Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

Personal Protective Equipment Monitoring in Construction Site Using Deep Neural Network

Mrs. G. Sharmila[1], Mrs. N. Pavithra[2], Mrs. D. Ramya Cauvery[3], Mrs. Y. Suganya[4]

[1]Assistant Professor, Department of Computer Science and Engineering, Mookambigai College of Engineering, Pudukkottai, TamilNadu, India. **EMail:** sharmisang@gmail.com

[2]Assistant Professor, Department of Computer Science and Engineering, Mookambigai College of Engineering, Pudukkottai, TamilNadu, India. **EMail:** pavithraneelu@gmail.com

[3]Assistant Professor, Department of Computer Science and Engineering, Mookambigai College of Engineering, Pudukkottai, TamilNadu, India. **EMail:** ramyac69@gmail.com

[4]Assistant Professor, Department of Computer Science and Engineering, Mookambigai College of Engineering, Pudukkottai, TamilNadu, India. **EMail:** suganyasuchithrra@gmail.com

Abstract

Safety has always been of paramount importance in all industrial endeavors, particularly construction. It is not an ordinary office position, and precautions are necessary. Accidents and injuries are less likely to occur on a construction site when workers are well-equipped with safety gear. The cranium is the only organ entirely enclosed in bone in the human body. The significance of safeguarding the brain, a component of the body with a vital role in bodily function, is a natural law. Hard headwear and safety headgear serve as the first line of defense against head injuries, but only when worn correctly. Thus, it is reasonable to assert that safety Helmets reduce the risk of brain injury and save lives. This paper presents a system for real-time detection based on video streaming analysis and Deep Neural Network (DNN). A new method in convolutional neural network predicts whether or not employees are donning helmets correctly. The paradigm of edge computing in which the application for image analysis and classification is deployed on an embedded system directly connected to the camera. The proposed system is developed using a low-cost commercial embedded system, namely a Raspberry PI with an Intel Neural Compute Stick. The system was tested with various convolutional neural networks (CNNs) that had been pre-trained and was optimized to monitor the worker's headgear. In terms of classification performance and inference latency, CNNs were contrasted. Then, each CNN was deployed on the actual system and the system's throughput was measured by analyzing video frames per second.

Keywords: - Safety Helmets, Deep Neural Network, Convolutional Neural Network, Edge Computing Model

I. Introduction

Construction is a high-risk industry in which construction laborers frequently sustain injuries. Traumatic brain injuries are often fatal. According to accident statistics released by the state administration of work safety between 2015 and 2018, 67.95% of the 78 construction accidents that occurred between 2015 and 2018 were the result of

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

workers not wearing safety helmets. [1] It is essential for construction site safety management to monitor the condition of the personal protective equipment worn by construction employees. Helmets can absorb and disperse the impact of falling objects and mitigate the injuries sustained by employees who fall from heights. Due to a lack of safety awareness, construction workers tend to disregard protective headgear. On the construction site, employees who inappropriately don safety helmets are significantly more likely to sustain an injury. Typically, traditional supervision of employees donning safety helmets on construction sites requires physical labor. There are issues such as a wide variety of operations and challenging administration of site personnel. These factors make manual supervision difficult and inefficient, and it is challenging to accurately monitor and manage all construction site laborers in real time. Therefore, it is difficult to meet the modern requirements of construction safety management using only manual supervision. In this context, studying the automatic detection and recognition of safety helmet donning conditions remains an important issue.

[2] In deep learning, a convolutional neural network (CNN or ConvNet) is a type of Artificial Neural Network (ANN) that is most frequently used to analyze visual imagery. Feature maps are the shared weight architecture of convolution kernels or filters that glide along input features and produce translation-invariant responses. Application areas include image and video recognition, recommender systems, image classification, image segmentation, medical image analysis, natural language processing, brain-computer interfaces, and financial time series. CNNs are regularized multilayer perceptron models. Typically, multilayer perceptrons denote completely connected networks, in which each neuron in one layer is linked to all neurons in the following layer. These networks' "full connectivity" makes them susceptible to data over-fitting. Normal methods of regularization or averting over-fitting include penalizing parameters during training (such as weight decay) or reducing connectivity (skipped connections, dropouts, etc.). Therefore, CNNs are at the bottom of the connectivity and complexity scale. Different neuronal receptive fields partially overlap so as to encompass the entire visual field. Figure 1.1 demonstrates that CNNs require less preprocessing than other image classification algorithms. This means that the network learns to optimize the filters (or kernels) through automated learning, whereas these filters are hand-engineered in conventional algorithms. This independence from prior knowledge and human intervention is a primary benefit of feature extraction.

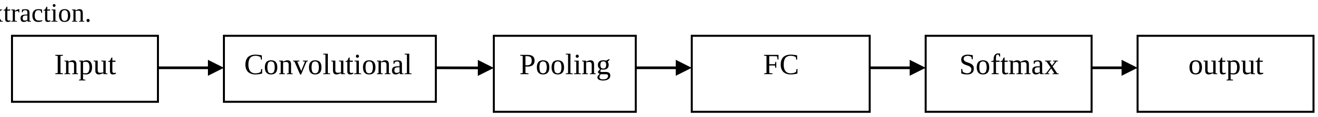


Fig.1.1 Layers in CNN

II. Related Work

The method used for detecting multiple objects in an image or video source is provided as input for the purpose of ensuring the safety of construction laborers.

[3] Baining Zhao and Haijuan Lan introduce the paper Detection and Location of Safety Protective Clothing in Power Substation Operations: In this work, gamma correction is used as a preprocessing technique to emphasize the operator's detail before data augmentation is performed. k means++ algorithm replaces k mean in wear-enhanced YOLOv3 to determine the optimal prior bounding box size and increase the speed of detection.On the basis of transfer learning, the proposed method can be rapidly trained, the mean average precision is improved by 2%, and frame per second is increased compared to conventional object detection methods. WE YOLO v3 not only outperforms the prevalent image detection method in terms of detection precision, but is also 50% quicker. Consideration is given to extreme conditions including backlight, complex background, and fragmentary object information.

---

[4] Ning li and Xin Lyu present the article Incorporate Online Hard Example Mining and Multi-Part Combination Into Automatic Safety Helmet Wearing Detection 2020. In this paper, the multi scale training and increasing anchors strategies are used to improve the robustness of the original RCNN algorithm for detecting objects of various scales and sizes.The objective of the OHEM is to optimize the model to avoid an imbalance between positive and negative samples. The quicker RCNN detects the individual donning the helmet and the correct elements, the better. The precision has enhanced by 7%.It also improves detection performance for partial occlusion and objects of various sizes, demonstrating excellent generalization. The pyramid method is used to acquire the multi-scale features of the image ,and the OHEM mechanism in introduced to select hard samples, and then transmit them to the network retraining so that the network can learn the hard samples. The method approximates the relationship between the safety headwear and other components.

[5] A Smart System for Personal Protective Equipment Detection in Industrial Environments is introduced by Gionatan Gallo and Francesco Di Rienzo. Based on Deep Learning at the Edge 2022, this work develops a system for real-time PPE detection in streaming videos using a deep neural network (DNN).The edge computing paradigm in which the image analysis and classification application is deployed on the embedded system installed in the proxy camera and is directly connected to it. With an Intel Neural Compute Stick, the proposed system creates low-cost commercial embedded systems comparable to Raspberry PI. To assess the system throughput in terms of the number of video frames that are analyzed per second. Based on the edge computing paradigm, the system analyzes images from a hazardous location in real time on an embedded system to determine whether or not employees are wearing protective equipment, such as a helmet, vest, and gloves. The YOLOv4 model has a lower classification accuracy of approximately 10.4%.

[6]In their 2001 paper Rapid Object Detection Using a Boosted Cascade of Simple Features, Paul Viola et al. introduced an object detection method based on a boosted cascade and a simple time frame. The approach used a sliding window using which they went through all the locations possible in an image and in the end it kept on verifying if any window in any scale consisted of the human visage. The calculations in this procedure were too complex for the computing capabilities of the time to handle. The author employs three methods: integral image to accelerate box filtering, features selection to select significant features, and detection. Multi-Stage Detection is an approach that uses a cascade to reduce the computational capacity of a concept.

III. System Design

The objective of a safety helmet detection initiative in industries is to ensure the correct use of personal protective equipment (PPE) by employees and to improve workplace safety. The initiative intends to autonomously detect and alert workers who are not wearing safety helmets. This can be done using computer vision and machine learning techniques to recognize the presence of a helmet.

Fig.3.1 USB Camera

A typical low-cost webcam is a video capture device that is connected to a computer or computer network, typically through a USB interface or, if connected to a network, Ethernet or Wi-Fi, or built-in for certain laptop models.

Features

As webcam capabilities have been introduced to instant messaging, text chat services such as AOL Instant Messenger, millions of mainstream PC users around the world now have access to live video communication over the Internet. As a result of enhanced video quality, webcams have begun to compete with traditional video conferencing systems. New features such as automatic illumination controls, real-time enhancements (retouching, wrinkle flattening, and vertical stretch), automatic face tracking, and autofocus make webcams significantly more user-friendly, thereby increasing their popularity.

Video security

Webcams are also employed as surveillance cameras. Software is available that enables PC-connected cameras to monitor for motion and sound, recording both when they are detected; these recordings can then be preserved, emailed, or uploaded to the Internet. In one widely reported case, a computer emailed images as the thief took it, allowing the owner to provide police with a distinct image of the thief's face even after the computer was stolen.

Input Control Device

Using the video broadcast from a webcam, specialized software can facilitate or enhance a user's control of applications and games. Faces, contours, models, and colors, among other video characteristics, can be observed and traced to generate a corresponding form of control. A head-mounted light would enable hands-free computing and significantly enhance computer accessibility. This can also be applied to video games to provide enhanced control, interactivity, and immersion.

Face Recognition

Face Identification is a type of computer vision that attempts to identify or verify the asserted identity of a person based on their visage. This method is used to identify a person's visage, whether or not they are donning a helmet.

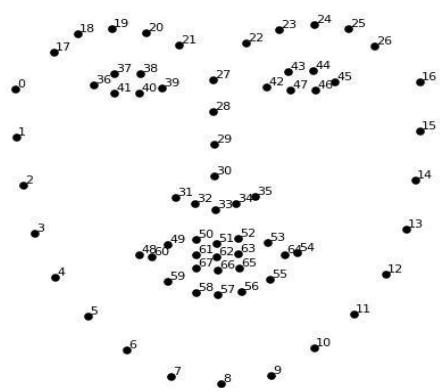


Fig.3.2.1 DLib Pre-Trained Model

DLib

Dlib is a contemporary Python toolkit comprising machine learning algorithms and tools for developing complex Python software to solve real-world problems. It is utilized by both industry and academia in a variety of domains, including robotics, embedded devices, mobile phones, and large environments for high performance computation. Dlib's open source licensing permits its use in any application without cost. Figure 3.2.1

IV. Proposed Work

In this study, the CNN method of deep learning is used to predict whether a helmet will be worn or not. As depicted in the proposed diagram, a Convolutional Neural Network is a type of deep learning, feed-forward artificial neural networks that is typically beneficial for multiple analyses. Figure.4.1

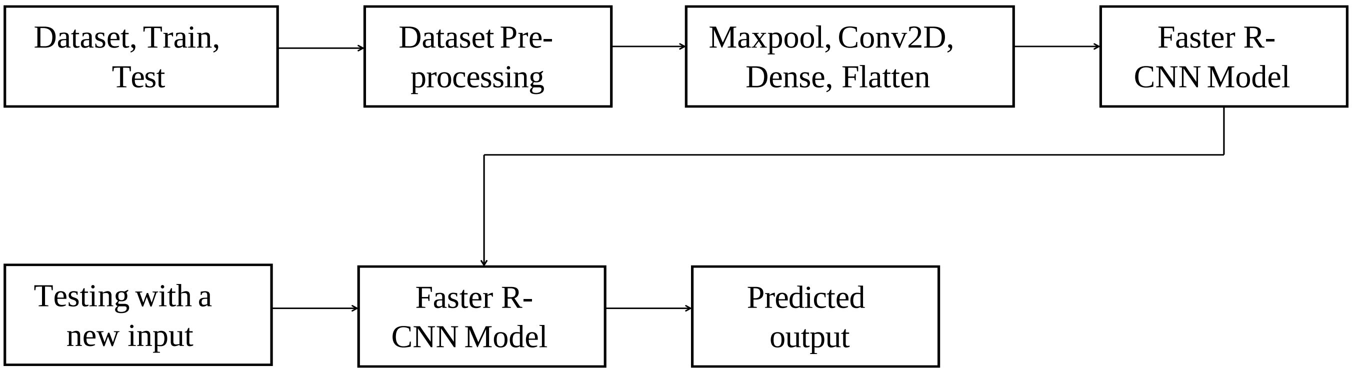


Fig.4.1 Proposed Block Diagram

A convolutional neural network (CNN) is a neural network with multiple layers. It is a technique for image recognition and classification based on deep learning. It can solve the problems of deep neural networks having too many parameters and being challenging to train, and it can improve classification effects. Most CNNs consist of an input layer, a convolutional layer, an activation function, a pooling layer, and a fully connected layer. CNNs are distinguished by their local connectivity and parameter sharing, which reduces the number of parameters and improves detection efficacy.

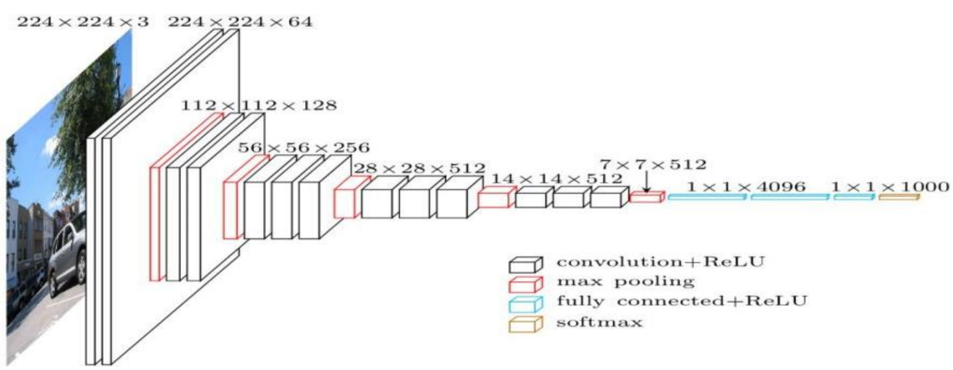


Fig.4.2 Extractor Layer

As shown in Figure No.4.2, the architecture consists of six layers: 1 Input Layer, 2 Pairs of Convolutional Layers, Max Pooling Layers, and 1 Output Layer.

Input Layer: This layer contains 786 x 786 neurons, which corresponds to the number of pixels in each image being transmitted. Here, the input layer receives the pixel values of the training images.

Convolutional Layer 1: This layer contains 32 neuronal cells. There is a connection between every neuron in this layer and every neuron in the layer beneath it. Convolution is performed on the input pixels, which is a process of performing dot product on the pixel values with arbitrary integers called as filters.

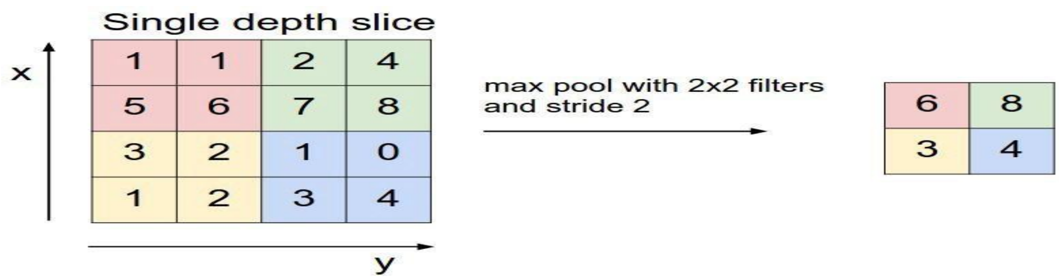
Max Pooling Layer 1: This layer is depicted in Figure 4.3, with the provided filters; a max pooling operation is performed on the received input, which identifies the highest value in each feature map region.

Convolution Layer 2: The output of the max pooling layer is concatenated to a convolution layer (convolution layer 2) with 16 filters, a kernel size of 4*4, and ReLu activation. This layer is transmitted to a maximum pooling layer next.

Max Pooling Layer 2: On the received input, the max pooling layer executes the max pooling operation. The output of the max pooling layer is then subsequently compressed. The procedure of flattening converts any matrix into a one-dimensional array.

Output Layer: The total number of neurons in this layer is equal to the number of disease classification levels. The neuron consisting of the maximal value ranging between 0-1 will be the output. This operation is carried out on every single training image.

Fig.4.3 Pooling Layer



Dataset

As depicted in Figure 4.2 below, a massive data set comprised of retina images with extremely high resolution has been collected from Kaggle under various imaging conditions. Then, a dataset containing 3,261 images of various helmets are constructed and divided into three portions for training and testing the model. Tensor Flow is selected as the training framework for the model.

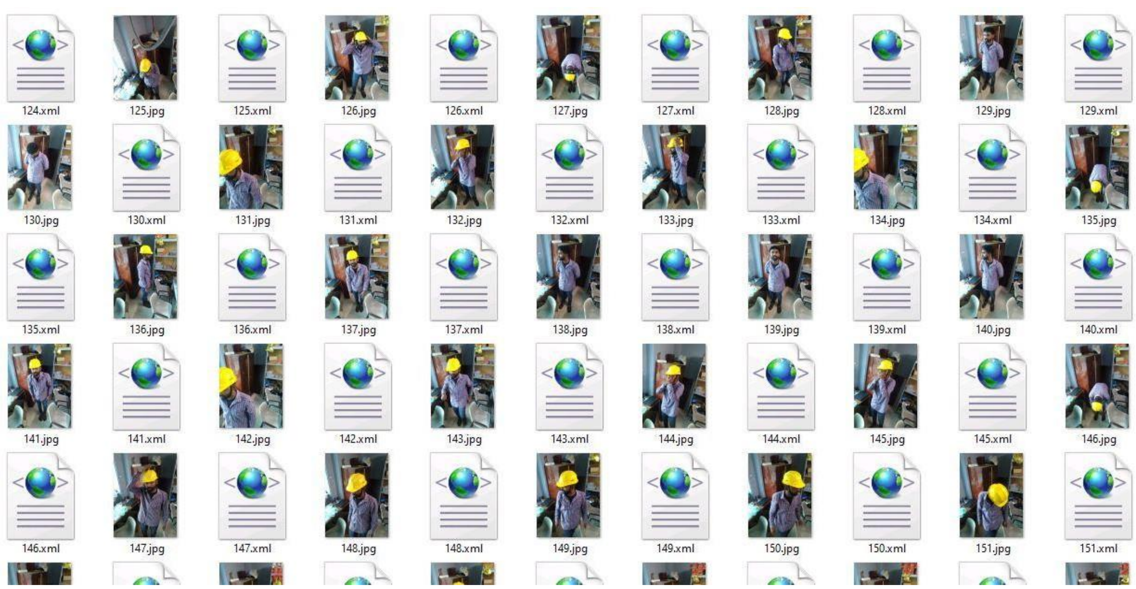


Fig.4.2 Training Images

Preprocessing

The data set was obtained from the Kaggle online platform. The number of images in the data set was reduced to 350. Before feeding the input directly into the model, the data which is the set of fundus images must endure some preprocessing steps which includes i. resizing of images size from 3888 *2951.to 786 * 786 dimension. ii. Execute flip-flop operations that result in a 90-degree rotation of the image. In order to exercise the model in an effective manner, images are flipped. The input data set is categorized into three distinct groups. Training datasets, which are used to train or exercise the model? This data is designated as data package. b. Testing data set, used to evaluate the model. c. Validation data set, which is the dataset used to validate the model. The validation data set is used to ensure that the model has not been over-fitted, while the training data set is used to minimize the loss function. When the training data set is exercised in the model, the weights are updated proportionately, whereas the validation data set does not entail an updating process. Training dataset, validation dataset, but not testing dataset is labeled. Additionally, one quick encoding is applied to the training labels.

Camera Interface for deep learning applications entails capturing images or videos from a camera and feeding them to the deep learning model as input. Here are the deep learning camera interface procedures. Install the required library files: Install OpenCV, a well-known library for computer vision, and other libraries required for deep learning. **Attach the Camera:** Connect the camera to the computer and verify that it is functioning properly. **Capture Photographs or Films:** Utilize the OpenCV libraries to acquire images or recordings from a camera. Face Recognition is a type of computer vision that attempts to identify or verify a person's asserted identity using facial features. This method is used to identify a person's visage, whether or not they are donning a helmet. **Alert System:** The Site Alert module is an adaptable solution that enables site administrators to post an alert on their site.

V. Result and Conclusion

The proposed method for detecting whether or not employees are donning protective headgear is based on convolutional neural networks. To detect safety headwear, this model employs the Faster RCNN-algorithm. Then, a dataset containing 3,261 images of various helmets are constructed and divided into three portions for training and testing the model. The Tensor Flow framework is selected for model training. The mean average precision (Map) of the detection model is stable after training and testing, and the helmet detection model is constructed. The results of the experiment indicate that the method can be used to detect safety headwear worn by construction workers on the job site. The presented method provides an alternative method for detecting safety headgear and enhancing the safety management of construction laborers on the job site.

References

[1] S. Rowlinson, Construction Safety Management Systems. Evanston, IL, USA: Routledge, 2004.

[2] W. Zhang, C.-F. Yang, F. Jiang, X.-Z. Gao, and X. Zhang, ‘‘Safety helmet wearing detection based on image processing and deep learning,’’ in Proc. Int. Conf. Commun., Inf. Syst. Comput. Eng. (CISCE), Jul. 2020, pp. 343–347.

[3] Baining Zhao,Haijuan Lan Detection and Location of Safety Protective Wear in Power Substation Operation Using Wear-Enhanced YOLOv3 Algorithm 2021

[4] Ning li,Xin Lyu Incorporate Online Hard Example Mining and Multi-Part Combination Into Automatic Safety Helmet Wearing Detection 2020

[5] Gionatan Gallo,Francesco Di Rienzo A Smart System for Personal Protective Equipment Detection in Industrial Environments Based on Deep Learning at the Edge 2022

[6] Paul Viola et al Rapid Object Detection using a Boosted Cascade of Simple Features in 2001

[7] S. Barro-Torres, T. M. Fernández-Caramés, H. J. Pérez-Iglesias, and C. J. Escudero, “Real-time personal protective equipment monitoring system,” Computer Communications, vol. 36, no. 1, pp. 42–50, 2012.