Article

Development of YOLOv5-Based Real-Time Smart Monitoring System for Increasing Lab Safety Awareness in Educational Institutions

Luqman Ali 1,2,3, Fady Alnajjar 1,3, Medha Mohan Ambali Parambil 1,3, Mohammad Issam Younes 4, Ziad Ismail Abdelhalim 4 and Hamad Aljassmi 2,4,*

1 Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University (UAEU), Al Ain 15551, United Arab Emirates
2 Emirates Center for Mobility Research, United Arab Emirates University (UAEU), Al Ain 15551, United Arab Emirates
3 AI and Robotics Lab (Air-Lab), United Arab Emirates University (UAEU), Al Ain 15551, United Arab Emirates
4 Department of Civil Engineering, College of Engineering, United Arab Emirates University (UAEU), Al Ain 15551, United Arab Emirates
* Correspondence: h.aljasmi@uaeu.ac.ae

Abstract: The term “smart lab” refers to a system that provides a novel and flexible approach to automating and connecting current laboratory processes. In education, laboratory safety is an essential component of undergraduate laboratory classes. The institution provides formal training for the students working in the labs that involve potential exposure to a wide range of hazards, including chemical, biological, and physical agents. During the laboratory safety lessons, the instructor explains the lab safety protocols and the use of personal protective equipment (PPE) to prevent unwanted accidents. However, it is not always guaranteed that students follow safety procedures throughout all lab sessions. Consequently, students may unintentionally commit unrecognized unsafe acts, which can lead to unwanted situations. Therefore, the aim of the research article was to propose a real-time smart vision-based lab-safety monitoring system to verify the PPE compliance of students, i.e., whether the student is wearing a mask, gloves, lab coat, and goggles, from image/video in real time. The YOLOv5 (YOLOv5l, YOLOv5m, YOLOv5s, and YOLOv5x) and YOLOv7 models were trained using a self-created novel dataset named SLS (Students Lab Safety). The dataset comprises four classes, namely, gloves, helmets, masks, and goggles, and 481 images, having a resolution of 835 × 1000, acquired from various research laboratories of the United Arab Emirates University. The performance of the different YOLOv5 and YOLOv7 versions is compared based on instances’ size using evaluation metrics such as precision, F1 score, recall, and mAP (mean average precision). The experimental results demonstrated that all the models showed promising performance in detecting PPE in educational labs. The YOLOv5n approach achieved the highest mAP of 77.40% for small and large instances, followed by the YOLOv5m model having a mAP of 75.30%. A report detailing each student’s PPE compliance in the lab can be prepared based on data collected in real time and stored in the proposed system. Overall, the proposed approach can be utilized to make laboratories smarter by enhancing the efficacy of safety in research settings; this, in turn, will aid the students in establishing a health and safety culture among students.

Keywords: IoT; smart academic laboratories; safety; deep learning; object detection; YOLOv5; PPE compliance

1. Introduction

Regular classroom instructions and academic research are facilitated by labs at institutions, which are responsible for training future scientists and uncovering the mysteries of
nature [1]. Several recent incidents in university laboratories have prompted an increased
focus on laboratory safety [2–5]. Universities and science departments have taken various
initiatives, such as performing multiple training sessions and safety plans and providing
online information to ensure the safety protocols in the lab environment. Maintaining
active involvement in safety training by principal investigators (PIs) or lab supervisors is
critical to fostering good attitudes toward academic laboratory safety. Traditionally, the
lab instructors arrange training sessions for newcomers to maintain a safe learning and
working environment for their students. In the training sessions, the lab instructors guide
the students about the lab safety protocol and the use of PPE. The personal protective
equipment (PPE), which includes safety eye wear, lab coat, gloves, and mask, provides
a direct protective layer to the students. The use of PPE in the laboratories indicates the
extent to which students are following the safety policies of the institution. Previous studies
showed that PPE compliance by researchers in academia was positively influenced when
their safety behavior was monitored [6].

Traditionally, the PPE is manually monitored by lab supervisors; the monitoring is
expensive, time consuming, and resource intensive. The instructors cannot keep a vigilant
eye on every student and track their movements during the lab sessions to ensure PPE
compliance. Additionally, it is challenging to fulfill the requirements of modern educa-
tional safety management by solely relying on manual monitoring approaches. To bridge
these problems, automatic vision-based deep learning monitoring and detection techniques
provide the solution. These approaches have shown promising performance in tackling
the challenge of accurate safety monitoring and hazard detection problems in various
applications [7–14]. Most studies focus on detecting and monitoring PPE compliance for
workers’ safety in the construction industry [15–17]. Several different industries, such as
construction, mining, and energy, have started to make investments in order to improve
the safety of their workers by incorporating new technology, sometimes known as “smart
technologies”, into the workplace environment. These technologies are in charge of keeping
an eye on workers and ensuring their safety in the workplace. Wu et al. [18] used the
Single Shot MultiBox Detector (SSD) [19] for the detection of construction workers’ helmets
and their associated colors. Fang et al. [20] proposed an end-to-end Faster RCNN [21]
-based approach for the non-hard-hat-use (NHU) detection in raw videos. Similarly,
Saudi et al. [22] used the Faster RCNN method for detecting multiple PPEs such as helmets
and vests for the safety of the worker in construction sites.

The YOLO architecture is becoming increasingly popular because of its speed and
precision in the recognition of objects from images. Nath et al. [23] proposed a real-time
You-Only-Look-Once (YOLO) architecture for the verification of PPE compliance of workers,
i.e., whether or not a worker was wearing a hard hat, vest, or both. Moreover, human
identity recognition and helmet detection was performed using YOLOv3 architecture in [24].
Wang et al. [25] compared the performance of various architectures of the YOLO family
(YOLOv3 [16], YOLOv4 [26], and YOLOv5 [27]) on a custom dataset, named as the CHV
dataset, and found that YOLOv5x had a superior performance as compared to other models.
Among all real-time object detectors with 30 FPS or higher on GPU V100, OLOv7 had the
best accuracy (56.8% AP) and was the fastest (up to 160 FPS) [28]. To the authors’ knowledge,
minimal efforts were made to use these object detection models for the detection of PPE
in educational laboratories. Therefore, this paper proposes a YOLOv5-based real-time PPE
compliance detection and monitoring system for academic laboratories using a custom dataset.
The main aim of the system is to create a reliable and real-time automated smart safety
detection system by having early warning systems, which will create a safety culture in the
institutional lab, enhance the lab safety awareness, and eliminate the occurrence of various
unwanted incidents, as depicted in Figure 1. The data of the student regarding the student
will be sent to the system and a safety report for the student will be generated. The system
can also be set up at the entrance door to the laboratory, allowing only properly geared-up
personnel access. Students are not allowed inside the lab and are informed of their error if
they are found to be lacking proper PPE. The main aims of the proposed study are:
(1) The creation of a novel, labeled PPE dataset named SLS (Student Lab Safety) containing four different classes, including mask, lab coat, safety glass, and gloves. The dataset contains 481 images and the corresponding annotations of these four classes.

(2) The performance evaluation of various versions of the YOLOv5 [27] (YOLOv5l, YOLOv5m, YOLOv5n, and YOLOv5x) and YOLOv7 (YOLOv7 and YOLOv7X) on the proposed dataset for the detection and monitoring of students’ PPE in academic laboratories.

(3) The performance evaluation of the YOLOv5 and YOLOv7 model variant based on instance size of the object, i.e., large instances (lab coat and gloves) and small instances (masks and goggles).

Figure 1. Role of safety culture in reducing incidents’ rate.

2. System Overview

The overview of the proposed system is depicted in Figure 2: (1) SLS dataset Preparation, (2) training of the proposed YOLOv5 and YOLOv7 frameworks, and (3) testing of the proposed system in real-time environments. The system takes images/videos as input and analyzes them using the trained versions of the YOLOv5 and YOLOv7 models. The output image/video from the system contains different detected classes, i.e., gloves, lab coat, goggles, and mask. Each section of the proposed system is explained in detail below.

Figure 2. Overview of the proposed system (correct the arrows).

2.1. Student Laboratory Safety (SLS) Dataset

In the proposed study, a novel dataset named SLS was created for PPE detection as there is currently no publicly accessible dataset for the detection of PPE in educational labs. The images were acquired from students working in research laboratories of the United
Arab Emirates University using a handheld Canon EOS 650D (40 mm) and surveillance (CCTV 2.0 Dome Camera PoE Onvif conformant) camera installed in the lab. The images were taken from various distances and viewpoints of students working in lab surroundings including the top camera installed in the lab. In order to protect the identity of the students working in the laboratory, their faces were obscured. The dataset contained 481 images having a resolution of 1600 × 1200 pixels collected from the handheld and surveillance cameras installed in the lab. The activities of the students were divided into two categories: those who followed PPE compliance were marked safe and those who did not were marked unsafe. The number of images in both categories was kept equal. After data acquisition, the images were manually labeled using the graphical images annotation tool LabelImg [29].

The labeled dataset contained 1485 instances of masks, gloves, goggles, and lab coat, each with a class label and bounding box. The number of instances for class gloves, lab coat, goggles, and masks were 421, 421, 322, and 321, respectively. The number of small-scale instances was greater than large-scale instances, which made the job of PPE detection more challenging. Object detection models require a large number of samples to train; therefore, data augmentation was applied on the original dataset to increase the size of the dataset using Roboflow [30]. After data augmentation, a total of 1164 images, which were 931, 116, and 116 images, were used for training, validation, and testing, respectively. After data augmentation, the images were resized to 416 × 416 pixels.

The sample images of the SLS dataset are shown in Figure 3. The acquired dataset was then given to various variants of the YOLOv5 and YOLOv7 models for training the PPE detection system.

Figure 3. Sample images of Student Laboratory Safety (SLS) dataset.

2.2. YOLOv5 Model

The R-CNN [31], Faster RCNN [21], and YOLO [32] series are currently the most popular object detection algorithms in research. The YOLO series is superior to the earlier models in terms of its increased speed and its capacity to detect small objects. In this work, both training and testing were performed by using various versions of the YOLOv5 models. YOLOv5 [27], released in 2020, provides a variety of object identification architectures that have already been trained using the MS COCO dataset. There are five distinct versions of YOLOv5, ranging from the tiny YOLOv5 nano version, designed for use on mobile and embedded devices, to the massive YOLOv5x large version. The YOLOv5 architecture is composed of various components, including the backbone, neck, and head, as shown in Figure 4.
2.2. YOLOv5 Model

The R-CNN [31], Faster RCNN [21], and YOLO [32] series are currently the most popular object detection algorithms in research. The YOLO series is superior to the earlier R-CNN [31], Faster RCNN [21], and YOLO [32] released in 2020, providing a variety of object identification architectures that have already been trained using the MS COCO dataset. There are five distinct versions of YOLOv5, both training and testing models in terms of its increased speed and its capacity to detect small objects. In this work, the focus is on YOLOv5, which is superior to the earlier R-CNN [31], Faster RCNN [21], and YOLO [32] series.

Figure 4. Network structure of YOLOv5.

The backbone consists of the focus structure [33] and Cross Stage Partial Networks (CSP) [34]. The focus structure downsamples the input data dimension while preserving the original information, as shown in Figure 5.

Figure 5. The processing flow of focus module.

The CSP Network extracts useful information, which improves the learning ability and reduces the memory cost of the model. The neck part combines the acquired features and forwards them to the prediction layer by using Feature Pyramid Networks (FPN) and the Path Aggregation Network (PAN). The FPN upsamples the high-level feature information through top-to-bottom communication and fusion for prediction. The underlying pyramid, PAN, conveys significant positional characteristics from a bottom-to-top manner, which helps in the differentiation of the same objects with different sizes and scales. The feature pyramids help the model to perform efficiently on new data. Figure 6 depicts how the feature extraction network upsamples its output feature map (F1, F2, and F3) by generating numerous new feature maps (P1, P2, and P3) for recognizing targets of varying scales. The output layer, the head, generates the final output vectors by applying anchor boxes on features and generates the final output vector, which includes class probabilities, object scores, and bounding boxes. The addition of the focus and CSP layers is the most notable improvement in YOLOv5. The focus layer reduces layers, parameters, FLOPS, and CUDA memory to increase forward and backward speeds. The CSP layer used as the backbone layer aims to extract detailed information and perform more comprehensive tasks. The meshing concepts of the original YOLO algorithm have been carried over into YOLO v5.
The network takes an RGB image as an input and produces a three-scale (small, medium, and large) output. The process of bounding box regression of YOLOv5 can be explained in detail by Equation (1) [35].

\[
\begin{align*}
    b_x &= 2\sigma(s_x) - \frac{1}{2} + r_x \\
    b_y &= 2\sigma(s_y) - \frac{1}{2} + r_y \\
    b_h &= p_h(2\sigma(s_h))^2 \\
    b_w &= p_w(2\sigma(s_w))^2
\end{align*}
\]

(1)

In the above equation, the coordinate value of the upper left corner of the feature map should be set to \((0, 0)\). The values \(r_x\) and \(r_y\) represent the distance between the center of the label bounding box and the upper left corner of the grid, respectively. The \(b_x\) and \(b_y\) are the center point coordinates while \(b_h\) and \(b_w\) represent the width and height of the label bounding box, as shown in Figure 7. The prior bounding box width and height are represented by \(p_h\) and \(p_w\). The \(s_x, s_y, s_h,\) and \(s_w\) are parameters related to the bounding box.

The most recent YOLO object detection model, YOLOv7 [28], was developed by Alexey Bochkovskiy. The architecture surpasses all the previous versions in terms of detection accuracy and speed. The authors’ primary contributions to the YOLOv7 model that allowed it to reach this pinnacle were: (1) their ultimate aggregation layer, E-ELAN, is an enhanced form of the efficient layer aggregation (ELAN) computational block; (2) model depth and breadth can be scaled in parallel by concatenating layers, an innovative approach to model scaling; and (3) the introduction of an auxiliary head network to enhance the training process and model re-parameterization technique to make the model more resilient and generalize well on fresh data.
3. Experimental Results

3.1. Environment Setup

The proposed system was trained using NVIDIA DGX-1, “The Fastest Deep Learning System” for AI Research based at the AI and Robotics Lab of United Arab Emirates University. The system consists of dual 20-core Intel® XEON® E5-2698 v4 2.2 GHz CPUs and 40,960 NVIDIA CUDA cores. The system has 8× Tesla V100 GPUs and a total of 256 GB GPU Memory. PyTorch library, Windows 10, and Python3.8 were used to train the YOLOv5 and YOLOv7 models and accomplish the predictions. The performance of the model was evaluated by using various evaluation metrics, each of which is explored in further depth in the next section.

3.2. Evaluation Metrics

In the proposed study, various evaluation metrics such as precision, recall, average precision (AP), mean average precision (mAP), and intersection over union (IoU) were used to compare the acquired experimental results. The term “intersection over union,” which is given in Equation (2), describes the degree to which two bounding boxes, i.e., predicted (PR) and ground-truth (GT), overlap one another. The higher the IoU, the larger the area of overlap should be.

\[
\text{IOU} = \frac{\text{Area of intersection}}{\text{Area of union}} = \frac{\text{Area (GT} \cap \text{PR)}}{\text{Area (GT} \cup \text{PR)}}
\]  

A recall is the true positive rate, which is also known as sensitivity and is a metric that determines how likely it is that ground-truth objects will be successfully recognized. A high recall is achieved by a model when it does not produce any false negatives, which means that there are no bounding boxes that are not detected but should be detected. The mathematical representation for the recall is given in Equation (3) below.

\[
R = \frac{TP}{TP + FN} = \frac{TP}{\text{Total Ground Truths}}
\]  

In the above equation, TP and TN represent the true positive and false negative, respectively. Precision, also known as the positive predictive value, defined in Equation (4), is the proportion of predicted positives that are correct. The precise model identifies only relevant objects and produces no false positives (FP).

\[
P = \frac{TP}{TP + FP} = \frac{TP}{\text{Total Predictions}}
\]  

The harmonic mean of the precision and recall scores is the \(F-1\) score as defined in Equation (5).

\[
F - 1 = 2 \times \frac{P \times R}{P + R}
\]  

\(AP\) is the area under the precision–recall curve while \(mAP\) is the average of all \(AP\) values over different classes/categories, as shown in Equation (6).

\[
mAP = \frac{1}{n} \sum_{i=1}^{n} AP_i
\]  

where \(n\) is the number of classes.

3.3. Analysis of Experimental Results and Discussion

The performance of various YOLOv5 and YOLOv7 versions on the proposed SLS dataset is summarized in Table 1. The number of epochs that were considered necessary for training all of the models was 300; however, an early stopping condition was utilized to prevent model overfitting. All the models that were trained on the SLS dataset showed promising performance. The YOLOv5n model achieved an mAP@0.5 of 0.774 with a
precision of 0.795 and a recall of 0.787. The individual class score of the YOLOv5 models and the performance of each model based on instance size is summarized in Table 2. The model (YOLOv5n) achieved the highest mAP@0.5, precision, and recall, of 0.943, 0.918, and 0.918, for the large-scale instance (lab coat) followed by the gloves, respectively. The YOLOv5n model size had the lowest size, of 3.9 MB, and had a faster inference time than the other compared models. The PR curve and confusion matrix of the model are shown in Figure 8. The YOLO5s model achieved an mAP@0.5, precision, and a recall of 0.717, 0.798, and 0.7022. In the individual class performance, the large-scale instances (lab coat and glasses) outperformed the other small-scale instances and achieved an individual mAP@0.5 of 0.907 and 0.952, respectively. Figure 9 depicts the confusion matrix and precision-recall curve of the YOLO5s model. The YOLOv5m model outperformed the YOLOv5s model in terms of mAP@0.5 and precision; however, the performance was lower than the YOLOv5n model. The model (YOLOv5m) achieved mAP@0.5, precision, and recall of 0.753, 0.776, and 0.837, respectively. The confusion matrix and PR curve of the model is shown in Figure 10. The YOLOv5l and YOLOv5x models achieved an mAP@0.5 of 0.707 and 0.725, respectively. The confusion matrices and PR curves of both models are shown in Figures 11 and 12. It can be seen from the results that increasing the number of parameters had a significant effect on the detection performance of the safety system. The performance of the YOLOv5 model degraded moving from the YOLOv5n model, having a lower number of parameters, to the YOLOv5x model, which had a large number of parameters. An increasing number of parameters also affected the detection speed of the model. In the proposed work, the YOLOv5n and YOLOv5m models outperformed the other variant size terms of mAP. YOLOv5n and YOLOv5m achieved the highest mAP, of 0.774 and 0.753, which showed that both models can detect objects more accurately compared to the other variants for our specific safety application trained on the SLS dataset.

On top of that, YOLOv7 and YOLOv7X were evaluated for how well they performed on the proposed SLS dataset. The YOLOv7 model’s two variations did not demonstrate promising results. The YOLOv7X model outperformed the YOLOv7 model while achieving an mAP@0.5 of 0.616. The YOLOv7 model achieved precision, recall, and mAP@0.5 of 0.700, 0.654, and 0.609, respectively. The confusion matrices of both variants of YOLOv7 are shown in Figures 13 and 14. YOLOv7’s accuracy was promising for the large-scale instances (lab coat and gloves classes); however, it was not as good for the small-scale instances (goggles and mask classes) due to the fact that the size of the class objects was small and there were fewer instances to train on. The YOLOv7 algorithm performed poorly on the proposed dataset when compared to the YOLOv5 algorithm on detecting small-scale instances, which is also in line with the literature [36]. Additionally, among the large-scale instances classes, gloves had the highest mAP@0.5, of 0.943, 0.952, 0.958, 0.954, 0.921, 0.860, and 0.855 corresponding to the model variants v5n, v5s, v5m, v5l, v5x, v7, and v7X, respectively. Among all the variants, YOLOv5m achieved the highest performance for all scale instances by achieving mAP@0.5 values of 0.958, 0.510, 0.872, and 0.672 for the gloves, goggles, lab coat, and mask classes, respectively. The two classes, mask and goggles, were not detected properly by all the variants of the YOLO models due to a small object size and lower number of instances of the classes in the data and a complex background of the training images.

The precision and recall curves of the models and the confusion metrics depict that the YOLOv5n model showed a high mAP@0.5 and true positives as compared to the other models. The real-time testing of the proposed system was performed on YOLOv5 models, and the results are shown in Figure 15. The images in the second row are the images acquired from the top camera installed in the lab; these show that the trained model can accurately predict the objects in the test dataset acquired from the surveillance camera. The acquired results showed that the system can be used in a lab environment for the compliance of student PPE. From the results, it was found that the YOLOv5n and YOLOv5m models outperformed the other variants of YOLOv5 in terms of performance and computational complexity. Both models have a smaller number of parameters and fast
detection speed in comparison with their later variants. When comparing the YOLOv5 and YOLOv7 based on large-scale and small-scale instances, the YOLOv5 versions’ detection capabilities stand out. The YOLOv7’s overall accuracy suffered because it could not detect the small-scale items efficiently. This was due to the limited number of small-scale instances in the SLS dataset and complex background in the images. It was also found that increasing the network complexity had a significant effect on the performance and the speed of the models. It is also evident from Table 2 that YOLOv5l had less weight and number of parameters as compared to the other compared variants of YOLOv5, which helped in reducing the computational time and complexity of the models. Increasing the weights and number of parameters not only increased the computational time of the models but also led to degradation of the performance on this particular SLS dataset. The YOLOv5l and YOLOv5x models had the highest weight and number of parameters while achieving a lower detection accuracy and speed as compared to other YOLOv5 variants. Nevertheless, the experimental results showed that the YOLOv5 model guarantees high performance for the PPE detection of students in educational labs. Due to the paucity of relevant literature and scant efforts in the research community about PPE monitoring in educational labs, this work does not provide a comparative analysis of the proposed algorithms vs. state-of-the-art approaches.

Table 1. Comparison of the various YOLOv5 versions for PPE detection.

|               | Precision | Recall | mAP@0.5 | mAP@0.5:0.95 | Weights  | No. of Parameters |
|---------------|-----------|--------|---------|--------------|----------|-----------------|
| YOLOv5n       | 0.795     | 0.787  | 0.774   | 0.485        | 3.9 MB   | 1.9 M           |
| YOLOv5s       | 0.798     | 0.702  | 0.717   | 0.476        | 14.5 MB  | 7.2 M           |
| YOLOv5m       | 0.837     | 0.776  | 0.753   | 0.481        | 42.3 MB  | 21.2 M          |
| YOLOv5l       | 0.805     | 0.725  | 0.707   | 0.482        | 92.9 MB  | 46.5 M          |
| YOLOv5x       | 0.794     | 0.688  | 0.725   | 0.488        | 173.2 MB | 86.7 M          |
| YOLOv7        | 0.700     | 0.654  | 0.609   | 0.366        | 74.8 MB  | 36.9 M          |
| YOLOv7X       | 0.775     | 0.652  | 0.616   | 0.400        | 142.1 MB | 71.3 M          |

Table 2. Individual class performance of the YOLOv5 detection models.

|               | Precision | Recall | mAP@0.5 | mAP@0.5:0.95 |
|---------------|-----------|--------|---------|--------------|
| Gloves        | 0.918     | 0.918  | 0.943   | 0.610        |
| Goggles       | 0.566     | 0.636  | 0.565   | 0.286        |
| Lab Coat      | 0.937     | 0.925  | 0.93    | 0.602        |
| Mask          | 0.761     | 0.67   | 0.659   | 0.440        |

|               | Precision | Recall | mAP@0.5 | mAP@0.5:0.95 |
|---------------|-----------|--------|---------|--------------|
| Gloves        | 0.942     | 0.902  | 0.952   | 0.638        |
| Goggles       | 0.519     | 0.455  | 0.366   | 0.247        |
| Lab Coat      | 0.968     | 0.775  | 0.907   | 0.620        |
| Mask          | 0.763     | 0.677  | 0.645   | 0.400        |
Table 2. Cont.

| Class       | Instance Size | Precision | Recall  | mAP@0.5 | mAP@0.5:0.95 |
|-------------|---------------|-----------|---------|---------|--------------|
| YOLOv5n     |               |           |         |         |              |
| Gloves      | L             | 0.934     | 0.929   | 0.958   | 0.629        |
| Goggles     | S             | 0.666     | 0.636   | 0.510   | 0.242        |
| Lab Coat    | L             | 0.93      | 0.825   | 0.872   | 0.622        |
| Mask        | S             | 0.819     | 0.713   | 0.672   | 0.431        |
| YOLOv5m     |               |           |         |         |              |
| Gloves      | L             | 0.957     | 0.951   | 0.954   | 0.668        |
| Goggles     | S             | 0.553     | 0.545   | 0.339   | 0.207        |
| Lab Coat    | L             | 0.966     | 0.719   | 0.915   | 0.644        |
| Mask        | S             | 0.743     | 0.684   | 0.621   | 0.408        |
| YOLOv5l     |               |           |         |         |              |
| Gloves      | L             | 0.963     | 0.902   | 0.921   | 0.617        |
| Goggles     | S             | 0.53      | 0.545   | 0.473   | 0.282        |
| Lab Coat    | L             | 0.906     | 0.675   | 0.866   | 0.612        |
| Mask        | S             | 0.776     | 0.632   | 0.641   | 0.439        |
| YOLOv5x     |               |           |         |         |              |
| Gloves      | L             | 0.795     | 0.803   | 0.860   | 0.505        |
| Goggles     | S             | 0.483     | 0.364   | 0.214   | 0.079        |
| Lab Coat    | L             | 0.892     | 0.825   | 0.807   | 0.522        |
| Mask        | S             | 0.628     | 0.622   | 0.555   | 0.357        |
| YOLOv7      |               |           |         |         |              |
| Gloves      | L             | 0.874     | 0.836   | 0.855   | 0.565        |
| Goggles     | S             | 0.599     | 0.544   | 0.327   | 0.162        |
| Lab Coat    | L             | 0.891     | 0.65    | 0.707   | 0.495        |
| Mask        | S             | 0.736     | 0.579   | 0.574   | 0.376        |

Figure 8. Confusion matrix and PR curve of the YOLOv5n model.
Figure 8. Confusion matrix and PR curve of the YOLOv5n model.

Figure 9. Confusion matrix and PR curve of the YOLOv5s model.

Figure 10. Confusion matrix and PR curve of the YOLOv5m model.

Figure 11. Confusion matrix and PR curve of the YOLOv5l model.

On top of that, YOLOv7 and YOLOv7X were evaluated for how well they performed on the proposed SLS dataset. The YOLOv7 model’s two variations did not demonstrate promising results. The YOLOv7X model outperformed the YOLOv7 model while achieving an mAP@0.5 of 0.616. The YOLOv7 model achieved precision, recall, and mAP@0.5 of 0.700, 0.654, and 0.609, respectively. The confusion matrices of both variants of YOLOv7 are shown in Figures 13 and 14. YOLOv7’s accuracy was promising for the large-scale instances (lab coat and gloves classes); however, it was not as good for the small-scale instances (goggles and mask classes) due to the fact that the size of the class objects was small and there were fewer instances to train on. The YOLOv7 algorithm performed poorly on the proposed dataset when compared to the YOLOv5 algorithm on detecting small-scale instances, which is also in line with the literature [36]. Additionally, among the large-scale instances classes, gloves had the highest mAP@0.5, of 0.943, 0.952, 0.958, 0.954, 0.921, 0.860, and 0.855 corresponding to the model variants v5n, v5s, v5m, v5l, v5x, v7, and v7X, respectively. Among all the variants, YOLOv5m achieved the highest performance for all scale instances by achieving mAP@0.5 values of 0.958, 0.510, 0.872, and 0.672 for the gloves, goggles, lab coat, and mask classes, respectively. The two classes, mask and goggles, were not detected properly by all the variants of the YOLO models due to a small
These detection capabilities stand out. The YOLOv7 and YOLOv5 models have a smaller number of parameters and computational complexity. Both models outperform their earlier variants. When comparing the YOLOv5 and YOLOv7 models, the YOLOv7X model outperforms the YOLOv5n model.

The acquired results show that the system can be used in surveillance applications due to the limited number of small scale instances, the YOLOv5 versions had the highest mAP@0.5, of 0.943, 0.952, 0.958, and 0.963, respectively. Among all the variants, YOLOv5m achieved the highest mAP@0.5 and true positives as compared to the model variants v5n, v5s, v5m, v5l, v5x, respectively. YOLOv7 performs better on the proposed dataset compared to the YOLOv5 models. YOLOv7X and YOLOv7 have high mAP@0.5 and true positives as compared to the model variants v5n, v5s, v5m, v5l, v5x, respectively. YOLOv7 is efficient and achieves promising results. The YOLOv7X model outperforms the other variants of YOLOv5 in terms of performance.

Figure 12. Confusion matrix and PR curve of the YOLOv7 model.

Figure 13. Confusion matrix and PR curve of the YOLOv7X model.

Figure 14. Confusion matrix and PR curve of the YOLOv7X model.
In the proposed work, YOLOv5- and YOLOv7-based PPE compliance monitoring systems were implemented to enhance the safety of academic labs. Firstly, a dataset consisting of four various classes, i.e., lab coat, gloves, goggles, and mask, was created. Secondly, various variants of YOLOv5 and YOLOv7 were trained, and their performance was compared based on various evaluation metrics such as precision, recall, mAP, weights, and computational time of the algorithm. From the above discussion, it can be concluded that YOLOv5 and YOLOv7 models can be used for PPE detection of the student to provide lab instructors with more efficient and intelligent safety strategies. It can also be concluded that lightweight variants of YOLOv5 such as YOLOv5n and YOLOv5m can be utilized to build a robust and fast PPE detection system. The YOLOv5n and YOLOv5m achieved the highest mAP@0.5, of 0.774 and 0.753, respectively. The YOLOv5 variants were able to perform well for various scale instances as compared to the YOLOv7 model. Among all YOLOv5 models, the highest @mAP was achieved by the YOLOv5n and YOLOv5m. The YOLOv7 was not able to show a promising performance for small-scale objects due to the limited number of instances, the size of the objects, and the complex background of the acquired data. The performance of the models can be enhanced by providing more efficiently labeled data with a sufficient number of instances for all the classes. In conclusion, training a lightweight model with a sufficient amount of data is the best option for practical PPE detection systems. The proposed system will considerably reduce the occurrence of safety-related incidents and accidents in the labs by creating a safety culture. In addition, the proposed system will enhance the traditional lab safety training process by providing insights to the lab instructors about the safety protocols of the students, creating a feedback loop in which the information is best absorbed. In the future, we will improve the performance of YOLOv5 and YOLOv7 variants by using various optimization techniques and we will add more classes to the data to expand the applicability of the proposed system. The suggested system will also leverage methods such as the Internet of Things (IoT) and big data to make educational labs safer places to learn.
Author Contributions: Conceptualization, L.A., F.A. and H.A.; data collection and labeling M.I.Y. and Z.I.A.; methodology, L.A., M.M.A.P., H.A., F.A., M.I.Y. and Z.I.A.; writing—original draft preparation, L.A., H.A. and F.A.; writing—review and editing, M.M.A.P., F.A., H.A. and L.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wu, T.-C.; Liu, C.-W.; Lu, M.-C. Safety Climate in University and College Laboratories: Impact of Organizational and Individual Factors. J. Saf. Res. 2007, 38, 91–102. [CrossRef]
2. Negligence Caused UCLA Death. Available online: https://cen.acs.org/articles/87/i19/Negligence-Caused-UCLA-Death.html (accessed on 16 September 2022).
3. Van Noorden, R. A Death in the Lab. Nature 2011, 472, 270–271. [CrossRef]
4. University of Hawaii Fined $115,500 for Lab Explosion. Available online: https://cen.acs.org/articles/94/web/2016/09/University-Hawaii-fined-115500-lab.html (accessed on 16 September 2022).
5. Texas Tech University Chemistry Lab Explosion | CSB. Available online: https://www.csb.gov/texas-tech-university-chemistry-lab-explosion/ (accessed on 16 September 2022).
6. Schröder, I.; Huang, D.Y.Q.; Ellis, O.; Gibson, J.H.; Wayne, N.L. Laboratory Safety Attitudes and Practices: A Comparison of Academic, Government, and Industry Researchers. J. Chem. Health Saf. 2016, 23, 12–23. [CrossRef]
7. Rubaiyat, A.; Toma, T.; Kalantari-Khandani, M.; Rahman, S.A.; Chen, L.; Ye, Y.; Pan, C. Automatic Detection of Helmet Uses for Construction Safety. In Proceedings of the 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW), Omaha, NE, USA, 13–16 October 2016; pp. 135–142.
8. Shrestha, K.; Shrestha, P.; Bajracharya, D.; Yantis, E. Hard-Hat Detection for Construction Safety Visualization. J. Constr. Eng. 2015, 2015, 721380. [CrossRef]
9. Qiu, Z.; Zhao, Z.; Chen, S.; Zeng, J.; Huang, Y.; Xiang, B. Application of an Improved YOLOv5 Algorithm in Real-Time Detection of Foreign Objects by Ground Penetrating Radar. Remote Sens. 2022, 14, 1895. [CrossRef]
10. Xiong, C.; Hu, S.; Fang, Z. Application of Improved YOLOv5 in Plate Defect Detection. Int. J. Adv. Manuf. Technol. 2022, 2022, 2315. [CrossRef]
11. Hindawi Swin-YOLOv5: Research and Application of Fire and Smoke Detection Algorithm Based on YOLOv5. Available online: https://www.hindawi.com/journals/cin/2022/681680/ (accessed on 17 September 2022).
12. Kumar, S.; Gupta, H.; Yadav, D.; Ansari, I.A.; Verma, O.P. YOLOv4 Algorithm for the Real-Time Detection of Fire and Personal Protective Equipments at Construction Sites. Multimed. Tools Appl. 2022, 81, 22163–22183. [CrossRef]
13. Damage detection and localization in masonry structure using faster region convolutional networks. Geomate J. 2019, 17. Available online: https://geomatejournal.com/geomate/article/view/270 (accessed on 17 September 2022).
14. Otgonbold, M.-E.; Gochoo, M.; Alnajjar, F.; Ali, L.; Tan, T.-H.; Hsieh, J.-W.; Chen, P.-Y. SHEL5K: An Extended Dataset and Benchmarking for Safety Helmet Detection. Sensors 2022, 22, 2315. [CrossRef]
15. Delhi, V.S.K.; Sankarlal, R.; Thomas, A. Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques. Front. Built Environ. 2020, 6, 136. [CrossRef]
16. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. arXiv 2018, arXiv:1804.02767.
17. Tang, S.; Roberts, D.; Golparvar-Fard, M. Human-Object Interaction Recognition for Automatic Construction Site Safety Inspection. Autom. Constr. 2020, 120, 103556. [CrossRef]
18. Wu, J.; Cai, N.; Shen, W.; Wang, H.; Wang, G. Automatic Detection of Hardhats Worn by Construction Personnel: A Deep Learning Approach and Benchmark Dataset. Autom. Constr. 2019, 106, 102894. [CrossRef]
19. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. arXiv 2016, 9905, 21–37. [CrossRef]
20. Fang, Q.; Li, H.; Luo, X.; Ding, L.; Luo, H.; Rose, T.M.; An, W. Detecting Non-Hardhat-Use by a Deep Learning Method from Far-Field Surveillance Videos. Autom. Constr. 2018, 85, 1–9. [CrossRef]
21. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Curran Associates, Inc.: Montreal, QC, Canada, 2015; Volume 28.
22. Saudi, M.; Hakim, A.; Ahmad, A.; Mohd Saudi, A.S.; Hanafi, M.; Narzullaev, A.; Ghazali, I. Image Detection Model for Construction Worker Safety Conditions Using Faster R-CNN. Int. J. Adv. Comput. Sci. Appl. 2020, 11, 246–250. [CrossRef]
23. Nath, N.D.; Behzadan, A.H.; Paal, S.G. Deep Learning for Site Safety: Real-Time Detection of Personal Protective Equipment. *Autom. Constr.* **2020**, *112*, 103085. [CrossRef]

24. Wang, J.; Zhu, G.; Wu, S.; Luo, C. Worker’s Helmet Recognition and Identity Recognition Based on Deep Learning. *Open J. Model. Simul.* **2021**, *9*, 135–145. [CrossRef]

25. Wang, Z.; Wu, Y.; Yang, L.; Thirunavukarasu, A.; Evison, C.; Zhao, Y. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors* **2021**, *21*, 3478. [CrossRef]

26. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.

27. Jocher, G. Yolov5. Code Repository. Available online: https://github.com/ultralytics/yolov5 (accessed on 17 September 2022).

28. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors 2022. *arXiv* **2022**, arXiv:2207.02696.

29. GitHub—Heartexlabs/LabelImg: LabelImg Is Now Part of the Label Studio Community. The Popular Image Annotation Tool Created by Tzutalin Is No Longer Actively Being Developed, but You Can Check Out Label Studio, the Open Source Data Labeling Tool for Images, Text, Hypertext, Audio, Video and Time-Series Data. Available online: https://github.com/heartexlabs/labelImg (accessed on 1 October 2022).

30. Roboflow: Give Your Software the Power to See Objects in Images and Video. Available online: https://roboflow.com/ (accessed on 1 October 2022).

31. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.

32. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

33. Yang, S.J.; Berndl, M.; Michael Ando, D.; Barch, M.; Narayanaswamy, A.; Christiansen, E.; Hoyer, S.; Roat, C.; Hung, J.; Rueden, C.T.; et al. Assessing Microscope Image Focus Quality with Deep Learning. *BMC Bioinformatics*. **2018**, *19*, 77. [CrossRef] [PubMed]

34. Guo, Y.; Zeng, Y.; Gao, F.; Qiu, Y.; Zhou, X.; Zhong, L.; Zhan, C. Improved YOLOV4-CSP Algorithm for Detection of Bamboo Surface Sliver Defects With Extreme Aspect Ratio. *IEEE Access* **2022**, *10*, 29810–29820. [CrossRef]

35. Zhang, Y.; Guo, Z.; Wu, J.; Tian, Y.; Tang, H.; Guo, X. Real-Time Vehicle Detection Based on Improved YOLO V5. *Sustainability* **2022**, *14*, 12274. [CrossRef]

36. Liu, H.; Sun, F.; Gu, J.; Deng, L. SF-YOLOv5: A Lightweight Small Object Detection Algorithm Based on Improved Feature Fusion Mode. *Sensors* **2022**, *22*, 5817. [CrossRef] [PubMed]