A Robust Deep Learning Architecture for FireFighter PPEs Detection

Achilleas Sesis
OINF, London, UK
achilleas@Oinfinity.net

Ilias Siniosoglou
University of Western Macedonia, Kozani, GR
isiniosoglou@uowm.gr

Yannis Spyridis
OINF, London, UK
yannis@Oinfinity.net

Georgios Efstathopoulos
OINF, London, UK
george@Oinfinity.net

Thomas Lagkas
International Hellenic Univ, Kavala, GR
tlagkas@cs.ihu.gr

Vasileios Argyriou
Kingston University, London, UK
Vasileios.Argyriou@kingston.ac.uk

Panagiotis Sarigiannidis
University of Western Macedonia, Kozani, GR
psarigiannidis@uowm.gr

***Abstract*—Personal Protective Equipment (PPE) is one of the primary defence mechanisms to reduce the exposure of the personnel to hazardous environments. It's significantly important to Fire Fighters as they are constantly exposed to dangerous elements such as fire, gas or chemicals. Unfortunately, in real-time emergencies, such as fires, it is very difficult to identify if a responder using PPE is fully equipped to reduce any accidents in the workplace or even coordinate response actions due to the high pace of the situation. A lack of a unified Fire Fighting PPE image dataset was also observed, which makes the task of training Machine Learning (ML) models to solve this problem a challenge. To that end, we first create a general purpose FireFighter Equipment Detection dataset. We then propose to utilise the widely used YoloV5 Deep Network architecture to detect different PPE components in real-time. This work leverages the pre-trained YoloV5 model, using transfer learning to fine-tune the model using the created detection dataset that contains targeted Fire Fighter PPE images. By employing the pre-trained model which requires substantially fewer training samples, we were able to achieve a considerably good performance on the Fire Fighter PPE object detection. The proposed method can distinguish four different PPE components such as a Helmet, Gloves, Mask or Insulated protective cloth, achieving high detection efficiency which is experimentally established.**

***Index Terms*—Object detection, PPE, FireFighter, YoloV5, Dataset, Data Collection**

I. Introduction

The history of emergency responders, their method and importantly the Personal Protective Equipment (PPE) they used can be traced throughout the ages until today. As the personal protection of first responders has been widely recognized, so the technology of PPE, as well as the on-site safety procedures, have significantly evolved. Especially in the field of Firefighting, where extreme conditions that are widely varying in nature prevail, PPE is of especially significant importance [1] [2] [3]. Unfortunately, in emergency situations, the use of PPE by personnel is difficult to be monitored and tracked since things are moving fast which results in two problems. First, the performance and the correct application of safety procedures is difficult to be monitored by the respective agency, which can result in misconduct in the field with severe repercussions [4], even loss of life. The second problem is an immediate extension of the former. In the case of an on-site accident it is difficult to audit and produce the respective forensics in the subsequent legal procedure [5] [6].

These aforementioned shortcomings begin to beg the question of how a system can be established to keep track of the PPE of firefighters in the line of duty. Artificial Intelligence (AI), and in particular Object Detection through the use of Deep Learning (DL) can effectively tackle the task of recognizing the equipment of first responders. Modern field equipment, that being wearable equipment or field monitoring and actuating equipment, usually carry high definition cameras in order to monitor the on-going situation and/or keep records for later auditing. Deep Learning algorithms can actively utilise the feed from those cameras to effectively track and identify object in the camera field of vision helping the on-site crew with operations. Deep Learning based PPE detection can realise on-site identification of the correct application of PPE equipment of first responders to accommodate the previously mentioned drawbacks.

The use of Deep Learning dictates the availability of a large quantity of field-specific data in order to train and validate the constructed models to perform their intended task. Unfortunately, as is widely recognized, case-specific datasets are a rare commodity since they require the arduous task of collecting, filtering and annotating the information in preparation to be used for the training of Deep Learning models. Firefighting PPE image data for the purpose of training PPE detection models are no exception, shown a clear lack of either organised or public data sources.

In response to the demand for a solid implementation of a firefighting PPE detection and classification algorithm in order to minimise on-site neglect leading to accidents and its subsequent auditing, as well as the corresponding PPE dataset targeting Firefighting protective equipment, and taking into account the aforementioned preconditions, this work proposes a novel robust Deep Learning architecture for Firefighter PPE detection. Specifically, this work presents a DL model based on the YoloV5 architecture that utilises Transfer Learning

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 883371 (RESPOND-A).

along with a custom Firefighter PPE annotated detection dataset in order to apply optimised domain adaption, actively conserving resources offering promising results. In summary, the main contributions of this work are:

- Presents a custom annotated Firefighting PPE image data collection for training and validating Deep Neural Network in the task of PPE detection and recognition.
- Proposes a robust Deep Learning Architecture for object detection on Firefighting PPE equipment for on-site PPE detection.
- Implements and validates in a quantitative manner the proposed methodology for optimized resource conservation based on Transfer Learning.

The rest of this paper is organized as follows. Section II presents similar work performed for PPE and object detection. Section III gives insight for the implemented methodology, the developed model and the collected PPE dataset. Section IV evaluates the presented methodology using quantitative metrics. Finally, Section V concludes this work.

II. Literature Review

A. Object Detection

Object Detection is one of the fundamental task that where sought of in the field of ML and DL and lunched a multitude of subsequent innovations in both the application and research sectors. Object Detection can be defined as the process of both recognising a certain object within an image while also localising it withing that image. One of the most prominent Deep Learning Object Detection algorithm that was widely adopted was the You Only Look Once (YOLO) [7] algorithm due to its efficiency and speed in recognising objects in a given image and localising it within a bounding box. Recent developments in advanced AI-enabled Object Detection have pushed the boundaries of this field with the release of newer more optimized versions of the YOLO algorithm [8] [9] [10] with the current version being YoloV5.

B. Personal Protective Equipment Datasets

Personal Protective Equipment detection has come along with advances in AI-oriented safety procedures. As there are a lot to be gained through PPE detection, such as, on-field surveillance, PPE oriented datasets have been increasingly been sought of. It is well known that one of the basic pylons and the same time the biggest bane of Deep and Machine Learning is the available data in order to train and validate the respective models. In this respect, some noteworthy datasets have emerged. In particular, in [11] the authors propose the CPPE-5 dataset that contains PPE of workers in the medical field. In particular, the CPPE-5 data collection contains a series of annotated images of medical workers wearing PPE in different context and environments. The labels of the presented data point to 5 object categories (coveralls, face shield, gloves, mask, goggles). The work also analyses the performance of state-of-the-art Deep Learning Object Detection models on the provided Dataset with the focus of recognising non-iconic Medical PPE objects within a scene.

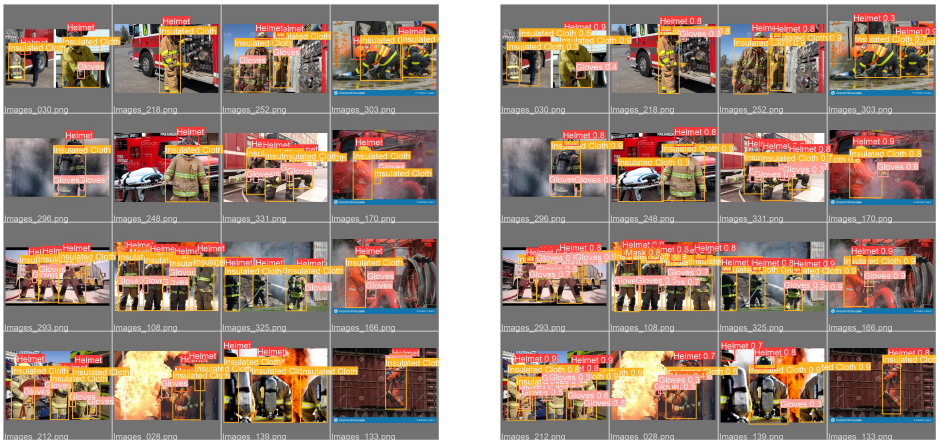
In [12] the authors present a concurrent implementation for Construction Site PPE compliance detection. In their work, the authors collect and utilise images from construction site surveillance video feeds construction a dataset of 2509 samples. The samples are annotated containing four different classes, namely, NOT SAFE, SAFE, NoHardHat, and NoJacket. The authors employ a Deep Learning Convolutional architecture, utilizing the pretrained YoloV3 object detection model. The implemented system produces alerts when non-safe flags are raised, proving feasibility in field deployment.

The work performed in [13] follows a similar principle. The authors create two datasets for Construction Site PPE object detection. The first dataset is a compilation of real images containing different scenes that include different PPE components. The second is a virtually created dataset of construction PPE images. The data generation of the virtual dataset takes advantage a 3D modeling and object gaming engine to construct scenes that contain PPE components. The authors utilise the virtual data with a domain adaption stem, using real images. Using these data a Yolo model was trained using both sets. The produced results show that this method can train Deep models successfully, using only a sub-portion of real images in conjuction with virtually created ones, for object detection.

C. Object Detection Dataset

As it is widely acknowledged, one of the most basic pillars of Machine and Deep Learning is the utilization of data. Although a lot of research has been directed in defining quality and rich datasets for the different problem categories that AI tries to solve there is still a considerable lack of datasets. Especially in Object Detection, where to solve a problem an algorithm needs, except for high quality images, the localized region of interest in the image. The discovery of quality datasets is a difficult task since the production of such a data collection is an arduous manual process and very difficult to automate. Albeit this problem, some widely utilised datasets for image and object detection has surfaced. A good example presents the ImageNet [14] data collection. ImageNet contains a set of hand-annotated images to train and validate a Deep Learning model for large-scale object detection. The dataset contains 14,197,122 annotated images in two categories, namely, a) image-level annotation for the presence or absence of an object and b) object level annotation and localization with bounding boxes. ImageNet is constitutes as a state-of-the-art benchmark dataset for model validation and pre-training.

Another interesting example are the CIFAR [15] datasets. They are distinguished in CIFAR-10 and CIFAR-100 where the former contains 10 and 100 object classes, respectively. The CIFAR-10 dataset contains 60000 colour images, of size 32x32, with 6000 images per class and without overlapping objects or classes. On the other hand, CIFAR-100 contains 60000 images but 600 images for each class. The CIFAR-100 contains also 20 super-classes with 5 non-overlapping sub-classes each. These two datasets are also considered



(a) Labelled Images (b) Predicted Images

Fig. 1: Proposed PPE Dataset

benchmark data collections and are widely used for the task of Object Detection on a wide scale.

Similarly, the Microsoft Common Objects in Context (MSCOCO) image collection [16] is a large-scale dataset for object detection. Specifically it is oriented in providing intuitive data for training Deep Learning models in the tasks of object recognition, object segmentation and extraction, landmark detection, and image/object captioning. The dataset includes 328K images. It contains around 200,000 annotated images, segregated in a Training set of 83,000 images, a Validation set of 41,000 images, and a Testing set of 41,000 images, as of the latest release, and around 123,000 non-annotated images. MSCOCO provides a variety of dependencies for the different object detection tasks. In particular, the dataset provides 80 object categories for training detection algorithms, which include bounding boxes, localization and segmentation masks for the classes contained in an image. Furthermore, it includes a large number of images containing 17 different key-points, including dense poses, for pose detection while it provides 91 and 80 categories for background and scene detection, respectively. Finally, the dataset includes image descriptions for Natural Language Processing (NLP) oriented problems.

This work leverages the YoloV5 detection algorithm, pre-trained on the MSCOCO dataset. Table I summarises the attributes of the different existing Object Detection dataset, along with the one produced in this work.

III. Methodology

A. Data Preparation and Generation

One of the two main contributions of this work is the creation of a targeted Firefighter PPE detection dataset. This work gives special care in collecting, processing and annotating images containing Firefighters on the field, in different states of operational preparedness using varying PPE components. This dataset aims to be used to train ML and DL algorithms in detecting the aforementioned PPE components that firefighters use in operational situations. To create this dataset, the following steps were followed. First, the data were scraped from online sources, collecting a big pool of images containing fully equipped firefighters in different situations, such as, training exercises, during the mitigation of incidents and in a variety of poses. This includes single firefighers or in teams. Subsequently, utilizing the LabelImg software [17], bounding boxes were manually drawn over the different PPE components, producing the coordinates of the desired ROIs in each picture. These were then annotated to contain the correct class. The produced dataset contains four distinct PPE classes, namely, i) Helmet, ii) Gloves, iii) Mask and iv) Insulated protective cloth. Finally, using these information, the ground-truth labels were produced in order to be used for training the Deep Learning detectors. The data were exported in the Yolo dataset format (1) in a txt file for each respective sample for the training, validation and test sets, respectively. The final collected dataset includes 342 annotated samples.

$$< \text{object-class} > \ < \text{x} >< \text{y} > \ < \text{width} >< \text{height} > \quad (1)$$

where each frame describes the class of the object of interest, the x and y coordinates within the image and the width and height of the object region of interest (ROI).

TABLE I: Publicly Available Object Detection Datasets

Dataset	No. of classes	No. of signs
Object Detection		
ImageNet [14]	1,000	14,197,122
CIFAR-10 [15]	10	60,000
CIFAR-100 [15]	100	5,184
MSCOCO [16]	91	328,000
PPE Detection		
CPPE-5 [11]	5	1,029
Construction Site PPE [12]	5	2,509
Safety Equipment Detection [13]	7	180 *(Real)*
This Work		
FireFighter PPE Dataset-5 [11]	4	342

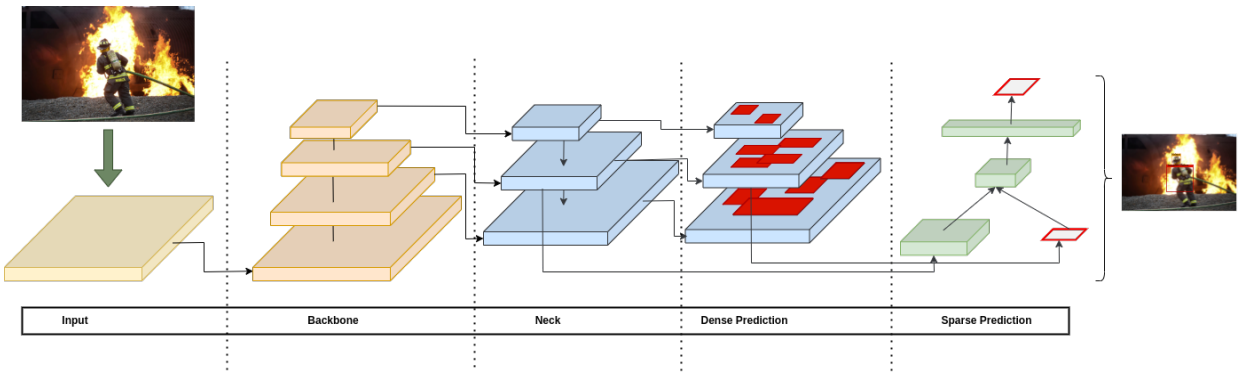


Fig. 2: Proposed YoloV5 Architecture

For the implementation of the proposed algorithm, the produced dataset was segmented into three slices, a Training, a Testing and a Validation one. Each of these portions included 277, 34 and 31 samples, respectively.

B. Model Training

The presented methodology trains the YoloV5 algorithm on the collected PPE dataset. The YoloV5 model is pre-trained on the MSCOSO large image collection, to enhance the generalization and precision of the model on the task of object detection. Since the MSCOCO dataset offers object detection, object from image segregation and pose identification, pre-training the model on this dataset will enhance the ability of the detection model to successfully identify the different PPE components, worn by firefighters on the field, from the information rich background (e.g. an image containing the deployment of rescue means, people and support equipment). The architecture of the pre-trained YoloV5 algorithm is depicted in Figure 2.

For the purpose of training the pre-trained YoloV5 model, the collected PPE dataset was segmented into a training set of 277 samples, a validations set of 34 samples and a testing set of 31 samples. The main aim of using an already trained model is to significantly reduce the data requirements for training the PPE detection model, as well as the computational cost. To test this, the collected dataset leverages a small number of images for detected a total pf 4 major firefighting equipment classes, namely, i) Helmet, ii) Gloves, iii) Mask and iv) Insulated Clothes. To reach this goal, this work leverages the YoloV5 model, using transfer learning, to train only on a small portion of PPE data. By applying a domain adaption rational, using the PPE data on top of the YoloV5 trained on the MSCOCO dataset, the model aims to use the semantic information extracted by the multitude of general purpose data of the MSCOCO image collection to be able to extrapolate PPE specific information.

On a technical perspective, to implement the model, the pytorch library was utilised along with the YoloV5 version v6.1, which is publicly available [18].After the training of the YoloV5 algorithm on the collected PPE dataset, the model is able to input generic firefighting scenes, for example firefighters trying to extinguish an urban fire, and produce the classification of the protective equipment worn by the firefighters. Specifically, as can be seen in Figure 1, the algorithm is able to localize, draw and denote each component of the PPE, from the aforementioned classes, on the image.

For the optimization of the bounding box localization, the $DIoU$ loss is effectively utilized since it shows an increased performance with the YoloV5 algorithm [19]. $DIoU$ is a direct extension of IoU (2) that optimizes the bounding box prediction. In particular, we take

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} = IoU(B_p, B_r) = \frac{B_p \cap B_r}{B_p \cup B_r} \quad (2)$$

and the distance

$$Loss_{IoU} = 1 - IoU \quad (3)$$

where B_p and B_r denote the predicted and real bounding box respectively. Then $DIoU$ optimizes IoU by taking in consideration the square of the diagonal d_B of smallest overlapping bounding box B_o which contains B_p and B_r. Thus, we have

$$DIoU = IoU - \frac{\sqrt{(B_p^2) - (B_r^2)}^2}{d_B^2} \quad (4)$$

and its equivalent distance

$$Loss_{DIoU} = 1 - DIoU = 1 - IoU - \frac{\sqrt{(B_p^2) - (B_r^2)}^2}{d_B^2} \quad (5)$$

As this function solves the non-intercecting bounding box problem of IoU it help the model converge faster.

IV. Evaluation

A. Evaluation Environment

To realize the outlined methods, a mid-range evaluation system was utilized. The experiments were performed on a Linux workstation consisting of 16GB RAM memory, an i7 Intel core processor, and an NVIDIA GeForce RTX 2080 Ti 11Gb GPU. Since the resource allocation and experiment times are relative to the evaluation environment, they are not presented here.

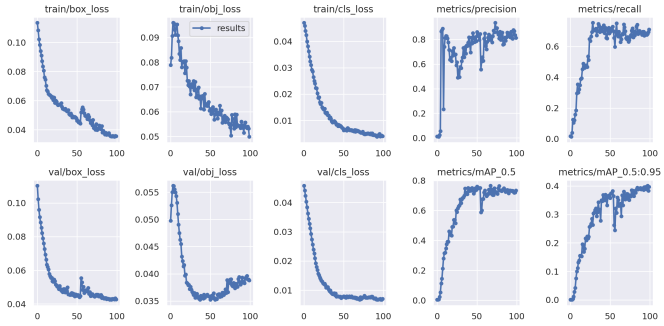


Fig. 3: Model Training

B. Metrics

To accurately measure the efficiency and efficacy of the developed model, this work takes advantage of targeted metrics, widely used in object detection and recognition tasks. In particular, this study leverages i) Recall, ii) F1-Score, iii) the produced confusion matrix, iv) Mean Average Precision (mAP) and IoU (intersection-over-union). These metrics are oriented in measuring the efficiency of the trained model on the collected dataset in a quantitative manner. The lower tier metrics depend on the fundamental measuring of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) that contribute to the establishment of the Confusion Matrix of the classification output of the model at hand. Based on these, the efficiency metrics are defined as:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

which defines the precision (6) of the neural classifier,

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

which defines the recall (7) or sensitivity of the classifier,

$$\textit{F1-Score} = 2 * \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{8}$$

and the F1-Score (8) which is the weight average of the combination of the two former. Furthermore, since we aim to optimize an object classification problems, some problem specific metrics are used. In particular, the Intersection over Union is utilised (IoU) (2),

that defines the threshold on which the model will optimize the process of matching the predicted bounding box tho the ground-truth one, on the task of object localisation. This threshold is used to make the final differentiation between the state of the classification of a certain image, i.e., if the predictions is a True Positive or a False Positive. Finally, to evaluate the produced model, the mean Average Precision (mAP) (9) is measured. The mean Average Precision is calculated as the mean of weighted average precision in each threshold.

$$mAP = \frac{1}{C} \sum_{i=1}^{C} AP_i \tag{9}$$

Where C is the number of classes, i denotes the inspected class and AP the weighted average precision of the i^{th} class.

C. Evaluation Results

To evaluate the proposed methodology and dataset, the pre-trained model was retrained on a sub-sample of the collected PPE data. The YoloV5 model was trained for a limited number of iterations, in this case 100, both due to the small number of PPE samples and for optimization purposes. In particular, this work aims to minimize the computational effort of the produced model in conjunction with the available data by using transfer learning. The model is trained on the training set of the PPE dataset and evaluated using the testing set. For the tuning of the detection process this implementation uses an IoU threshold of 0.6 for normalizing the bounding box correlation for the evaluation process. Figure 3 depicts the performance of the detection model through the training procedure for 100 epoch. As can be seen, the models shows a steady loss decline in the localization, detection and classification of the PPE components within the provided images. During the training a stady increase in the precision, recall and mAP[0.5:0.95] can be seen which is also establish by the validation steps throughout the training process.

The effect of the training on the small PPE sample can be initially seen in Figure 4. In the quantified ratios, it can be seen that most classes archive a low false positive to a false negative ratio, with the exception of the *Gloves* class. This can be attributed to the fact that gloves present the most volatile characteristics in the PPE pictures. The gloves/hands of the first responders are often found holding or maneuvering something and so are hidden or merged/crossed with other artefact in the image, resulting in a lower detection rate. Nevertheless, this class detection shows a small error with the rest of the classes. Similarly, the confidence of the prediction on the subsequent classes can be seen in Figure 5. We can see that the confidence of the network is high, reaching a peak at about 80% F1-Score. Again we see that the class with the lower confidence is the *Gloves* class, as mentioned.

On a more detailed evaluation of the results of the developed model, in Figure 6 the Confusion Matrix of the predictions can be seen. The produced model shows promising results, as the detection for each class is 87% for the *Helmet*, 65% for the *Gloves*, 78% for the *Mask* and 91% for the *Insulated Cloth* samples, respectively, as the rest is attributed to the background surroundings within the inspected image. The evaluation results are summarized on Table II, showing the measured efficiency of the prediction for each label, with the overall mAP of the model reaching 81%.

V. Conclusion

In the field of critical operations, and especially in the field of Firefighting, the application of PPE detection and monitoring is crucial to ensure the correct use of the first responders' protective equipment to avoid possible accidents safeguarding the integrity of the Firefighter's health. Adversely, in the case of an accident, the outcome of the PPE detection can be used

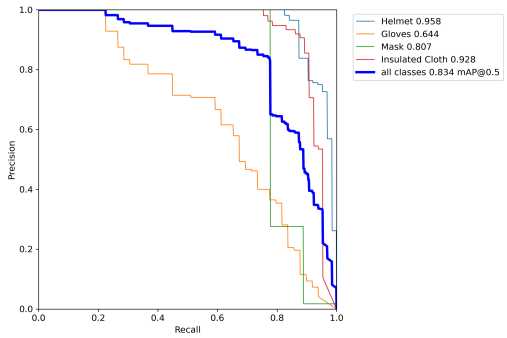


Fig. 4: Precision Recall Curve

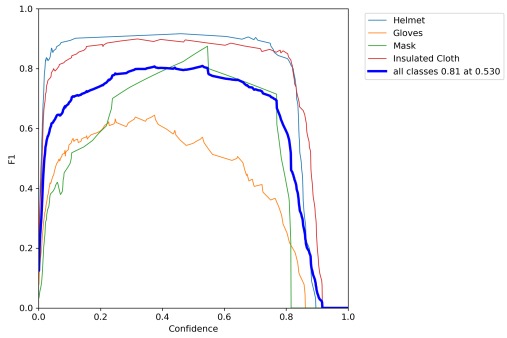


Fig. 5: F1-Score

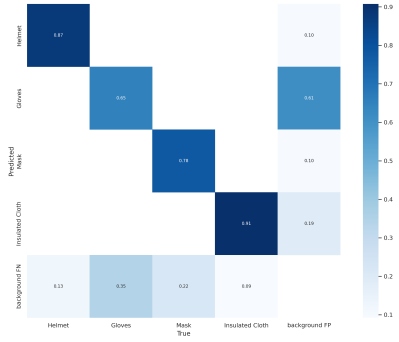


Fig. 6: Confusion Matrix

TABLE II: Overall results of FRs PPE detection

Class	Labels	Precision	Recall	mAP@0.5
Helmet	63	0.965	0.865	0.958
Gloves	49	0.785	0.448	0.644
Mask	9	0.974	0.778	0.807
Insulated Cloth	65	0.932	0.847	0.928
all	186	0.914	0.735	0.834

as forensics evidence in the subsequent legal auditing of that accident. This paper first produces a custom dataset containing images of Firefighters using PPE equipment. Subsequently, a methodology to robustly detect Firefighting PPE equipment is implemented by leveraging the YoloV5 algorithms, pre-trained on the MSCOCO image collection. This aims at utilising a model containing the semantic information of the MSCOCO collection, applying domain adaption by using a small portion of PPE images, thus conserving computational resources while using a small data sample. The evaluation of the proposed algorithm shows promising results, showing high accuracy of detection for all classes.

VI. Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 883371 (RESPOND-A).

References

[1] S.-Y. Son, I. Bakri, S. Muraki, and Y. Tochihara, "Comparison of firefighters and non-firefighters and the test methods used regarding the effects of personal protective equipment on individual mobility," *Applied ergonomics*, vol. 45, 01 2014.

[2] Y. Wu and M. M. Islam, "Function design of firefighting personal protective equipment: A systematic review," 09 2020.

[3] R. Nayak, S. Houshyar, and R. Padhye, "Recent trends and future scope in the protection and comfort of fire-fighters' personal protective clothing," *Fire Science Reviews*, vol. 3, pp. 1–19, 12 2014.

[4] A. Stec, K. Dickens, M. Salden, F. Hewitt, D. Watts, P. Houldsworth, and F. Martin, "Occupational exposure to polycyclic aromatic hydrocarbons and elevated cancer incidence in firefighters," *Scientific Reports*, vol. 8, 02 2018.

[5] M. Maglio, C. Scott, A. Davis, J. Allen, and J. Taylor, "Situational pressures that influence firefighters' decision making about personal protective equipment: A qualitative analysis," *American journal of health behavior*, vol. 40, pp. 555–567, 09 2016.

[6] L. Osvaldová and M. Petho, "Occupational safety and health during rescue activities," *Procedia Manufacturing*, vol. 3, pp. 4287–4293, 12 2015.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 06 2016, pp. 779–788.

[8] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," 07 2017, pp. 6517–6525.

[9] ——, "Yolov3: An incremental improvement," 04 2018.

[10] A. Bochkovskiy, C.-Y. Wang, and H.-y. Liao, "Yolov4: Optimal speed and accuracy of object detection," 04 2020.

[11] R. Dagli and A. M. Shaikh, "Cppe-5: Medical personal protective equipment dataset," 12 2021.

[12] V. Delhi, S. Lal, and A. Thomas, "Detection of personal protective equipment (ppe) compliance on construction site using computer vision based deep learning techniques," *Frontiers in Built Environment*, vol. 6, 09 2020.

[13] M. Di Benedetto, E. Meloni, G. Amato, F. Falchi, and C. Gennaro, "Learning safety equipment detection using virtual worlds," 09 2019, pp. 1–6.

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, 09 2014.

[15] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," *Computer Science Department, University of Toronto, Tech. Rep*, vol. 1, 01 2009.

[16] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. Zitnick, "Microsoft coco: Common objects in context," 05 2014.

[17] Tzutalin, "Tzutalin/labelimg: Labelimg is a graphical image annotation tool and label object bounding boxes in images." [Online]. Available: https://github.com/tzutalin/labelImg

[18] Ultralytics, "Ultralytics/yolov5: Yolov5 in pytorch." [Online]. Available: https://github.com/ultralytics/yolov5

[19] Z. Wang, L. Wu, T. Li, and P. Shi, "A smoke detection model based on improved yolov5," *Mathematics*, vol. 10, no. 7, 2022. [Online]. Available: https://www.mdpi.com/2227-7390/10/7/1190