Enhancing Safety Measure at Construction Sites through Automated PPE Kit Detection

Abstract- Safety measures at construction sites are crucial to prevent accidents and ensure the well-being of workers. In this study, we propose an automated system for Personal Protective Equipment (PPE) kit detection using computer vision techniques. The system aims to enhance safety by detecting whether workers at construction sites are wearing the necessary PPE, such as helmets, vests, and boots. We compare the performance of two popular object detection models, YOLOv9 and Faster R-CNN, to determine their effectiveness in accurately identifying PPE items in images. Our experiments demonstrate the potential of these models in automating PPE detection, thereby improving safety compliance and reducing the risk of workplace injuries.

1.INTRODUCTION

Construction sites are inherently hazardous environments where workers are exposed to various risks and dangers. Ensuring the proper utilization of Personal Protective Equipment (PPE) is crucial to mitigating these risks and safeguarding the well-being of workers. However, manual monitoring of PPE compliance at construction sites is labor-intensive, time-consuming, and prone to errors.

According to statistics from the Ministry of Labor in Taiwan, the construction industry accounted for a disproportionately high incidence rate of fatal occupational injuries in 2020, with up to 71% attributed to unsafe equipment usage [1]. Similarly, the U.S. Occupational Safety and Health Administration (OSHA) emphasizes the importance of personal protective equipment (PPE) in preventing occupational injuries and fatalities. PPE items such as helmets, safety vests, steel-toe boots, goggles, and gloves are crucial for mitigating risks associated with falling objects, electrocution, and other common hazards at construction sites [2]. For instance, the use of helmets significantly reduces the impact of falling objects and electrocution from hanging cables, while high-visibility safety vests help prevent collisions with heavy construction equipment.

Efficient management of PPE usage among construction workers is therefore paramount to enhancing workplace safety and reducing the incidence of occupational injuries and fatalities. In this context, leveraging advanced technologies such as computer vision offers promising solutions for automating the detection of PPE compliance. Object detection models, such as Version 9 of the YOLO (You Only Look Once) architecture, present an opportunity to accurately identify and monitor the usage of various PPE items in real-time.

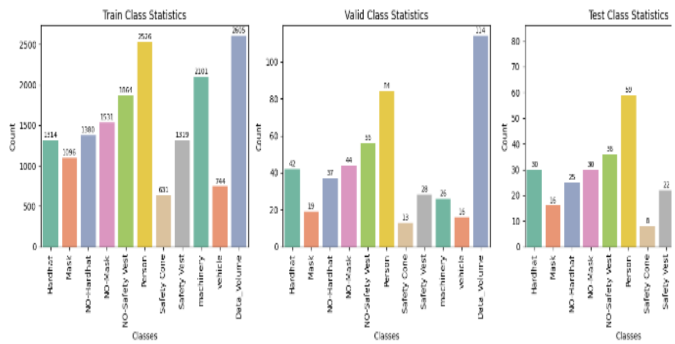
This study aims to explore the effectiveness of YOLOv9, a state-of-the-art object detection model, in detecting and monitoring the usage of PPE items among construction workers. By analyzing the performance of YOLOv9 in comparison to traditional methods and previous iterations, we seek to identify the most efficient approach for managing PPE usage in construction environments. The findings of this research have the potential to inform safety management practices, empower supervisors with real-time insights, and ultimately contribute to a safer working environment for construction workers.

2.PROBLEM STATEMENT

The aim of this project is to develop a robust computer vision system capable of automatically detecting the presence and correct utilization of PPE kits among workers at construction sites. By leveraging state-of-the-art algorithms such as YOLO (You Only Look Once), faster-RCNN (Region-based Convolutional Neural Networks), and SSD (Single Shot MultiBox Detector), we seek to achieve high accuracy and efficiency in identifying essential PPE components including helmets, safety vests, goggles, gloves, and boots.

3.DATASET

The dataset consists of images collected from construction sites with annotations for detecting various types of personal protective equipment (PPE), including hard hats, safety vests, and goggles. Each image is annotated with bounding boxes indicating the location of different PPE items.



4.RELATED WORK

4.1 Object Detection Using Deep Learning, CNNs and Vision Transformers

The research paper "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review" provides a comprehensive literature review on the evolution of object detection in the era of deep learning. It examines state-of-the-art algorithms, classifies them into different groups, and compares major convolutional neural networks for object detection. The paper covers the strengths and limitations of various object detection models and identifies future research areas in the field. It categorizes object detection models into four groups: two-stage models based on anchors, one-stage models based on anchors, anchor-free methods, and transformer-based models. The paper's findings offer valuable insights for researchers and engineers seeking to advance the field of object detection.[1] 109/ACCESS.2023.3266 093.

4.2 Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques

The research paper primarily focuses on the use of a robust CNN-based algorithm called YOLO (You Only Look Once) for real-time prediction of safety hazards on construction sites. The study experimented with YOLO, which is comparable to Faster R-CNN and SSD algorithms, and tested transfer learning on an already trained YOLO model to customize it to construction safety. The PPE detection was restricted to a hard hat and safety jacket, and the algorithm was validated on video feeds from Indian construction sites to understand its applicability in an Indian scenario. The results showed an overall accuracy of 96.27% and 96.51% in different stages of the study, indicating the effectiveness of the YOLO algorithm in detecting PPE compliance on construction sites. Additionally, the study compared the accuracy of the developed model to earlier models and found it to be comparable, demonstrating the potential of YOLO in real-time safety monitoring on construction sites.[2]

4.3 Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning

The research paper introduces a deep learning approach for detecting personal protective equipment (PPE) on construction sites using You Only Look Once (YOLO) architectures. The study presents a high-quality dataset, CHV, for PPE detection and evaluates the performance of the proposed approach. The YOLO model is used for real-time processing and detecting small instances, such as faces and helmets, with high accuracy . The results show that the proposed model achieves better performance in terms of correctness and speed compared to the state-of-the-art, with a mean average precision of 91.91% and a detection speed of 10 frames per second . The study also compares the proposed model with other deep learning methods based on the same datasets, demonstrating its superior performance.[3]

4.4 Personal Protective Equipment Kit Detection using Yolo v5 and TensorFlow

This research paper explores the use of Yolo v9 and TensorFlow for detecting Personal Protective Equipment (PPE) kits, focusing on object detection in public places for security and surveillance purposes. The study uses the MobileNet SSD, a TensorFlow model, and Yolo v9, a PyTorch framework, to achieve PPE object identification. The researchers obtained the "COVID-19 Personal Protective Equipment (PPE) Detection Dataset" from Kaggle and used performance metrics such as mAP, Loss Function, and Learning Rate to compare the two models. The findings indicate that the Yolo v9 model is more effective at categorizing data, with a significantly lower classification loss compared to the TensorFlow model . The study also emphasizes the importance of focusing on the deployability of the model, in addition to accuracy, and provides insights into the models' performance metrics.[4]

4.5 Design and Implementation of an object detection system using faster R-CNN

Recent developments in object detection are greatly driven by the success of region proposal approaches and region-based convolutional neural networks (R-CNNs). In this paper, we designed and implemented an object detection system using a faster-CNN method that shares full-image convolutional features with a detection network, so as to enable nearly cost-free region proposals. Development of this system is based on the previous work on Faster R-CNN. Results show that with this method, we could achieve high accuracy while detecting objects[5]

4.6 Deep Learning for site safety: Real-time detection of personal protective equipment

This section presents generic frameworks for verifying safety compliance for multiple PPE from visual data (image or video). In particular, three different DL-based approaches are proposed to perform the verification in real-time. Although the proposed techniques are designed to work for any number (≥1) and type of PPE (e.g., hard hat, safety vest, gloves, safety goggles, and steel toe shoes), for the scope of this study, the technical discussions, data analysis, and validation are particularly. Since YOLO-v3 requires nine anchor boxes, for each approach, all boxes in the training subset (regardless of class labels) of the *Pictor-v3* dataset are clustered into nine groups using *k*-means clustering ($k = 9$) [41]. For example, in Fig. 11, points represent geometric shapes of the training boxes (X- and Y-axis are the width and height in pixels, respectively) and colors represent different clusters. Next, a representative (centroid) from each group is selected as the anchor box, and anchor.This paper introduced, tested, and evaluated three DL-based approaches for detecting PPE attire (i.e., if a worker is wearing hat, vest, or both). The proposed methods are designed to be scalable to any number and type of PPE (e.g., not only hard hat, and safety vest but also gloves, safety goggles, and steel toe shoes). This can be achieved by modifying the last layer of the YOLO-v3 models to accommodate object classes of interest in each application. Most existing vision-based methods for monitoring PPE compliance merely focus on identifying hard hats. For example, Fang et al. used Region-based CNNs (R-CNNs) to detect if a worker is not wearing a hard hat. Wu et al. proposed a Single Shot Detector (SSD)-based algorithm to detect hard hats, and Mneymneh et al. isolated moving workers (by detecting motion) in videos and identified if any hard hat was located in or around the top area of a worker's detection box[6]

BLE I. SUMMARISATION OF EXISTING WORKS RELATED TO CONSTRUCTION SITES

Detection Method	Dataset	Challenges	Strength
Used: ResNet-152 and Faster R-CNN.	Training: ImageNet & MS COCO 2014 dataset. Testing: 3,241 images	Limited to features of construction worker's body only	Considering varying poses of the images
Used: Deep CNN and Faster R-CNN.	Training: 81,000 Testing: 19,000 images. Both were self-collected from 25 different construction projects.	Dedicated to existing construction workers.	Produce highly accurate results based on image size with average of more than 95%
Used: Google Inception v3	Training: 1208 images Testing: 27 images	It needs bigger image dataset.	Produced accuracy rate of 90%.
Used: Enhanced Faster R-CNN	Training: 9,500 images Testing: 1500 images	Full coverage of image sources.	Produced accuracy rate of more than 95%.
Used: Histogram of oriented gradients (HOG)	Training: 100 images (hardhat) & 1,800 dataset (people condition) Testing: Hundreds of self-collected images.	Limited to hardhat and worker images in standing position.	Achieved overall 94.3% precision.
Used: HOG and Circle Hough Transform (CHT).	Training: 954 images. Testing: 200 images.	Improvement for image detection.	Detection based on colours.
Used: Faster R-CNN.	Training: MIT Places Database 1129 images Testing: 333 images (263 MIT, 65 self- collected, 5 google)	Improvement of accuracy rate based on picture size refinement.	Detection safety conditions based on 3 combinations of PPEs in the form of hardhats, vests and boots.

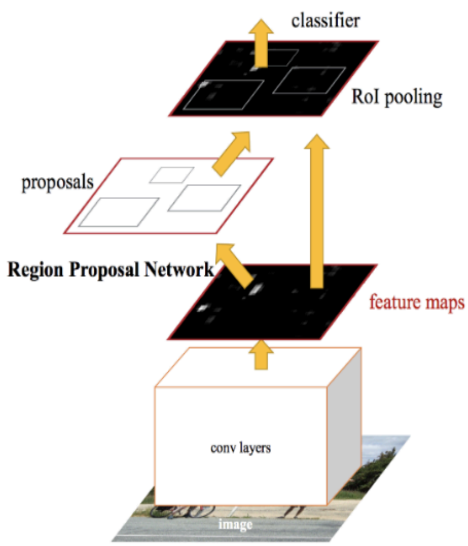
5.ALGORITHM

Faster R-CNN

Identifying and localizing objects within images or video streams is one of the key tasks in computer vision. With the arrival of deep learning, there is significant growth in this field. Faster R-CNN, a major breakthrough, has reshaped how objects are detected and categorized in real-world images. It is the advancement of R-CNN architecture. The R-CNN family operates in the following phases:

- Region proposal networks to identify possible locations in an image where objects might be present.
- CNN to extract important features.
- Classification to predict the class.
- Regression to fine-tune object bounding box coordinates

The strength of Faster R-CNN is based on its ability to reuse the CNN results for the regional proposal process. Hence, only one CNN needs to be trained, and regional proposals can be made almost cost-free computationally. Once the image has been inserted, the Faster R-CNN produces the classifications and bounding box coordinates of the specified classes in the images. In our research, this algorithm helps us to identify and to assign the safety condition based on PPE compliance. The safety condition was decided upon, as being either safe or unsafe (dangerous) based on worker compliance in terms of wearing a hardhat, vest and boots at the construction site. For safety conditions, the formulation was as follows.



- Take an input image and pass it to the ConvNet which returns feature maps for the image
- Apply Region Proposal Network (RPN) on these feature maps and get object proposals
- Apply ROI pooling layer to bring down all the proposals to the same size
- Finally, pass these proposals to a fully connected layer in order to classify any predict the bounding boxes for the image

For PPE, there are three main variables, which are hardhat, vest and boots. The formula is as follows:

Let α_i be a hardhat I, and $\alpha = \bigcap_{i=1}^{n} \alpha_i$, β_j be a vest j, and $\beta = \bigcup_{i=1}^{m} \beta_i$, γ_k be as boots $\gamma = \bigcap_{i=1}^{p} \gamma_i$

Let PPE be the PPE classification and T be the target image. S is the safety model and it can be defined as the following function:

$$f(PPE,T) = S \tag{1}$$

where, $PPE(\alpha,\beta,\gamma) = \alpha \cap \beta \cap \gamma$

$$f\ (PPEi, Tj) = Sij \tag{2}$$

where, PPE represents the PPE classification, T represents the target image and S is the safety model.

$$PPE\ (\alpha, \beta, \gamma) = \ \alpha \cap \beta \cap \gamma \tag{3}$$

$$\alpha = \ \alpha_1 \ \cup \ \alpha_2$$

$$\beta = \ \beta_1 \ \cup \ \beta_2$$

$$\gamma = \gamma_1 \ \cup \gamma_2$$

$$\begin{bmatrix} PPE_i & \alpha\,\beta & \gamma \\ \vdots & \ddots & \vdots \\ PPE_n & \cdots & \delta_n \end{bmatrix}$$

where, α_1, α_2 : with hardhat, without hardhat

β_1 , β_2 : with vest, without vest

$\gamma_1 \ \cup \gamma_2$: with boots, without boots

The evaluation of this research is based on accuracy as follows.

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \times 100 \tag{4}$$

where

TP= True positive (number of workers correctly classified as safe).
FP= False positive (number of workers incorrectly classified as unsafe).
TN= True negative (number of workers correctly classified as unsafe).
FN= False negative (number of workers incorrectly detected as safe.

YOLO (you only look once)

The YOLO (You Only Look Once) algorithm, introduced in 2016, revolutionized object detection in computer vision by offering a streamlined approach to simultaneously predicting bounding boxes and class probabilities in a single forward pass of the image through the neural network. YOLO divides the input image into a grid of cells, with each cell responsible for predicting bounding boxes and associated class probabilities. By framing object detection as a regression problem, YOLO achieves remarkable speed and accuracy compared to traditional methods. The algorithm predicts bounding boxes relative to the spatial location of each grid cell, utilizing a confidence score to represent the probability of containing an object. Through non-max suppression, YOLO efficiently filters duplicate detections, ensuring each object is detected only once. YOLO's iterations, such as YOLOv2, YOLOv3, and YOLOv4, have further refined its performance, making it a widely adopted solution for real-time object detection in various applications, including autonomous driving, surveillance, and video analysis.

GITHUB LINK

https://github.com/AKSRathor/Deep-learning/

RESULTS

REFERENCES

[1] A. B. Amjoud and M. Amrouch, "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review," in IEEE Access, vol. 11, pp. 35479-35516, 2023, doi: 10.

[2] AUTHOR=Delhi Venkata Santosh Kumar, Sankarlal R., Thomas AlbertDOI=10.3389/fbuil.2020.00136 ISSN=2297-3362

[3] Wang, Z.; Wu, Y.; Yang, L.; Thirunavukarasu, A.; Evison, C.; Zhao, Y. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. Sensors 2021, 21, 3478. https://doi.org/10.3390/s21103478

[4] Udatewar, Pallavi, Desai, Abhishek, Godghase, Gauri, Nair, Arunkumar, Kosamkar, Pranali, 2021/10/29 Personal Protective Equipment Kit Detection using Yolo v5 and TensorFlow10.1109/SMARTGENCON51891.2021.9645811

[5] Published in: 2019 International Conference on Robots & Intelligent System (ICRIS)

[6] Nipun D. Nath, Amir H. Behzadan, Stephanie G. Paal, Deep learning for site safety: Real-time detection of personal protective equipment,Automation in Construction,

[7] H. Son, H. Choi, H. Seong, and C. Kim, "Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks," Autom. Constr., vol. 99, pp. 27–38, Mar. 2019.

[8] 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," Adv. Eng. Informatics, vol. 37, pp. 139–149, Aug. 2018.

[9] D. Gil, G. Lee, and K. Jeon, "Classification of Images from Construction Sites Using a Deep-Learning Algorithm," in 35th International Symposium on Automation and Robotics in Construction (ISARC 2018), 2018, pp. 176–181.

[10] Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T.M. and An, W., "Detecting non-hardhat-use by a deep learning method from far-field surveillance videos," Autom. Constr., vol. 85, pp. 1–9, Jan. 2018.

[11] M.-W. Park, N. Elsafty, and Z. Zhu, "Hardhat-Wearing Detection for Enhancing On-Site Safety of Construction Workers," J. Constr. Eng. Manag., vol. 141, no. 9, p. 04015024, Sep. 2015.

[12] A. H. M. Rubaiyat et al., "Automatic Detection of Helmet Uses for Construction Safety," in 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW), 2017, pp. 135–142.

[13] Saudi, Madihah, Hakim, Aiman, Ahmad, Azuan, Mohd Saudi, Ahmad Shakir, Hanafi, Mohd, Narzullaev, Anvar, Ghazali, Ifwat 2020/01/01 10.14569/IJACSA.2020.0110632 International Journal of Advanced Computer Science and Applications