Towards Automation in 5s – Classification and Detection of Images from Construction Sites

Viral Makwana[1] and Ganesh Devkar[2]

[1]Faculty of Technology, CEPT University, India
[2]Faculty of Technology, CEPT University, India
ganesh.devkar@cept.ac.in, viral.pcm21404@cept.ac.in

Abstract –

Traditional construction management practices often suffer from inefficiencies and waste, leading to increased costs and project delays. Lean Construction (LC), by emphasizing waste reduction and customer value prioritization, has emerged as an effective construction management system. In recent years, efficiency enhancement approaches such as Lean are being further augmented by automation and AI tools in the Architecture, Engineering, and Construction (AEC) industry.

This work develops an ML-based image processing algorithm that can be used within the Lean Construction framework. This algorithm classifies construction images based on the 5S Lean tool's sort and set in order principles, utilizing image analysis techniques. The algorithm that leverages convolutional neural networks and TensorFlow/Keras libraries, shows promise in effective image classification under the 5S framework. This study also proves that YOLO models can integrate lean concepts with AI Tools to foster improvements in productivity, safety, and sustainability in the construction industry.

Keywords –

5s;Image Classification; YOLO Lean Construction; Automation; Object Detection; Image Classification;

1 Introduction

The construction industry is facing increasing pressure to improve efficiency, productivity, and safety, and the avenues offered by automation, and Artificial Intelligence tools such as machine learning, object detection, and image analysis are seen as key to achieving these goals. Construction projects around the world are confronted with numerous challenges. Automation and AI tools can help streamline and optimize construction processes through data-driven decisions [1]. Object detection and image analysis can be used to improve the accuracy and efficiency of tasks such as classification and site inspection.

For the past two decades, Lean Construction is being used to overcome the inefficiencies and waste associated with traditional construction practices[2]. The efficacy of Lean construction can be greatly enhanced with the implementation of AI tools and there is increasing academic and industrial interest in this synergy[3]. The integration of automation, machine learning, and advanced analytics aligns with the lean principles of waste minimization by facilitating real-time monitoring, continuous improvement, and data-driven decision-making [2].

Cisterna et al., (2022) explored the benefits of integrating of Lean Construction (LC) methodologies and Artificial Intelligence (AI) approaches within the realm of production processes in projects and corporate settings. They used a theoretical framework and provided illustrative examples of application scenarios specifically within the Architecture, Engineering, and Construction (AEC) industry. The study's findings showed that the incorporation of Lean principles in construction initiatives can enable the assimilation of AI technologies. This was attributed to the capacity of AI-driven automation and support solutions to alleviate individuals from tasks characterized by tedium or complexity. Furthermore, the research showed that the interplay between AI and Lean methodologies can generate mutual growth and value addition[3].

One Lean construction methodology that can benefit from AI integration is 5S. 5S is a cyclical methodology based on a visual management approach, that emphasizes on organization and cleanliness in the workplace. To ensure the site is following 5S principles, personnel trained in the 5S tool periodically inspect the site and provide suggestions for improvement. Site photographs are frequently taken and reviewed in meetings to identify areas for improvement[4]. These improvements are categorized as sort, set in order, shine, standardize, and sustain. The identification of areas of improvements (which is also called as Kaizens) requires considerable involvement of human beings for analysis of site photographs from the perspective of 5S principles. Often, this process of analysis would be time consuming and tiresome. Utilizing advanced AI-driven image processing and machine learning tools can significantly enhance efficiency, and can be used in the sort and set in order categories[3].

The objective of this study was to develop an image processing algorithm that can classify construction images according to the sort and set in order principles of 5S Lean tool. The algorithm was designed to use image analysis techniques to identify specific features in the construction images and categorize them based on the 5S methodology. The goal was to provide a useful tool for construction professionals to improve their processes and increase productivity.

2 Literature Review

The Lean philosophy emphasizes the importance of continuous improvement, and 5S lean tool is considered a crucial prerequisite for implementing lean principles. According to Obulam & Rybkowski (2021), 5S can serve as a valuable tool for organizations to improve their operations by reducing waste and optimizing costs. The 5S method is based on five principles, each of which starts with the letter "S" phonetically. 1) The Seiri principle involves sorting by getting rid of unneeded items. 2) The Seito principle involves putting things in order by arranging them based on their purpose. 3) The Seiso principle is centered around cleaning or shining by eliminating sources of dirt. 4) The Seiketsu principle involves standardizing a system by establishing a consistent approach. 5) Finally, the Shitsuke principle emphasizes the importance of sustaining the system to ensure its ongoing maintenance. Obulam & Rybkowski (2021) The literature related to 5S principles mainly focuses on their practical application and execution in the workplace. Various checklists are used to evaluate the 5S process, which may take the form of questionnaires, marking or numbering checklists, or evaluation sheets. These evaluations may be considered as appraisals or assessments of goal achievement or the percentage of 5S implementation achieved.

Over the past decade, automated processes have become popular in a variety of industries. Periodic and repeatable processes can be optimized via automation. Repetitive and mundane jobs in construction include monitoring progress and evaluating quality of various processes. Currently, a team of experts completes these tasks by visually inspecting the construction site and frequently making use of measuring equipment. Unfortunately, given the reliance on human factors, this can result in discrepancies between reports that are issued at different times [6].

Construction companies are adopting digital imaging to enhance their visual inspection processes [7]. With digital imaging, high-resolution images and videos of the construction site can be quickly and easily captured, enabling potential issues or problems to be identified. The use of digital images also facilitates a more thorough and consistent inspection process, as they can be easily stored and reviewed at a later time. By utilizing digital imaging, construction companies can save time, reduce the risk of human error, and improve the overall quality of their construction projects.

The construction industry has seen an increase in the use of deep learning-based object detection methods, which are primarily focused on identifying components, detecting damage and defects, classifying objects, monitoring worker activities, and detecting collisions on construction sites. These methods use computer vision algorithms to recognize and detect objects There are several object detection models that are commonly used, including the R-CNN family [8], that includes R-CNN, Fast R-CNN, and Faster R-CNN, the YOLO family [9], and the SSD models [10]. The Yolo series of models is known for its high speed of detection. You Only Look Once (YOLO) is an advanced object detection algorithm used in computer vision. The latest Yolov5 model has been applied in various fields such as pedestrian detection, vehicle detection and medical applications used the Yolov5 model to develop a real-time detection method for unauthorized intrusion on construction sites [11].

3 Research Methodology

3.1 Image Classification Model

Based on studies and research, it is possible to automate the Sort principle of 5S using image classification. The Sort principle involves identifying the items that are necessary and those that are not and eliminating the unnecessary ones from the workplace. One potential application of image classification for the Sort principle is in the steel yard. Steel reinforcement bars are commonly used in construction, and they are often stored in large quantities in steel yards. However, these bars can become disorganized, making it difficult to find the required bars when needed.

To automate the sorting process in the steel yard, we can use image classification to identify and separate the steel reinforcement bars. The classification process can be achieved by training a machine learning algorithm to recognize two classes of images: one with steel bars placed in an unorganized manner, and another with steel bars placed in an organized manner, tied together in a bundle.

To train the machine learning algorithm, a large dataset of images of steel bars in the steel yard is required. The images should include examples of both organized and unorganized placement of the steel bars. Once a sufficient dataset is obtained, deep learning algorithms can be used to train the machine to recognize the two classes of images.

Once the machine learning algorithm has been trained, it can be used to classify new images of steel bars. As new images are fed into the algorithm, it would identify the placement of the steel bars and classify them as either organized or unorganized. This would enable construction workers to quickly and easily identify the steel bars that are needed, improving the efficiency and productivity of the construction process.

A methodology was proposed to create an image processing algorithm for classifying construction images for both modal Image classification as well as Object detection model according to the principles of 5S, which consists of the following steps: (1) Collecting a dataset of construction images and annotating them to identify the objects of interest; (2) Extracting relevant features from the images, such as object shapes, textures, and colors; (3) Training a machine learning model, such as a Convolutional Neural Network (CNN) or YOLO, using the annotated data to classify objects in the images; (4) Evaluating the model's performance on a test dataset to determine its accuracy in object classification; (5) Integrating the image processing algorithm into a construction workflow, such as the 5S system, to automate the classification of objects in construction images; (6) Continuously improving the algorithm's performance by incorporating new data and adjusting the model as required.

3.2 Object Detection model

The YOLO (You Only Look Once) model is a popular object detection algorithm that has been widely used for the identification of labeled objects in images. This model is designed to recognize objects in real-time, making it an effective tool for a wide range of applications. In order to train the YOLO model to recognize labeled objects in images, it is necessary to provide a dataset of annotated images that includes both labeled and unlabeled images. The annotated dataset should include a file that specifies the location of each labeled object within the image, as well as its class label.

According to the set-in-order principle of the 5S system, visually marked, specified places are designated for everything, and visual aids such as signboards, color-coding, labeling, or outlining may be used. The research methodology adopted for Yolo model development is shown in Figure 1.

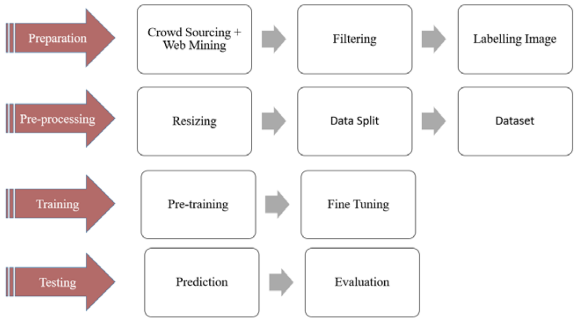

Figure 1 Methodology-Yolo model development

1. Preparation

To start building the YOLO object detection model, a dataset must be prepared, which will be utilized to train and test model. This dataset consists 50 to 60 nos of images from Web mining and construction site. This involves selecting and annotating the dataset, which includes choosing images for training and testing, and labeling the objects within the images indicating their presence and location accurately and consistently. The annotation process can be time-consuming and laborious, and there are annotation tools available like LabelImg and VGG Image Annotator (VIA) to aid in this task. Once the dataset is annotated, it is divided into the training data and testing data sets. The YOLOv5 model is trained using the training dataset, and its effectiveness is assessed using the testing set.

2. Pre-processing

Before training the YOLOv5 model, the prepared dataset must be preprocessed. This involves several steps such as resizing the images to a fixed size, normalization of the pixel values, and the application of data augmentation techniques.

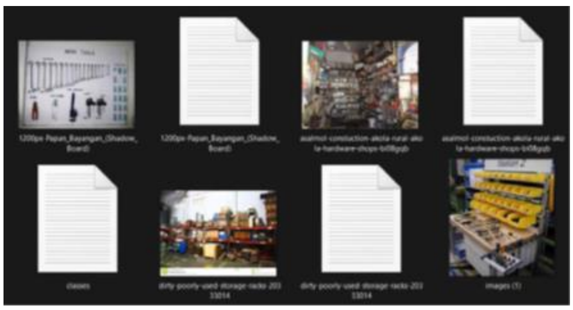

Figure 2 Image annotation

The purpose of resizing the images is to make sure that they are of uniform size, which is a requirement for training the YOLOv5 model. Normalization of the pixel values is necessary to scale the pixel values within a range of 0 to 1, which increases the YOLOv5 model's efficiency. In order to expand the training set and improve the model's generalizability, data augmentation approaches are used. Random cropping, flipping, and rotation are often used methods of data augmentation.

3. Training

To train the YOLO model with the prepared dataset, certain steps must be followed. First, the training environment is set up by installing essential software and libraries like PyTorch and CUDA and configuring hardware like GPUs. Next, the YOLO model is configured by setting hyperparameters like batch size, number of epochs, and learning rate, which significantly impact the model's performance. The preprocessed dataset is then fed into the model, and the model's parameters are iteratively changed depending on the loss function to begin the training process. The goal of this function is to minimize the difference between the expected output and the ground truth labels during training. It calculates the difference between the predicted output and the ground truth labels.

4. Testing

The performance of the YOLO model is assessed using the testing set by setting up the testing environment, configuring the YOLO model, and running the testing process. The testing environment is created by installing necessary software and libraries, like PyTorch and CUDA, and configuring the hardware, like GPUs, for testing YOLOv5 and YOLOv8 models. Configuring the YOLO model involves selecting the most suitable checkpoint, which is the saved model that performed best during training on the validation set, and then using the model to predict objects in the testing set. Various metrics like mean average precision (mAP) and intersection over union (IoU) are used to determine the accuracy and overlap of the predicted bounding boxes with the ground truth bounding boxes.

5. Fine-tuning and Deployment

Once the YOLO model has been tested, it is fine-tuned to enhance its performance. This process requires adjusting the hyperparameters and retraining the model on the complete dataset, including the validation and training sets.

Once the model is fine-tuned, it can be deployed in various applications, such as security camera systems and self-driving cars. The process involves integrating the YOLO model into a larger software system or application, enabling efficient object detection.

4 Data collection

4.1 Image Classification Dataset

Image classification is an important task in the field of computer vision, and it has numerous applications in various industries. One such application is the automated sorting of steel bars based on whether they are organized or not. To carry out this task, it is necessary to collect a significant amount of data in the form of images of steel bars. The images can be collected using various methods, such as web crawling and crowdsourcing.

Web mining is a method of collecting data from the internet. In this study, web crawling was used to collect images of steel bars from various web platforms and search engines. The images were collected using specific keywords such as "steel bars," "construction steel bars," "reinforcement steel bars," and so on. The images collected from web crawling varied in terms of quality, resolution, and orientation. Therefore, it was necessary to manually screen the images to remove any irrelevant or low-quality images. This process ensured that only high-quality images were included in the final dataset.

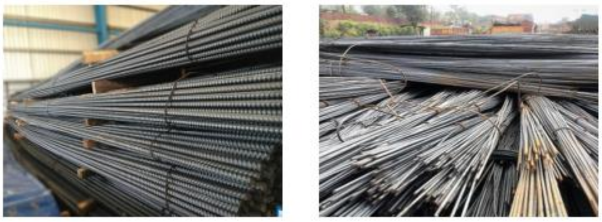

Figure 3 Reinforcement Bar Images for Training

In this study, crowdsourcing was used to collect images of steel bars from ongoing construction sites. The images were collected by contacting various construction companies and asking them to provide images of steel bars used in their ongoing projects. The images collected through crowdsourcing were of high quality and provided a wide variety of steel bar configurations. This method of data collection was particularly useful as it provided a more realistic representation of the steel bars used in real-life construction projects.

Once the images were collected, the next step was to clean and prepare the data. This process involved removing irrelevant and low-quality images and organizing the remaining images into relevant categories. The images were categorized based on the type of steel bars, such as organized or unorganized steel bars. The images were also resized and standardized to ensure that they were all of the same size and resolution. This step ensured that the images were processed efficiently during the image classification stage.

Techniques for data augmentation were utilized to expand the dataset. Data augmentation entails modifying the original photos to produce new images that are similar to the originals but not exactly the same. In this study, new pictures were produced using data augmentation methods including flipping, rotating, and zooming.

4.2 Object Detection Dataset

The principle of "Set in Order" involves organizing workspaces and equipment for improved efficiency and productivity. In this study, the aim was to automate the process of identifying labels as objects that are set in order using object detection techniques. To achieve this goal, a significant amount of data was required to train and evaluate the performance of the object detection model. As before, two methods were used for data collection: web mining and crowdsourcing.

In this study, web mining was used to collect images of labels from various websites and search engines. The images were collected using specific keywords such as "labels," "organized labels," "inventory labels," and so on. The images collected through web mining were of varying quality, resolution, and orientation. Therefore, the images were manually screened to remove any irrelevant or low-quality images. This step helped to ensure that only high-quality images were included in the final dataset.

Crowdsourcing was also used to collect images of labels from various industries and workplaces. The images collected through crowdsourcing provided a more realistic representation of labels used in real-life work environments. This method of data collection provided a diverse range of labels and configurations. The images were collected by contacting various contracting companies and asking them to provide images of their labeled equipment and workspaces.

After the images were gathered, LabelImg software was used to annotate them. Annotation entails manually highlighting the subject matter of a picture. The labels were the study's items of interest, and they were marked using bounding boxes. The annotation method contributed to the creation of a labelled dataset, which was necessary for the object detection model's training and performance evaluation.

The next step was to clean and prepare the data. This process involved removing irrelevant or low-quality images and ensuring that the remaining images were appropriately labeled. The images were organized into relevant categories based on the type of labels, such as inventory labels, equipment labels, and so on. The images were also resized and standardized to ensure that they were all of the same size and resolution. This step ensured that the images were processed efficiently during the object detection stage.

Data augmentation entails modifying the original photos to produce new images that are similar to the originals but not exactly the same. In this study, new pictures were produced using data augmentation methods including flipping, rotating, and zooming.

5 Data analysis

5.1 Image classification

This study implemented an image classification model using deep learning architectures such as TensorFlow Keras, Inception ResNet v2, and EfficientNet v2 B0. The dataset was split into training and validation sets with a validation split of 20%, and the model was trained using a batch size of 32 across 10 epochs. The history variable, which holds the loss and accuracy values for the training and validation sets for each epoch, was used to analyze the model's performance throughout training. A visual representation of the model's performance during training was made using these values. Similarly, the EfficientNetV2B0 model was tested on the dataset and the results were generated. One example of the EfficientNetV2B0 model is shown below, in Figure 4 where the accuracy graph is plotted.

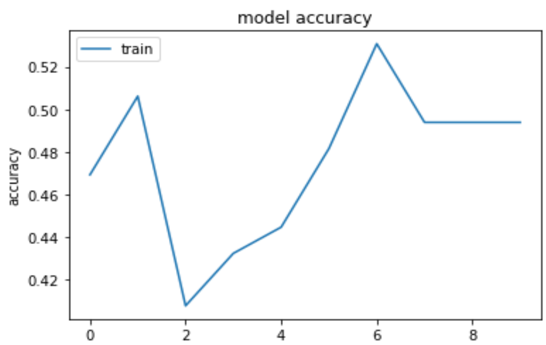

Figure 4 Model accuracy- EfficientNetV2B0

Model accuracy- EfficientNetV2B0

As shown in Figure 4, a graph displaying model accuracy is a visual representation of how well a machine learning model can predict the correct output for a given dataset. The Y-axis of the graph shows the accuracy of the model, while the X-axis indicates the number of training epochs or iterations. The training accuracy represents the model's performance on the data it was trained on, while the validation accuracy indicates how well the model is generalizing to new data.

The ideal situation is for both training accuracy and validation accuracy to increase steadily and converge over time, indicating that the model is learning and generalizing well. However, if the training accuracy continues to increase while the validation accuracy plateaus or decreases, this indicates overfitting, where the model is too complex and has learned to fit the noise in the training data instead of the underlying patterns. Conversely, if both the training and validation accuracy are low, this suggests underfitting, where the model is too simple and cannot capture the complexity of the data.

Figure 5 presents the EfficientNetV2B0 model, and the validation loss graph generated during training. The training loss and validation loss are shown on the Y-axis, while the number of epochs or iterations is shown on the X-axis. When plotted against epochs or iterations, the training loss and validation loss graph generally depicts the fluctuation in training loss and validation loss over time. The validation loss gauges how successfully the model generalizes to fresh, untested data, whereas the training loss assesses how well the model fits the training data. When training a model, the ideal situation is for the training loss to reduce over time while the validation loss stays the same or reduces. Overfitting of the model to the training data is indicated by a rising validation loss while the training loss stays declining.

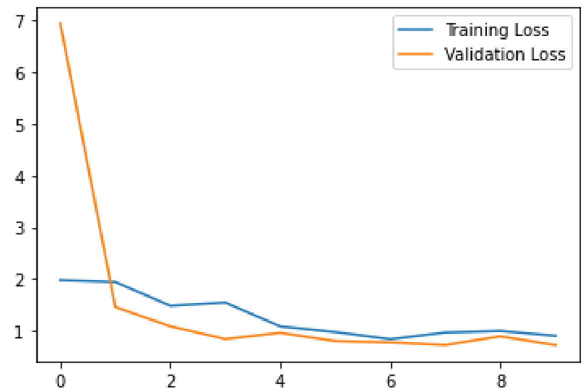

Figure 5 Train Val loss- EfficientNetV2B0

After training and validating the model on the training and validation sets respectively, the performance of each model was visualized by plotting the training and validation loss curves. The EfficientNet v2 B0 model achieved the highest accuracy of 0.8981%, followed by Inception ResNet v2 with 0.8276%, and TensorFlow Keras with 0.6969%. This indicates that the choice of architecture is crucial for achieving high accuracy in image classification tasks, as well as the quality of the dataset used for training and validation. Overall, the image classification model developed in this project successfully differentiated between organized and disorganized steel bar images.

5.2 Object Detection Model

The validation was performed using the best model saved during training. the validation results obtained from training the YoLOv5 model on the designated dataset. Subsequent to the validation process, the model underwent testing on a distinct dataset comprising 20 images, encompassing a collective total of 184 object instances. The model was tested on 20 images, containing a total of 184 instances of objects. The precision (P) and recall (R) values for the 'all' class are 0.813 and 0.732, respectively.

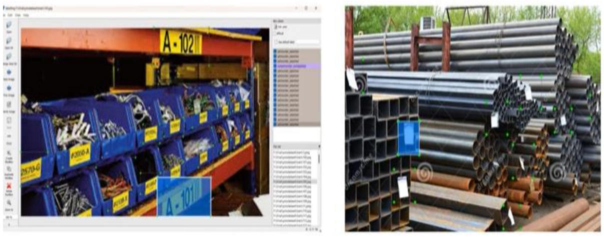

Figure 6 Image annotation for Labelled images

Furthermore, the validation results were broken down by two additional classes: 'setinorder_labelled' and 'notsetinorder &_notlabelled'. These classes may have been manually labeled by the user and refer to images where the objects are either set in order or not labeled in order as shown in Figure 6.

Overall, the mAP50 value for the 'all' class was 0.751, indicating a moderate performance in object detection. The mAP50 values for the additional classes were 0.849 and 0.653, respectively. These results suggest that the model may be better at detecting objects in labeled images that are set in order.

It is important to remember that the YOLOv5 model's accuracy can vary based on a wide range of variables, including the caliber and volume of training data, hyperparameters, and network design. To assess the model's performance in various circumstances, more testing and analysis are advised.

An essential tool for assessing a classification model's performance is the confusion matrix. It provides a thorough analysis of the predictions made by the model and how well they correspond to the actual labels. A confusion matrix was created for the YOLOv5 model during the validation procedure to examine the model's performance on the validation set. For each class the model was trained on, the matrix shows the quantity of true positives, false positives, true negatives, and false negatives.

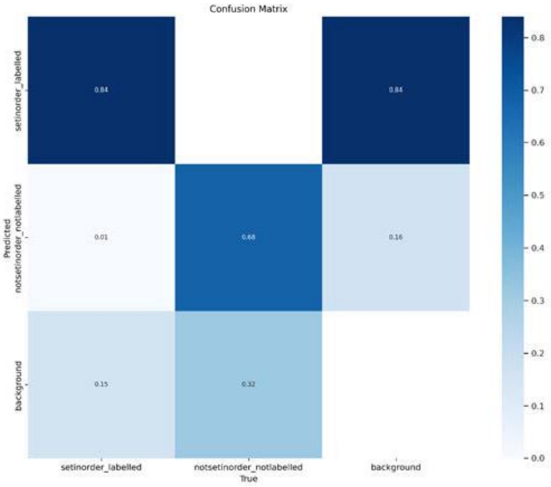

Figure 7 Confusion Matrix for YOLOv5

In the example Figure 7, the confusion matrix shows that the model correctly classified 156 of 184 instances of the "setinorder_labelled" class, resulting in a precision of 0.847 and a recall of 0.833. Similarly, the model correctly classified 22 of 28 instances of the "notsetinorder_notlabelled" class, resulting in a precision of 0.779 and a recall of 0.631. Thus, the confusion matrix generated for the YOLOv5 model provided a comprehensive overview of the performance of the model and helped to identify the areas that need improvement.

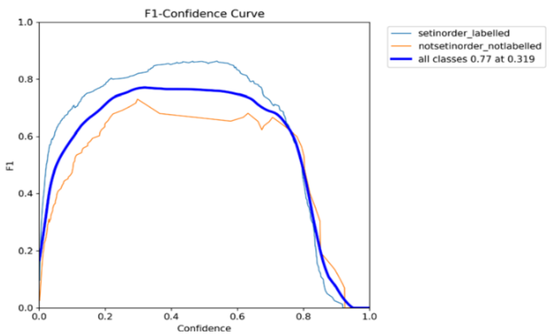

Figure 8 F1-Confidence Curve

The Figure 8 F1 confidence curve is a visual representation of the relationship between the confidence level of a machine learning model's predictions and its F1 score. The F1 score is a commonly used metric for evaluating the accuracy of binary classification models, and it takes into account both precision and recall. The F1 confidence curve is useful for identifying the optimal confidence threshold for a model, which can improve its overall performance. By analyzing the curve, we can determine at what confidence level the model achieves its maximum F1 score, allowing us to make informed decisions about how to tune the model's hyperparameters.

The YOLOv5 model generates a results.png file that gives a comprehensive overview of the model's performance on the validation dataset. The graph in the results.png file displays the precision, recall, and F1 score of the model at different confidence thresholds. Precision refers to the percentage of correctly identified objects out of all the predicted objects, while recall refers to the percentage of correctly identified objects out of all the actual objects in the image. The F1 score is a balanced measure of precision and recall and is calculated as their harmonic mean.

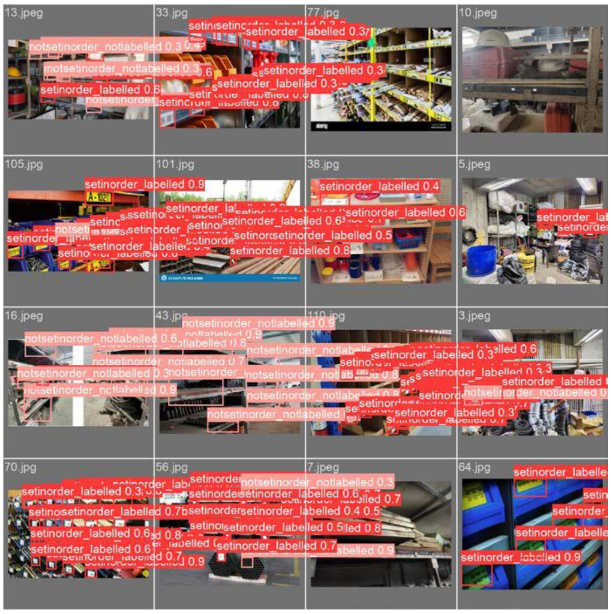

Figure 9 Yolov8 Object Detection model

Similarly, to the yolov5 model and yolov8 dataset trained on same dataset similarly result received as shown here. The YOLOv8 model as shown in Figure 9 achieved an overall mAP50 of 0.904 and mAP50-95 of 0.688 with a precision of 0.92 and recall of 0.879 on the validation set after 300 epochs of training. The model was able to detect 384 instances of objects in 34 images, out of which 304 instances were correctly identified and labeled as "setinorder_labelled" with a precision of 0.924 and recall of 0.908. The remaining 80 instances were labeled as "notsetinorder_notlabelled" with a precision of 0.917 and recall of 0.85. The model was able to perform inference at a speed of 19.1ms per image.
The confusion matrix shows the model achieved an mAP50 of 0.904 and an mAP50-95 of 0.688 across all classes. The precision and recall for the 'setinorder_labelled' class were 0.924 and 0.908, respectively, while for the 'notsetinorder_notlabelled' class, they were 0.917 and 0.85, respectively. The model performed better for the 'setinorder_labelled' class than the 'notsetinorder_notlabelled' class, with higher

precision, recall, and mAP50 values.

Similarly, to yolov5 model and yolov8 model generated result graph which shows coming with more accuracy. The result graph generated from the yolov8 result shows the relationship between the confidence threshold and precision, recall, and F1 score for each class in the dataset.

6 Conclusion

This study underscores the transformative potential of integrating Lean principles with automation in the construction industry. By taking the first step towards integrating Lean 5S methodologies with automation.

Although the preliminary algorithm's limitations underscore the necessity for refinement, the methodology's fundamental approach, utilizing convolutional neural networks and TensorFlow/Keras libraries, demonstrates promising potential for enhancing efficiency within the construction sector. This study, conducted with a constrained dataset sourced from both web mining and on-site observations, presents a significant analysis of implementing automation alongside Lean practices in practical settings. For that further more image for training and validation as well as case studies will be require. Also considering current data This study also proves that YOLO models can integrate lean concepts with AI Tools to foster improvements in productivity, safety, and sustainability in the construction industry.

The research also suggests avenues for future work, including the verification of AI model accuracy, algorithm development for "Set in Order," comprehensive data collection, and potential revisions to optimize outcomes. Future research may address the limitations of YOLO models, such as challenges with small or distant objects and interpretability concerns.

Exploring integration of lean principles with automation in construction processes emerges as a promising avenue to further enhance the transformative capabilities of the industry.

References

[1] Dumrak, J., & Zarghami, S. A. (2023). The role of artificial intelligence in lean construction management. *Engineering, Construction and Architectural Management*, *June*. https://doi.org/10.1108/ECAM-02-2022-0153

[2] Koskela, L., Howell, G., Ballard, G., & Tommelein, I. (2007). The foundations of lean construction. Design and Construction, January 2014, 211–226. https://doi.org/10.4324/9780080491080

[3] Cisterna, D., Lauble, S., Haghsheno, S., & Wolber, J. (2022). Synergies Between Lean Construction and Artificial Intelligence: AI Driven Continuous Improvement Process. Proc. 30th Annual Conference of the International Group for Lean Construction (IGLC), 49, 199–210. https://doi.org/10.24928/2022/0122

[4] Gil, D., Lee, G., & Jeon, K. (2018). Classification of images from construction sites using a deep-learning algorithm. ISARC 2018 - 35th International Symposium on Automation and Robotics in Construction and International AEC/FM Hackathon: The Future of Building Things,Isarc. https://doi.org/10.22260/isarc2018/0024

[5] Obulam, R., & Rybkowski, Z. K. (2021). Development and Testing of the 5S Puzzle Game. IGLC 2021 - 29th Annual Conference of the International Group for Lean Construction - Lean Construction in Crisis Times: Responding to the Post-Pandemic AEC Industry Challenges, 309–319. https://doi.org/10.24928/2021/0152

[6] Prieto, S. A., Giakoumidis, N., & de Soto, B. G. (2021). AutoCIS: An Automated Construction Inspection System for Quality Inspection of Buildings. Proceedings of the International Symposium on Automation and Robotics in Construction, 2021-Novem(November),669–676. https://doi.org/10.22260/isarc2021/0090

[7] Zhang, X., Bakis, N., Lukins, T. C., Ibrahim, Y. M., Wu, S., Kagioglou, M., Aouad, G., Kaka, A. P., & Trucco, E. (2009). Automation in Construction Automating progress measurement of construction projects. 18, 294–301. https://doi.org/10.1016/j.autcon.2008.09.004

[8] Girshick, R., Donahue, J., Darrell, T., Malik, J., & Berkeley, U. C. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation.580–587. https://doi.org/10.1109/CVPR.2014.81

[9] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-Decem, 779–788. https://doi.org/10.1109/CVPR.2016.91

[10] Liu, J., Jennesse, M. ;, & Holley, P. (2016). Utilizing Light Unmanned Aerial Vehicles for the Inspection of Curtain Walls: A Case Study. 2651–2659.

[11] Wan, H., Asce, M., Zhang, W., Ge, H., Luo, Y., Asce, M., & Todd, M. D. (2023). *Improved Vision-Based Method for Detection of Unauthorized Intrusion by Construction Sites Workers*. *149*(2015), 1–12. https://doi.org/10.1061/JCEMD4.COENG-13294