Field Worker Safety Detection System

Tushar MoolChanadani
Dept. of Computer Science (AI & ML)
Vellore Institute of Technology (VIT)
Chennai, India
tushar.moolchandani2020@vitstudent.ac.in

Akshat Keshav Dhamale
Dept. of Computer Science (AI & ML)
Vellore Institute of Technology (VIT)
Chennai, India
akshatkeshav.dhamale2020@vitstudent.ac.in

Kartik Deepu
Dept. of Computer Science & Engineerimg
Vellore Institute of Technology (VIT)
Chennai, India
kartik.deepu2020@vitstudent.ac.in

Siddesh Kushare
Dept. of Computer Science & Engineerimg
Vellore Institute of Technology (VIT)
Chennai, India
siddeshvikas.kushare2020@vitstudent.ac.in

Ayush Kapri
Dept. of Computer Science & Engineering
Vellore Institute of Technology
Chennai, India
ayush.kapri2020@vitstudent.ac.in

Geetha S
School of Computer Science & Engineering
Vellore Institute of Technology (VIT)
Chennai, India
geetha.s@vit.ac.in

***Abstract*— A field worker safety detection system is a piece of technological equipment that is intended to monitor and, more importantly, safeguard the safety of employees who are in distant or outside work environments. This system often consists of a mix of hardware and software that can identify possible risks and immediately notify employees as well as management of any potential dangers, so assisting in the prevention of accidents and injuries that may occur on the job. Before beginning work in mines, building sites, or any other hazardous environment, our product will check to see if personnel are clad in all the necessary protective gear and have attained a certain level of safety. A gas sensor that can alert employees in the event of a gas leak and a timer that can record when workers enter and exit the building are also included. Field worker safety detection systems have applications in a wide number of sectors, including agriculture, construction, oil & gas, and mining. They may give financial savings to businesses by minimizing the number of accidents and injuries that occur on the job, which in turn helps to enhance the safety of employees in hazardous or remote work environments. They can also assist improve worker safety in general.**

Keywords: safety equipment detection; object detection; deep learning; YOLOv5; CNN

I. INTRODUCTION

Safety has become a major issue for organizations, governments, and people alike in today's fast-paced and continuously changing world. Workplace safety is critical, particularly in high-risk situations like construction sites, industrial facilities, and chemical plants. Safety equipment detection systems have become a crucial instrument for ensuring the safety of employees in these circumstances.

The use of cameras to monitor employees and their usage of personal protective equipment is a critical component of safety equipment detection systems. (PPE). These cameras are carefully positioned at the worksite's entrance to record photographs and videos of the employees and their surroundings. The system can determine whether employees are wearing the proper PPE, such as hard helmets, safety boots, gloves, and high-visibility vests, before approaching a potentially hazardous location by studying this video.

To protect employees' privacy, cameras in safety equipment detection systems are often intended to collect just the information required to identify the presence of PPE and possible safety threats. In most cases, the film is processed in real time by software that use artificial intelligence and machine learning algorithms to identify and warn safety staff to any possible safety breaches.

Furthermore, typical safety monitoring approaches, such as periodic safety inspections and spot checks, are sometimes inadequate for identifying and responding to safety infractions in real time. This lag time might result in major accidents and injuries. As a result, a more effective and efficient safety monitoring system that can identify and warn safety workers to PPE infractions in real-time is required. This issue may be addressed using a safety equipment detection system that employs cameras and sensors to monitor employees' usage of PPE and flag possible safety concerns.

The objective is to create a system that can continuously monitor the health and safety of field personnel and provide real-time warnings in the case of an emergency or potentially hazardous situation. The system should be able to track workers' whereabouts, movements, and activity levels, and then use this data to flag any potential hazards or incidents. Furthermore, workers should be able to quickly and easily convey emergency alerts through the system, and managers should have real-time visibility into their employees' health and safety via the system. The system should be designed to be user-friendly, sturdy, and reliable, and it should be able to perform in either distant or difficult environments. Field safety is essential for both worker and corporate safety. Many systems attempt to address this challenge, but they all use the same methodical technique of employing image detection to do it.

In Safety Helmet Detection Using YOLOv5, as well as other comparable works such as Construction Safety Equipment Detection System, Visual Detection of Personal Protective Equipment and Safety Gear on Industry Workers, As previously mentioned in the literature review, A Smart System for Personal Protective Equipment Detection in Industrial Environments Based on Deep Learning, Deep Learning Detection of Personal Protective Equipment to Maintain Safety Compliance on Construction Sites, and others all discuss the same a similar generalised approach to tackling this problem.

II. OVERVIEW

A. Convolutional Neural Network (CNNs)

Convolutional Neural Networks or CNNs are deep learning frameworks that have recently popped up in the field of computer vision. They are mostly used in image based applications, wherein these networks detect and identify features, objects, perform image segmentation, classification and various other uses autonomously [2].

The general structure of the convolutional neural network is as follows: input layer, hidden layers and output layer. The hidden layers might commonly involve convolutional, pooling, activation or RELU layers, normalization, fully connected layers. They are called as "hidden" as a result of them masking the input and the output, through the activation function and the final convolution [3].

Convolutional Neural Networks typically consists of multiple convolutional layers wherein the convolutional layers which contain several kernels of various sizes, apply the convolutional or dot product operation on their input to produce feature maps. The feature maps are further sent to a pooling layer where maximum activations are selected from a small neighborhood region of features. These two results in a reduction of feature dimensions. The fully connected layer is usually at the near end layer of the convolutional neural network architecture, and the final high level classification or information derivation from the input image is performed. Here, the convolutional, fully connected layers have neurons, whose weights are altered during training to get the desired outputs. CNNs that are commonly deployed across multiple computer vision applications include ResNet[4][5], AlexNet[6] and GoogLeNet[7]. Since each of these CNNs have a varying architecture, applying every CNN for the same application may produce varying outputs and results. Thus, the CNN to be deployed should be analyzed for maximizing performance of the use case.

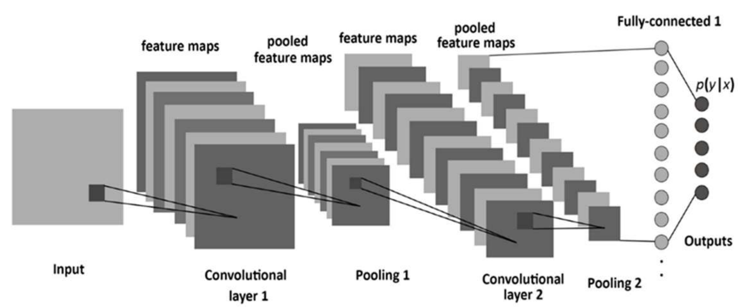
B. Image Processing Techniques

Image processing techniques are the traditionally used methods for computer visions applications. They involve basic processing techniques such as segmentation, extraction, colour based identification techniques, shape and texture based segmentation amongst a variety of other techniques that target a specific feature of the object or objects that are to be identified in the image. These techniques are applied on a training image set to generate a standard feature set of the target object to be detected. These feature sets are compared to features extracted from images which have objects to be detected or classified using similarity comparison metrics or machine learning techniques such as SVMs.

C. Image Processing Techniques VS CNN

Deep learning solves image detection difficulties. Detecting and classifying visual characteristics is image detection. Multiple-layered artificial neural networks classify photographs using deep learning algorithms. CNNs are utilized in image detection applications. CNNs have convolutional, pooling, and fully connected layers. Convolutional layers extract visual characteristics using learnable filters that travel across the picture. Pooling layers down sample feature maps to minimize computational complexity and dimensionality. Fully connected layers classify traits extracted by convolutional and pooling layers.

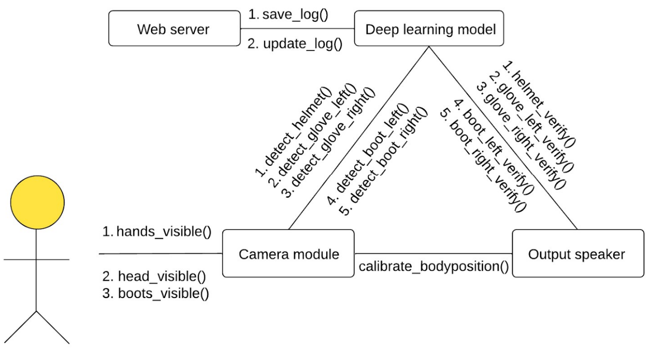
The network's weights are adjusted to lessen the discrepancy between anticipated and real labels while training a CNN. Backpropagation, which sends errors across the network, changes weights. After training, a CNN can recognize pictures by putting fresh photographs through the network and predicting objects or features. Face recognition, object identification, and medical image analysis are deep learning image detection applications.



CNN Architecture

D. Model Architecture

The UML communication diagram is a kind of interaction diagram that demonstrates how objects communicate with one another. The object diagram may be extended into a communication diagram, which displays the objects as well as the messages that go from one to another. A communication diagram will not only display the relationships between items, but it will also illustrate the messages that the objects send to one another.



Camera module: Basic camera module attached to arduino circuit.

hands_visible() : Checks if person's hands are visible in the camera.

head_visible() : Checks if person's head is visible in the camera.

boots_visible() : Checks if person's boots are visible in the camera.

Deep learning model: AI model deployed as an api endpoint where the camera will send images once all essential body parts are visible.

detect_helmet() : Detects if person is wearing helmet or not.

detect_glove_right() , detect_glove_left() : Detects if person is wearing gloves or not.

detect_boot_right(), detect_boot_left() : Detects if person is wearing boots or not.

Web server: Remote server where data recorded of a person and time will be saved and be updated regularly.

save_log() : Saves current data into the server.

update_log() : Updates current data according to ID.

Output Speaker: Speaker module attached to arduino board which will give outputs based on the processing done by deep learning model.

helmet_verify(), glove_left_verify(), glove_right_verify(), boot_left_verify(), boot_right_verify() : Returns true or false for detection of helmet, gloves and boots respectively.

III. RELATED WORK

Ref [1], the authors propose an IoT-based system that uses IoT devices to collect data on air quality pollutants like carbon monoxide and methane in the mine and sends it to the Azure machine learning platform for analysis. The authors have implemented a machine learning model that can predict pollutant levels in real time. Data preparation, feature extraction, and model selection were suggested to increase prediction model performance. IoT devices capture mine air quality data, including carbon monoxide and methane. The Azure machine learning platform analyses this data and improves regression model by 16.9%.

In [2], Fangbo Zhou et al. present a method to recognise safety helmets in construction and industrial worker photos and videos to increase safety. Safety helmets were detected in photos and videos using the YOLOv5 object detection technique. The authors argue that this is an important task because it can help identify workers who are not wearing safety helmets, which can prevent accidents and injuries in these environments. object detection algorithms, such as Faster R-CNN and RetinaNet, and found that YOLOv5 performed better in terms of speed and accuracy. It should be noted that the various models improved by 9.75% on average.

Kang Li et al. [3]. proposes a construction and industrial safety helmet detection system. The technology alerts employees and supervisors of helmetless workers to increase safety. Using ViBe backdrop modelling. After motion object segmentation, real-time human classification framework C4 locates pedestrians. C4 enhances object detection.

Arjya Das Mohammad, et al. found that their deep learning model could correctly recognise face masks [4]. The model worked well in all real-world circumstances, including various lighting and face emotions. In conclusion, this deep learning model may be used in diverse scenarios to enforce face mask wear and avoid COVID-19. Uses a dataset of persons wearing and not wearing face masks to train their deep learning model for real-time face mask identification in public areas. The report does not specify the amount of photos, their sources, or their variety. The study's deep learning model was trained using a dataset designed for face mask identification. 95.77% accuracy.

[5] by Venkata Santosh Kumar et al. utilises machine vision and deep learning to recognise construction workers using PPE such safety helmets, gloves, and high-visibility vests. Identifying employees without PPE and reporting them or a supervisor improves safety. The research uses a

transfer learning-based Convolutional Neural Networks model.

Nath, et al. [6] suggest utilising deep learning to identify PPE in construction worker pictures and videos. A deep neural network and computer vision algorithm detects PPE in photos and movies.

Yogesh Kawade, et al., [7]. The authors suggest a system that employs computer vision and deep learning to automatically recognise and categorise employees wearing or not wearing safety gear such hard helmets, vests, and glasses. Convolutional neural networks (CNNs) were used to train the system on a huge dataset of construction workers wearing and not wearing safety gear. The research found that the suggested system could precisely identify safety equipment presence or absence. The technology worked effectively on real-world building sites under variable lighting and face expressions. This paper detects hard helmets and PPE vests using YOLOv3. Geometric relationships of OpenPose and YOLOv3 outputs were significant to identify suitable PPE.

Jonathan Karlsson, et al. [8] propose a way to identify industrial employees using PPE and safety gear. The authors employ computer vision and deep learning to automatically recognise and categorise workers using PPE and safety gear including hard helmets, safety vests, and safety glasses. A vast collection of photos of industrial employees wearing and not wearing PPE and safety gear educated the system. The research found that the suggested system could precisely identify PPE and safety gear. The technology functioned effectively in real-world industrial environments, including variable lighting and face expressions. Kaggle dataset includes positive and negative hardhat and safety vest samples. YOLOv4 shows good accuracy scores and metrics.

In [9], Gioatan Gallo et al. employ computer vision and deep learning to automatically recognise workers with or without helmets, safety vests, and eyewear. A huge dataset of industrial workers with and without PPE was used to train a CNN. The investigation found that the suggested approach properly detected her necessary PPE. In a genuine industrial setting, the system functioned effectively under varying lighting and face expressions. Finally, their PPE detecting technology may increase workplace safety. Computer vision and deep learning can monitor and enforce industrial PPE usage in real time.

ND Nath, et al. [10] developed a deep learning-based system to identify PPE on building sites to ensure safety. Computer vision detects employees' PPE to increase safety compliance. The article details the deep learning model's creation, testing, accuracy, and prospective applications in real-world building sites. Deep learning (DL) has enabled convolutional neural network (CNN) algorithms to more accurately detect PPE components using RFID tags, LANs, and short-range transponders (Kelm et al., 2013). able to recognise workers without hat or vest (W), with hat (WH), with vest (WV) and with both (WHV) with 90% accuracy, and the colour of individual PPE components with 77% accuracy.

Using green edge computing and deep learning, Xiao Ke et al. provide a real-time PPE detection system [11]. The system can detect PPE at over 100 frames per second (FPS) and improve worker safety in industries. A channel pruning algorithm based on the BN layer scaling factor reduces the size of the detection model by 32% and detection by 25%.

Pedro Torres et al. [12] describe a robust, real-time PPE detector for industrial settings. The author used YOLO-v3 for hard hat identification, which increased by 21.52% and 13.96% compared to YOLO-v4-AP1 and AP2.

Saudi, et al. [13] present a Faster R-CNN-based image detection model to assess construction workers' safety on a construction site. ResNet152, Faster RCNN, Deep CNN, Google Inception v3 70% accuracy.

Moohialdin et al. offer a real-time computer vision (CV) system to identify construction workers' PPE and postures. Python data-labeling tool was used to annotate the chosen datasets, and the labelled datasets were utilised to develop a detection model in TensorFlow. Model testing and validation showed posture classification accuracy of 72% and 64%, respectively.

Deep Learning-Based Safety Helmet identification in Engineering Management Based on Convolutional Neural Networks [15] suggests computer vision-based real-time PPE identification. The SSD-MobileNet technique employs convolutional neural networks and has 95% accuracy and 77% recall.

Ferdous et al. [16] employ a computer vision (CV)-based automated PPE identification system to identify different forms of PPE. YOLOX's anchor-free design. YOLOX-m outperforms 3.29%.

[17] allows real-time PPE detection. The detector uses the real-time object detection-optimized YOLOv4 computer vision model. YOLOv4 computer vision model detector weights to TensorFlow format for live detection performance testing and live object count and record keeping. Object detector mAP is 79%.

In “Applying the Haar-cascade Algorithm for Detecting Safety Equipment in Safety Management Systems for Multiple Working Environments” [18], Phuc et al. use the algorithm to calculate a score based on the safety equipment and working environment to determine the danger of the current working environment. Based on this data, the system chooses whether to alert. This work employed principal component analysis Haar cascade technique. CARs ranged from 66.8% to 68.2% with minimal error rate.

Qiu et al. [19] present a technique for identifying and assessing safety protection equipment to guarantee its good operation and avoid accidents. Sensors and algorithms identify safety threats and assess protective equipment performance. Protective equipment must work properly to increase safety. This target identification technique uses linear distance and angle restrictions.

A deep learning-based system for construction equipment detection: from development to deployment [20] shows a real-world solution. The solution is created, tested, and deployed on a work site to identify various construction equipment. The deep learning algorithm classifies equipment in photos and videos to improve construction safety and efficiency. Results demonstrate the solution's efficacy and practicality. Above deep learning techniques are 90% accurate.

Marks, et al. [21] provide a technique for evaluating construction equipment proximity detection and alarm systems. The system alerts operators to probable crashes to increase safety. The technology's ability to identify risks and notify operators is tested in real-world scenarios. The results of the tests are used to determine the technology's effectiveness and identify areas for improvement to ensure safe and efficient operation of construction equipment emerging radio frequency (RF) remote sensing technology to demonstrate the test method's ability to evaluate proximity detection and alert systems' ability to alert when heavy construction equipment and workers are too close. Authors achieved 89% accuracy.

A Convolutional Neural Network (CNN) methodology is proposed in "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach" [22]. CNN trains on construction site pictures and videos to locate employees and heavy equipment. Real-time personnel and equipment tracking improves construction site safety and efficiency. The research found that the CNN-based personnel and equipment identification technology may be used on construction sites. IFaster R-CNN algorithms automatically recognise objects in real time.

Wang et al. [23] use computer vision and deep learning to forecast construction worker and equipment safety concerns. A deep learning model is trained using building site data to identify and forecast safety issues. Real-time alerts and danger avoidance advice increase building site safety. The findings suggest that the proposed method may forecast safety issues and be used on building sites. Faster R-CNN detects scaffold workers.

“Substation Safety Awareness Intelligent Model: Fast Personal Protective Equipment Detection using GNN Approach” [24] by Zhao, et al. offers a Graph Neural Network (GNN)-based intelligent model for substation PPE detection. Identifying employees without PPE improves safety awareness and compliance. The GNN model is trained on substation pictures and videos to recognise employees and their PPE for real-time alerts and PPE awareness. The findings suggest that the GNN-based model can identify PPE and be used in substations. PPE is detected using a few-shot graph neural network (GNN).

IV. METHODOLOGY

The dataset include photos of safety gear that field workers may need, marked for the elements we want to identify. To evaluate the model's training photos, a testing set was provided.

We present a low-cost, effective IOT-based architecture with computer vision technologies to identify employees' safety equipment. We implemented YOLOv5 with keras backend for object detection. Keras is a Python-based neural network API that makes Tensorflow algorithm development easy. You only look once (YOLO) is a cutting-edge CNN-based object identification system for real-time processing. YOLO grids the picture. Grid cells anticipate one item and a predetermined number of border boxes. Each grid cell forecasts B boundary boxes with one box confidence score. No matter how many boxes B, it finds one thing. Finally, it predicts C conditional class probabilities. (one per class for the likeliness of the object class). Each boundary box has x, y, w, h, and a box confidence score. The confidence score indicates the box's item likelihood.

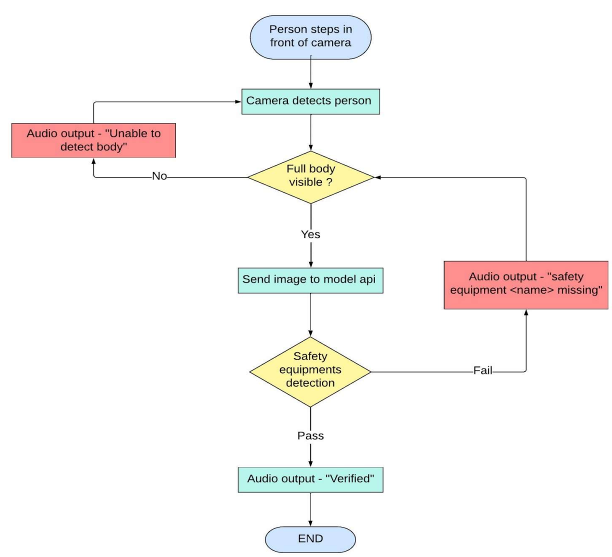


Fig 4.1: Proposed Detection System Diagram

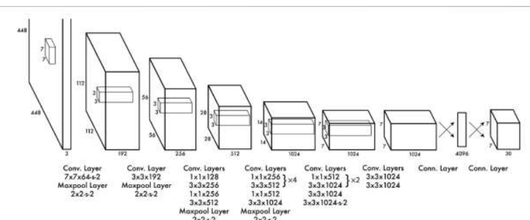
Our camera module will use the mediapipe api pipeline for human posture recognition when a person enters the picture. Once all body points are in frame, one snapshot of the individual will be captured and submitted to the YOLOv4 model api endpoint, which will evaluate it and report any missing safety equipment. The individual is confirmed through a speaker node once all safety equipment is validated. The speaker node alerts users if person or safety equipment calibration fails.

A. *Object Detection and Recognition Process*

Object detection finds and distinguishes objects in an image or scene. Computers can do a number of tasks on their own using object detection methods in computer vision. However, image processing-based procedures, which have been traditionally employed, should be avoided due to their longer processing time, more difficult methods, and lesser accuracy. Deep Learning-based object identification algorithms are now considered the best. These algorithms can detect things more precisely and learn picture and object properties on their own.

Computer vision-based item identification tasks involve picture categorization, localisation, and object detection. Image classification involves analysing a photo and classifying it into one of several categories. Computer vision's basic application. This approach predicts the picture's genre but not its contents. Categorization with localization is the result. Numerous annotations identify and locate an item in this category. However, it can only recognise one type of items, which is its main limitation. Labelling and categorising objects requires object detection.

Deep learning performs well in computer vision applications, which are increasingly focused on object recognition. Region-based CNNs are the best solution for item recognition in this area. An end-to-end neural network that predicts bounding boxes and class probabilities is recommended by the You Only Look Once (YOLO) algorithm. It differs from past object identification methods that reused classifiers to recognise objects.



The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1 × 1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution (224 × 224 input image) and then double the resolution for detection.

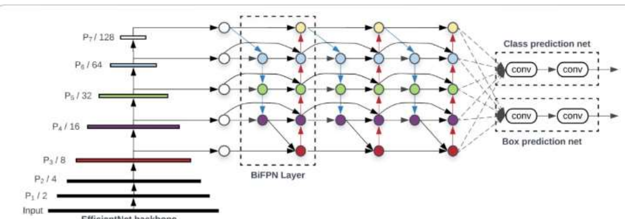
YOLO outperforms real-time object detection algorithms by a large margin. A novel object recognition method achieved this. Unlike Faster RCNN, which uses the Region Proposal Network to identify possible regions of interest and then performs recognition on each region, YOLO makes all of its predictions using a single fully connected layer. YOLO just needs one iteration each photo, unlike Region Proposal Network methods.

YOLO development

- The real-time YOLO system identifies things. One neural network finds objects in photos and videos.
- Safety wear detection is one of YOLO's most essential uses. It checks construction workers and industrial employees for helmets, gloves, and vests.
- YOLO can identify several objects in a single picture and accurately locate them, making it ideal for safety wear detection. It is quicker and more accurate than typical object detecting systems.
- For safetyware detection, YOLO needs large picture and video datasets. These databases must include photographs of employees in different settings and wearing different safety gear.
- YOLO can recognise safety equipment in construction site and industrial camera feeds after training on a dataset. Ensuring employees use proper safety gear may increase worker safety.
- YOLO detects dangerous substances and building site safety issues for worker safety.
- Computer vision technologies will improve YOLO and other object detection systems. This might boost worker safety and other industrial and occupational health and safety measures.

B. *YOLOv5 Development*

YOLO uses a simple deep convolutional neural network to recognise objects in photos. This graphic shows YOLO's CNN model's structure.



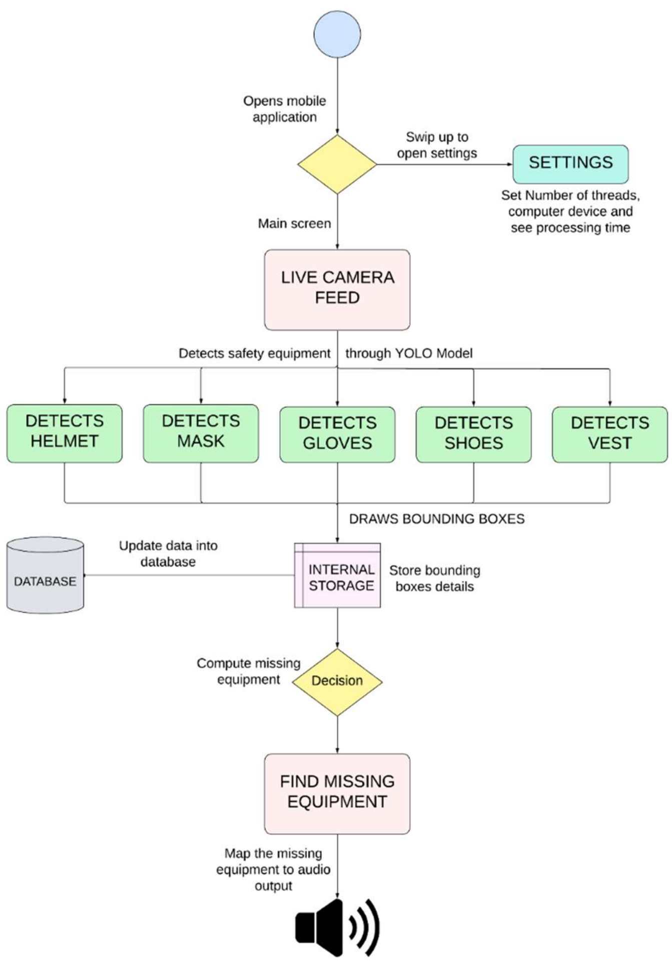
EfficientDet architecture – It employs EfficientNet [39] as the backbone network, BiFPN as the feature network, and shared class/box prediction network. Both BiFPN layers and class/box net layers are repeated multiple times based on different resource constraints as shown in Table 1.

ImageNet pre-trains the model's first 20 convolution layers, which are implemented using a temporary average pooling and completely connected layer. A pre-trained network is repurposed for detection after convolution and connected layers improve its performance. Last fully connected layer in YOLO predicts class probabilities and bounding box locations.

YOLO turns pictures into square grids. The middle grid cell detects an object. Grid cells provide predicted B bounding boxes and confidence ratings. These ratings indicate how well the model predicts an item is in the box.

YOLO predicts several bounding boxes per grid cell. Train one bounding box predictor per item. YOLO assigns "responsibility" for predicting an item to the predictor with the highest current IOU with the ground truth. Thus, bounding box predictors specialise. By accurately predicting object sizes, shapes, and classifications, predictors increase their recall.

YOLO models need non-maximal suppression. (NMS). NMS post-processing improves object detection accuracy. In object detection, many bounding boxes may be constructed. Even if its bounding boxes overlap or vary, it's the same. NMS eliminates duplicate and faulty bounding boxes to create a single bounding box per picture item.



C. *Mobile App/Detection System Integration*

The Python-based YOLOv5 and SQL database combine the java android application with safety equipment detection. This software was developed in many stages.

The YOLOv5 model must be trained on a large dataset of safety equipment photos. PyTorch or TensorFlow are used to train the model using a labelled dataset. The model is trained on batches of photos and weights are modified to minimise loss across numerous epochs.

After training, the model must be exported and incorporated into the Java Android app. ONNX, TorchScript, and TensorFlow Lite can export the model. The application's codebase may utilise the exported model for object detection.

Object Detection Algorithm: The Java Android application requires an object detection algorithm that examines the device's camera feed and provides the bounding boxes around discovered items. OpenCV or TensorFlow Lite may build the object detection technique using the YOLOv5 model.

Real-Time Object Detection: Java Android applications must be optimised for real-time object detection. NMS and anchor boxes are used to increase model accuracy and efficiency.

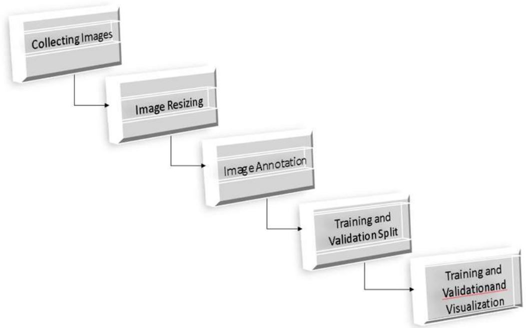
User Interface: The Java Android app must show the device's camera feed, detection results, and notification messages. The user interface should be straightforward, simple to use, and provide feedback.

Testing: To make sure the Java Android app works, test it. This entails testing the app's real-time safety equipment detection on numerous mobile devices and platforms.

V. EXPERIMENTAL TEST & ANALYSIS

The purpose-specific collections of text, pictures, and videos. Data sets are crucial for building deep learning algorithms like CNNs. Without them, networks would fail. CNNs self-learn through cost reduction and backpropagation utilising dataset pictures. If the datasets are low-quality or small, the network's learning and training will fail, resulting in useless or harmful false or true negative detections. Datasets also underpin methodology comparisons. Datasets determine which method is faster, more efficient, and more precise when using photos to create outputs or results for other uses. Selecting and curating datasets for a project is essential for deep learning algorithm success.

This work collects gloves, shoes, vest, helmet, and mask photos from multiple categories to train the convolutional neural network. The neural network is highly durable and reliable since it uses a range of picture kinds.



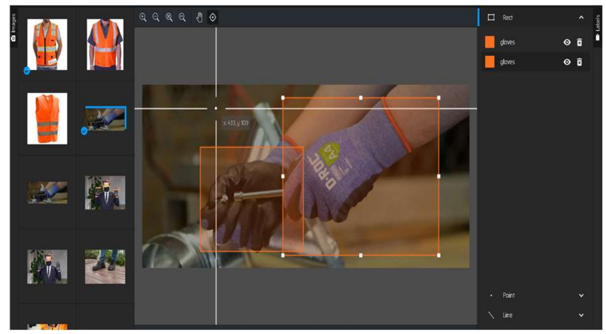
Preparing the dataset includes reducing the photographs to a model-appropriate size to save processing time. Picture annotation is also included. Annotating an image requires drawing enclosing boxes and labelling interesting things. Without annotation, computer vision object recognition fails. Annotated data informs a training neural network on object differences. After proper training, the neural network can learn the features of annotated and tagged objects and distinguish and identify the essential items in a test picture. Figure 1 shows the dataset preparation procedure.

This study employed numerous datasets to create the optimum training dataset for detecting a helmet, vest, mask, mittens, and shoes. This project used roboflow datasets and makesense.ai to label photos. A vast number of photos from various datasets were collected into a single dataset and validated to contain just helmet, mitten, vest, mask, and shoe shots. Figure 4.2 illustrates these photos. 11793 of these photos were hand-selected and annotated for training, validation, and testing.

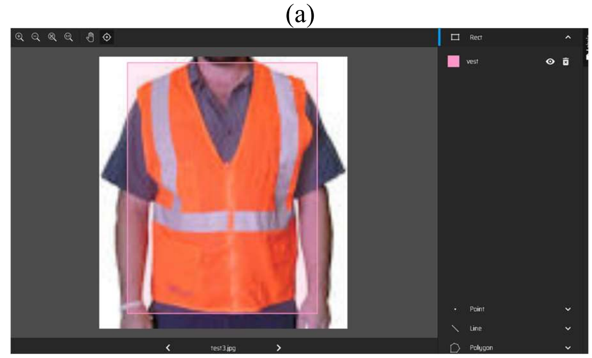
Fig. 5.1 Sample Images in dataset

Makesense.ai annotated images. Computer vision company MakeSense.ai offers a variety of annotation and labelling services for machine learning systems. They annotate images, videos, texts, and voices. MakeSense.ai uses several tools and methodologies to reliably classify items in photos. LabelImg is a popular item detection application. LabelImg users can create bounding boxes around items of interest in each training set image. Customers may train and verify machine learning models to better distinguish and label things.

MakeSense.ai optimises input data for machine learning model training and validation using TF Record, a binary file storage structure used by Tensorflow. It's Tensorflow. This format reduces storage space and read/copy time, improving functionality. TF Records may integrate several datasets into one input data repository, making huge dataset processing easier. Machine learning applications handle huge amounts of data, making this method very useful. MakeSense.ai uses LabelImg and TF to deliver high-quality labelling and annotation services to its customers.



(a)



(b)

Figure 5.2 (a) & (b): Dataset Annotation Process

VI. RESULT OF EXPERIMENTAL TEST

Experiments used entire colab resources, including jupyter notebook, an Intel core CPU, 12 GB of virtual memory, and a Google GPU. To test the algorithms, Python, Tensorflow, and Jupyter Notebook were configured.

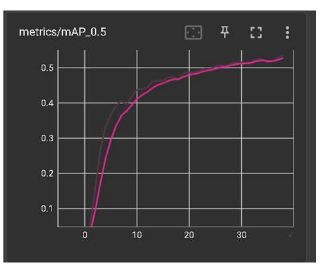
Python and the CNN-based Yolov5 algorithm created the model. Experimentation separated the annotated dataset into three. The training dataset has 9700 pictures, the validation dataset 1000, and the test dataset 400. During training, rotation, flip, colour saturation, hue, and contrast are programmed. After selecting 64, training began. Thus, training took 100 epochs. The surgery took seven hours. Training, validation, and testing show the suggested technique is exact and effective. The network detects 98%, 93%, and 91% of the training, validation, and test datasets, respectively. Table 4.4 summarises results.

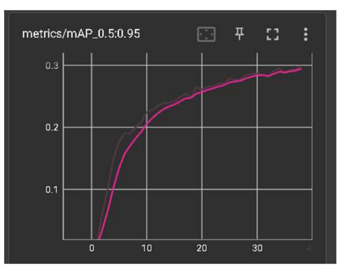
Table 5.3: Results of the Proposed System

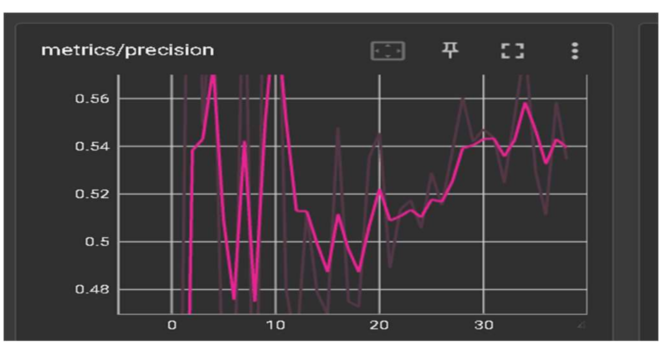
Dataset	Total Number of Images	Accuracy Rate
Train	9700	98 %
Validation	1000	93%
Test	400	91 %

Following training, an Android application was developed. Using Java, the application was developed. Tensorflowlite was derived from the trained model.(About application)

Fig. 5.4, Fig. 5.5, shows YOLOv5 detection results, Fig. 5.6, Fig. 5.7 and Fig 5.8 represents the evaluation metrics obtained after training the model. The Application functions are displayed in Fig. 5.9 and Fig. 5.10.







VII. CONCLUSION

To sum up, machine learning algorithms have shown to be an YOLO and YOLO v5 differ in the object detection model training data. The 20-object PASCAL VOC dataset trained YOLO. YOLO v5 was learned on D5, a much larger and more diverse dataset with 600 item categories. Dynamic anchor boxes are introduced in YOLO version 5. It involves clustering ground truth bounding boxes and using their centroids as anchor boxes. Click here. This lets anchor boxes match recognised items in size and shape. SPP is also included in YOLO version 5. SPP pooling layers reduce feature map spatial resolution. SPP improves small object recognition by letting the model view objects of various sizes. YOLO v5 uses SPP like YOLO v4, but it improves the design to accomplish more and yield better outcomes. The loss function used to train YOLO v4 and v5 is similar. YOLO version 5 introduces "CIoU loss," a modification of the IoU loss function that improves model performance on imbalanced datasets.

REFERENCES

[1] Ding, W., Liu, J., Zhang, W., Zhang, W., & Wang, Q. (2020). An Internet of Things System for Underground Mine Air Quality Pollutant Prediction Based on Azure Machine Learning. Journal of Sensors, 2020, 1-10.

[2] F. Zhou, H. Zhao and Z. Nie, "Safety Helmet Detection Based on YOLOv5," 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), Shenyang, China, 2021, pp. 6-11, doi: 10.1109/ICPECA51329.2021.9362711.

[3] Kang Li and Xiaoguang Zhao and Jiang Bian and Min Tan Automatic Safety Helmet Wearing Detection CoRR abs/1802.00264 2018

[4] A. Das, M. Wasif Ansari and R. Basak, "Covid-19 Face Mask Detection Using TensorFlow, Keras and OpenCV," 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, India, 2020, pp. 1-5, doi: 10.1109/INDICON49873.2020.9342585.

[5] Delhi, Venkata Santosh Kumar, R. Sankarlal, and Albert Thomas. "Detection of personal protective equipment (PPE) compliance on construction site using computer vision based deep learning techniques." Frontiers in Built Environment 6 (2020): 136.

[6] Nath, Nipun D., Amir H. Behzadan, and Stephanie G. Paal. "Deep learning for site safety: Real-time detection of personal protective equipment." Automation in Construction 112 (2020): 103085.

[7] Construction Safety Equipment Detection System Yogesh Kawade, Mayur Kasture, Yash Mallawat, Aniruddh Dubal Student, B.Tech., Construction Engineering and Management, Symbiosis Skills and Professional University, Pune, Maharashtra, India Asst. Prof., Construction Engineering, Symbiosis Skills and Professional University, Pune, Maharashtra, Pune, India

[8] Karlsson, Jonathan, et al. "Visual Detection of Personal Protective Equipment and Safety Gear on Industry Workers." arXiv preprint arXiv:2212.04794 (2022).

[9] Gallo, Gionatan, et al. "A smart system for personal protective equipment detection in industrial environments based on deep learning." 2021 IEEE

International Conference on Smart Computing (SMARTCOMP). IEEE, 2021.
[10] Nath, Nipun D., and Amir H. Behzadan. "Deep learning detection of personal protective equipment to maintain safety compliance on construction sites." Construction Research Congress 2020: Computer Applications. Reston, VA: American Society of Civil Engineers, 2020.
[11] Ke, Xiao, Wenyao Chen, and Wenzhong Guo. "100+ FPS detector of personal protective equipment for worker safety: A deep learning approach for green edge computing." Peer-to-peer networking and applications (2022): 1-23.
[12] Torres, Pedro, et al. "A Robust Real-time Component for Personal Protective Equipment Detection in an Industrial Setting." ICEIS (1). 2021.
[13] Saudi, Madihah Mohd, et al. "Image detection model for construction worker safety conditions using faster R-CNN." International Journal of Advanced Computer Science and Applications (2020).
[14] Moohialdin, Ammar, et al. "A Real-Time Computer Vision System for Workers' PPE and Posture Detection in Actual Construction Site Environment." EASEC16: Proceedings of The 16th East Asian-Pacific Conference on Structural Engineering and Construction, 2019. Springer Singapore, 2021.
[15] Li, Yange, et al. "Deep learning-based safety helmet detection in engineering management based on convolutional neural networks." Advances in Civil Engineering 2020 (2020): 1-10.
[16] Ferdous, Md, and Sk Md Masudul Ahsan. "PPE detector: a YOLO-based architecture to detect personal protective equipment (PPE) for construction sites." PeerJ Computer Science 8 (2022): e999.
[17] Protik, Adban Akib, Amzad Hossain Rafi, and Shahnewaz Siddique. "Real-Time personal protective equipment (PPE) detection using YOLOv4 and TensorFlow." 2021 IEEE Region 10 Symposium (TENSYMP). IEEE, 2021.
[18] Phuc, Le Tran Huu, et al. "Applying the Haar-cascade Algorithm for detecting safety equipment in safety management systems for multiple working environments." Electronics 8.10 (2019): 1079.
[19] Qiu, Jianzhong, et al. "A Target Detection and Evaluation Method for Safety Protection Equipment." Journal of Physics: Conference Series. Vol. 1966. No. 1. IOP Publishing, 2021.
[20] Arabi, Saeed, Arya Haghighat, and Anuj Sharma. "A deep learning based solution for construction equipment detection: from development to deployment." arXiv preprint arXiv:1904.09021 (2019).
[21] Marks, Eric Daniel, and Jochen Teizer. "Method for testing proximity detection and alert technology for safe construction equipment operation." Construction Management and Economics 31.6 (2013): 636-646.
[22] Fang, Weili, et al. "Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach." Advanced Engineering Informatics 37 (2018): 139-149.
[23] Wang, Mingzhu, et al. "Predicting safety hazards among construction workers and equipment using computer vision and deep learning techniques." ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction. Vol. 36. IAARC Publications, 2019.
[24] Zhao, Meng, and Masoud Barati. "Substation Safety Awareness Intelligent Model: Fast Personal Protective Equipment Detection using GNN Approach." IEEE Transactions on Industry Applications (2023).