ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 9 (2022) No.3, pp.97-111
DOI:10.33168/LISS.2022.0308

A YOVO5 Based Real-time Helmet and Mask Detection System

Zu Jun Khow, Kah Ong Michael Goh, Connie Tee, Check Yee Law

Faculty of Information Science and Technology, Multimedia University, Melaka 75450, Malaysia

Abstract. The COVID-19 pandemic has brought unimaginable damage to the globe; It had brought people we love away and forced the government to lock down the cities to prevent the infection of COVID-19 from spreading. This stopped various industries from working, especially the engineering and construction industries. Fortunately, the effectiveness of the COVID-19 vaccine had enabled the industries to resume their operation back to normal. However, the mask now is an essential equipment to be worn by all workers while on site, they also required to wear helmet for safety reasons. Therefore, the aim of research is to detect the helmets and masks worn by workers, if there the workers were found not for not wearing the mask properly an alert will be triggered. Five classes were determined namely 'Head,' 'Helmet,' 'Incorrect Mask,' 'No Wearing Mask', and 'Wearing Mask'. A total of 1711 images of construction workers scenes were collected, and augmentation was applied on these images to generate 4733 images. All images were annotated corresponding to the defined classes. The experiments have split the training and validation dataset into a ratio of 9:1. The result obtain a 65.235% Mean Average Precision (mAP) as a result.

Keywords: deep learning, object detection, you only look once (YOLO), mask detection, helmet detection.

1. Introduction

In the past two years, COVID-19 has brought severe damage globally, and it has damaged the global economy and brought death to the public. According to the research (Punsalan et al., 2021), wearing face masks is the key to reducing virus transmission among humans. To make sure all citizens wear face masks during this epidemic period and adapt to this situation, plenty of mask detection and face detection experiments have been implemented like this research (Chavda et al., 2021).

Thanks to the arrival of the COVID-19 vaccine, there is a lot of economic exercises have started operating again, which included the construction and industry field. The industry field operation already had essential equipment to be equipped while operating. And one of the vital pieces of equipment that needs to be fitted is the safety helmet. There are also plenty of safety helmet detection experiments using deep learning that have been implemented.

Now, here is the important part: due to the COVID-19, the essential equipment that is compulsorily needed to be equipped while operating in construction and industry is no longer only the safety helmet anymore but also included the face mask to prevent the spread of COVID-19. Wearing both mask and helmet while operating in construction is no longer a choice for workers. Still, it is a necessary behaviour for workers to be done to protect themselves and those around them. Apart from that, during the Movement Control Order (MCO) period, a lot of organizations or companies have strictly stipulated their employee must wear a mask while doing their job.

Wearing a face mask is a very efficient way to against the spread of COVID-19. This study pointed out that wearing a face mask is a very efficient way to reduce the spread and transmission of the COVID-19 virus (Habib et al., 2022). But even though many reports and research or even media have warned the public about the importance of wearing a mask during this COVID-19 pandemic period, some countries even make wearing a mask when you are going out a law during this pandemic period. But there is still a group of people who fail to follow this basic and straightforward order during the pandemic. According to the research (He et al., 2021), this group of people always use uncomfortable while wearing a mask, which might bring adverse effects wearing the mask, lack of effectiveness or it is an unnecessary behavior etc. as the reasons for standing the opposition side of wearing the mask.

Moving forward, let us look at the safety helmet side. Wearing a safety helmet while operating in the construction is always the priority safety issue that companies and workers should alert from. Unlike the mask during the pandemic period, a safety helmet is always a piece of significant safety equipment that workers should be equipped with while operating in construction. But just like a mask, there is still a group of persons who fail to follow this instruction and do irrecoverable damage.

According to the Health and Safety Executive report (Fawzi et al., 2016), construction workers suffer 10% of major injuries and 31% of fatal injuries. A considerable number of workers get injured while operating in the construction.

By investigating the safety issue consisting of the mask and safety helmet, we can observe that there is always a very difficult mission to keep all people following one rule by giving them orders. But during this pandemic period, some action must consistently be implemented to keep people safe. Simpeh et al., (2022) suggested that the company should strictly stipulate at the workers wear both masks and helmets while working to keep everyone safe.

2. Literature Review

Chavda et al. (2021) have divided the process of building a face mask detection model into two stages; first, when the image is inputted, the model will run stage 1, which is face detection; in this stage, they are using RetinaFace as the face detection model. First, the Region of Interest (R.O.I.) of every identified face is anticipated to emerge as a consequence of step 1; in this R.O.I., the bounding boxes are stretched up to 120 per cent to minimize any overlap with other faces. Second, simultaneously, the image's enlarged bounding boxes are extracted to retrieve the R.O.I. for each recognized face and batch it together to save processing time. As the core function of this dual-stage mask detection model, stage 2 uses the Convolutional Neural Network (CNN) model as the face mask classifier. By comparing different CNN model, this experiment has chosen NASNetMobile as the final model, since it obtains an accuracy of 99.23% in the test.

Zhou et al. (2021) have developed a helmet detection system with YOLOv5 for a better working environment. There are a total of 6045 pictures in the Dataset, and the Dataset contains the large scale, medium, and small scale as positive samples and some classroom scenes as negative samples. This study has proved that different YOLOv5 model will bring different result, but it not always brings a very significant improvement. The improvement between YOLOv5s and YOLOv5m is 1.3%, but the different between YOLOv5l and YOLOv5m is only 0.1%.

Ieamsaard et al. (2021) have used 853 images as the Dataset, and those images are labelled with ‘With_Mask’ and ‘Without_Mask’, and ‘Incorrect_Mask’. The photos are divided into a training set with 682 illustrations, a testing set with 96 images, and a validation set with 85 images. This study using YOLOv5 as the training model. This study obtains the greatest result in 300 epochs instead of 500 epochs, which prove that the higher epochs doesn’t always mean the greater performance.

Y. Li et al. (2020) used 3261 images as the Dataset, and the Dataset is divided into a training set, testing set, and validation set with a ratio of 6:2:2. The Dataset is pre-processed, such as rotation and zooming, to increase object detection performance. The TensorFlow framework is used to train the model in this

experiment. The pictures' attributes are extracted using the pre-trained SSD mobile net v1 COCO model and the COCO dataset. Through these experiments, we observed that the object background complexity will hugely affect the object detection result. This study obtains a precision of 95% and 77% recall at the end of the study.

Basha et al. (2021) captured 5,000 portraits of 525 individuals wearing masks and 90,000 pictures of the same 525 subjects with no pretences, which is a total of 95000 datasets. They start to pre-process it; in this experiment, they have performed random resized cropping and a vertical flip to each image dataset. Besides that, they also make the Dataset square images with a usual pixel size of 224 x 224 since most deep neural networks, including ResNet, require square images as input. In this experiment, data augmentation explained by Fawzi et al., (2016) also be introduced and implemented. After the pre-processing stage, 75% of the images were selected for training and the rest were used as testing set. The Adam Optimizer using binary cross-entropy was used to generate the model. The mark detection system was able to show excellent result with accuracy of 97.8% with ResNet50. We also have observed the model trained very well on an extensive dataset which is Real-World Masked Face Recognition Dataset (RMFRD).

Gu et al. (2019) has proposed a model based on Faster Region-Based Convolutional Neural Networks (RCNN) but with some model optimization. Firstly, the multi-scale technique has been applied to raise the resilience of the item‘s size and increase the number of anchors to improve the accuracy in a small-scale thing. It is furthermore utilizing Online Hard Example Mining (OHEM) to identify sample sand train networks automatically. Lastly, offer a multi-part technique to eliminate false detection. The Dataset involved in this experiment comprises the VOC2012 Dataset, self-collection, and internet collection; there are 7000 photos incorporated, and obtained an outcome with 91.51% recall and 94.56% precision as a result.

Jian and Lang (2021) have created a face mask detection model based on PaddlePaddle – You only look once (PP-YOLO) and enhanced the model with transfer learning due to insufficient data samples. This experiment contains 7959 images for the Dataset. In addition, this experiment uses Fine-tune as the transfer learning strategy. They first pre-trained the PP-YOLO with the Dataset annotated in PascalVoc format, and a set of the Dataset with Mix-up data is sent to migration training. This experiment obtains the highest 89.69% mAP by using PP-YOLO-mask. Through these experiments, we also can conclude that some enhancement strategy is recommendable to enhance our model.

Wang et al. (2020) proposes a cross-stage YOLOv3 (CSYOLOv3) model to satisfy the need for a particular safety helmet dataset. The Dataset used in this experiment is self-built with a COCO dataset format. The experiment shows that the CSYOLOv3 performs better than other models. It obtains a mAP of 67.05%, while another model don’t even surpass 50%. In conclusion, some of the original models

might provide a stable platform to experiment, but the original is the best to use; some improvements might require modification to improve the result.

3. Proposed Model (YOLOV5)

YOLOv5 is a deep learning object detection developed by Jocher et al., (2020) using PyTorch. Thus, it can benefit from the ecosystem of PyTorch, which includes but is not limited to more superficial support and easier deployment. One of the biggest reasons that make YOLOv5 popular is the productivity; YOLOv5 is very fast in terms of processing speed, but even though it is speedy, it still can balance the accuracy, which makes it even better.

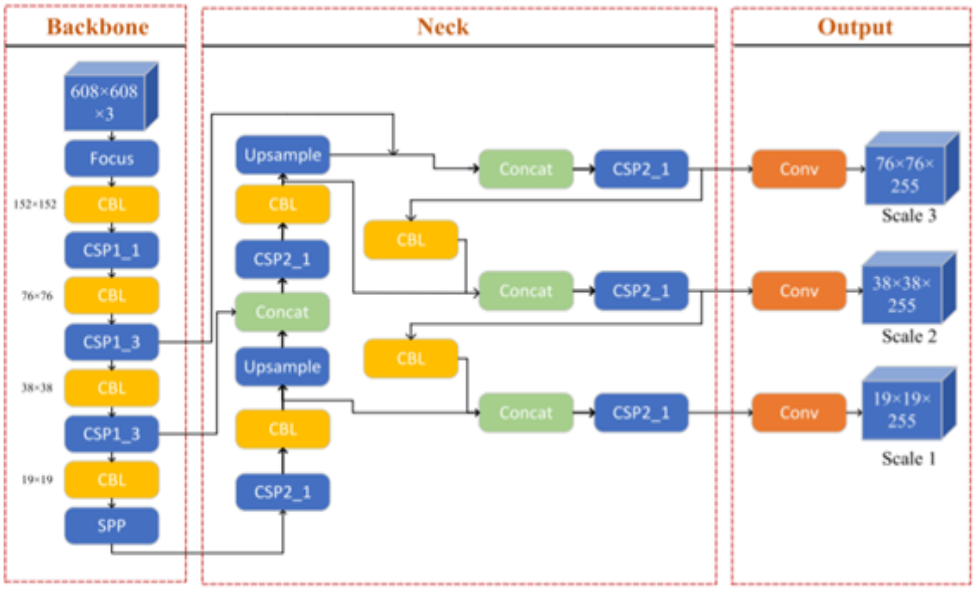

Fig. 1: YOLOv5 architecture (Zhu et al., 2021).

Figure 1 comes from this research (Zhu et al., 2021) shows the architecture of YOLOv5; YOLOv5 consists of three different parts, each one responsible for other work; in general, the backbone is accountable for generating the feature map, the Neck accountable for the feature extraction on the feature map, and Prediction responsible for the output. After this study has a brief introduction to YOLOv5 architecture, this study will go into detail for further understanding. The image will first get into the focus layer in the backbone, which will slice and concatenate the images for better feature extraction. A whole backbone network is CNN, which extracts feature maps of the images by multiple and pooling. The neck part will analyse the feature maps that come from the backbone to obtain more details and gain more information by doing this process and reducing information loss. The Feature Pyramid Network (F.P.N.) and Path Aggregation Network (P.A.N.) will be used in this process, the F.P.N. responsible for conveying solid semantic features from the

top features maps into lower feature maps. While the P.A.N. is accountable for conveying robust localization features from lower feature maps into higher feature maps. After everything is done, the prediction part will be responsible for the classification and prediction part.

4. Project Implementation

This study describes the implementation of this project. This chapter cover the dataset collection and annotation processes.

4.1. Dataset and annotation

The images of construction scenes were needed to be annotated and train the YOLOv5 model. Five classes and the number of annotations were identified in Table 1.

Table 1: Table of different classes.

Classes	Number of Annotation	Description
Head	2172	Detect all head that appears in each frame
Helmet	1015	Detects all helmets that appear in each frame
Incorrect Mask	918	Detect faces wearing the mask wrongly
No Wearing Mask	1114	Detect a person's face that doesn't wear a mask
Wearing Mask	1212	Detect persons wearing a mask

After introducing our dataset classes, this study shall move forward to the dataset collection part. This study has successfully collected and annotated a total of 1711 images. Besides that, thanks to the convenience of Roboflow (Solawetz and Nelson 2020), this study has obtained 4733 images which are three times the original images, after completing the data augmentation and without costing any time. Roboflow offers an easy data augmentation function for users by clicking and setting up a few things. After introducing the number of Dataset, this study would like to introduce some details about the annotation and the measurement to choose the Dataset.

This study has focused on choosing images that contain either helmet or mask, and not only simple as that, and this study also focuses on searching images that contain the object in different angles and colors or scales to train a model that can fit into different condition.

Fig. 2: Dataset of helmet.

This study has collected helmet datasets that contain multiple scales, angles and even colours. The data annotation also follows the size of the helmet to produce a better YOLOv5 weight. An example of a helmet dataset can be found in Figure 2.

Fig. 3: Dataset of wearing the mask.

The helmet and the mask dataset have also been collected under different measurements, which included but were not limited to the colour scale or angle of those images in this study. Figure 3 has the example images for reference.

Fig. 4: Dataset of not wearing the mask

Fig. 5: Dataset of Incorrect wearing mask

Figure 4 shows the different scales or angles of that person not wearing the mask. Still, at the same time, this study also annotated the wearing mask category in the same images, which means this study is flexible to do the data annotation if there is an object that fulfils the classes.

Figure 1 shows the images of that incorrect mask. Just like the previous, this study also collects images of the incorrect wearing mask under different measurements to develop a better object detection system.

Finally, this study should talk about the "Head" classes, and the head classes are the easiest Dataset that can be found among all resources. Every image must contain at least a head. This study will label those heads as "Head" classes. But why some of the example images can't see the head annotation is due to the data being overrepresented. The overrepresented data might affect the real-time model performance, so therefore some of the head has been annotated as the class 'Head', but some of them are not annotated as the class 'Head'.

4.2. The difficulty of data collection

Even though the overall dataset collection and performance are acceptable, this study still faces some difficulties while collecting the data—for example, the Incorrect wearing mask class. Unlike the helmet or mask dataset, it is hard to search Dataset that specifically describes a person who wears a mask incorrectly. In this case, this study has adopted the MaskedFace-Net (Cabani et al., 2021) a dataset of correctly or incorrectly wearing the mask, as the main incorrectly wearing mask dataset resource to be imported in this study. Figure 6 shows the incorrectly wearing mask dataset that collects from MaskedFace-Net. The Dataset from MaskedFace-Net is editing a mask into a human face so that the model can detect it as a wearing mask object.

Fig. 6: Incorrectly wearing mask dataset from MaskedFace-Net (Cabani et al., 2021).

4.3. Train the YOLOv5 model

As mentioned before, this study has labelled the dataset into 5 classes, which are "Head", "Helmet", "Incorrect Mask", "No Wearing Mask" and "Wearing Mask". Consider the training environment and how to find the optimal batch size.

This study completed the model training with 3 different batch sizes. The training per epochs in different batch sizes and the training loss are considerable values while choosing the best batch size. The batch size result is shown in Table 2.

Table 2: Small experiment result through different batch size.

Batch Size	Training Speed per epochs (mins)	Training Loss
16	14	0.015912
32	11	0.016435
46	10	0.016749

Even though the result performs better in batch size 32, but the training speed should also be one of the concern factors might be needed to focus. After weighing the benefits, this study decides to adopt a batch size equal to 46 as the final decision.

After decided the batch size, this study also go through a comparison between different epochs to find the optimal epochs. In epochs, the mAP is the considerable value while choosing the epochs. The result of different epochs is shown is Table 3.

Table 3: Small experiment result through the different epoch.

Epochs	Precision	Training Loss
40	0.69431	0.019337
50	0.71449	0.017703
60	0.69572	0.016749

Considering the time consumed and the accuracy and the training loss, this study found out that an epoch equal to 60 is the most suitable parameter for this study to use. Even though precision resulted better in epochs 50, but the overall score shows that the epochs 60 are consider good as well.

To find the most suitable YOLOv5 model to be used in this study, this study has go through a comparison between different YOLOOv5, the result is shown in table 4.

Table 4: Comparisons between YOLOv5m, YOLOv5s and YOLOv5n.

Model	Precision	Training Loss	mAP
YOLOv5n	0.63424	0.02256	0.62755
YOLOv5s	0.66112	0.019868	0.63222
YOLOv5m	0.69572	0.016749	0.65235

Table 4 shows the overall result between three YOLOv5 models, and these comparisons are running in the same parameter. By looking at the table already can conclude that YOLOv5m has an absolute benefit compared with YOLOv5n and YOLOv5s. After completed the comparison between different YOLOv5 model.

4.4. Difficulty while training the YOLOv5 model

This study has faced a major training environment problem. Due to the low-level hardware, the training progress shall also consider some parameters that will affect the hardware memory.

5. Experimental Results & Outcome

This study will simply explain the outcomes throughout this project, which include the results of the experiments and some of the visualizations generated throughout this project.

5.1. Results obtain through dataset training

Table 4 also show the experiment result at the end of this study. We adopt the result and weights trained by YOLOv5m as the real-time detection weights. To evaluate the model, we shall focus on mAP as the overall results since mAP represent the accuracy in YOLOv5. This study has received mAP 0.65235 which 1 represent the highest value. We can conclude even though, it performs good overall, but there is still a greatest improvement can be made. Now, we shall look at to section 5.2 in term of real-time detection.

5.2. Real-time detection

This section shows some real-time mask and helmet detection using web camera (Figure 7 to 10) and videos (Figure 11 and 12). From the observation, the system works well on detecting the proper and improper wearing of mask and helmet. It also shows good detection performance on video scene with more than one worker as shown in Figure 11 and 12.

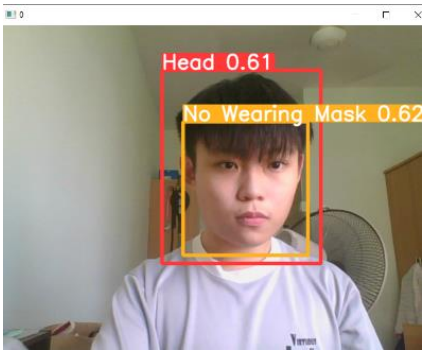

Fig. 7: Not wearing mask result in webcam.

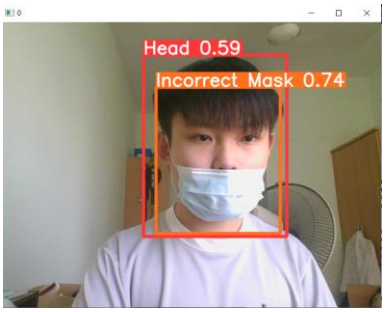

Fig. 8: Wearing mask result in webcam.

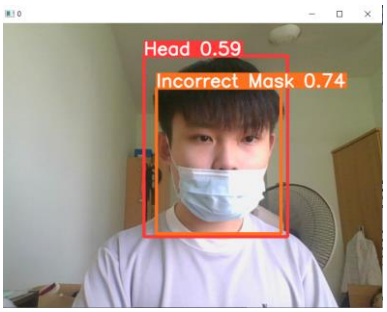

Fig. 9: Incorrect wearing mask result in webcam.

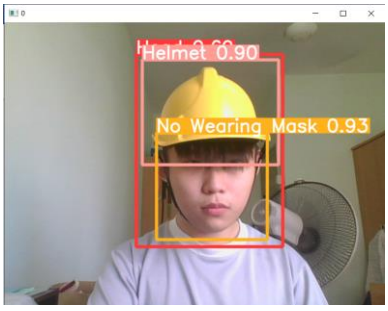

Fig. 10: Wearing helmet result in webcam.

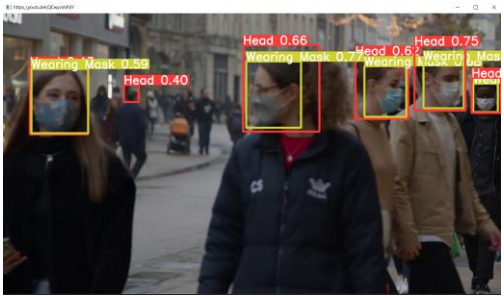

Fig. 11: Multiple object detection video with mask.

Fig. 12: Multiple object detection video with helmet

6. Conclusion

This study has introduced a real-time helmet and mask detection system using YOLOv5. This study has complete a process of image dataset collection, dataset annotations, parameter comparison, as well as system testing. This study has first collected a dataset up to 1711 images and labelled those images into the classes which are “Head”, “Helmet”, “Incorrect Mask”, “No Wearing Mask” and “Wearing Mask” according to the YOLOv5 format. After completed the dataset annotation part, this study has applied the data augmentation strategy into those images, which make this study obtain an image dataset up to 4733 images.

This study also completed some parameter comparison to obtain the optimal training parameter, but unfortunately, due to the restrictions of training environment, the training time are also considered as one of the considerable factors while training the model. Finally, this study obtains a mAP of 65.235% as a result, and a great performance in term of webcam or video that contain multiple objects in the same scene.

Acknowledgments

This project is supported by the TM R&D (Project no. RDTC/221054) and MMU IR Fund (Project ID MMUI/210108).

References

Basha, C. Z., Pravallika, B. N. L., & Shankar, E. B. (2021). An efficient face mask detector with pytorch and deep learning. EAI Endorsed Transactions on Pervasive Health and Technology, 7(25), 1–8. DOI:https://doi.org/10.4108/eai.8-1-2021.167843.

Cabani, A., Hammoudi, K., Benhabiles, H., & Melkemi, M. (2021). MaskedFace-Net – A dataset of correctly/incorrectly masked face images in the context of COVID-19. Smart Health, 19(November 2020), 100144. DOI:https://doi.org/10.1016/j.smhl.2020.100144.

Chavda, A., Dsouza, J., Badgujar, S., & Damani, A. (2021). Multi-stage CNN architecture for face mask detection. 2021 6th International Conference for Convergence in Technology, I2CT 2021, 1–8. DOI:https://doi.org/10.1109/I2CT51068.2021.9418207.

Executive, H. S. (2015). Health and safety in construction sector in Great Britain, 2014/15. *Health and Safety Executive.*

Fawzi, A., Samulowitz, H., Turaga, D., & Frossard, P. (2016). Adaptive data augmentation for image classification. *Proceedings - International Conference on Image Processing, ICIP*, 3688-3692. DOI:https://doi.org/10.1109/ICIP.2016.7533048.

Gu, Y., Xu, S., Wang, Y., & Shi, L. (2019). An advanced deep learning approach for safety helmet wearing detection. Proceedings - 2019 IEEE International Congress on Cybermatics: 12th IEEE International Conference on Internet of Things, 15th IEEE International Conference on Green Computing and Communications, 12th IEEE International Conference on Cyber, Physical and So, 669–674. DOI:https://doi.org/10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00128.

Habib, S., Alsanea, M., Aloraini, M., Al-Rawashdeh, H. S., Islam, M., & Khan, S. (2022). An efficient and effective deep learning-based model for real-time face mask detection. *In Sensors*, 22(7). DOI:https://doi.org/10.3390/s22072602.

He, L., He, C., Reynolds, T. L., Bai, Q., Huang, Y., Li, C., Zheng, K., & Chen, Y. (2021). Why do people oppose mask wearing? A comprehensive analysis of U.S. tweets during the COVID-19 pandemic. *Journal of the American Medical Informatics Association*, 28(7), 1564–1573. DOI:https://doi.org/10.1093/jamia/ocab047.

Heikal Ismail, M., Ghazi, T. I. M., Hamzah, M. H., Manaf, L. A., Tahir, R. M., Mohd Nasir, A., & Ehsan Omar, A. (2020). Impact of movement control order (MCO) due to coronavirus disease (COVID-19) on food waste generation: A case study in Klang Valley, Malaysia. *In Sustainability*, 12(21). DOI:https://doi.org/10.3390/su12218848.

Ieamsaard, J., Charoensook, S. N., & Yammen, S. (2021). Deep learning-based face mask detection using YoloV5. *Proceeding of the 2021 9th International Electrical*

Engineering Congress, IEECON 2021, 428–431. DOI:https://doi.org/10.1109/iEECON51072.2021.9440346.

Jian, W. & Lang, L. (2021). Face mask detection based on Transfer learning and PP-YOLO. *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, ICBAIE 2021*, *Icbaie*, 106–109. DOI:https://doi.org/10.1109/ICBAIE52039.2021.9389953.

Jocher, G., Nishimura, K., Mineeva, T., & Vilariño, R. (2020). Yolov5. Code Repository Https://Github. Com/Ultralytics/Yolov5.

Karthi, M., Muthulakshmi, V., Priscilla, R., Praveen, P., & Vanisri, K. (2021). Evolution of YOLO-V5 algorithm for object detection: Automated detection of library books and performace validation of dataset. *Proceedings of the 2021 IEEE International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES 2021*, 1–6. DOI:https://doi.org/10.1109/ICSES52305.2021.9633834.

Li, H., Li, X., Luo, X., & Siebert, J. (2017). Investigation of the causality patterns of non-helmet use behavior of construction workers. *Automation in Construction*, 80, 95–103. DOI:https://doi.org/10.1016/j.autcon.2017.02.006.

Li, Y., Wei, H., Han, Z., Huang, J., & Wang, W. (2020). Deep learning-based safety helmet detection in engineering management based on convolutional neural networks. *Advances in Civil Engineering*, 9703560. DOI:https://doi.org/10.1155/2020/9703560

Punsalan, M. L. D. & Salunga, A. T. (2021). Mask is a must: the need of protection and safety against COVID-19. *Journal of Public Health*, 43(2), e379–e380. DOI:https://doi.org/10.1093/pubmed/fdab077.

Severance, C. (2015). Guido van Rossum: The early years of python. *Computer*, 48(2), 7–9. DOI:https://doi.org/10.1109/MC.2015.45.

Simpeh, F. & Amoah, C. (2022). COVID-19 guidelines incorporated in the health and safety management policies of construction firms. *Journal of Engineering, Design and Technology*, 20(1), 6–23. DOI: https://doi.org/10.1108/JEDT-01-2021-0042.

Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., & Agha, R. (2020). World health organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery*, 76, 71–76. DOI: https://doi.org/10.1016/j.ijsu.2020.02.034.

Solawetz, J. & Nelson, J. (2020). How to train yolov5 on a custom dataset. Roboflow Blog.

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

Zhou, F., Zhao, H., & Nie, Z. (2021). Safety helmet detection based on YOLOv5. *2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA)*, 6–11. DOI: https://doi.org/10.1109/ICPECA51329.2021.9362711.

Zhu, L., Geng, X., Li, Z., & Liu, C. (2021). Improving yolov5 with attention mechanism for detecting boulders from planetary images. *Remote Sensing*, 13(18), 1–19. DOI: https://doi.org/10.3390/rs13183776.