stored in the dataset folder. Additionally, the other parameters mentioned above were systematically varied - in conjunction with the practical application of non-maximum suppression (NMS) - during each run to elucidate their impact on the accuracy of the respective model. During the training of each model, in addition to mAP50, additional metrics such as precision, recall, and mAP50-95 were obtained. However, these metrics were not automatically saved. Precision is a measure of the accuracy of a model's positive predictions and is derived from the following relationship:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

On the other hand, recall, also known as sensitivity or true positive rate, is a term used to evaluate the performance of a classification or object detection model and is calculated as follows:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

By using these two terms, it is possible to calculate the F1 Score, another widely used metric for evaluating classification models. The formula for the F1 score is:

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

The term 'mAP' (mean Average Precision) is a metric that assesses the precision-recall tradeoff of a model. It evaluates how well a model performs at different confidence levels in its predictions. Specifically, 'mAP50' evaluates the model's precision and recall at a specific 50% Intersection over Union (IoU) threshold. Higher mAP50 values indicate better performance, with a maximum value of 1.0 representing perfect precision and recall at the specified IoU threshold. IoU is a metric that measures the overlap between the predicted bounding box and the actual location of the object. A 50% IoU means there is at least a 50% overlap between the predicted and actual contexts. This evaluation system is commonly used in assessing object detection models, including those trained for custom object detection tasks. At a later stage, this model was employed in another code, where its effectiveness in detecting the examined and subsequently trained objects was evaluated using random photos. A summary flowchart of this methodology is presented in Figure 2.

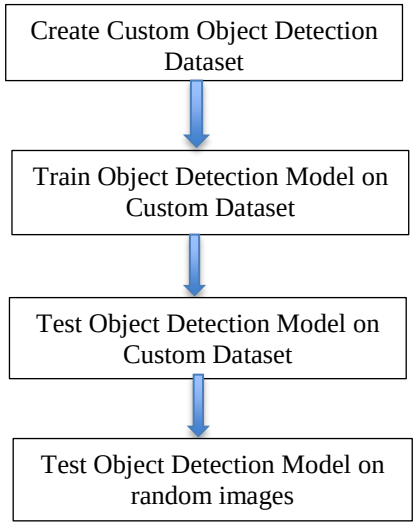

Figure 2: Object detection methodology flowchart

Additionally, the capability was provided by creating a Python code in combination with trained detection models to identify health and safety issues in photographs from construction sites, specifically printing a relevant warning message for the absence of part or all of the necessary safety equipment (protective helmet and reflective jacket) in case a person is detected in those areas.

Case Studies, Findings and Discussion of Results

In the pursuit of fulfilling this work's objectives, a series of tests were conducted through the execution of custom training Python programming language codes, as described. Throughout these tests, specific parameters were systematically varied in each training code, including the dataset itself, to generate two models - one for each task - with the highest accuracy and optimal performance. These models aimed to best fulfill the intended purpose for which they were created.

The final analysis results for the nine classes described in the previous stage are as follows. For image classification, considering the case of nine classes, a *MobileNetV2* model achieved an accuracy of 81.06%. This relatively high accuracy indicates the near certainty of the model in the correctness of its predictions, specifically in successfully predicting the nine trained classes in any given photo. This result was further validated by the model's performance on various photos, consistently

yielding generally high probabilities for correctly predicting the depicted object. The model was tested on both trained and random images, and during the conducted tests, no significant change in performance was observed between these two categories of images. It is thus evident that the model's performance was proportional to its accuracy rate. Some example results are provided below.

The figure below presents the outcomes of a random image illustrating various objects at a construction site, including columns, beams and slabs. Utilizing the aforementioned image classification model, a probability of 52.24% was assigned to the depiction of a beam in the photo, 26.02% for column and 21.38% for slab. Predictions for the remaining classes were notably low, aligning with expectations given that only these three classes were prominently featured in the image under examination. The difference between the three main predicted classes is based mainly on the extent of viewing each class from the angle of the photo. This result affirms the high predictive capability of the trained model.

```
beam : 52.2404
column : 26.02
slab : 21.3805
masonry : 0.1741
person : 0.1149
Safety_Helmet : 0.0305
excavator : 0.0174
Reflective_Jacket : 0.0142
window : 0.0081
```

Figure 3: Result of random image in image classification (example 1)

Figure 4 displays corresponding outcomes from a random photo depicting masonry and window, which are equally positive. In this photo the prediction percentage is relatively high and almost equal between the two mainly predicted classes and very low for the rest classes that are not represented. Similar results were obtained for the rest of the classes among random photographs with the characteristic of combining several classes in the same examined image with the results being generally satisfactory as can be seen in Figures 5 & 6.

```
window : 50.7926
masonry : 44.4139
Safety_Helmet : 1.6045
Reflective_Jacket : 1.513
person : 0.6469
beam : 0.5646
column : 0.315
slab : 0.1471
excavator : 0.0023
```

Figure 4: Result of random image in image classification (example 2)

The image in Figure 5 depicts a female worker at a construction site, with appropriate personal construction protection gear. The results of the applied model to this case confirm the successful prediction of the three main classes that appear in the photo in question, with the corresponding prediction percentages showing significant

```
Reflective_Jacket : 49.1006
person : 38.3086
Safety_Helmet : 12.3392
excavator : 0.2281
beam : 0.0159
masonry : 0.0066
slab : 0.0006
window : 0.0003
column : 0.0001
```

Figure 5: Result of random image in image classification (example 3)

fluctuations. Specifically, a higher percentage was given to the reflective jacket class (49.10%), with the person class following (38.31%) and the safety helmet class

registering a significantly lower percentage (12.34%) among these three classes.

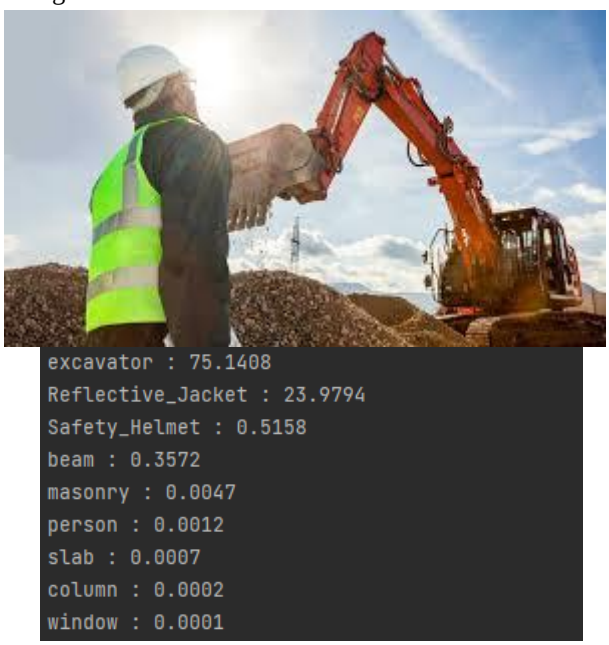

Figure 6: Result of random image in image classification (example 4)

Similarly, Figure 6 shows the results for an image showing a worker with appropriate personal protection gear near an excavator. The trained model correctly identified the existence of the excavator and the safety vest but not the worker and the safety helmet. One possible explanation for the very low prediction rates of the mentioned existing classes in the respective photograph is the intense lighting precisely at the point where the safety helmet is located, along with the posture of the human figure.

Frequently, models such as the one developed for image prediction encounter challenges related to their real-world effectiveness when presented with random images or images combining the examined objects. For instance, a model may perform well on the trained dataset but struggle to generalize its effectiveness to new, random data. However, as previously mentioned, no such phenomena were observed for this particular model.

Accordingly, for object detection, a *YOLOv3* model with an average accuracy (mAP) of 67.41 % was achieved. This figure indicates the relatively average to good accuracy of the specific model in terms of detecting and successfully classifying the objects under study in examined photographs, however efforts are being made to enhance the performance of this model to achieve even higher success rates. This result was further validated by the model's performance on various photos, where several satisfactory results were observed in terms of the true positive detection and classification of objects. The *YOLOv3* model was tested on both trained and random images, and during the conducted tests, no significant change was observed in terms of the model's performance between these two categories of images. Therefore, in this case as well, it is evident that the performance of the model is proportional to its accuracy rate. Some example results are provided below.

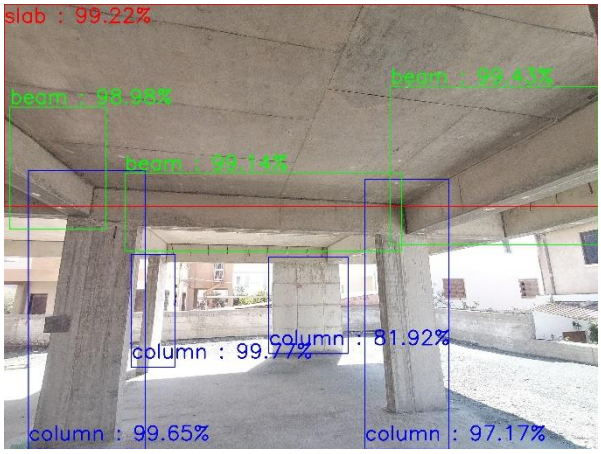

Figure 7: Result of rabdom image in object detection (example 1)

The figure above displays the outcomes of a random image depicting various objects on a construction site, including columns, beams and slabs. Utilizing the aforementioned object detection model, several load-bearing elements were detected with a probability of successful classification exceeding 81% affirming the high performance of the trained model. However, some other objects that would be expected to be detected by the model were not detected. These objects may have been influenced by the phenomenon of occlusion due to the relative angle of the photograph's capture. Nevertheless, the results largely coincide with those obtained using the classification model, which is deemed satisfactory. Additionally, Figure 8 demonstrates corresponding outcomes from a random photo of a worker on a construction site with appropriate personal protection measures. From the results of the model for this case, it emerged the successful detection of the three main classes that appear in the photo in question, with the corresponding confident scores showing very high (>96%) as in the case of the classification model for the same photo, the results of which were previously presented.

Figure 8: Result of random image in object detection (example 2)

Equally satisfactory results were obtained from tests conducted on other random photographs depicting the examined classes, as evident from Figures 9 & 10. The photo in question, presented in Figure 9, shows that using the model, masonry and a window were successfully detected with a confidence score exceeding 94%. Therefore, the behavior of the model under these conditions is considered satisfactory.

Figure 9: Result of random image in object detection (example 3)

Furthermore, in Figure 10, an example is provided depicting multiple objects (of those under examination), with the detection model yielding satisfactory results for most of them and moderate results for a single class (person). However, a better performance of the detection model is observed in the said image compared to its classification counterpart (Figure 6), as noticeably higher prediction rates are evident.

Figure 10: Result of random image in object detection (example 4)

Using the same detection model, results were obtained from other construction site photographs, with an additional capability introduced: the detection and notification of safety and health issues concerning the necessary and recommended personal protective measures on the construction site. Specifically, if a person was detected in these photographs without either or both of the reflective jacket and safety helmet, a corresponding warning was issued, as demonstrated in Figure 11.

Figure 11: Result of random image in object detection (health & safety example)

In conclusion, Figure 12 presents the confusion matrix of the detection model examined in the study. In a confusion matrix, each row represents the actual labels depicted in the validation set images, while the columns represent the corresponding labels predicted by the detection model. From the presented matrix, it is observed that some classes are positively evaluated due to a high number of true positive detections (e.g., excavator, window, etc.), while others are characterized as moderate to negative. The results of the matrix are to some extent expected, as the examined detection model did not achieve particularly high levels of accuracy. However, an NMS implementation with value equal to 0.4 led to an increase in the overall accuracy of the model in terms of F1-Score by approximately 30%. Essentially, the matrix provides insights into which classes the model struggles to predict accurately and can guide further improvements in the model, such as fine-tuning class-specific features or collecting more diverse training data for those classes. Therefore, there is room for significant future improvements.

Complete Confusion Matrix:

	column	excavator	beam	masonry	slab	window	person	Helmet	Jacket
column	349	0	111	10	67	3	3	0	0
excavator	0	76	0	0	0	0	3	0	0
beam	144	0	223	6	51	7	3	1	0
masonry	5	0	6	130	5	4	6	1	0
slab	66	0	41	3	78	1	2	0	1
window	13	0	5	7	2	277	26	0	0
person	7	5	4	6	4	15	298	58	74
Helmet	0	2	1	1	0	0	95	107	62
Jacket	0	0	0	1	0	0	85	70	65

Figure 12: Results of confusion matrix based on object detection model

Conclusions

The utilization of artificial intelligence, particularly technologies such as Machine Vision (MV) and Deep Learning (DL), in the construction industry is deemed imperative. The applications and benefits that can arise from these technologies are crucial, especially during the transition to a new era fraught with challenges. The real-time application of MV and DL can enhance the monitoring of safety and health issues on construction sites, extending to the broader and more essential oversight of labor management, mechanical equipment, vehicles, and materials, all while considering the relatively low costs resulting from the use of these technologies.

The present study focused on the automated detection and classification of construction elements at construction sites using the *ImageAI* library, built on the foundation of Python's *TensorFlow* and *Keras* libraries. The entire process was based on the integration of Machine Vision and Deep Learning technologies, combined with a dataset collected for the objects under consideration. The extracted results were analyzed in relation to the accuracy of the corresponding models from which they were derived. As part of future work, the following actions are to be taken to enhance the performance and accuracy of the relevant models based on *ImageAI*:

- Improvement of the annotation functions for bounded frames and labels, using advanced rendering practices in conjunction with the practical application of non-maximum suppression (NMS).
- Use of a balanced dataset with respect to all examined objects to prevent overfitting and the memorization of specific objects by the trained model for each activity.
- Exploration and testing of other custom activities offered by the *ImageAI* library for accuracy and usefulness, particularly by using video streams of related content.

References

Ahmadzada, A. (2020). People Image Dataset, many pictures of people performing different activities. https://www.kaggle.com/datasets/ahmadahmadzada/images2000/data

B Naik, N. (2023). Safety Helmet and Reflective Jacket, images of Individuals Wearing Safety Helmets and Reflective Jackets. https://www.kaggle.com/datasets/niravnaik/safety-helmet-and-reflective-jacket

Brilakis, I., Lourakis, M., Sacks, R., Savarese, S., Christodoulou, S., Teizer, J. and Makhmalbaf, A. (2010). Toward automated generation of parametric BIMs based on hybrid video and laser scanning data. Advanced Engineering Informatics, 24(4), pp.456-465.

Czerniawski, T. & Leite, F. (2020). Automated digital modeling of existing buildings: A review of visual object recognition methods. Automation in Construction, 113, p.103131.

Deshmukh, R., Wenguang, M. & Wei, M. (2020). Window Detection in Street Scenes, selected images from Paris Street-View Dataset with Window Annotations. https://www.kaggle.com/datasets/rude009/window-detection-in-street-scenes

Duan, R., Deng, H., Tian, M., Deng, Y. & Lin, J. (2022). SODA: site object detection dataset for deep learning in construction. arXiv preprint arXiv:2202.09554.

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

Jog, G.M., Brilakis, I.K. & Angelides, D.C. (2011). Testing in harsh conditions: Tracking resources on construction sites with machine vision. Automation in construction, 20(4), pp.328-337.

Lambrides, E., & Christodoulou, S.E. (2023). Human action detection and ergonomic risk assessment at construction sites, by use of machine vision and deep learning. In: EC3 Conference 2023 (Vol. 4). European Council on Computing in Construction, Crete, Greece.

Moses, O. (2018). ImageAI, an open source python library built to empower developers to build applications and systems with self-contained computer vision capabilities. https://github.com/OlafenwaMoses/ImageAI.

Nath, N. & Behzadan, A.H. (2020). Deep generative adversarial network to enhance image quality for fast object detection in construction sites. In: 2020 Winter Simulation Conference (WSC) (pp. 2447-2459). IEEE.

Paneru, S. & Jeelani, I. (2021). Computer vision applications in construction: Current state, opportunities & challenges. Automation in Construction, 132, p.103940.

Tzutalin (2015). LabelImg, a graphical image annotation tool. https://github.com/HumanSignal/labelImg

Umer Yasin, M. (2022). Bricks Under Construction or Old Building / Houses, an image dataset that contains pictures of buildings and houses under construction. https://www.kaggle.com/datasets/mumeryasin/bricks-under-construction-or-old-building-houses/data

Wang, Y., Xiao, B., Bouferguene, A., Al-Hussein, M. & Li, H. (2022). Vision-based method for semantic information extraction in construction by integrating deep learning object detection and image captioning. Advanced Engineering Informatics, 53, p.101699.

Zhou, Q., Liu, H., Qiu, Y. & Zheng, W. (2022). Object Detection for Construction Waste Based on an Improved YOLOv5 Model. Sustainability, 15(1), p.681.