Tech Science Press

Real-Time Safety Helmet Detection Using Yolov5 at Construction Sites

Kisaezehra[1], Muhammad Umer Farooq[1,*], Muhammad Aslam Bhutto[2] and Abdul Karim Kazi[1]

[1]Department of Computer Science and Information Technology, NED University of Engineering & Technology, Karachi, 75270, Pakistan
[2]Department of Civil Engineering, NED University of Engineering & Technology, Karachi, 75270, Pakistan
*Corresponding Author: Muhammad Umer Farooq. Email: umer@neduet.edu.pk

Abstract: The construction industry has always remained the economic and social backbone of any country in the world where occupational health and safety (OHS) is of prime importance. Like in other developing countries, this industry pays very little, rather negligible attention to OHS practices in Pakistan, resulting in the occurrence of a wide variety of accidents, mishaps, and near-misses every year. One of the major causes of such mishaps is the non-wearing of safety helmets (hard hats) at construction sites where falling objects from a height are unavoidable. In most cases, this leads to serious brain injuries in people present at the site in general and the workers in particular. It is one of the leading causes of human fatalities at construction sites. In the United States, the Occupational Safety and Health Administration (OSHA) requires construction companies through safety laws to ensure the use of well-defined personal protective equipment (PPE). It has long been a problem to ensure the use of PPE because round-the-clock human monitoring is not possible. However, such monitoring through technological aids or automated tools is very much possible. The present study describes a systematic strategy based on deep learning (DL) models built on the You-Only-Look-Once (YOLOV5) architecture that could be used for monitoring workers' hard hats in real-time. It can indicate whether a worker is wearing a hat or not. The proposed system uses five different models of the YOLOV5, namely YOLOV5n, YOLOv5s, YOLOv5 m, YOLOv5l, and YOLOv5x for object detection with the support of PyTorch, involving 7063 images. The results of the study show that among the DL models, the YOLOV5x has a high performance of 95.8% in terms of the mAP, while the YOLOV5n has the fastest detection speed of 70.4 frames per second (FPS). The proposed model can be successfully used in practice to recognize the hard hat worn by a worker.

Keywords: Object detection; computer-vision; personal protective equipment (PPE); deep learning; industry revolution (IR) 4.0; safety helmet detection

1 Introduction

The construction industry is a vital part of the economy of a country with several backward and forward links to other industries. This industry makes a considerable contribution to economic growth and

employment opportunities, but several common difficulties affect the industry in developing and emerging countries.

Pakistan's construction industry has been contributing tremendously towards economic recovery as well as growth through increased employment opportunities in particular. Fig. 1 shows the employability rate of the industry that touched 25.79 percent in 2020 [1]. The construction industry equally contributes up to 2.53 percent of the country's Gross Domestic Product (GDP) to the tune of up to 380 billion PKR and 7.61% of Pakistan's employment in this area [2]. However, there are a lot of problems and challenges being faced by the construction industry in Pakistan, with a prominent one being the occurrences of accidents, mishaps, and near misses. One of the major reasons for this occurrence is a little or negligible attention to the occupational safety and health (OSH) practices at the worksites both in the public and private sectors. Atif Mehmood (2020) studied the causes of accidents in Pakistan's construction industry, their impacts, and preventative measures. It has been found that the highest among the top ten accidents reasons is lack of safety rules and work procedures with an alarming 74% [3]. In 2012, a large fire overtook the factory located in the Baldia area of Karachi, causing the deaths of 258 workers for the obvious reason that safety rules were completely ignored in terms of adequate fire exits [4]. As per the International Labor Organization (ILO) data, 2.20 million people die every year as a result of work-related accidents or illnesses, with over 270 million employees injured and approximately 160 million affected by work-related illnesses. The Decent Work Country Programs (DWCP) focuses on the production and deployment of workplace safety and health systems for unprotected workers [5].

Figure 1: Pakistan industry employment [1]

The US economy relies heavily on construction, accounting for $1.3 trillion in annual spending 6.3 percent of GDP and 7.2 million workers (5 percent of the overall workforce). However, because of the high incidence of work and worker injuries, this industry is considered one of the deadliest of all [6]. According to the latest research from the Center for Construction Training and Research (CPWR), the number of workers who died on the job hit its highest level in at least nine years in 2019. Results showed 1,102 worksite deaths in 2019 using data from the Census of Fatal Occupational Injuries from 2011 to 2019. This represents a 41.1% increase over the first year of the observation period [7].

It can be pretty hard to avoid personal injury when you're surrounded by heavy machinery, goods being loaded and unloaded, and seemingly new hazards and impediments at every step. Construction injuries and fatalities are caused by a variety of factors, such as fires, explosions, electrocutions, and falling objects. If a worker gets stuck in or between any of these objects, he or she risks being crushed and maybe dying. Another is Repetitive strain injuries can be caused by a repetitive motion that puts pressure on the joints and muscles of the body [8]. From a security point of view, the bulk of these injuries and deaths in the construction world

could have been avoided if workers had been wearing proper personal protective equipment (PPE) because they are used to ensure the health, safety, and protection of workers. Workers are generally considered responsible for enforcing, monitoring, and ensuring proper PPE on the job site, according to the governing laws and safety standards of OSHA. The helmet, or hard hat, is an essential PPE that protects laborers by resisting items and absorbing damage from direct object hits to the head [9]. Earlier research has shown that wearing a helmet while falling from a height can minimize the risk of traumatic injuries, neck sprains, and concussions. According to the study, construction workers suffer more traumatic brain injuries than laborers in any other sector of work in the United States.

Traumatic Brain Injury (TBI) accounted for 11%–61% of all head injuries suffered by employees in Sweden and Germany between 2014 and 2018. Falling objects accounted for 35% of all TBIs, making them the most common type of fatal injury. In North American studies, falling objects are also thought to be the cause of 52%–78% of TBI [10]. Aside from the face, helmet energy absorption primarily covers the head, with TBI accounting for 49%–62% of all injuries, highlighting the importance of helmet testing standards that include TBI protection in addition to skull fractures. According to a report from the United Kingdom, 142 people were killed in workplace accidents in 2020 and 21. Falls, as illustrated in Fig. 2, are the most common type of fatal workplace accident [11]. So, monitoring the helmet can be a promising step to reducing injuries and even fatalities in case of accidents. This protective equipment can also prevent TBI occurrences and improve the management's efficiency.
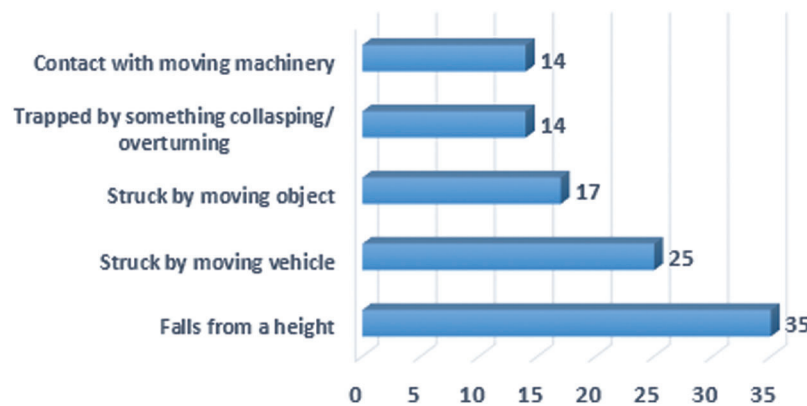

Figure 2: Fatal accident (2020 and 2021) [11]

There are two types of automated PPE compliance monitoring techniques, namely: sensor based and vision based. Details of background work are explained in the literature review section. The key safety factors linked to work related TBI in the construction industry can largely be overcome by using helmets or hard hats. The YOLOv5 models can be used to detect whether or not workers are wearing PPE helmets during the experiment.

In summary, this paper has the following contributions:

1. There are 7063 images from the Harvard Dataverse, and they were annotated. Annotated as "head" if workers are not wearing hard hats and "helmet" if they are.
2. Manual monitoring of a large workforce for hard hats is not a viable solution because it is costly, time-consuming, human-dependent, and resource-intensive. Therefore, this research has looked into the possibility of automating helmet detection.

3. The model has presented an automated as well as a cost-effective solution to overcome the high death rate during work by detecting hard hats and heads due to falling from heights, being struck by objects, and any other circumstances.

4. Making real-time safety through automated hardhat detection is more accessible and easier on construction sites.

5. Both the YOLOV3 and YOLOV4 models have been extensively used but still need improvement in their work. So, to determine whether the worker was wearing a helmet or not, apply the most advanced model to fill this gap, using the deep learning (DL) method YOLOV5 (n, s, m, l, x) detection algorithm and train different models of it, then make comparisons based on mAP, precession, and recall.

If workers are not wearing hard hats, automatic detection of the head alerts them as well as their monitors that their lives are in danger. Therefore, the detection of hard hats is of significant importance. The rest of this work is arranged in the following manner: Section 2 introduces the literature review, Section 3 presents the YOLOv5 method, and Section 4 describes the experimental details and outcomes followed by the conclusions in Section 5.

2 Literature Review

The literature review presented in this section has laid the groundwork for detecting personal protective equipment and developing guidelines to improve construction sector safety. Traumatic brain injuries (TBI) and collisions are the main causes of construction deaths. As a preventive measure, OSHA calls for contractors to ensure that employees' proper personal safety equipment (PPE) is used and monitored.

Sensor-based techniques involve the installation of a sensor and the analysis of its signals. BarroTorres et al. [12] present a cyber-physical system (CPS). Each worker uses a body area network (BAN) of sensors in the system. To identify whether each employee uses the required PPE, check their attendance, and warn the employee if they have not used it adequately. Meanwhile, Kelm et al. [13] set up Radio Frequency Identification (RFID) tags on each PPE component and inspected them with a scanner at the job site's gate to ensure that workers were carrying proper PPE. Similarly, Naticchia et al. [14] for the feasibility of the real-time monitoring system in terms of health and safety management, every PPE item is fitted with short-range transponders and a wireless system analyses them. This is communicating via radio frequency (RF). The sensor-based approach, however, requires considerable investment in the procurement, installation, and maintenance of complicated sensor networks to prevent its practical implementation.

Since vision-based approaches use cameras to record images or videos from the worksite and then analyze them to verify compliance with the PPE system, there has been rapid advancement in deep learning research in vision-based fields in recent years.

Although Du et al. [15] proposed a combined method for hard hat detection in video using machine learning and image processing, the first one is face sensing based on Haar like features. The second is the filtering of movement, and the third is the color detection of hard hats. False alarms can hardly be distinguished from faces. However, it is easy to filter out by color information. Meanwhile, Shrestha et al. [16] proposed a set of CCTV cameras inserted at the site, first detecting the face with a Haar like feature, and second, detecting the hard hat by filling in two conditions: the hat is red and the outline of the hat, or when a worker is detected on the site with no hard hat, warning messages are sent out from the program. Furthermore, to gain a better understanding of the current state of computer vision techniques. Sun et al. [17] discover that it is difficult to identify small objects collected by UAVs. To address this problem, they invented a small object detection system relying on YOLOv3, which enhances small object detection performance through the use of location prediction and the Feature Pyramid Network (FPN).

Similarly, Redmon et al. [18] proposed YOLO, a model for object detection. Unlike classifier-based methods, Yolo trained in the full picture right away. The input image is divided into an S × S grid by our method. If the object's center falls inside a grid cell, that grid cell is in charge of detecting the object. Every grid cell has bounding boxes and confidence scores for those boxes. Fast YOLO is a new framework. It speeds up YOLOv2 to conduct object identification in a video in real-time on embedded devices. Furthermore, Shafiee et al. [19] the major goal is to offer a framework for object recognition in videos that can be performed quickly on embedded devices while consuming fewer resources, resulting in a significant reduction in power consumption. They suggested Fast YOLO can be divided into the following two major components: 1) Optimization of the YOLOv2 architecture; 2) Motion-adaptive inference. It receives an image stack comprised of the video frame and a reference frame. It is ± 3.3X faster than the previous YOLOv2 framework. Meanwhile, Buric et al. [20] work on the comparison of Yolo and Mask R-CNN. YOLO's working method is to divide the input into possible boundary boxes, which extract convolutional features from it. Mask R-CNN provides instance segmentation, which means that the object is not localized with a bounding box, but rather that the given output consists of the exact pixel number and location of the wanted object and each object is marked with its color. The YOLO F1 score achieved in the dataset increased from 6% to 34%. Mask R-CNN's recall did not improve as much as YOLO's.

Du [21] projected the CNN family and YOLO are two modern object detection algorithms that are discussed briefly. YOLOv2 offers optimal speed and precision in terms of real-time object recognition, reaching up to 78.6 mAP. Furthermore, Fang et al. [22] create an improved, Faster R-CNN that can be used to indicate the existence of objects on construction sites on a real-time basis. There are two modules: one is an improved faster R-CNN (IF R-CNN) and the second module is a Fast R-CNN detector. Although the Faster R-CNN is not the fastest calculation method, it is the most accurate. The Faster R-CNN cannot detect small objects accurately. The IF-RCNN method finds that a high confidence threshold is obtained. Mneymneh et al. [23] aim to evaluate current computer vision techniques, particularly object detection and color-based tools. The processes are: 1) detection of features, removal, and matching 2) template matching and 3) cascade models. The number of characteristics recognized was low because of the consistent shape and color of a hard hat. The problem was solved by adding a sticker to the hardhat. Furthermore, Wu et al. [24] aim to propose a convolutional neural network to automatically detect whether construction workers are wearing hardhats and identify the corresponding colors. They propose a one-stage, data-driven CNN method based on the SSD framework. In the trained model SSD-RPA512 with a 512*512 input, the best result was an 83.89% mAP with a considerably better than the SSD512 frame. Similarly, Hung et al. [25] presented a new framework for automatic monitoring of the complete staffing of the required PPE. The way you identify pictures of real-time protection products is based on the YOLO algorithm. Meanwhile, Chuan et al. [26] proposed that YOLOV3 be based on an enhanced method. Improvement can be done by following these steps. The first is to forecast and acquire a candidate box of safety helmets by the k-mean clustering algorithm, the second step is Combining a deep residual network with multi-scale detection training improves accuracy, and the last step is to adjust the weight of the loss function to optimize target box selection. It boosts the accuracy of small object detection. Another work done by Ayrapetov et al. [27] is to set up and operate the DarkNet Yolo V3 and V2 neural networks' software and hardware, which allows them to recognize things quickly.

Similarly, Saudi et al. [28] suggested the Faster Region-based Convolutional Neural Networks (R-CNN) algorithm provides an image scanning model for worker protection based on conformity with the PPE. Meanwhile, Wang et al. [29] Cases of COVID-19 are identified using a chest X-ray. The authors have proposed a convolutional neural network, CFW-Net, based on the channel feature weight extraction (CFWE) module, to accomplish this. For COVID-19 cases, it has extremely high accuracy and sensitivity. Additionally, Long et al. [30] present a new object implementation of the Paddle Paddle detector, known

as the PP-YOLO detector. DarkNet-53 appears for the first time in the original YOLOv3 used to extract a feature, so replacing it with ResNet50-vd results in PPYOLO being faster. So the RestNet model is better than DarkNet. Furthermore, Zhang et al. [31] present an enhanced region-based fully convolutional network (R-FCN) for improving small object accuracy rate and eliminating localization misalignment by using position-sensitive precise region of interest (PS-Pr-RoI) pooling. In addition, Zhao et al. [32] suggest that the YOLOv3 technique takes time to handle huge datasets, so it is time-consuming. Thus, use the innovative approach of clusters. First, a few height and width values are selected randomly; second, selects cluster centers. It has a shorter runtime, specifically for huge datasets, as compared to the k-mean technique. Furthermore, Tun et al. [33] projected the use of a hybrid approach to integrate deep learning and traditional methods for effective safety. YOLO detects the entire image, and YOLO's ROI is converted to HOG, and the trained SVM model is to determine whether the workers are wearing a security helmet. Additionally, Shukla et al. [34] suggested the YOLOv5 model, to recognize people in a video. The extraordinary thing about this deep neural network is that it is very simple to retrain the network on your custom dataset. Furthermore, Zhou et al. [35] to establish a digital safety helmet monitoring system, suggest a method based on YOLOv5. The results show that the mAP of YOLOv5x reaches 94.7%, demonstrating the effectiveness of YOLOv5-based helmet detection. Wang et al. [36] used You Only Look Once (YOLO) architecture to train eight deep learning detectors, including helmets with four colors, people and vests. The result shows that YOLO v5x has the best mAP (86.55%), and YOLO v5s has the fastest speed on GPU.

In 2020, Yang et al. [37] showed a comparison of the results of mask detection on different models and found that the different accuracy rates are 70.40%, 77.60%, 84.60%, and 97.90% for Faster RCNN, R-FCN, SDD, and YOLOV5 respectively. In addition, Wu et al. [38] offer a novel neural network structure called Yolov5-Ghost, which is based on the current Yolo v5s neural network architecture. It loses about 3% of its mAP value, and the frame rate has been increased. Furthermore, Protik et al. [39] proposed a method by using the YOLOv4 object detection model to find some PPE for COVID-19. Similarly, Torres et al. [40] proposed four possible classes to perform the classification for which two approaches are used. The first one is YOLO-v4-AP1, which combines the identification and verification processes, and the second approach, YOLO-v4-AP2, applies a multi-stage strategy to identify and verify proper PPE usage by workers. Yet, the aforementioned detection-based can only be employed in limited conditions and has a low level of accuracy. To address the shortcomings of earlier efforts, this study provides a novel automatic hardhat and head detection method based on the Yolov5 framework.

3 Methodology

In real-time object detection, there are a variety of algorithms and models to choose from. Among them, the YOLOv5 model has been chosen for this study because it provides excellent and reliable results in terms of real-time object detection, as well as this deep neural network is very simple, user-friendly, and customized.

When compared to YOLO v3, YOLO v5 provides a faster speed of up to 40 frames per second (FPS) with one-third of the targeted model size and a 60%–70% accuracy in the results. There are five main network models in Yolov5 namely Yolov5n, Yolov5s, Yolov5l, Yolov5m, and Yolov5x [41]. The approach employed throughout this study comprises four primary steps, namely: data collection, development of the proposed model, model training, and finally, testing of the proposed approach. In addition to the methods, the entire structure of the YOLOV5 network has also been explained.

3.1 Network Architecture of Yolov5

In this article, the YOLOV5 model is used as the fundamental model for hard-hat detection, and its framework is given in Fig. 3. YOLOv5 is small and fast, requiring fewer computing resources than other

existing models. It is significantly faster than previous YOLO versions [42]. Yolo is used for real-time object detection. YOLOv5 is ultralytics' most recent object detection model. It was released on June 25, 2020, by Glenn Jocher [43]. The majority of YOLOv5's performance gains come from PyTorch training procedures. YOLOv5's most significant contribution is the conversion of the Darknet methodological framework to the PyTorch framework. On PyTorch the training process is more user-friendly than it was on Darknet. YoloV5's network is divided into three sections: backbone, neck, and output head.

The basic purpose of Model Backbone is to extract key features from an input image. The CSP (Cross Stage Partial Networks) as a backbone is utilized in YOLO v5 to extract rich features out of a given image. The input image with $640 \times 640 \times 3$ resolution passes via the focus structure in the backbone. It becomes a $320 \times 320 \times 12$ feature map after the slicing process, and after that, a $320 \times 320 \times 32$ feature map after a convolution operation with 32 convolution kernels. The CBL is a simple convolution module. In the hidden layers of YOLO v5, only the leaky-relu activation function (CBL module) is used [44].

The Model Neck is primarily used to create feature pyramids. Feature pyramids aid models in generalizing well when it comes to object scaling. It aids in the identification of the same object in various sizes and scales. Path Aggregation Network (PANET) is used as a feature pyramids throughout the YOLOV5 model. The model head is primarily used for the last stage of detection. It creates final output vectors with class probabilities and bounding boxes after applying anchor boxes to features [45].

The bounding box regression loss function used by YOLO V5 is the Generalized Inter-section over Union loss (GIoU loss). GIoU is used to correct the prior IOU loss function's incorrect estimation of non-overlapping bounding boxes. The GIoU loss maximizes the area of overlap between the ground-truth and predicted bounding boxes. It is calculated by using Eqs. (1) and (2) [38].
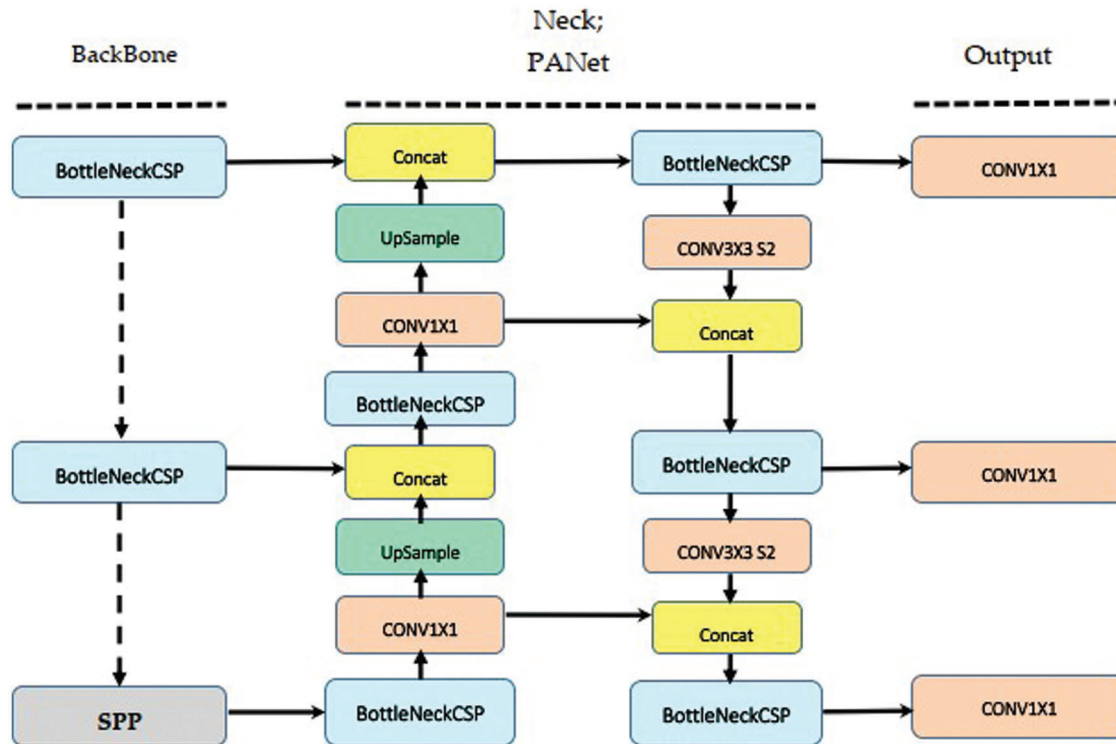

Figure 3: The network architecture of yolov5 [46]

$$IoU = \frac{|A \cap B|}{|A \cap B|} \tag{1}$$

$$GIoU = 1 - IoU + \frac{|C \setminus (A \cap B)}{|C|} \tag{2}$$

In Eqs. (1) and (2), A is the predicted box, B is the ground-truth box and C is the smallest box covering A and in Fig. 4 depicts the system's operational flow.
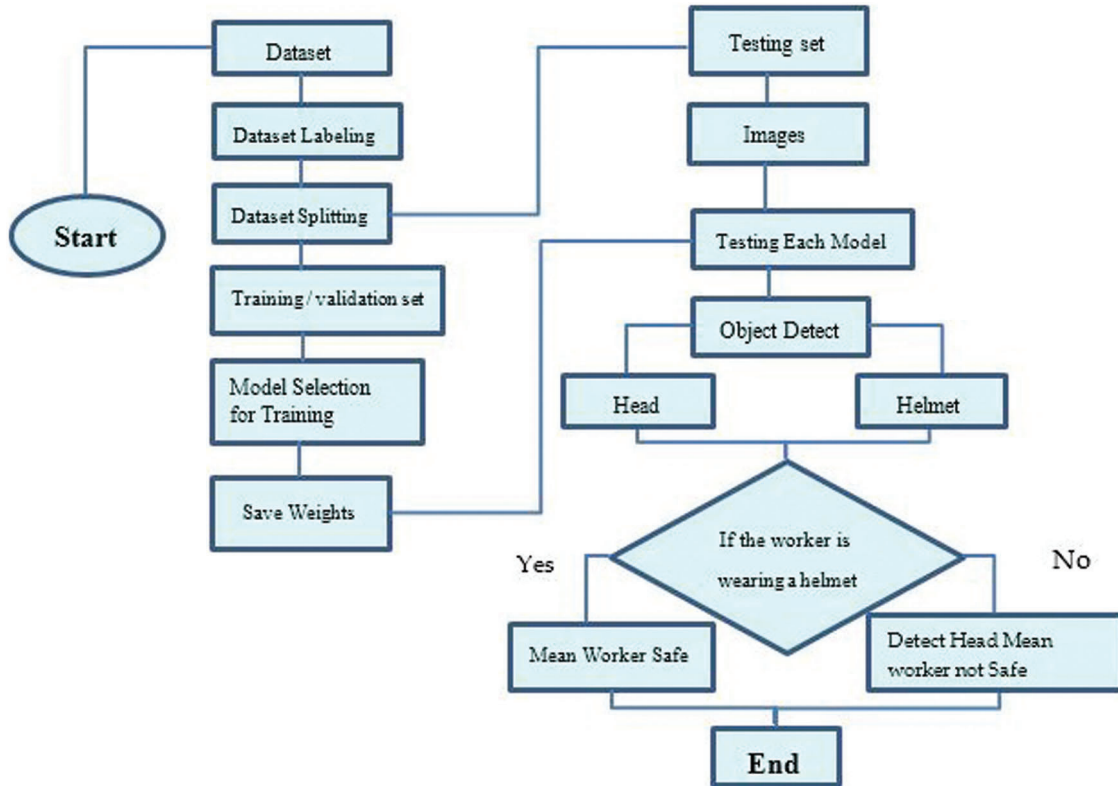

Figure 4: Working flow-chart of helmet detection

3.2 Dataset Collection

Data collection is the first step in creating a hardhat detection model. All images in a study were collected from the Harvard Dataverse shared by Northeastern University China. Image labelling and annotation were done online by Roboflow. The details of the dataset will be explained in the experiment section.

3.3 Proposed Model

To build the Personal Protective Equipment (PPE) Detector that detects hard hats. The use of hard hats helps to eliminate vital factors associated with work-related traumatic brain injury (TBI) in the construction industry. When workers do not wear hard hats, automatic detection of the head warns them that their lives are at risk. For this, we used the yolov5 model with PyTorch. PyTorch is an open-source machine learning library based on the Torch library that may be used for computer vision and natural language processing applications [47]. The proposed system uses different tools for working: Anaconda, which works as a virtual environment for Python code, Sublime, which is used as a text editor for Python code, and labelling and annotation done

online by Roboflow (https://roboflow.com). This method makes the entire procedure faster and easier to implement in projects. The main task is to train and test YOLOv5 (n, s, m, l, x) models and compare them.

To make a classification change, there are also weights for 80 item classifications, and thus the resulting size is 3 (5 + 80) = 255, where 3 indicates every grid prediction's three template boxes. The coordinates (x, y, w, and h) and confidence for every prediction box are represented by the number 5. This is essential when training our dataset, hence it must be modified when training a custom dataset. Determine how many classes are available in the detection strategy by defining how many classes are present. Two classes must be detected: head or helmet. First and foremost, to clarify, head detection is used to alert workers who are not wearing hard hats, which are required for the safety of any incident. As in the hard-hat identification scene, I have two sorts of objects: a helmet and a head (without a hard hat). Hence, we need to adjust the YOLOv5 classifier as 3 (5 + 2) = 21 is becoming the output size.

3.4 Training the Model

The next step is to train the model after it has been set up. Deep learning training is a time-consuming and resource-intensive process. Because the overall outcome of the process is primarily determined by the quality of the training process, for this project, I have already created the training dataset, which includes information such as staff who wear hard hats and staff who do not wear hard hats. The model's training process has been completed by running it on different models of Yolov5 (YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x) with PyTorch and adhering to the rules. For training, computer vision models require a powerful GPU. As a result, we trained by using the GPU (NVIDIA GeForce GTX 1080 Ti) and other resources. For the training process, we give a batch size of 20, an epoch of 50, and an image size of 416.

3.5 Testing the Model

The final step in the procedure is to test the produced model. The generated models have been evaluated using the test images from the dataset, as well as internet video and real-time video. The test dataset is processed in the same way as the training dataset by the system. The model then calculates the coefficient value and compares it to the value that was previously trained. The system detects the helmet or head based on this data.

4 Experiment and Results

This section first goes through the details of an experiment, such as the dataset and system, before showing the training outcomes.

4.1 Hard-Hat Dataset

The collection of data is the first step in the creation of a hard-hat detection model. The model is being trained to detect whether or not someone is wearing hard hats. If workers are not wearing hard hats, automatic detection of the head alerts them that their lives are in danger. A novel dataset contained 7063 images. All the images for the study were collected from the Harvard Data-verse shared by Northeastern University-China [48]. The dataset is divided into three parts, the training set, the validation set and the testing set. The training set consists of 5297 images, the validation set consists of 1059 images and the testing test consists of 707 images. The pictures in the dataset depict a variety of scenes and helmets of varying sizes. We tagged the head parts of each worker in the photographs with a bounding box and the label "helmet," or if they were not wearing a hardhat, then labelled "head". The following characteristics can be found in our dataset:

- The photographs were taken both outside and inside.
- Worker gestures come in a variety of shapes and sizes (walking, standing, sitting, bending, among others).
- Workers in the photos are photographed from various perspectives, like from the front, back, side, top, and bottom.

4.2 Experimental Environment

Our experiment runs on Windows 10, with Pytorch 1.10.2, Python 3.8.5, and Cuda 11.3 installed, and training and testing are performed on a single NVIDIA GeForce GTX 1080 Ti GPU. The YOLOv5 (n, s, m, l, x) different network was trained and tested for this experiment of helmet detection, with a batch size of 20, an epoch of 50, and an image size of 416*416.

4.3 Metrics for Performance Evaluation

This study evaluates system performance by calculating the mAP (Mean Average Precision) metric at 0.5, recall, F1, precision, and FPS for each category of YOLOV5 (n, s, m, l, x) models to analyze model performance. The calculation equation for these is as follows [22].

$$mAP = \sum_{q=1}^{Q} \frac{AveP\,(q)}{Q} \tag{3}$$

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1\ score = 2* \frac{(precision * recall)}{precision + recall} \tag{6}$$

In Eq. (3), mAP is the average of APs in all categories, which represents the model's overall performance in class detection. In Eqs. (4) and (5), a true positive (TP) sample has an IOU greater than a certain threshold. Negative samples are false positive (FP) samples that have an IOU of less than a certain threshold, and FN is a false negative example, which means that the negative class is predicted inaccurately by the model. In Eq. (6), the F1-score combines a classifier's precision and recall by calculating their harmonic mean.

4.4 Experimental Results and Analysis

This section will discuss extensive experimental results based on the YOLOv5 (n, s, m, l, and x) different models. Changes during the training process are printed out separately for comparison to the effectiveness and superiority of the proposed method based on the evolution metric obtained by training on the dataset. Tab. 1 displays the experimental results.

Tab. 1 shows that YOLOV5x performs better in terms of mAP and F1 than any other model, as Yolov5n and Yolov5s require less time for training. The training processes for Yolov5l, Yolov5m, and Yolov5x take time to complete. As a result of the findings, we can conclude that as training duration increases, model performance based on metrics increases as well. The evaluation results show that YOLOV5 s has a 1.1% better output than YOLOV5n, YOLOv5m has a 0.2% higher performance than YOLOV5s, YOLOV5l has a 0.1% higher performance than YOLOV5m, and YOLOV5x has a 0.2% higher performance than YOLOV5l. As training time increases, the F1 values of the five models gradually converge.

Table 1: Comparison of evaluation metrics

Model	Batch-Size	Image-size	Recall (%)	Precision (%)	F1	mAP@.5
YOLOV5n	20	416*416	87.6	93.4	0.904	0.942
YOLOV5s	20	416*416	89.9	93.6	0.917	0.953
YOLOV5m	20	416*416	90.4	94.8	0.925	0.955
YOLOV5l	20	416*416	90.3	94.9	0.925	0.956
YOLOV5x	20	416*416	91.2	93.9	0.925	0.958

Tab. 2 displays the results of head and helmet precision and recall for the five models.

Table 2: Precision and Recall of head and helmets

Model	Helmet		Head	
	Recall (%)	Precision (%)	Recall (%)	Precision (%)
YOLOV5n	89.3	93.5	85.9	93.5
YOLOV5s	91.6	94.2	88.4	93
YOLOV5m	92.3	95.4	88.5	94.3
YOLOV5l	92.5	95.2	88.1	94.6
YOLOV5x	92.3	94.5	90.2	93.3

Now, as indicated in Tab. 3, compare the detection times of all five distinct models to images, video of hard hats gathered from the internet and real time video. The best weights of these five models were tested and evaluated, and the findings demonstrate that YOLOV5n is faster than other models in both image and video tests. Also consider that as the training time increases, the detection speed decreases.

Table 3: Detection speed comparison

Model	FPS (images)	FPS (video)	FPS (real-time video)
YOLOV5n	62.5	70.4	65.3
YOLOV5s	55.8	54	43.8
YOLOV5m	41.1	42.1	30.9
YOLOV5l	37.1	35.4	18.9
YOLOV5x	25.6	27.1	10.8

The model's progress is shown in the graphs in Figs. 5, 6 and 7, which demonstrate that the model improved quickly in terms of precision, recall, and mean average precision. Precision is one metric for assessing the performance of a machine learning model. Recall is a way of measuring how well our model identifies true positives. To fully assess a model's efficiency, you must look at both precision and recall. The f1 metric assesses the trade-off between precision and recall. When the value of f1 is large, it indicates that the precision and recall are both large. From the training process, the validation and training

data also show the rapid decline in the box, object, and classification losses; therefore, early stopping was used to determine the best weights and to avoid over-fitting.
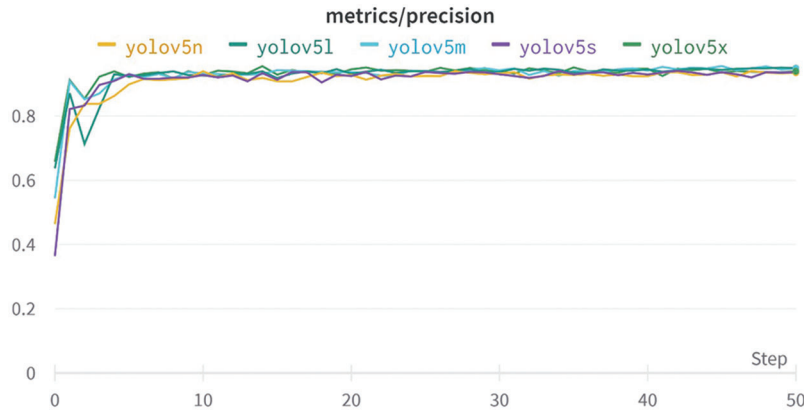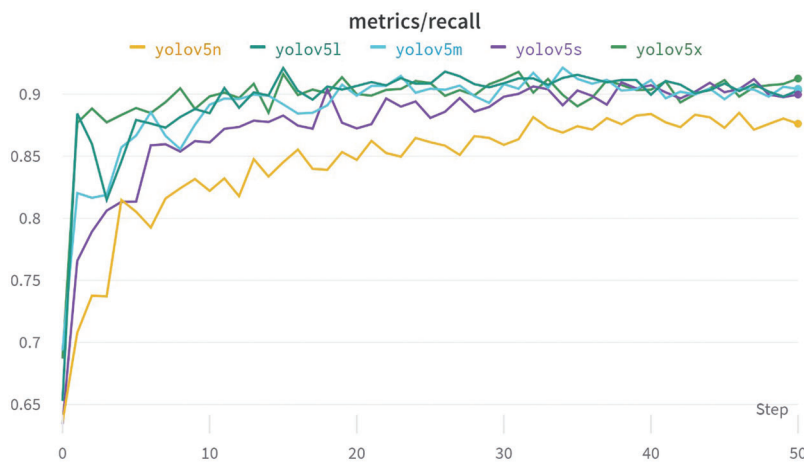

Figure 5: Precision graph of all models

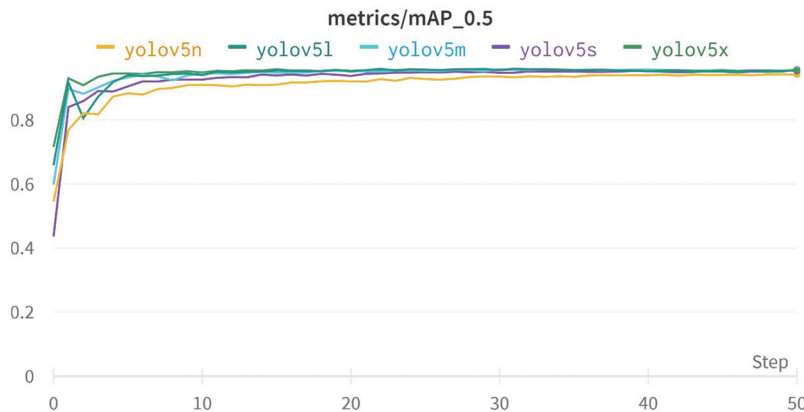**Figure 6:** Recall graph of all model

Figure 7: Map graph of all models

A system for detecting helmet use can greatly reduce the chance of traumatic brain injury while also improving construction site safety. Prior research has highlighted the need for using deep learning techniques. Prior research has highlighted the need for using deep learning techniques rather than sensor-based methods to detect helmets. The focus of this research was to use a yolov5-based approach to identify any worker's helmet. As a result, rigorous tests on the proposed dataset show that the proposed strategy is effective.

Research conducted on past studies done on the topic at hand, it is pertinent to mention that our research has yielded better results using the same algorithms than those of previous research. The work of Zhou et al. [35] was described in the above literature section. They demonstrate the efficiency of YOLOv5-based helmet detection by evaluating the mAP of all YOLOv5 models (s, m, l, x), which shows the YOLOV5x model having the highest mAP of 94.7 percent. According to the research we conducted on the same algorithm YOLOV5 (n, s, m, l, x), the model YOLOV5x has the highest mAP value of 95.8%. In comparison to the prior search, each model has improved in terms of its mAP value effectively. The dataset has 7063 photos, while Zhou, F., Zhao, H. et al. have 6045. The progress in the task has been demonstrated using the graph in Fig. 8.
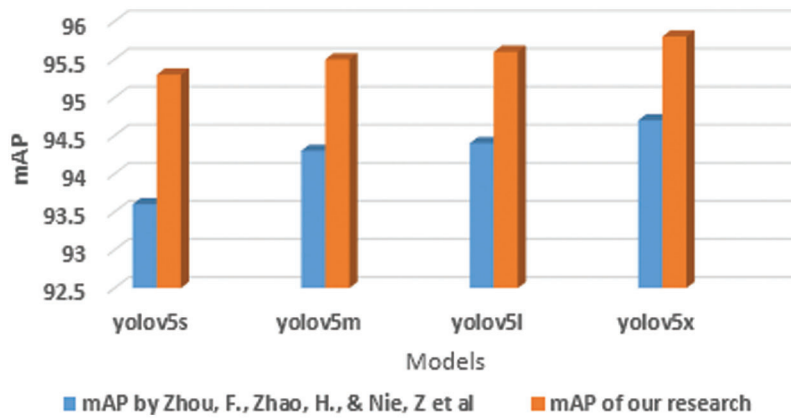

Figure 8: Comparison of work

We found a way to show the results on the images in real-time to better display the results of our studies as depicted in Fig. 9.

(a)   (b)

Figure 9: (Continued)

(c)   (d)

Figure 9: Detection results in real-time (a), (b), (c), (d)

5 Conclusion

This study has focused on the use of the YOLOV5 model, although the literature's findings concentrated on using the YOLOV4 and YOLOV3 models. The Yolov5 model comparatively has a higher rate of accuracy, less model size, and a relatively faster speed. The majority of the significant improvements in YOLOv5 can be attributed to the PyTorch training processes. As a result, a hard-hat detection method based on the YOLOv5 models was developed. The approach is equally valuable and useful both in the fields of civil (construction) engineering and ICT. A method consists of four key steps. Data collection comes first, followed by the suggested model, model training, and finally testing the performance of the proposed method. Experiments were carried out using YOLOv5 models (s, m, l, and x) for comparison of results. It has the potential to check whether people are wearing helmets or not. Experiments using the YOLOv5 method yielded excellent results in terms of evaluation metrics. The results show that the YOLOV5n is faster than other models in terms of FPS for both image and video test data, which are 62.5 and 70.4, respectively. The YOLOVX is the best performer in terms of mAP (95.8%). When processing speed is taken into account, then YOLOv5n is a bit superior to YOLOv5x. When mAP is considered, then YOLOV5x is better than YOLOV5n in performance. Conclusively, the proposed Deep Learning (DL) model using YOLOV5 can successfully be used in practice to recognize the helmet or hard hat worn by a worker at construction sites with adequate accuracy and speed.

6 Future-Work

However, there are some limitations to the current work that could lead to future advancements to-ward Industry Revolution (IR) 4.0 involves AI connectivity with new emerging technologies such as the Industrial Internet of Things (IIoT). The concept of Industrial Revolution (IR) 4.0 embodies digitization, automation, and increased use of information and communications technology (ICT) across all industries. Information and communication technologies (ICTs) have assisted in advancements in occupational health and safety and worker security. The usage of ICT-based Personal Protective Equipment (PPE) minimizes workplace accidents.

In the construction industry (IR 4.0), it makes it difficult for the construction industry to adjust to changing technologies, which are mostly focused on the internet of things (IOT). However, because the recent deadly COVID-19 outbreak has largely opened the construction industry's eyes to IR 4.0, the

adaptation of Industry 4.0 in the field of construction includes physical and digital technologies. Like in civil engineering, wireless sensors are used to remotely monitor the physical, structural, and environmental aspects of projects. Just like a smart helmet is capable of detecting hazardous events in the construction industry and designing a safety system based on wireless sensor networks, if there is any danger, then the sensor detects it and vibrates the safety helmet to inform the worker to keep him safe. Sensors are also used to find out the humidity and temperature of the environment. In short, the smart helmet model monitors the conditions in the workers' environment and assesses danger in real-time. These models closely monitor the workplace and inform relevant staff of any anomalies or risks.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] "Pakistan-employment in industry (%of total employment)," 2020, [online]. Available: https:// tradingeconomics. com/pakistan/employment-in-industry-percent-of-total-employment-wb-data.html.

[2] "Housing and construction board of investment," 2021, [online]. Available:https://invest.gov.pk/housing-and-construction>.

[3] A. Mehmood, "An investigation into causes of accidents in construction projects of Pakistan," M.S. Thesis, Department of Mechanical Engineering, Capital University of Science & Technology, Islamabad, Pakistan, 2020.

[4] T. Siddiqui, "Rangers' report blames MQM for baldia factory fire," 2015, [online]. Available: https:// www.dawn. com/news/1162049.

[5] "Safety and health at work in Pakistan," 2022, [online]. Available: https://www.ilo.org/islamabad/areasofwork/ safety-and-health-at-work/lang–en/index.htm.

[6] N. D. Nath, A. H. Behzadan and S. G. Paal, "Deep learning for site safety: Real-time detection of personal protective equipment," *Automation in Construction*, vol. 112, no. 4, pp. 1–20, 2020.

[7] "Construction worker deaths rose 41% between 2011 and CPWR report 2019," 2021, [online]. Available: https:// www.safetyandhealthmagazine.com/articles/21056-cpwr-report-looks-at-fatal-injury-trends-in construction.

[8] "8 common causes of personal injury on construction sites," 2017, [online]. Available: <https:// www. thornycroftsol icitors.co.uk/blog/8-common-causes-personal-injury-construction-sites/>.

[9] R. Sehsah, A. Gilany and A. M. Ibrahim, "Personal protective equipment (PPE) use and its relation to accidents among construction workers," *La Medicina del Lavoro Work, Environment and Health*, vol. 111, no. 4, pp. 285–295, 2020.

[10] K. Brolin, D. Lanner and P. Halldin, "Work-related traumatic brain injury in the construction industry in Sweden and Germany," *Safety Science*, vol. 136, no. 4, pp. 1–11, 2021.

[11] "Workplace fatal injuries in Great Britain," 2021, [online]. Available: https://www.hse.gov.uk/statistics/pdf/ fatalinjuries.pdf.

[12] S. Barro-Torres, T. M. Fernández-Caramés, H. J. Pérez-Iglesias and C. J. Escudero, "Real-time personal protective equipment monitoring system," *Computer Communications*, vol. 36, no. 1, pp. 42–50, 2012.

[13] A. Kelm, L. Laußat, A. Meins-Becker, D. Platz, M. J. Khazaee *et al.,* "Mobile passive radio frequency identification (RFID) portal for automated and rapid control of personal protective equipment (PPE) on construction sites," *Automation in Construction*, vol. 36, no. 8, pp. 38–52, 2013.

[14] B. Naticchia, M. Vaccarini and A. Carbonari, "A monitoring system for real-time interference control on large construction sites," *Automation in Construction*, vol. 29, no. 1, pp. 148–160, 2013.

[15] S. Du, M. Shehata and W. Badawy, "Hard hat detection in video sequences based on face features, motion and color information," in *Proc. of the 3rd Int. Conf. on Computer Research and Development*, Shanghai, China, pp. 25–29, 2011.

[16] K. Shrestha, P. P. Shrestha, D. Bajracharya and E. A. Yfantis, "Hard-hat detection for construction safety visualization," *Journal of Construction Engineering*, vol. 2015, no. 3, pp. 1–8, 2015.

[17] W. Sun, L. Dai, X. R. Zhang, P. S. Chang and X. Z. He, "RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring," *Applied Intelligence*, vol. 52, no. 8, pp. 1–16, 2021.

[18] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 779–788, 2016.

[19] M. J. Shafiee, B. Chywl, F. Li and A. Wong, "Fast YOLO: A fast you only look once system for real-time embedded object detection in video," *Journal of Computational Vision and Imaging System*, vol. 3, no. 1, pp. 1–3, 2017.

[20] M. Buric, M. Pobar and M. Ivasic-Kos, "Ball detection using yolo and mask R-CNN," in *Proc. of the Int. Conf. on Computational Science and Computational Intelligence*, Las Vegas, NV, USA, pp. 319–323, 2018.

[21] J. Du, "Understanding of object detection based on CNN family and Yolo," *Journal of Physics: Conference Series*, vol. 1004, no. 212, pp. 1–9, 2018.

[22] W. Fang, L. Ding, B. Zhong, P. E. D. Love and H. Luo, "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach," *Advanced Engineering Informatics*, vol. 37, no. 3, pp. 139–149, 2018.

[23] B. E. Mneymneh, M. Abbas and H. Khoury, "Evaluation of computer vision techniques for automated hardhat detection in indoor construction safety applications," *Frontiers of Engineering Management*, vol. 5, no. 2, pp. 227–239, 2018.

[24] J. Wu, N. Cai, W. Chen, H. Wang and G. Wang, "Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset," *Automation in Construction*, vol. 106, no. 10, pp. 1–7, 2019.

[25] H. M. Hung, L. T. Lan and H. S. Hong, "A deep learning-based method for real-time personal protective detection," *Journal of Science and Technique-Le Quy Don Technical University*, vol. 6, no. 13, pp. 24–34, 2019.

[26] F. Chuan and W. RongXin, "Research on safety helmet wearing YOLO-V3 detection technology improvement in mine environment," *Journal of Physics*, vol. 1345, no. 4, pp. 1–7, 2019.

[27] D. Ayrapetov and B. Buyanov, "Analysis of work of yolo v.3 and yolo v.2 neural networks," in *Proc. of the Int. Conf. on Technology & Entrepreneurship in Digital Society*, Moscow, Russia, pp. 22–25, 2019.

[28] M. M. Saudi, A. Hakim, A. Ahmad, A. Shakir, M. Hanafi *et al.,* "Image detection model for construction worker safety conditions using faster R-CNN," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 6, pp. 246–250, 2020.

[29] W. Wang, H. Liu, J. Li, H. Nie and X. Wang, "Using CFW-net deep learning models for X-ray images to detect COVID-19 patients," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 199–207, 2021.

[30] X. Long, K. Deng, G. Wang, Y. Zhang, Q. Dang *et al.,* "PP-YOLO: An effective and efficient implementation of object detector," ArXiv, vol. 3, no. 12099, pp. 1–8, 2020.

[31] D. Zhang, J. Hu, F. Li, X. Ding, A. K. Sangaiah *et al.,* "Small object detection via precise region-based fully convolutional networks," *Computers, Materials and Continua*, vol. 69, no. 2, pp. 1503–1517, 2021.

[32] L. Zhao and S. Li, "Object detection algorithm based on improved YOLOV3," *Electronics*, vol. 9, no. 3, pp. 537–548, 2020.

[33] W. N. Tun, J. H. Kim, Y. Jeon, S. Kim and J. W. Lee, "Safety helmet and vest wearing detection approach by integrating YOLO and SVM for UAV," in *Proc. Korean Society for Aeronautical and Space Sciences Spring Conf.*, Korea, pp. 1–4, 2020.

[34] R. Shukla, A. K. Mahapatra, A. Kumar and J. S. P. Peter, "Social distancing tracker using Yolo V5," *Turkish Journal of Physiotherapy and Rehabilitation*, vol. 32, no. 2, pp. 1785–1793, 2021.

[35] F. Zhou, H. Zhao and Z. Nie, "Safety helmet detection based on YOLOV5," in *Proc. of the IEEE Int. Conf. on Power Electronics, Computer Applications*, Shenyang, China, pp. 6–11, 2021.

[36] Z. Wang, Y. Wu, L. Yang, A. Thirunavukarasu, C. Evison *et al.,* "Fast personal protective equipment detection for real construction sites using deep learning approaches," *Sensors*, vol. 21, no. 10, pp. 1–22, 2021.

[37] G. Yang, F. Wei, J. Jintao, L. Qujiang, L. Xiuhao *et al.,* "Face mask recognition system with YOLOV5 based on image recognition," in *Proc. of the IEEE 6th Int. Conf. on Computer and Communications*, Chengdu, China, pp. 1398–1404, 2020.

[38] T. H. Wu, T. W. Wang and Y. Q. Liu, "Real-time vehicle and distance detection based on improved Yolo V5 network," in *Proc. of the 3rd World Symp. on Artificial Intelligence*, Guangzhou, China, pp. 24–28, 2021.

[39] A. A. Protik, A. H. Rafi and S. Siddique, "Real-time personal protective equipment (PPE) detection using YOLOV4 and tensorflow," in *Proc. of the IEEE Region 10 Symp.*, Jeju, Korea, pp. 1–6, 2021.

[40] P. Torres, A. Davys, T. Silva, L. Schirmer, A. Kuramoto *et al.,* "A robust real-time component for personal protective equipment detection in an industrial setting," in *Proc. of the 23rd Int. Conf. on Enterprise Information Systems, online streaming*, Rio de Janeiro, Brazil, pp. 639–700, 2021.

[41] R. Shukla, A. K. Mahapatra and J. S. P. Peter, "Social distancing tracker using YOLOV5." *Turkish Journal of Physiotherapy and Rehabilitation*, vol. 32, no. 2, pp. 1785–1793, 2021.

[42] T. F. Dima and E. Ahmed, "Using YOLOV5 algorithm to detect and recognize American sign language," in *Proc. of the Int. Conf. on Information Technology*, Amman, Jordan, pp. 603–607, 2021.

[43] G. Jocher, "Ultralytics/YOLOV5," GitHub, 2020, [online]. Available: https://github.com/ultralytics/yolov5.

[44] Q. Song, S. Li, Q. Bai, J. Yang, X. Zhang *et al.,* "Object detection method for grasping robot based on improved YOLOv5," *Micromachines*, vol. 12, no. 11, pp. 1–18, 2021.

[45] "YOLOV5‐Explained and demystified," Towards AI, 2020, [online]. Available: https://towardsai.net/p/computer-vision/yolo-v5%E2%80%8A-%E2%80%8Aexplained-and-demystified.

[46] SeekFire, "Overview of model structure about YOLOv5 #280," GitHub, 2020, [online]. Available: https ://git hub.com/ultralytics/yolov5/issues/280.

[47] A. Paszke, S. Gross, S. Chintala and G. Chanan, 2022, [online]. Available: https://en.wikipedia.org/wiki/PyTo rch..

[48] L. Xie, "Hardhat-harvard-dataverse," 2019, [online]. Available: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi.org/10.7910/DVN/7CBGOS.