Evolving Safety Protocols: Deep Learning-Enabled Detection of Personal Protective Equipment

Mustafa Alahmid, Kishankumar Bhimani, Khushbu Saradva, Sushil Ghildiyal, Sayed Aryan Saeedi, Saraa Ali, Han Yan, and Evgeny Filippov

Abstract To give shift in safety protocols, we have employed advanced deep learning algorithms and frameworks (Shrestha and Mahmood in IEEE Access 7:53,040–53,065, 2019 [25]) to construct an innovative AI model. The designed model detects the usage of personal protective equipment (PPE) (Personal protective equipment. Geneva: World Health Organization, 2020 [18]) by workers in high-risk industries such as construction and manufacturing. We have used Google's TensorFlow object detection API (Sai and Sasikala in Object detection and count of objects in image using tensor flow object detection API, pp 542–546, 2019 [22]) to modify and train a model for dual purposes: PPE detection and face recognition. The state-of-the-art of this research is to substantially enhance safety compliance by addressing the prevalent issue of PPE non-compliance. To emphasis this, we have developed a pioneering software prototype that synergizes PPE detection with a face recognition-based clock-in system. This prototype demonstrates impressive object detection metrics with a mean average precision (mAP) of 0.9 for vests and 0.85 for helmets. Moreover, it exhibited efficient face recognition with a successful

M. Alahmid · S. A. Saeedi · H. Yan · E. Filippov
Faculty of Computer Science and Technology, Saint Petersburg Electrotechnical University "LETI", Saint Petersburg, Russian Federation

K. Bhimani (✉) · K. Saradva
Laboratory on AI for Computational Biology, Faculty of Computer Science, HSE University, Moscow, Russian Federation
e-mail: info.bhimani@gmail.com; kbhimani@hse.ru

S. Ghildiyal
Department of Computer Science and Engineering, Indian Institute of Technology Ropar, Punjab, India

S. Ali
Laboratory of Methods for Big Data Analysis, Faculty of Computer Science, HSE University, Moscow, Russian Federation

S. A. Saeedi
Deggendorf Institute of Technology, Bavaria, Germany

B. Annappa et al. (eds.), *High Performance Computing, Smart Devices and Networks*, Lecture Notes in Electrical Engineering 1262,
https://doi.org/10.1007/978-981-97-7794-5_7

threshold range of 17–20%. The implementation of AI in our system promises significant enhancements to worker safety, while concurrently reducing the financial burden associated with big hazards and accidents. Beyond the development and performance of the system, this paper provides a thorough exploration of the encountered challenges, potential real-world applications (particularly in employee monitoring and clock-in systems), and the future implications of this study on research and practical applications in the field of AI-integrated safety compliance.

Keywords Personal protective equipment (PPE) · AI · Object detection · Open CV · TensorFlow object detection API · Faster R-CNN

1 Introduction

1.1 Safety Protocols and History

Safety is a paramount concern for both day laborers and skilled workers. Every year, the construction industry is plagued by numerous accidents, many of which can be attributed to a lack of personal protective equipment (PPE) or by intensely avoiding to utilize the provided PPE. PPE is an equipment designed to shield employees from health and safety hazards at work. It is deployed to mitigate employee exposure to risks when administrative and engineering interventions are not feasible or effective in reducing these hazards to acceptable levels. Such risks may include falling debris, slippery surfaces, and various industrial hazards. PPE typically includes protective vests, helmets, and safety glasses.

Construction workers have a high rate of injury; there are roughly 340 million accident injuries per year, according to the International Labour Organization Statistics [11]. It is crucial to ensure PPE compliance. Therefore, we propose the design of deep learning model capable of detecting whether workers are equipped with the necessary safety gears and PPE.

1.2 Personal Protective Equipment (PPE) Challenges

Significant compliance issues with the use of personal protective equipment (PPE) have been revealed by a survey [34] of one hundred professionals in high-risk industries such as construction, which is the problem of interest in our research, the employees non-compliance issue was one of the main issues that 38% of the professionals have mentioned the problem according to the survey, and it is on the top of the PPE concerning challenges.

The report by the artificial intelligence company Cortexica [30]said that the vast majority (84%) of businesses that have work in a high-risk environment incurred losses from injuries due to the PPE non-compliance in the past year. The report also claims that 84% of them still rely on manually checking for employees for PPE non-compliance.

1.3 History and Evolution of Object Detection

Object detection, a crucial component of machine learning and deep learning, has transformed drastically over the years. Presently, deep learning-based approaches lead the field, thanks to the robustness and accuracy they provide. The rise of deep learning in object detection [32] coincides with advancements in neural network architectures [23], especially convolutional neural networks (CNNs) [15]. Deep learning algorithms, typically defined as layered models of inputs, aim to emulate the human understanding of the world. This allows them to perform tasks like image and video recognition, image classification, natural language processing, and many more.

Nowadays, the approaches of deep learning are the state-of-arts [17] in this field. CNN is the cornerstone of most deep learning models today. Its fundamental components include convolution layers and pooling operations, both contributing to its high efficacy in image and video processing.

Today's object detection systems boast accuracy, speed, and efficiently classifying objects. This precision is achieved through extensions of image classification models and the introduction of new APIs, like Google's TensorFlow. API includes pre-built architectures and weights for models like SSD with Inception V2 [1], Single Shot Multibox Detector [13], faster RCNN with Inception ResNet V2 [3], faster RCNN with ResNet 101, and region-based fully convolutional networks (R-FCN) [2].

1.4 Face Detection and Recognition

Face recognition, a computer vision problem consisting of face detection and face identification [33], is dominant in modern user authentication mechanisms. As an example, tech giant Baidu leverages face recognition instead of traditional ID cards for employee office access [14].

The 1990s saw a significant leap with Eigenfaces, implemented by Sirovich and Kirby [26], which developed a low-dimensional representation of facial images. This approach established that less than a hundred values were needed to accurately code a normalized image of a face.

More recently, AlexNet [12], using a deep convolutional neural network (CNN), achieved an impressive error rate of 15.3%, surpassing other algorithms [9] of its time by over 10% in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [21]. The technology further improved, reaching an error rate of just 3.6% by 2015 with ResNet [6].

FaceNet [24], a system for face recognition, verification, and clustering, utilizes a Euclidean embedding for image representation through a deep convolution network. The architecture facilitates the matching of faces based on minimized distances for the same person and maximized distances for different individuals. Its structure consists of a batch input layer, a convolutional neural network, L2 normalization, and a triplet loss function. After creating the embeddings represented by a 128-feature vector, the recognition task becomes a KNN classification problem, which could be solved using a SVM classifier.

2 Holistic Methodology for PPE Identification and Face Recognition

2.1 Role and Importance of PPE Monitoring in Workplace Safety

In an effort to mitigate workplace accidents, particularly in the construction industry, constant monitoring of PPE usage among employees is vital. However, manual monitoring presents issues such as cost, time, and human error. This paper proposes an AI-based model to detect PPE usage among workers, thus ensuring continuous safety compliance.

The use of PPE such as helmets, vests, gloves, and safety footwear (see Fig. 1) plays a critical role in protecting employees from various on-site hazards. Employers have an obligation to ensure the provision and correct usage of PPE. Monitoring is a primary duty within many industries, where workers are mandated to wear PPE.

2.2 Detection Model: Faster R-CNN, SSD, and Inception

Transfer Learning: Transfer learning [31] is a technique where a pre-trained model for a specific task is leveraged as an initiation for another related task. In the context of deep learning and computer vision, transfer learning allows reuse of learned features from one model to benefit a new model trained on a different task. Two common approaches to transfer learning are feature extraction and weight initialization.

TensorFlow Object Detection API: The TensorFlow object detection API [29] offers a comprehensive framework for building, training, and deploying object detection models. This research utilizes the TensorFlow object detection API to develop a PPE detection model.

Faster R-CNN: The faster R-CNN model [20] comprises two interconnected modules: the region proposal network (RPN) and the fast R-CNN detector. The

Fig. 1 Personal protective equipment spectrum: helmets to accessories (essential gloves, cones, safety glasses, protective boots, and vests)

RPN predicts object presence and generates bounding boxes, while the fast R-CNN uses these characteristics to classify objects and refine their bounding boxes. This unified architecture enables efficient and accurate object detection.

Inception Model: The Inception model [28] can tackle the challenge of handling objects with varying sizes in an image. By employing filters of different sizes at the same network level, the model can capture both global and local information. Inception models, including the widely-known GoogleNet [27], demonstrate improved computational efficiency through smart factorization methods.

Single Shot Detector (SSD): The SSD model represents a single deep neural network capable of performing object detection and classification in a single feed-forward pass. By producing fixed-size bounding boxes for objects and associated scores, the model streamlines the detection process. Suppression techniques are then applied to select high-scoring detections above a threshold.

These methodologies form the basis for practical implementation of PPE detection.

2.3 Dataset

The dataset consisted of images collected from the Internet source using three different methods including Google Search PPE Images, Google Open Dataset Images [16], and video frames extracted from construction documentaries using Python script. In total the dataset contains 8351 objects distributed as shown in the (see Fig. 2).

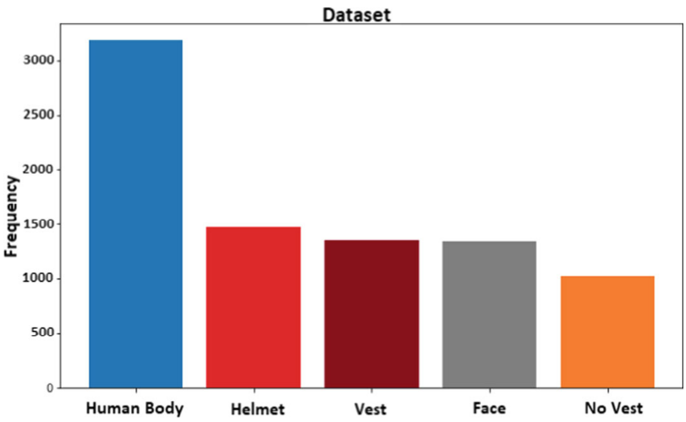
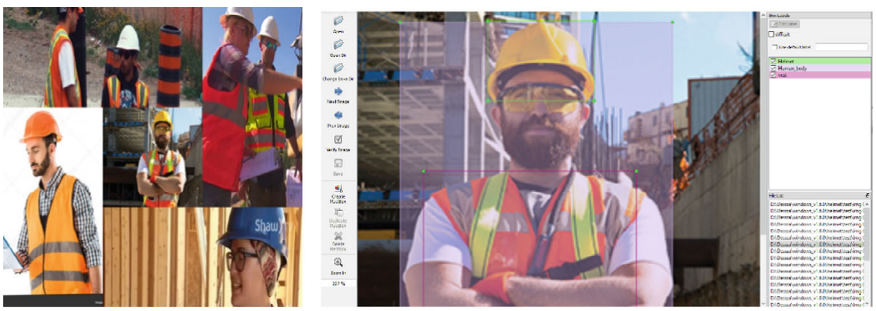


Fig. 2 Dataset statistics—with 8351 objects—helmets, vests, and construction workers faces from Google Images, Google Open Dataset, and video frame extraction

Labeling the Dataset: In order to feed the model with data, the data should be labeled, and we use LabelImg (from Label Studio) [7] tool for labeling images. The final dataset was annotated with around 1300 images. The figure (see Fig. 3a) displays random examples of images from the dataset and their corresponding annotations achieved through the LabelImg tool in (see Fig. 3b).

Computational Environment: Python 3.10 or above version with the following libraries latest version: TensorFlow, Protosun, Pillow, NumPy, lamp, OpenCV, and Pandas.



(a) Dataset Images **(b)** Dataset Labeled Images

Fig. 3 **a** Shows random images examples from the dataset, **b** labeled using the LabelImg tool [7], resulting in around 1,300 annotated images

3 Experiment

3.1 TensorFlow Object Detection API

Experiment 1: Worker Classification in Overlapping Body Images

- The dataset contained images of workers with and without helmets and vests.
- The objective was to detect human bodies and classify them as workers.
- Challenged by significant overlap between worker and non-worker bodies, leading to difficulty in accurate classification.

Experiment 2: Model Performance on Synthetic Dataset

- The model trained on a dataset of 500 images from Google (vest and helmets).
- Achieved accurate detection of helmets, but low accuracy in predicting vests.
- The dataset lacked real-life images, resulting in reduced accuracy when applied to real-life videos or images.

Experiment 3: Model Training and Optimization

- Trained faster R-CNN Inception v2 and SSD Inception v2 models on an image dataset with five classes: vests, no vests, helmets, persons, and faces.
- Both models required approximately 50,000 training steps to achieve the lowest total loss.

3.2 Evaluation

Mean Average Precision (mAP): [8] is a widely used evaluation metric for object detection and segmentation models. It assesses the accuracy of predicting both the class and location of objects in images. The ground truth includes class labels and precise bounding box coordinates for each object. During training and validation, images are annotated similarly. The model generates multiple predictions, but only those above a confidence threshold are considered. The resulting detections are then evaluated using Intersection over Union (IoU) to measure the correctness of bounding boxes. mAP provides a performance measure for accurately predicting object classes and locations in images (see Fig. 4).

Interpolated AP: PASCAL VOC [4] is a popular object detection dataset. In PASCAL VOC, a prediction is considered positive if IoU ≥ 0.5. Multiple detections of the same object are treated as negatives, except for the first one. The average precision (AP) in Pascal VOC2008 [10] is calculated using 11-point interpolation, where precision is computed for 11 equally spaced recall values from 0 to 1.

Table 1 gives a summary of models performance with mAP score.

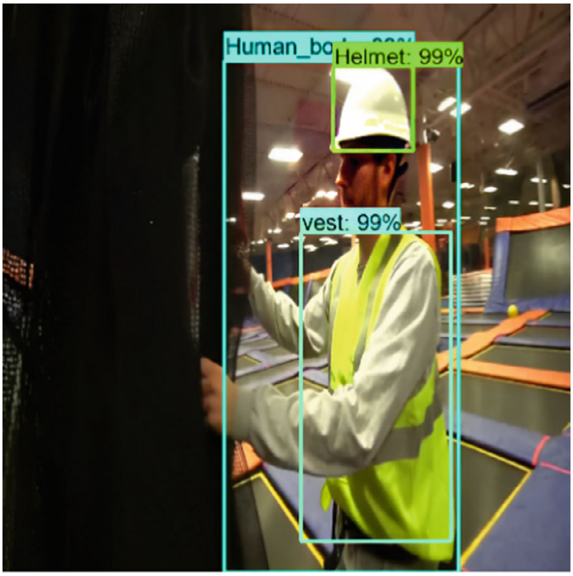


Fig. 4 mAP: evaluating object detection and segmentation with boundary box coordinates

Table 1 mAP score with various models

Model	Vest	No vest	Face	Human body	Helmet	mAP
No. of images	1347	1019	1338	3179	1468	–
SSD model	0.9056	0.43	0.66	0.59	0.721	0.661
Faster R-CNN	0.96	0.69	0.78	0.72	0.85	0.804

3.3 Face Detection and Recognition

Here, we focus on developing a face detection and recognition system, comprising these two main components. Faster R-CNN [20], a highly effective face detection model, is utilized to identify faces as objects of interest. Additionally, FaceNet [24] is employed to extract feature embeddings for each face image, enabling accurate face recognition.

Face Detection: Faster R-CNN [20], originally designed for PPE detection, demonstrates exceptional performance in detecting faces. By incorporating face detection into our system, we leverage the capabilities of this model.

Face Recognition: The face recognition process consists of two steps: creating embeddings (feature vectors) and predicting the face ID using SVM or KNN algorithms. FaceNet, a face recognition system, generates embeddings that map face images to a Euclidean space where distances reflect face similarity. These embeddings enable accurate face identification and can be further classified using an SVM algorithm.

Data Preprocessing: The face dataset utilized for training includes images of 19 different celebrity classes collected from Google Images. The faces are cropped using the faster R-CNN model, and the resulting cropped images are stored for subsequent processing. FaceNet is then employed to create 128-dimensional feature embeddings for each face image. The dataset is divided into training and testing sets.

Model Selection: This paper presents two powerful PPE detection models. Figure (see Fig. 5) shows the SSD model's accuracy in face detection (a) and PPE detection on data images without and with vests (b), respectively. Similarly, (see Fig. 6) demonstrates the faster R-CNN model's proficiency in face detection (c) and PPE detection without and with vests (d). Notably, the faster R-CNN model excels in both PPE (vest) detection and face recognition.

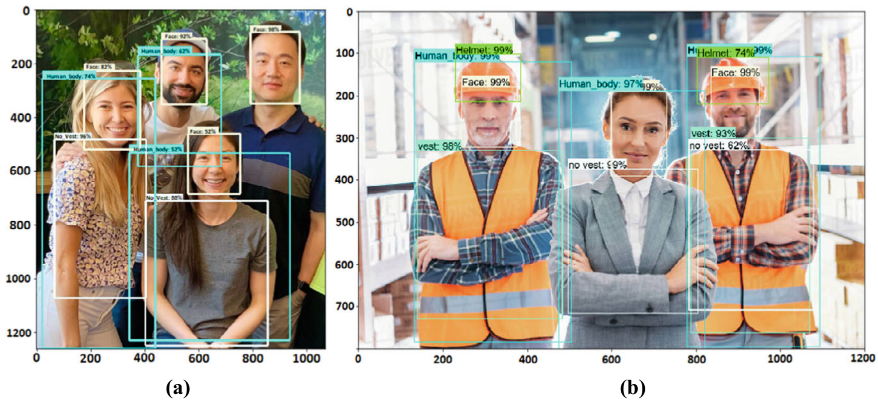


Fig. 5 PPE detection using SSD model

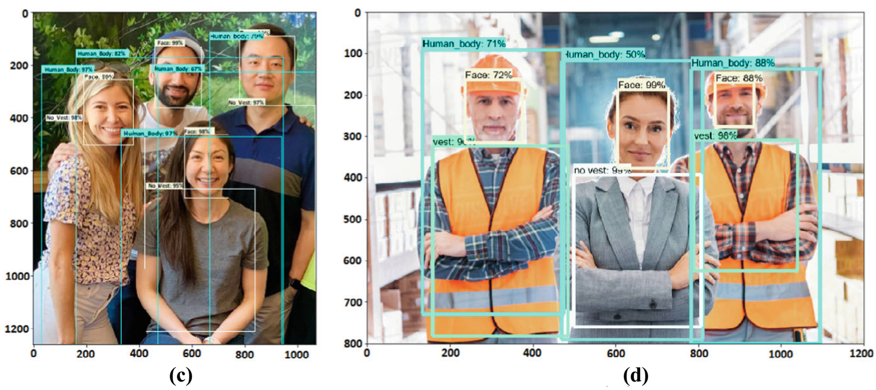


Fig. 6 PPE detection with the faster R-CNN model

Face Recognition system Evaluation: The evaluation process involves calculating the false acceptance rate (FAR) and the false rejection rate (FRR) using a test set of 200 face images, including 100 imposter pictures (50%). The evaluation is performed by varying the threshold (predicting probability) from 1 to 100%. The obtained results are used to generate a FAR-FRR curve, enabling the determination of the equal error rate (EER). In our experiments, the EER is nearly zero, indicating optimal performance at a threshold of 16%. The confusion matrix [19] illustrates the correct and misclassified face instances.

The faster R-CNN model demonstrates high accuracy and speed compared to other models, making it suitable for our requirements. The face recognition module achieves good accuracy within the threshold range of 17–20%.

4 Result: Monitoring and Clock-in System

PPE is necessary at construction sites, and many countries mandate its use. AI vision can assist in monitoring workers and implementing efficient systems for PPE compliance.

4.1 Monitoring and Clock-in System

A PPE detector can aid in identifying and monitoring PPE usage. A clock-in system can be implemented using the PPE detector and biometric technologies such as fingerprint or face recognition (see Fig. 7a).This system ensures that workers can only sign in (see Fig. 7b) when wearing the required PPE, enforcing compliance.

4.2 Application Objective

The objective is to create an accurate attendance record and enable precise payroll calculations. The clock-in system combines PPE detection with face recognition to verify worker identity and enforce PPE compliance effectively. By integrating AI and biometrics, the application enhances safety measures and enforces the PPE mandate.

1. **PPE Detection**: The system utilizes a camera to capture images and employs a detector to identify safety equipment such as vests, helmets, and faces. By analyzing the image, the system ensures that workers are wearing the required PPE.

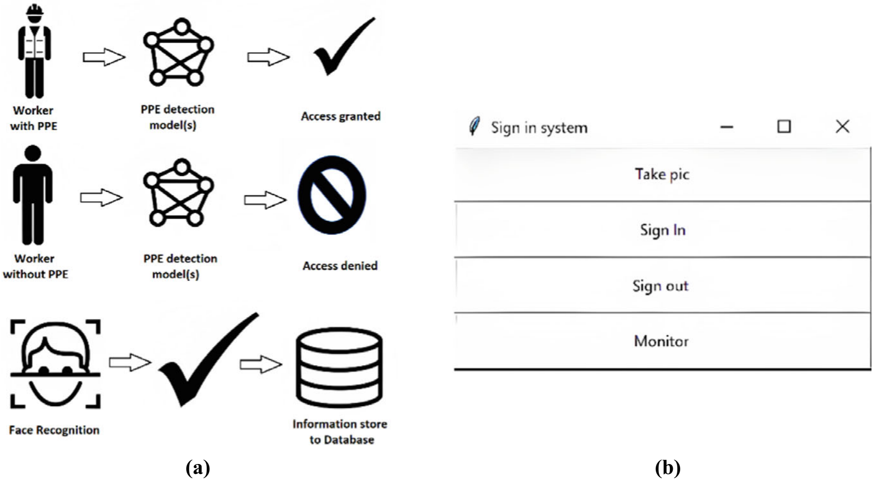


Fig. 7 Application pipeline and functioning (**a**) access protocol, (**b**) clock-in system

2. **Face Recognition and ID**: The application extracts the face from the image and utilizes facial recognition technology to predict the worker's ID [5]. It records the ID along with the time of clock-in and clock-out, creating an accurate attendance record.
3. **Field Monitoring**: The system uses a camera and a PPE detection model to monitor the field (see Fig. 8). By detecting PPE, it assists in monitoring worker compliance and ensuring safety protocols are followed.

Fig. 8 Real-time field monitoring and PPE detection on workers

5 Conclusion

Construction work is a hazardous activity that often requires the use of personal protective equipment (PPE). However, manual monitoring of PPE compliance is challenging and not cost-efficient. In this study, we developed a system that utilizes AI to recognize and identify PPE, determining whether workers are wearing it or not.

Our research demonstrates a promising approach for developing a PPE detection system and highlights the challenges involved, as well as the potential real-world applications. By employing the faster R-CNN model, we achieved satisfactory results in detecting vests and helmets, with a mean average precision (mAP) of 0.9 for vests, 0.85 for helmets, and an average mAP of 0.8 for all detected objects.

Furthermore, we developed a face recognition-based clock-in system that not only detects PPE but also verifies worker identities and records their login times in a database. The proposed system consists of two parts: PPE detection, where we utilized the faster R-CNN to train the model to recognize both faces and PPE, achieving a respectable mAP of 0.78; and face recognition, which identifies worker IDs and stores them along with login timestamps.

6 The Future of AI in PPE

The future of AI in PPE holds great promise, particularly in addressing challenges related to worker density and camera or surveillance distance from workers. By leveraging advancements in AI technology, especially deploying liquid neural network, Detectron2, and InsightFace in detection applications significantly improves result.

References

1. Bose SR, Kumar VS (2020) Efficient inception V2 based deep convolutional neural network for real-time hand action recognition. IET Image Proc 14(4):688–696
2. Dai J et al (2016) R-FCN: object detection via region-based fully convolutional networks. In: NIPS'16. Curran Associates Inc., Red Hook, NY, USA, pp 379–387
3. Devi SK, Subalalitha CN (2022) Assisting the visually challenged people using faster RCNN with inception ResNet V2 based object detection model. In: Ranganathan G et al (eds) Pervasive computing and social networking, pp 171–181. https://doi.org/10.1007/978-981-16-5640-8_14
4. Everingham M, VanGool L, Williams CKI, Winn J, Zisserman A (2008) The PASCAL visual object classes challenge 2008 (VOC2008) results. http://www.pascal-network.org/challenges/VOC/voc2008/workshop/index.html
5. Hassan N, Ramli DA, Suandi SA (2014) Fusion of face and fingerprint for robust personal verification system. Int J Mach Learn Comput 4:371–375. https://doi.org/10.7763/ijmlc.2014.v4.439

6. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, Las Vegas, NV, USA, pp 770–778. https://doi.org/10.1109/CVPR.2016.90
7. Heartexlabs: Tzutalin: LabelImg. https://github.com/heartexlabs/labelImg
8. Henderson P, Ferrari V (2017) End-to-end training of object class detectors for mean average precision. In: Lai S-H et al (eds) Computer vision—ACCV, pp 198–213. https://doi.org/10.1007/978-3-319-54193-8_13
9. Hochreutiner C (2020) The history of facial recognition technologies: how image recognition got so advanced. https://anyconnect.com/blog/the-history-of-facial-recognition-technologies
10. Hoiem D, Divvala SK, Hays JH (2009) Pascal VOC 2008 challenge. World Literature Today Citeseer 24:1
11. International Labour Organization (ILO). World statistic. occupational safety and health (OSH). https://www.ilo.org/moscow/areas-of-work/occupational-safety-and-health/WCMS_249278/lang--en/index.htm
12. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60(6):84–90
13. Liu W, Anguelov D, Erhan D, Szegedy C, Reed SM, Fu C-Y, Berg AC (2016) SSD: single shot multibox detector. In: Leibe B et al (eds) Computer vision—ECCV, pp 21–37. https://doi.org/10.1007/978-3-319-46448-0_2
14. Marr B (2018) How Chinese internet giant Baidu uses artificial intelligence and machine learning. https://www.forbes.com/sites/bernardmarr/2018/07/06/how-chinese-internet-giant-baidu-uses-artificial-intelligence-and-machine-learning/?sh=6b66fdc2d557
15. O'Shea K, Nash R (2015) An introduction to convolutional neural networks
16. Open Images Dataset V7 and Extensions. https://storage.googleapis.com/openimages/web/index.html
17. Pal SK, Pramanik A, Maiti J, Mitra P (2021) Deep learning in multi-object detection and tracking: state of the art. ApplIntell. 51(9):6400–6429
18. Personal protective equipment. Geneva: World Health Organization; Licence: CC BY-NC-SA 3.0 IGO (2020)
19. Ren J, Wang Y (2022) Overview of object detection algorithms using convolutional neural networks. J Comput Commun 10(1):115–132
20. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: NIPS'15, vol 1. MIT Press, pp 91–99. https://doi.org/10.1109/TPAMI.2016.2577031
21. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein MS, Berg AC, Fei-Fei L (2015) ImageNet large scale visual recognition challenge. Int J Comput Vision 115(3):211–252
22. Sai BNK, Sasikala T (2019) Object detection and count of objects in image using tensor flow object detection API. In: 2019 ICSSIT, Tirunelveli, India, pp 542–546. https://doi.org/10.1109/ICSSIT46314.2019.8987942
23. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003
24. Schroff F, Kalenichenko D, Philbin J (2015) FaceNet: a unified embedding for face recognition and clustering. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, Boston, MA, USA, pp 815–823
25. Shrestha A, Mahmood A (2019) Review of deep learning algorithms and architectures. IEEE Access 7:53040–53065. https://doi.org/10.1109/ACCESS.2019.2912200
26. Sirovich L, Kirby M (1987) Low-dimensional procedure for the characterization of human faces. J Opt Soc Am A 4(3):519. https://doi.org/10.1364/JOSAA.4.000519
27. Szegedy C, Liu W, Jia Y, Sermanet P, Reed SM, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), Boston, USA, pp 1–9
28. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, Las Vegas, NV, USA, pp 2818–2826

29. TensorFlow Object Detection API. https://github.com/tensorflow/models
30. Tomorrow's Health and Safety Team, Cortexica (2018) Overcoming the barriers of PPE compliance with the use of artificial intelligence. https://www.tomorrowshs.com/improper-ppe-cost-businesses-seventy-nine-billion--1038.html
31. Torrey L, Shavlik J (2010) Transfer learning. IGI Global, pp 242–264
32. Xiao Y, Tian Z, Yu J, Zhang Y, Liu S, Du S, Lan X (2020) A review of object detection based on deep learning. Multimed Tools Appl 79(33–34):23729–23791
33. Zhafran F, Ningrum ES, Tamara MN, Kusumawati E (2019) Computer vision system based for personal protective equipment detection, by using convolutional neural network. In: 2019 international electronics symposium (IES), Indonesia, pp 516–521. https://doi.org/10.1109/ELECSYM.2019.8901664
34. Zou Z, Chen K, Shi Z, Guo Y, Guo Y (2023) Object detection in 20 years: a survey. Proc IEEE 111(3):257–276. https://doi.org/10.1109/JPROC.2023.3238524