Personal protective equipment detection using YOLOv8 architecture on object detection benchmark datasets: a comparative study

Alibek Barlybayev, Nurzada Amangeldy, Bekbolat Kurmetbek, Iurii Krak, Bibigul Razakhova, Nazira Tursynova & Rakhila Turebayeva

cogent

COMPUTER SCIENCE | RESEARCH ARTICLE

OPEN ACCESS · Check for updates

Personal protective equipment detection using YOLOv8 architecture on object detection benchmark datasets: a comparative study

Alibek Barlybayev[a,b] (iD), Nurzada Amangeldy[a] (iD), Bekbolat Kurmetbek[a] (iD), Iurii Krak[c] (iD), Bibigul Razakhova[a] (iD), Nazira Tursynova[a] (iD) and Rakhila Turebayeva[a] (iD)

[a]Faculty of Information Technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan; [b]Higher School of Information Technology and Engineering, Astana International University, Astana, Kazakhstan; [c]Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

ABSTRACT

Over the past decade, global industrial and construction growth has underscored the importance of safety. Yet, accidents continue, often with dire outcomes, despite numerous safety-focused initiatives. Addressing this, this article introduces a novel approach using YOLOv8, a rapid object detection model, for recognizing personal protective equipment (PPE). This method, leveraging computer vision (CV) instead of traditional sensor-based systems, offers an economical, simpler and field-friendly solution. We established the Color Helmet and Vest (CHV) and Safety HELmet dataset with 5K images (SHEL5K) datasets, comprising eight object classes like helmets, vests and goggles, to detect worker-worn PPE. After categorizing the dataset into training, testing and validation subsets, diverse YOLOv8 models were assessed based on metrics including precision, recall and mAP50. Notably, YOLOv8x and YOLOv8l excelled in PPE detection, particularly in recognizing person and vest categories. This innovative CV-driven method promises real-time PPE detection, fortifying worker safety on construction sites.

1. Introduction

In the last 10 years, there has been some significant growth in the industrial and construction sectors all over the world. Safety is now a major priority and yet, accidents keep occurring there and are often undetected until it is too late. Injuries, regardless of the severity, can have a huge impact on the worker, their family and the project's timeline and its budget. Consequently, various initiatives have been introduced in recent years to enhance job site safety and efficiency. Furthermore, these initiatives are becoming smarter and more innovative than ever before. According to the Committee on Labour, Social Protection and Migration of the Ministry of Labor and Social Protection of the Population of the Republic of Kazakhstan, there were workplace accidents at the end of 2021 that resulted in injuries to 1465 workers, of which 200 were fatal

(https://kz.kursiv.media/2022-01-29/v-kazakhstane-rabochie-kalechatsya-na-proizvodstve-v-tri-raza-chaschechem/).

Several contributing factors can lead to accidents in the workplace, such as an employee's unawareness and inexperience, inadequate safety training, not properly using PPEs, not having a safety officer in hazardous areas and malfunctioning machinery. Working in any industry carries inherent risks of accidents and injuries. These risks range from falls from ladders and structures, electrocution, being struck by moving machinery or falling materials, getting caught in equipment, and much more – each depending on the specific work environment. Preventing accidents in the workplace can be achieved by diligently following safety protocols and regulations. This could include offering employees suitable safety

---

training, strictly enforcing PPE compliance, performing regular maintenance checks on work machines, marking danger areas and appointing a safety supervisor for each risky area. Despite the authorities' ongoing efforts to enhance workplace safety, managing such safety remains a complex task that necessitates manual intervention. This complex endeavor requires continuing due diligence and constant attention in order to give everyone the protection they deserve. Although the use of PPE can prevent workplace accidents and injuries, some workers still neglect to wear protective gear when working. This not only jeopardizes their safety, but it also has a negative impact on company productivity and leads to financial losses. The timely identification of workers not utilizing PPE is crucial for maintaining a safe production environment. This is usually done manually by examining surveillance footage, however, AI-driven solutions are being developed to automate this process and detect unsafe human behavior more efficiently and accurately.

For a project or venture to be successful, prioritizing the safety and well-being of employees is pertinent. Regular assessments by relevant authorities should take the physical and mental health of employees into account to ensure their optimal performance and satisfaction. Mishaps or accidents can easily lead to prolonged complications or possible failures for both parties involved. Potential risks, however, can be better understood and managed through careful examination of multiple factors that are difficult to assess manually. To reduce the chances of accidents, an AI-assisted approach is necessary for efficient risk prediction.

Currently, PPE detection can be divided into two distinct categories – sensor-based and vision-based. Sensor-based methods typically use positioning technology to identify personnel and protective equipment. Vision-based techniques involve computer vision (CV) algorithms to detect the presence of PPE.

Initially, researchers employed sensor-based techniques for PPE identification on construction sites (Dong et al., 2015; Kelm et al., 2013). However, these methods entailed additional expenses and posed hazards to workers' safety. Therefore, it was not a practical solution, as it added extra costs to production without supplying the desired outcome. Before the advent of deep learning, image processing, in conjunction with machine learning, was primarily utilized to ascertain whether employees were properly donned in their protective equipment (Fang, Ding, et al., 2018; Park & Brilakis, 2012; Cai & Qian, 2011).

Regrettably, this approach exhibits efficacy only when distractions in the surrounding environment are minimal and performs suboptimally in situations involving complex background scenes. To ensure that the personal protective equipment (PPE) is used properly, Zhang et al. (2015) utilized the Global Position System (GPS) to identify workers and helmets. In addition, Kelm et al. (2013) developed a mobile Radio Frequency Identification (RFID) platform to confirm compliance with PPE regulations. Wearing RFID PPE, workers' information can be documented when they go through verification gates. Regrettably, this approach exhibits efficacy only when distractions in the surrounding environment are minimal and performs suboptimally in situations involving complex background scenes. Although the use of this technology requires workers to wear an extra device to send and receive data, sensor-based helmet detection methods utilize equipment that are not influenced by external elements such as weather, illumination and humidity. Sensor-based approaches have the advantage of giving reliable results and making them applicable for most construction sites. Nonetheless, such systems require a considerable initial and long-term investment in its acquisition, installation and maintenance. While individual sensors are relatively affordable, installing them for every piece of PPE and each employee accumulates costs quickly, suggesting that scalability might be constrained. In addition to this, traditional RFID approaches require workers to wear an end device for connection with the network, which adds more weight and causes discomfort. RFID technology is being implemented to track when an employee utilizes their PPE. This system utilizes Zigbee mesh network communication to do this (Barro-Torres et al., 2012). However, investing in sensors and setting them up can be difficult and expensive which makes it hard to scale up. Also, troubleshooting (Stojanovic et al., 2020) and maintenance add to the complexity of the process.

Utilizing cameras to capture images of construction sites, vision-based methods afford a more comprehensive insight into complex areas. These pictures can provide abundant information which can be processed for PPE detection quickly, accurately and more comprehensively (Seo et al., 2015). Zhu et al. employed HOG to detect head features, which were subsequently submitted to an SVM (Support Vector Machines) for classification to discern whether individuals were wearing helmets (Zhu et al., 2015; Park et al., 2015). Rubaiyat et al. (2016) came up with an innovative method by combining

HOG (Histogram of Oriented Gradients) and SVM for human detection and Circle Hough Transform (CHT) to detect helmets. This method has proven successful in detecting both humans and helmets. Wu and Zhao (2018 implemented K-nearest neighbors (KNN) to pick out moving objects from videos and fed them into convolutional neural networks (CNNs) for categorizing of the pedestrian, head and helmet. Pradana et al. developed a CNN-based model (Pradana et al., 2019) to classify 12 situations, which consisted of 5 PPE items including glasses and helmets. Although experiments were conducted on images with plain indoor backgrounds (not on actual outdoor building sites), this might limit further adaptation for outdoor environments.

Region CNN (R-CNN) models are becoming the go-to family for object recognition. This includes R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2017), each successive method offering improvements over its predecessor in terms of performance. R-CNN uses a region proposal algorithm called 'Selective Search' to generate 2000 regions from an image. Subsequently, visual descriptors are automatically extracted using convolutional layers, with each region being classified *via* a one-versus-all SVM (Liu and Zheng, 2005) classifier. Fast R-CNN was introduced to improve the time complexity of the model. Instead of computing visual descriptors from 2000 regions, this approach extracts visual descriptors from the entire image first. Subsequently, a region of interest (ROI) pooling layer is applied to aggregate the contextual descriptions from the concluding feature map. This is then followed by a SoftMax layer that classifies this area. Building upon R-CNN (Girshick et al., 2014) and Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2017) was developed, extending the capabilities of the prior models. In contrast with the 'Selective search' algorithm, this technology utilizes a region proposal network (RPN) instead. As opposed to manually extracting a limited number of regions which may be either empty or partially include the object, Faster R-CNN (Ren et al., 2017) uses a mini CNN known as the RPN to learn the location of the region in question. These region-based solutions offer two outputs. These coordinates provide the boundaries of the object of interest, along with its respective class. Utilizing this data will help to identify and classify various objects. (Akbarzadeh et al. 2020) implemented two Faster R-CNN approaches to identify violations of safety regulations. The first one recognized human presence at the construction site whereas the second one identified helmet and vest use. The Faster R-CNN approach (Ren et al., 2015) is

an effective solution for detecting workers wearing helmets under remote surveillance (Fang, Li, et al., 2018). A study by Fan et al. (2020) compared multiple object detection algorithms and highlighted that Faster R-CNN had the best performance for detecting large-scale targets such as helmets.

Deep learning has emerged as the predominant method for PPE detection, witnessing notable success with techniques such as object detection (Li et al., 2023). It provides incredibly accurate results with a faster processing speed which is one of the reasons why it is being adopted in industrial production (Wang et al., 2020; Han et al., 2021). Vision-based methods can be bifurcated into two categories: the conventional approach combining image processing and machine learning (Fang, Ding, et al., 2018; Park & Brilakis, 2012; Cai & Qian, 2011; Li et al., 2017; Bo et al., 2019), and the utilization of deep learning technologies, such as object detection (Huang et al., 2021; Nath et al., 2020; Xiong & Tang, 2021; Wang, Wu, et al., 2021; Shen et al., 2021; Iannizzotto et al., 2021; Gallo et al., 2022; Ferdous & Ahsan, 2022). Traditionally, image processing techniques have been utilized to detect the ROI and extract pertinent features. Following this, machine learning methods can be employed to train a classifier which is capable of determining if the region is a helmet or workwear (Lowe, 2004; Dalal & Triggs, 2005; Lienhart & Maydt, 2002). Li et al. (2017) employed the ViBe background modeling algorithm and a pedestrian classification framework to accurately identify workers. Subsequently, they pinpointed the head region and applied color space transformations and color feature recognition to detect helmets. Cai and Qian (2011) developed edge images of safety helmets from different perspectives and extracted four directional features. To create a classifier for recognizing safety helmets *versus* non-safety helmets, they modeled the feature dispersion with a Gaussian function.

The advent of deep learning, object detection and related technologies has propelled significant innovations in PPE detection. Improvements such as R-CNN (He et al., 2017) have helped break through a bottleneck period for object detection algorithms and have increased both speed and accuracy significantly. Numerous object detection techniques, constructed on the foundation of potential boxes, have been created and are being used for intelligent video security purposes. The Faster R-CNN (Ren et al., 2015) based method has proved to be effective in detecting construction workers wearing helmets under remote monitoring (Fang, Li, et al., 2018). However, while the candidate frame-based

object detection algorithm exhibits high accuracy, it does not perform optimally in real-time scenarios. Consequently, numerous researchers have adopted single-stage object detectors for recognizing safety helmets and workwear.

Furthermore, YOLO algorithms (Redmon et al., 2016; Redmon & Farhadi, 2017; Redmon & Farhadi, 2018) have become popular for helmet detection. Fan et al. (2020) and Wang et al. (2020) have found ways to enhance the YOLOv3 algorithm (Redmon and Farhadi, 2018), particularly for industrial helmet detection. Wang, Wu, et al. (2021) created a high-quality dataset and used several versions of YOLO to detect six classes of objects (helmets in four colors, person and vest). This demonstrated that YOLOv5x is truly effective for PPE detection. The Covid-19 health crisis has caused a need for stricter measures to ensure everyone is wearing the necessary PPEs. This has resulted in various research studies that have developed systems based on YOLO technology, capable of detecting masks and gloves (Loey et al., 2021; Protik et al., 2021; Avanzato et al., 2020). Xie et al. (2018) carried out an evaluation of various detection models using the same datasets. The results showed that you only look once (YOLO) had the highest mean average precision (mAP) (53.8%) and the fastest speed (10 FPS), outperforming SSD and faster R-CNN in both respects.

Table 1 shows a summary of surveyed literature where machine vision has been used to detect PPE.

To this end, this article proposes a new method for utilizing YOLOv8 in PPE recognition tasks. The YOLOv8 model is employed as the main framework for PPE detection. YOLOv8 is lightweight and fast, requiring fewer computational resources compared to other models. It is an improved version of previous YOLO models in terms of speed. YOLO is used for real-time object detection. YOLOv8 is the latest object detection model developed by ultralytics.

2. Materials and methods

Managing safety measures and ensuring compliance on construction and production sites have become increasingly challenging due to the extensive workforce, complicating effective safety management and monitoring. Consequently, this study proposes an automatic PPE detection system, rooted in CV, capable of identifying various PPE types. Employing the Color Helmet and Vest (CHV) dataset – encompassing 1330 annotated images across five classes – the YOLOv8 architecture was applied for detection, outperforming other object detection models and delivering satisfactory execution times. Furthermore, the YOLOv8x model demonstrated superior mAP compared to other YOLOv8 variations.

The initial YOLOv1 algorithm, unveiled in 2016, employed a CNN for simultaneous bounding box and object class prediction, offering rapid and efficient detection but struggling with small object detection and class limitation. Its successor, YOLOv2, introduced in 2017, incorporated anchor boxes and batch normalization, enhancing algorithm performance. YOLOv3, revealed in 2018, advanced the network architecture and training methodology, utilizing three different scales for object detection to adeptly identify varying object sizes, alongside implementing the Leaky ReLU activation function and the Darknet-53 architecture. The 2020 edition, YOLOv4, brought forth further advancements, such as network architecture optimization, CSPNet and PANet utilization, and new data augmentations, ensuring notable performance and accuracy alongside swift processing. Subsequently, YOLOv5 incorporated architectural and training alterations like adopting PyTorch and introducing different model sizes (S, M, L and X) for varied performance and accuracy demands, alongside pioneering training improvements like new data augmentations and AutoAugment. Released in July 2022,

Table 1. Summary of surveyed literature.

Model	Goal
R-CNN	To detect if a worker is not wearing a hard hat (Girshick et al., 2014; Fang, Li, et al., 2018; He et al., 2017)
Fast R-CNN	To detect if a worker is not wearing a hard hat (Girshick, 2015)
Faster R-CNN	To detect if a worker is not wearing a hard hat (Ren et al., 2015, 2017; Fang, Li, et al., 2018)
SSD	To detect if a worker is not wearing a hard hat (Ferdous & Ahsan, 2022; Liu et al., 2016; Wu et al., 2019; Long et al., 2019)
FCN	To detect if a worker is not wearing a hard hat (Xie et al., 2018; Dai et al., 2016; Park et al., 2019)
CNN-LSTM	Detection of PPE (Márquez-Sánchez et al., 2021; Ding et al., 2018)
HOG	Detection of PPE compliance (Zhu et al., 2015; Park et al., 2015; Rubaiyat et al., 2016)
YOLO	Detection of PPE compliance (Redmon et al., 2016)
YOLO v2	Detection of medical PPE (Loey et al., 2021)
YOLO v3	Detection of PPE compliance (Redmon & Farhadi, 2018; Avanzato et al., 2020)
YOLO v4	Real-time detection of fire and PPE (Protik et al., 2021; Bochkovskiy et al., 2020; Kumar et al., 2022; Ma et al., 2022; Ji et al., 2023)
YOLO v5	Fast detection of PPE in construction sites (Wang, Wu, et al., 2021; Alateeq et al., 2023; Kwak & Kim, 2023)
YOLO v6	To detect PPE for construction sites (Ferdous & Ahsan, 2022)
YOLO v7	Detection safe behavior like wearing safety helmets and vests along with detecting any nonsafe activities such as not wearing protective gear is an important part of ensuring workplace safety (Shahin et al., 2023)

YOLOv7 surpassed its predecessors in both speed and accuracy across a 5 FPS to 160 FPS range, despite training solely on the MS COCO dataset without pre-trained backbone layers, and proposed numerous architectural alterations. Lastly, YOLOv8 emerges as the pinnacle of YOLO models for object detection, image classification, and instance segmentation. Developed by Ultralytics, creators of the pivotal YOLOv5, YOLOv8 introduces numerous architectural enhancements and improvements, notably prioritizing developer experience compared to YOLOv5.

2.1. Data preparing

In recent years, while sensor-based systems for detecting PPE have been developed, this study pivots toward utilizing a CV-based system, owing to its cost-effectiveness, simplicity and convenient applicability in field conditions. Utilizing the CHV dataset – originated by Wang et al. in 2021 – PPE worn by workers, inclusive of helmets and vests in four distinct colors, alongside the individuals themselves, were detected. However, the practicality of a CV alarm system extends beyond merely helmet detection; head detection also holds pivotal importance, enabling the system to identify users without helmets (Wang, Wu, et al., 2021). Given the prevalent concern regarding eye injuries among workers, detection capabilities for protective goggles and human heads were integrated into the CHV dataset, culminating in the CHVG dataset, which encompasses eight object classes. This dataset was compiled through internet image searches, significantly leaning on the prior work of Xi and Wang. Helmets, vests, and protective goggles stand out as primary PPE on construction sites, and their detection via CV can play a crucial role in safeguarding workers. The 'SHEL5K' dataset (Wang, Yeh, et al., 2021) was also developed with an aim to train and validate machine learning models for the automated detection and localization of objects in images, thereby enhancing safety on work sites. Comprising 5000 annotated images in the PASCAL VOC format, each image indicates the object class (human or safety helmet) and provides bounding box coordinates, thereby furnishing a sturdy foundation for developing and testing object detection algorithms.

Ensuring a strategic distribution of objects within the SHEL5K dataset (Figure 1) among the training, testing and validation sets is pivotal to certify that the model is trained and evaluated on a representative data sample. Ideal distribution necessitates randomness and assures representation of all classes
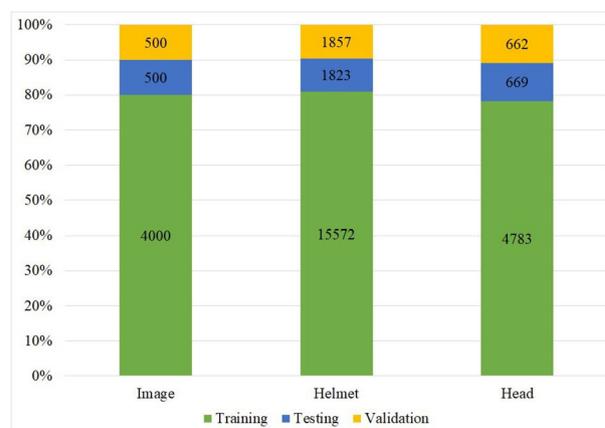

Figure 1. Object distribution into training, testing and validation set of SHEL5K.

and subclasses in each set. The training set must encompass sufficient data for model training, while the testing and validation sets should be adequately expansive to offer a reliable evaluation of the model's performance. The data were segmented as follows: 80% for training, 10% for testing and 10% for validation. This distribution was selected to afford ample images for training while simultaneously ensuring reliable model performance evaluation on unseen data.

To ensure that all classes are well-represented in the training set, an analysis of the distribution of objects by class in each set was performed (Figure 2). This is an important step in ensuring that the dataset is properly prepared for training and that the resulting model can generalize well to new data.

If one class is underrepresented in the training set, the model may not learn to recognize that class, which can lead to poor performance on the test set. Therefore, it is important to ensure a balanced distribution of objects for each class in all sets.

2.2. Experiments with YOLOv8

In order to compare different versions of the YOLOv8 architecture for PPE detection utilizing the CHV and SHEL5K datasets, we explored the following versions: YOLOv8n, YOLOv8s, YOLOv8l, YOLOv8m and YOLOv8x. When comparing these versions of YOLOv8, several metrics were utilized, including precision, recall, mAP at various Intersection over Union (IoU) thresholds, and training time. These metrics facilitate the evaluation of model performance across different classes and assist in determining which version is optimally suited for a particular project. Precision assesses the frequency with which the model accurately classifies objects, whereas recall evaluates the number of objects the model
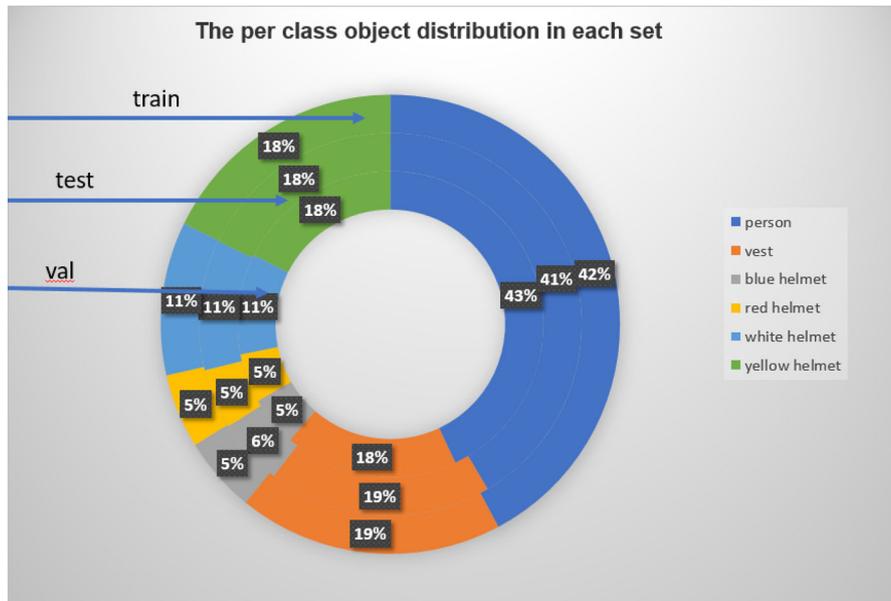
Figure 2. The per class object distribution in each CHV set.

categorizes into a given class. The mAP50 and mAP50-95 metrics illustrate the model's efficacy in recognizing objects in images at distinct IoU levels, while training time reflects the model's learning speed across a large dataset.

Table 2 provides a comparative analysis of metrics across different versions of the YOLOv8 architecture. Each row of the table presents metric values for a specific object class, along with average values across all classes. The table encompasses versions YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large) and YOLOv8x (extra-large). The results indicate that all models achieve high accuracy and recall in detecting PPE within images. However, the YOLOv8x and YOLOv8l models demonstrate superior metric values, indicative of enhanced performance.

For tasks where rapid model learning from a minimal amount of data is pivotal, YOLOv8n or YOLOv8s may be selected due to their shorter training times. However, they manifest slightly lower accuracy compared to other models. If high accuracy is paramount, then YOLOv8x or YOLOv8l might be the optimal choices as they exhibit the highest precision and mAP50 and mAP50-95 metrics, albeit with lengthier training durations. For endeavors where both training time and accuracy bear significance, YOLOv8m might be a prudent choice, offering higher accuracy than YOLOv8n and YOLOv8s, while also training more rapidly than YOLOv8x and YOLOv8l.

The confusion matrices (Figure 3) delineate the performance of various YOLOv8 models on an object

Table 2. Metrics comparison of different YOLOv8 architectures.

Object	YOLOv8n (nano)			
Class	P	R	mAP50	mAP50-95
All	0.913	0.807	0.884	0.546
Person	0.959	0.866	0.932	0.573
Vest	0.852	0.775	0.841	0.525
Blue helmet	0.947	0.731	0.857	0.513
Red helmet	0.863	0.789	0.869	0.55
White helmet	0.941	0.879	0.925	0.626
Yellow helmet	0.918	0.804	0.88	0.491
Object	**YOLOv8s (small)**			
All	0.93	0.844	0.907	0.562
Person	0969	0.88	0.938	0.588
Vest	0.884	0.813	0.88	0.531
Blue helmet	0.968	0.796	0.917	0.55
Red helmet	0.835	0.854	0.862	0.525
White helmet	0.98	0.899	0.954	0.666
Yellow helmet	0.944	0.812	0.893	0.511
Object	**YOLOv8l (large)**			
All	0.917	0.856	0.918	0.58
Person	0.928	0.872	0.933	0.954
Vest	0.876	0.789	0.874	0.549
Blue helmet	0.975	0.857	0.927	0.584
Red helmet	0.832	0.875	0.897	0.554
White helmet	0.948	0.918	0.958	0.664
Yellow helmet	0.944	0.826	0.918	0.534
Object	**YOLOv8x (x large)**			
All	0.945	0.869	0.929	0.591
Person	0.955	0.874	0.949	0.61
Vest	0.898	0.784	0.888	0.566
Blue helmet	0.999	0.878	0.935	0.588
Red helmet	0.866	0.971	0.915	0.579
White helmet	0.993	0.939	0.967	0.678
Yellow helmet	0.957	0.822	0.919	0.522
Training time	442.5 ms			
Object	**YOLOv8m (medium)**			
All	0.923	0.87	0.915	0.538
Person	0.949	0.905	0.945	0.606
Vest	0.855	0.807	0.867	0.543
Blue helmet	0.977	0.857	0.924	0.585
Red helmet	0.86	0.875	0.895	0.566
White helmet	0.959	0.934	0.964	0.671
Yellow helmet	0.939	0.84	0.897	0.519

detection task across six classes: person, vest, blue helmet, red helmet, white helmet and yellow helmet. The true labels are presented in rows, whereas the predicted labels are illustrated in columns.

Overall, the YOLOv8m (medium) and YOLOv8n (nano) models demonstrate analogous performance, each displaying an average precision of around 0.3. The YOLOv8s (small) model performs slightly inferiorly, achieving an average precision of approximately 0.2, while the YOLOv8l (large) model exhibits the best performance among all the models, with an average precision nearing 0.4. These models manifest superior performance in detecting the 'person' and 'vest' classes as compared to the various 'helmet' classes. Analyzing the confusion matrix for YOLOv8m, it is evident that the model provides the most balanced precision-recall (PR) trade-off for all classes, with the notable exception of the 'blue helmet' class. Although

the precision for this class is the highest among all models, its recall is the lowest, indicating that the model often overlooks objects within this class.

The 'precision-recall' (PR) curve, illustrated in Figure 4, facilitates choosing the optimal evaluation metric for the model. Displaying precision on the Y-axis and recall on the X-axis, the PR curve enables the determination of optimal values for both metrics, which are found in the top-right corner.

The results reveal that the YOLOv8l (large) model boasts the highest precision and recall values across all classes, suggesting that it might be the optimal model for object detection. Nevertheless, when selecting a model for a particular use case, considerations such as computational resources and model size should not be overlooked.

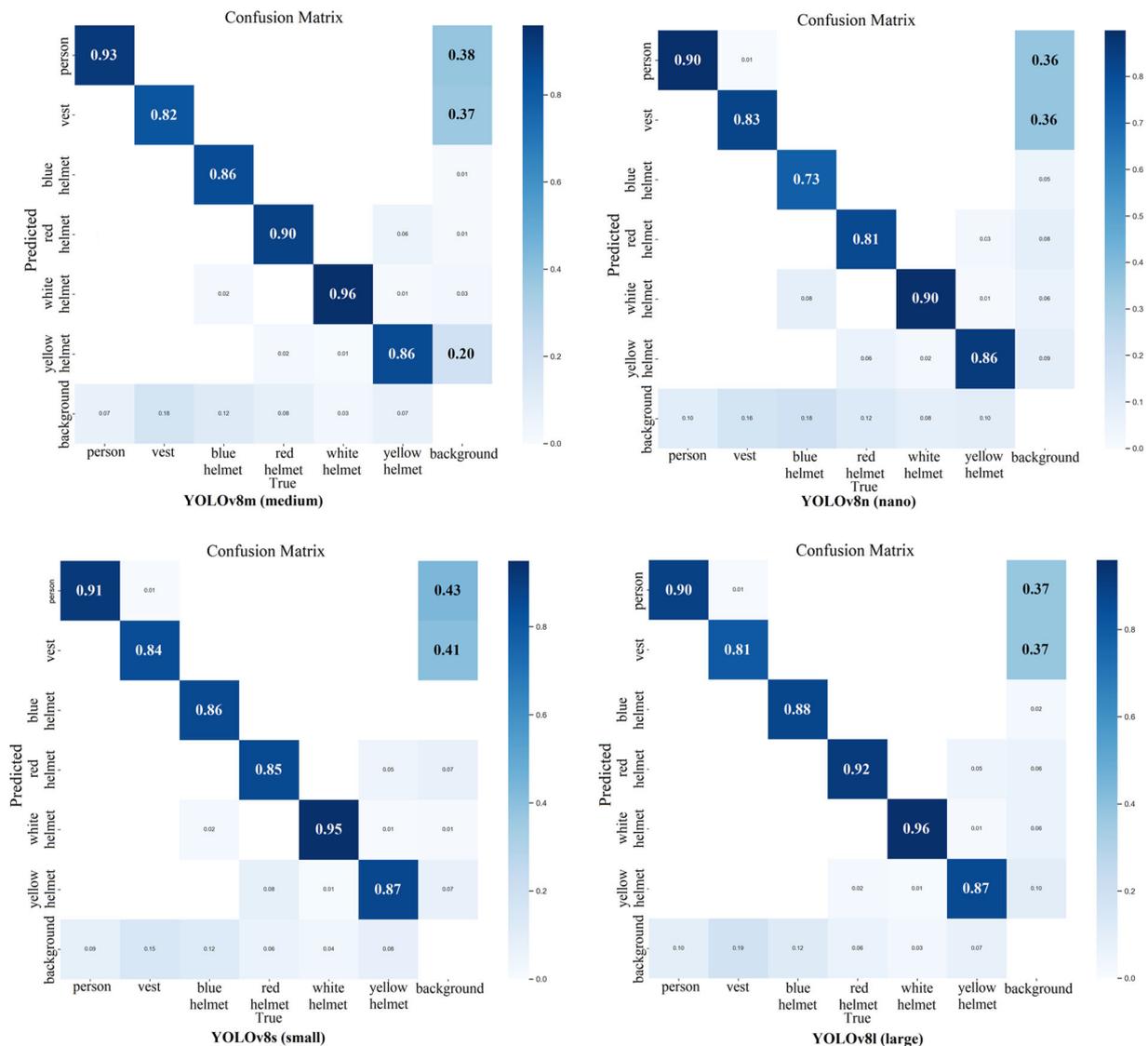Table 3 provides a comparative analysis between the YOLO models utilizing the CHV and SHEL5K

Figure 3. Performance of YOLOv8m, YOLOv8n, YOLOvV8s and YOLOvV8l models.
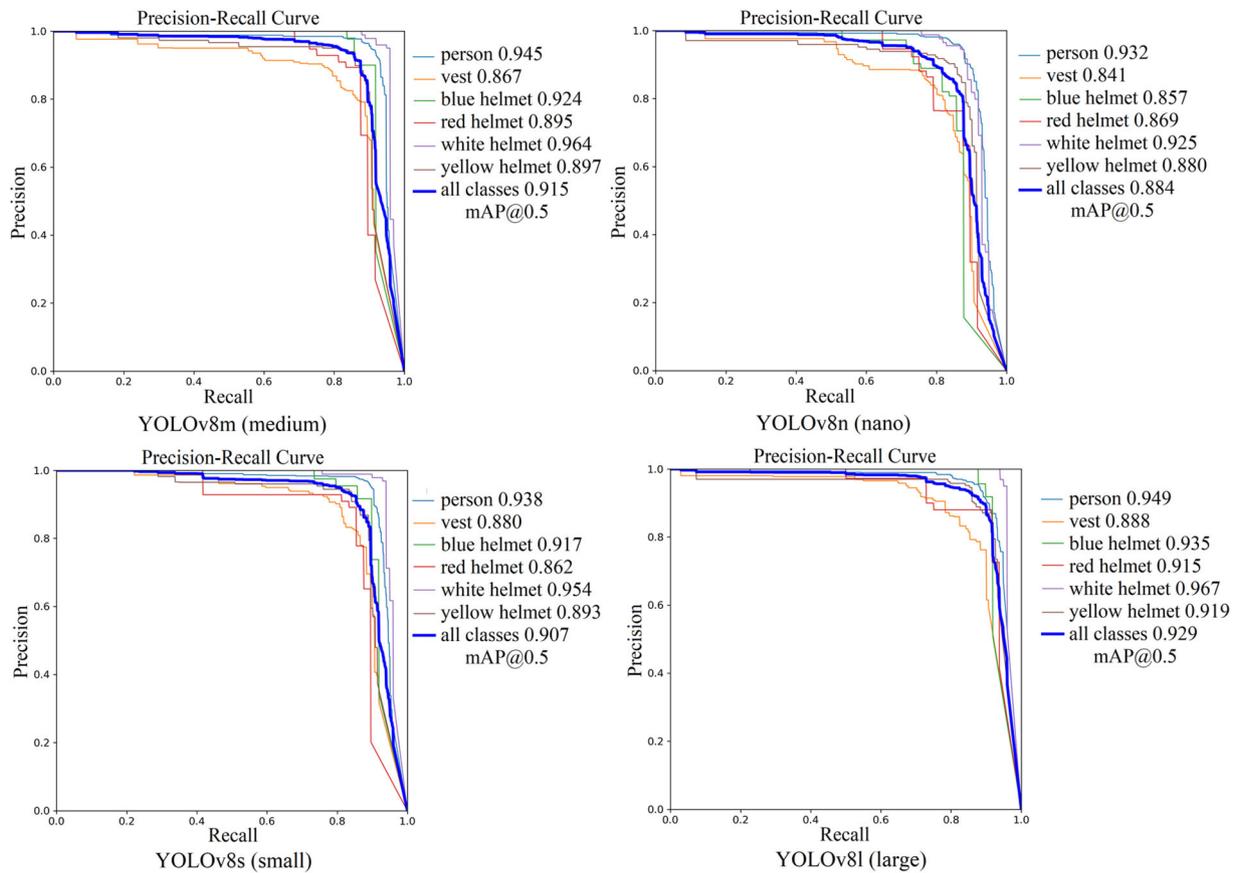
Figure 4. PR curve of YOLOvV8m, YOLOv8n, YOLOvV8s and YOLOvV8l.

Table 3. Comparison results between YOLO models.

Model performance	Precision	Recall	mAP 50
YOLOv3-tiny/SHEL5K (Otgonbold et al., 2022; Adarsh et al., 2020)	0.7695	0.4225	0.3779
YOLOv3/SHEL5K (Redmon & Farhadi, 2018; Otgonbold et al., 2022)	0.8509	0.4482	0.417
YOLOv3-SPP/SHEL5K (Zhang et al., 2020; Otgonbold et al., 2022)	0.8851	0.5848	0.5572
YOLOv4/SHEL5K (Bochkovskiy et al., 2020; Otgonbold et al., 2022)	0.925	0.7798	0.7693
YOLOv4pasp-x-mesh/SHEL5K (Bochkovskiy et al., 2020; Otgonbold et al., 2022)	0.9195	0.8036	0.7915
YOLOv5s/SHEL5K (Jocher et al., 2021; Otgonbold et al., 2022)	0.9205	0.774	0.861
YOLOv5m/SHEL5K (Jocher et al., 2021; Otgonbold et al., 2022)	0.9251	0.7851	0.8687
YOLOv5x/SHEL5K (Jocher et al., 2021; Otgonbold et al., 2022)	0.9188	0.817	0.8826
YOLOvR/SHEL5K (Otgonbold et al., 2022; Wang, Yeh, et al., 2021)	0.9322	0.8066	0.8828
YOLOv7/CHV (Shahin et al., 2023)	0.833	0.714	
YOLOv8n (nano)/SHEL5K [proposed model]	0.911	0.781	0.86
YOLOv8l (large)/SHEL5K [proposed model]	0.92	0.824	0.891
YOLOv8x (x large)/SHEL5K [proposed model]	0.92	0.832	0.896
YOLOv8m (medium)/SHEL5K [proposed model]	0.919	0.833	0.89
YOLOv8n (nano)/CHV [proposed model]	0.913	0.807	0.884
YOLOv8s (small)/CHV [proposed model]	0.93	0.844	0.907
YOLOv8l (large)/SH CHV [proposed model]	0.917	0.856	0.918
YOLOv8x (x large)/CHV [proposed model]	0.945	0.869	0.929
YOLOv8m (medium)/CHV [proposed model]	0.923	0.87	0.915

datasets. This table delineates the performance metrics of various models concerning the detection and recognition of distinct objects. The evaluation leverages precision, recall and mAP scores, which are standard metrics in object detection tasks.

Precision, recall, and mAP stand as pivotal metrics in evaluating the performance of object detection models. Precision quantifies how many of the objects identified by the model truly belong to the

target class. For instance, a high precision of 0.945 for YOLOv8x (x large) indicates that the model yields 0.065 false positives. Conversely, Recall assesses the proportion of target class objects the model detects relative to all objects of that class in the dataset; a robust Recall of 0.87 implies that the YOLOv8m (medium) model effectively identifies objects of the target class. mAP, representing the mAP across all classes and IoU thresholds, provides a comprehensive

assessment of the object detection model's quality. An evaluation score of 0.929, as demonstrated by YOLOv8m (medium), underscores its aptitude to correctly classify and precisely localize objects.

Evident from the table, each successive version of YOLO – from v3 to v8 – generally exhibits enhancements across all three metrics (precision, recall, mAP@50), signifying that algorithms and network architectures have evolved and refined with each iteration. Particularly, variants YOLOv8x (x large) and YOLOv8m (medium) on the CHV dataset showcase exemplary results across all metrics among all versions. This underscores the superior efficiency of these models in tackling object detection tasks. Figure 5 graphically represents the results articulated in Table 3.

In Figure 6, recall is employed as an auxiliary axis since its values do not surpass those in the precision column. The main line, MAP 50, showcases the fluctuations in the performance of various YOLO models. From the MAP50 curve, it is evident that each iteration of YOLO technology has brought about enhancements, progressively improving object identification within images as technology has advanced. Among the YOLO versions evaluated, our proposed model, YOLOv8x (x large), demonstrates superior performance on the CHV dataset. Specifically, it attains the highest precision of 0.945, a recall of 0.869, and an impressive mAP of 0.929 at an IoU threshold of 0.5.

To meticulously oversee the training evolution of the YOLOvl model, we strategically analyzed data pertaining to training epochs, loss values, and other pertinent metrics on the SHEL5K validation dataset.

A visual representation of this analysis is provided in Figure 6, which systematically delineates the progression observed in each epoch of the model.

Diving into the nuanced examination of the training curves, our focus navigates through the tracking of the loss trajectory spanning across all training epochs. The initial phase is characterized by an anticipated and pronounced decline in loss values. This sharp descent signifies the model's adeptness at swiftly discerning prominent patterns within the data. Subsequently, this rapid decline morphs into a more gradual deceleration of loss reduction, gradually approaching a point of stability or equilibrium.

Discerning whether and when this equilibrium is attained emerges as a pivotal aspect, fundamentally assisting in determining an apt juncture to conclude the training process. Such strategic termination of training circumvents both, unnecessary computational expenditure and the risk of inadvertently veering into overfitting.
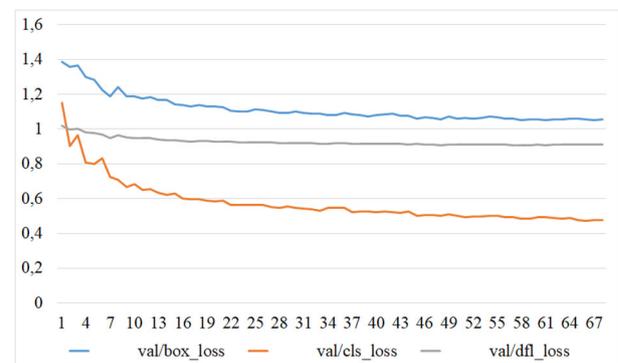

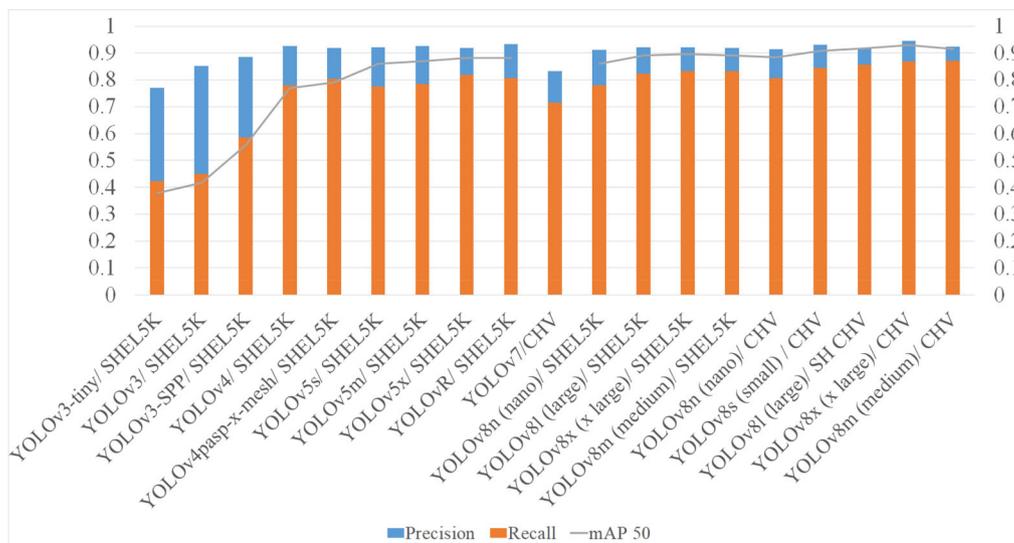**Figure 6.** Loss trends across epochs of YOLOv1 model.

Figure 5. Comparison mean average precision 50 between models.

In a further breakdown of the losses – explicitly categorized as val/box_loss, val/cls_loss and val/dfl_loss – on the validation dataset for object detection, a consistent diminishing trend becomes apparent throughout the training process. Nearing the culmination of training, these losses ostensibly stabilize or plateau, which can be inferred as an indicator of the model's iterative improvements and refinements throughout the learning process.

3. Discussion

This article presents a comparison of various versions of the YOLOv8 architecture for detecting PPE using the CHV dataset. The comparative analysis utilizes metrics such as precision, recall, and mAP at different IoU thresholds, along with training time.

The results illustrate that all models attain high precision and recall in detecting PPE in images. Nonetheless, the YOLOv8x and YOLOv8l models exhibit superior metric values, signifying enhanced performance.

The YOLOv8n (nano) model yields the lowest precision and recall values across all classes, indicating potential unsuitability for more intricate object detection tasks. The YOLOv8m (medium) and YOLOv8s (small) models showcase similar performance; however, YOLOv8m (medium) slightly edges out in precision and recall values for most classes.

In an effort to boost model performance, object distribution within the image was derived for the YOLOv8m (medium) model for object detection, and the optimal location for the bounding box around each object was determined. Names of the object classes to be detected and recognized in the image are presented as labels in the figure. Object distribution within the image is illustrated as a histogram, where each column represents one class of objects, and each row signifies an image divided into cells, as demonstrated in the figure. A darker column indicates a higher count of objects of that class situated in the corresponding cells of the image (refer to Figure 7).

Selecting the optimal model hinges on specific project requirements, necessitating a thorough evaluation to determine which model aligns best with set criteria. If rapid learning from a minimal data volume is imperative, YOLOv8n or YOLOv8s may be optimal choices due to their abbreviated training times. Conversely, if high precision is prioritized, YOLOv8x or YOLOv8l may emerge as the top choices, exhibiting the highest precision alongside mAP50 and mAP50-95 metrics, albeit with extended training durations. When both precision and training time are pivotal, YOLOv8m might be the model of choice, boasting superior precision compared to YOLOv8n and YOLOv8s while also achieving faster training than YOLOv8x and YOLOv8l.

Further, the confusion matrix for YOLOv8m was scrutinized. It was observed that the model proffers the most favorable PR ratio for all classes, with 'blue helmet' being an exception. Although the precision for this class surpasses that of all other models, its recall is the lowest, indicating a tendency for the model to miss objects within this class. The research findings can assist in determining the optimal YOLOv8 model for PPE detection tasks on the CHV dataset and offer insights into model selection based on specific project prerequisites.
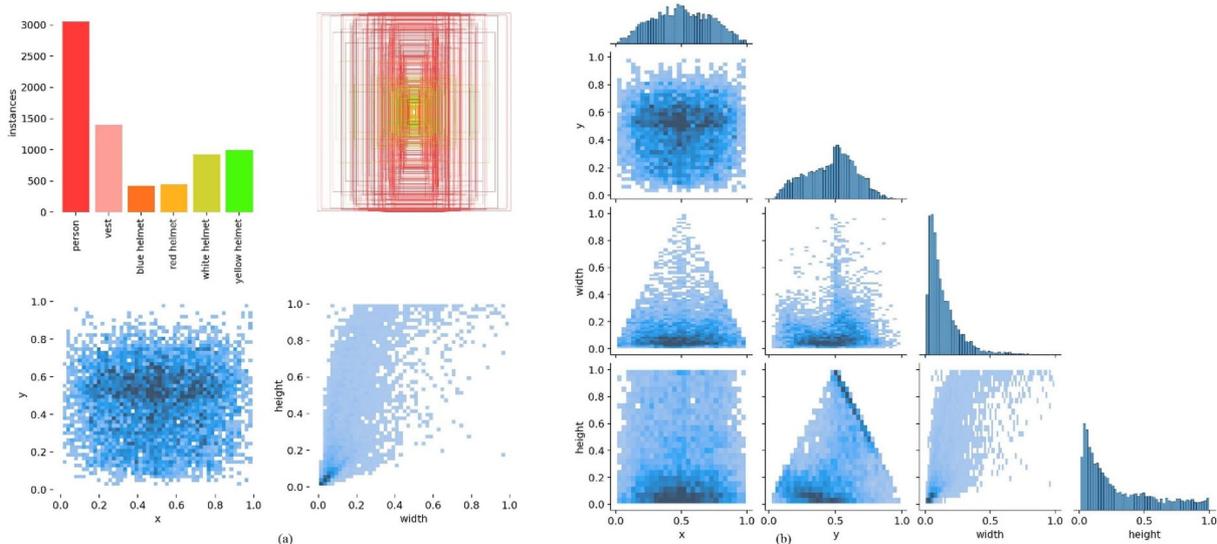

Figure 7. CHV dataset: (a) labels and (b) correlogram.

4. Conclusion

In conclusion, the findings from our study underscore the robust efficiency of the YOLOv8 architecture in detecting PPE in images, with all models showcasing high accuracy and recall. Notably, the YOLOv8l (large) model demonstrated superior performance across all models, achieving the highest precision and mAP50 and mAP50-95 metrics, albeit with extended training periods. The YOLOv8n (nano) and YOLOv8s (small) models emerge as suitable choices for tasks necessitating swift learning from limited data, despite presenting slightly diminished accuracy compared to other models. YOLOv8m (medium), offering a balance between training duration and accuracy, stands out as an optimal choice for scenarios where both factors are pivotal, as it surpasses YOLOv8n and YOLOv8s in accuracy and trains more rapidly than YOLOv8x and YOLOv8l.

Figures 8 and 9 illustrate several commendable results from the YOLOv8m (medium) architecture. It is observable that the objects, even while in a densely packed state and thus occluded, are correctly detected. The objects, depicted in natural working poses such as kneeling and spine-bending at various angles, are accurately identified by the YOLOv8m (medium) model, among others. The comparative results of all models are accessible on GitHub (https://github.com/NurzadaEnu/Personal-Protective-Equipment-Detection-using-YOLOv8).

Our examination of the distribution of objects per class in each set underscores the imperative of maintaining a balanced distribution across all classes in every set to forestall suboptimal performance on the test set. The confusion matrices facilitated an evaluation of the performance of various YOLOv8 models in an object detection task across six classes. The findings illuminated that the models exhibited superior performance in detecting 'person' and 'vest' classes in contrast to helmet classes. A more granulated analysis of the confusion matrix for YOLOv8m disclosed that the model offers the most advantageous PR trade-off for all classes, with the exception of 'blue helmet'.

In a bid to enhance the performance of the YOLOv8m model, we ascertained the distribution of objects within the image and determined the optimal location for the bounding box around each object. The resultant histogram depicts the distribution of objects in the image, where darker columns denote a higher concentration of objects of that class in the corresponding image cells.

Our study proffers invaluable insights into the performance of diverse YOLOv8 models in the detection

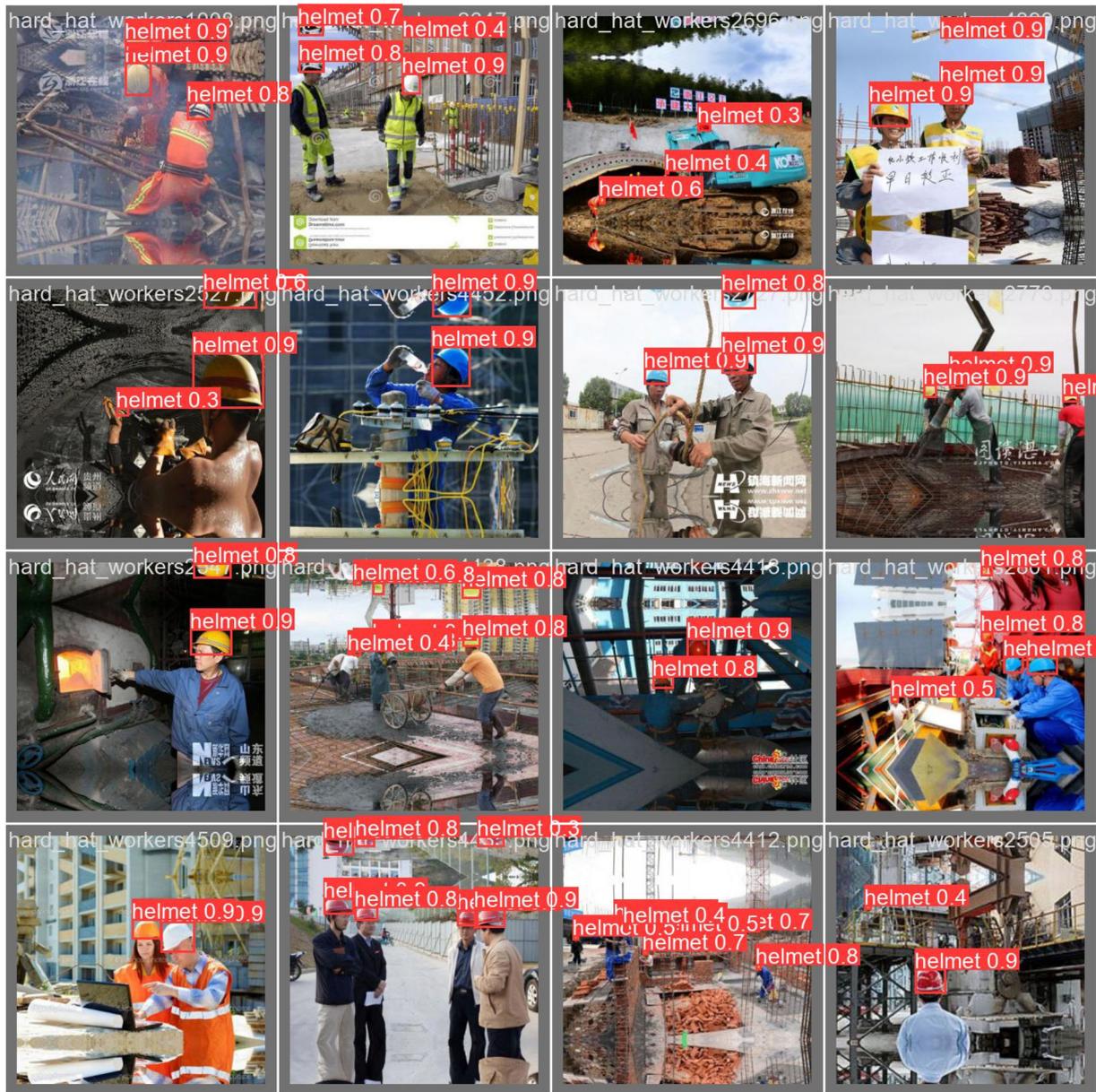Figure 8. Several satisfactory results of the YOLOvV8m/CHV.

Figure 9. Several satisfactory results of the YOLOvV8m/SHELK5K.

of PPE and underlines the criticality of adeptly preparing datasets for model training and evaluation. The insights derived from our findings can serve to inform the selection of the optimal model for specific use cases, cognizant of considerations such as computational resources and model size.

4.1. Future works

In the planning phase for advancing a methodology concerning safety knowledge modeling, particularly considering workers' PPE, our intention is to cultivate a framework that meticulously ensures the automated assignment of safety attributes based on PPE detection through YOLO, integrates pivotal monitoring rules to ensure on-site worker safety, and facilitates the analysis of PPE data. Moreover, it aims at evaluating safety compliance, issuing timely alerts and notifications upon safety violations, and conducting real-time monitoring with prompt responsive actions to mitigate any arising risks. As a crucial aspect of our forthcoming work, we also plan to gather data directly from actual construction sites, which will substantively ground our research and development, providing tangible, real-world insights and augmenting the practical applicability and efficacy of our proposed methodology.

Disclosure statement

The authors declare that they have no conflicts of interest to report regarding this study.

About the authors

Nurzada Amangeldy a postdoctoral fellow at L.N. Gumilyov Eurasian National University, specializes in object recognition, computer vision and artificial intelligence. Her leadership in spearheading several key projects, including the creation of the Kazakh sign language recognition system, underscores her innovative approach. Her involvement in the study of detecting personal protective equipment using the YOLOv8 architecture points to her commitment to elevating safety standards through cutting-edge technologies. Set against a backdrop of global trends, Nurzada's work addresses the pressing need for automated safety solutions, making her contributions timely and relevant. Her research endeavors not only foster technical advancements in the field but also address pressing contemporary challenges.

Alibek Barlybayev received the B.Eng. degree in information systems from L.N. Gumilyov Eurasian National University, Kazakhstan, in 2009 and the M.S. and Ph.D. degrees in computer science from L.N. Gumilyov Eurasian National University, Kazakhstan, in 2011 and 2015, respectively. Currently, he is a Director of the Research Institute of Artificial Intelligence, L.N. Gumilyov Eurasian National University. He is also an Associate Professor of the Department of Artificial Intelligence Technologies, L.N. Gumilyov Eurasian National University, and Higher School of Information Technology and Engineering, Astana International University. His research interests are NLP, the use of neural networks in word processing, smart textbooks, fuzzy logic, stock market price prediction, information security. He can be contacted at email: frank-ab@mail.ru

ORCID

Alibek Barlybayev http://orcid.org/0000-0002-0188-5336
Nurzada Amangeldy http://orcid.org/0000-0002-4669-9254
Bekbolat Kurmetbek http://orcid.org/0009-0001-7510-2445
Iurii Krak http://orcid.org/0000-0002-8043-0785
Bibigul Razakhova http://orcid.org/0000-0002-8152-8661
Nazira Tursynova http://orcid.org/0000-0002-1857-9028
Rakhila Turebayeva http://orcid.org/0009-0006-4530-1300

References

Adarsh, P., Rathi, P., & Kumar, M. (2020, March). *YOLO v3-Tiny: Object detection and recognition using one stage improved model* [Paper presentation]. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 687–694). IEEE. https://doi.org/10.1109/ICACCS48705.2020.9074315

Akbarzadeh, M., Zhu, Z., & Hammad, A. (2020) *Nested network for detecting PPE on large construction sites based on frame segmentation* [Paper presentation]. Creative Construction e-Conference 2020 (pp. 33–38). Budapest University of Technology and Economics.

Alateeq, M. M., Pp, F. R., & Ali, M. A. (2023). Construction site hazards identification using deep learning and computer vision. *Sustainability*, *15*(3), 2358. https://doi.org/10.3390/su15032358

Avanzato, R., Beritelli, F., Russo, M., Russo, S., & Vaccaro, M. (2020). Yolov3-based mask and face recognition algorithm for individual protection applications. *CEUR Workshop Proceedings*. Volume 2768 (pp. 41–45). https://www.scopus.com/record/display.uri?eid=2-s2.0-85097903936&origin=inward&txGid=79805e76b928adf2df3aa196caa0bcd1

Barro-Torres, S., Fernández-Caramés, T. M., Pérez-Iglesias, H. J., & Escudero, C. J. (2012). Real-time personal protective equipment monitoring system. *Computer Communications*, *36*(1), 42–50. https://doi.org/10.1016/j.comcom.2012.01.005

Bo, Y., Huan, Q., Huan, X., Rong, Z., Hongbin, L., Kebin, M., … Lei, Z. (2019, October). *Helmet detection under the power construction scene based on image analysis* [Paper presentation]. 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT) (pp. 67–71). IEEE. https://doi.org/10.1109/ICCSNT47585.2019.8962495

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. https://doi.org/10.48550/arXiv.2004.10934

Cai, L., & Qian, J. (2011). A method for detecting miners based on helmets detection in underground coal mine videos. *Mining Science and Technology (China)*, *21*(4), 553–556. https://doi.org/10.1016/j.mstc.2011.06.016

Dai, J., Li, Y., He, K., & Sun, J. (2016). R-fcn: Object detection via region-based fully convolutional networks. *Advances in neural information processing systems* (p. 29). https://doi.org/10.48550/arXiv.1605.06409

Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 1, pp. 886–893). IEEE. DOI: 10.1109/CVPR.2005.177

Ding, L., Fang, W., Luo, H., Love, P. E., Zhong, B., & Ouyang, X. (2018). A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Automation in Construction*, *86*, 118–124. https://doi.org/10.1016/j.autcon.2017.11.002

Dong, S., He, Q., Li, H., & Yin, Q. (2015). *Automated PPE misuse identification and assessment for safety performance enhancement* [Paper presentation]. ICCREM 2015 (pp. 204–214). https://doi.org/10.1061/9780784479377.024

Fan, Z., Peng, C., Dai, L., Cao, F., Qi, J., & Hua, W. (2020). A deep learning-based ensemble method for helmet-wearing detection. *PeerJ. Computer Science*, *6*, e311. https://doi.org/10.7717/peerj-cs.311

Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M., & An, W. (2018). Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in*

Construction, *85*, 1–9. https://doi.org/10.1016/j.autcon.2017.09.018

Fang, W., Ding, L., Luo, H., & Love, P. E. (2018). Falls from heights: A computer vision-based approach for safety harness detection. *Automation in Construction*, *91*, 53–61. https://doi.org/10.1016/j.autcon.2018.02.018

Ferdous, M., & Ahsan, S. M. M. (2022). PPE detector: A YOLO-based architecture to detect personal protective equipment (PPE) for construction sites. *PeerJ Computer Science*, *8*, e999. https://doi.org/10.7717/peerj-cs.999

Gallo, G., Di Rienzo, F., Garzelli, F., Ducange, P., & Vallati, C. (2022). A smart system for personal protective equipment detection in industrial environments based on deep learning at the edge. *IEEE Access 10*, 110862–110878. https://doi.org/10.1109/ACCESS.2022.3215148

Girshick, R. (2015). Fast r-cnn. *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440–1448). IEEE.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580–587). IEEE.

Han, G., Zhu, M., Zhao, X., & Gao, H. (2021). Method based on the cross-layer attention mechanism and multiscale perception for safety helmet-wearing detection. *Computers and Electrical Engineering*, *95*, 107458. https://doi.org/10.1016/j.compeleceng.2021.107458

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2961–2969). IEEE.

https://kz.kursiv.media/2022-01-29/v-kazakhstane-rabochie-kalechatsya-na-proizvodstve-v-tri-raza-chasche-chem/.

Huang, L., Fu, Q., He, M., Jiang, D., & Hao, Z. (2021). Detection algorithm of safety helmet wearing based on deep learning. *Concurrency and Computation: Practice and Experience*, *33*(13), e6234. https://doi.org/10.1002/cpe.6234

Iannizzotto, G., Bello, L. L., & Patti, G. (2021). Personal protection equipment detection system for embedded devices based on DNN and fuzzy logic. *Expert Systems with Applications*, *184*, 115447. https://doi.org/10.1016/j.eswa.2021.115447

Ji, X., Gong, F., Yuan, X., & Wang, N. (2023). A high-performance framework for personal protective equipment detection on the offshore drilling platform. *Complex & Intelligent Systems*, *9*(5), 5637–5652. https://doi.org/10.1007/s40747-023-01028-0

Jocher, G., Stoken, A., Borovec, J., Changyu, L., Hogan, A., Chaurasia, A., Abhiram, V., Hogan, A., Hajek, J., Diaconu, L., Kwon, Y., Defretin, Y., Lohia, A., Milanko, B., Fineran, B., Khromov, D., Yiwei, D., Ingham, F. (2021). *ultralytics/yolov5: v4. 0-nn. SiLU activations, Weights & Biases logging, PyTorch Hub integration*. Zenodo.

Kelm, A., Laußat, L., Meins-Becker, A., Platz, D., Khazaee, M. J., Costin, A. M., Helmus, M., & Teizer, J. (2013). Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites. *Automation in Construction*, *36*, 38–52. https://doi.org/10.1016/j.autcon.2013.08.009

Kumar, S., Gupta, H., Yadav, D., Ansari, I. A., & Verma, O. P. (2022). YOLOv4 algorithm for the real-time detection of fire and personal protective equipments at construction sites. *Multimedia Tools and Applications*, *81*(16), 22163–22183. https://doi.org/10.1007/s11042-021-11280-6

Kwak, N., & Kim, D. (2023). Detection of worker's safety helmet and mask and identification of worker using deeplearning. *Computers, Materials & Continua*, *75*(1), 1671–1686. https://doi.org/10.32604/cmc.2023.035762

Li, J., Liu, H., Wang, T., Jiang, M., Wang, S., Li, K., & Zhao, X. (2017, February). *Safety helmet wearing detection based on image processing and machine learning* [Paper presentation]. 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI) (pp. 201–205). IEEE. https://doi.org/10.1109/ICACI.2017.7974509

Li, X., He, M., Liu, Y., Luo, H., & Ju, M. (2023). SPCS: A spatial pyramid convolutional shuffle module for YOLO to detect occluded object. *Complex & Intelligent Systems*, *9*(1), 301–315. https://doi.org/10.1007/s40747-022-00786-7

Lienhart, R., & Maydt, J. (2002, September). An extended set of haar-like features for rapid object detection. *Proceedings International Conference on Image Processing* (Vol. 1, pp. I–I). IEEE.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). *SSD: Single shot multibox detector* [Paper presentation]. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21–37). Springer International Publishing.

Liu, Y., & Zheng, Y. F. (2005, July). One-against-all multi-class SVM classification using reliability measures. *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005* (Vol. 2, pp. 849–854.). IEEE.

Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustainable Cities and Society*, *65*, 102600. https://doi.org/10.1016/j.scs.2020.102600

Long, X., Cui, W., & Zheng, Z. (2019, March). *Safety helmet wearing detection based on deep learning* [Paper presentation]. 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (pp. 2495–2499). IEEE. https://doi.org/10.1109/ITNEC.2019.8729039

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Ma, L., Li, X., Dai, X., Guan, Z., & Lu, Y. (2022). A combined detection algorithm for personal protective equipment based on lightweight YOLOv4 model. *Wireless Communications and Mobile Computing*, *2022*, 1–11. https://doi.org/10.1155/2022/3574588

Márquez-Sánchez, S., Campero-Jurado, I., Herrera-Santos, J., Rodríguez, S., & Corchado, J. M. (2021). Intelligent platform based on smart PPE for safety in workplaces. *Sensors*, *21*(14), 4652. https://doi.org/10.3390/s21144652

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. https://doi.org/10.1016/j.autcon.2020.103085

Otgonbold, M. E., Gochoo, M., Alnajjar, F., Ali, L., Tan, T. H., Hsieh, J. W., & Chen, P. Y. (2022). SHEL5K: An extended dataset and benchmarking for safety helmet detection. *Sensors*, *22*(6), 2315. https://doi.org/10.3390/s22062315

Park, M. W., & Brilakis, I. (2012). Construction worker detection in video frames for initializing vision trackers. *Automation in Construction*, 28, 15–25. https://doi.org/10.1016/j.autcon.2012.06.001

Park, M. W., Elsafty, N., & Zhu, Z. (2015). Hardhat-wearing detection for enhancing on-site safety of construction workers. *Journal of Construction Engineering and Management*, 141(9), 04015024. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000974

Park, S., Yoon, S., & Heo, J. (2019). Image-based automatic detection of construction helmets using R-FCN and transfer learning. *KSCE Journal of Civil and Environmental Engineering Research*, 39(3), 399–407. https://doi.org/10.12652/Ksce.2019.39.3.0399

Pradana, R. D. W., Adhitya, R. Y., Syai'in, M., Sudibyo, R. M., Abiyoga, D. R. A., Jami'in, M. A., Rochiem, N. H. (2019, October). *MIdentification system of personal protective equipment using Convolutional Neural Network (CNN) method* [Paper presentation]. 2019 International Symposium on Electronics and Smart Devices (ISESD) (pp. 1–6). IEEE. https://doi.org/10.1109/ISESD.2019.8909629

Protik, A. A., Rafi, A. H., & Siddique, S. (2021, August). *Real-time personal protective equipment (PPE) detection using YOLOv4 and TensorFlow* [Paper presentation]. 2021 IEEE Region 10 Symposium (TENSYMP) (pp. 1–6). IEEE. https://doi.org/10.1109/TENSYMP52854.2021.9550808

Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7263–7271). IEEE. DOI: 10.1109/CVPR.2017.690

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779–788). IEEE.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems* (p. 28). https://doi.org/10.48550/arXiv.1506.01497

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards realtime object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Rubaiyat, A. H., Toma, T. T., Kalantari-Khandani, M., Rahman, S. A., Chen, L., Ye, Y., & Pan, C. S. (2016, October). *Automatic detection of helmet uses for construction safety* [Paper presentation]. 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW) (pp. 135–142). IEEE. https://doi.org/10.1109/WIW.2016.045

Seo, J., Han, S., Lee, S., & Kim, H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, 29(2), 239–251. https://doi.org/10.1016/j.aei.2015.02.001

Shahin, M., Chen, F. F., Hosseinzadeh, A., Khodadadi Koodiani, H., Bouzary, H., & Shahin, A. (2023). Enhanced safety implementation in 5S+ 1 via object detection algorithms. *The International Journal of Advanced Manufacturing Technology*, 125(7–8), 3701–3721. https://doi.org/10.1007/s00170-023-10970-9

Shen, J., Xiong, X., Li, Y., He, W., Li, P., & Zheng, X. (2021). Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning. *Computer-Aided Civil and Infrastructure Engineering*, 36(2), 180–196. https://doi.org/10.1111/mice.12579

Stojanovic, V., He, S., & Zhang, B. (2020). State and parameter joint estimation of linear stochastic systems in presence of faults and non-Gaussian noises. *International Journal of Robust and Nonlinear Control*, 30(16), 6683–6700. https://doi.org/10.1002/rnc.5131

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. https://doi.org/10.48550/arXiv.2207.02696

Wang, C. Y., Yeh, I. H., & Liao, H. Y. M. (2021). You only learn one representation: Unified network for multiple tasks. https://doi.org/10.48550/arXiv.2105.04206

Wang, H., Hu, Z., Guo, Y., Yang, Z., Zhou, F., & Xu, P. (2020). A real-time safety helmet wearing detection approach based on CSYOLOv3. *Applied Sciences*, 10(19), 6732. https://doi.org/10.3390/app10196732

Wang, Z., Wu, Y., Yang, L., Thirunavukarasu, A., Evison, C., & Zhao, Y. (2021). Fast personal protective equipment detection for real construction sites using deep learning approaches. *Sensors*, 21(10), 3478. https://doi.org/10.3390/s21103478

Wu, H., & Zhao, J. (2018). Automated visual helmet identification based on deep convolutional neural networks. *Computer aided chemical engineering* (Vol. 44, pp. 2299–2304). Elsevier.

Wu, J., Cai, N., Chen, W., Wang, H., & Wang, G. (2019). Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset. *Automation in Construction*, 106, 102894. https://doi.org/10.1016/j.autcon.2019.102894

Xie, Z., Liu, H., Li, Z., & He, Y. (2018, December). *A convolutional neural network based approach towards real-time hard hat detection* [Paper presentation]. 2018 IEEE International Conference on Progress in Informatics and Computing (PIC) (pp. 430–434). IEEE. https://doi.org/10.1109/PIC.2018.8706269

Xiong, R., & Tang, P. (2021). Pose guided anchoring for detecting proper use of personal protective equipment. *Automation in Construction*, 130, 103828. https://doi.org/10.1016/j.autcon.2021.103828

Zhang, S., Teizer, J., Pradhananga, N., & Eastman, C. M. (2015). Workforce location tracking to model, visualize and analyze workspace requirements in building information models for construction safety planning. *Automation in Construction*, 60, 74–86. https://doi.org/10.1016/j.autcon.2015.09.009

Zhang, X., Gao, Y., Wang, H., & Wang, Q. (2020). Improve YOLOv3 using dilated spatial pyramid module for multi-scale object detection. *International Journal of Advanced Robotic Systems*, 17(4), 172988142093606. 1729881420936062. https://doi.org/10.1177/1729881420936062

Zhu, Z., Park, M. W., & Elsafty, N. (2015 *Automated monitoring of hardhats wearing for onsite safety enhancement* [Paper presentation]. Proceedings of the 11th Construction Specialty Conference, Vancouver, UK, 8–10 June 2015 (pp. 1–9).