*A report submitted*

*by*

**PHAROS SOPHY SAMUEL T J (20BAI1049)**

*in partial fulfilment for the award of the degree*

*of*

## B.tech in Computer Science and Engineering with Specialization in Artificial Intelligence and Machine Learning

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**



July, 2024

**ABSTRACT**

In the context of industrial and safety-critical environments, the implementation of effective safety protocols is paramount to safeguarding the well-being of individuals. Personal Protective Equipment (PPE) plays a pivotal role in mitigating occupational hazards, and ensuring its proper utilization is of utmost importance. This research project delves into the domain of automated detection of essential safety gear, namely safety helmets, masks, and vests, leveraging sophisticated deep learning techniques. The study places a particular emphasis on two state-of-the-art object detection algorithms, Faster R-CNN and YOLOv5, with the overarching goal of developing robust models that exhibit a high degree of accuracy in identifying and precisely localizing safety equipment within images. The initiation of the project involves a meticulous process of compiling a diverse and well-annotated dataset. This dataset comprises images featuring individuals from various occupational settings donning safety helmets, masks, and vests send to mail. To enhance the model's ability to generalize across different scenarios, data augmentation techniques are systematically applied, enriching the dataset with variations in lighting conditions, perspectives, and environmental contexts. Subsequently, both Faster R-CNN and YOLOv5 models undergo comprehensive training using this augmented dataset, wherein the specific characteristics and performance requirements of each algorithm are meticulously considered. The choice between Faster R-CNN and YOLOv5 is a nuanced decision, influenced by factors such as real-time processing demands, detection accuracy, and computational efficiency. Comparative analyses of the two models in terms of precision, recall, and overall detection performance guide this decision-making process. The outcomes of this research, beyond their theoretical significance, hold substantial practical implications for the domain of occupational safety. By providing a reliable and automated means of monitoring the correct deployment of safety equipment, the developed models contribute significantly to the enhancement of safety compliance and surveillance measures. This research project uses advanced deep learning algorithms, specifically YOLOv5, to automate the detection of critical safety equipment in industrial situations, such as helmets, masks, and vests. The goal of the project is to create extremely accurate models that can locate and identify safety equipment within photos with precision through the painstaking collecting and augmentation of datasets. A thorough comparison of YOLOv5 is made taking into account several aspects such as efficiency, accuracy, and real-time processing. The process of making decisions is guided by comparative studies, with the ultimate goal of improving safety compliance and surveillance by means of dependable automated monitoring of the deployment of safety equipment.

**CONTENTS**

**CONTENTS…………………………………………………………….**iii

**CHAPTER 1**

**INTRODUCTION**

* 1. BACKGROUND AND MOTIVATION………………………………………………......7
  2. PROBLEM STATEMENT………………………………………………………………...8
  3. OBJECTIVES OF THE STUDY…………………………………………………………..9

**CHAPTER 2**

**LITERATURE REVIEW**

1. REVIEW OF EXISTING LITERATURE ON OBJECT DETECTION…………......................................................................................................10
   1. OVERVIEW OF ARTIFICIAL INTELLIGENCE AND DEEP LEARNING…………………………………………………………................................14
   2. SUMMARY OF KEY FINDINGS FROM LITERATURE REVIEW…………………...18

**CHAPTER 3**

**METHODOLOGY**

* 1. DESCRIPTION OF DATASET ………………………………………………………....23
  2. ALGORITHMS…………………………………………………………………………..24
  3. PROPOSED METHODS………………………………………………………………...26

**CHAPTER 4**

**RESULTS**

* 1. PRESENTATION OF EXPERIMENTAL RESULTS…………………………………..30

**CHAPTER 5**

**DISCUSSION**

* 1. ANALYSIS AND INTERPRETATION OF RESULTS………………………………...33
  2. ADVANTAGES………………………………………………………………………….36
  3. LIMITATIONS AND CHALLENGES ENCOUNTERED……………………………...36

**CHAPTER 6**

**CONCLUSION**

* 1. SUMMARY OF KEY FINDINGS………………………………………………………37
  2. RECOMMENDATIONS FOR FUTURE WORK……………………………………….37

**CHAPTER 7**

**REFERENCE**

**CHAPTER 8**

**APPENDICES**

**CHAPTER 1**

**INTRODUCTION**

The combination of deep learning and computer vision has become a powerful tool in the field of occupational safety, helping to increase workplace security. In order to tackle the crucial problem of real-time safety helmet, mask, and vest detection inside industrial settings, this research explores the construction and comparative analysis of the YOLOv5 deep learning model. Taking into account that these products are essential parts of personal protective equipment and that they are essential for reducing risk and preventing injuries, the study aims to give a detailed understanding of the performance characteristics of each model. Examining the one-stage detection strategy, anchor box clustering, and feature pyramid network of YOLOv5, the research tries to uncover the complexities by exploring architecture components, region proposal mechanisms, and feature extraction methodologies.

Conventional object identification techniques often depend on a region selection strategy based on sliding windows, which is popular but has limitations including non-targeted Ness and high computing cost. Furthermore, the manually- crafted feature extractor utilized in these techniques frequently demonstrates inadequate resilience to the wide range of targets encountered in practical situations. But with significant improvements in computer speed, especially with regard to GPU capabilities, the deep learning paradigm has drawn interest from researchers once more. Specifically, convolutional neural networks (CNNs) have proven remarkably effective in a wide range of computer vision applications. When it comes to object detection, modern techniques primarily use CNN-based approaches because of their built-in capacity to automatically extract pertinent visual elements without requiring human design.

Numerous academics have put forth a number of deep learning-based object detection methods in recent years. The YOLO algorithm was proposed by Redmon et al. in 2015 and is significantly faster than previous techniques. The YOLOv3 object detection technique, which Redmon et al. presented in 2018, considerably increased detection accuracy and speed. By 2020, YOLOv5, the fifth algorithm in the YOLO family, has been developed. The requirements for helmet detection from the product safety field are combined in this document.

In the domain of occupational safety, the intersection of computer vision and deep learning has emerged as a powerful solution for enhancing workplace security. This research delves into the development and comparative analysis of two prominent deep learning models, Faster R-CNN and YOLOv5, to address the imperative task of real-time detection of safety helmets, masks, and vests in industrial environments. As pivotal components of personal protective equipment, these items play a crucial role in minimizing risks and preventing injuries. The study aims to offer a nuanced understanding of the performance characteristics of each model, encompassing metrics such as precision, through an exploration of architecture components, region proposal mechanisms, and feature extraction techniques, the research aims to unravel the nuances of Faster R-CNN. Simultaneously, it scrutinizes the one-stage detection approach, anchor box clustering, and feature pyramid network of YOLOv5. Sent to mail by evaluating detection accuracy, speed, and resource utilization, the study aspires to provide valuable insights for practitioners seeking to implement robust safety gear detection systems. This comparative analysis, rooted in real-world scenarios and diverse datasets, endeavours to contribute to the ongoing discourse on leveraging deep learning for workplace safety, offering recommendations for optimal model selection based on specific application requirements and laying the groundwork for future advancements in the field.

* 1. **BACKGROUND AND MOTIVATION**

Industrial safety is a paramount concern, with various regulations and guidelines in place to protect workers. However, manual monitoring of compliance is labor-intensive and prone to human error, leading to gaps in safety enforcement. In recent years, technological advancements, especially in artificial intelligence and deep learning, have opened new avenues for enhancing workplace safety.

This study proposes the implementation of a deep learning-based system for the detection of safety outfit non-compliance in industrial settings. Such a system can autonomously monitor compliance, thereby reducing the reliance on manual checks. By utilizing facial recognition, the system can identify individuals and record specific instances of non-compliance. This capability is not only essential for immediate corrective action but also for long-term safety compliance analysis and policy development.

The significance of this study lies in its potential to revolutionize safety protocol enforcement in industrial environments. By ensuring higher compliance rates, the system can significantly reduce the risk of accidents and injuries, leading to a safer workplace.

* 1. **PROBLEM STATEMENT**

In industrial environments, safety is of utmost importance. A critical aspect of ensuring worker safety is the mandatory wearing of specific safety outfits, including helmets, vest and mask, in designated areas. However, a significant problem that industries face is the non-compliance of employees with these safety regulations. This non-compliance leads to an increased risk of accidents and injuries, posing a serious threat to individual safety and overall workplace security.

Existing systems for safety helmet, mask, and vest detection often leverage deep learning technologies to enhance workplace safety and compliance. These systems typically employ pre-trained convolutional neural network (CNN) architectures such as Faster R-CNN, YOLO, or SSD, which have demonstrated effectiveness in object detection tasks. The training process involves using annotated datasets containing images of individuals wearing safety equipment and background images without the specified items. safety helmets, masks, and vests in images or video streams. Deployment of these systems can be integrated into existing surveillance infrastructure, enabling real-time monitoring and alerting when safety violations are detected. Continuous improvement through model retraining ensures adaptability to changing environments and evolving safety standards

The challenge is twofold. Firstly, there is the difficulty in continuously monitoring and enforcing the compliance of safety regulations across large and complex industrial zones. Manual monitoring is not only labor-intensive but also prone to human error, making it an unreliable method for ensuring safety compliance. Secondly, there is an absence of a systematic approach to record and analyze instances of non-compliance, which is essential for understanding patterns, identifying frequent violators, and developing targeted safety interventions.

* 1. **OBJECTIVES OF THE STUDY**

The primary objective of this research is to design, develop, and validate a deep learning-based system for detecting safety outfit in industrial zones. This system aims to enhance workplace safety by ensuring adherence to safety gear regulations, thereby reducing the risk of accidents and promoting a culture of safety. The specific goals of this research are:

* Develop a Deep Learning Model: To create an advanced deep learning model capable of accurately detecting and classifying safety outfits, such as helmets, vest and mask. This model should function effectively under various industrial conditions, including different lighting, angles, and distances.
* Implement Reliable Facial Recognition: To integrate a robust facial recognition system that can identify individuals in diverse and challenging industrial environments, ensuring accurate association of safety outfit compliance with specific employees.
* Ensure Data Privacy and Ethical Compliance: To address privacy and ethical concerns by developing a system that complies with data protection laws. This involves secure handling of personal data and ensuring transparency in data usage.
* Achieve Real-Time Processing and High Reliability: To ensure the system processes data in real time and maintains high reliability and accuracy in detecting non-compliance, which is crucial for immediate corrective action.
* Integrate with Existing Industrial Systems: To design the system for easy integration with existing safety protocols and IT infrastructure in industries, ensuring compatibility and minimal disruption.

**CHAPTER 2**

**LITERATURE REVIEW**

1. **REVIEW OF EXISTING LITERATURE ON OBJECT DETECTION**

* **[1]** Title: Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches

Authors: Arjun Thirunavukarasu

Description:

The existing deep learning-based Personal Protective Equipment (PPE) detectors can only detect limited types of PPE and their performance needs to be improved, particularly for their deployment on real construction sites. This paper introduces an approach to train and evaluate eight deep learning detectors, for real application purposes, based on You Only Look Once (YOLO) architectures for six classes, including helmets with four colours, person, and vest. Meanwhile, a dedicated high-quality dataset, CHV, consisting of 1330 images, is constructed by considering real construction site background, different gestures, varied angles and distances, and multi PPE classes. The comparison result among the eight models shows that YOLO v5x has the best mAP (86.55%), and YOLO v5s has the fastest speed (52 FPS) on GPU. The detection accuracy of helmet classes on blurred faces decreases by 7%, while there is no effect on other person and vest classes. And the proposed detectors trained on the CHV dataset have a superior performance compared to other deep learning approaches on the same datasets. The novel multiclass CHV dataset is open for public use.

### [2] Title Deep learning for site safety: Real-time detection of personal protective equipment

# Authors: Nipun D. Nath

# Description:

A major obstacle faced when developing convolutional neural networks (CNNs) for medical imaging is the acquisition of training labels: most current approaches rely on manually prescribed labels from physicians, which are time consuming and labour intensive to attain. Clinical biomarkers, often measured alongside medical images and used in diagnostic workup, may provide a rich set of data that can be collected retrospectively and utilized to train diagnostic models. In this work, we focused on the blood serum biomarkers BNP and BNPP, indicative of acute heart failure (HF) and cardiogenic pulmonary edema, paired with the chest X-ray imaging modality. We investigated the potential for inferring BNP and BNPP from chest radiographs. For this purpose, a CNN was trained using \textcolor{black}{27748} radiographs to automatically infer BNP and BNPP, and achieved strong performance (AUC=0.90, \textcolor{black}{SEN=0.88}, \textcolor{black}{SPEC=0.81}, r=0.79). Since radiographic features of pulmonary edema may not be visible on low resolution images, we also assessed the impact of image resolution on model learning and performance, comparing CNNs trained at five image sizes (64×64 to 1024×1024). With comparable AUC values obtained at different resolutions, our experiments using three activation mapping techniques (saliency, Grad-CAM, XRAI) revealed considerable in-lung attention growth with increased resolution. The highest resolution models focus attention on the lungs, necessary for radiographic diagnosis of pulmonary edema. Our results emphasize the need to utilize radiographs of near-native resolution for optimal CNN performance, not fully captured by summary metrics like AUC.

# [3] Title: Learning Safety Equipment Detection using Virtual Worlds

# Authors: Marco di Benedetto

Description:

Plant diseases are a major threat to farmers, consumers, environment and the global economy. In India alone, 35% of field crops are lost to pathogens and pests causing losses to farmers. Indiscriminate use of pesticides is also a serious health concern as many are toxic and biomagnified. These adverse effects can be avoided by early disease detection, crop surveillance and targeted treatments. Most diseases are diagnosed by agricultural experts by examining external symptoms. However, farmers have limited access to experts. This project is the first integrated and collaborative platform for automated disease diagnosis, tracking and forecasting. Farmers can instantly and accurately identify diseases and get solutions with a mobile app by photographing affected plant parts. Real-time diagnosis is enabled using the latest Artificial Intelligence (AI) algorithms for Cloud-based image processing.

* [4] Title: AI-powered banana diseases and pest detection

# Authors: [Michael Gomez Selvaraj](https://link.springer.com/article/10.1186/s13007-019-0475-z#auth-Michael_Gomez-Selvaraj-Aff1)

# Description:

Nowadays, the possibilities offered by state-of-the-art deep neural networks allow the creation of systems capable of recognizing and indexing visual content with very high accuracy. Performance of these systems relies on the availability of high quality training sets, containing a large number of examples (e.g. million), in addition to the machine learning tools themselves. For several applications, very good training sets can be obtained, for example, crawling (noisily) annotated images from the internet, or by analysing user interaction (e.g.: on social networks). However, there are several applications for which high quality training sets are not easy to be obtained/created. Consider, as an example, a security scenario where one wants to automatically detect rarely occurring threatening events. In this respect, recently, researchers investigated the possibility of using a visual virtual environment, capable of artificially generating controllable and photo-realistic contents, to create training sets for applications with little available training images. We explored this idea to generate synthetic photo-realistic training sets to train classifiers to recognize the proper use of individual safety equipment (e.g.: worker protection helmets, high-visibility vests, ear protection devices) during risky human activities. Then, the performed domain adaptation to real images by using a very small image data set of real-world photographs. They show that training with the generated synthetic training set and using the domain adaptation step is an effective solution to address applications for which no training sets exist.

# [5] Title: A Smart System for Personal Protective Equipment Detection in Industrial Environments Based on Deep Learning at the Edge

# Authors: Gionatan Gall

# Description:

Real-time object detection is currently used to automate various tasks in industrial environments. One of the most important tasks is to improve the safety of workers by monitoring the correct use of Personal Protective Equipment (PPE) in dangerous areas. In this context, usually, a monitoring system analyses the stream of videos from surveillance cameras to assess PPE usage in real time. When a worker not wearing the appropriate PPE is detected, an acoustic or visual alarm is triggered automatically to raise attention and awareness. The solutions proposed so far are mostly cloud-based systems: images from the site are continuously offloaded to the cloud for analysis. This centralized architecture requires significant network bandwidth to transmit the video feeds through an internet connection that must be reliable, as a network outage would disrupt the service. In this work, we propose a system for real-time PPE detection based on video streaming analysis and Deep Neural Network (DNN). We adopt the edge computing model in which the application for image analysis and classification is deployed on an embedded system installed in proximity of the camera and directly connected to it. The system does not require continuous image transmission towards a cloud system, thus ensuring bandwidth efficiency, reliability, and workers’ privacy. A prototype of the proposed system is developed exploiting a low-cost commercial embedded system, i.e. a Raspberry PI, equipped with an Intel Neural Compute Stick 2. We tested the system with five different pre-trained convolutional neural networks (CNNs), fine-tuned to detect different PPEs, namely helmets, vests, and gloves. In our experimental evaluation, we first compared the five CNNs in terms of classification performance and inference latency. Then, we deployed each CNN on the real system and evaluated the system’s throughput regarding the number of video frames analysed each second.

* 1. **OVERVIEW OF ARTIFICIAL INTELLIGENCE AND DEEP LEARNING**

**2.2.1 ARTIFICIAL INTELLIGENCE**

Artificial intelligence (AI) is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers "smart". As machines become increasingly capable, mental facilities once thought to require intelligence are removed from the definition. AI is an area of computer sciences that emphasizes the creation of intelligent machines that work and reacts like humans. Some of the activities computers with artificial intelligence are designed for include: Face recognition, Learning, Planning, Decision making etc.,

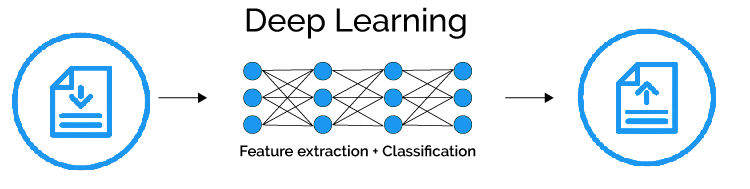
Artificial intelligence is the use of computer science programming to imitate human thought and action by analysing data and surroundings, solving or anticipating problems and learning or self-teaching to adapt to a variety of tasks.

* + 1. **DEEP LEARNING**

A subset of machine learning techniques called "deep learning" is based on representation learning in artificial neural networks. The use of multiple layers in the network is indicated by the adjective "deep" in deep learning. The employed techniques can be unsupervised, semi-supervised, or supervised.

In a variety of fields, including computer vision, speech recognition, natural language processing, machine translation, bioinformatics, drug design, medical image analysis, climate science, material inspection, and board game programming, deep-learning architectures such as deep neural networks, deep belief networks, deep reinforcement learning, recurrent neural networks, convolutional neural networks, and transformers have produced results on par with, if not better than, human expert performance.

The information processing and distributed communication nodes found in biological systems served as the model for artificial neural networks, or ANNs. ANNs are not like biological brains in a number of ways. In particular, the biological brain of the majority of living things is dynamic (plastic) and analog, whereas artificial neural networks typically exhibit static and symbolic behavior. The below block diagram explains the working of Deep Learning algorithm:



**Features of Deep Learning:**

* Deep learning systems can perform feature extraction automatically, meaning they don't require supervision to add new features.
* Deep learning systems can process both structured and unstructured data.
* Accuracy
* Deep learning systems can analyse large amounts of data and uncover complex patterns in images, text and audio and can derive insights that it might not have been trained on.

**Classification of Deep Learning**

At a broad level, Deep learning can be classified into three types:

1. Supervised learning
2. Unsupervised learning
3. Partially Supervised (semi-supervised)

### **1) Supervised Learning**

Supervised learning is a type of machine learning where the algorithm is trained on a labeled dataset, which means that the input data is paired with the corresponding correct output. In other words, the algorithm is provided with input-output pairs, and the goal is to learn a mapping function from the input to the output.

In the context of deep learning, which is a subfield of machine learning, supervised learning involves using neural networks to learn complex mappings from inputs to outputs. These neural networks are composed of layers of interconnected nodes (neurons) that process the input data and produce an output. During the training process, the network adjusts its internal parameters (weights and biases) based on the difference between its predictions and the true outputs in the labeled training data.

The training process typically involves an optimization algorithm (e.g., gradient descent) that minimizes a loss function, which measures the difference between the predicted outputs and the true outputs. The goal is to find the optimal set of parameters that minimizes this loss, allowing the model to generalize well to new, unseen data.

Supervised learning in deep learning is widely used in various applications, such as image recognition, natural language processing, speech recognition, and many others. It is called "supervised" because the process involves a "teacher" (the labeled data) guiding the learning algorithm to make accurate predictions.

### **2) Unsupervised Learning**

Unsupervised learning is a type of machine learning where the algorithm is given input data without explicit instructions on what to do with it. Unlike supervised learning, there are no labeled outputs provided during training. The goal of unsupervised learning is to find patterns, relationships, or structures in the data without explicit guidance. In the context of deep learning, unsupervised learning encompasses various approaches, and common types are:

* **Clustering**
* **Dimensionality Reduction**
* **Generative Models**
  + 1. **GUI (Graphical User Interface)**

A Graphical User Interface (GUI) in Python refers to a visual way of interacting with a computer program. Instead of relying on text-based commands, GUIs utilize graphical elements such as windows, buttons, menus, and other visual components to enable user interaction. Python provides several libraries for creating GUI applications, with Tkinter being the default and widely used option. GUIs enhance user experience by offering an intuitive and visually appealing environment for interacting with software. Developers can design interfaces that make complex functionalities accessible to users through mouse clicks, keyboard input, or touch interactions, catering to a broad range of applications from desktop software to mobile apps. The design and implementation of GUIs in Python involve the use of specific libraries that simplify the creation and management of graphical elements, allowing developers to build interactive and user-friendly applications.

* + 1. **CNN**

CNN typically stands for Convolutional Neural Network, which is a class of deep neural networks primarily designed for image recognition and processing. It has become a fundamental technology in various computer vision tasks, including image and video analysis, object detection, and classification.

**Convolutional Layers:** CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input images. Convolutional layers use filters (also known as kernels) to convolve over the input image, capturing local patterns and features.

**Pooling Layers:** After convolution, pooling layers are often employed to downsample the spatial dimensions of the data. Max pooling, for example, retains the maximum value within a certain region, reducing the computational load and helping the network focus on the most relevant features.

**Activation Functions:** Non-linear activation functions, such as ReLU (Rectified Linear Unit), are applied to the output of convolutional layers to introduce non-linearity into the network and enable it to learn more complex patterns.

**Fully Connected Layers:** Toward the end of the network, fully connected layers are typically used for classification tasks. These layers connect every neuron to every neuron in the previous and subsequent layers, allowing the network to learn complex relationships between high-level features.

**Training and Backpropagation:** CNNs are trained using a process called backpropagation, where the network adjusts its parameters (weights and biases) based on the error between predicted and actual outputs. This is usually done through optimization algorithms like stochastic gradient descent.

**Pre-trained Models:** Due to the computational intensity of training deep networks, pre-trained models on large datasets (like ImageNet) are often used as a starting point. These models can be fine-tuned on smaller, task-specific datasets.

* 1. **SUMMARY OF KEY FINDINGS FROM LITERATURE REVIEW**

We mainly look at target identification algorithms and techniques for mask, vest, and helmet detection in this part.

The detection and recognition of personal protective equipment (PPE) in industrial settings have garnered significant attention in recent years due to their critical role in ensuring workplace safety. Various deep learning-based methods have been proposed to address this challenge, ranging from helmet and safety vest detection to more comprehensive approaches covering multiple types of PPE. For instance, Wang et al. (2020) introduced a safety helmet and protective clothing detection method based on an improved YOLOv3 model [1]. Similarly, Nath and Behzadan (2020) presented a deep learning approach for real-time detection of PPE to maintain safety compliance on construction sites [4]. These studies highlight the importance of leveraging advanced technologies to enhance safety protocols and reduce the risk of accidents in industrial environments. Furthermore, recent research by Chen et al. (2023) focused on real-time detection algorithms for helmets and reflective vests based on an improved YOLOv5 model, underscoring ongoing efforts to refine detection techniques for specific types of PPE [25].

In addition to PPE detection, researchers have explored vision-based monitoring systems to ensure site safety compliance and worker identification. Cheng et al. (2022) proposed a method based on worker re-identification and PPE classification for vision-based monitoring of site safety compliance [10]. Moreover, Li et al. (2022) introduced an intelligent vision-based method for worker identification in the context of the industrial Internet of Things (IoT), further enhancing safety measures through automated identification processes [11]. These studies demonstrate the interdisciplinary nature of safety research, combining computer vision, deep learning, and IoT technologies to create comprehensive safety management systems. By leveraging these advancements, industries can improve safety standards, reduce accidents, and create safer working environments for employees [10].

Continuing the exploration of deep learning applications in safety management, recent studies have extended beyond traditional PPE detection to address more nuanced challenges in occupational safety. Antwi-Afari et al. (2022) delved into the automation of recognition and classification of awkward working postures in construction using wearable insole sensor data, showcasing the potential of wearable technology combined with deep learning for comprehensive safety monitoring [19]. Moreover, advancements in gesture recognition have been investigated for real-time recognition of dynamic finger gestures, as demonstrated by Lee and Bae (2020) using a data glove, indicating the potential for gesture-based safety protocols and interactions in industrial settings [21]. These developments underscore the evolution of safety management from passive monitoring to proactive intervention, where real-time analysis and feedback contribute to a dynamic safety ecosystem. By harnessing the power of deep learning and wearable technology, industries can not only detect non-compliance but also proactively mitigate risks and promote safer work practices, ultimately fostering a culture of safety and well-being in the workplace.

Further advancements in safety management through deep learning techniques include the exploration of intelligent surveillance systems for industrial work environments. Ahn et al. (2023) introduced SafeFac, a video-based smart safety monitoring system designed to prevent industrial accidents through real-time analysis and intervention [12]. This represents a shift towards proactive safety measures, where predictive analytics and AI-driven insights enable preemptive actions to mitigate potential hazards. Moreover, Li et al. (2019) investigated standardized use inspection of workers' personal protective equipment, highlighting the role of deep learning in ensuring adherence to safety protocols and regulations [13]. By integrating AI-driven inspection systems into regular safety audits, industries can enhance compliance monitoring and quickly address any deviations from safety standards. Additionally, the application of deep learning extends beyond traditional safety equipment detection, as seen in Rosero-Montalvo et al. (2018), who developed a sign language recognition system based on an intelligent glove using machine learning techniques [22]. This innovative approach not only enhances communication accessibility but also contributes to safety by enabling effective communication in noisy or hazardous environments. These multifaceted applications of deep learning underscore its transformative potential in revolutionizing safety management practices across diverse industrial sectors, ultimately leading to safer and more productive work environments.

Research in construction site safety and personal protective equipment (PPE) detection has witnessed significant advancements through deep learning techniques. Wang et al. (2021) introduced a fast PPE detection system tailored for real construction environments, emphasizing the critical need for swift and accurate detection to bolster safety measures [2]. Concurrently, Bao et al. (2021) delved into the realm of offshore platform safety, leveraging deep learning for damage detection [3]. Nath et al. (2020) and Ahmed et al. (2023) focused on real-time PPE detection, highlighting the transformative potential of deep learning in enhancing safety protocols [5]. Additionally, Hayat and Morgado-Dias (2022) pioneered an automatic safety helmet detection system, showcasing the efficacy of deep learning in ensuring compliance with safety standards [15]. Long et al. (2019) contributed to this field by developing a deep learning-based system for detecting safety helmet wearing, further reinforcing the importance of technology-driven safety solutions [9]. Vukicevic et al. (2022) extended the application of deep learning to ensure generic compliance of industrial PPE, reflecting the versatility of these techniques in addressing safety challenges across diverse industrial settings [7]. Moreover, Nain et al. (2021) demonstrated the utility of computer vision and deep learning in creating a comprehensive safety and compliance management system [14]. Li et al. (2019) explored deep learning approaches for pedestrian detection in challenging weather conditions [17], while Rebekah et al. (2020) focused on dress code surveillance using similar techniques [20]. Seong et al. (2018) conducted a comparative study on machine learning classification for safety vest detection, enriching the understanding of deep learning's applicability in construction safety [23]. Lastly, Chen et al. (2023) proposed a real-time detection algorithm for helmets and reflective vests, further advancing the practical implementation of deep learning in safety-critical scenarios [24]. Together, these studies underscore the transformative impact of deep learning in revolutionizing construction site safety and PPE detection practices.

In summary, the proliferation of deep learning techniques in safety management has ushered in a new era of proactive safety protocols and comprehensive risk mitigation strategies. From the detection of personal protective equipment to intelligent surveillance systems and gesture recognition technologies, researchers have leveraged AI- driven insights to enhance safety standards and promote a culture of well-being in industrial settings. By integrating deep learning into various aspects of safety management, industries can not only detect non-compliance but also predict and prevent potential hazards, ultimately fostering safer and more productive work environments for employees.

**CHAPTER 3**

**METHODOLOGY**

* Data collection
* Data Pre processing
* Model implementation
* Loading the trained model
* Prediction

**Data Collection:**

* In this module, a comprehensive and diverse dataset is compiled, featuring individuals from various occupational settings wearing safety helmets, masks, and vests.
* The dataset should be well-annotated and represent different lighting conditions, perspectives, and environmental contexts to ensure the robustness of the model.

**Data Pre-processing:**

* The collected dataset undergoes preprocessing to enhance its quality and suitability for training.
* This involves tasks such as resizing images, normalizing pixel values, and applying data augmentation techniques to introduce variations.
* The goal is to prepare a refined dataset that optimally trains the model while ensuring its ability to generalize across different scenarios.

**Model Implementation:**

* This module focuses on the actual implementation of the chosen deep learning model, whether it be Faster R-CNN or YOLOv5.
* The model architecture is configured, and training parameters are set. The code for the chosen algorithm is written or configured to adapt to the specific requirements of the safety gear detection task.

**Loading the trained model:**

* Once the model is trained successfully, it is saved and prepared for deployment.
* This module involves creating a mechanism to load the trained model efficiently during the operational phase. The loading process ensures that the model is ready to make predictions on new data.

**Prediction:**

* The final module involves utilizing the trained model to make predictions on live or new data captured by the webcam.
* The webcam feed is continuously analyzed, and the model predicts the presence and correct usage of safety equipment in real-time.
* This prediction module is crucial for the system to provide immediate feedback on safety compliance, contributing to the overall goal of enhancing safety measures in the monitored environment.
  1. **DESCRIPTION OF DATASET**

We use YOLOv5 as a detector to ascertain if workers are donning helmets, as shown in Figure 1. To sum up, the following contributions are made by this paper:

We gathered and annotated 3080 photos from the Internet. Examples in which the head and body are not wearing a helmet, vest, or mask are marked as "NOT SAFETY," but those that are wearing the proper protective gear are marked as "SAFETY."

We accomplish exceptional results by properly detecting if workers are wearing helmets by utilizing the most sophisticated target detection algorithm. Using different depths and widths, we trained and assessed the YOLOv5 model and compared its performance.

****

Fig. 1. The structure of the helmet and vest detector. Input a picture, pass the helmet and vest detector, output the detected helmet and vest and its position information, and output its confidence

* 1. **ALGORITHMS**

**FASTER R-CNN**

Faster R-CNN (Region-based Convolutional Neural Networks) is an advanced object detection algorithm that improves the speed and accuracy of object detection. The algorithm integrates region proposal generation and object detection into a single, unified network, making it faster than its predecessors. Here is a high-level overview of the Faster R-CNN algorithm:

* Input Image:
  + An image is provided as input to the network.
* Feature Extraction:
  + A Convolutional Neural Network (CNN) such as VGG16 or ResNet is used to extract feature maps from the input image.
* Region Proposal Network (RPN):
  + Anchor Boxes: The RPN generates anchor boxes of different scales and aspect ratios at each location of the feature map.
  + Sliding Window: A sliding window of size \(n \times n\) is used on the feature map to generate region proposals.
  + Classification and Regression: Binary Classification Layer: Determines whether an anchor box contains an object or not (objectness score).
  + Bounding Box Regression Layer: Refines the coordinates of the anchor boxes to fit the objects more accurately.
* Region of Interest (RoI) Pooling:
  + The proposals generated by the RPN are used to crop regions from the feature map.
  + These regions are resized to a fixed size using RoI Pooling to ensure compatibility with the fully connected layers.
* Object Detection:
  + The regions obtained from RoI Pooling are passed through fully connected layers for final object classification and bounding box regression.
  + Softmax Layer: Determines the class of the object within each proposal.
  + Bounding Box Regression Layer: Refines the coordinates of the bounding boxes.
* Output:
  + The final output consists of the class labels and the refined bounding boxes for the detected objects in the image.

**YOLOv5:**

YOLOv5 (You Only Look Once version 5) is a single-stage object detection algorithm known for its speed and accuracy. Unlike region-based methods like Faster R-CNN, YOLOv5 processes the entire image in one go, making it faster. Below is a high-level overview of the YOLOv5 algorithm:

* Input Image:
  + An image is provided as input to the network.
* Feature Extraction:
  + A Convolutional Neural Network (CNN) backbone (e.g., CSPDarknet) extracts feature maps from the input image.
* Feature Pyramid Network (FPN) and Path Aggregation Network (PAN):
  + These networks combine feature maps from different scales to help detect objects of various sizes.
* Detection Head:
  + For each scale, the detection head predicts:
    - Bounding boxes (coordinates)
    - Objectness scores (confidence that an object is present)
    - Class probabilities (which class the object belongs to)
* Non-Maximum Suppression (NMS):
  + Removes redundant bounding boxes based on the objectness score to produce the final set of detections.

YOLOv5 processes the entire image in a single forward pass, making it extremely fast. By using anchor boxes and a grid system, it can effectively detect multiple objects in real-time. The FPN and PAN further enhance its capability to detect objects of various sizes, making it one of the most efficient and accurate object detection algorithms available.

* 1. **PROPOSED METHODS**

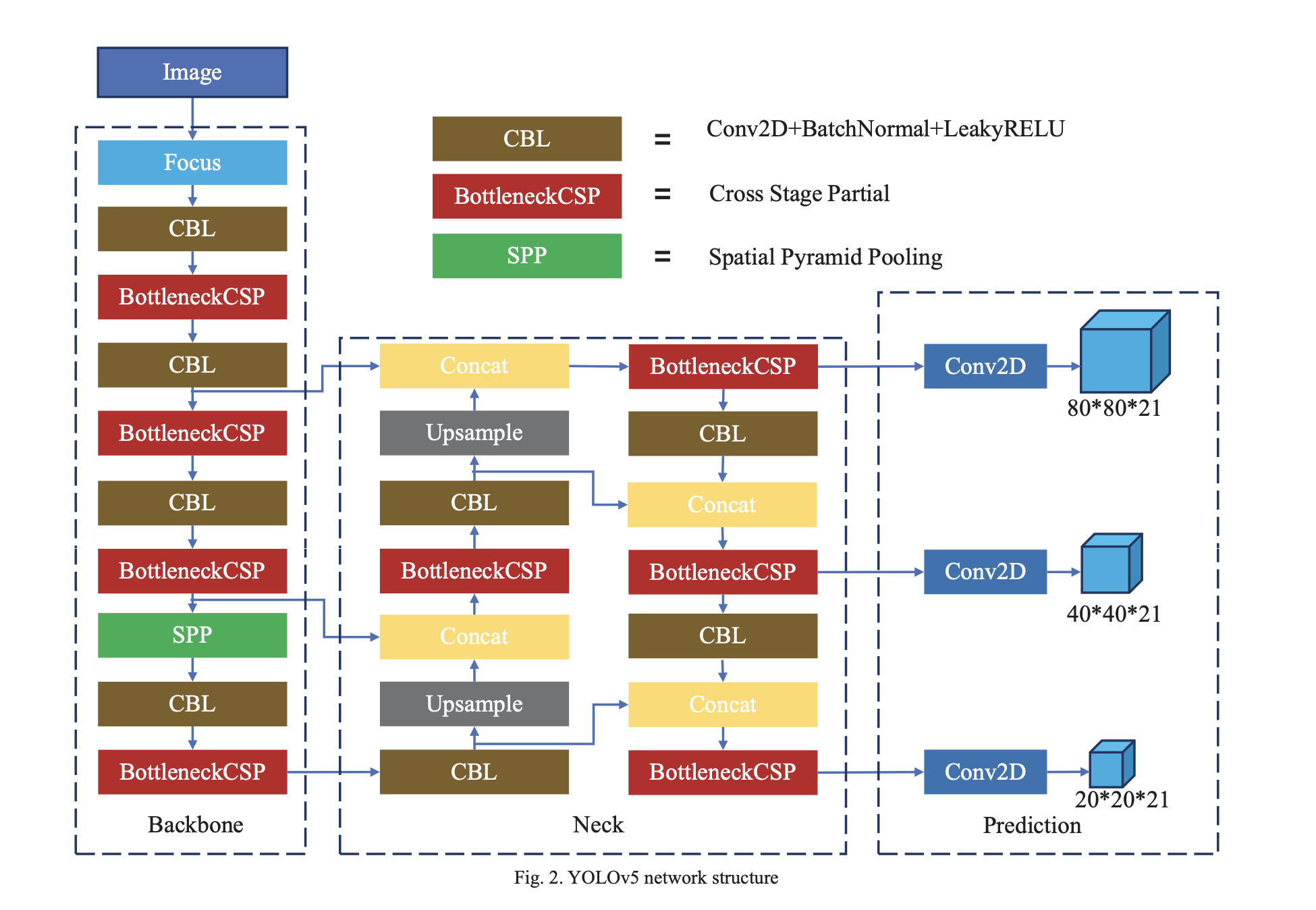
The purpose of the suggested system is to improve industrial safety compliance by using a webcam-based method for real-time monitoring. The technology scans the camera feed using advanced computer vision algorithms to find and recognize people wearing vital safety equipment including masks, vests, and helmets. This is accomplished by incorporating cutting-edge deep learning algorithms, like YOLOv5, which are well-known for their effectiveness and accuracy in object detection tasks. The technology offers immediate feedback on the use of safety equipment by continuously monitoring the live feed. This helps to improve safety protocols and ensures a proactive approach to maintaining a secure working environment.

The system's operational framework consists of multiple modules that are intended to optimize its functionality. First, during the data collecting stage, a large dataset of people in different work environments wearing protective gear is assembled. This dataset is carefully annotated and representation of various environmental circumstances is made to guarantee the robustness of the resulting model. The gathered dataset is then enhanced using techniques like scaling, normalization, and data augmentation during the data preprocessing stage, making it ideal for training and generalization in a variety of scenarios. The model implementation module then concentrates on defining the deep learning architecture of choice YOLOv5—and establishing training parameters specific to the goal of detecting safety gear.

The loading module sets the model up for deployment by designing a way to effectively load the trained model during the operational phase, after the model has been trained successfully. Lastly, real-time predictions are made using the trained model on live webcam data during the prediction phase. This allows for ongoing analysis of safety compliance and prompt feedback to improve safety protocols in the monitored area. The suggested system provides an all-encompassing solution for automating safety compliance monitoring through various methodical elements, ultimately leading to a safer working environment in industrial settings.

1. **YOLOv5 network:**

We provide an overview of the YOLOv5 network architecture in this part, emphasizing its main elements and features. The backbone, neck, and output are the three primary structural components of the YOLOv5 network. The input image is processed through the Focus structure, starting with the backbone, and ends with a sequence of operations that produce a 320x320x32 feature map. In this stage, the core convolutional module is the CBL module, which consists of Conv2D + BatchNormal + LeakyRELU. Furthermore, the BottleneckCSP module is essential to feature extraction since it minimizes the duplication of gradient information during convolutional neural network optimization, hence improving network performance. because of this module's flexibility, four other models— YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x—can be created by varying its width and depth.



1. **Classifier modification:**

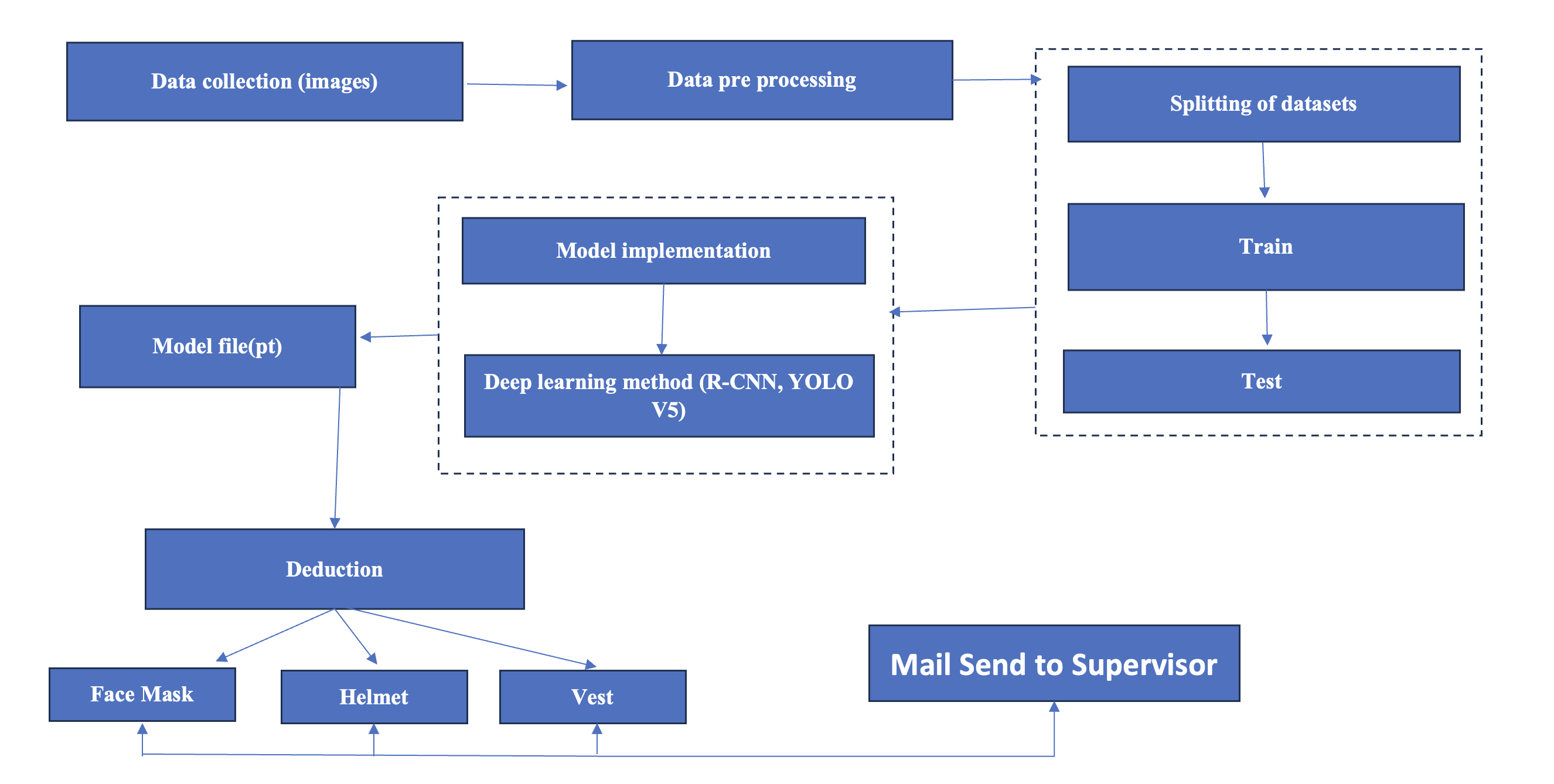
There are eighty object categories for the COCO dataset. Three × (5 + 80) = 255 is the computed dimension of the output tensor, where three stands for the three template boxes for each grid prediction. Coordinates (x, y, w, h) and confidence (confidence, c) are included in every prediction box. There are two sorts of objects in the context of helmet detection: those wearing safety helmets and those that don't. This means that the YOLOv5 classifier needs to be changed. When the output dimension is adjusted for the helmet detection scene, it becomes 3 × (5 + 2) = 21. With this change, we can improve detection speed and accuracy while lowering computational overhead and the amount of network parameters.

1. **Dataset introduction:**



We have 3080 images in our dataset that were gathered by crawlers. These photos, which are shown in Figure 3, feature a variety of settings and scales of helmets. We also added scenes from classrooms to increase the quantity of negative samples. We tagged objects and designated categories using labeling software, designating "Helmet" as positive samples for people wearing helmets and "Alarm" as negative examples for people not wearing helmets. Lastly, a training set and a test set were created from the labeled images in a 2:1 ratio.

1. **Proposed Architecture:**

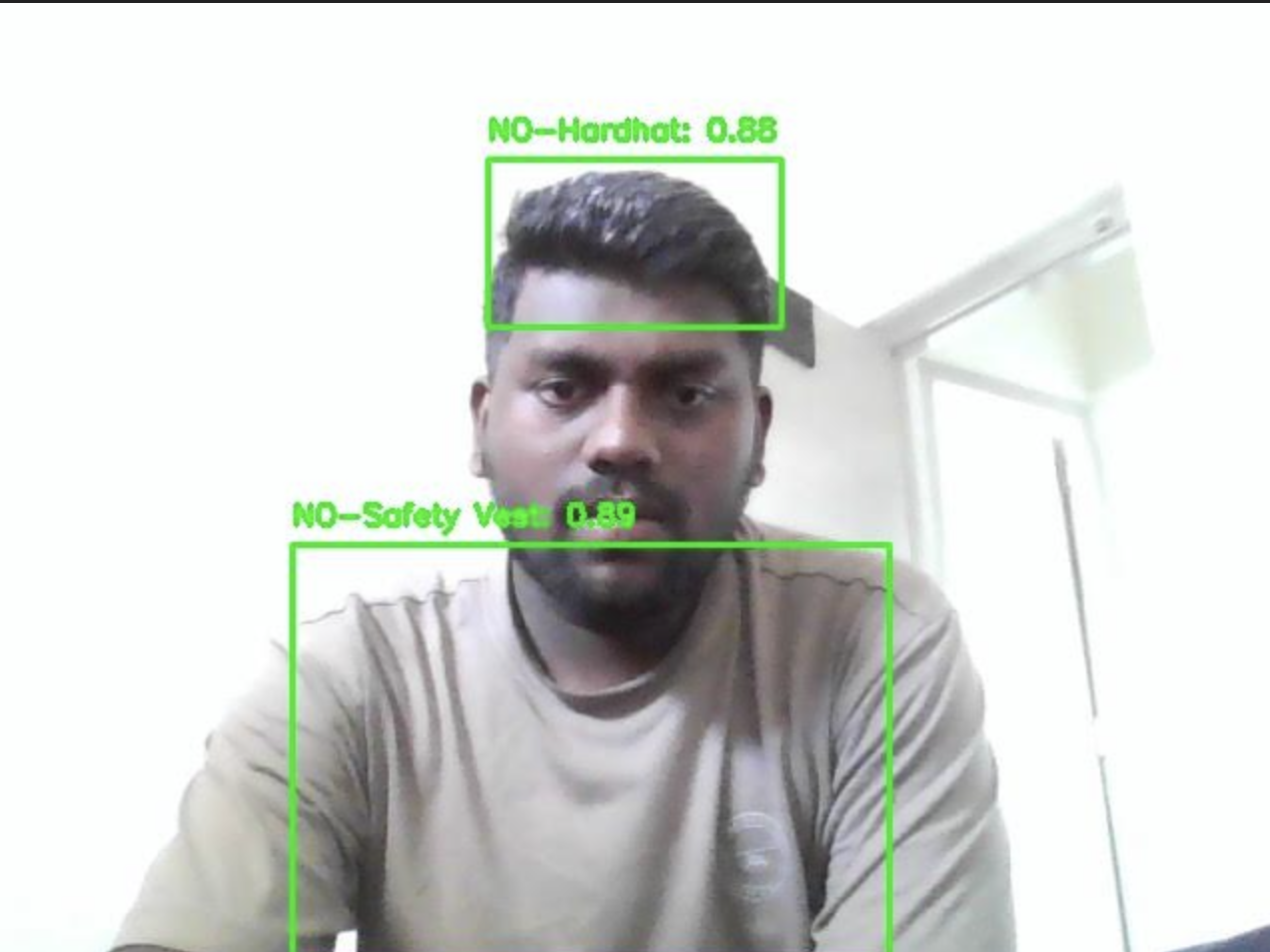
****

**CHAPTER 4**

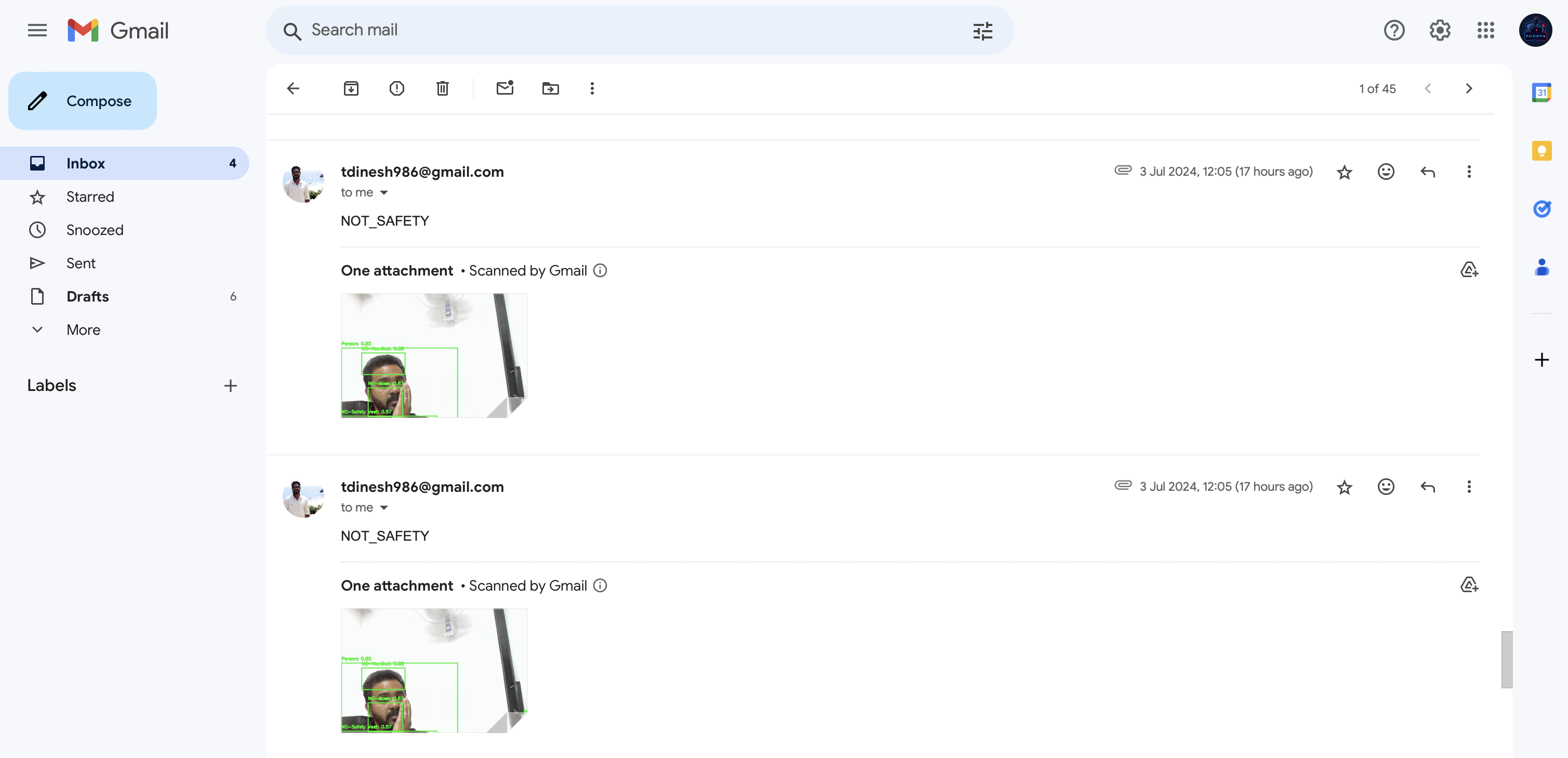
**RESULTS**

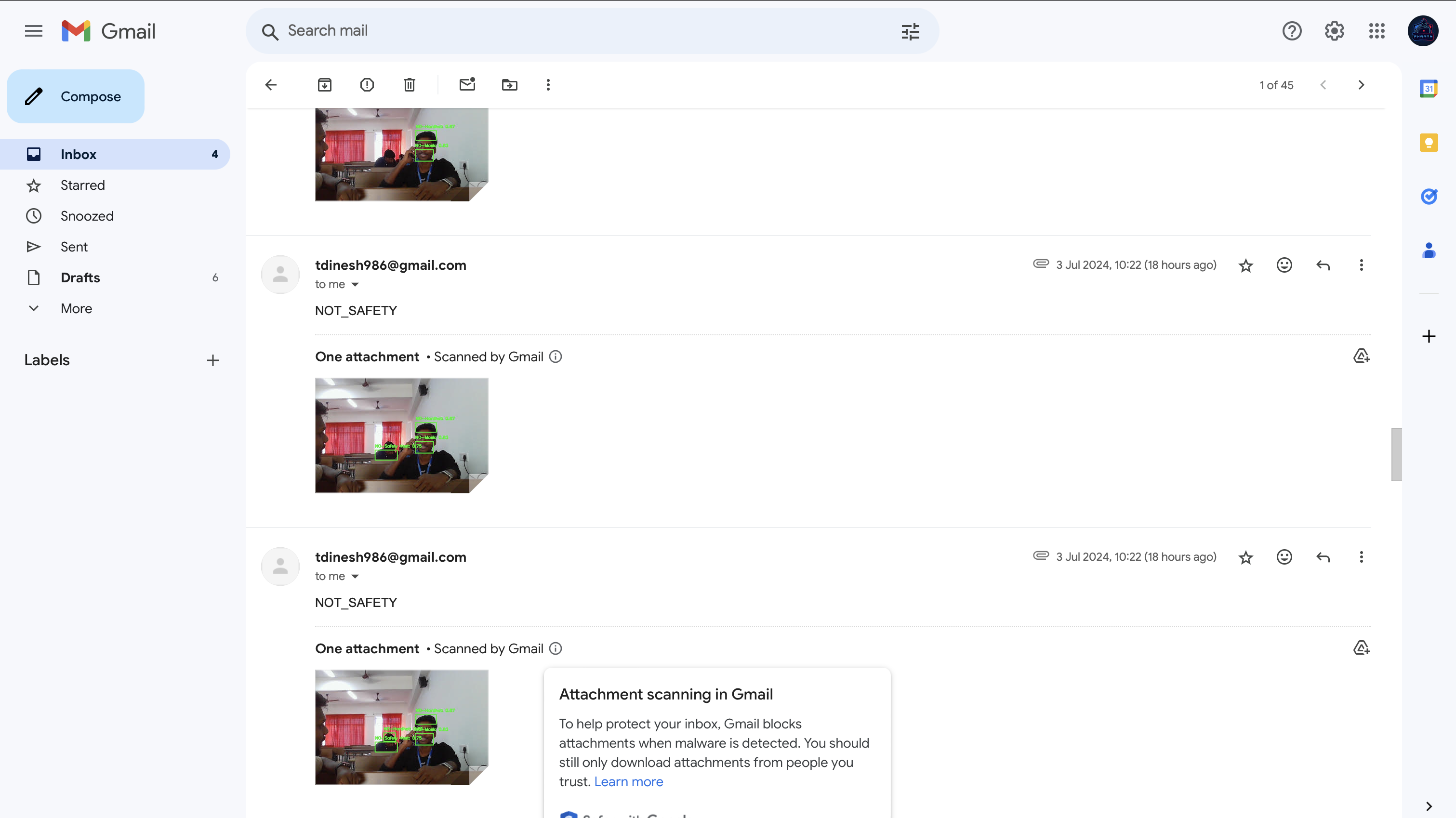
* 1. **PRESENTATION OF EXPERIMENTAL RESULTS**

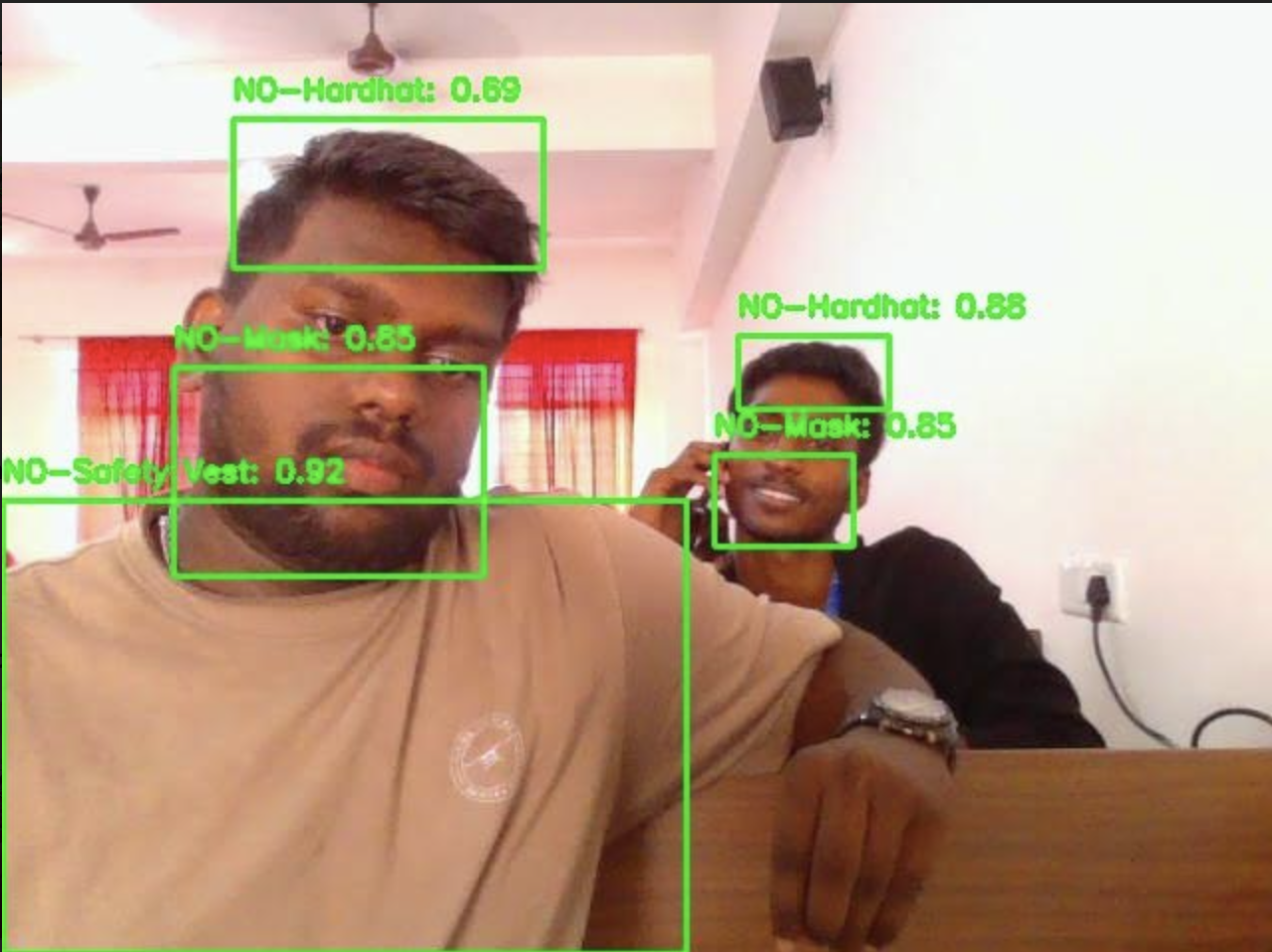
As a result, the project is able to detect people those who are not wearing the safety outfit.

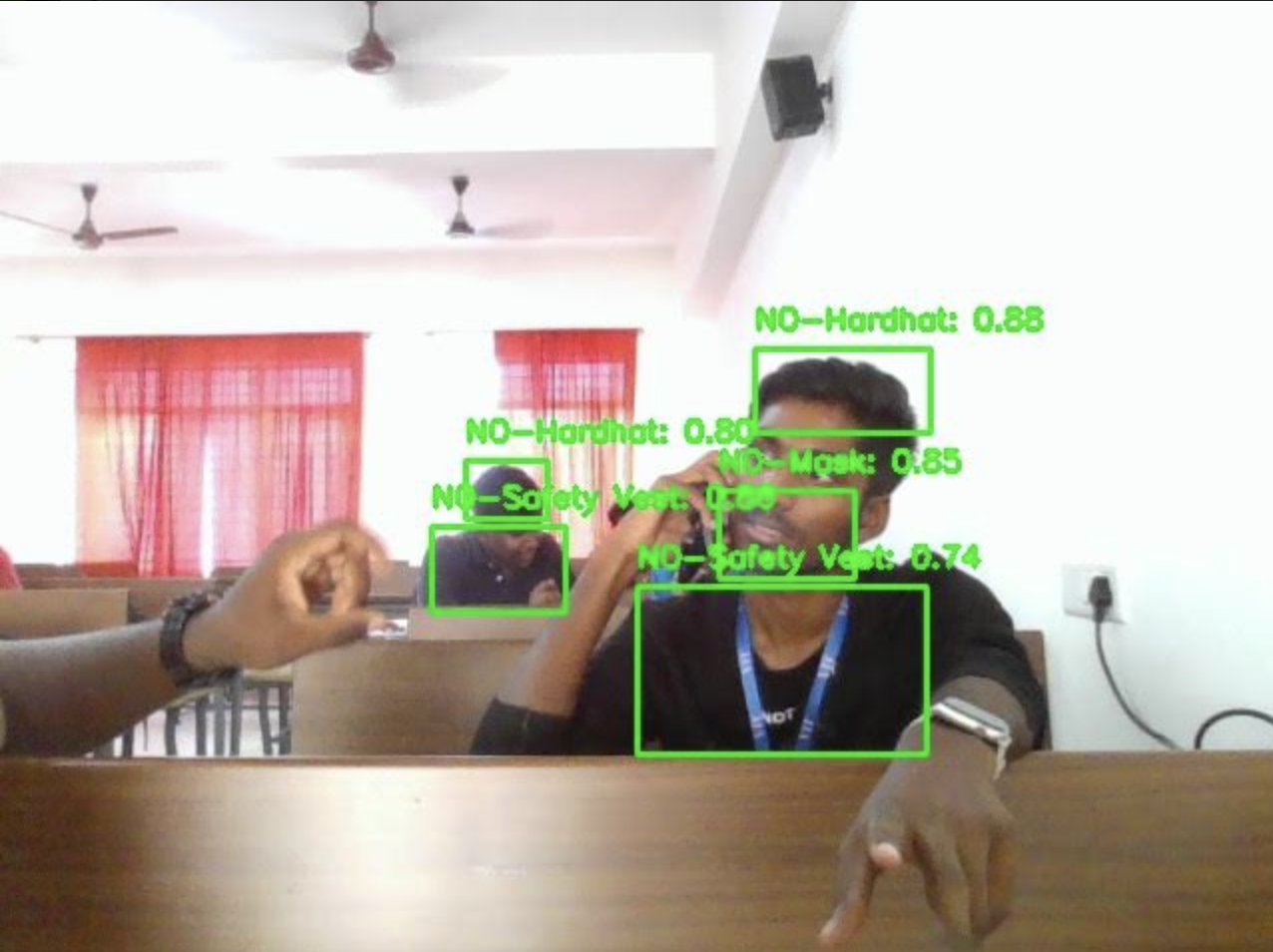


Those who are not wearing the safety outfit, their Image, the safety outfit which they were not wearing with time will be mailed to the necessary person









**CHAPTER 5**

**DISCUSSIONS**

* 1. **ANALYSIS AND INTERPRETATION OF RESULTS**

This part includes a description of the experiment, training outcomes acquired without pre-training weights, and a comparison of the four YOLOv5 models' performances. Lastly, we present the transfer learning-achieved experimental outcomes.

1. **Experimental details:**

Every training set image in our experiment was arbitrarily cropped to 640 by 640 pixels. Additionally, a variety of data augmentation techniques were used, including random flipping, cutting, mixup, random erase, geometric distortion, lighting distortion, and image occlusion. In order to facilitate network evaluation during the testing phase, the images were scaled to 640x640. Using the PyTorch framework and the Ubuntu 18.04 operating system, our experiment was carried out on a single NVIDIA GTX1660 GPU for training and testing. Our network was optimized using the SGD optimizer, and Table I shows several experimental hyperparameters.

**TABLE I SOME EXPERIMENTAL HYPERPARAMETER**S

|  |  |
| --- | --- |
| Parameter | value |
| Learning rate | 0.01 |
| Learning rate decay | 0.999 |
| Learning rate decay step | 1 |
| Weight rate decay | 5e-4 |
| Momentum | 0.937 |
| Batch size | 16 |
| Number of iterations | 100 |

1. **Experimental results:**

For helmet training, we employ YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x models. Additionally, Fig. 4 displays the evaluation findings following each training round.

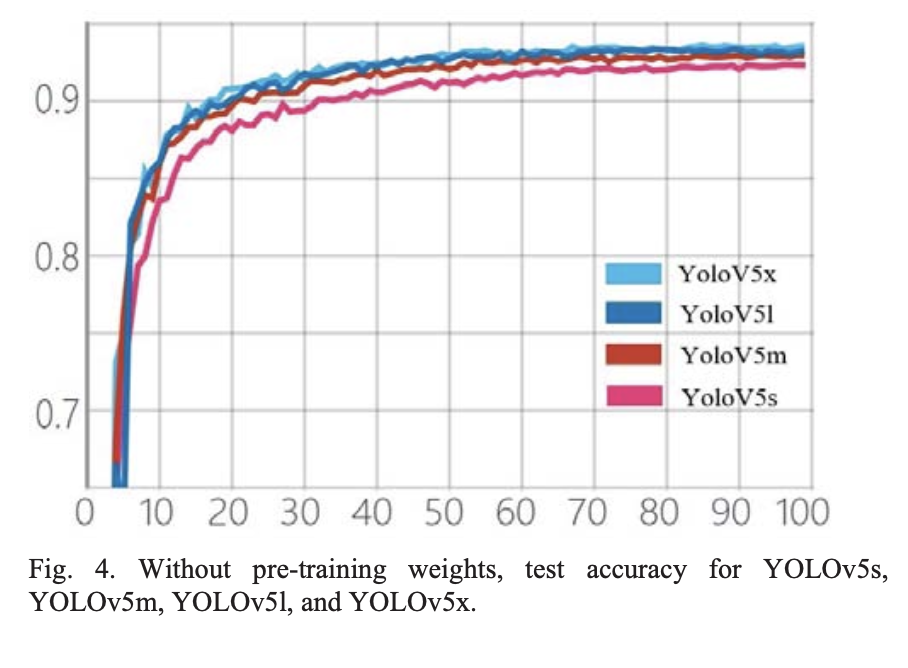


Fig. 4. Without pre-training weights, test accuracy for YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x.

All four models show significant increases in mean Average Precision (mAP) during the first training phase. Of the models, YOLOv5x, which has the greatest parameters, shows the fastest improvement. The mAP values of the four models progressively converge over time. We analyze each model's optimal weight and show the findings in Table II to compare the model flaws quantitatively. It's interesting to see that YOLOv5m performs 0.8% better than YOLOv5s. But the benefits become smaller as the number of parameters rises, as YOLOv5x only beats YOLOv5l by 0.1% even with much more parameters and twice as slow inference time.

**TABLE II. THE PERFORMANCE OF DIFFERENT MODELS WITHOUT PRE- TRAINING WEIGHTS.**

|  |  |  |
| --- | --- | --- |
| Model | mAP | FPS |
| YOLOv5s | 92.3 | 110 |
| YOLOv5m | 93.1 | 64 |
| YOLOv5l | 93.5 | 37 |
| YOLOv5x | 93.6 | 21 |

We took pictures from the test set and used them to visualize the findings of our investigations. YOLOv5 exhibits efficient detection skills in a range of settings, as illustrated in Figure 5. YOLOv5 demonstrates precise detection in image (a) with multicolored helmets and image (b) with a grayscale image. This is made possible by its convolutional neural network approach, which reduces the requirement for manual feature extraction and improves generalization capacity. Furthermore, as can be shown in photos (d) and (c), which show several targets, YOLOv5 successfully identifies the objects, even in situations where there is mutual occlusion. In image (d), for example, the model correctly ascertains whether or not people are wearing helmets. Furthermore, YOLOv5 is able to identify people who are not wearing helmets in photos (e) and (f). Notably, YOLOv5 still detects the lack of a helmet in image (f), even though the subject is donning a hat.



**TABLE III. THE PERFORMANCE OF DIFFERENT MODELS WITH PRE- TRAINING WEIGHTS.**

|  |  |  |
| --- | --- | --- |
| Model | mAP | Improved |
| YOLOv5s | 93.6 | 1.3 |
| YOLOv5m | 94.3 | 0.8 |
| YOLOv5l | 94.4 | 0.9 |
| YOLOv5x | 94.7 | 0.9 |

To improve the network's generalization ability, deep learning frequently needs a large amount of data, yet method to improve the Model's Generalization Capability. We choose and assess the four models with the best weights. Table III presents the evaluation results. It can be seen that after using and training The Weights, Map increases with the increase of the number of parameters. When pre- training weights are used as opposed to not using them, the outcome is improved.

* 1. **ADVANTAGES**
* Utilizing the webcam feed enables continuous, real-time monitoring of safety compliance
* The integration of advanced deep learning algorithms automates the surveillance process, eliminating the need for manual inspection.
* The system's reliance on webcam technology enhances scalability and adaptability, making it easily deployable across various settings and industries.
* The flexibility of integrating different deep learning algorithms allows customization to meet specific requirements, ensuring the system's effectiveness in diverse occupational scenarios.
  1. **LIMITATIONS AND CHALLENGES ENCOUNTERED**
* Deep learning models, especially complex ones used for object detection, often require significant computational resources during both training and inference.
* Addressing data bias is crucial for ensuring the model's effectiveness across different contexts.
* Annotating a large dataset with bounding boxes for helmets, masks, and vests can be time-consuming and expensive.

**CHAPTER 6**

**CONCLUSION**

In summary, this research delves into the realm of computer vision and deep learning, particularly focusing on the real-time detection of safety helmets, masks, and vests in industrial settings—an essential aspect for ensuring workplace safety. Through a comparative analysis of Faster YOLOv5, the study sheds light on their performance characteristics, considering metrics such as precision, detection accuracy, speed, and resource utilization. By evaluating these models across real-world scenarios and diverse datasets, the research attains practical relevance, making its findings applicable to the intricate and varied environments of industrial workplaces. Ultimately, by bridging theory with practical implementation, this study contributes significantly to the ongoing dialogue surrounding the utilization of deep learning techniques for enhancing workplace safety.

* 1. **RECOMMENDATIONS FOR FUTURE WORK**

Additionally, improvements in anomaly detection algorithms may make it possible to spot minute departures from safety procedures, strengthening preventative risk reduction tactics. Investigating cutting-edge architectures like graph neural networks and attention processes may provide more efficient means of capturing temporal and spatial relationships in safety monitoring activities. Furthermore, studies could concentrate on creating energy-efficient and lightweight models that can be deployed on edge devices to enable real-time safety monitoring in contexts with limited resources. To ensure consistency and dependability across various implementations, industry, academic institutions, and regulatory agencies could work together to develop standardized standards and processes for assessing safety gear detection systems. Ultimately, there is a great deal of promise for improving occupational safety procedures and creating safer work environments if advancements in the fields of computer vision and deep learning continue.

**CHAPTER 7**

**REFERENCES**

1. X. Wang, D. Niu, P. Luo, C. Zhu, L. Ding and K. Huang, "A Safety Helmet and Protective Clothing Detection Method based on Improved-Yolo V 3," 2020 Chinese Automation Congress (CAC), Shanghai, China, 2020, pp. 5437-5441, doi: 10.1109/CAC51589.2020.9327187.
2. Wang Z, Wu Y, Yang L, Thirunavukarasu A, Evison C, Zhao Y. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. Sensors. 2021; 21(10):3478. <https://doi.org/10.3390/s21103478>
3. Xingxian Bao, Tongxuan Fan, Chen Shi, Guanlan Yang, Deep learning methods for damage detection of jacket-type offshore platforms,Process Safety and Environmental Protection,Volume 154,2021
4. Nath, N. D., & Behzadan, A. H. (2020, March). Deep learning detection of personal protective equipment to maintain safety compliance on construction sites. In Construction Research Congress 2020 (pp. 181-190). Reston, VA: American Society of Civil Engineers.
5. Nath, N. D., Behzadan, A. H., & Paal, S. G. (2020). Deep learning for site safety: Real-time detection of personal protective equipment. Automation in Construction, 112, 103085.
6. Wang, Z., Wu, Y., Yang, L., Thirunavukarasu, A., Evison, C., & Zhao, Y. (2021). Fast personal protective equipment detection for real our data set is small. Thus, We Use Yolov5's Pre-construction sites using deep learning approaches. Sensors, 21(10), 3478.
7. Vukicevic, A. M., Djapan, M., Isailovic, V., Milasinovic, D., Savkovic, M., & Milosevic, P. (2022). Generic compliance of industrial PPE by using deep learning techniques. Safety science, 148, 105646.
8. Ahmed, M. I. B., Saraireh, L., Rahman, A., Al-Qarawi, S., Mhran, A., Al-Jalaoud, J., ... & Gollapalli, M. (2023). Personal protective equipment detection: A deep-learning-based sustainable approach. Sustainability, 15(18), 13990.
9. Long, X., Cui, W., & Zheng, Z. (2019, March). Safety helmet wearing detection based on deep learning. In 2019 IEEE 3rd information technology, networking, electronic and automation control conference (ITNEC) (pp. 2495-2499). IEEE.
10. Cheng, J. P., Wong, P. K. Y., Luo, H., Wang, M., & Leung, P. H. (2022). Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification. Automation in Construction, 139, 104312.
11. Wang, S., Li, X., Chen, W., Fan, W., & Tian, Z. (2022). An intelligent vision-based method of worker identification for industrial internet of things (IoT). Wireless Communications and Mobile Computing, 2022.
12. Ahn, J., Park, J., Lee, S. S., Lee, K. H., Do, H., & Ko, J. (2023). SafeFac: Video-based smart safety monitoring for preventing industrial work accidents. Expert Systems with Applications, 215, 119397.
13. Li, J., Zhao, X., Zhou, G., & Zhang, M. (2022). Standardized use inspection of workers' personal protective equipment based on deep learning. Safety science, 150, 105689.
14. Nain, M., Sharma, S., & Chaurasia, S. (2021, March). Safety and compliance management system using computer vision and deep learning. In IOP Conference Series: Materials Science and Engineering (Vol. 1099, No. 1, p. 012013). IOP Publishing.
15. Hayat, A., & Morgado-Dias, F. (2022). Deep learning-based automatic safety helmet detection system for construction safety. Applied Sciences, 12(16), 8268.
16. Nath, N. D., Behzadan, A. H., & Paal, S. G. (2020). Deep learning for site safety: Real-time detection of personal protective equipment. Automation in Construction, 112, 103085.
17. G., Yang, Y., & Qu, X. (2019). Deep learning approaches on pedestrian detection in hazy weather. IEEE Transactions on Industrial Electronics, 67(10), 8889-8899.
18. Ling, X., Ji, S., Zou, J., Wang, J., Wu, C., Li, B., & Wang, T. (2019, May). Deepsec: A uniform platform for security analysis of deep learning model. In 2019 IEEE Symposium on Security and Privacy (SP) (pp. 673-690). IEEE.
19. Antwi-Afari, M. F., Qarout, Y., Herzallah, R., Anwer, S., Umer, W., Zhang, Y., & Manu, P. (2022). Deep learning-based networks for automated recognition and classification of awkward working postures in construction using wearable insole sensor data. Automation in construction, 136, 104181.
20. Rebekah, J., Wise, D. J. W., Bhavani, D., Regina, P. A., & Muthukumaran, N. (2020, July). Dress code Surveillance Using Deep learning. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 394-397). IEEE.
21. Lee, M., & Bae, J. (2020). Deep learning based real-time recognition of dynamic finger gestures using a data glove. IEEE Access, 8, 219923-219933.
22. Rosero-Montalvo, P. D., Godoy-Trujillo, P., Flores-Bosmediano, E., Carrascal-García, J., Otero-Potosi, S., Benitez-Pereira, H., & Peluffo- Ordonez, D. H. (2018, October). Sign language recognition based on intelligent glove using machine learning techniques. In 2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM) (pp. 1-5). IEEE.
23. Seong, H., Son, H., & Kim, C. (2018). A comparative study of machine learning classification for color-based safety vest detection on construction-site images. KSCE Journal of Civil Engineering, 22, 4254-4262.
24. Chen, Z., Zhang, F., Liu, H., Wang, L., Zhang, Q., & Guo, L. (2023). Real-time detection algorithm of helmet and reflective vest based on improved YOLOv5. Journal of Real-Time Image Processing, 20(14).
25. Chen, Z., Zhang, F., Liu, H., Wang, L., Zhang, Q., & Guo, L. (2023). Real-time detection algorithm of helmet and reflective vest based on improved YOLOv5. Journal of Real-Time Image Processing, 20(14).

**CHAPTER 8**

**APPENDICES**

import cv2

import torch

import numpy as np

import time

import smtplib

from email.mime.multipart import MIMEMultipart

from email.mime.text import MIMEText

from email.mime.base import MIMEBase

from email import encoders

fromaddr="tdinesh986@gmail.com"

toaddr= "samuelpharossophy@gmail.com"

def mail(text):

print(text)

msg=MIMEMultipart()

msg['From']=fromaddr

msg['To']=toaddr

msg['Subject']="NOT\_SAFETY"

body=text

msg.attach(MIMEText(body,'plain'))

filename="output/img.jpg"

attachment=open("output/img.jpg","rb")

p=MIMEBase('application','octet-stream')

p.set\_payload((attachment).read())

encoders.encode\_base64(p)

p.add\_header('Content-Disposition',"attachment; filename=%s"%filename)

msg.attach(p)

s=smtplib.SMTP('smtp.gmail.com',587)

s.starttls()

s.login(fromaddr,"zfyvrsnrvlyuuruz")

text=msg.as\_string()

s.sendmail(fromaddr,toaddr,text)

s.quit()

def detect\_objects\_live(weights\_path='best.pt', conf\_threshold=0.2):

# Load YOLOv5 model

model = torch.hub.load('ultralytics/yolov5', 'custom', path=weights\_path, force\_reload=True)

# Set device

device = torch.device('cuda' if torch.cuda.is\_available() else 'cpu')

model.to(device)

# Open video capture device (webcam)

cap = cv2.VideoCapture(0)

while cap.isOpened():

ret, frame = cap.read()

if not ret:

break

# Perform inference

results = model(frame)

# Get bounding boxes, confidence scores, and class labels

boxes = results.xyxy[0] # Bounding boxes in (x1, y1, x2, y2) format

confidences = boxes[:, 4] # Confidence scores

class\_labels = boxes[:, 5] # Class labels

# Filter detections based on confidence threshold

detections\_above\_threshold = boxes[confidences > conf\_threshold]

# Draw bounding boxes for detections above threshold

for detection in detections\_above\_threshold:

label = int(detection[5])

score = float(detection[4])

bbox = detection[:4].cpu().numpy().astype(int)

cv2.rectangle(frame, (bbox[0], bbox[1]), (bbox[2], bbox[3]), (0, 255, 0), 2)

cv2.putText(frame, f'{model.names[label]}: {score:.2f}',

(bbox[0], bbox[1] - 10), cv2.FONT\_HERSHEY\_SIMPLEX, 0.5, (0, 255, 0), 2)

# print(model.names[label])

klass=model.names[label]

# NO-Safety Vest,NO-Mask,NO-Hardhat

if klass == "NO-Hardhat":

cv2.imwrite('output/img.jpg', frame)

print("Without Hardhat")

mail("NOT\_SAFETY")

elif klass == "NO-Safety Vest":

cv2.imwrite('output/img.jpg', frame)

print("Without Safety Vest")

mail("NOT\_SAFETY")

elif klass == "NO-Mask":

cv2.imwrite('output/img.jpg', frame)

print("Without Mask")

mail("NOT\_SAFETY")

# Display the frame

cv2.imshow('Waste Detection', frame)

# Break the loop on 'q' key press

if cv2.waitKey(1) & 0xFF == ord('q'):

break

# Release resources

cap.release()

cv2.destroyAllWindows()

# Example usage

detect\_objects\_live()

**YOLOv5 Code:**

[**https://colab.research.google.com/drive/1OkRHGsgjRiCoSbV6rsDG1Cnmq0BMl4KB?usp=sharing**](https://colab.research.google.com/drive/1OkRHGsgjRiCoSbV6rsDG1Cnmq0BMl4KB?usp=sharing)