 https://doi.org/10.24191/jeesr.v25i1.006

Hardhat10K: A Large-scale Dataset for Deep Learning-based Hardhat-wearing Detection

Wanbo Luo, Khairul Khaizi Mohd Shariff, Rajeswari Raju and Ahmad Ihsan Mohd Yassin*

***Abstract*—Hardhats are crucial personal protective equipment in various industries, including construction and manufacturing. Utilizing deep learning methods to develop unattended detection systems can ensure compliance with hardhat-wearing. However, a deep-learning-based hardhat-wearing detection model requires enormous training images. Unfortunately, existing public datasets lack sufficient images and labels to train a hardhat-wearing detection model with good robustness. The introduction of the Hardhat10K dataset, comprising over 10,000 images, is a significant feature of this research. Images containing long-distance, occluded, dense, and low-light objects were collected to enhance the model's robustness. Furthermore, images from various weather conditions and periods were added to improve the model's generalization ability. Finally, background images were supplemented to enhance the model's accuracy. Compared to the state-of-the-art SHEL5K dataset, the number of images and labels increased by over one time and approximately 69.2%, respectively. The proposed dataset was evaluated using three types of models. Each model achieved good accuracy and usability on the proposed dataset. The dataset is publicly available at *https://github.com/bobo504/hardhat10k*.**

***Index Terms*—Dataset, object detection, deep learning, hardhat-wearing detection.**

I. Introduction

Hardhats are designed to safeguard workers against critical head injuries from impacts, falls, and other accidents, significantly enhancing safety on job sites. Despite this, the adherence to wearing hardhats is often compromised by a lack of safety awareness among some workers, leading to severe consequences including fatalities [1].

From 2017 to 2021, China reported a total of 3,622 construction-related accidents in the housing and municipal engineering sectors, culminating in 4,198 deaths [2]. These incidents were categorized by the nature of the accidents, with falls from heights being the most prevalent at 53%, followed by impacts from objects at 14%, and accidents involving lifting machinery and earthwork collapses at 7% and 8%, respectively. Additionally, the International Powered Access Federation (IPAF) Global Safety Report of 2023 indicated a rising trend in reported incidents across 34 countries, with 759 incidents in 2022, marking a 15% increase from the previous year, though fatalities decreased by 19% [3].

The structural components of a hardhat, including the shell, lining, chin strap, and accessories, protect the head by absorbing shock and distributing stress [4]. The effectiveness of hardhats in preventing head injuries underscores the necessity of wearing them correctly as mandated in many industrial regulations worldwide [5].

Many countries have regulations that require employees to wear hardhats when necessary. In China, wearing hardhats is a basic requirement in the construction industry [6]. The revised Work Safety Law of the People's Republic of China, which took effect on September 1, 2021, aims to ensure the safety of employees in the workplace. According to Article 102 of this law, enterprises are required to provide their employees with protective equipment that meets national standards and ensure that personnel wear it properly [7]. The U.S. Department of Labor mandates that employers conduct a hazard assessment at their job sites to determine whether head protection is necessary based on workplace hazards. When head protection is necessary, employers should require employees to wear hardhats to provide the best protection against occupational head injuries [8].

Despite these regulations, there remains a gap in compliance, primarily due to insufficient safety awareness and inadequate supervision. This gap highlights the need for innovative solutions to enhance compliance monitoring. Traditional object detection methods have proven inadequate for ensuring the consistent use of hardhats. In contrast, deep learning-based detection models offer a more robust alternative, provided they are trained on diverse and extensive datasets to enhance their accuracy and generalizability.

Therefore, to address this need, this paper introduces an advanced dataset for hardhat-wearing detection, the Hardhat10K dataset, which extends and enhances the existing Safety Helmet Detection dataset.

The main contributions of this paper are:

1) The Safety Helmet Detection dataset has been meticulously re-annotated into six distinct classes using the YOLO format to standardize annotation protocols with added images, significantly expanding its scope and utility.

This manuscript is submitted on 16 May 2024, revised on 20 August 2024, accepted on 5 September 2024 and published on 31 October 2024. Wanbo Luo, Khairul Khaizi Mohd Shariff and Ahmad Ihsan Mohd Yassin are with the College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia. Rajeswari Raju is with the College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 21080 Kuala Terengganu, Terengganu, Malaysia.

*Corresponding author
Email address: ihsan_yassin@uitm.edu.my

2) Images, including challenging scenarios such as low visibility, high density, and varied weather conditions, were incorporated into the Safety Helmet Detection dataset to form the Hardhat10K dataset, which comprises over 10,000 images.
3) This paper details the evaluation of three differently scaled models on this dataset to assist enterprises in selecting the most effective and economical solutions for hardhat-wearing detection.

The classes of the Hardhat10K dataset are the "*head*", "*head_with_hardhat*", "*person_with_hardhat*", "*head*", "*person_no_hardhat*", and "*face*", where the "*head*" and "*person_no_hardhat*" classes represent workers without hardhats, the "*head_with_hardhat*" and "*person_with_hardhat*" classes represent workers with hardhats and the "*face*" class was used for facial recognition.

The rest of this paper is organized as follows: Section 2 presents a review of recent relevant literature, followed by the methodology in Section 3. Results are detailed in Section 4. Section 5 presents discussions. Finally, concluding remarks are displayed in Section 6.

II. Related Works

Recent scholarly efforts have been directed toward developing public datasets for detecting hardhat usage, yet these datasets have encountered several challenges that diminish their utility for training effective detection models.

The Safety Helmet Detection dataset, which includes 5,000 images annotated in the PASCAL VOC format, is divided into three classes: "*helmet*", "*person*", and "*head*". A significant limitation of this dataset is its incomplete annotations, with many "person" objects not being labelled, rendering it suboptimal for model training [9].

The SHEL5K dataset was an extended version of the Safety Helmet Detection dataset [10]. The original three-class label was insufficient to represent all objects in images. For instance, objects of people wearing hardhats were labelled to the "*person*" class, while all objects of heads with hardhats were labelled as the "*helmet*" class. Original classes could not distinguish objects of heads and persons without hardhats. Therefore, the number of label classes was increased from three to six (*helmet, head_with_helmet, person_with_helmet, head, person_no_helmet*, and *face*) in the SHEL5K dataset. Furthermore, the missed-labelled objects in the safety helmet detection dataset are fully labelled using the Pascal VOC format.

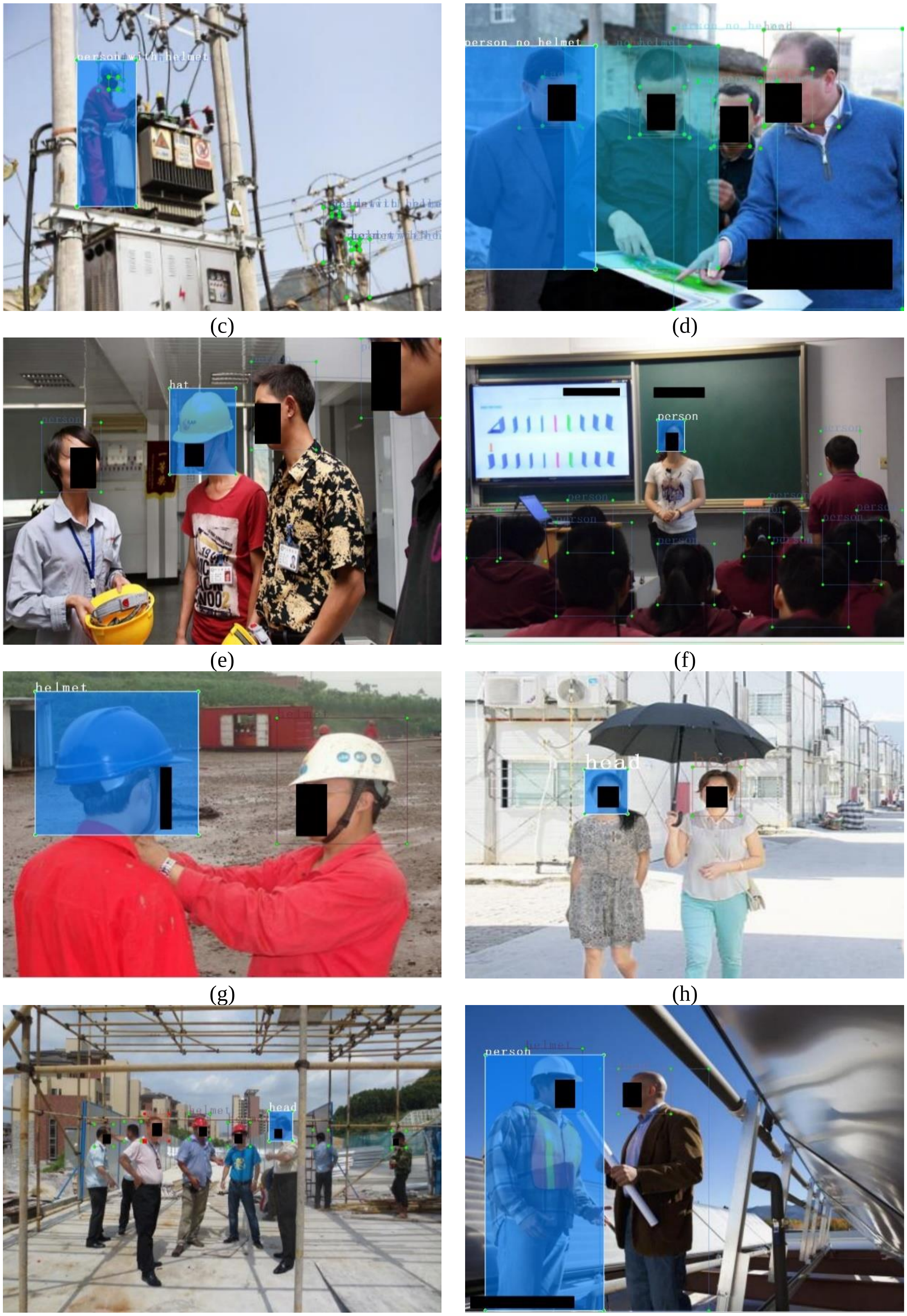
The Safety Helmet Wearing dataset comprised 7,581 images, annotated as two classes: "*hat*" and "*person*" using the PASCAL VOC format [11]. Upon examination of the dataset, it was determined that 3,241 images originated from work sites, whereas 4,341 images were captured from single scenes, such as classrooms and conference rooms. Consequently, the model trained on this dataset demonstrated insufficient generalization ability.

Furthermore, the Hardhat dataset includes 7,063 images specifically annotated for object detection in workplace environments, where safety regulations mandate hardhat wearing. This dataset is annotated in the PASCAL VOC format and labels individuals not wearing hardhats as "*head*", thereby facilitating specific object recognition in compliance-focused applications [12].

Lastly, the Hard Hat Workers v2 dataset with 7035 images was shared on the Internet in 2020 [13]. The dataset was a modified version of the Hardhat dataset, which was annotated as three classes: "*head*", "*helmet*", and "*person*". The difference between the two versions was the format used, one using the YOLO format and the other using the PASCAL VOC format. Similar to other datasets, many "*person*" objects in this dataset were not labelled.

Fig. 1 displays examples of labelled images from the five datasets. The Safety Helmet Detection dataset labelled heads with hardhats as the "*helmet*" class and workers with hardhats as the "*person*" class, as shown in Fig. 1(a) and Fig. 1(b). In the SHEL5K dataset, the six-class label can fully represent all objects from images, as shown in Fig. 1(c) and Fig. 1(d). The Safety Helmet Wearing dataset labelled heads with hardhats as the "*hat*" class and heads as the "*person*" class, as shown in Fig. 1(e) and Fig. 1(f). The Hardhat dataset labelled heads with hardhats as the "*helmet*" class and individuals without hardhats as the "*head*" class, as shown in Fig. 1(g) and Fig. 1(h). The labelling strategy of the Hard Hat Workers v2 Dataset is similar to that of the Hardhat dataset. However, it classifies workers with hardhats and individuals without hardhats as the "*person*" class, as shown in Fig. 1(i) and Fig. 1(j).

(a) (b)



(c) (d)
(e) (f)
(g) (h)
(i) (j)

Fig. 1. Examples of labelled images from five datasets.

III. Hardhat10K Dataset

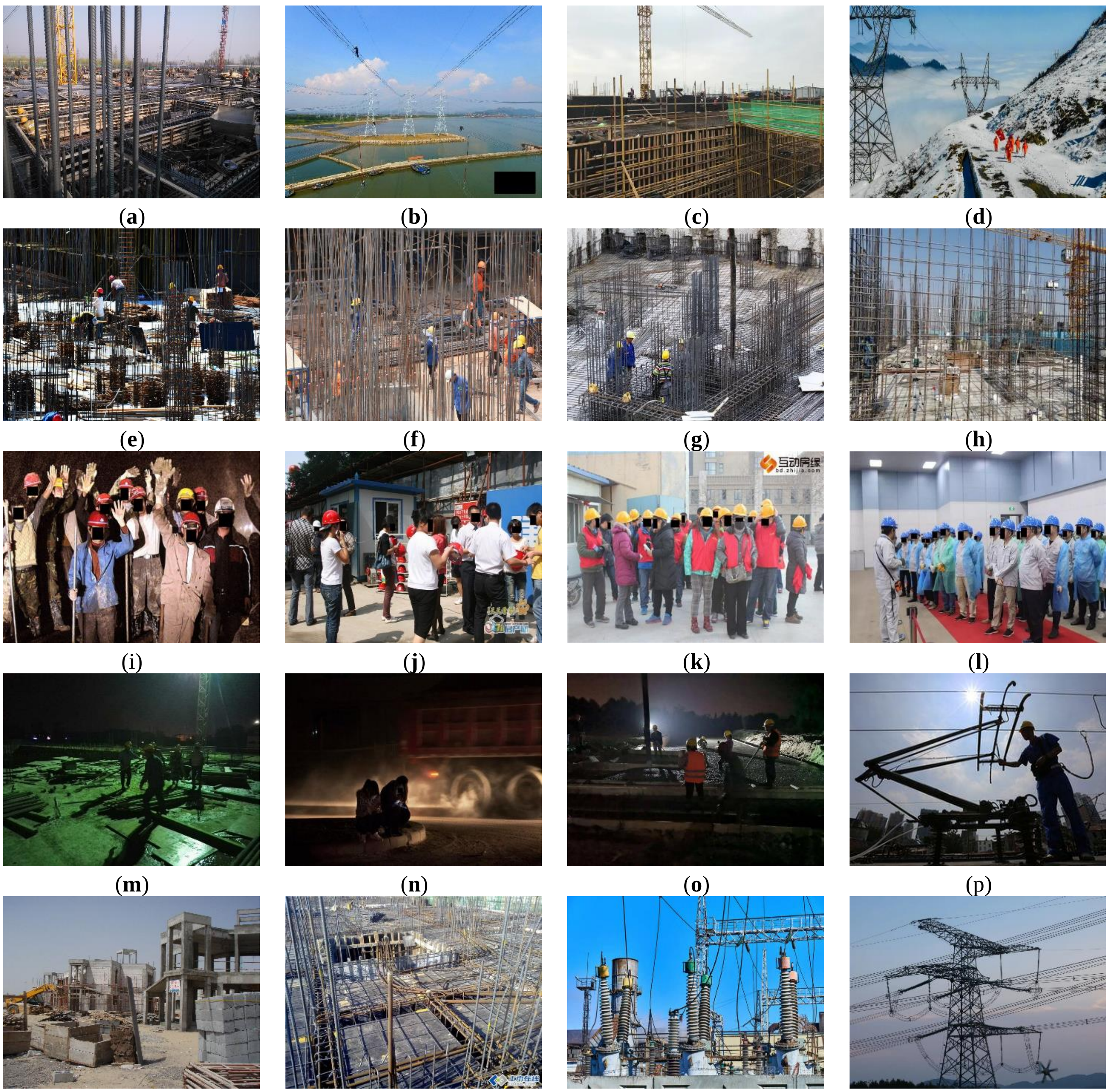
Real-time hard-hat detection algorithms are the subject of active research. To illustrate, the Safety Helmet Detection dataset consists of images with a resolution of 416 × 416 pixels, whereas cameras at work sites typically capture higher resolutions, such as 1920 × 1080 or 1280 × 720 pixels. The Hardhat10K dataset included images of various resolutions to address this discrepancy and meet practical application requirements. In addition, comparative experiments were conducted between the Hardhat10K and the state-of-the-art

SHEL5K datasets to assess the validity of each type of image.

A. Building the Hardhat10K Dataset

To enhance the model's robustness and generalization ability, these steps were followed to create the Hardhat10K dataset. First, the Safety Helmet Detection dataset was divided into a training dataset of 4500 images and a validation dataset of 500 images. Second, images containing long-distance, occluded, dense, low-light objects and backgrounds were added to the training dataset to train the model, respectively. The number of images in each category is as follows: 113, 103, 75, 160, and 300, respectively. Third, 3287 images from various work sites, weather conditions, and periods were added to the training dataset, bringing the number of images to a total of 8538 images. Fourth, 1500 images were incorporated into the validation dataset, bringing the number of images to a total of 2000 images. Finally, the Hardhat10K dataset contained 10,538 images, of which about 80% were used for training and about 20% were used for validation. Most images were sourced from the Internet, with a small number being captured by authors.

Fig. 2 presents a series of sample images that encompass a diverse range of types, including long-distance (Fig. 2(a) to Fig. 2(d)), occluded (Fig. 2(e) to Fig. 2(h)), dense (Fig. 2(i) to Fig. 2(l)), and low-light objects (Fig. 2(m) to Fig. 2(p)); background images from work sites (Fig. 2(q) to Fig. 2(t)); and images captured under various weather conditions and periods (Fig. 2(u) to Fig. 2(x)).



(**a**) (**b**) (**c**) (**d**)

(**e**) (**f**) (**g**) (**h**)

(i) (**j**) (**k**) (**l**)

(**m**) (**n**) (**o**) (p)

(**q**) (**r**) (**s**) (**t**)

(u) (v) (w) (x)

Fig. 2. Sample images in the Hardhat10K dataset.

A comparison of five public datasets with the Hardhat10K dataset is presented in Table I. The Safety Helmet Detection dataset comprised 5,000 images, with 18,966 "*helmet*", 751 "*person*" and 5,785 "*head*" instances. The SHEL5K dataset undertook a reannotation of the Safety Helmet Detection dataset, which had 19,252 "*hardhat*", 16,048 "*head_with_hardhat*", 14,767 "*person_with_hardhat*", 6,120 "*head*", 5,248 "*person_no_hardhat*", and 14,135 "*face*" instances. The Safety Helmet Wearing dataset comprised 7,581 images with 8,437 "*hat*" and 111,514 "*person*" instances. The Hardhat dataset comprised 7,063 images with 19,852 "*helmet*" and 6,781 "*head*" instances. The Hard Hat Workers v2 Dataset comprised 7,035 images with 6,677 "*head*," 19,747 "*helmet*," and 612 "*person*" instances. The Hardhat10K dataset exhibited the highest number of labels among the datasets under consideration, reaching 127,891. Furthermore, it comprised six classes, analogous to the SHEL5K dataset, which fully represents all objects in work-site images. Compared to the SHEL5K dataset, the number of labels exhibited a notable increase. Where * represents the class in the original dataset.

TABLE I. COMPARISON OF FIVE PUBLIC DATASETS WITH THE HARDHAT10K DATASET

Class	Safety Helmet Detection	SHEL5K	Safety Helmet Wearing	Hardhat	Hard Hat Workers v2	Hardhat10K
image number	5,000	5,000	7,581	7,063	7,035	10,538
label number	3	6	2	2	3	6
hardhat	18,966 (*helmet)	19,252	8437 (*hat)	19,852 (*helmet)	19,747 (*helmet)	34,363
head_with_hardhat		16,048				28,880
person_with_hardhat	751 (*person)	14,767			612 (*person)	27,955
head	5,785	6,120	111,514 (*person)	6781	6677	7,651
person_no_hardhat		5,248				6,653
face		14,135				22,389
total	25,502	75,570	119,951	26,633	27,039	**127,891**

Where * represents the class in the original dataset.

Fig. 3 compares the number of labels for each class between the SHEL5K and the Hardhat10K datasets. Compared to the SHEL5K dataset, the total number of labels in the Hardhat10K dataset increased by approximately 69.2%. The number of labels for each class increased by 77.4%, 80%, 89.3%, 25%, 26.8%, and 58.4%, respectively.

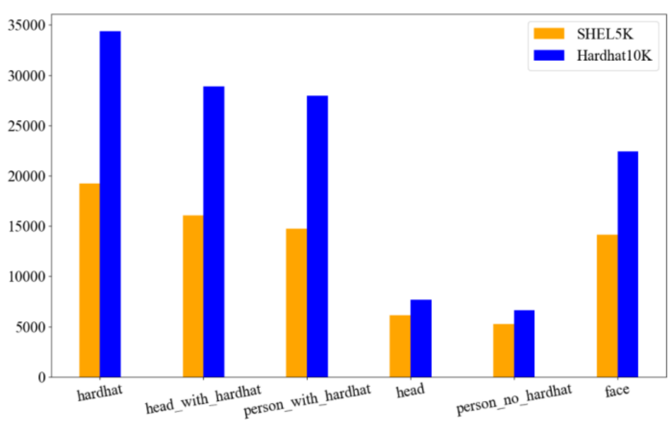


Fig. 3. Comparison of the number of labels for each class.

B. *Data Validity Assessment*

The YOLO series of object detection algorithms achieves a good balance between accuracy and speed. It includes YOLOv1 [14], YOLOv2 [15], YOLOv3 [16], YOLOv4 [17], and YOLOv5. In particular, the YOLOv5 series is widely used in the industry for object detection tasks. Therefore, this paper utilized YOLOv3 and YOLOv5 algorithms to conduct experiments.

Table II presents the validation results of models trained using different types of images based on the YOLOv5s algorithm. The primary metric for evaluating model performance is mAP50, which indicates the model's detection accuracy. The model trained using 4500 images from the SHEL5K dataset achieved a mAP50 of 78.6%. After incorporating various types of images, the trained models achieved mAP50 values of 78.8%, 79.3%, 80.1%, 79.6%, 79%, and 86.3%, respectively.

TABLE II. VALIDATION RESULTS OF MODELS TRAINED USING DIFFERENT TYPES OF IMAGES

Training images	Validation images	mAP50 (all classes)
4500 (the SHEL5K dataset)	2000	78.6%
4500 + 113 (long-distance objects)	2000	78.8%
4500 + 101 (occluded objects)	2000	79.3%
4500 + 75 (dense objects)	2000	80.1%
4500 + 160 (low-light objects)	2000	79.6%
4500 + 300 (background images)	2000	79.0%
8538 (+ 4038 images from various work sites, weather conditions, and periods, Hardhat10K)	2000	86.3%

Fig. 4 illustrates the mAP50 results over 200 training epochs for each model configuration.

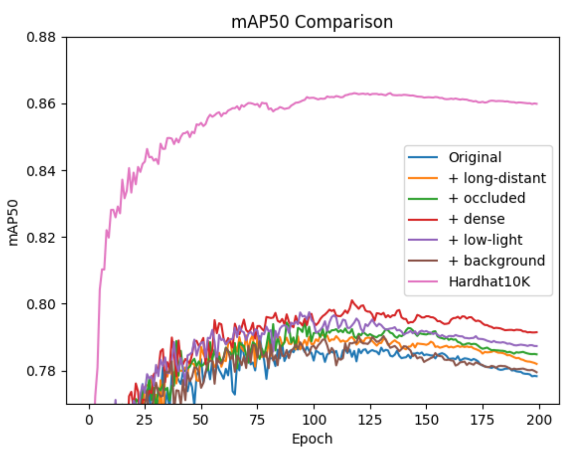


Fig. 4. Comparison of the mAP50 over 200 training epochs.

It has been demonstrated that the incorporation of a diverse range of images leads to an increase in mAP50 compared to the model trained solely with original images from the SHEL5K dataset. This comparison underlines the robustness and validity of the data in the Hardhat10K dataset for enhancing model performance.

IV. USABILITY OF THE HARDHAT10K DATASET

Three types of models, differentiated by their parameter counts and computational demands, were trained on the Hardhat10K dataset to evaluate their usability. Models in the first category, which include YOLOv3-SPP [18], YOLOv5s-EfficientNetv2 [19], and YOLOv5l, each have over 20 million parameters. The second category consists of models like YOLOv3-tiny [20] and YOLOv5s, each with over 7 million parameters. The third category includes YOLOv5s-MobileNetv3 [21], YOLOv5s-ShuffleNetv2 [22], and YOLOv5s-LCNet [23], all having over 4 million parameters. Each model was trained for 300 epochs with the images resized to 640 × 640 pixels.

A. Training Results

Fig. 5 displays the training results of the YOLOv5s model over 300 epochs. The smooth training curves across various metrics indicate well-labelled objects in the Hardhat10K dataset. During training and validation, losses for bounding boxes, objectness, and class categories consistently decreased, while metrics such as Precision, Recall, mAP50, and mAP50-95 gradually improved. The training curves for other models were similar to those observed with the YOLOv5s model.

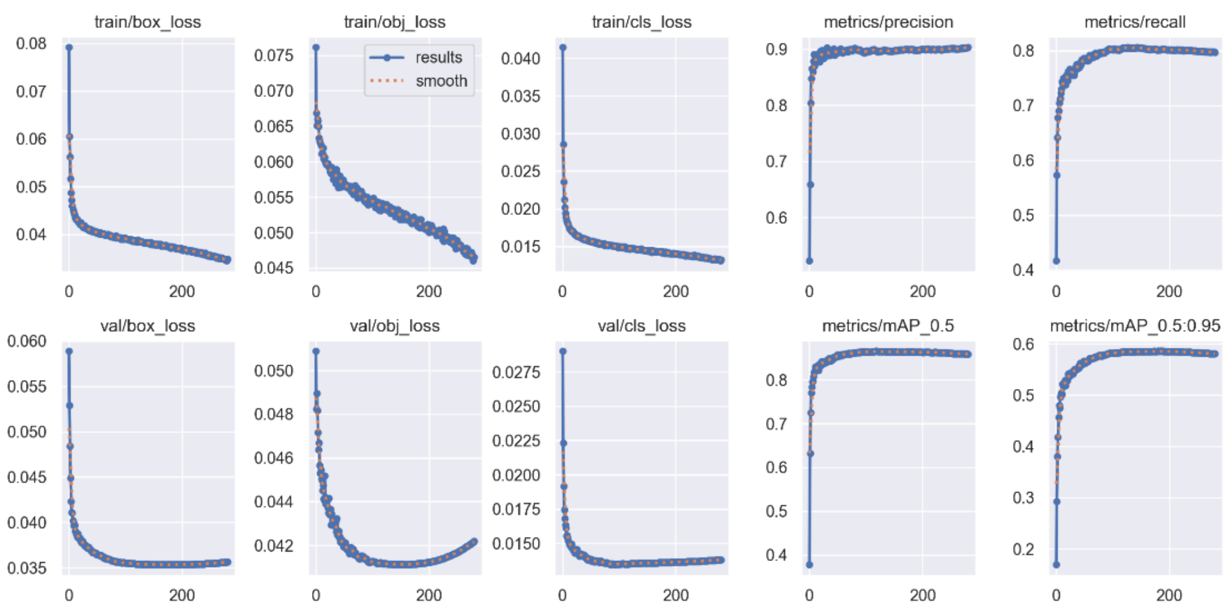


Fig. 5. Training results of the YOLOv5s model.

B. Validation Results

The performance of the trained models was evaluated using 2000 images from the validation dataset. The results are presented across multiple tables, reflecting the performance of models with varying computational complexities and parameter sizes.

Table III shows the validation results of the YOLOv3-SPP model. The model achieved a mAP50 of 88.3% for all six classes. The recall of the "*hardhat*" class was 0.785, indicating that some "*hardhat*" objects were missed detection. However, the model achieved high precision, recall, and mAP50 for the other classes, with these metric values exceeding 80%. In particular, the model achieved mAP50s of over 92% for the "*head_with_hardhat*" and "*person_with_hardhat*" classes.

TABLE III. VALIDATION RESULTS OF YOLOV3-SPP MODEL

Class	Precision	Recall	mAP50
all	0.896	0.830	**0.883**
hardhat	0.935	0.785	0.882
head_with_hardhat	0.941	0.843	0.921
person_with_hardhat	0.891	0.893	0.920
head	0.889	0.809	0.842
person_no_hardhat	0.817	0.809	0.845
face	0.901	0.840	0.888

Table IV shows the validation results of the YOLOv5s-EfficientNetv2 model, which is trained by the YOLOv5s network using the EfficientNetv2 backbone replacing its backbone. The model achieved a mAP50 of 86.9% for all six classes. Additionally, it had a higher detection accuracy rate for the "*head_with_hardhat*" and "*person_with_hardhat*" classes.

TABLE IV. VALIDATION RESULTS OF YOLOV5S-EFFICIENTNETV2 MODEL

Class	Precision	Recall	mAP50
all	0.908	0.799	**0.869**
hardhat	0.965	0.760	0.863
head_with_hardhat	0.947	0.819	0.910
person_with_hardhat	0.897	0.867	0.909
head	0.904	0.789	0.839
person_no_hardhat	0.840	0.754	0.828
face	0.892	0.804	0.867

Table V shows the validation results of the YOLOv5l model. The model achieved the highest mAP50 of 88.5% for all six classes than other models.

TABLE V. VALIDATION RESULTS OF YOLOV5L MODEL

Class	Precision	Recall	mAP50
all	0.903	0.816	**0.885**
hardhat	0.943	0.760	0.878
head_with_hardhat	0.952	0.817	0.920
person_with_hardhat	0.905	0.879	0.925
head	0.898	0.809	0.850
person_no_hardhat	0.809	0.821	0.857
face	0.910	0.808	0.878

Table VI shows the validation results of the YOLOv3-tiny model. The model achieved a mAP50 of 75.3% for all six classes. The recalls of the "*hardhat*", "*head*", "*person_no_hardhat*", and "*face*" classes showed that the model had a higher missed detection rate for these objects.

TABLE VI. VALIDATION RESULTS OF YOLOV3-TINY MODEL

Class	Precision	Recall	mAP50
all	0.843	0.676	0.753
hardhat	0.913	0.644	0.754
head_with_hardhat	0.934	0.727	0.840
person_with_hardhat	0.799	0.759	0.800
head	0.880	0.638	0.713
person_no_hardhat	0.701	0.608	0.675
face	0.830	0.682	0.735

Table VII shows the validation results of the YOLOv5s model. The model achieved a mAP50 of 86.4% for all six classes and the recall of the "*face*" class indicated that the model had a higher missed detection rate for the "*face*" objects.

TABLE VII. VALIDATION RESULTS OF YOLOV5S MODEL

Class	Precision	Recall	mAP50
all	0.899	0.802	0.864
hardhat	0.953	0.764	0.859
head_with_hardhat	0.940	0.824	0.905
person_with_hardhat	0.882	0.779	0.829
head	0.819	0.783	0.825
person_no_hardhat	0.911	0.799	0.864
face	0.830	0.682	0.735

Tables VIII, IX, and X show the validation results of lightweight models with small amounts of parameters and computations, such as the YOLOv5s-MobileNetv3, YOLOv5s-ShuffleNetv2, and YOLOv5s-LCNet models. These models were trained by the YOLOv5s network using the MobileNetv3, ShufflueNetv2, and LCNet backbone replacing its backbone.

Table VIII shows the validation results of the YOLOv5s-MobileNetv3 model. The model achieved a mAP50 of 78.7% for all six classes. The recalls for the "*hardhat*", "*head*", and "*person_no_hardhat*" classes were 67.3%, 65.7%, and 64.1%, respectively, indicating that the model had a higher missed detection for these three class objects.

TABLE VIII. VALIDATION RESULTS OF YOLOV5S-MOBILENETV3 MODEL

Class	Precision	Recall	mAP50
all	0.880	0.709	**0.787**
hardhat	0.945	0.673	0.773
head_with_hardhat	0.941	0.758	0.861
person_with_hardhat	0.861	0.794	0.852
head	0.869	0.657	0.721
person_no_hardhat	0.776	0.641	0.713
face	0.891	0.729	0.801

Table IX shows the validation results of the YOLOv5s-ShuffleNetv2 model. The model achieved a mAP50 of 82.3% for all six classes. The recall for the "*person_no_hardhat*" class was 68.8%, indicating over 30% of these objects were not detected.

TABLE IX. VALIDATION RESULTS OF YOLOV5S-SHUFFLENETV2 MODEL

Class	Precision	Recall	mAP50
all	0.892	0.748	**0.823**
hardhat	0.956	0.725	0.824
head_with_hardhat	0.946	0.781	0.883
person_with_hardhat	0.879	0.821	0.872
head	0.886	0.714	0.771
person_no_hardhat	0.832	0.688	0.760
face	0.851	0.760	0.829

Table X shows the validation results of the YOLOv5s-LCNet model. The model achieved a mAP50 of 80.5% for all six classes. Like the YOLOv5s-MoblieNetv3 model, it exhibited a higher rate of missed detections for the "*hardhat*", "*head*", and "*person_no_hardhat*" objects.

TABLE X. VALIDATION RESULTS OF YOLOV5S-LCNET MODEL

Class	Precision	Recall	mAP50
all	0.894	0.706	**0.805**
hardhat	0.896	0.678	0.796
head_with_hardhat	0.957	0.758	0.871
person_with_hardhat	0.908	0.784	0.860
head	0.918	0.659	0.753
person_no_hardhat	0.857	0.631	0.744
face	0.826	0.725	0.807

Fig. 6 compares the mAP50s for 300 training epochs of the three-type models. The YOLOv5l model achieved the highest mAP50 of 88.5%, while the YOLOv3-tiny model had the lowest mAP50 of 75.3%. All trained models achieved high detection accuracy on the Hardhat10K dataset. Therefore, the Hardhat10K dataset exhibited high usability in training hardhat-wearing detection models. Due to the early stopping mechanism, YOLOv3-SPP, YOLOv5l, and YOLOv5s models were trained for 222, 172, and 281 epochs, respectively.

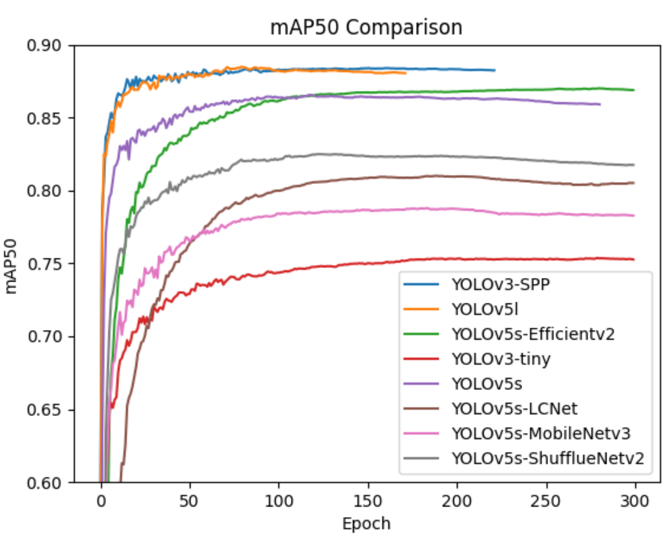


Fig. 6. Comparison of the mAP50s of the three-type models in 300 training epochs.

C. Model Complexity Assessment

The relationship between model complexity and detection accuracy is pivotal in deploying object detection systems in practical applications, particularly for safety equipment detection in work environments. Model complexity in this context is quantified by two key metrics: Parameters, which indicate the memory requirement for model operation, and Giga Floating Point Operations (GFLOPs), which measure the computational effort required for model inference.

Table XI compares the Parameters, GFLOPs, and mAP50 of the three types of models. The larger YOLOv3-SPP, YOLOv5l, and YOLOv5s-EfficientNetv2 models achieved a higher mAP50 with more parameters and computations. On the other hand, the smaller YOLOv5s-LCNet, YOLOv5s-MobileNetv3, and YOLOv5s-ShuffleNetv2 models achieved a lower mAP50 with fewer parameters and computations. Among the medium-sized models, the YOLOv5s model performed well with a higher mAP50 of 86.4%, while the YOLOv3-tiny model had the lowest mAP50 of 75.3%. Compared to other models, the YOLOv5s-ShuffleNetv2 model achieves 82.3% mAP50 with the fewest parameters and the second least amount of computation.

TABLE XI. COMPARISON OF PARAMETERS, GFLOPS, AND MAP50 OF THREE-SIZE MODELS

Model	Parameters	GFLOPs	mAP50
YOLOv5l	46,135,203	107.7	**0.885**
YOLOv3-SPP	62,573,443	155.5	0.883
YOLOv5s-EfficientNetv2	25,135,923	57.2	0.869
YOLOv3-tiny	8,678,242	12.9	0.753
YOLOv5s	7,026,307	15.8	0.864
YOLOv5s-LCNet	4,663,491	8.7	0.805
YOLOv5s-MobileNetv3	4,477,091	7.1	0.787
YOLOv5s-ShuffleNetv2	**4,084,871**	7.2	0.823

In summary, the accuracy of the model's detection is generally proportional to the number of parameters and computations. Highly accurate models require significant hardware resources, such as graphic memory and specialized computing circuits, resulting in high costs for practical applications. Therefore, three-size hardhat-wearing detection models are provided to enterprises for options with different cost budgets.

V. DISCUSSION

The hardhat dataset can significantly advance hardhat detection and safety monitoring by providing high-quality, annotated data critical for training deep learning models for accurate detection. Liao et al. developed a smart surveillance system for helmet detection [24]. Using their helmet dataset, they reduced the number of safety violations by 30% in the monitored zones. Liu et al. developed an AI-powered drone system trained on a safety helmet dataset [25]. This system can autonomously detect and report safety violations across a large construction site, enhancing monitoring efficiency greatly.

However, existing hardhat-wearing detection datasets are relatively few. Concurrently, the trained models with these datasets lack robustness and generalization capabilities. To address this issue, the Hardhat10K dataset is proposed.

First, the Safety Helmet Detection dataset has been meticulously re-annotated into six classes using the YOLO format. The newly added images were labelled following the same protocols. This was done to enhance the generalization ability of the trained model in real-world work sites.

Second, Images, including challenging scenarios such as low visibility, high density, long distance, occlusion, and varied weather conditions, were incorporated into the Safety Helmet Detection dataset to form the Hardhat10K dataset. This was done to enhance the robustness of the trained model.

Third, each type of image was evaluated to demonstrate these images' validity. Experimental results showed that the model's accuracy was improved by 0.2%, 0.7%, 1.5%, 1%, and 0.4% in the type of long-distance, occluded, dense, low-light, and background images, respectively. The image number proportions of each type are approximately 2.4%, 2.2%, 1.6%, 3.4%, and 6%, respectively.

Finally, the comprehensive testing of Hardhat10K across various object detection models, including YOLOv3-SPP, YOLOv5s-EfficientNetv2, and YOLOv5l, etc., demonstrates its high usability for training hardhat-wearing detection algorithms aimed at improving workplace safety.

VI. Conclusion

This paper introduces Hardhat10K, a large-scale dataset for deep learning-based hardhat-wearing detection. The dataset is an extension of the Safety Helmet Detection dataset, with completed missing and fixed incorrect labels. The proposed dataset is formed by adding many types of images to the original dataset. Compared to the other datasets, the proposed dataset contains a significantly larger number of images and labels. Each type of image is effective for hardhat-wearing detection. The proposed dataset is tested on multiple object detection models, and the experimental results demonstrate its high usability for training deep learning-based hardhat-wearing detection models. Finally, this paper explores three types of models to aid enterprises in selecting hardhat-wearing detection solutions with varying costs for practical applications.

References

[1] Y. Guan, W. Li, T. Hu, and Q. Hou, "Design and Implementation of Safety Helmet Detection System Based on YOLOv5," in *Proceedings of 2nd Asia Conference on Computers and Communications (ACCC)*, Singapore, September 24-27, 2021, pp.69–73.

[2] Z. Yi, G. Wu, X. Pan, and J. Tao, "Research on Helmet Wearing Detection in Multiple Scenarios Based on YOLOv5," in *Proceedings of 33rd Chinese Control and Decision Conference (CCDC)*, Kunming, China, May 22-25, 2021, pp.769–773.

[3] Z. Li, S. Liu, F. Li, J. Wang, J. Zhang, and Y. Li, "A helmet detection scheme based on improved YOLOv3," in *Proceedings of 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC)*, Qingdao, China, December 2-4, 2022, pp.14–17.

[4] P. Ge and Y. Chen, "An Automatic Detection Approach for Wearing Safety Helmets on Construction Site based on YOLOv5," in *Proceedings of IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS)*, Sichuan, China, August 3-5, 2022, pp.140–145.

[5] M. Zhou, Z. Fang, B. Zhao, and P. Li, "Safety Helmet Wearing Detection and Recognition Based on YOLOv4," in *Proceedings of 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST)*, Guangzhou, China, December 10-12, 2021, pp.798–802.

[6] G. Yang and Q. Lei, "The system of detecting safety helmets based on YOLOv5," in *Proceedings of International Conference on Electronic Information Engineering and Computer Science (EIECS)*, Yanji, China, September 27-29, 2021, pp.750–755.

[7] X. Ai, C. Chen, Y. Wang, and Y. Guo, "Multi-Scene Safety Helmet Detection with Multi-Scale Spatial Attention Feature," in *Proceedings of 7th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC)*, Beijing, China, November 17-19, 2021, pp.21–25.

[8] Y. Gao, Y. Liu, D. Nie, Z. Huang, B. Li, and B. Qian, "An Efficient Safety Helmet Detection based on Attentional Mechanism," in *Proceedings of 7th International Symposium on System and Software Reliability (ISSSR)*, Chongqing, China, September 23-24, 2021, pp.97–98.

[9] Larxel, "Safety Helmet Detection," Available: https://www.kaggle.com/andrewmvd/hard-hat-detection.

[10] M.-E. Otgonbold, M. Gochoo, F. Alnajjar, L. Ali, T.-H. Tan, J.-W. Hsieh, and P.-Y. Chen, "SHEL5K: An Extended Dataset and Benchmarking for Safety Helmet Detection," *Sensors*, vol. 22, no. 6, pp. 2315-2337, 2022.

[11] JUNWIDE, "SafetyHelmetWearing," Available: https://www.kaggle.com/datasets/junwide/safetyhelmetwearing.

[12] L. B. Xie, "Hardhat," Available: https://doi.org/10.7910/DVN/7CBGOS.

[13] L. B. Xie, "Hard Hat Workers Dataset," Available: https://public.roboflow.com/object-detection/hard-hat-workers.

[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 27-30, 2016, pp.779–788.

[15] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, July 21-26, 2017, pp. 6517–6525.

[16] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv*, 2018.

[17] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv*, 2020.

[18] X. Zhang, Y. Gao, H. Wang, and Q. Wang, "Improve YOLOv3 using dilated spatial pyramid module for multi-scale object detection," *International Journal of Advanced Robotic Systems*, vol. 17, no. 4, 2020.

[19] M. X. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," In *Proceedings of the 38th International Conference on Machine Learning*, Virtual Event, July 18–24, 2021, pp.10096-10106.

[20] P. Adarsh, P. Rathi, and M. Kumar, "YOLO v3-Tiny: Object Detection and Recognition using one stage improved model," In *Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Tamil Nadu, India, March 6–7, 2020.

[21] A. G. Howard, M. Sandler, and G. Chu, "Searching for MobileNetV3," In *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*, Seoul, South Korea, October 27 – November 2, 2019, pp.1314–1324.

[22] X. Y. Zhang, X. Y. Zhou, M. X. Lin, and J. Sun, "ShuffleNet: an extremely efficient convolutional neural network for mobile devices," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 18-23, 2018, pp.6848-6856.

[23] C. Cui, T. Q. Gao, S. Y. Wei, Y. N. Du, R. Y. Guo, and B. Lu, et al, "PP-LCNet: a lightweight CPU convolutional neural network," *arXiv*, 2021.

[24] B. Liao, T. Chen, and P. Xue, "Improved YOLOv5 Algorithm for Helmet Wearing Detection," In *Proceedings of the ACM International Conference Proceeding Series*, 2023, pp. 266–271.

[25] Q. Liu and F. Han, "Research on an improved yolov5s algorithm for detecting helmets on construction sites," In *Proceedings of the ACM International Conference Proceeding Series*, 2023, pp. 104–110.