iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61402

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Implementing Improved YOLOv5s for Immediate Identification of Helmetless Forklift Drivers

Dr S.Lilly Sheeba[1], Shaurya Singh[2], Aritra Karan[3], Sudipta Jana[4]
[1]*Associate Professor, Department of Computer Science Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai.*
[2]*Department of Computer Science Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai*

Abstract: ***Construction workers are exposed to a multitude of health and safety risks on job sites. Despite authorities' repeated efforts to enhance safety management, incidents persist, negatively impacting both worker well-being and project progress. Therefore, it is imperative to devise effective strategies aimed at improving construction site safety management. Currently, one of the challenges in ensuring safety is the lack of a more precise method for detecting objects, particularly safety helmets. To address this issue, techniques such as Generalized Intersection Over Union (GIOU) and Soft Non-Maximum Suppression (Soft-NMS) are employed to refine the regression mechanism of bounding boxes. Furthermore, advancements have been made in training the Faster RCNN model using a multiscale approach, thereby enhancing its overall robustness.***
Index terms: Safety risks, safety helmets, Soft - NMS.

I. INTRODUCTION

Safety helmets are crucial in various indoor and outdoor workplaces like metallurgical high-temperature operations and high-rise building construction. They prevent injuries and ensure production safety. However, manual supervision is costly and prone to enforcement gaps and human factors' interference. Additionally, small target object detection lacks precision, highlighting the need for improved helmet detection algorithms. Occupational injuries in construction are a global concern, with high fatality rates despite safety regulations. The dynamic nature of construction sites increases the risk of collision accidents due to work zone complexity and limited visibility of heavy vehicles.

Given the rapid growth of the construction industry, safety has become a critical topic alongside traditional concerns like design and technology. The industry experiences a higher accident rate than others, with head injuries being a major concern. Helmets are vital for reducing head injuries by dispersing impact. Implementing helmet-wearing rules faces challenges due to supervision workload. Monitoring systems, especially those using moving cameras, offer flexibility but can suffer from data transmission issues and image quality degradation during compression.

To address these challenges, ongoing research focuses on improving the efficiency and effectiveness of helmet monitoring systems. One approach is the integration of artificial intelligence (AI) and machine learning algorithms into these systems, enabling automated analysis of helmet usage and detection of non-compliance.

(a) Forklift truck operator with helmet.

(b) Forklift truck operator without helmet

Figure 1. Typical scenes of helmetless detection on the lift truck in factory environment.

AI-powered systems can provide real-time alerts and notifications, reducing the need for manual oversight and enhancing overall safety management on construction sites. In addition to technological advancements, promoting a safety culture within the construction industry is essential. This involves continuous training and education on the importance of wearing safety helmets and adhering to safety protocols. By combining technological innovation with a proactive safety mindset, the construction industry can further enhance worker protection and reduce the incidence of occupational injuries.

II. LITERATURE REVIEW

Workers Safety Helmet Wearing Detection on Construction Sites Using Multi-Scale Features 2022 [16]. This paper proposes a deep learning method for detecting safety helmet wearing on construction sites, achieving 92.2% mean average precision with high speed and accuracy using multi-scale features.

Investigation Into Recognition Algorithm of Helmet Violation Based on YOLOv5-CBAM-DCN, 2022 [18]. This paper presents an enhanced YOLOv5-based method for recognizing safety helmet violations, addressing challenges like complex backgrounds, dense targets, and irregular helmet shapes, achieving 91.6% accuracy at 29 fps.

A Smart System for Personal Protective Equipment Detection in Industrial Environments Based on Deep Learning at the Edge, 2022 [20]. This paper presents a real-time PPE detection system using Deep Learning at the edge, enhancing safety by monitoring PPE usage in industrial settings efficiently and privately.

Detection and Location of Safety Protective Wear in Power Substation Operation Using WearEnhanced YOLOv3 Algorithm, 2021 [24]. This paper proposes a wear - enhanced YOLOv3 algorithm for real-time detection of safety protective equipment and workers in power substations, improving mean average precision by over 2%.

Helmet Use Detection of Tracked Motorcycles Using CNN-Based Multi-Task Learning 2020 [15] ,The paper proposes a CNN-based multi-task learning method for detecting helmet use in tracked motorcycles.

III. PRINCIPLES AND ENHANCEMENTS

A. YOLOv5s Network Structure

YOLOv5s represents a more compact and efficient iteration of the full-scale YOLOv5 model, meticulously designed to optimize processing speed and minimize model size without substantially compromising on detection accuracy or overall performance. This particular variant of the YOLO family is engineered to cater to applications that demand real-time object detection capabilities, even when operating on devices with constrained computational power or limited memory capacity.

1) Input Stage

Mosaic Data Augmentation: In the input stage, YOLOv5s applies Mosaic data augmentation, a technique where four images are randomly combined by scaling, cropping, and arranging them into a single image. This augmentation strategy significantly increases dataset diversity, exposing the model to a wider range of object compositions and backgrounds. As a result, YOLOv5s becomes more robust and capable of handling varied real-world scenarios during training and inference.

Adaptive Image Scaling: After Mosaic augmentation, the input image undergoes adaptive scaling to a fixed size. This process ensures that the image dimensions are suitable for processing by the neural network while minimizing the addition of black edges, which can occur during traditional resizing methods.
Adaptive Anchor Box Calculation: YOLOv5s employs adaptive anchor box calculation, a method that allows the network to dynamically learn and adjust the optimal anchor box values based on the characteristics of the training data.

2) Backbone
Flexibility in Backbone Selection: YOLOv5s provides flexibility in selecting the backbone network structure, offering options such as Darknet, CSP Darknet, or Efficient Net. Each backbone architecture has unique strengths and characteristics, allowing researchers to tailor the model's performance and efficiency based on specific application requirements and computational resources available.
Darknet: The Darknet backbone comprises multiple convolutional layers and spatial sampling operations, enabling effective feature extraction at different levels of abstraction. It forms the core foundation of YOLOv5s and plays a vital role in capturing informative features from input images for subsequent object detection tasks.
CSP Darknet: CSP Darknet introduces the Cross-Stage Partial (CSP) module, which enhances feature representation by facilitating cross-stage connections. This architecture improves information flow and enables the model to learn more discriminative features, leading to enhanced detection accuracy and robustness.
Efficient Net: Efficient Net is a scalable backbone network architecture that optimizes network structure and depth scaling coefficients through automated structure search. It achieves superior performance with relatively fewer computational resources, making it an efficient choice for applications requiring high accuracy with limited computational overhead.

3) Neck
Feature Fusion and Enhancement: The neck module, which may include architectures like Feature Pyramid Network (FPN) or Path Aggregation Network (PAN), plays a crucial role in enhancing feature representation and spatial information integration.
FPN or PAN: FPN and PAN utilize multiple convolutional layers, up sampling, and down sampling operations to fuse and connect feature maps from different levels of the backbone network. This fusion process enables the generation of feature maps with increased semantic information and multiscale feature representation.
Semantic Information Fusion: By merging and connecting feature maps from various backbone levels, the neck module synergizes low-level and high-level feature information. This fusion enhances the model's capability to represent complex spatial structures and object contexts, contributing to improved detection performance and accuracy.

4) Output Detection
Anchor-Based Object Detection: YOLOv5s adopts an anchor-based approach for object detection, which involves predicting bounding box positions and object categories across different scales of feature maps.
Convolutional and Fully Connected Layers: The output detection head of YOLOv5s utilizes a combination of convolutional and fully connected layers to generate predictions. These layers work collaboratively to process refined features from the neck module and produce accurate object detection results.
Adaptability and Performance Enhancement: Through the incorporation of data augmentation, adaptive image scaling, adaptive anchor box calculation, and advanced loss functions, YOLOv5s demonstrates enhanced performance and adaptability across different hardware configurations and real-world scenarios. These optimizations contribute to the model's ability to achieve robust object detection results in real-time applications with limited computational resources.

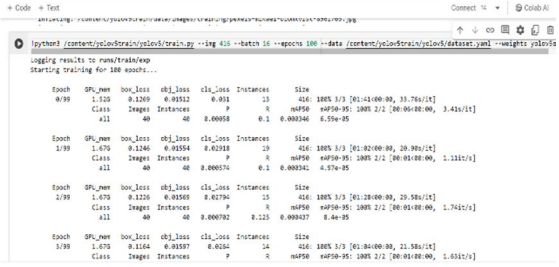


Figure 2: Training the model.

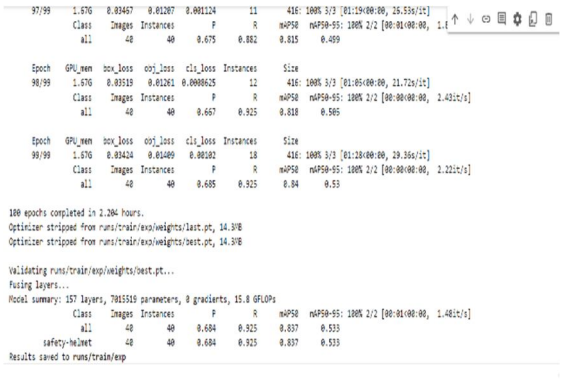


Figure 3: Model trained and results are stored.

B. Project Modules

1) Module 1: Image Preprocessing

Preprocessing encompasses various image operations at the fundamental level of abstraction, where both input and output are intensity images. These iconic images mirror the original sensor-captured data, typically represented as a matrix of brightness values. The primary goal of preprocessing is to refine image data by mitigating distortions or enhancing crucial image features for subsequent processing steps. Histogram equalization stands out as a widely recognized contrast enhancement technique, valued for its effectiveness across various image types. It offers a sophisticated means of adjusting an image's dynamic range and contrast by reshaping its intensity histogram. Unlike simple contrast stretching, histogram equalization employs nonlinear and non-monotonic transfer functions to map pixel intensities between input and output images.

In addition to histogram equalization, other preprocessing techniques play vital roles in image enhancement and analysis.

For instance, noise reduction algorithms are crucial for improving image quality by minimizing unwanted artifacts introduced during image acquisition or processing.

2) Module 2: Feature Vector Processing

The proposed fully connected network relies heavily on the inverted residual structure, incorporating a bottleneck level connected with a residual connection. Mobile Net architecture is leveraged in this phase to extract features. The CNN architecture comprises input, fully connected (FC), convolutional, pooling, and output layers. Unlike standard NNs, it emphasizes weighted sharing, local connection, and down sampling, which can enhance efficiency by eliminating local features, mitigating overfitting, and reducing network parameters. The convolutional layer serves as a fundamental building block in CNNs, facilitating local feature extraction by interconnecting each neuron's input with the local sensing area of the previous layer. Within the convolutional layer, functions are divided into activation and convolution, each contributing to the computation process. This approach optimizes feature vector processing, leading to improved performance and accuracy in extracting meaningful features from input data.

$$T = f_k\left(\sum_{x,y,z=1}^{r} C_{x,y,z}\ w^s_{x,y,z} + b^s\right)$$

In the above T and C represent the input and resultant of the convolution layer; f indicates the activation function of kth layers; r and s characterize the sequence number of convolutions and the channel count; w and b denotes the weight and bias
of convolution, x, y, and z characterize the dimensional of the input dataset.

3) Module 3: Context-Aware Multi-Feature Learning Network

The network, based on VGG-16, serves as the backbone architecture for learning image features and is initialized with pretrained weights from the ImageNet dataset. The core principle of the faster R-CNN is its Region Proposal Network (RPN). Unlike the fast R-CNN, it eliminates selective search and calculates the region of interest (ROI) directly through the RPN. Initially, the CNN extracts a feature map, which is then utilized by the RPN to compute the ROI.

The RPN, being fully convolutional, efficiently generates a variety of region proposals for the detection network, which refines these proposals to produce a fixed-length feature vector for each ROI via an ROI pooling layer.

This network operates by sliding a small window across the shared convolutional feature map to create region proposals. In the final convolutional layer, 3x3 windows are chosen to reduce dimensionality, followed by two 1x1 convolutional layers—one for localization and the other for categorization (background or foreground). The faster R-CNN provides probabilities for vehicle and pedestrian classes observed from UAVs, along with their bounding box coordinates. The RPN generates a list of candidate bounding boxes from the RGB aerial image, each with a confidence score reflecting its likelihood of belonging to object classes versus the background.

4) Module 4: Evaluation Index and Loss Function

This paper utilizes various evaluation metrics to assess the detection performance of deep neural network algorithms. These metrics include precision, recall rate, intersection over union (IoU), mean precision (mP), and mean average precision (mAP). The evaluation criteria assess the strengths and weaknesses of each attribute based on the corresponding evaluation index values. Larger values indicate better attribute performance. The accuracy, recall, IoU, and mAP formulas are defined as follows:

1. **Accuracy:**

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. **Precision:**

$$Precision = \frac{TP}{TP + FP}$$

3. **Recall:**

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

4. **f1-score:**

$$f1 - score = \frac{2 * Precision * Recall}{Precision + Recall}$$

5. **mAP:**

$$mAP = \sum_{k=1}^{n} P(k)\ \Delta r(k)$$

C. Implementation Of Realtime Detection System

The real-time monitoring system is used for detecting whether personnel operating the lift truck wear safety helmets. It operates on a server architecture and offers real-time viewing and statistical analysis functions as shown in figure 4. The system consists primarily of three components: cameras, streaming media server, and clients.

Cameras capture real-time video streams, while streaming media servers receive, process, store, and transmit these streams. Clients are used for real-time viewing and conducting statistical analysis. In achieve real-time recognition and viewing effects, cameras need to support common streaming protocols like real-time streaming protocol (RTSP). RTSP, an application layer control protocol, is based on the client/server model and controls the transmission of real-time streaming media data. Cameras push real-time captured video streams to streaming media servers or clients using the RTSP protocol to achieve real-time monitoring and viewing functions.

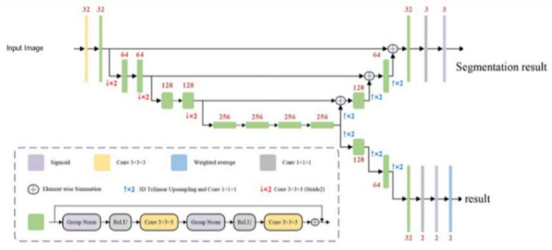


Figure 4: Architecture diagram

Videos are then encoded and transcoded to meet the requirements of different transmissions and devices. Subsequently, the server stores and transmits real-time video streams, establishing connections with clients through wireless RTSP stream or RTSP stream. The server acts as an intermediary between cameras and clients.

IV. ALGORITHM

```
import cv2
import torch
import numpy as np

path='C:/Users/shaur/Downloads/yolov5safetyhelmet-main/best.pt'
model=torch.hub.load('ultralytics/yolov5','custom',path, force_reload=True)
cap=cv2.VideoCapture('helmet.mp4')
count=0
while True:
    ret,frame=cap.read()
    if not ret:
        break
    count += 1
    if count % 3 != 0:
        continue
    frame=cv2.resize(frame,(1020,600))
    results = model(frame)
    frame = np.squeeze(results.render()
    results=model(frame)
    cv2.imshow("FRAME",frame)
    if cv2.waitKey(1)&0xFF==27:
        break
cap.release()
cv2.destroyAllWindows()
```

V. OUTPUT

Figure 5: Helmet detection in Real time

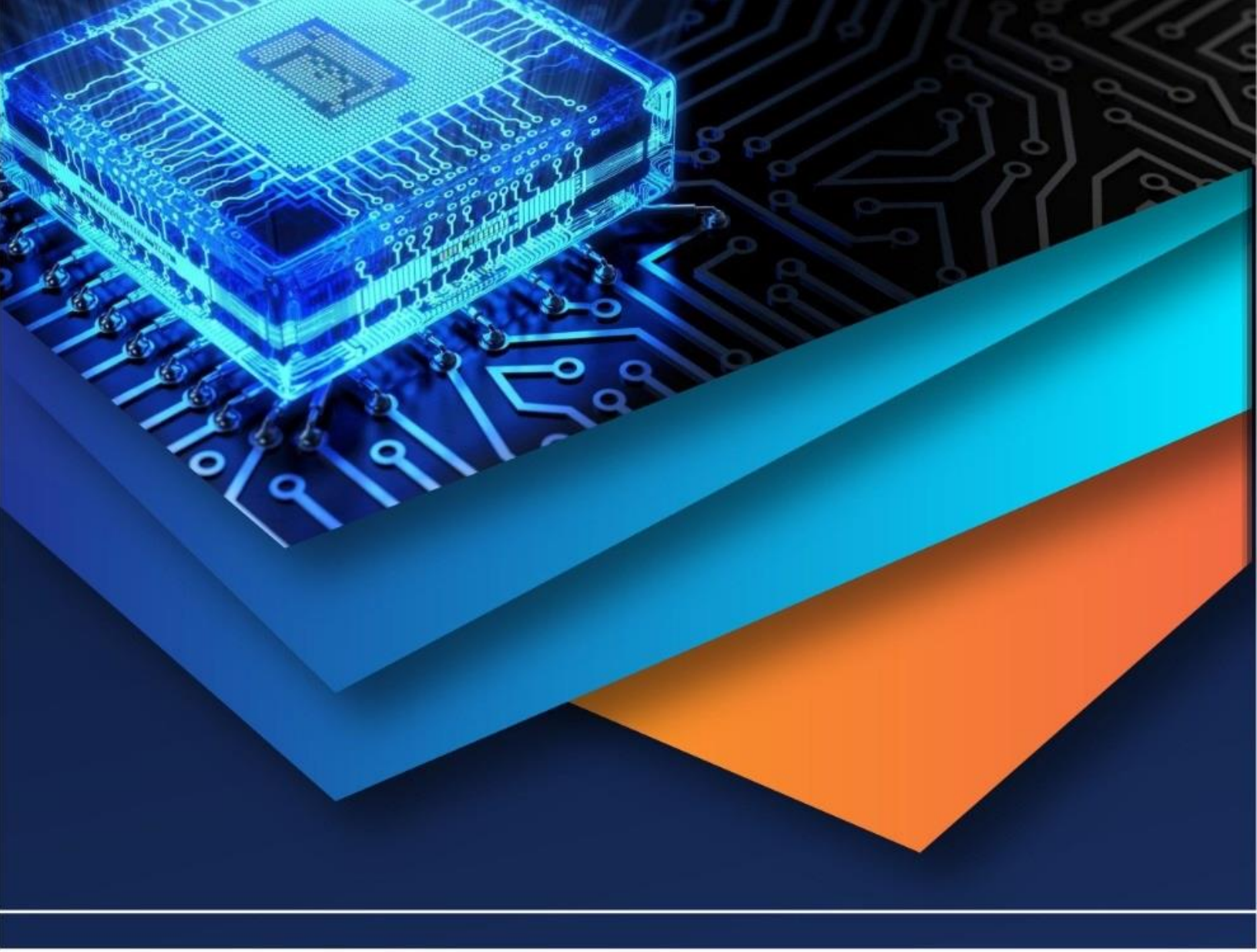
VI. CONCLUSION

Based on the chosen Faster-RCNN model for feature extraction, improvements have been made by eliminating down sampling and substituting the convolution kernel. This effectively minimizes information loss in the original image, especially regarding safety helmet objects.

The two-stage Faster RCNN algorithm has been enhanced specifically for safety helmet object detection tasks. Improvements include refining the backbone network, addressing imbalanced positive and negative samples, enhancing bounding box regression mechanisms, and implementing multi-scale training strategies. Comparative experiments validate the feasibility of using the Res2Net101 network module as the feature extraction structure in the improved Faster RCNN. Additionally, ablation experiments are conducted to systematically evaluate each improvement component of the Faster RCNN underwater object detection algorithm.

REFERENCES

[1] Hayat and F. Morgado-Dias, ‘‘Deep learning-based automatic safety helmet detection system for construction safety,’’ Appl. Sci., vol. 12, no. 16,p. 8268, 2022.
[2] M. D. Benedetto, F. Carrara, E. Meloni, G. Amato and C. Gennaro, Learning accurate personal protective equipment detection from virtual worlds, Multimedia Tools Appl., vol. 80, pp. 1-13, Aug. 2020.
[3] Y. Li, H. Wei, Z. Han, J. Huang and W. Wang, Deep learning-based safety helmet detection in engineering management based on convolutional neural networks, Adv. Civil Eng., vol. 2020, pp. 1-10, Sep. 2020
[4] P. H. M. de Andrade, J. M. M. Villanueva and H. D. D. M. Braz, An outliers processing module based on artificial intelligence for substations metering system, IEEE Trans. Power Syst., vol. 35, no. 5, pp. 3400-3409, Sep. 2020.
[5] Z. Ma, Detection on wearing behavior of safety helmet based on machine learning method, Urban rural Stud., vol. 7, no. 3, pp. 52-55, Feb. 2022.
[6] Y.-Z. Xiao, Z.-Q. Tian, J.-C. Yu, Y.-S. Zhang, S. Liu, S.-Y. Du, et al., A review of object detection based on deep learning, Multimedia Tools Appl., vol. 79, no. 33, pp. 23729-23791, Sep. 2020.
[7] H. Pan, Y. Li and D. Zhao, Recognizing human behaviors from surveillance videos using the SSD algorithm, J. Supercomput., vol. 77, no. 7, pp. 6852-6870, Jan. 2021.
[8] T. Umar, Applications of drones for safety inspection in the Gulf Cooperation Council construction, Eng. Construct. Archit. Manage., vol. 28, no. 9, pp. 2337-2360, Dec. 2020.
[9] B. Wang, W. Li, and H. Tang, ‘‘Improved YOLOv3 algorithm and its application in helmet detection,’’ Comput. Eng. Appl., vol. 56, no. 9, pp. 33–40, 2020.
[10] J. Shen, X. Xiong, Y. Li, W. He, P. Li, and X. Zheng, ‘‘Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning,’’ Comput.-Aided Civil Infrastruct. Eng., vol. 36, no. 2, pp. 180–196, Feb. 2021.
[11] M. Tan, R. Pang, and Q. V. Le, ‘‘EfficientDet: Scalable and efficient object detection,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 10778–10787.
[12] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Liao, ‘‘YOLOv4: Optimal speed and accuracy of object detection,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Apr. 2020, pp. 1–17.
[13] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, ‘‘CSPNet: A new backbone that can enhance learning capability of CNN,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2020, pp. 390–391.
[14] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, ‘‘Distance-IoU loss: Faster and better learning for bounding box regression,’’ in Proc. AAAI Conf. Artif. Intell., Apr. 2020, vol. 34, no. 7, pp. 12993–13000.
[15] Hanhe Lin, Jeremiah D. Deng , Deike Albers and Felix Wilhelm Siebert, Helmet Use Detection of Tracked Motorcycles Using CNN-Based Multi-Task Learning, 02 Sep 2020 - IEEE Access - Vol. 8, pp 162073-162084.
[16] Kun Han and Xiangdong Zeng, Deep Learning-Based Workers Safety Helmet Wearing Detection on Construction Sites Using Multi-Scale Features 01 Jan 2022 - IEEE Access - Vol. 10, pp 718-729.
[17] K. Han and X. Zeng, ‘‘Deep learning-based workers safety helmet wearing detection on construction sites using multi-scale features,’’ IEEE Access, vol. 10, pp. 718–729, 2022.
[18] Lijun Wang , Yunyu Cao , Song Wang , Xiaona Song , Shenfeng Zhang , Jianyong Zhang andJinxing Niu, Investigation Into Recognition Algorithm of Helmet Violation Based on YOLOv5-CBAM-DCN01 Jan 2022- IEEE Access - Vol. 10, pp 60622-60632
[19] W Tai, Z. Wang, W. Li, J. Cheng, and X. Hong, ‘‘DAAM-YOLOv5: A helmet detection algorithm combined with dynamic anchor box and attention mechanism,’’ Electronics, vol. 12, no. 9, p. 2094, 2094.
[20] Gionatan Gallo , Francesco Di Rienzo , Federico Garzelli , Pietro Ducange and Carlo Vallat ,A Smart System for Personal Protective Equipment Detection in Industrial Environments Based on Deep Learning at the Edge, 01 Jan 2022 - IEEE Access - Vol. 10, pp 110862-110878
[21] Z. Jin, P. Qu, C. Sun, M. Luo, Y. Gui, J. Zhang, and H. Liu, ‘‘DWCAYOLOv5: An improve single shot detector for safety helmet detection,’’ J. Sensors, vol. 2021, Oct. 2021, Art. no. 4746516.
[22] H. Cai, J. Li, M. Hu, C. Gan, and S. Han, ‘‘EfficientViT: Lightweight multi-scale attention for on-device semantic segmentation,’’ 2023, arXiv:2205.14756.
[23] Y. Sun, G. Chen, T. Zhou, Y. Zhang, and N. Liu, ‘‘Context-aware cross-level fusion network for camouflaged object detection,’’ 2021, arXiv:2105.12555.
[24] Baining Zhao, Haijuan Lan, Zhewen Niu, Huiling Zhu , Tong Qian and Wenhu Tang, Detection and Location of Safety Protective Wear in Power Substation Operation Using Wear-Enhanced YOLOv3 Algorithm, 13 Aug 2021 - IEEE Access - Vol. 9, pp 125540-125549.
[25] B. Koonce and B. Koonce, ‘‘EfficientNe convolutional neural networks with swift for TensorFlow: Image recognition and dataset categorization,’’ Tech. Rep., 2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY