Indonesian Journal of Electrical Engineering and Computer Science
Vol. 34, No. 3, June 2024, pp. 1935~1943
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v34.i3.pp1935-1943

Improving industrial security device detection with convolutional neural networks

Orlando Iparraguirre-Villanueva[1], Josemaria Gonzales-Huaman[1], Jose Machuca-Solano[1], John Ruiz-Alvarado[2]
[1]Facultad de Ingeniería y Arquitectura, Universidad Autónoma del Perú, Lima, Perú
[2]Facultad de Ingeniería, Universidad Tecnológica del Perú, Lima, Perú

Article Info

Article history:

Received Jan 8, 2024
Revised Feb 25, 2024
Accepted Mar 4, 2024

Keywords:

CNNs
Machine vision
Security
Sensing
YOLOv5

ABSTRACT

Employee safety is paramount in the manufacturing industry to ensure their well-being and protection. Technological advancements, particularly convolutional neural networks (CNN), have significantly enhanced this safety aspect by facilitating object detection and recognition. This project aims to utilize CNN technology to detect personal protective equipment and implement a safety implement detection system. The CNN architecture with the YOLOv5x model was employed to train a dataset. Dataset videos were converted into frames, with resolution scale adjustments made during the data collection phase. Subsequently, the dataset was labeled, underwent data cleaning, and label and bounding box revisions. The results revealed significant metrics in safety equipment detection in industrial settings. Helmet precision reached 91%, with a recall of 74%. Goggles achieved 85% precision and an 87% recall. Mask absence recorded 92% precision and an 89% recall. The YOLOv5x model exhibited commendable performance, showcasing its robust ability to accurately locate and detect objects. In conclusion, the utilization of a CNN-based safety equipment detection system, such as YOLOv5x, has yielded substantial improvements in both speed and accuracy. These findings lay a solid foundation for future industrial security applications aimed at safeguarding workers, fostering responsible workplace behavior, and optimizing the utilization of information technology resources.

Corresponding Author:

Orlando Iparraguirre-Villanueva
Facultad de Ingeniería y Arquitectura, Universidad Autónoma del Perú
Panamericana Sur Km. 16.3, Villa EL Salvador, Lima, Perú
Email: oiparraguirre@ieee.org

1. INTRODUCTION

In industry, improper use or lack of safety equipment is one of the main causes of injuries and accidents, which can be avoided if safety equipment is used while working. According to the ILO [1], some 317 million people are affected by occupational injuries each year, and 234 million die from work-related accidents or diseases [2]. In 2021, the U.S. Bureau of Labor Statistics reported 5,190 cases of fatal work injuries [3], an increase of 8.9% over 2020. The fatal work injury rate was 3.6 deaths per 100,000 full-time equivalent workers, up from 3.4 in 2020 and 3.5 before the 2019 pandemic. These figures were obtained from the census of fatal occupational injuries [4]. Manufacturing has a 24.96% share, followed closely by real estate, business, and renting activities with 16.72%.

Machine vision is currently used as an effective method to capture and identify images, videos, actions, and shapes. Machine vision algorithms are more accurate than sensors and can be implemented in video surveillance systems, saving time and money [5], [6]. With proper training, these algorithms can prevent workplace accidents by detecting whether workers are wearing their safety equipment. Many industries, such as architecture, engineering, and construction, have shown interest in machine vision to address challenges and recognize interactions between collaborators and their work items, providing robust data to support and achieve high-level understanding [7].

Convolutional neural networks (CNNs) are a type of artificial neural network consisting of four convolutional layers and other reduction layers, which are alternatively assigned [8]. Finally, fully connected layers are added, like a multilayer perceptron network. Many high-performance methods in computer vision software are based on CNNs, which allow computers to learn about the context of visual data from images and/or videos [9]. There are several ways to detect whether workers are wearing their safety equipment. First, the algorithm is trained to detect objects quickly and accurately based on their geometry, appearance, and features [10], [11]. Computer vision is a discipline that is used in different everyday tasks and industries, mainly to improve the capability of machines, to control, and track people’s movements and to ensure people’s safety [12], [13].

The potential of this research lies in improving the safety of industrial workers. Making it possible to accurately detect whether a worker is wearing safety equipment to ensure their well-being at work, can lead to the development of more effective safety protocols and the prevention of accidents and injuries. In turn, this may lead to the development of more effective safety protocols and the prevention of accidents and injuries. Ultimately, the impact of this study can be felt in the lives of workers, as it provides a means to protect them from harm while on the job. It is a crucial step toward creating a safer work environment for all. This work aims to create a detection system that uses CNNs to identify personal protective equipment in the industrial sector.

In recent years, researchers and scholars have devoted their efforts to exploring technologies aimed at improving industrial safety. For example, Chen *et al.* [14] developed a real-time hard hat detection algorithm using a variant of faster R-CNN. The network was trained on a labeled dataset that included images of workers wearing and without hard hats. It is evaluated by testing and compared to other existing methods, showing better performance compared to faster R-CNN and you only live once (YOLO). On the other hand, research in [15] developed a real-time hard hat detection system using deep learning technology. In addition, they used a CNN trained on an annotated dataset of workers with and without hard hats. The high accuracy achieved by the system shows its great potential to improve safety in the construction industry by preventing accidents and protecting the integrity of workers. Li *et al.* [16] developed a protective helmet detection system using CNN in computer vision. The system was evaluated under various lighting and positioning conditions, demonstrating its ability to detect protective helmets in real environments. Providing effective solutions to increase safety in the construction industry, help prevent accidents, and protect workers. Very similarly, Park *et al.* [17] developed a system to detect patterns of people wearing hard hats in different environments and under different lighting conditions using CNN. In addition, Vasanthakumar *et al.* [18] developed a deep learning-based system to detect hard hats in an industrial environment. Use deep CNNs and annotated datasets to demonstrate its effectiveness in improving industry safety by protecting workers. Also, Kong *et al.* [19] used CNNs to develop an industrial construction site hard hat identification system. To do this, they used a significant dataset that included images with hard hats on these sites. This they appreciated by mimicking different lighting conditions, and the test environment demonstrated their ability to identify hard hats according to the facts. Similarly, Krizhevsky *et al.* [20] developed a system for detecting security obstacles with CNN; the system was trained using a dataset containing images from security cameras. They adjusted the parameters and network structure to achieve higher accuracy in detecting blockages in buildings under construction. These results confirm the strength and reliability of the system to verify that the lack of road blockage provides valuable solutions to prevent accidents and protect employees. Finally, Bitirgen and Filik [21] use computer vision and long short-term memory (LSTM) networks to predict the uncertain behavior of the construction industry. The developed model shows high accuracy in identifying and predicting dangerous situations in construction sites. These results support the use of these technologies to improve construction safety.

The research is organized as follows: section 2 discusses related work regarding personal protective equipment using CNNs as an identification technique. Section 3 describes the methodology used to develop the system, while section 4 presents the training findings. Section 5 provides a brief discussion of previous work and, finally, section 6 presents the study’s conclusions.

2. METHOD

In this section, we propose to use the CNN architecture with the YOLOv5 model to detect whether workers have their safety equipment on when entering the production area. This is done by analyzing images with a previously trained computer vision system and using the object detection algorithm to classify workers "with safety equipment" or "without safety equipment". The CNN architecture is designed to process data with a grid structure, such as images. It is a multilayer neural network that extracts dependencies from structured grid inputs such as images and text [22]. The convolution operation applied at many intermediate layers is the most important property of CNNs and involves the scalar product of a set of grid-structured weights and another set of similarly structured inputs [23]. The architecture of a CNN is shown in Figure 1.

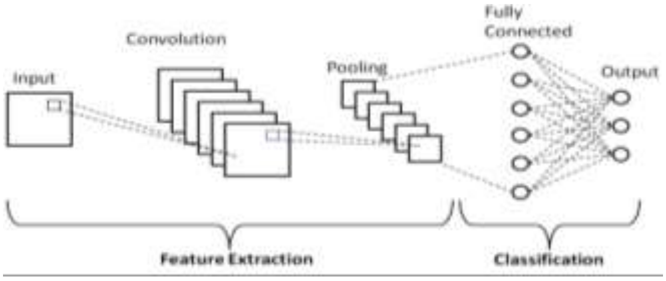

Figure 1. Basic CNN architecture

Convolutional layers perform local convolutions to obtain attributes, while layer clustering reduces spatial resolution and highlights invariant features. Here is an overview of a typical CNN architecture:
- Input layer: the image is represented as a multidimensional tensor, usually with height, width, and channels.
- Convolutional layers: CNN is based on a convolutional layer, each of which applies a series of filters (also called kernels) to the input image to extract the corresponding features. Each filter moves through the image and performs a convolution operation. The following are nonlinear activation functions such as ReLU. These layers learn to recognize simple image patterns and features such as edges, texture, and color [24].
- Reduction layer: a reduction layer is usually added after the convolution layer to improve computational efficiency and reduce the size of the resulting features. The subsampling operation highlights the most important features and reduces the spatial dimension of the features [25].
- Fully connected layer: a fully connected layer, also known as a dense layer, is used after the convolution and reduction layers to perform the final classification. These layers combine the obtained features to obtain the desired result, such as classification probability [26].
- Output layer: the result of the CNN is provided by the output layer. Depending on the problem, one neuron with an appropriate activation function (e.g., sigmoid functions for binary classification problems) or multiple neurons with a SoftMax activation function for multiclass classification problems can be used [27]. Several models and algorithms are available for real-time object recognition. In this study, we chose the YOLOv5 model because it provides reliable results in real-time video object detection. In addition, this deep neural network is characterized by its ease of use and adaptability [28].

2.1. YOLOv5

It is a computer vision detection model that uses a CNN-based architecture and is an enhanced version of YOLO designed to improve detection accuracy and speed. Table 1 shows a comparison of the architectures [29]. YOLOv5 can handle images of different sizes without resizing them beforehand. It facilitates the detection of objects in high-resolution images without loss of accuracy and can be implemented on different platforms. In Figure 2, the results show that YOLOv5x offers significant improvements in detection speed and accuracy compared to YOLOv5s, YOLOv5m, and YOLOv5, especially when detecting small or distant objects. Although it can be observed that EfficientDet has higher performance, it consumes more computer time and resources, unlike YOLOv5x, which has similar performance and lower time and resource consumption.

2.2. Data set creation

This dataset comes from an industrial company in Peru and consists of 750 short videos. The videos are carefully selected and show how employees use safety equipment in their daily activities. The safety equipment seen in the video varied in type, color, and proximity, providing different scenarios for analysis. The videos were filmed in a variety of conditions, resulting in significant differences in image quality, tones, and environment. To facilitate understanding and classification of the data, the videos were divided into individual images by capturing frames from each video. Manual labeling is then performed to properly annotate each image with relevant information about the site's security tools. A representation of the dataset is presented in Figure 3.

Table 1. Comparison of architectures

Model	YOLOv3	YOLOv4	YOLOv5
Model weight	237 MB	244MB	167 MB
Parameters	63 million	63 million	86.7 million
GLOFPs	157.3	187.3	205.7
Latency	11.78	14.3	16.5
Neural network architecture	Darknet-53	(CSPNet, SPP)	CSPDarknet-53
Accuracy	Enhanced	Highly accurate	Media
Small object detection	Flexible	Flexible	Very flexible
Multiplatform implementation	Yes	Yes	Yes

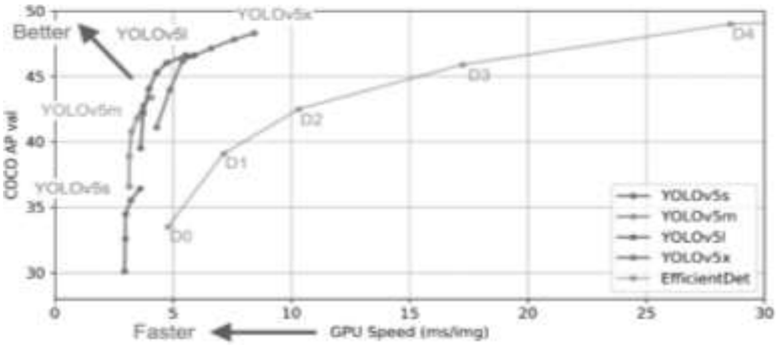

Figure 2. YOLOv5 performance comparison

Figure 3. Characteristics of the data set

2.3. Conversion and labeling of the dataset

At this stage, the video is converted into still images. To extract frames from each of the videos several programs perform this process, in this case, the video to JPG converter tool is used, which allows extracting all the frames from the video, but the results were not convincing, therefore. Considering this problem, it was decided to extract frames per second, so that the system works without false positives and false negatives. All images have been resized to 600×600 to match the pixel resolution of YOLO.

It is important to label each object with the appropriate class and place the object in a bounding box. Each video frame is given labels so that the model can understand what it is trying to detect. This phase is the most time-consuming because the labeling of the various personal safety equipment is done manually. Artificial Intelligence tools were used which allow tagging one or more personal protective equipment over an image, given that the image may contain one or more people wearing helmets, gloves, and masks. Figure 4 shows the labeling of objects. Next, the images are exported in YOLO format, where the name attached to the box label and the coordinate name can be seen. This is important so that, at the time of training, it provides the necessary information during the training so that the model can understand and identify different objects related to the classes.

Figure 4. Data set labeling

2.4. Training

In the training phase, you work with the labeled data set, which is divided into training (1,051) and validation (375). The training set is used to train the model and the validation set is used to evaluate the performance during training. Next, the model is configured by defining the classes to be instrumented, the parameterization of the model is done by specifying the number of images, the batch size, and the training period. Also, during training, the data set is passed to the model and a loss is calculated for each image, which can be interpreted as the difference between the model predictions and the actual marks.

To perform the training, we used a computer with an Intel Core i7 7400 processor, 16 GB of RAM, and a dedicated GTX 1080 graphics card with 8 GB of VRAM. Then we proceeded with the training of the model by loading the data set that will help the model to improve its accuracy in each epoch. The system can associate patterns found in these features with appropriate categories during training. Initially, these associations may be incorrect, but as training progresses, the model will adapt and improve the accuracy of the assigned tasks.

3. RESULTS AND DISCUSSION

Detection of safety equipment in industrial environments is a critical challenge to ensure worker integrity and operational efficiency. Worker safety in the commercial sector is a growing concern, and early detection of safety equipment such as hard hats, reflective vests, and safety glasses is important to prevent accidents and ensure a safe working environment. In the past, these tasks were performed by manual methods, which have limitations in terms of accuracy and efficiency. This study used CNN, a deep learning technique that shows great potential in detecting and classifying objects in images, to create a detection solution for identifying personal protective equipment in industrial environments. The main findings of this study are presented, focusing on the performance of the YOLOv5 model in CNN-based security equipment detection compared to traditional methods, as well as evaluating the performance of the model on the dataset and industrial scenarios by analyzing metrics such as accuracy, recall, mean accuracy value (mAP), and F1-scores. Table 2 shows the results of the training.

Table 2. Training results

Classes	Validation	Instance	Accuracy	Recall	mAP
Helmet	375	297	0.91	0.74	0.74
Lenses	375	89	0.85	0.87	0.87
Mask	375	84	0.93	0.60	0.64
Without helmet	375	186	0.91	0.92	0.94
Without glasses	375	133	0.79	0.74	0.71
Without mask	375	187	0.92	0.89	0.91

Table 2 shows the results in terms of metrics such as accuracy, recall, and mAP for the different types of safety equipment, such as helmets, goggles, masks, and the absence of these elements. For example, for helmet identification, a high accuracy of 91% was obtained. However, the count of 74% indicates that it fails to identify all positive cases of helmets, which opens an area for improvement. Concerning eyeglasses, the model obtained solid results, with an accuracy of 85% and a recall of 87%. These values are quite significant, indicating that the model is efficient in detecting instances of spectacles. The model also showed high accuracy in detecting workers who were not wearing their safety equipment, such as, for example, people who were not wearing a protective mask, in this case, the absence of a mask obtained a high accuracy of 92% and a recall of 89%, which indicates that the model is efficient in identifying people who do not wear safety equipment in industrial environments.

To reinforce this analysis can be seen in Figure 5, Figure 5(a) displays the performance per epoch for the mAP 0.5 series, and it is the highest among the four series. Performance steadily increases over the 50 epochs, reaching a peak of 0.8. This indicates that the model is successfully learning to classify images accurately. Figure 5(b) illustrates the performance per epoch for the mAP 0.5:0.95 series, which is lower than that of Figure 5(a) but still relatively high. Performance steadily increases over the 50 epochs, reaching a maximum of 0.7. This indicates that the model is also successfully learning to classify images accurately, though not as precisely as the series presented in Figure 5(a). Figure 5(c) presents performance per epoch for the series, which is lower than that of the series presented in Figures 5(a) and 5(b) but still reasonable. Performance steadily increases over the 50 epochs, reaching a peak of 0.65. This indicates that the model is learning to classify images accurately. In Figure 5(d), the performance per epoch for the series is presented, which is the lowest among the four series shown in Figures 5(a) to 5(c). Performance steadily increases over the 50 epochs, reaching a maximum of 0.55. This indicates that the model is learning to classify images accurately, but not as precisely as the models presented in Figures 5(a) to 5(c).

The different metrics used provide varying perspectives on the model's performance. The mAP 0.5 is a measure of the overall accuracy of the model, while the mAP 0.5:0.95 is a measure of the model's accuracy in detecting small objects. Precision is a measure of the proportion of objects correctly classified by the model, while recall is a measure of the proportion of actual objects that are correctly classified by the model.

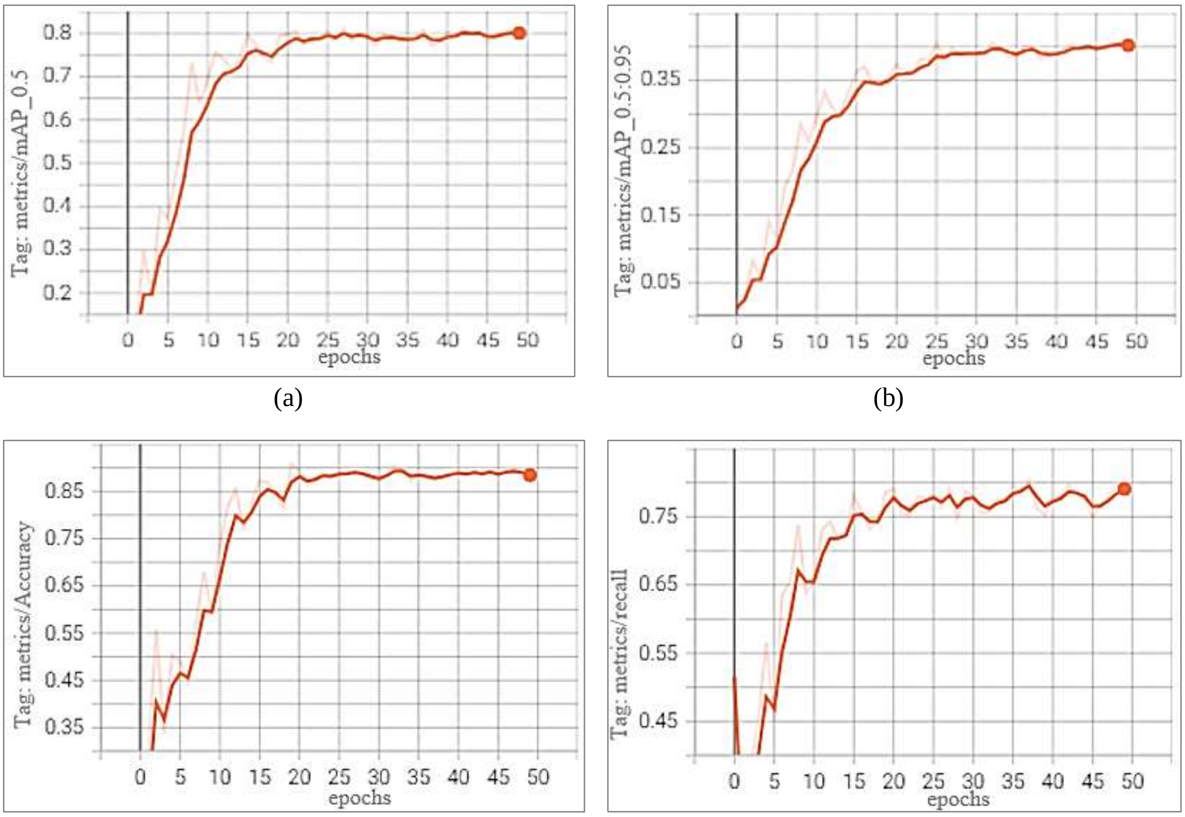

Figure 5. Performance per epoch for the metric; (a) mAP 0.5, (b) mAP 0.5:0.95, (c) accuracy, and (d) recall

After training the YOLOv5 detection model, its performance was evaluated using Tensorboard and Pytorch. A total of 1051 images were analyzed over 50 epochs, and two variations were calculated: mAP(0.5:0.95) and mAP(0.5). The former follows the common objects in context (COCO) parameterization, while the latter uses pascal visual object classes (VOC). The results obtained were 0.80 for mAP (0.5) and 0.40 for mAP(0.5:0.95). When analyzing the figures, it is seen that the accuracy values stabilize after epoch number 25, as shown in Figure 5(c). The precision values are considered high at this point. On the other hand, the recall values stabilize after epoch 30, as shown in Figure 5(d). The recall values are high, but not as stable as the precision values.

These results suggest that the YOLOv5 model has good accuracy and can detect objects with high accuracy. However, further improvements can be made to increase the stability of the recall values and improve the performance. In this project, a system for detecting safety equipment in an industrial environment has been developed with an emphasis on comprehensive evaluation compared to previous research. The results of this work correlate with metrics obtained in previous research such as [14], [16], which focused on safety helmet detection at construction sites using CNN. While these studies obtained stable results, they also faced challenges in detection due to variability in object distances and sizes, which made accurate identification difficult. In the case of [16], it achieved an accuracy of 95%, but with a low recovery rate of 77% and a mAP of 36.82%. On the other hand, [14] did not provide detailed evaluation metrics but reported a mAP of 84%. Similarly, [18] used the YOLO model for the detection of protective equipment at construction sites, achieving a mAP of 86%. In contrast to the studies, the results of the present project are characterized by being robust and significant. On average, an accuracy of 88%, a recall of 79%, and an acceptable mAP of 80% were obtained. The results will not always coincide, and this is due to distinct reasons, such as the size of the data set, and the techniques used, among others.

Detection of safety equipment in the industrial sector presents considerable challenges. The presence of obstacles and occlusions hinders the precise identification of equipment, while variable lighting and humidity conditions can adversely affect model accuracy. To address these challenges, this study used a CNN in combination with image processing techniques to optimize model training. This approach significantly improved accuracy in equipment detection, even in the presence of obstacles and adverse identification conditions. In the case of the work [15], they proposed a technique based on deep learning with the single shot detector (SSD)-MobileNet algorithm for real-time detection of safety helmets. Although this method demonstrated the ability to detect helmets under adverse weather conditions, the accuracy metrics obtained did not reach the expected levels. This could be attributed to numerous factors, such as the selection of the dataset and the techniques used. On the other hand, the study [19] trained a deep CNN for the classification of a high-quality dataset consisting of approximately 1.3 million ImageNet images. The error rate yielded an accuracy of 18.9%, indicating significant performance. These findings are correlated with the results of the present study. It is important to highlight that the YOLOv5 model used in this study has a great capacity for this type of task, which contributed to obtaining representative accuracy metrics. Furthermore, this study may have a direct impact on the development of industrial safety solutions. Its integration into work environments could contribute to improving employee safety, resulting in increased company credibility.

4. CONCLUSION

This section presents the conclusions drawn from this study on the detection of safety equipment in industrial environments using CNN. Achieving precise and reliable detection of safety equipment in industrial settings through the CNN model YOLOv5 necessitates ample data and intensive training over multiple epochs. This study utilized 1,051 images, divided into 70% for training and 30% for validation, undergoing 50 epochs of training using the YOLOv5x model for safety equipment detection. Results reveal a helmet precision of 91%, yet a recall of 74%, indicating areas for improvement. For goggles, precision stood at 85% with a recall of 87%, demonstrating the model’s effectiveness. The absence of masks yielded a precision of 92% and a recall of 89%. This study’s significant contribution lies in successfully detecting safety equipment using the YOLOv5 model with six distinct classes, validating proper safety equipment usage. This system can identify individuals working without the required safety gear, crucial for industrial safety. In the future, this model could be integrated into real-time video surveillance systems to mitigate accident risks by alerting authorities of potential incidents. However, it’s essential to acknowledge certain study limitations, including the necessity for a robust dataset to enhance precision and the importance of high-quality images for optimal results. Additionally, powerful processing resources such as CPUs and GPUs are required for efficient system operation.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my fellow researchers, whose collaboration and support were fundamental to the success of this project. Your contributions significantly enriched our work.

REFERENCES

[1] B. H. W. Guo, Y. Zou, Y. Fang, Y. M. Goh, and P. X. W. Zou, “Computer vision technologies for safety science and management in construction: a critical review and future research directions,” *Safety Science*, vol. 135, p. 105130, Mar. 2021, doi: 10.1016/j.ssci.2020.105130.

[2] R. K. Chandana and A. C. Ramachandra, "Real time object detection system with YOLO and CNN models: a review," *arxiv*, 2022, [Online]. Available: http://arxiv.org/abs/2208.00773.

[3] CFOI, "Censo de muertes por lesiones ocupacionales," 2019. https://estadisticas.pr/files/Inventario/publicaciones/DTRH_CensodeMuertesporLesionesOcupacionales_2019.pdf (accessed Jul. 03, 2023).

[4] M. Alruwaili, M. H. Siddiqi, M. N. Atta, and M. Arif, "Deep learning and ubiquitous systems for disabled people detection using YOLO models," *Computers in Human Behavior*, vol. 154, p. 108150, May 2024, doi: 10.1016/j.chb.2024.108150.

[5] G. Iannizzotto, L. L. Bello, and G. Patti, "Personal protection equipment detection system for embedded devices based on DNN and fuzzy logic," *Expert Systems with Applications*, vol. 184, p. 115447, Dec. 2021, doi: 10.1016/j.eswa.2021.115447.

[6] Gohari, R. Jafari, L.a Aliahmadipour, and M. K. Rafsanjani, "Deep learning-based intrusion detection systems: a comprehensive survey of four main fields of cyber security," *Jmmrc.Uk.Ac.Ir*, 2022, doi: 10.22103/jmmr.2022.19961.1305.

[7] M. Arashpour, V. Kamat, A. Heidarpour, M. R. Hosseini, and P. Gill, "Computer vision for anatomical analysis of equipment in civil infrastructure projects: Theorizing the development of regression-based deep neural networks," *Automation in Construction*, vol. 137, p. 104193, May 2022, doi: 10.1016/j.autcon.2022.104193.

[8] O. Iparraguirre-Villanueva, V. Guevara-Ponce, O. R. Paredes, F. Sierra-Liñan, J. Zapata-Paulini, and M. Cabanillas-Carbonell, "Convolutional neural networks with transfer learning for pneumonia detection," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, pp. 544–551, 2022, doi: 10.14569/IJACSA.2022.0130963.

[9] E. Chian, W. Fang, Y. M. Goh, and J. Tian, "Computer vision approaches for detecting missing barricades," *Automation in Construction*, vol. 131, p. 103862, Nov. 2021, doi: 10.1016/j.autcon.2021.103862.

[10] S. B. Jha and R. F. Babiceanu, "Deep CNN-based visual defect detection: Survey of current literature," *Computers in Industry*, vol. 148, p. 103911, Jun. 2023, doi: 10.1016/j.compind.2023.103911.

[11] A. Allmendinger, M. Spaeth, M. Saile, G. G. Peteinatos, and R. Gerhards, "Precision chemical weed management strategies: a review and a design of a new CNN-based modular spot sprayer," *Agronomy*, vol. 12, no. 7, p. 1620, Jul. 2022, doi: 10.3390/agronomy12071620.

[12] L. Mohammadpour, T. C. Ling, C. S. Liew, and A. Aryanfar, "A survey of CNN-based network intrusion detection," *Applied Sciences (Switzerland)*, vol. 12, no. 16, p. 8162, Aug. 2022, doi: 10.3390/app12168162.

[13] A. S. Dina, A. B. Siddique, and D. Manivannan, "A deep learning approach for intrusion detection in internet of things using focal loss function," *Internet of Things (Netherlands)*, vol. 22, p. 100699, Jul. 2023, doi: 10.1016/j.iot.2023.100699.

[14] S. Chen, W. Tang, T. Ji, H. Zhu, Y. Ouyang, and W. Wang, "Detection of safety helmet wearing based on improved faster r-CNN," in *Proceedings of the International Joint Conference on Neural Networks*, Jul. 2020, pp. 1–7, doi: 10.1109/IJCNN48605.2020.9207574.

[15] 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, doi: 10.1109/ACCESS.2021.3138407.

[16] 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," *Advances in Civil Engineering*, vol. 2020, pp. 1–10, Sep. 2020, doi: 10.1155/2020/9703560.

[17] J. Park, H. Lee, and H. Y. Kim, "Risk factor recognition for automatic safety management in construction sites using fast deep convolutional neural networks," *Applied Sciences (Switzerland)*, vol. 12, no. 2, p. 694, Jan. 2022, doi: 10.3390/app12020694.

[18] G. U. Vasanthakumar, D. K. Kumar, K. P. Reddy, S. B. Shankar, and S. K. P. Reddy, "HDRT: helmet detection system using convolutional neural network in real time," in *MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference*, Oct. 2022, pp. 1–5, doi: 10.1109/MysuruCon55714.2022.9972413.

[19] T. Kong, W. Fang, P. E. D. Love, H. Luo, S. Xu, and H. Li, "Computer vision and long short-term memory: learning to predict unsafe behaviour in construction," *Advanced Engineering Informatics*, vol. 50, p. 101400, Oct. 2021, doi: 10.1016/j.aei.2021.101400.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 2, pp. 1097–1105, 2012.

[21] K. Bitirgen and Ü. B. Filik, "A hybrid deep learning model for discrimination of physical disturbance and cyber-attack detection in smart grid," *International Journal of Critical Infrastructure Protection*, vol. 40, p. 100582, Mar. 2023, doi: 10.1016/j.ijcip.2022.100582.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[23] O. Iparraguirre-Villanueva *et al.*, "Disease identification in crop plants based on convolutional neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, pp. 519–528, 2023, doi: 10.14569/IJACSA.2023.0140360.

[24] O. Iparraguirre-Villanueva, K. Espinola-Linares, R. O. Flores Castañeda, and M. Cabanillas-Carbonell, "Application of machine learning models for early detection and accurate classification of type 2 diabetes," *Diagnostics*, vol. 13, no. 14, p. 2383, Jul. 2023, doi: 10.3390/diagnostics13142383.

[25] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 8697–8710, doi: 10.1109/CVPR.2018.00907.

[26] M. Hussain, C. Cheng, R. Xu, and M. Afzal, "CNN-fusion: an effective and lightweight phishing detection method based on multi-variant ConvNet," *Information Sciences*, vol. 631, pp. 328–345, Jun. 2023, doi: 10.1016/j.ins.2023.02.039.

[27] X. Song, S. Gao, and C. Chen, "A multispectral feature fusion network for robust pedestrian detection," *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 73–85, Feb. 2021, doi: 10.1016/j.aej.2020.05.035.

[28] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019, vol. 1, pp. 4171–4186.

[29] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-Decem, pp. 779–788, doi: 10.1109/CVPR.2016.91.

BIOGRAPHIES OF AUTHORS

Orlando Iparraguirre-Villanueva Systems Engineer with a master's degree in Information Technology Management, and a Ph.D. in Systems Engineering from the Universidad Nacional Federico Villarreal - Peru. ITIL® Certified, Specialization in Business Continuity Management (SBCM), SCRUM Certified. National and international speaker/panelist (Panama, Colombia, Ecuador, Venezuela, México). Extensive experience in undergraduate and postgraduate teaching in different universities in the country. Thesis advisor and jury in different universities. Professional experience in management positions in the field of Information Technology. Research professor with publications in Scopus and WoS-indexed journals (Q1, Q2, Q3, and Q4) of high impact. Topics of interest: Open-source software, IoT, Augmented Reality, Machine Learning, AI, CNN, text mining, virtual environments, scientific research methodology, and thesis. He can be contacted at email: oiparraguirre@ieee.org.

Josemaria Gonzales-Huaman he is a Systems Engineering graduate from the Universidad Autónoma del Perú. He is a proactive person, with a capacity for teamwork and easy adaptability in the work area. As a professional in training, my goal is to be in constant evolution and learning. Interested in being part of a technology company that allows me to achieve my aspirations and professional development and contribute with my knowledge. HeI know technological tools such as databases in SQL Server, Oracle, PHP, Python, augmented reality, and neural networks, among other technologies. He can be contacted at email: jgonzalesh@autonoma.edu.pe.

Jose Machuca-Solano he is a graduate of Systems Engineering at the Autonomous University of Peru. With knowledge in electronics, industrial safety, electricity, and programming languages such as HTML, PHP, JS, Python, and SQL. In addition to having knowledge and certification in Scrum. He is a proactive person, with teamwork skills and easy adaptability in the work area. He can be contacted at email: wmachuca@autonoma.edu.pe.

John Ruiz-Alvarado Doctor, and Master in University Pedagogy from the National University of Trujillo. Renacyt Researcher. With extensive professional experience at undergraduate and graduate levels. Thesis Advisor and Jury. Speaker on topics of Educational Accreditation, Leadership, and Coaching. Currently, he is a work as a Professor of Higher Education Writing and Scientific Research. He can be contacted at email: C20326@utp.edu.pe.