YOLOv4-based Deep Learning Approach for Personal Protective Equipment Detection

Mahmoud Abouelyazid
Exodia AI Labs

Abstract

Detecting Personal Protective Equipment (PPE) has become essential for assuring worker safety and regulatory compliance in numerous industries. This study presents a deep learning approach for PPE detection using the YOLOv4 architecture. The primary objective is to develop a robust model capable of identifying ten PPE classes. The dataset used for training and evaluation consists of 2,605 images for training, 114 images for validation, and 82 images for testing, with checks performed to prevent data leakage. The proposed model architecture is based on YOLOv4 and comprises 225 layers. It incorporates convolutional layers, spatial pyramid pooling, skip connections, and data augmentation techniques to enhance detection performance. The model is trained for 100 epochs using Stochastic Gradient Descent (SGD) optimization with a learning rate of 0.01. Evaluation metrics, including precision, recall, and mean Average Precision (mAP), are employed to assess the model's effectiveness. Experimental results demonstrate the model's proficiency in detecting certain PPE classes, such as Mask and machinery, with high precision and recall scores. However, challenges are encountered in accurately detecting the absence of safety items and localizing vehicles. Precision-recall curves reveal trade-offs between precision and recall for safety-related objects, while precision-confidence and F1-confidence curves indicate performance improvements at higher confidence thresholds. A comprehensive analysis of class-wise performance metrics reveal that the vehicle and Person classes exhibit higher box, object, and classification losses, indicating difficulties in accurate localization and classification. Conversely, the Mask class achieves the highest precision, and the machinery and Mask classes demonstrate strong recall performance. This study contributes to the advancement of PPE detection by presenting a deep learning approach using YOLOv4 and conducting a thorough performance analysis across various PPE classes. The findings highlight the importance of detailed performance evaluation to identify class-specific challenges and guide future research efforts in enhancing PPE detection accuracy and robustness. The proposed approach can be integrated into real-world safety monitoring systems, promoting worker safety and compliance in industrial settings.

Indexing terms: Personal Protective Equipment (PPE) detection, deep learning, YOLOv4, object detection, worker safety, performance analysis, industrial safety compliance

Introduction

Workplace injuries continue to be a significant concern in modern society, with a lack of properly worn safety equipment often cited as a primary contributing factor [1], [2]. This issue is particularly prevalent in the construction industry, where job sites are inherently hazardous environments, and workers are at an elevated risk of sustaining injuries or experiencing falls [3], [4]. Despite the well-documented dangers associated with construction work, many employees still fail to consistently and correctly utilize the personal protective equipment (PPE) provided to them, thus increasing their vulnerability to accidents and health risks.

Table 1. Occupational injuries and illnesses per 100 full-time equivalent workers for selected industry groups in the private industry in the year 2011. Source: **Source:** **U.S. Bureau of Labor Statistics.**

Industry Group	Injury rate Rate per 100 Full-Time Equivalent Workers
Private Industry	3.5
Construction	3.9
Manufacturing	4.4
Financial Activities	1.4
Health Care and Social Assistance	5.0

In recent years, there has been a growing recognition of the need to prioritize worker safety within the construction sector. This renewed focus on safety not only benefits individual employees but also has a significant positive impact on the reputation of construction companies. Firms can demonstrate their commitment to the well-being of their workforce through actively promoting and enforcing the use of PPE [5], [6], which can help to attract and retain talented professionals, as well as foster positive relationships with clients and the broader community. The consistent use of safety equipment, such as harnesses and helmets, has been shown to substantially reduce the incidence of falls, which are among the most common and devastating types of accidents on construction sites.

Although there are clear benefits of wearing PPE and the increasing availability of monitoring systems, some workers may still fail to comply with safety regulations. This non-compliance can be temporary, with workers removing their PPE for brief periods due to discomfort, inconvenience, or a perceived lack of immediate risk. In other cases, workers may reject the use of safety equipment altogether, often due to a lack of awareness about the importance of PPE or a belief that it is unnecessary [7], [8]. Regardless of the underlying reasons, the failure to wear PPE consistently can have severe consequences, ranging from minor injuries to life-altering disabilities or even fatalities [9], [10].

Several factors have been identified as contributing to PPE non-compliance in the construction industry. One of the most significant issues is inadequate safety supervision, which can occur when managers and supervisors fail to prioritize safety or do not actively enforce PPE requirements. This lack of oversight can create a culture in which workers feel that PPE use is optional or unimportant, leading to widespread non-compliance. Another critical factor is poor risk perception among workers, who may underestimate the hazards associated with their tasks or overestimate their ability to avoid accidents. This misperception of risk can be particularly problematic when combined with a lack of safety training, as workers may not fully understand the proper use and limitations of their PPE [11], [12].

In addition to these factors, the lack of climate adaptation can also contribute to PPE non-compliance. Construction workers are often required to perform their duties in a wide range of weather conditions, including extreme heat, cold, and humidity. When PPE is not designed or selected with these environmental factors in mind, workers may find it uncomfortable or impractical to wear, leading to increased instances of non-compliance. For example, workers may remove their helmets or eye protection in hot, sunny conditions to cool off, or they may forego gloves in cold weather to maintain dexterity . A lack of management support can also play a significant role in PPE non-compliance. When company leaders and supervisors do not prioritize safety or fail to allocate sufficient resources to safety training and equipment, workers may perceive PPE use as a low priority. This lack of support can also manifest in the form of inadequate communication about safety policies and procedures, as well as a failure to involve workers in the selection and implementation of PPE [13], [14].

To address the challenge of ensuring that workers consistently wear their PPE, many construction companies and safety organizations have invested in the development of monitoring systems. These systems employ various technologies, such as cameras, sensors, and wearable devices, to track the use of safety equipment during working hours. These monitoring solutions enable supervisors and managers to quickly identify instances of non-compliance and intervene, by providing real-time data on PPE compliance to correct the situation before an accident occurs. The data collected by these systems can be analyzed to identify patterns and trends in PPE usage, allowing companies to target their safety training and awareness efforts more effectively.

Artificial Intelligence (AI) has been applied to develop automated, advanced, and cost-effective monitoring systems that can recognize Personal Protective Equipment (PPE) and workers to determine if the worker is complying with safety regulations. These systems help ensure employee safety by detecting and classifying objects using Computer Vision and Deep Learning (DL) technologies. DL is designed to mimic the human brain and has the ability to improve and learn by itself.

Computer Vision's effectiveness is derived from Convolutional Neural Networks (CNNs), which automatically perform feature extraction for targeted objects. The extracted features are then fed into the DL model to support decision-making capabilities. Transfer learning can also be used to increase the reliability of the solution by applying knowledge from previously trained models on related problems.

There are two main categories of detectors: one-stage detectors and two-stage detectors. One-stage detectors, such as the YOLO (You Only Look Once) family, perform both localization and recognition of desired objects in the same phase, enabling near real-time detection [15]. Two-stage detectors, like the Regional-based Convolutional Neural Network (RCNN) family, execute the object detection procedure in two phases to achieve accurate and reliable results. In the first phase, object localization is performed to propose regions with a high probability of containing objects. The second phase then identifies objects from the characteristics extracted from the localized regions. Both types of detectors aim to detect targeted objects, but there is a notable trade-off between real-time detection and accuracy. One-stage detectors prioritize speed, while two-stage detectors focus on precision.

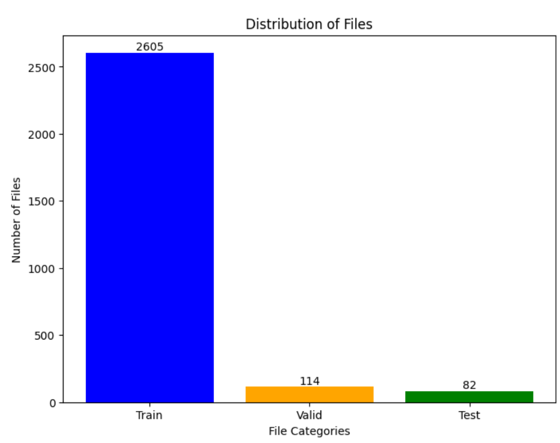
Implementing AI-based monitoring systems for PPE compliance can significantly improve workplace safety by automating the process of identifying workers who are not wearing the required protective equipment. This technology can be applied in various industries, such as construction, manufacturing, and healthcare, where adherence to safety regulations is crucial for preventing accidents and injuries.

Methods

Dataset and PPE Classes

The study focuses on detecting Personal Protective Equipment (PPE) using the YOLOv4 architecture. The dataset comprises ten PPE classes as shown Figure 2. The dataset is split into three subsets: training, validation, and testing. The training set consists of 2,605 images, while the validation set contains 114 images, and the testing set includes 82 images.

Figure 1. distribution of the training, validation, and test data



Data Integrity Check

To ensure the reliability of the training process, there is a need to verify that no filenames are present in more than one subset of the dataset. This issue is known as data leakage and can lead to unreliable training results. In this study, a thorough check is performed to confirm that each filename is unique across the training, validation, and testing sets.

Data Annotation Quality

When dealing with large datasets, especially those obtained from different competitions, it is common to have manual annotators involved in the labeling process. To ensure the quality and consistency of the annotations, three key aspects are investigated before proceeding with Exploratory Data Analysis (EDA) or data pre-processing. These aspects include verifying the accuracy of the annotations, checking for consistency in labeling across different annotators, and ensuring that the annotations adhere to the defined PPE classes and their corresponding definitions.

Figure 2. 10 classes

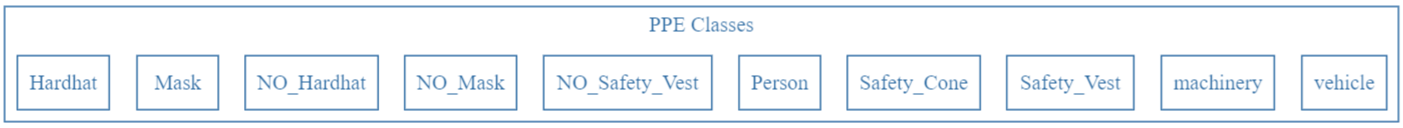


Table 2. Class distribution

	Train	Valid	Test
0	3145	79	110
1	1651	21	28
2	2317	69	41
3	3097	74	79
4	3962	106	90
5	9532	166	174
6	3366	44	92
7	3033	41	61
8	5247	55	44
9	1545	42	41

The training set consists of a total of 36,895 instances across all PPE classes, with Class 5 having the highest representation of 9,532 instances and Class 9 having the lowest representation of 1,545 instances. The validation set contains a total of 697 instances, with Class 5 having the highest representation of 166 instances and Class 1 having the lowest representation of 21 instances. The testing set comprises a total of 760 instances, with Class 5 having the highest representation of 174 instances and Class 1 having the lowest representation of 28 instances.

The distribution of instances across the different PPE classes varies within each subset. In the training set, Class 5 has the highest number of instances, followed by Class 8, Class 4, Class 6, Class 0, Class 3, Class 7, Class 2, Class 1, and Class 9. The validation set follows a similar pattern, with Class 5 having the highest representation and Class 1 having the lowest. In the testing set, Class 5 remains the most represented class, while Class 1 has the lowest number of instances.

YOLOv4

The YOLOv4 architecture is a state-of-the-art object detection system that combines the best practices and techniques from various object detection models to achieve optimal speed and accuracy. It is designed to be a one-stage detector, which means that it does not rely on a preliminary stage to identify regions of interest before classifying objects within those regions [16]. Instead, YOLOv4 directly predicts the presence and location of objects in a single pass through the network, making it faster and more efficient than two-stage detectors.

The architecture of YOLOv4 consists of several key components. The first stage is the "*Input*" stage, where the dataset is introduced to the network. The input data can be of any size, but it is often resized to a consistent resolution to facilitate efficient processing. Once the input data is prepared, it is passed to the "Backbone" stage [17].

The *backbone* of YOLOv4 is a convolutional neural network (CNN) that is responsible for extracting features from the input images. CNNs are well-suited for image processing tasks. They consist of multiple layers of interconnected nodes for performing a specific operation on the input data. The layers in a CNN typically include convolutional layers, which apply a set of learnable filters to the input image to detect specific features, and pooling layers, which downsample the feature maps to reduce computational complexity and increase robustness to small variations in the input.

There are several popular CNN architectures that can be used as the *backbone* for YOLOv4, including VGG16, ResNet-50, SpineNet, EfficientNet-B0/B7, CSPResNeXt50, and CSPDarknet53. Among these options, CSPDarknet53 has been shown to provide the best results in terms of both speed and accuracy, according to the paper "Yolov4: Optimal speed and accuracy of object detection." As a result, CSPDarknet53 is the default *backbone* used in the YOLOv4 architecture [18].

After the *backbone*, the feature maps are passed through the "*Neck*" stage, which consists of additional layers that enhance the discriminability and robustness of the features. The neck stage employs various techniques, such as Feature Pyramid Network (FPN), Path Aggregation Network (PAN), Spatial Pyramid Pooling (SPP), BiFPN, and NAS-FPN. In YOLOv4, the two primary components used in the neck stage are SPP and PAN.

Spatial Pyramid Pooling (SPP) allows the network to extract the most important contextual features from the input image without affecting the overall speed of the network. It works by dividing the feature maps into multiple spatial bins of different sizes and then pooling the features within each bin. This approach helps to capture multi-scale information and improves the network's ability to detect objects of different sizes.

Path Aggregation Network (PAN) is used in the *neck* stage of YOLOv4. PAN provides a way to aggregate feature maps from different levels of the *backbone* network and combine them for use in the detector stage. PAN helps to improve the network's ability to detect objects at different scales and increases the overall robustness of the object detection system [19].

The final stage in the YOLOv4 architecture is the "*Head*" stage, which is responsible for predicting the classes and bounding boxes of the detected objects. In the one-stage detector model used by YOLOv4, the head stage is related to dense prediction, which means that it directly predicts the presence and location of objects without relying on a preliminary region proposal stage. This is in contrast to two-stage detector models, which use a sparse prediction stage (such as Faster R-CNN, R-FCN, or RepPoint) to generate region proposals before classifying the objects within those regions [20].

The *head* of YOLOv4 is based on the YOLOv3 architecture, which has been proven to be effective for dense object detection. YOLOv3 uses a series of convolutional layers to predict the class probabilities and bounding box coordinates for each cell in the output feature map. It also employs anchor boxes, which are predefined bounding boxes of different sizes and aspect ratios, to improve the network's ability to detect objects of various shapes and sizes.

YOLOv4 architecture can achieve real-time object detection while maintaining high accuracy. This is made possible by the careful selection and combination of techniques used in each stage of the network. For example, the use of CSPDarknet53 as the backbone provides a good balance between speed and accuracy, while the incorporation of SPP and PAN in the neck stage helps to enhance the discriminability and robustness of the features.

Multi-scale object detection for YOLOv4 is achieved through the use of feature pyramids, which allow the network to detect objects at different scales by processing the input image at multiple resolutions. The feature maps from different levels of the backbone are combined and upsampled in the neck stage. This provides the head stage with a rich set of features that can be used to detect objects of various sizes.

In addition to its architectural innovations, YOLOv4 also incorporates several training techniques that help to improve its performance. These include data augmentation, which involves applying random transformations to the input images during training to increase the diversity of the training data, and multi-scale training, which involves training the network on images of different resolutions to improve its capacity to handle objects at different scales.

Model configuration and training setup

Model Architecture:

The proposed model for PPE detection is based on the YOLOv4 architecture and consists of a total of 225 layers. The architecture incorporates various components, including convolutional layers, spatial pyramid pooling, concatenation layers, and detection layers. These layers are to extract meaningful features from the input images and perform object detection. To facilitate the flow of information and improve the model's performance, skip connections are employed for allowing the model to leverage features from different scales and depths.

Optimizer and Training Parameters:

The model is trained using the Stochastic Gradient Descent (SGD) optimization algorithm with a learning rate of 0.01. The learning rate determines the step size at which the model's parameters are updated during the training process. The parameters of the model are grouped into three categories: weights without decay, weights with decay, and biases. This grouping is for different regularization strategies to be applied to different sets of parameters for aiding in preventing overfitting and improving generalization.

Data Augmentation Techniques:

Various data augmentation techniques are employed during training to enhance the model's robustness and generalization ability,. These techniques include blur, median blur, grayscale conversion, and Contrast Limited Adaptive Histogram Equalization (CLAHE). After applying these augmentations to the training images, the model is exposed to a wider range of variations and becomes more resilient to different lighting conditions, noise, and image transformations encountered in real-world scenarios.

Data Preparation and Loading:

The training and validation datasets are sourced from specific directories, ensuring a clear separation between the two sets. During the data preparation phase, duplicate labels are identified and removed from the training data to avoid any inconsistencies or redundancies. To efficiently load the data during training, two workers are utilized, for parallel processing and faster data retrieval. The model undergoes a training process that spans a total of 100 epochs.

Model Initialization:

Prior to training, the model's weights are initialized using pretrained weights. These pretrained weights is for allowing the model to leverage the knowledge learned from previous tasks or datasets. The initialization process is successful for most of the model's components for further fine-tuning and adaptation to the specific PPE detection task.

Logging and Visualization:

Throughout the training process, important information and results are logged for monitoring and analysis purposes. This includes the labels associated with the detected objects, as well as the progress of the training process. To facilitate visual inspection and interpretation, the detected labels are plotted and saved for reference. The proposed YOLOv4-based approach aims to achieve robust and accurate PPE detection.

Results

Following results are achieved.

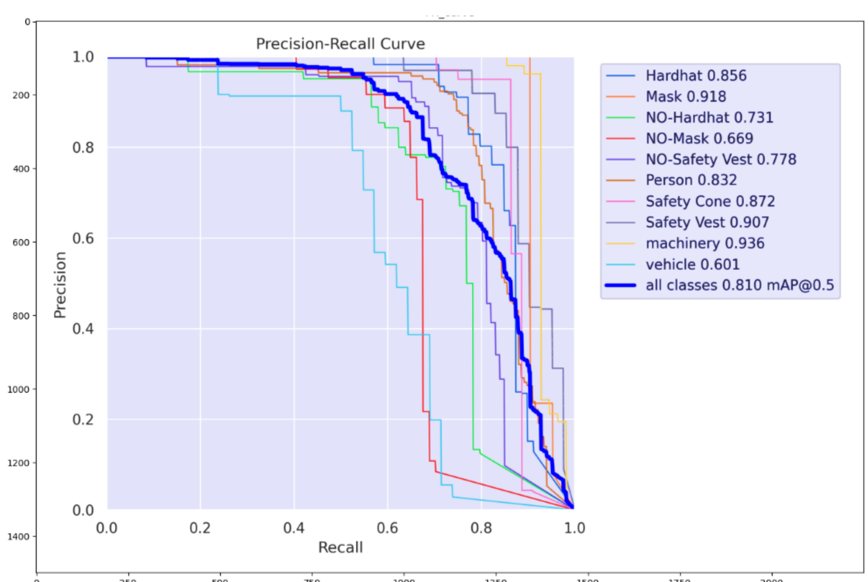
- **Decreasing Loss**: The loss values for bounding box regression, class prediction, and possibly other losses are decreasing over epochs. This suggests that the model is learning and improving its performance.
- **Consistent GPU Memory Usage**: The GPU memory usage seems consistent throughout the training process, indicating that the model architecture and batch size might remain stable.
- **Increasing Number of Instances**: The number of instances detected in each epoch seems to fluctuate but generally increases towards the end of training. This could mean that the model is becoming more capable of detecting objects in the dataset.
- **Evaluation Metrics**: The evaluation metrics (Box Precision, Recall, mAP50, etc.) for all classes are improving or remaining stable over epochs, indicating

overall improvement in the model's performance in terms of object detection accuracy.

Table 3. **Evaluation Metrics by Class**

Class	Images	Instances	Box(P)	R	mAP50	mAP
all	114	697	0.919	0.729	0.81	0.507
Hardhat	114	79	0.921	0.735	0.856	0.56
Mask	114	21	0.966	0.905	0.918	0.665
NO-Hardhat	114	69	0.939	0.565	0.731	0.412
NO-Mask	114	74	0.894	0.595	0.669	0.343
NO-Safety Vest	114	106	0.913	0.651	0.778	0.45
Person	114	166	0.905	0.744	0.832	0.515
Safety Cone	114	44	0.892	0.864	0.872	0.52
Safety Vest	114	41	0.923	0.78	0.907	0.604
Machinery	114	55	0.954	0.927	0.936	0.65
Vehicle	114	42	0.88	0.522	0.601	0.347

Figure 3. precision-recall curve



The precision-recall curves reveal the performance characteristics of the object detection model for different classes. The "Mask" class stands out with a high and relatively stable precision across a wide range of recall values, indicating excellent performance in detecting masks. Similarly, the "machinery" class maintains high precision for most recall levels, suggesting reliable detection of machinery objects.

On the other hand, the "NO-Mask" and "vehicle" classes exhibit lower precision values, especially at higher recall levels. This implies that the model struggles more with accurately detecting the absence of masks and identifying vehicles compared to other classes.

The "Hardhat", "Safety Vest", and "Safety Cone" classes show similar precision-recall curves, with precision gradually decreasing as recall increases. This indicates a trade-off between precision and recall for these safety-related objects, where the model's ability to detect all relevant instances comes at the cost of some false positive predictions.

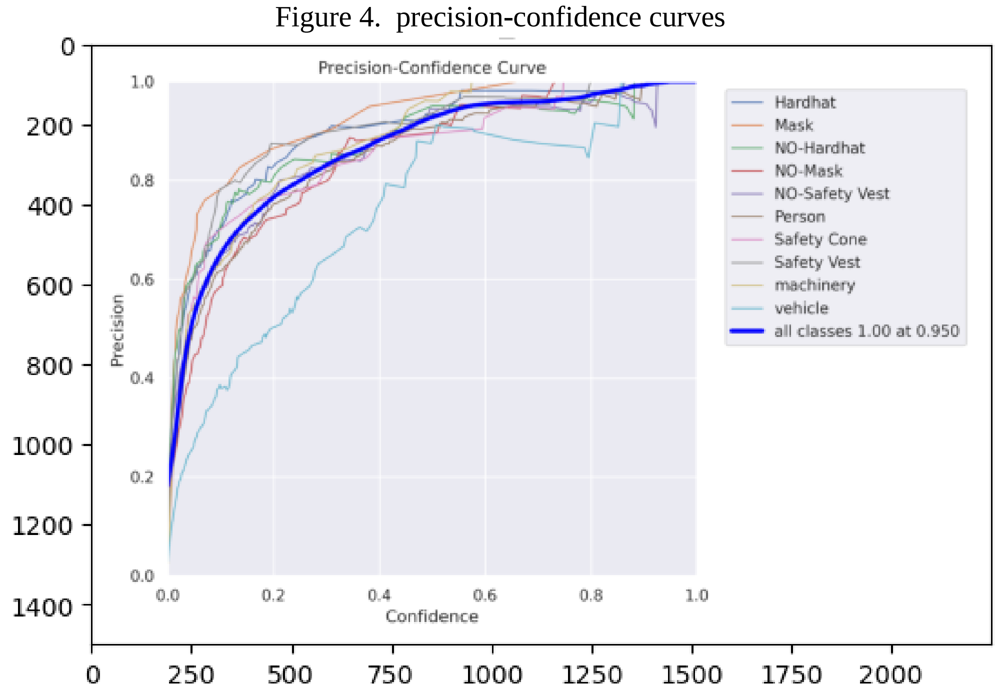
The "Person" class has a precision-recall curve that lies in the middle of the pack, suggesting moderate performance in detecting people compared to other classes.

Notably, the "NO-Hardhat" and "NO-Safety Vest" classes have lower overall precision compared to their counterparts ("Hardhat" and "Safety Vest"), indicating that the model may have more difficulty accurately identifying the absence of these safety items.

The dashed black line represents the overall performance across all classes, providing an aggregate view of the model's precision-recall characteristics. It shows a balance between precision and recall, with the curve maintaining relatively high precision values up to a recall level of around 0.6 before starting to decline more rapidly.

In summary, the precision-recall curves highlight the model's strong performance in detecting classes like "Mask" and "machinery", while also revealing challenges in accurately detecting the absence of certain safety items and identifying vehicles. The curves provide valuable insights into the model's behavior and trade-offs between precision and recall for each class.

Figure 4. precision-confidence curves



The precision-confidence curves shows that as the confidence scores increase, the precision values for most classes tend to improve, suggesting that the model's predictions become more accurate when it has higher confidence in its detections.

However, there are notable differences in the precision-confidence curves among the classes. The "Mask" and "machinery" classes maintain high precision values across a wide range of confidence scores, indicating strong and consistent performance in detecting these objects. On the other hand, classes like "NO-Mask", "NO-Safety Vest", and "vehicle" show lower precision values, especially at lower confidence scores, suggesting room for improvement in detecting these objects accurately.

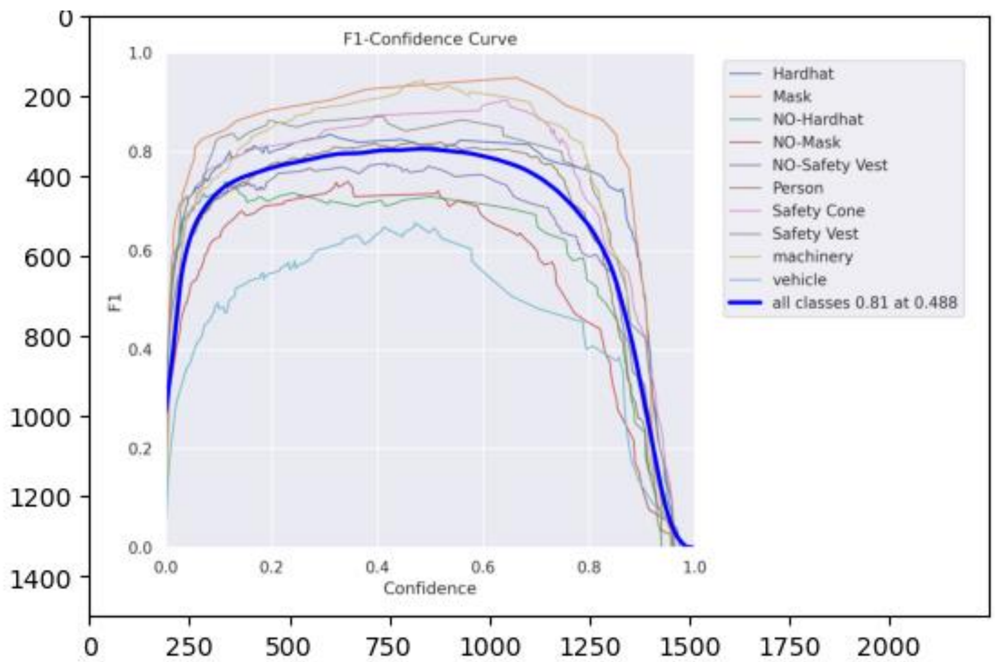
The "Hardhat", "Safety Vest", and "Safety Cone" classes exhibit similar precision-confidence curves, with precision values increasing steadily as confidence scores rise. This indicates that the model performs relatively well in detecting these safety-related objects.

The "Person" class has a precision-confidence curve that falls somewhere in the middle, suggesting moderate performance in detecting people compared to other classes.

Interestingly, the "NO-Hardhat" and "NO-Mask" classes have lower precision values compared to their counterparts ("Hardhat" and "Mask"), indicating that the model may struggle more in detecting the absence of these safety items.

Overall, the precision-confidence curves highlight the model's strengths in detecting certain classes like "Mask" and "machinery", while also revealing areas where the model's performance could be improved, particularly for classes such as "NO-Mask", "NO-Safety Vest", and "vehicle".

Figure 5. F1-confidence curves curve



The F1-confidence curves provide insights into the balance between precision and recall for the object detection model across different classes. As confidence scores increase, the F1 scores for most classes tend to improve, indicating a better balance between precision and recall when the model has higher confidence in its predictions.

However, there are notable differences in the F1-confidence curves among the classes. The "Mask" and "machinery" classes maintain high F1 scores across a wide range of confidence levels, suggesting a strong and consistent balance between precision and recall in detecting these objects. On the other hand, classes like "NO-Mask", "NO-Safety Vest", and "vehicle" show lower F1 scores, especially at lower confidence levels, indicating room for improvement in achieving a good balance between precision and recall for these classes.

The "Hardhat", "Safety Vest", and "Safety Cone" classes exhibit similar F1-confidence curves, with F1 scores increasing steadily as confidence levels rise. This suggests that the model maintains a relatively good balance between precision and recall in detecting these safety-related objects.

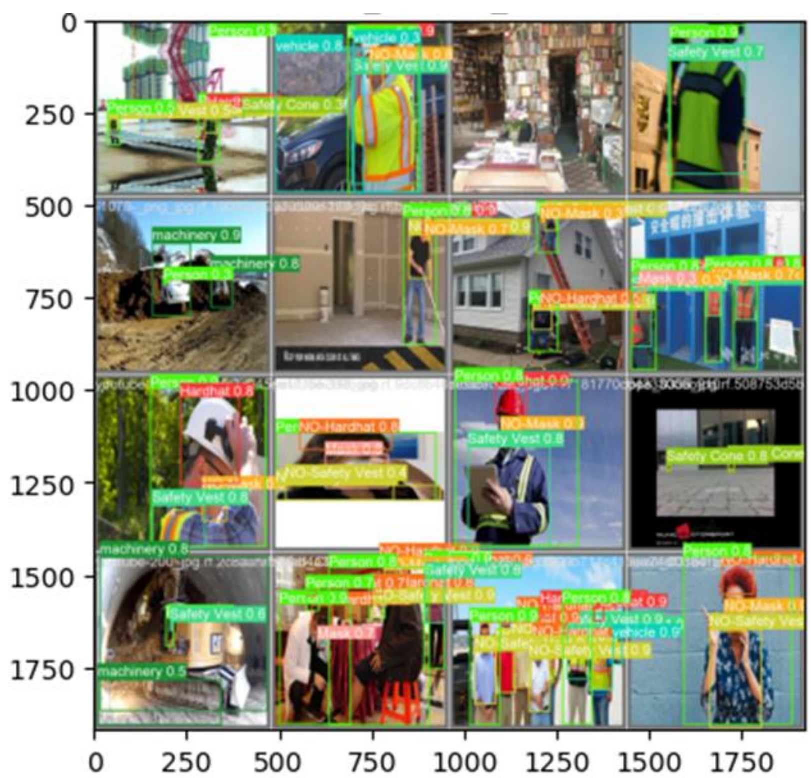
The "Person" class has an F1-confidence curve that falls in the middle range compared to other classes, indicating a moderate balance between precision and recall in detecting people.

Interestingly, the "NO-Hardhat" and "NO-Mask" classes have lower F1 scores compared to their counterparts ("Hardhat" and "Mask"), suggesting that the model may have more difficulty achieving a good balance between precision and recall when detecting the absence of these safety items.

The dashed black line represents the overall F1 score across all classes at an IoU threshold of 0.5, which reaches a value of 0.488 at the highest confidence level. This indicates that, on average, the model achieves a moderate balance between precision and recall when considering all classes together.

Overall, the F1-confidence curves highlight the model's strengths in maintaining a good balance between precision and recall for certain classes like "Mask" and "machinery", while also revealing areas where the model's performance could be improved to achieve a better balance, particularly for classes such as "NO-Mask", "NO-Safety Vest", and "vehicle".

Figure 6. prediction demonstration



Conclusion

This study demonstrates the effectiveness of the YOLOv4-based deep learning approach for detecting Personal Protective Equipment (PPE) across various classes. The proposed model architecture incorporates convolutional layers, spatial pyramid pooling, skip connections, and data augmentation techniques. They exhibited good performance in identifying certain PPE classes, such as Mask and machinery. Challenges persisted in accurately detecting the absence of safety items and localizing vehicles, as evidenced by the class-wise performance metrics and precision-recall trade-offs.

The analysis of class-specific performance metrics shows the model's strengths and weaknesses. Although the Mask class achieves high precision and the machinery class demonstrates strong recall, the vehicle and Person classes face difficulties in accurate localization and classification, as indicated by their higher box, object, and classification losses.

The findings of this study have significant implications for real-world safety monitoring systems in industrial settings. Organizations can enhance worker safety, ensure compliance with regulations, and proactively identify potential safety hazards by integrating the proposed YOLOv4-based approach. It is crucial to recognize the limitations and continually strive for improvements in PPE detection accuracy across all classes.

To effectively address the problem of PPE non-compliance in the construction industry, a multi-faceted approach is necessary. This should include safety training and education for workers, with a focus on raising awareness about the importance of PPE and the specific hazards associated with different tasks. Training should also cover the proper selection, use, and maintenance of safety equipment, as well as the potential consequences of non-compliance. In addition, companies should invest in high-quality, comfortable, and adaptable PPE that is well-suited to the specific needs of their workers and the environments in which they operate.

Equally important is the need for strong leadership and management support for safety initiatives. Company leaders and supervisors must prioritize safety as a core value and actively promote a culture of compliance. This can involve regular safety audits and inspections, as well as the implementation of incentive programs that reward workers for consistent PPE use. Managers should also be trained to recognize and address instances of non-compliance promptly and effectively, using a combination of education, coaching, and, when necessary, disciplinary action.

Adoption of modern monitoring systems can significantly improve PPE compliance rates. These technologies can assist businesses in promptly and efficiently identifying and addressing noncompliance issues through offering real-time data on PPE usage and enabling targeted responses. The implementation of these systems is accompanied by clear communication about their purpose and benefits, as well as safeguards to protect worker privacy and prevent the misuse of data.

References

[1] L. I. Boden, E. A. Biddle, and E. A. Spieler, "Social and economic impacts of workplace illness and injury: current and future directions for research," *Am. J. Ind. Med.*, vol. 40, no. 4, pp. 398–402, Oct. 2001.
[2] R. T. Reville, J. Bhattacharya, and L. R. Sager Weinstein, "New methods and data sources for measuring economic consequences of workplace injuries," *Am. J. Ind. Med.*, vol. 40, no. 4, pp. 452–463, Oct. 2001.
[3] S. F. Conti and L. Silverman, "Epidemiology of foot and ankle injuries in the workplace," *Foot Ankle Clin.*, vol. 7, no. 2, pp. 273–290, Jun. 2002.
[4] D. McCaughey, A. Kimmel, G. Savage, T. Lukas, E. Walsh, and J. Halbesleben, "Antecedents to workplace injury in the health care industry: A synthesis of the literature," *Health Care Manage. Rev.*, vol. 41, no. 1, pp. 42–55, Jan-Mar 2016.
[5] A. Patel, M. M. D'Alessandro, K. J. Ireland, W. G. Burel, E. B. Wencil, and S. A. Rasmussen, "Personal Protective Equipment Supply Chain: Lessons Learned from Recent Public Health Emergency Responses," *Health Secur*, vol. 15, no. 3, pp. 244–252, May/Jun 2017.
[6] S. Hegde, "Which type of personal protective equipment (PPE) and which method of donning or doffing PPE carries the least risk of infection for healthcare workers?," *Evid. Based. Dent.*, vol. 21, no. 2, pp. 74–76, Jun. 2020.
[7] National Research Council (US) and Institute of Medicine (US) Panel on Musculoskeletal Disorders and the Workplace, *Musculoskeletal Disorders and the Workplace: Low Back and Upper Extremities*. Washington (DC): National Academies Press (US), 2001.
[8] M. S. Christian, J. C. Bradley, J. C. Wallace, and M. J. Burke, "Workplace safety: a meta-analysis of the roles of person and situation factors," *J. Appl. Psychol.*, vol. 94, no. 5, pp. 1103–1127, Sep. 2009.
[9] S. Rendeki *et al.*, "An Overview on Personal Protective Equipment (PPE) Fabricated with Additive Manufacturing Technologies in the Era of COVID-19 Pandemic," *Polymers* , vol. 12, no. 11, Nov. 2020.
[10] W. A. Fischer 2nd, D. Weber, and D. A. Wohl, "Personal Protective Equipment: Protecting Health Care Providers in an Ebola Outbreak," *Clin. Ther.*, vol. 37, no. 11, pp. 2402–2410, Nov. 2015.
[11] P. L. Murphy, G. S. Sorock, T. K. Courtney, B. S. Webster, and T. B. Leamon, "Injury and illness in the American workplace: a comparison of data sources," *Am. J. Ind. Med.*, vol. 30, no. 2, pp. 130–141, Aug. 1996.
[12] C. M. Barnes and D. T. Wagner, "Changing to daylight saving time cuts into sleep and increases workplace injuries," *J. Appl. Psychol.*, vol. 94, no. 5, pp. 1305–1317, Sep. 2009.
[13] S. Ambigapathy *et al.*, "How should front-line general practitioners use personal protective equipment (PPE)?," *Malays Fam Physician*, vol. 15, no. 1, pp. 2–5, Mar. 2020.
[14] Z. Tian, M. Stedman, M. Whyte, S. G. Anderson, G. Thomson, and A. Heald, "Personal protective equipment (PPE) and infection among healthcare workers - What is the evidence?," *Int. J. Clin. Pract.*, vol. 74, no. 11, p. e13617, Nov. 2020.
[15] M. I. H. Azhar, F. H. K. Zaman, N. M. Tahir, and H. Hashim, "People Tracking System Using DeepSORT," in *2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2020, pp. 137–141.

[16] J. Deng, X. Xuan, W. Wang, Z. Li, H. Yao, and Z. Wang, “A review of research on object detection based on deep learning,” *J. Phys. Conf. Ser.*, vol. 1684, no. 1, p. 012028, Nov. 2020.
[17] J. Fan, T. Huo, and X. Li, “A Review of One-Stage Detection Algorithms in Autonomous Driving,” in *2020 4th CAA International Conference on Vehicular Control and Intelligence (CVCI)*, 2020, pp. 210–214.
[18] Y.-J. Liang, X.-P. Cui, X.-H. Xu, and F. Jiang, “A Review on Deep Learning Techniques Applied to Object Detection,” in *2020 7th International Conference on Information Science and Control Engineering (ICISCE)*, 2020, pp. 120–124.
[19] C.-Y. Wang, A. Bochkovskiy, and H. Liao, “Scaled-YOLOv4: Scaling cross stage partial network,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 13024–13033, Nov. 2020.
[20] Z. Jiang, L. Zhao, S. Li, and Y. Jia, “Real-time object detection method based on improved YOLOv4-tiny,” *arXiv [cs.CV]*, 09-Nov-2020.