Facial Mask Detection and Energy Monitoring Dashboard Using YOLOv5 and Jetson Nano

Nur Asyiqin binti Amir Hamzah[1], Hadhrami bin Abd Ghani[2](✉), Hatem Fahd Al-Selwi[1], Nawaid Hassan[1], and Azlan bin Abd Aziz[1]

[1] Multimedia University, Cyberjaya, Malaysia
[2] Universiti Malaysia Kelantan, Kota Bharu, Malaysia
hadhrami.ag@umk.edu.my

Abstract. Health and energy are two important elements of life that must be constantly monitored. In the present pre-endemic or post-pandemic age, it is critical to monitor individuals to ensure that crucial standard operating procedures, such as wearing face masks, are followed. Furthermore, any measures to reduce energy use, particularly in public buildings that are now being utilised again in the post-pandemic context, are critical. As a result, this study presents a live dashboard method for tracking both energy and face mask wear. The detection system is built on a Jetson Nano device and is then interfaced to a live dashboard to monitor people who wear masks, including those who wear them incorrectly. The observed findings for the confusion matrix and F1 score are also shown. At the same time, the suggested visual detection system is also capable of monitoring the energy use in the building where the device is mounted.

Keywords: Dashboard · Face Mask Detection · Counting · YOLOv5 · Jetson Nano

1 Research Background and Related Works

In order to minimize the spread of disease, social distancing, mask wearing and temperature screening are helpful methods. Many organizations, including the World Health Organization (WHO), have recommended them. The effects of social distancing measures on the spread of coronavirus were investigated by Russell et al. [1] and, Nabil Kahale [2].

The goal of the study was to come up with an approximation that indicates how early social distancing measures can significantly reduce economic loss and the rate of new infections. When the coronavirus initially began spreading among people and society, researchers and scientists began to look for the best way to stop the pandemic from spreading [3]. Jennifer Berglund [4] proposed using GPS and developed software in smartphones to track an affected person with corona virus. However, this technology has limitation when it comes to detect those who don't have availability of wireless network or smartphone. Some authorities, on the other hand, use drone technology to monitor

M. Y. bin Alias (Ed.): MECON 2022, AER 214, pp. 49–57, 2023.
https://doi.org/10.2991/978-94-6463-082-4_7

people in the open area [5]. This method is appropriate for screening COVID-19, which may be prominent during the coronavirus crisis.

Recently, due to advancement and improvements in the field of computer vision and artificial intelligence, the challenge of detecting and classifying elements in an image has now become solved. Computer vision research has emphasized on a range of interesting and challenging tasks, such as neural extracting features, categorization, and monitoring, and also, the image detection [6]. In deep learning techniques for object detection, a convolutional neural network (CNN) is frequently employed. CNN is a deep learning technique that takes an input image and adds learnable weights and biases to multiple classes in the image, enabling them to be separated from each other.

The development of a convolutional neural network that can be implemented on an integrated system with low-resolution input and high efficiency [7]. R-CNN, Single-shot detector (SSD), and YOLO are some of the deep learning models used in computer vision in variety of areas. These models have excellent algorithmic approaches for estimating movement in video frames and object detection. Ebrahim et al. [8] proposed a technique for detecting persons by analysing video frames. For persons detection, the author used a deep learning detector and a background removal and Gaussian combination. The author implemented a deep learning (CNN) algorithm for human detection in [9]. They used a hybrid of deep learning and machine learning techniques to detect persons with high precision and limited processing. However, this approach met challenges with real-time detection due to its slow efficiency. Researchers proposed a technique based on crowds that remained in the same spot over an extended period of time in approach [10]. They used a support vector machine (SVM) to classify patches as essential crowds, and text features were used to extract these patches.

Recent research has shown that a person's facial [11] and walking style can be used to identify them through video surveillance cameras. However, detecting a person in a crowd is difficult and difficult to optimise. In method [12], the authors presented a method for recognizing pedestrians with a low-resolution camera using background reduction and classifying them in real time.

Object detection techniques create a frame around the objects and associate each frame with the appropriate object category. Deep learning is an efficient way for detecting objects. Ross Girshick researched a regional convolutional neural network (R-CNN) detector in [13]. The same researcher of the R-CNN model has developed new version of the regional proposal method, Fast R-CNN [14]. Fast R-CNN enhanced the drawbacks from R-CNN to build faster object detection algorithm.

Other techniques, such as YOLO, have recently improved object detection deep learning. You only look at one or (YOLO) is a cutting-edge deep learning object detection algorithm. Joseph Redmon et al. were the ones who presented it [15]. To process the entire image, YOLO employs a single neural network. It separates the image into regions and predicts each region's bounding frames and its probability. These bounding frames are weighted by predicted probabilities. YOLO has a significant and effective record of actual object detection. It is fast, precise, and simple to train.

Therefore, this project presents a visual detection system based on YOLOv5 algorithm for detecting the face mask wearing as well as monitoring the energy usage in the surrounding areas of the building in which the visual detection system is placed. The visual detection scheme is installed on a Jetson Nano device, which is also connected to a dashboard for live monitoring.

2 Methodology

In this project, we developed an end-to-end monitoring system. This monitoring system is placed in a Masjid, in which many people are observed to enter and exit this building. The system consists of three modules which are detection, server and visualization. The detection module is responsible for detecting and counting people as well as detecting facemask.

The module mainly consists of Jetson Nano computing board equipped with a camera. The server module is responsible for reserving video stream and detection data from the detection module and storing detection data in a database. Meanwhile, the visualization module is a dashboard which is responsible for visualizing video stream, detection data and statistics. Figure 1 shows the framework of the developed system.

2.1 Detection Module

In the detection module, YOLOv5 for people and facemask detection has been applied. The model is trained on facemask dataset. A pre-trained YOLOv5 model is used and then trained using the custom dataset to finetune the model for the task.

2.1.1 Dataset Processing

We utilize a public facemask dataset to train the model. However, since the dataset is not readily designed for YOLOv5 implementation, the data is formatted using Roboflow service. The service allows the user to label datasets and reformat annotations for YOLOv5. The dataset is categorized as training, validation and testing, where each category is 87%, 8% and 4% respectively.

2.1.2 Training

The model is trained for 100 epochs using YOLOv5s as initial weights. The model is trained on a computer with the following configuration: 16GB RAM and Nvidia GTX 6GB GPU. For training and testing, PyTorch framework is applied.

2.1.3 Inferencing

After the model have been trained, the model is run with some inference on our Jetson Nano on video stream in real-time. The model will detect facemask status either facemask is worn correctly or not.

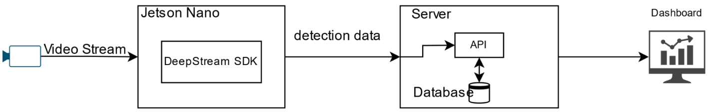

Fig. 1. The framework of the proposed Masjid monitoring system

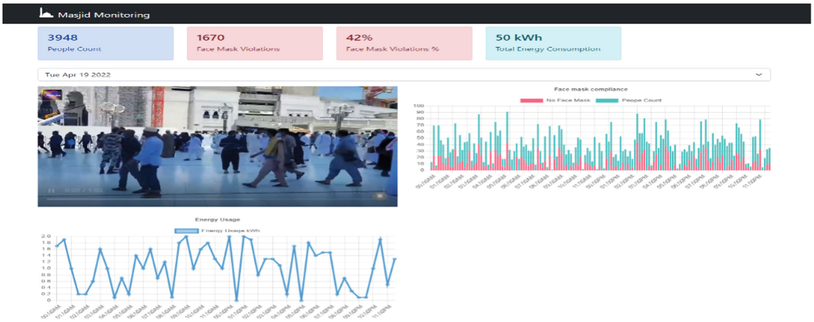

Fig. 2. The Face Mask Detection and Counting Dashboard for Masjid Monitoring

2.1.4 People Counting

To perform people counting, a pretrained YOLOv5 model is used to detect people in a video frame. The model will detect people from a video stream, and at the same time counting number of the people.

2.1.5 Video Streaming

Nvidia DeepStream SDK is used to stream video over Real Time Streaming Protocol (RTSP) to display the video on the dashboard and send detection data to the server to process and store it in the database.

2.1.6 Light Detection for Energy Saving

The system gathers the light data from the video stream to detect light in specific area and then store the data for the purpose of energy usage analysis as well as energy saving.

2.2 Server Module

In the server module, basic server is utilized to retrieve data from detection module as video stream over RTSP and detection data over MQ Telemetry Transport (MQTT), the retrieve data is then stored in a real-time database, so that the stored data can be used for analysis and visualization.

2.2.1 Data Collection

The server retrieves detection data from detection module in JSON format in real-time with one minute interval, The retrieved data consists of detection data of mask detection coming from mask detection model, and people count from people detection model as well as light detection data. The data is stored in database with timestamps.

2.2.2 Data Analysis

The data stored of both facemask detection and people count are used to statistically analyze people compliance with facemask rules. Moreover, the stored light data is used to analyze energy usage based on the time for lights and the count of detected people in the area. This will also allow us to control lights based on occupancy of the area.

2.3 Visualization Module

In the visualization module, we apply a JavaScript web application to retrieve data and video stream from the server and visualize it on a dashboard using charts to show statistical and historical data as well as the video stream in real-time from the cameras.

2.3.1 Data Visualization

ChartJs library is used to visualize our historical data as well as statistical data in a comprehensive way. The dashboard consists of people count chart, facemask chart and energy chart. The people counting chart shows historical data of the number of people that visit the Masjid each day. Meanwhile, the facemask chart shows number of people wearing mask against facemask violations in each minute. On the other hand, the energy chart shows the on-time for lights against number of people detected as well as time of day.

2.3.2 Video Stream

The video stream is shown over the RTSP channel in a dashboard to show video stream in real-time from detection module.

3 Results

In this section, the results of the face detection and counting technique based on YOLOv5 algorithm are presented with a live dashboard for monitoring. As seen in Fig. 2, a snapshot of the dashboard of the proposed monitoring system is given. The dashboard has four sections namely, total stats, video stream, face mask compliance chart, and energy usage chart. The dashboard also allows us to observe selected days from the select menu.

The dashboard shows the total statistics for detected people, face mask violations as well as energy consumption detected using our vision system. The energy consumption is calculated based on pre-defined lights usage and then our vision system will detect the status of light in real-time and calculate usage based on camera area and assigned lights.

The facemask chart in Fig. 3 shows the facemask violations against the number of people. By plotting these facemask violations observation in the chart, the person-in-charge of the building will attain an insight of the number of violations as well as the corresponding violations percentage based on the number of people present in each time interval.

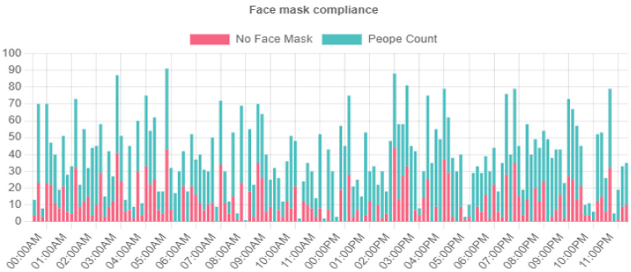

Fig. 3. The Facemask Chart

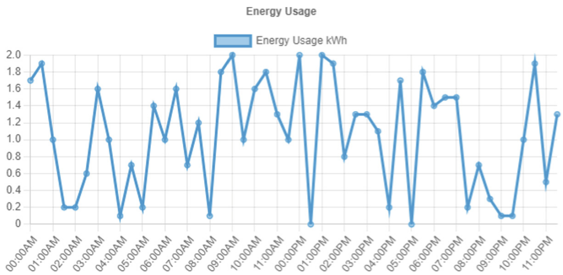

Fig. 4. The Energy Usage

The energy usage chart in Fig. 4 shows the energy consumption every 30 min based on the detected illumination and the pre-defined lights usage in the observed area by the proposed vision system. This is a useful approach to monitor the energy usage in order to save the total energy utilized in the building.

For further analysis on the performance of the face detection and counting model implemented in this paper, the confusion matrix of the detection results is produced, as shown in Fig. 5. From this figure, it can be learnt that the proposed face detection model has a high confusion value when detecting apart between the people who wear the masks correctly and those who wear the masks incorrectly. The recorded confusion value is 0.43. This result is no surprise due to the close similarity between the people who wear masks correctly and those who wear incorrectly. However, when comparing between the people who wear masks and those who are not wearing masks, the model has performed well, as proved by the low confusion value of 0.02 recorded in the confusion matrix between these two classes.

The next result is about the F1 score of the proposed model, as shown in Fig. 6. Of the four classes observed, the with_mask class, which represents the people who wear the masks correctly, has the highest F1 score, which is more than 0.8 in most of the confidence value range. As for the mask_weared_incorrect class, which represents the people who wear masks incorrectly, the F1 score recorded is the lowest amongst all classes.

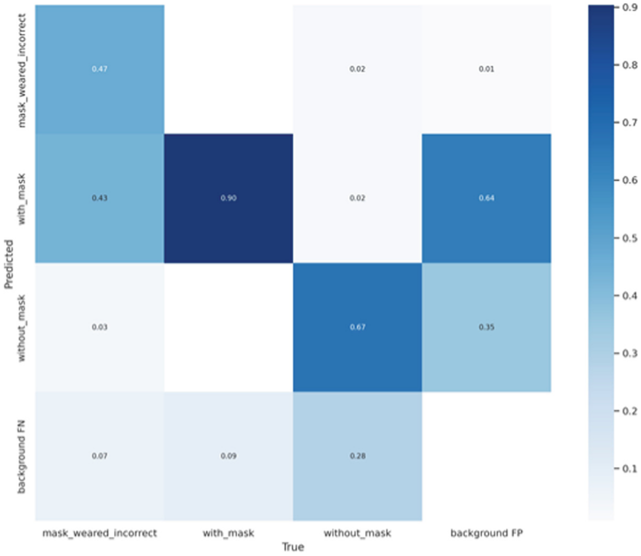

Fig. 5. The Confusion Matrix

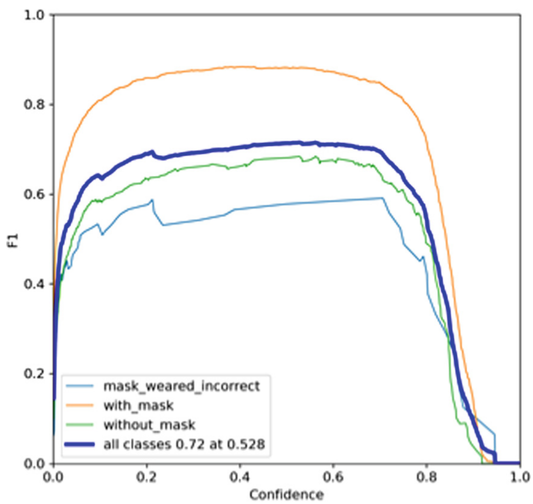

Fig. 6. F1 Score

4 Summary

This paper presents a new dashboard for live recording of the detection and counting results of the people wearing the masks in the area of the buildings such as mosques and offices. Furthermore, the energy usage in the building is also monitored to avoid any energy wastage. The YOLOv5 algorithm applied for detecting and counting the people who wear masks, correctly or not, is found effective with acceptable accuracy values and relatively low confusion values.

Acknowledgments. The authors acknowledge and thank all the support from the Tabung Amanah Zakat Multimedia University for funding this research under the FRDGS fund, with the grant code of No. 11 (MMUE/210032). Not to forget the owner of Ultralytics Github (available at: https://github.com/ultralytics/YOLOv5) for YOLOv5 coding guidelines. Finally, we appreciate the support from our respective faculties, as well as all other individuals who are directly or indirectly involved in preparing this paper.

References

1. K. Prem, Y. Liu, T.W. Russell, A.J. Kucharski, R.M. Eggo, N. Davies, S. Flasche, S. Clifford, C.A.B. Pearson, J.D. Munday, S. Abbott, H. Gibbs, A. Rosello, B.J. Quilty, T. Jombart, F. Sun, C. Diamond, A. Gimma, K. van Zandvoort, S. Funk, C.I. Jarvis, W.J. Edmunds, N.I. Bosse, J. Hellewell, M. Jit, P. Klepac, The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study, Lancet Public Heal. 5 (2020) e261–e270. https://doi.org/10.1016/S2468-2667(20)30073-6.
2. S. Chakraborty, M.C. Biswas, Pr ep rin t n ot er re vie we ep rin t n ot er re vie we, 9 (2020) 4006–4010.
3. S.K. Sonbhadra, S. Agarwal, P. Nagabhushan, Target specific mining of COVID-19 scholarly articles using one-class approach, Chaos, Solitons and Fractals. 140 (2020) 110155. https://doi.org/10.1016/j.chaos.2020.110155.
4. J. Berglund, Tracking COVID-19 : There ' s an App for That, (2020) 2020–2023.
5. M. Robakowska, A. Tyranska-Fobke, J. Nowak, D. Slezak, P. Zuratynski, P. Robakowski, K. Nadolny, J.R. Ładny, The use of drones during mass events, Disaster Emerg. Med. J. 2 (2017) 129–134. https://doi.org/10.5603/demj.2017.0028.
6. P. Atkinson, Social Distancing, Des. J. 23 (2020) 327–330. https://doi.org/10.1080/14606925.2020.1747891.
7. X. Wang, H.W. Ng, J. Liang, Lapped convolutional neural networks for embedded systems, 2017 IEEE Glob. Conf. Signal Inf. Process. Glob. 2017 - Proc. 2018-January (2018) 1135–1139. https://doi.org/10.1109/GlobalSIP.2017.8309138.
8. E.N. Kajabad, S. V. Ivanov, People Detection and Finding Attractive Areas by the use of Movement Detection Analysis and Deep Learning Approach, Procedia Comput. Sci. 156 (2019) 327–337. https://doi.org/10.1016/j.procs.2019.08.209.
9. A. Brunetti, D. Buongiorno, G.F. Trotta, V. Bevilacqua, Computer vision and deep learning techniques for pedestrian detection and tracking: A survey, Neurocomputing. 300 (2018) 17–33. https://doi.org/10.1016/j.neucom.2018.01.092.
10. M. Manfredi, R. Vezzani, S. Calderara, R. Cucchiara, Detection of static groups and crowds gathered in open spaces by texture classification, Pattern Recognit. Lett. 44 (2014) 39–48. https://doi.org/10.1016/j.patrec.2013.11.001.
11. P. Huang, A. Hilton, J. Starck, Shape Similarity for 3D Video Sequences of People, Int. J. Comput. Vis. 89 (2010) 362–381. https://doi.org/10.1007/s11263-010-0319-9.
12. A. Alahi, M. Bierlaire, P. Vandergheynst, Robust real-time pedestrians detection in urban environments with low-resolution cameras, Transp. Res. Part C Emerg. Technol. 39 (2014) 113–128. https://doi.org/10.1016/j.trc.2013.11.019.
13. R. Girshick, J. Donahue, T. Darrell, J. Malik, U.C. Berkeley, J. Malik, 1043.0690, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 1 (2014) 5000. https://doi.org/10.1109/CVPR.2014.81.

14. S. Ren, K. He, R. Girshick, X. Zhang, J. Sun, Object detection networks on convolutional feature maps, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 1476–1481. https://doi.org/10.1109/TPAMI.2016.2601099.
15. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2016. https://doi.org/10.1109/CVPR.2016.91.