Detection of safety equipment to prevent hazards that threaten construction workers

MSc Research Project
Data Analytics

Daeun Sim
Student ID: x21209162

School of Computing
National College of Ireland

Supervisor: Hicham Rifai

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Daeun Sim
Student ID:	x21209162
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Hicham Rifai
Submission Due Date:	14/08/2023
Project Title:	Detection of safety equipment to prevent hazards that threaten construction workers
Word Count:	4701
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Daeun Sim
Date:	17th September 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Detection of safety equipment to prevent hazards that threaten construction workers

Daeun Sim
x21209162

Abstract

Construction industry sites have a high occurrence of accidents compared to other sectors. Wearing personal protective equipment can minimise injuries of workers and legally workers have to wear it at the construction site. Workers can underestimate the importance of properly wearing protective gear and it can be unfeasible for companies to inspect workers on a daily basis. In this case, an automated monitoring system based on computer vision can be used to detect any improper use of PPE equipment. In this work, an automated system is developed to detect if workers on a construction site are wearing a helmet, a vest, gloves, and boots. The automated system is based on the YOLOv8 model. The PPE detection model has achieved an mAP of 0.74 accuracy in detecting four safety equipment, with an F1 score of 0.6 for helmet and vest and 0.4 for gloves and boots.

1 Introduction

Safety regulations ensure that workers are protected and safe in workplaces. Even though these regulations are in place, serious injuries or life-threatening accidents continue to occur. In 2020, there were around 2.7 million accidents at work sites in European countries. Most work accidents occur at construction sites and can result in death or lifetime disability. According to statistics from Eurostat, from 2010 to 2020, the construction industry had the highest risk of workers suffering fatal accidents on the job in European countries.[1] In the United States, over a period of about 10 years, from 2010 to 2020, the death rate for construction workers has remained higher than the average rate for all workers. The statistics show that construction workers are more likely to have accidents than non-construction workers and are five times more likely to be fatal (Jalil Al-Bayati et al.; 2023). The Occupational Safety and Health Administration (OSHA) issued a list of hazardous working conditions in workplaces. As far as accidents in construction sites are concerned five possible risks were listed by OSHA. The five risks are related to hazardous material touching eyes or face, falling from ladders, falling from a high position, and scaffolding accidents.[2] To minimise risk, the regulation imposes on workers to properly wear Personal protective equipment (PPE) (Jalil Al-Bayati et al.; 2023).

Personal protective equipment (PPE) refers to the gear used to protect the human body from external materials which can cause accidents. It includes a safety helmet, boots,

[1]https://ec.europa.eu/eurostat
[2]https://www.osha.gov/top10citedstandards

goggles, mask, gloves, and boots (Table 1). Workers who are exposed to dust, extreme noise, or falling objects, must wear protective equipment at workplaces. PPE is critical equipment for workers on construction sites that can have numerous risks to the safety of workers. PPE protects workers from fatalities and provides a safe working environment. Training is usually provided to wear PPE correctly and to demonstrate the importance of the protection items needed for construction workers (Gunduz and Ahsan; 2018; Shamsuddin et al.; 2015). Wearing PPE equipment might be neglected by workers, so companies should continuously inspect the working environment to ensure compliance with safety rules and minimise the risk of accidents. However, construction businesses might have limited resources to invest in safety supervision (Wong et al.; 2020). The recent development of computer vision technology can offer an alternative to human inspection. A system for automatically detecting if workers are wearing PPE. It should detect multiple protective equipment at the same time. Thus, the system is ultimately reducing the high accident rate among construction workers in a feasible and efficient way.

This research in particular attempts to address the following question:
To what extent can a deep learning-based model be used to automatically detect if a worker is wearing PPE at the construction site. The examination area is divided into four parts: the head, the upper body, the hands, and the feet. You Only Look Once (YOLO), which is a convolutional neural networks (CNN) based algorithm is used to identify protective equipment such as helmets, highly visible clothing, gloves, and boots. The deep learning model is the main part of the PPE inspection system capable of detecting multiple protective gears.

Table 1: Functions of personal protective equipment

Device	**Description**
Helmet	Gear that protects the head and neck from objects
Ears plugs and muffs	Gear that protects ears in noisy and dusty environments
Goggles	Gear that protects eyes in environments that handle metal and chemical materials, and dust
Mask	Gear that protects the respiratory tract from dust and chemical gases, and prevents infection from viruses
Gloves	Gear that protects hands from cuts, electrocutions, and biochemicals
Boots	Gear that protects feet from heavy objects, slips, and chemicals
Protective clothing	Gear that regulates body temperature and high visibility, protects the whole body from electrocution, and chemicals

This report is structured as follows: Section 2 reviews the object detection papers using deep learning. From the point of view of machine vision, it consists of papers on the detection of defects in products and the detection of safety equipment. Section 3 describes the process for the implementation of a safety equipment detection model. Section 4 and Section 5 describe the object detection technology that is used in this project. Section 6 analyzes the results of the PPE detection model.

2 Related Work

Deep learning technology has become the key machine learning method in machine vision. It allowed the automation of object detection with high accuracy, therefore saving a considerable amount of time and effort. Previous work used deep learning techniques for the management of the working environment (Bhattarai et al.; 2020), detecting defects in products (Souza et al.; 2023), and inspecting safety equipment (Nath et al.; 2020). Previous research showed that deep learning methods performed well in the identification and detection of defective materials and safety equipment.

In this section, the outcome of research related to the management of the environment and identification of the defect is presented.

2.1 Inspecting defects using deep learning

Non-automated defect inspection, i.e., manual inspection by humans is prone to errors, and the quality of the inspection depends on the skills and experience of the operator. Several studies employed deep learning to detect defects in material. Yin et al. (2020) proposed a framework for defect inspection inside sewer pipes. The YOLOv3 model was used to inspect defects in pipes. The inside of a pipe is a specialized space that is difficult to see with the human eye, hence there is a need for an automated system. The system used a closed-circuit television (CCTV) to take videos on the inside of the pipe. The model was trained using image data extracted from the videos and identified six types of defects. Data management is essential since numerous images are generated from videos. The outcome of the training showed the average F1 score is around 0.8, and the mAP is 85.37%. The automated system provides the date and time of the inspection, the type and number of defects, and the number of defects is recorded in a text file to the user.

Wang et al. (2022) used an improved version of YOLOv7 for steel surface flaw detection. The modified YOLOv7 model uses a de-weighted bidirectional feature pyramid network (BiFPN) structure, ECA attention mechanism, and SIoU loss function. The model was trained using two datasets and used to detect ten classes of defects. It was noted that the change to the SIoU loss function resulted in improved detection rates for small-sized defects. The enhanced model showed mAP of 0.802 and 0.819 on the two test datasets, respectively. It showed high accuracy and was concluded to have better performance than the SSD algorithm, Faster R-CNN, and YOLOv5. However, for this training, grayscale images were used to detect the surface of the steel. The model was found to be less able to detect brightly colored flaws.

Souza et al. (2023) used a YOLOv5 model and a ResNet-18 classifier to detect the insulators, which are located on transmission lines. Transmission lines can be installed in remote locations, and usually, they are placed above the ground making it challenging for manual inspection. This work used unmanned aerial vehicles (UAVs) and an automated system for efficient insulator inspections. Five YOLOv5 models were compared and the YOLOv5x model showed superior performance. The results showed an F1 score of 0.962 and an mAP value of 0.992. The proposed model showed a better detection rate than those shown by the YOLOv7 model, the next model in the YOLO family.

2.2 Safety validation of equipment using deep learning

The inspection of safety equipment to protect against external hazards has been studied from different perspectives. This section reviews two types of studies. One is the detection of whether workers are wearing protective equipment. The other is the detection of whether workers are wearing protective equipment correctly.

Nath et al. (2020) explained that a protective equipment inspection system can efficiently verify the number of pieces of personal protective equipment for numberless workers, which ultimately has a positive impact on reducing worker injuries efficiently. Two pieces of protective equipment were selected: a helmet and a vest. Three different machine vision models were used Based on CNN and YOLOv3. The three models were trained to detect helmets and vests. Whereas Yung et al. (2022) used three versions of the YOLO algorithm for stable equipment detection. Different hyperparameter settings were applied to each model. Three datasets were used, and the results showed that the accuracy varied by about 7%, which shows that the dataset affects the accuracy results. The difference between datasets was the level of lightning used to capture images. In low intensity light, it was difficult to determine whether the worker was wearing a helmet or not.

Regarding the improper use of protective equipment, Li et al. (2022) demonstrated that the management of workers' protective equipment is the most efficient way to protect workers from threats on construction sites. In addition, it is one of the least expensive ways to manage the safety of workers. However, it is not enough to only provide PPE, the management of worker safety should also include the inspection of whether the workers are properly wearing the equipment provided. They proposed a monitoring system installed at a construction site to inspect whether workers are wearing PPE correctly. Two items for the inspection of the safety equipment were selected are hard hat and a harness. The suggested system was tested to check whether the worker is loosely belted on the hard hat and whether the harness is hooked. This is because the accidents that frequently occur on construction sites are those involving collisions. Workers fall or objects fall on the field, causing a significant impact on the worker and increasing the risk of injury. The monitoring system performs the inspection in two steps. First, the protective gear is detected, and then the incorrect wearing of the safety equipment is checked. The YOLOv5 algorithm and the OpenPose algorithm were used for this purpose. The YOLOv5 algorithm was used to detect helmets and harnesses. Improper use of helmets and harnesses was determined by the OpenPose algorithm. The data was obtained from a video. To check for looseness of the hat angles were measured. The threshold angle for proper use of a safety helmet is between 30 and 45 degrees.

Chen and Demachi (2021) also emphasized the need for a monitoring system for the construction site, ensuring that workers on the construction field are using protective equipment properly. This study concentrated only one gear, which is worn on the upper body. Since a mask and glasses are gears to protect against dust or chemical materials. Faster R-CNN based model and SDD-RPA model was used to identify helmets, masks, glasses, and harness. OpenPose algorithm was used to detect improper use of PPE. For identifying the posture of the worker, the system represented the body of the worker as points linked with each other to form the structure of the body.

Bhattarai et al. (2020) proposed a system that identifies hazards to keep firefighters safe while dealing with fires from image data obtained from video transmitted in real time. This research employed used RGB, thermal, and depth maps in order to detect any risk of possible falling objects during a fire. Two deep learning methods were suggested: A faster R-CNN and Mask R-CNN. Ladders, windows, doors, firefighters, and humans could be identified using Faster R-CNN. For segmenting the five items are applied Mask R-CNN. However, this study does not describe any indicators that can show the result of detection and segmentation results of the five objects. Given the environment in which the system will be used, it seems necessary to have evaluation metrics such as detection accuracy or processing speed that can be evaluated quickly in time.

Summary

From the reviewed studies, the mainstream detection object algorithm can be categorised into two categories: the one-stage detection model and the two-stage detection model (Diwan et al.; 2023). The two-stage model identifies the object location and then performs a classification. Detection models in two-stage are computationally slow but have high accuracy. The algorithms in the two-stage type use region based convolutional neural networks (R-CNN) (Girshick et al.; 2014), Fast R-CNN (Girshick; 2015), and Faster R-CNN (Ren et al.; 2015). On the other hand, in one-stage models, the process is all conducted in one step. It can have faster performance than R-CNN based models, but it has the disadvantage of being less accurate in detecting small objects (Diwan et al.; 2023). One of the models on the one-stage is YOLO (Redmon et al.; 2016).

The latest research in the field of safety equipment detection has been studied using the YOLO model. When comparing the performance of the YOLOv7 model with the Fast R-CNN model, the YOLOv7 model showed better accuracy in detecting objects (Wang et al.; 2022). The target to be detected is protective equipment that protects workers at construction sites. This is associated with a high risk of injury or death. A technology that accurately detects threats is needed. Therefore, the YOLOv8 model, which is the state-of-the-art technology is adopted.

This study considers inspecting multiple objects to protect workers in the workplace. Personal Protective Equipment (PPE) will be detected at construction sites with a higher accident rate than other industries. It will be conducted by increasing the number of safety gear to be inspected. A total of four types of personal protective gear will be detected using the YOLO algorithm. The four pieces of safety gear are a helmet, vest, gloves, and boots. Increased the range of equipment inspection to all parts of the body. That equipment can protect the head, hands, upper part, and feet of workers.

3 Methodology

This research follows the knowledge discovery in database (KDD) process (Fayyad et al.; 1996). The process consists of four steps as shown in Figure 1. A dataset is obtained from an open-source platform. It is stored locally and in the cloud. And the dataset is pre-processed in two phases: data labelling and data augmentation. Both tasks are performed on the platform that is provided in the form of a website. Once the dataset is pre-processed, use it to train the object detection model and evaluate its performance. Finally, the trained model is used to make predictions to detect which object is in the image.

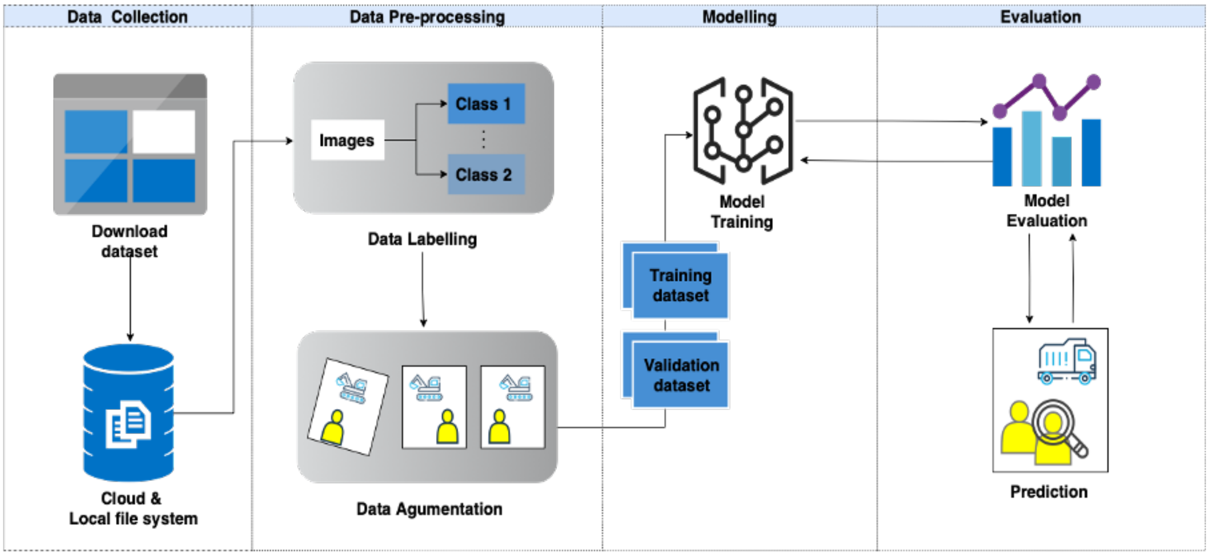

Figure 1: Research process

3.1 Data collection

The dataset is obtained from an open-source platform, which is called Roboflow (Isailovic et al.; 2022). It can be stored locally or in the cloud. In this work, the image dataset is stored both in the cloud and locally for backup. The dataset includes JPEG 640 by 640 pixel images of personal protective equipment (PPE) and people who wear a helmet, a high-visible vest, gloves, and boots on construction sites. The people in the images are positioned at different angles. They are in a certain pose, such as standing facing forward, sideways or bent over at 90 degrees, or sitting on the ground. The dataset consists of 13 classes. The classes are helmet, no-helmet, vests, vest, no-vest, head, and person, including the unnamed classes.

The image dataset is managed by 'data.yaml' so that it can be used for building the YOLO model. The dataset includes one configuration file 'data.yaml', and three directories used to store images and labels used for training, validating, and testing the model. The labels are stored in a text file that contains information about the coordinates, height, and width of the bounding box (Sesis et al.; 2022). Each row in the text file means the bounding boxes of the object. Class ID 3 in Table 2 indicates the vest which, is the yellow box, and Class ID 2 refers to the helmet, which is the blue box in Figure 2. The width and height of the yellow box are 0.52 and 0.46. X and Y represent the center value of the yellow box.

Table 2: Example of YOLOv8 data format

Class ID	X	Y	Width	Height
2	0.5578125	0.209375	0.32578125	0.253125
3	0.54609375	0.76953125	0.52265625	0.4609375

Figure 2: Bounding boxes and coordinates in YOLOv8 format

3.2 Data pre-processing

The second step is to capture bounding boxes, which is called labelling. The existing dataset has 13 classes, but only four classes are used for this study. This process is conducted manually using an open-source tool (Mostafa et al.; 2022). The classes are helmet, vest, glove, and boots. The dataset contains images with low quality resolutions that make it challenging to identify objects clearly. The bounding boxes are applied only to objects with clear boundaries. The number of images is 1,490 after labelling. In order to increase the size of the dataset, data augmentation is used. Several types of data augmentation techniques can be applied to enhance reasoning capabilities, such as adding noise or changing the brightness or angle of an image. Rotating and flipping images are typical data augmentation techniques (Sun et al.; 2022). Three techniques are applied for the dataset: Flipping, Rotation, and Noise as shown in Figure 3. In the case of the flipping technique, it is applied to the left and right flips. In fact, the angle at which the workers would be captured at the construction site is considered. Rotation also considers and uses the default range of -15 degrees to 15 degrees. The noise addition uses the default value of 5%, which means the amount of noise. This is because there are images that seem to have been blurred. By applying the augmentation techniques, the dataset is expanded as illustrated in Table 3. The total number of classes has been increased to approximately 48% from the previous one.

After the data pre-processing step, the dataset is split into three sub datasets. Train and valid datasets are for the model training. The training dataset is 3,120 images, the validation dataset is 152 images, and the test dataset is 298 images. The dataset is divided into about 85% (training), 5% (validation), and 10% (test).

Table 3: Comparison of the number of classes before and after applying augmentation

Class Name	**Before applying augmentation**	**After applying augmentation**
Helmet	2332	5424
Vest	1457	3012
Glove	1040	1812
Boots	800	1343
Total	5629	11591

Figure 3: Three data augmentation techniques

3.3 Modelling

The third step is to build a model. As noted in the previous section, previous research used the You Only Look Once (YOLO) object detection system to identify real objects in real time. In this study a YOLOv8 version which is the latest model. It is a pre-trained model based on the COCO dataset. The COCO dataset consists of 80 different categories of images, and the number of images used to train YOLOv8 is over 140,000 images. YOLOv8 model uses the size of image 640 by 640, so the images are used in their original size.

3.4 Evaluation

The last step is to evaluate the model. During the training process, the model is evaluated using the validation dataset to examine the loss function, precision, recall, and mAP metrics. The loss function allows the model's training result to be evaluated. The test dataset is used to examine how well the model predicts. Mean Average Precision (mAP), and F1 score are used as performance indicators.

4 Design Specification

As shown in Figure 4, it is designed to build the PPE detection system. The labelled images are trained using the YOLOv8 model, the predictions are performed with the trained model, and accuracy is calculated. The YOLOv8 model uses the Stochastic Gradient Descent (SGD) optimizer by default. AdamW optimizer, which is one of the SGD methods, learning rate of 0.01, momentum of 0.937, and weight decay of 0.0005 as basic parameters to start training. The batch size is specified as 8 and epochs is 100. YOLOv8 has five models: YOLOv8n (Nano), YOLOv8s (Small), YOLOv8m (Medium), YOLOv8l (Large), and YOLOv8x (Extra Large). The YOLOv8x model has the highest accuracy among the YOLO models (Talaat and ZainEldin; 2023). Therefore, the YOLOv8x model is adopted in this project.

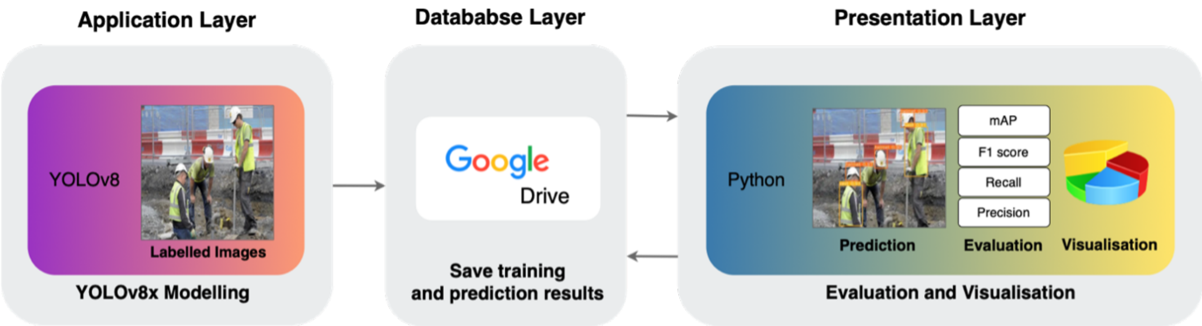

Figure 4: Design of the PPE detection application

	Index	From	Repeats	Module
Backbone	0	-1	1	conv.Conv
	1	-1	1	conv.Conv
	2	-1	3	block.C2f
	3	-1	1	conv.Conv
	4	-1	6	block.C2f
	5	-1	1	conv.Conv
	6	-1	6	block.C2f
	7	-1	1	conv.Conv
	8	-1	3	block.C2f
	9	-1	1	block.SPPF
Head	10	-1	1	upsampling.Upsample
	11	[-1, 6]	1	conv.Concat
	12	-1	3	block.C2f
	13	-1	1	upsampling.Upsample
	14	[-1, 4]	1	conv.Concat
	15	-1	3	block.C2f
	16	-1	1	conv.Conv
	17	[-1, 12]	1	conv.Concat
	18	-1	3	block.C2f
	19	-1	1	conv.Conv
	20	[-1, 9]	1	conv.Concat
	21	-1	3	block.C2f
	22	[15, 18, 21]	1	head.Detect

Figure 5: Layers of YOLOv8x model

Figure 5 shows the number of layers and types of the YOLOv8x model. It has a total of 365 layers. YOLOv8 is composed of two sections: Backbone and Head network. Backbone is the extraction of features from the images, and Head is to find where objects might be located. From element in Figure 5 is the index of the previous layer which is the input for the current layer. Repeat element means how many times the current layer will be repeated. Module element shows their use in the current layer, and they use different parameters. For example, the Conv module in the backbone section has [the number of input channels, the number of output channels, kernel size, stride] arguments. The channel parameter is related to the shape of an image, the kernel size parameter is the size of the convolutional matrix, which is used to extract features, and the stride parameter is how many spaces the matrix is moved horizontally and vertically. In YOLOv8, the C2f module has adapted to optimize the network and execution speed.

5 Implementation

The training and prediction process is conducted on Kaggle Kernels and Google Colab Notebook platforms respectively. They are both open-source platforms and give access to GPUs which accelerate computational time during training.

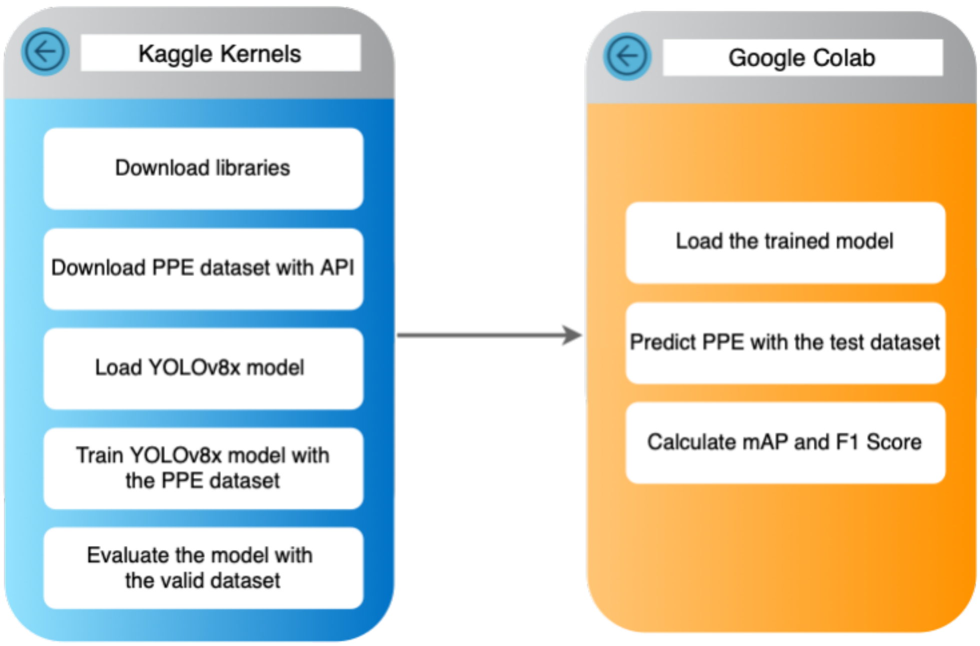

Figure 6: PPE detection model implementation diagram

The overall components of the suggested system to detect multiple objects are in Figure 6. All the components of the system are implemented in Python language. The packages used to preprocess the dataset are Roboflow and Ultralytics libraries. As stated earlier the dataset is downloaded directly into the Kaggle Notebook. A copy of the original dataset is kept in locally a personal workspace. It was found that image files might be broken or missed during the uploading process and might cause problems during the training phase of the model. In case of a missing image in the Kaggle cloud storage, the system fetches it from the local machine. The model is trained using the default parameter values as mentioned in the previous section. In addition, 10 of the save_period parameter is used. This option saves the training result in the form of a PT file every ten epochs. After the training step of the YOLOv8x model, it is loaded into the Google Colab.

mAP and F1 score are calculated to evaluate prediction performance. Ultralytics library that has the YOLOv8 model does not support the calculation function of the mAP and F1 score metrics. By converting the YOLOv8 format in Table 2 to the PASCAL VOC format, the mAP result is obtained with an external library, including Precision and Recall results. The result from the library contains Recall and Precision values per class. These are used to calculate the F1 score as shown in algorithms in Figures 7 and 8.

```
Algorithm 1: Getting Precision and Recall values from text file pseudocode
Input: S
Output: N1 … Nn
Def GetMetricsValues(S)
    '[' and ']' in S are replaced by ''
    S is split by ':' into the size of 2 array.
    N ← Second element of S is split by ','
    Convert the type of N into integer type
return N
```

Figure 7: Getting Precision and Recall values pseudocode

```
Algorithm 2: Calculating F1 score pseudocode
Input: P, R
Output: Class name, sum of score S, number of score T, F1 score F
C ← Read all text from the results.txt file
C ← Remove white spaces from the lead and tail of C
P ← []
R ← []
for i in 0 to length(C)-1 do
    if 'Precision:' in C[i] then
        P ← P + GetMetricsValues(C[i])
    if 'Recall:' in C[i] then
        R ← R + GetMetricsValues(C[i])
end
for i in 0 to length(P)-1 do
    T ← length(P)
    for j in 0 to T-1 do
        p ← P[i][j]
        r ← R[i][j]
        S ← S + (2 * p * r) / (p + r)
    end
    F ← S/T
    print(i, S, T, F)
end
```

Figure 8: Calculating F1 score pseudocode

6 Evaluation

For evaluating the YOLOv8x model four metrics are calculated after validation and test steps. The four metrics are Recall, Precision, mAP, and F1 score calculated using Equations (1), (2), (3) and (4). Recall metric shows how well the four pieces of protective gear, helmet, vest, glove, and boots, are detected by the YOLOv8x model. Precision shows the accuracy of the detected results. True Positive can be described as a case where the workers are wearing personal protective equipment (PPE), and the prediction also indicates that they are wearing PPE. False Positive means when the workers are not wearing PPE, but the prediction indicates that they are. False Negative is when the workers are wearing PPE, but the prediction shows that they are not. They are defined by an Intersection Over Union (IOU) and a specific threshold. The range is from 0 to 1. A higher value indicates that the prediction is accurate. IOU shows a region that is intersected between the bounding box that describes the location of the object and the predicted bounding box, as shown in Figure 2. If the threshold value is set to 0.5, and the overlapping area of the two boxes is greater than 0.5, it is a case of True Positive.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{1}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2}$$

$$F1\ score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{3}$$

The Average Precision (AP) is calculated based on Recall and Precision metrics. The Mean of the Average Precision (mAP) in Equation 4 is calculated using the AP values, which are calculated for all classes. n element means the number of classes, and k means classes. In this study, k corresponds to the four classes: helmet, vest, glove, and boots. YOLO model uses mAP50 and mAP50-95 evaluation. They show the detection results with numeric values. 50 and 50-95 means the IOU threshold. mAP50-95 is when the IOU threshold is from 0.5 to 0.95.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \tag{4}$$

6.1 Result & Discussion

The PPE detection system is based on the YOLOv8x model. Figure 9 and Table 4 are the training results. Figure 9 shows the loss value plotted against epochs for the training and validation step. It includes YOLOv8 three different loss functions: bounding box loss, class loss, and dual focal loss (DFL) loss. The rate of decrease of the loss function is fast during the first 20 epochs for both the training and validation plots. The rate of decrease of the loss function slows down after epoch 20 in the training while the validation function fluctuates, and dual focal loss shows an increasing trend from around the 25th epoch. This means that the accuracy of the prediction in the bounding box is not constant and shows the potential for imbalance or misclassification of data. This suggests that if the training exceeds 25, it might have an overfitting situation. Meanwhile, in Table 4, all four classes have a value of 0.9 or higher for mAP50 which verifies the performance of the model with a specific number.

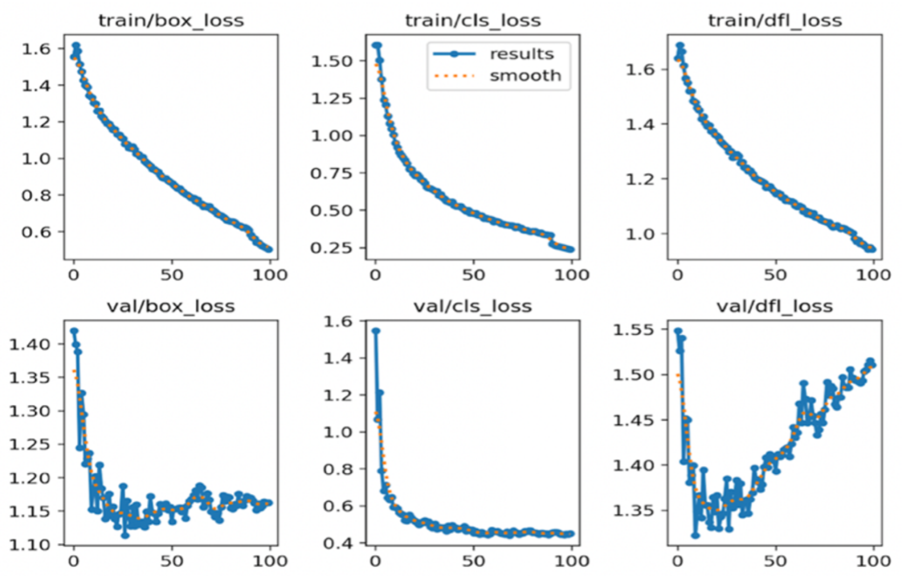

Figure 9: The class percentages of the training dataset

Table 4: Model performance metrics

Class	Precision	Recall	mAP50	mAP50-95
Helmet	0.972	1	0.989	0.74
Vest	0.978	0.968	0.982	0.738
Glove	0.981	0.965	0.986	0.652
Boots	0.892	0.846	0.9	0.652
All	0.956	0.945	0.965	0.695

The final model evaluation is to use the test dataset to perform the prediction of the different classes. The PPE detection system classified the four pieces of safety equipment and the result of prediction performance as shown in Table 5.

The overall mAP value is 0.74. The accuracy of helmet and vest classes is approximately 0.9 AP, and the F1 score is 0.6. However, glove and boots classes obtained 0.54 and 0.50 AP respectively, and they have under 0.5 F1 score. It indicates that the detection performance of gloves and boots is inferior to that of helmets and vests. Figure 10 shows an example of True Positive and False Positive cases. The left-side and centre images are predicted correctly. The workers on the construction sites are wearing a helmet, a vest, and boots. The prediction bounding boxes show that the workers are wearing PPE with the confidence score. The level is a value between 0 and 1, explaining that the closer to 1, the more likely it is that the predicted value is equal to the actual value. The right-side image is the model's incorrect predictions. A non-gloved object is predicted to be a glove. The score is 0.69, which is a high score. This is because the glove objects are not detected correctly, and the model seems to need more training with more glove classes.

Table 5: Prediction result evaluation

Class	AP	F1 score
Helmet	0.922	0.601
Vest	0.933	0.591
Glove	0.540	0.420
Boots	0.507	0.426

Figure 10: Example of PPE detection

7 Conclusion and Future Work

This project is conducted to detect threats to workers working on construction sites. Failure to wear safety equipment is determined as a hazard on construction sites that can lead to injury to workers and, in serious cases, death. This is done by defining non-wearing personal protective equipment (PPE) as a threat to workers. Four items are detected to protect the entire body. These four pieces of protective gear protect the head and feet from falling objects, protect the hands from debris and chemicals, and identify workers on the construction site where construction materials and heavy equipment are mixed. The state-of-the-art technology, the YOLOv8x model, is trained to detect four types of PPE. Results ranged from a high of 0.6 to a low of 0.42 for the F1 score, which is one of the object detection model evaluation metrics. According to the latest research was used the YOLOv7 model, the F1 score reached 0.9.

As shown in Figure 10, boots objects with relatively small sizes are identified. However, the overall accuracy is low, and there are differences in accuracy between classes. When the results are analyzed, it can be assumed that this is due to an imbalance of the dataset as shown in Figure 11 or the influence of manual labelling. Labeling can be performed using automatic ways rather than manually, and datasets can be collected with class balancing is considered for the future experiment. In addition, this study is modeled and tested using image files. In the future project, image data will be acquired from the video and the trained model can be tested using the video. Videos allow the PPE inspection system to monitor workers in real time. The system can be installed at the entrance of construction sites to detect workers with PPE or to be used before executing certain tasks.

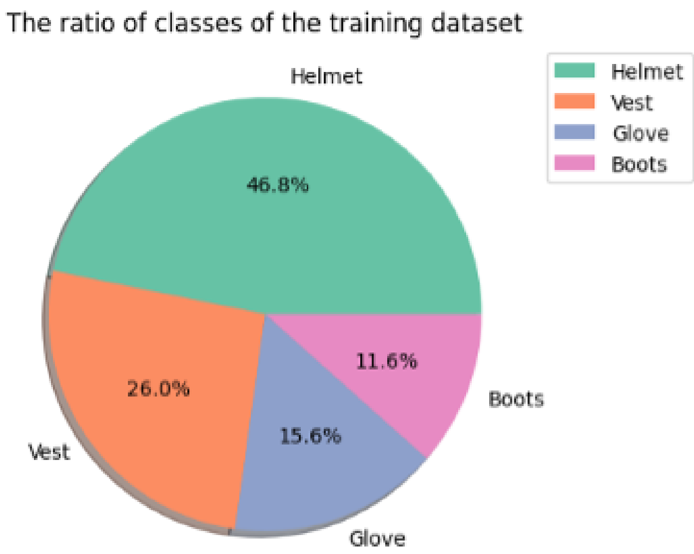

Figure 11: The class percentages of the training dataset

References

Bhattarai, M., Jensen-Curtis, A. R. and Martinez-Ramon, M. (2020). An embedded deep learning system for augmented reality in firefighting applications, *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, Miami, FL, USA, pp. 1224–1230.
URL: *https://ieeexplore.ieee.org/document/9356175/*

Chen, S. and Demachi, K. (2021). Towards on-site hazards identification of improper use of personal protective equipment using deep learning-based geometric relationships and hierarchical scene graph, *Automation in Construction* **125**: 103619.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0926580521000704*

Diwan, T., Anirudh, G. and Tembhurne, J. V. (2023). Object detection using YOLO: challenges, architectural successors, datasets and applications, *Multimedia Tools and Applications* **82**(6): 9243–9275.
URL: *https://link.springer.com/10.1007/s11042-022-13644-y*

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data, *Communications of the ACM* **39**(11): 27–34.
URL: *https://dl.acm.org/doi/10.1145/240455.240464*

Girshick, R. (2015). Fast r-cnn, *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448.

Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587.

Gunduz, M. and Ahsan, B. (2018). Construction safety factors assessment through Frequency Adjusted Importance Index, *International Journal of Industrial Ergonomics*

64: 155–162.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0169814117303347*

Isailovic, V., Peulic, A., Djapan, M., Savkovic, M. and Vukicevic, A. M. (2022). The compliance of head-mounted industrial PPE by using deep learning object detectors, *Scientific Reports* **12**(1): 16347.
URL: *https://www.nature.com/articles/s41598-022-20282-9*

Jalil Al-Bayati, A., Rener, A. T., Listello, M. P. and Mohamed, M. (2023). PPE non-compliance among construction workers: An assessment of contributing factors utilizing fuzzy theory, *Journal of Safety Research* **85**: 242–253.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0022437523000233*

Li, J., Zhao, X., Zhou, G. and Zhang, M. (2022). Standardized use inspection of workers' personal protective equipment based on deep learning, *Safety Science* **150**: 105689.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0925753522000297*

Mostafa, T., Chowdhury, S. J., Rhaman, M. K. and Alam, M. G. R. (2022). Occluded Object Detection for Autonomous Vehicles Employing YOLOv5, YOLOX and Faster R-CNN, *2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, Vancouver, BC, Canada, pp. 0405–0410.
URL: *https://ieeexplore.ieee.org/document/9946565/*

Nath, N. D., Behzadan, A. H. and Paal, S. G. (2020). Deep learning for site safety: Real-time detection of personal protective equipment, *Automation in Construction* **112**: 103085.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0926580519308325*

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016). You only look once: Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788.

Ren, S., He, K., Girshick, R. and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks, *Advances in neural information processing systems* **28**.

Sesis, A., Siniosoglou, I., Spyridis, Y., Efstathopoulos, G., Lagkas, T., Argyriou, V. and Sarigiannidis, P. (2022). A Robust Deep Learning Architecture for FireFighter PPEs Detection, *2022 IEEE 8th World Forum on Internet of Things (WF-IoT)*, IEEE, Yokohama, Japan, pp. 1–6.
URL: *https://ieeexplore.ieee.org/document/10152263/*

Shamsuddin, K. A., Mohd Norzaimi Che Ani, A. K. I. and Ibrahim, M. R. (2015). Investigation the safety, health and environment (SHE) protection in construction area, *International Research Journal of Engineering and Technology (IRJET)* **2**(6).

Souza, B. J., Stefenon, S. F., Singh, G. and Freire, R. Z. (2023). Hybrid-YOLO for classification of insulators defects in transmission lines based on UAV, *International Journal of Electrical Power & Energy Systems* **148**: 108982.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S014206152300039X*

Sun, X., Jia, X., Liang, Y., Wang, M. and Chi, X. (2022). A Defect Detection Method for a Boiler Inner Wall Based on an Improved YOLO-v5 Network and Data Augmentation Technologies, *IEEE Access* **10**: 93845–93853.
URL: *https://ieeexplore.ieee.org/document/9880480/*

Talaat, F. M. and ZainEldin, H. (2023). An improved fire detection approach based on YOLO-v8 for smart cities, *Neural Computing and Applications* .
URL: *https://link.springer.com/10.1007/s00521-023-08809-1*

Wang, Y., Wang, H. and Xin, Z. (2022). Efficient Detection Model of Steel Strip Surface Defects Based on YOLO-V7, *IEEE Access* **10**: 133936–133944.
URL: *https://ieeexplore.ieee.org/document/9994733/*

Wong, T. K. M., Man, S. S. and Chan, A. H. S. (2020). Critical factors for the use or non-use of personal protective equipment amongst construction workers, *Safety Science* **126**: 104663.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0925753520300606*

Yin, X., Chen, Y., Bouferguene, A., Zaman, H., Al-Hussein, M. and Kurach, L. (2020). A deep learning-based framework for an automated defect detection system for sewer pipes, *Automation in Construction* **109**: 102967.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0926580519307411*

Yung, N. D. T., Wong, W. K., Juwono, F. H. and Sim, Z. A. (2022). Safety Helmet Detection Using Deep Learning: Implementation and Comparative Study Using YOLOv5, YOLOv6, and YOLOv7, *2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, IEEE, Miri Sarawak, Malaysia, pp. 164–170.
URL: *https://ieeexplore.ieee.org/document/10010490/*