2023 26th International Conference on Computer and Information Technology (ICCIT)
13-15 December, Cox's Bazar, Bangladesh

Safety Helmet Detection of Workers in Construction Site using YOLOv8

Syed Shakil Mahmud[1*], Md. Ashraful Islam[1], Khandaker Jannatul Ritu[1], Mahmudul Hasan SMIEEE[1,2]
Yoshinori Kobayashi[2], and Md Mohibullah[1,2*]
[1]Department of Computer Science and Engineering, Comilla University, Cumilla, Bangladesh
[2]Interactive Systems Lab., Saitama University, Saitama, Japan
* syedshajib55@gmail.com, * mohibullah@cou.ac.bd

2023 26th International Conference on Computer and Information Technology (ICCIT) | 979-8-3503-5901-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCIT60459.2023.10441212

***Abstract*—The safety of construction workers is a paramount concern in the modern construction industry. A significant proportion of injuries and fatalities on construction sites are attributed to a lack of adherence to safety regulations, often resulting from the non-utilization of Personal Protective Equipment (PPE) by workers. Additionally, effective monitoring of construction areas by site supervisors can be challenging due to the vastness and complexity of construction sites. Manual monitoring, while a crucial aspect of construction safety management, is often hindered by time constraints and associated costs. To address these challenges, the application of computer vision and Convolutional Neural Network(CNN) techniques has led to the development of automated helmet and jacket detection systems. This research employs the YOLOv8 object detection algorithm to explore the efficacy of safety helmet detection, aiming to enhance the accuracy of automated safety helmet detection systems. The YOLOv8 object detection algorithm is a robust tool for recognizing objects in images. Our dataset comprises a total of 4,200 images, including 4,000 images of hard hats and jackets obtained from Kaggle [25], supplemented by 200 images depicting foggy, rainy, and nighttime conditions from our custom collection sourced from the internet. The dataset was meticulously divided into training, testing, and validation sets, following a ratio of 60%, 20%, and 20%, respectively. The experimental results revealed that the YOLOv8m architecture achieved an accuracy of 92%, demonstrating its effectiveness in detecting safety helmets under various lighting conditions, including low-light environments. This performance highlights the potential of the proposed approach for real-world applications in construction site safety monitoring. This developed model primarily focuses on detecting helmets and jackets in real-time in adverse weather conditions such as rainy, foggy, and low-light environments. It triggers an alarm if any workers are not wearing their helmets and jackets. We have also discussed model training, system structure, and performance measures.**

***Index Terms*—computer vision, object detection, YOLOv8, image processing, openCV, helmet detection, helmet, and jacket.**

I. Introduction

Recent years have witnessed a concerning rise in construction site accidents, often attributed to non-compliance with safety regulations regarding personal protective equipment (PPE). These accidents primarily stem from inadequate site supervision and the failure of workers to wear safety helmets. The cumulative effects of hazardous workplaces and the resulting worker fatalities have exacerbated the issue of construction site safety. In 2020 alone, Bangladesh recorded 112 construction-related fatalities [12]. Conventional construction site monitoring methods, such as manual inspections and video surveillance, are often hindered by camera angles and the limitations of human observation. The remarkable advancements in deep learning technologies over the past few years have opened up new possibilities for target recognition, particularly in the realm of construction site safety. Several methodologies have been developed for safety helmet detection in construction sites, including the Safety Helmet Detection Method Based on YOLOv4 [6], Safety Helmet Wearing Detection Based on Image Processing and Machine Learning [11], and Detection of Safety Helmet Wearing Based on Improved Faster R-CNN [16].

This paper aims to analyze construction surveillance images to identify the presence or absence of helmets among construction workers on the job site. We first employ a combination of gathered and self-collected images to identify construction workers and subsequently utilize computer vision, machine learning, and the YOLOv8 method to determine whether or not the worker is wearing a helmet. Construction surveillance images can be employed to detect and identify construction workers, including their attire, such as helmets and jackets. Computer vision and image processing techniques offer various methods for detecting and recognizing objects in images along with their distinctive features. The deep learning CNN model is one of the most widely used approaches for this purpose. To address the issue of missed detections and enhance the detection of small objects like safety helmets in construction site environments, we will employ the YOLOv8 method. Introduced by Ultralytics in 2020, the YOLOv5 target detection method offers the advantages of being compact, fast, and accurate. Its implementation on the PyTorch platform further facilitates deployment, implementation, and ecological maturity. The YOLOv8 family comprises five models: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. With a weight of just 6MB, YOLOv8n is the smallest among them, while YOLOv8x stands as the most powerful variant. The YOLOv8n-based method presents a well-developed algorithm that can be considered the preferred choice due to its accuracy and stability. The YOLOv8 algorithm has demonstrated its capability in identifying risk factors in various scenarios. However, its performance can be affected by low-light conditions and occlusions. To effectively contribute to hazard identification, the YOLOv8s algorithm requires further

979-8-3503-5901-5/23/$31.00 ©2023 IEEE

tuning. In this study, we train the network using a dataset of images depicting construction site workers and select the YOLOv8m model to identify and detect workers' safety helmets. Our system's primary objective is to detect helmets and jackets in foggy, rainy, and low-light environments and trigger an alarm in case of non-compliance. The trained YOLOv8 network is then applied to identify helmets and jackets in various construction sites. The utilization of YOLOv8 in this paper significantly enhances the detection speed.

II. Study of related works

J. Redmon introduced the YOLO detection algorithm for target detection, employing a single convolutional neural network instead of the intermediate stages of candidate regions. By explicitly regressing each bounding box and predicting the probability of the corresponding category, it achieved a final detection speed of 45 frames per second (f/s) while maintaining detection accuracy. Redmon subsequently proposed YOLOv2 [1] and v3 [2] based on YOLO, further improving detection performance and speed. In 2020, Alexey A.B. [3] took over from Redmon and developed modifications based on YOLOv3, further enhancing detection performance. On the COCO dataset, it achieved 43.5% AP, with real-time speed approaching 65 FPS. Moreover, detection accuracy increased by approximately 10% compared to YOLOv3. The YOLOv4 requires half the time of EfficientDet [7] to perform as well as 43AP. YOLOv4 can detect targets more accurately and quickly than previous versions while also producing better detection outcomes. Yan Rongrong [4] combined the Adaboost algorithm with skin color segmentation to detect human faces and identify helmets. Feng Guochen [5] explored automatic helmet recognition using OpenCV color recognition techniques. Zhang Bo [6] utilized OpenPose to generate sub-images representing the location of the human head and neck, and employed Faster R-CNN to identify helmets within these sub-images. Rattapoom et al. [10] employed machine vision techniques to detect motorcycle safety helmets using a K-Nearest Neighbor (KNN) classifier. Chiverton et al. [9] proposed segmenting motorbike riders in traffic videos using background subtraction and subsequently classifying the segments using support vector machines (SVM). Li et al. [11] implemented feature extraction based on histograms of oriented gradients (HOG) and SVM to detect safety helmets. Geng et al. [13] proposed an improved helmet detection method that utilized an unbalanced dataset of 7581 images, primarily featuring individuals wearing helmets against complex backgrounds. Upon evaluation using a test set of 689 images, the proposed method achieved label confidence of 0.982. Long et al. [14] developed a deep-learning-based safety helmet-wearing detection method utilizing 5229 images sourced from the Internet and various power plants (including those under construction). In the test pictures, the proposed system's mAP0.5 of 78.3 per was attained using SSD. Bochkovskiy et al. [15] introduced YOLOv4 in April 2020. This model replaces the FPN (Feature Pyramid Networks) algorithm in the YOLOv3 network with PANet (Path Aggregation Networks) and employs CSP (Cross Stage Partial) Darknet-53 as the backbone network. The literature presents a safety helmet detection approach based on Faster R-CNN, achieving a detection accuracy of 90%. In 2020, Xu Shoukun et al. [17] proposed a helmet detection method utilizing an enhanced Faster R-CNN and a combination of multiple elements. This method improved detection accuracy by 7% over the original Faster R-CNN and exhibited greater environmental adaptability.

In 2016, Liu et al. [18] introduced the SSD (Single Shot Multi-Box Detector) detection algorithm, pioneering the multi-scaled detection approach and enhancing detection accuracy. Bochkovskiy et al.'s [19] contribution was YOLOv4, a detection network that employs the PANet (Path Aggregation Network) path aggregation algorithm and utilizes CSP (Cross Stage Partial) Darknet-53 as its backbone network, thereby improving the model's detection precision. Jocher et al. [20] introduced YOLOv5 in 2020, a network model that enhances the YOLOv4 backbone network with a focused structure, achieving a balance between detection speed and accuracy. Transformers were introduced into the object detection fields by Carion et al. [21], who proposed DETR (Detection Transformer) for end-to-end object detection. Recently, Wang et al. [22] proposed YOLOv7, a model that surpasses YOLOv5 in terms of both accuracy and speed.

III. System training

The process of how the model has been trained is stated below:

A. *YOLOv8 network structure:*

The YOLO (You Only Look Once) object recognition technique utilizes a single neural network to simultaneously predict bounding boxes and class probabilities for each object in an image. The latest iteration of the YOLO algorithm, known as YOLOv8, outperforms its predecessors due to the integration of modules for spatial attention, feature fusion, and context aggregation. The convolutional neural network employed by YOLOv8 comprises two primary components: the head and the backbone. The head of the YOLOv8 algorithm consists of multiple convolutional layers and a series of fully connected layers. This architecture enables the model to dynamically adjust its focus on various image components based on their relative importance, allowing for multi-scaled object detection. To identify objects within an image, the model employs a feature pyramid network. The multiple layers of this feature pyramid network enable the detection of both large and small objects in an image by performing object detection at different scales. CBS (Cross Stage Partial Connections) is composed of convolution, batch normalization, and SiLu activation functions. SPPF (Spatial Pyramid Pooling) is composed of three Maxpool layers and two CBS [23].

B. *Dataset collection*

This study employed a total of 4200 construction worker images, sourced from the Kaggle Safety Helmet and Reflective Jacket dataset [25], and an additional 200 images of construction workers in foggy, rainy, and nighttime conditions, collected from the internet. As depicted in Figure 2, the dataset was divided into three subsets: a training set comprising 60% of the images (2520 images), a testing set accounting for 20% of the images (840 images), and a validation set consisting of the remaining 20% of the images (840 images).

C. Dataset preprocessing

The effectiveness of an object detection system hinges on the data preparation process. Data preprocessing plays a crucial role in augmenting the training data pool and enhancing the object detection system's ability to recognize objects from diverse perspectives. This process typically involves resizing images to the required resolution (640x640), cropping them, and converting them to a common format, such as JPG or PNG (in this case, JPG). To augment the diversity and robustness of the dataset, various distortions and transformations are applied to the images, including rotations, horizontal flips, vertical flips, and random cropping. This process of data augmentation enriches the training data by introducing variations in the images, enabling the object detection system to better generalize to unseen scenarios. The labeling tool was employed to label each image within the dataset about construction site workers. A representative example of labeling an image is shown below:

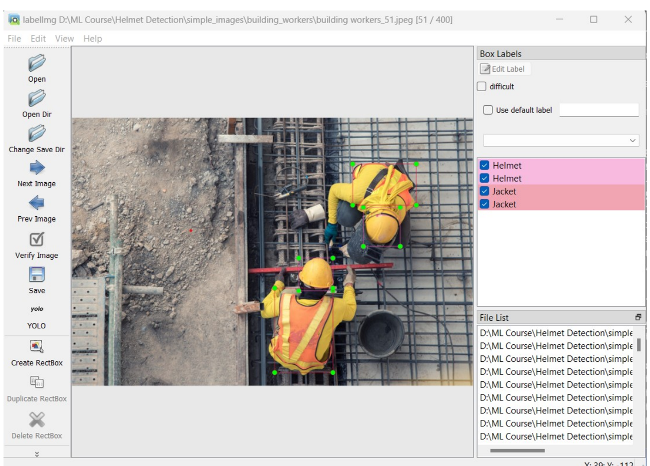

Fig. 1. Pre-processing of dataset[25]

D. Dataset folder structure

The data set for YOLOv8 is organized into three main folders: training, testing, and validation. Each of these folders contains two subfolders: images and labels. The images subfolder stores the raw images of construction workers, while the labels subfolder contains the corresponding label files. Label files are text documents that describe the location and class of each object in an image. In addition to the training, testing, and validation folders, there is a data.yaml file located in the data directory. This file is a configuration file that specifies various parameters related to the data set, such as the names of classes, paths to training and validation sets, and other relevant settings. Here's the folder structure illustrated in the figures below:

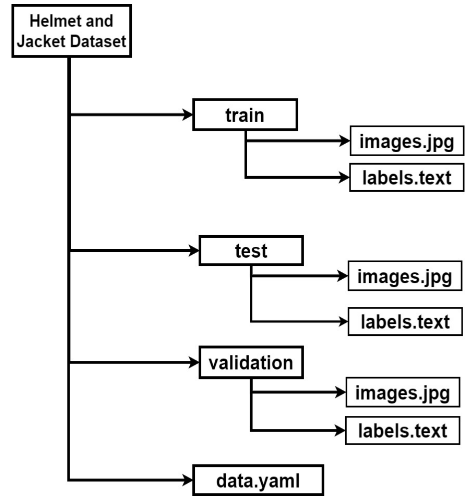

Fig. 2. Folder structure of dataset

E. Training process

The YOLOv8 settings employed in this experiment were optimized based on a series of empirical trials. The experimental batch size was set to 8, the momentum to 0.949, and the initial learning rate to 0.0013. The batch size of 8 strikes a balance between efficient training and memory constraints, while the momentum of 0.949 facilitates faster training progress by smoothing out weight updates. The initial learning rate of 0.0013 ensures that the model's weights are updated gradually, preventing overfitting the training data. A total of 20 training iterations were performed. The choice of 20 training iterations provides sufficient exposure to the training data without incurring excessive training time or computational resources.

F. Model evaluation

The mean average precision (mAP) and inference speed of each model were evaluated using the test set, comprising 20% of the collected dataset. This testing phase is crucial as it allows for the assessment of the model's performance on unseen data, providing insights into their overall effectiveness in recognizing helmets and jackets worn by construction workers. Upon the model (YOLOv8) predicting "not helmet" or "not jacket," a Python package named the "Alert System" is immediately invoked, triggering an alert on the computer display. This real-time notification mechanism ensures timely identification of potential safety hazards.

IV. Methodology

A. Initialize the webcam

The system captures real-time video footage from a CCTV camera and subsequently segments the video into individual images. The electric power source will be needed to run the webcam consistently.

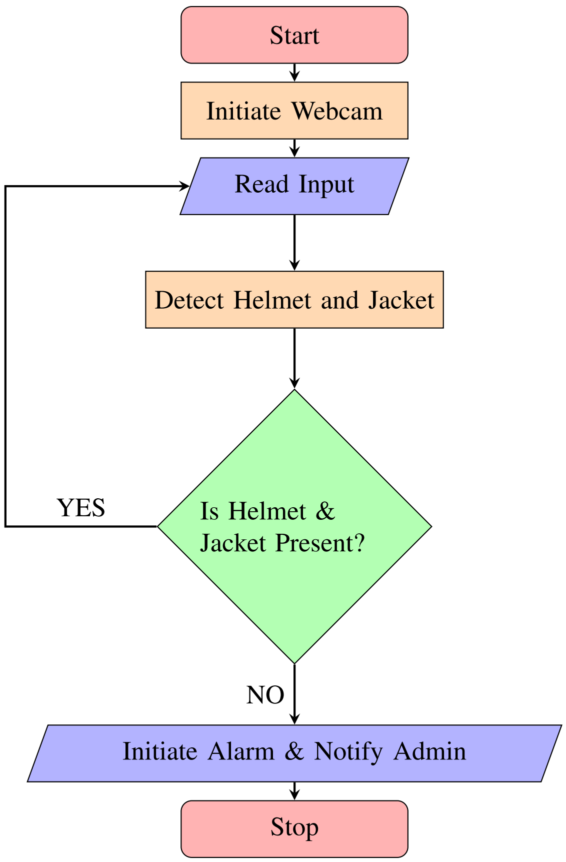

Fig. 3. Workflow of The System

B. Read input

The system acquires real-time video footage from a CCTV camera, subsequently segmenting the video into individual images. System will track the environment and consistently monitor the workers.

C. Detection of helmet

Our trained model YOLOv8m effectively detects the presence of helmets on construction workers, regardless of the helmet's color or design. This capability extends to noisy, foggy, rainy, and low-light environments, demonstrating the model's robustness and adaptability to real-world conditions. The improved detection accuracy in these challenging environments enhances the model's suitability for practical safety applications.

D. Detection of jacket

Our trained model YOLOv8m reliably identifies whether construction workers are wearing jackets, regardless of the jacket's color or style. This detection capability extends to noisy, foggy, rainy, and low-light environments, demonstrating the model's resilience and adaptability to real-world conditions. The enhanced jacket detection accuracy in these challenging environments underscores the model's practical value for safety applications.

E. Alarm initiate and Notify the server

On the construction site, if a worker is not wearing a helmet or safety jacket, the system will activate an alarm and notify the designated operator responsible for monitoring the system. This real-time alert mechanism ensures prompt identification of potential safety hazards and timely intervention to prevent accidents.

V. Results and Discussion

A. Result evaluation

The detection tasks are measured using Precision (P), Recall (R), and Average Precision (AP). One of them, called Precision, measures the proportion of positive examples that are anticipated to occur to all positive examples and it is calculated by[24]:-

$$P = \frac{TP}{TP + FP} \times 100\%$$

where, respectively, P, TP, and FP stand for precision, true positives, and false positives. Recall, which is expressed as where FN denotes false negatives, is used to describe how many of the positive samples were found in the prediction[24].

$$R = \frac{TP}{TP + FN} \times 100\%$$

Calculated from the area under the precision-recall curve, Average Precision combines P and R and can be given by[24]:-

$$AP = \frac{TP + TN}{TP + TN + FN + FP} \times 100\%$$

We have calculated the loss function of precision and recall. A loss function is a machine learning model that quantifies the difference between predicted and actual values in a machine learning model. It's a method of evaluating how well the algorithm models our data set. The loss function of precision and recall is shown below by using Fig. 4:

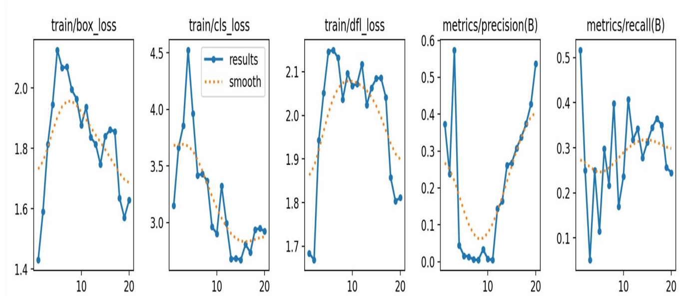

Fig. 4. Loss Precision-Recall curve

The actual measurement of confidence, precision, recall, and f1 value for detection of the jacket and helmet of our model has been shown below by using Fig. 5:

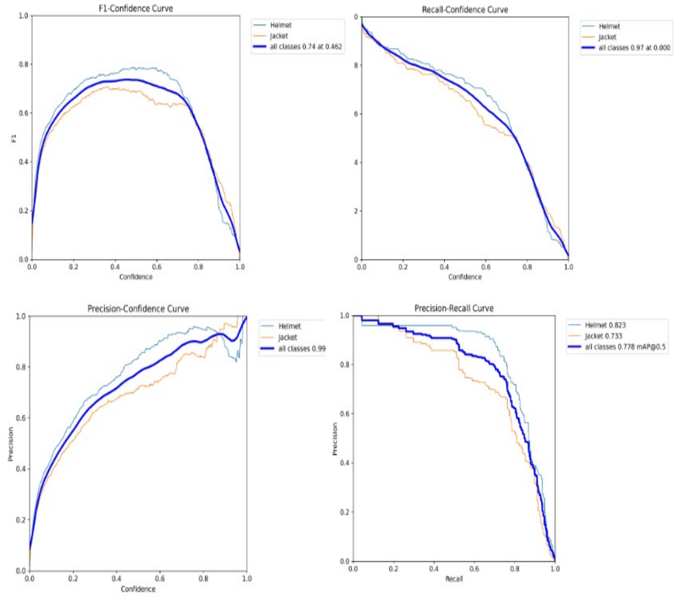

Fig. 5. P_curve, R_curve, PR_curve, F1_curve

TABLE I
COMPARISON OF OUR SYSTEM WITH THE EXISTING ONE

	Existing Systems Accuracy		
Paper No.	***Author***	***Method used***	***Accuracy(%)***
Our	Proposed System	YOLOv8	92%
[16]	Chen, S., Tang, W., Ji, T., Zhu, H., Ouyang, Y. and Wang, W	Faster R-CNN	90%
[24]	Charoenporn Bouyam, Yunyong Punsawad	YOLOv5	85.5%
[14]	Long, X., Cui, W. and Zheng, Z.	SSD	78.3%
[3]	Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M	YOLOv3	43.5%

B. Experimental results

We have tested our model on a 20% test set including rainy,low-light, foggy, and bright-light images. Here we compare the current system with the existing one, which is shown in TABLE-1

The table in Figure I provides a comprehensive comparison of the accuracy achieved by different object detection systems. The proposed system, leveraging YOLOv8, outperforms its counterparts with an impressive accuracy of 92%. In contrast, the existing methods, such as Faster R-CNN [16] by Chen et al. and YOLOv5 [24] by Charoenporn Bouyam and Yunyong Punsawad, demonstrate accuracies of 90% and 85.5%, respectively. Additionally, the SSD method [14] by Long et al. achieves an accuracy of 78.3%, while YOLOv3 [3] by Bochkovskiy et al. lags with a comparatively lower accuracy of 43.5%. This comparative analysis highlights the superior performance of the proposed YOLOv8-based system in object detection accuracy when compared to other state-of-the-art methods.

C. Sample output

The model has been tested on many images. Here is the sample output for the detection of helmets and jackets, shown below:

Fig. 6. Sample output of our trained model[25]

Fig. 7. Sample output of our trained model[25]

Upon completion of model training, the test dataset was introduced to the model, and the test outcomes are depicted in Figures 9 and 10. The value displayed on the target box represents the level of confidence in the helmet and jacket category designation. The YOLOv8 algorithm exhibited enhanced precision in identifying the target under challenging construction site conditions, irrespective of the color contrast between the helmet and the background.

VI. CONCLUSION AND FUTURE WORK

The successful integration of advanced computer vision techniques, such as YOLOv8, holds immense promise for enhancing the safety of construction workers and paves the way for the continuous development of intelligent safety solutions in high-risk work environments. To ensure the effectiveness of our system in challenging conditions, we utilized foggy, rainy weather datasets and night images of construction workers. This system will empower site operators to implement effective worker safety helmet protocols and enforce helmet usage on construction sites. Consequently, a significant number of workers engaged in megaprojects across Bangladesh will benefit from enhanced safety measures.

In this paper, we present a comprehensive investigation into the application of the YOLOv8 algorithm for the detection of helmets and jackets worn by construction workers in sensitive environments. Through a series of experiments and analyses, we have convincingly demonstrated the effectiveness and potential of YOLOv8 in accurately identifying these essential safety gear items. Our findings underscore the transformative impact of advanced computer vision techniques in promoting workplace safety and preventing accidents in the construction industry. Subsequent to the successful implementation of our helmet and jacket detection system, we were inspired to expand its functionality by incorporating additional features. We envision integrating an emergency button and a tracking device into the helmet, enabling workers to send immediate distress alerts in the event of an accident. Upon receiving an alert, the system will promptly relay the injured worker's precise location to facilitate swift rescue operations.

References

[1] Redmon, J. and Farhadi, A., 2017. YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).

[2] Redmon, J. and Farhadi, A., 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[3] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[4] Benyang, D., Xiaochun, L., and Miao, Y. (2020, November). Safety helmet detection method based on YOLO v4. In 2020 16th International Conference on Computational Intelligence and Security (CIS) (pp. 155-158). IEEE.

[5] Zhou, H., Wang, A., Li, M., Zhao, Y., and Iwahori, Y. (2021). Epidemic prevention system based on voice recognition combined with intelligent recognition of mask and helmet. 2021 3rd International Conference on Video, Signal and Image Processing. https://doi.org/10.1145/3503961.3503963

[6] Benyang, D., Xiaochun, L., and Miao, Y. (2020, November). Safety helmet detection method based on YOLO v4. In 2020 16th International conference on computational intelligence and security (CIS) (pp. 155-158). IEEE.

[7] Ahmed, S., Sobuz, M. H. R., and Haque, M. I. (2018, February). Accidents on construction sites in Bangladesh: A review. In 4th International Conference on Civil Engineering for Sustainable Development (pp. 9-11).

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

[9] Chiverton, J. (2012). Helmet presence classification with motorcycle detection and tracking. IET Intelligent Transport Systems, 6(3), 259-269.

[10] Waranusast, R., Bundon, N., Timtong, V., Tangnoi, C., and Pattanathaburt, P. (2013, November). Machine vision techniques for motorcycle safety helmet detection. In 2013 28th International conference on image and Vision Computing New Zealand (IVCNZ 2013) (pp. 35-40). IEEE.

[11] Li, J., Liu, H., Wang, T., Jiang, M., Wang, S., Li, K., and Zhao, X. (2017, February). Safety helmet wearing detection based on image processing and machine learning. In 2017 Ninth international conference on advanced computational intelligence (ICACI) (pp. 201-205). IEEE.

[12] https://safetyandrights.org/wp-content/uploads/2021/08/Death-Report-2020-1.pdf

[13] Geng, R., Ma, Y. and Huang, W., 2021, April. An improved helmet detection method for YOLOv3 on an unbalanced data set. In 2021 3rd International Conference on Advances in Computer Technology, Information Science and Communication (CTISC) (pp. 328-332). IEEE.

[14] Long, X., Cui, W. and Zheng, Z., 2019, March. Safety helmet wearing detection based on deep learning. In 2019 IEEE 3rd information technology, networking, electronic and automation control conference (ITNEC) (pp. 2495-2499). IEEE.

[15] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[16] Chen, S., Tang, W., Ji, T., Zhu, H., Ouyang, Y. and Wang, W., 2020, July. Detection of safety helmet wearing based on improved faster R-CNN. In 2020 International joint conference on neural networks (IJCNN) (pp. 1-7). IEEE.

[17] Shoukun, X., Yaru, W., Yuwan, G., Ning, L., Lihua, Z. and Lin, S., 2020. Research on helmet wearing detection based on improved Faster RCNN [J]. Computer Application Research, 37(03), pp.901-905.

[18] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016. Ssd: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing.

[19] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[20] Jocher, G., Nishimura, K., Minerva, T. and Vilariño, R., 2020. YOLOv5 https://github. com/ultralytics/yolov5. Accessed March, 7, p.2021.

[21] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S., 2020, August. End-to-end object detection with transformers. In European conference on computer vision (pp. 213-229). Cham: Springer International Publishing.

[22] Wang, C.Y., Bochkovskiy, A. and Liao, H.Y.M., 2023. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7464-7475).

[23] Guo, J., Lou, H., Chen, H., Liu, H., Gu, J., Bi, L. and Duan, X., 2023. A new detection algorithm for alien intrusion on highway. Scientific reports, 13(1), p.10667.

[24] Han, J., Liu, Y., Li, Z., Liu, Y. and Zhan, B., 2023. Safety helmet detection based on YOLOv5 driven by super-resolution reconstruction. Sensors, 23(4), p.1822.

[25] NIRAV B NAIK. (2023). Safety Helmet and Reflective Jacket. (Version 1). Kaggle. Available at: https://www.kaggle.com/code/harpdeci/yolo-nas-safety-helmet-and-vest-detection