Evolution in Electrical and Electronic Engineering Vol. 2 No. 2 (2021) 658-666

EEEE

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/eeee
e-ISSN : 2756-8458

Autonomous Human Recognition System for Various Challenging Postures in Various Postures in Various Environments

Lim Wei Ping[1], Abu Ubaidah Shamsudin[1]*, Sharifah Maryam Alhabshee[1]

[1]Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/eeee.2021.02.02.079
Received 04 July 2021; Accepted 04 August 2021; Available online 30 October 2021

Abstract: This paper introduces a deep learning method by using YOLOv3 as a medium to train human data samples and verify the presence of humans. The existing automated vehicles are designed and created for improving people's everyday life. However, they have the limitation in term of safety technologies such as slow human detection speed and inaccuracy and unstable sensing capability [1]. Therefore, there is a need to improvise the current safety technologies by reducing the span time for recognition and more accurate data detection. This enhances the reliability of the system which will able to lower the accidence rate. This project aims to develop a reliable system, YOLOv3 that can recognize the human being with different postures indoors and outdoors environments. Besides, to evaluate the performance of YOLOv3 toward human detection in the image, video, and real-time. Lastly, to determine the best training last weight in performing human recognition. Google Colaboratory is used as a core training environment. It offers powerful Nvidia k80 and T4 with GPU memory 12 GB and 16 GB respectively that provided rapid training speed but there have two limitations which are maximum execution time is only 12 hours and maximum idle time is only 90 minutes. In order to detect the amount of light intensity surrounds the environment, the Arduino Uno R3 with LDR photosensor is used. As a result, the overall accuracy of the YOLOv3 human recognition model has achieved above 88 percent in bright environment. However, the system does not run efficiently when it is dark. In the future, more data could be used in the training to increase the accuracy of the YOLOv3 model. Further possible attempts also can include the use of an IR camera or a Learning-to-see-in-the-dark model with YOLOv3 to solve the problem of recognizing low-light images.

Keywords: Human Recognition System, YOLOv3, Deep Learning

Corresponding author: ubaidah@uthm.edu.my

1. Introduction

A person recognition system is a biometric technology process of identifying or authenticating a person from a large pool of trained datasets based on python programming. It is eminently useful in plenty of applications such as implementation in autonomous driving vehicles or surveillance camera systems.

The person recognition system has recently been drawn to a new realm of application of autonomous driving. An autonomous car is a vehicle capable of sensing its surroundings including an incoming car and frontal obstacle and functioning without human intervention [1]. The main application of the sensing intelligence achieved by an in-vehicle camera is used for person recognition purposes. In reality, the appearance of human bodies is much more challenging to be recognized than other rigid objects, such as cars or motorbikes. There are non-rigid and high articulated human bodies. This implies a high degree of difficulty determines a high range of different poses and postures. Besides that, due to the variability of clothes, precise textures and color information cannot be taken advantage of in human detection. The problem of detection is also complicated by the absence of constraints on the image background [2]. Moreover, terrible weather and heavy traffic jam can all adversely affect the accuracy of the sensing capability.

In recent years, the growing need for mobility has led to major changes in the transportation system and infrastructure. Inefficiencies will definitely cause a tremendous loss of time, a decline in the quality of life, and a reduction in the degree of protection between humans and cars. Every year more than 1.3 million people died worldwide and tens of millions are seriously injured from road accidents [3]. In order to achieve level 5 that is fully autonomous according to Society of Automotive Engineer (SAE) which effectively control traffic condition and solve the problems like traffic crash accidents, a lot of research in terms of car safety and performance are required. Therefore, the YOLOv3 model is used as a practical solution to human detection applications in real-time [4]. With this kind of technology, the YOLOv3 can potentially reduce car road accidents per year and protect human life from accidents.

YOLOv3 is an algorithm for object detection using a convolution neural network. It can do deliver real-time detection and class classification well when compared to Fast-RCNN, Faster-RCNN, RFCN. Besides, it also makes predictions for a single network assessment as opposed to RCNN system that takes thousands of single images. This makes the YOLOv3 model extremely fast and accurate. In addition, it can process the image at around 40-90 FPS. This means streaming video can be processed within a few milliseconds [5]. Moreover, it can recognize multi-object in one image. In order to perform identification, it uses an entirely different approach. For the full picture, it applies a single convolution neural network for both classification and localization. For each region, this network splits the image into regions and predicts bounding boxes and probabilities. These bounding boxes are weighted by the predicted probabilities.

In this paper, to determine the best training weight by referring to a graph of current average loss against the iteration number, the higher the training iterations epochs between 0 and 20000, the lower the current average loss. Hence, the overall accuracy rate of human detection will be higher. Furthermore, the more quantity of trained data samples on different environments will produce a more accurate person detection system with higher variations.

2. Materials and Methods

2.1 Materials

A total of 2000 images were prepared by using a 'Samsung note 8' device with the rear camera specification 12-megapixel (f/1.7, 1.4-micron) and 12-megapixel (f/2.4, 1.0-micron). All the sample photos were related to the frontal and real sample images of the selected persons with various postures on 360 degrees and their formats are consisting of JPEG or .jpg with 640x311 pixels and 4032x1960 pixels. Logitech C920 HD Pro Webcam is an autofocus camera and it supports maximum resolution up to 1080 p/30 fps or 720p/ 30fps. The type of lens is glass and the diagonal field of view (dFoV) is 78°. It was used in the real-time detection method. In order to detect the amount of light intensity, the materials used were Arduino Uno R3, 10 cm Male to Male Breadboard Dupont Jumper Wires, breadboard, LDR Sensor Photosensor 5528, 1 KΩ resistor. The simulation scheme was shown in Figure 1 by using Tinkercad.

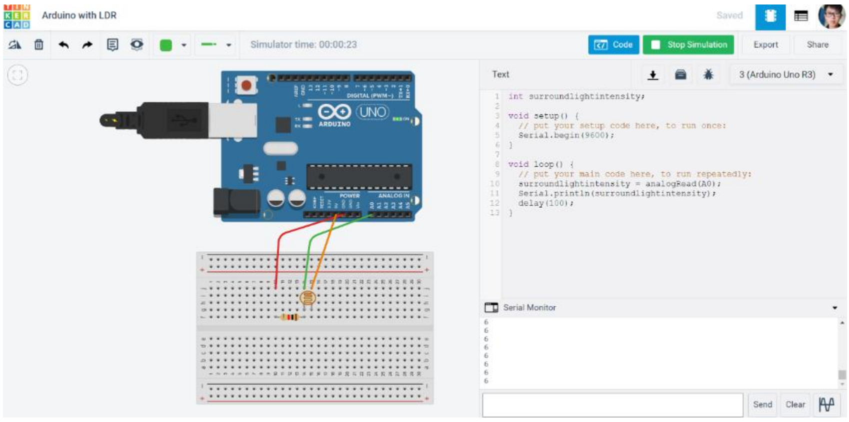

Figure 1: Circuit scheme

2.2 Methods

Google Colaboratory was a formidable online python software that can be utilized to solve the slow training speed faced by the CPU. The use of the YOLOv3 involving in training human sample images and verifying the human presence with free access to CUDA and GPU in a limited running time by using Google Colaboratory. The complete workflow was shown in Figure 2. There was a total of three methods utilized for human verification and prediction like image, video, real-time.

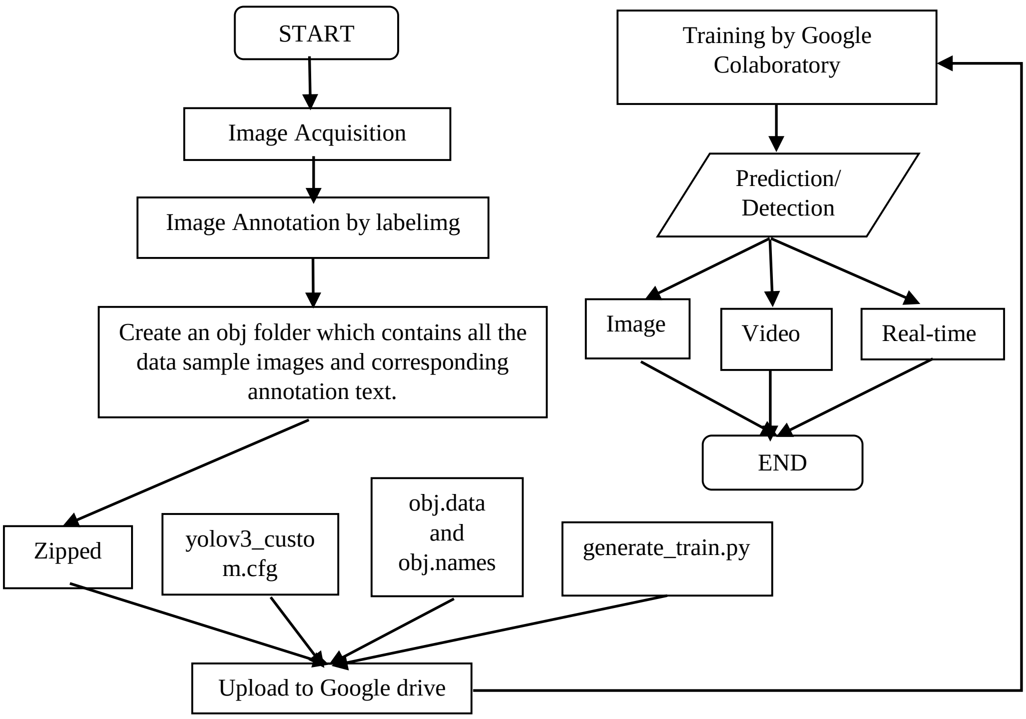

Figure 2: Flow chart of the project

3. Results and Discussion

3.1 Results

Based on Figure 3, the four images show the different standing positions of a little girl (person 1) who wearing a pink T-shirt and red trousers beside the road with the amount of light intensities between 741 and 744 as shown in Figure 4. The three images she was looking at in front of the camera, behind the camera, and right side show the excellent accuracy rate of person detection, which are 99 percent, 97 percent, and 99 percent respectively. However, the accuracy rate is slightly lower when the girl was looking at the left side, which is only 92 percent. As a result, the YOLOv3 person recognition system was working perfectly with high accuracy of over 91 percent on 4 main directions such as frontal, rear, left side, and right side.

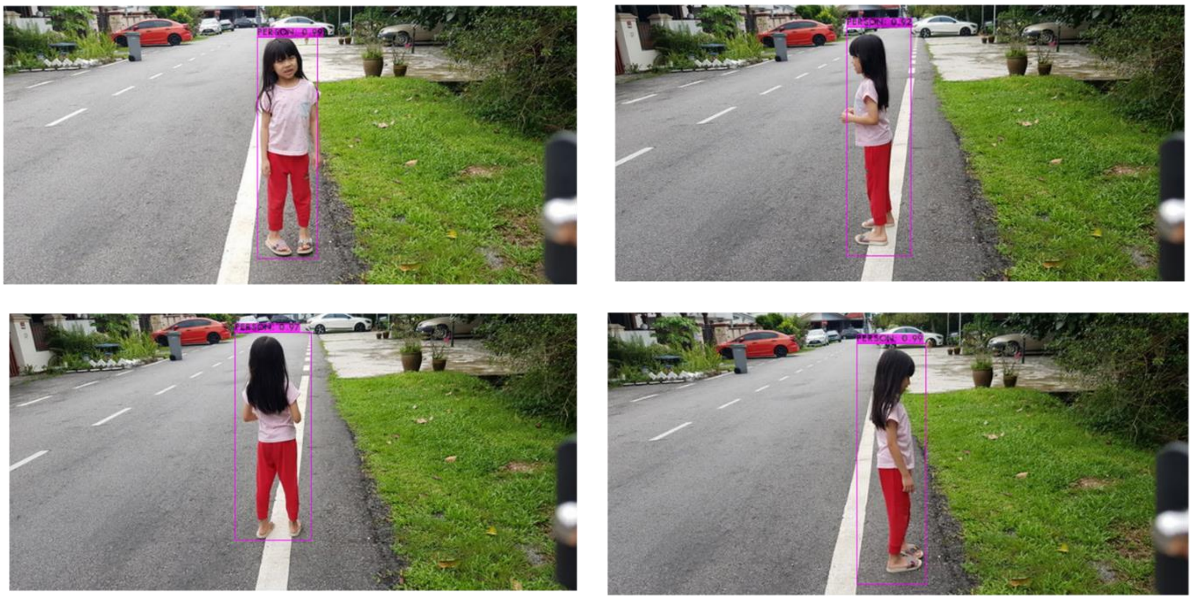

Figure 3: Person 1 was standing on the bright roadside

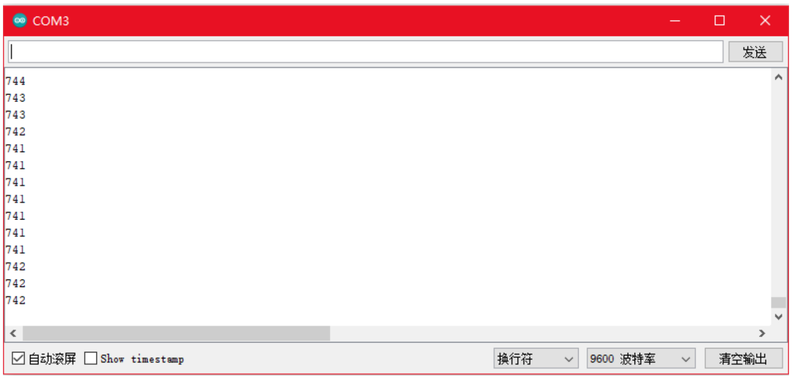

Figure 4: Values shown in serial monitor

Based on Figure 5, the four images show the different standing positions of a little girl (person 1) who wearing a pink T-shirt and red trousers beside the road with the amount of light intensities between 21 and 22 as shown in Figure 6. There are no created bounding boxes without accuracy rate shown in the four images she was looking at in front of the camera, behind the camera, left side, and right side. Therefore, no result of human recognition is indicated.

Figure 5: Person 1 was standing on the dark roadside

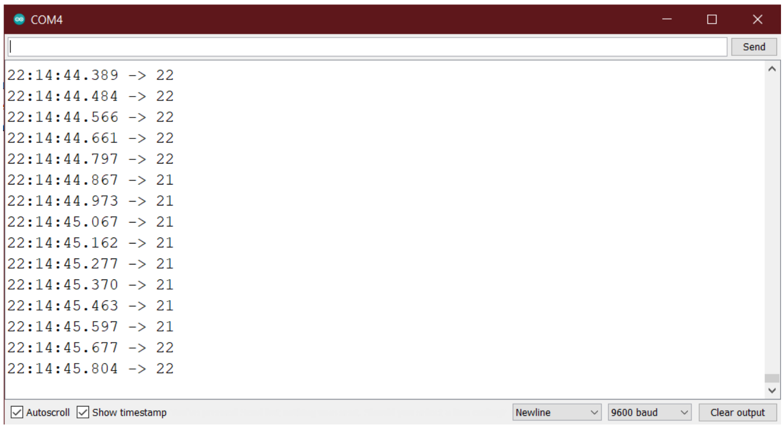

Figure 6: Values shown in serial monitor

Based on Figure 7, the author (person 3) wearing a yellow polo T-shirt, spectacles, and mask facing the camera in the room with a gray background and the amount of light intensities between 136 and 137 as shown in Figure 8. The created red bounding box has fully covered the person with an accuracy rate of 99 percent. As a result, the YOLOv3 person recognition system working perfectly toward the person who wears the mask in a bright indoor environment.

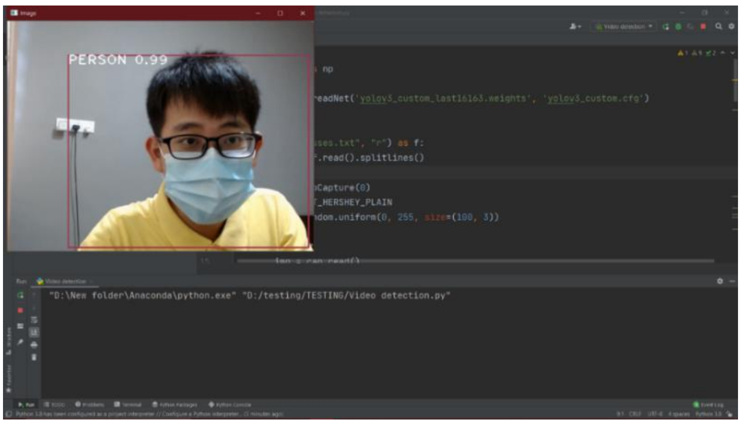

Figure 7: Person 3 was in the bright indoor environment

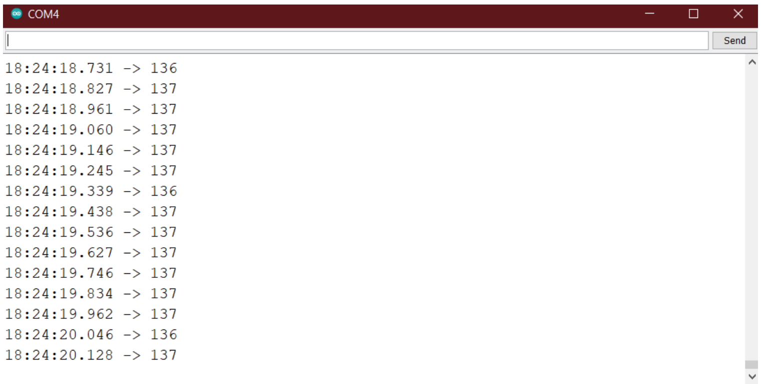

Figure 8: Values shown in serial monitor

Based on Figure 9, the author (person 3) wearing a yellow polo T-shirt, and spectacles facing the camera in the darkroom with a gray background and the amount of light intensities between 3 and 4 as shown in Figure 10. There is no created bounding box in Figure 9. Therefore, the system is unable to detect a person who wearing the mask in a dark environment.

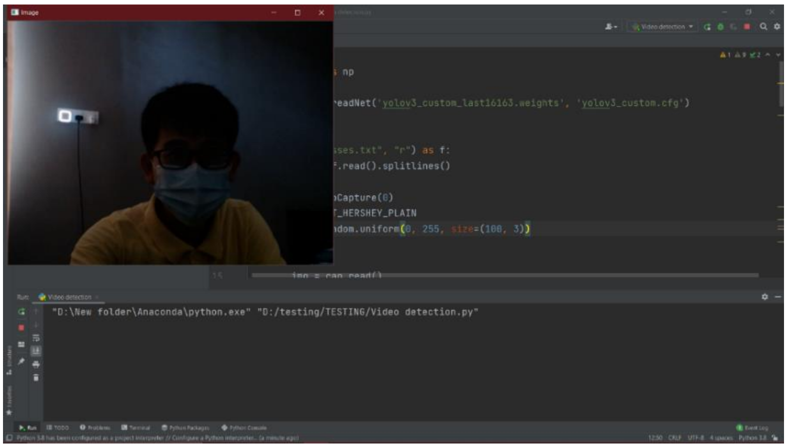

Figure 9: Person 3 was in the dark indoor environment

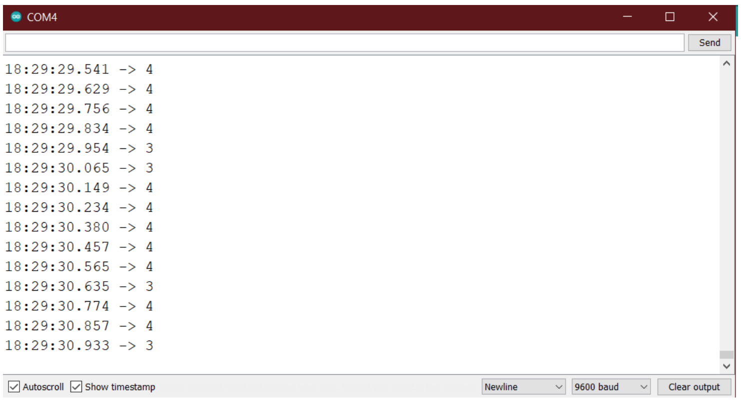

Figure 10: Values shown in serial monitor

3.2 Tables

Table 1 tabulates the summaries of the reading from Figure 3 to Figure 10. From the table, the YOLOv3 model is only functioning to detect person 1 and person 3 in the environment with the interval values of light intensity which are shown in the serial monitor are 741-744 and 136-137 respectively.

However, person 1 and 3 are not be recognized by the YOLOv3 model in the dark environments with the interval values of light intensity which are shown in the serial monitor are 21-22 and 3-4 respectively.

Table 1: Different Postures against Accuracy Rate of YOLOv3 Person Recognition

Postures	Values are shown in serial monitor	Accuracy rate of YOLOv3 person recognition (%)	Is YOLOv3 working?
Person 1 was standing on the bright roadside.	741-744	92-99	Yes
Person 1 was standing on the dark roadside.	21-22	None	None
Person 3 was in the bright indoor environment.	136-137	99	Yes
Person 3 was in the dark indoor environment.	3-4	None	None

3.3 Charts

Figure 11 shows the chart of accuracy rate against different direction of standing postures. There are a total of 5 individual persons who are located in different environments.

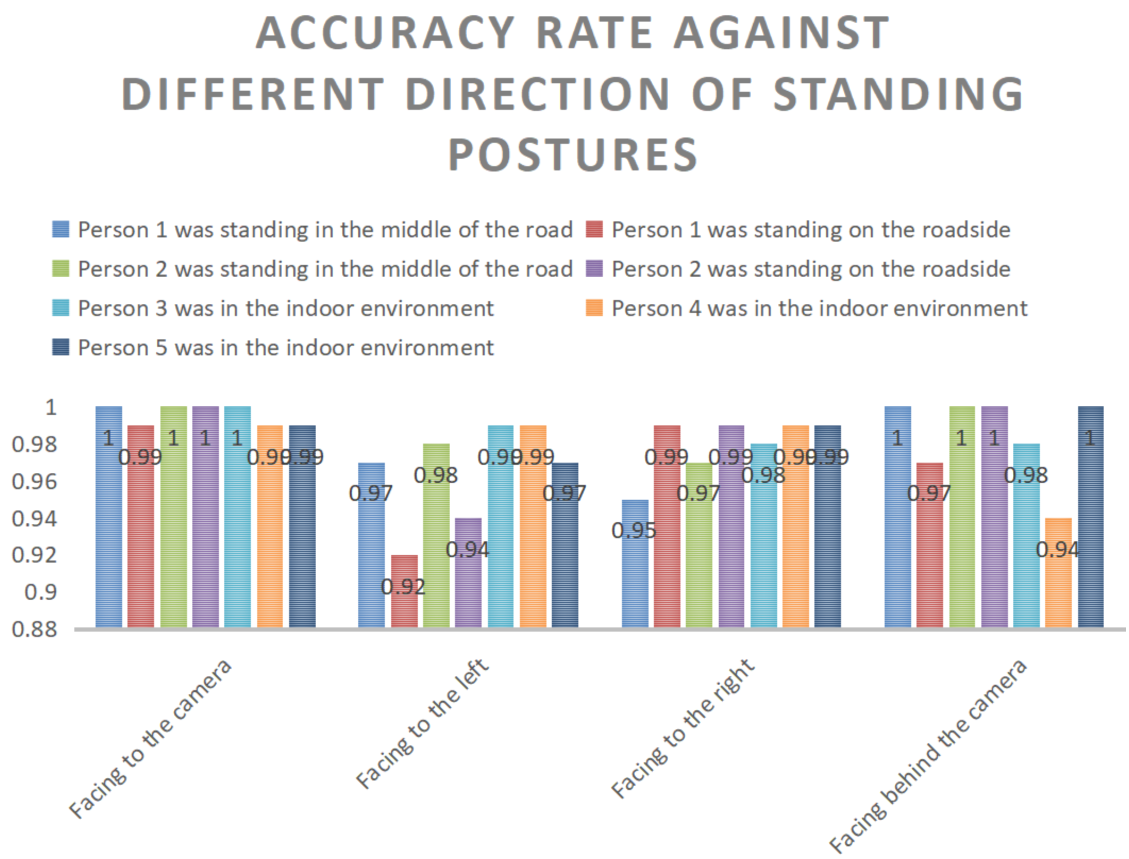

Figure 11: Chart of Accuracy Rate against Different Direction of Standing Postures

Figure 11 illustrates the chart of accuracy rate against the different directions of standing postures. Generally, there are a total of 5 individual persons in seven situations. The person who facing to the camera has the highest average accuracy rate of person recognition among four main different directions of standing postures, which is 0.996, followed by the person who facing behind the camera (0.984), facing to the right (0.98), and facing to the left (0.966). This is because the person facing the camera and facing behind the camera showed the complete body curve that has been well recognized by the YOLOv3 model. Despite that, the lowest accuracy rate of person recognition is the person who facing to the left, only 0.966 or 96.6 percent, but this value is in the range of acceptable tolerance. Due to facing to left and right, incomplete body figures lead to the YOLOv3 failing to achieve a 100 percent of person accuracy rate.

According to the person facing the camera, person 1 and person 2 standing in the middle of the road, person 2 who standing on the roadside, and person 3 who was in the indoor environment have the highest accuracy rate of person recognition, 100 percent. However, the lowest accuracy rate is 99 percent, specify states for the person 1 who standing on the roadside, person 4 and person 5 were in the indoor environment.

As stated by the person facing to the left, person 3 and person 4 were in the indoor environment have the outstanding accuracy rate of person recognition, 99 percent whereas the poorest accuracy rate is 0.92 or 92 percent, performed by person 1 who standing on the roadside. Overall, the accuracy rate of the person recognition system is over 91 percent.

According to the person facing to the right, person 1 and person 2 standing on the roadside, person 4, and person 5 were in the indoor environment have achieved 99 percent of the accuracy rate, which are the highest accuracy rate among seven situations. However, the worst accuracy rate is 95 percent, which states that person 1 who standing in the middle of the road.

According to the person facing behind the camera, person 1 and person 2 standing in the middle of the road, person 2 who standing on the roadside, and person 2 who was in the indoor environment have obtained the highest accuracy rate of person recognition, 100 percent. However, 0.94 or 94 percent is the poorest accuracy rate, performed by person 4 who was in the indoor environment.

As a result, the overall accuracy rate was over 91 percent against 4 different directions of standing postures based on seven situations. In conclusion, the YOLOv3 model is a reliable real-time person recognition system. Furthermore, it is appropriate for the identification and verification of human features with non-human features.

3.4 Discussions

Based on Table 1, person 1 and person 3 are successfully detected in the daylight with the amount of light intensities which are 741-744 and 136-137 respectively by the YOLOv3 model because the prepared human data samples are trained perfectly and the YOLOv3 model able to identify and distinguish between human and non-human regions that extract the pixels that contained human presence while discarding the non-human region accurately. As a result, the YOLOv3 person recognition system works perfectly in a bright environment.

However, all the proposed images which contained the person 1 and 3 are failed to be detected in the dark environments with the amount of light intensities which are 21-22 and 3-4 respectively by the YOLOv3 model. Therefore, the YOLOv3 model is unable to recognize the regions between human and non-human in the dark environments. This due to the statistical distribution of the pixel intensities shown in a dark image not being sufficient for the YOLOv3 model to recognize. In addition, the texture and silhouette information are extremely large losses. As a result, the YOLOv3 person recognition system failed to work in a dark environment.

4. Conclusion

In our study, we have introduced an approach, YOLOv3 to detect humans in different environments. Capturing multiple sample images related to the various human postures in different environments with different light intensities will definitely affect the accuracy rate of human detection. In this paper, the created 2000 images were trained perfectly using state of art YOLOv3 architecture. In order to determine the best weight training weight in performing human recognition, the average loss should below than 0.1. There are a total of three methods for human prediction like image detection, video detection, and real-time detection. All the three proposed methods are successfully conducted. As a result, the human verification is successfully proceeding in the daylight by the YOLOv3 model which can produce an overall accuracy rate of over 88 percent. However, the system does not run efficiently when it is dark.

Acknowledgement

The authors would like to thank the Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia for its support.

References

[1] Fafoutellis P., Mantouka E.G. (2019) Major Limitations and Concerns Regarding the Integration of Autonomous Vehicles in Urban Transportation Systems. In: Nathanail E., Karakikes I. (eds) Data Analytics: Paving the Way to Sustainable Urban Mobility. CSUM 2018. Advances in Intelligent Systems and Computing, vol 879. Springer, Cham. https://doi.org/10.1007/978-3-030-02305-8_89

[2] K. Balani, S. Deshpande, R. Nair and V. Rane, "Human detection for autonomous vehicles," 2015 IEEE International Transportation Electrification Conference (ITEC), 2015, pp. 1-5, doi: 10.1109/ITEC-India.2015.7386891

[3] International Traffic Safety Data and Analysis Group (IRTAD) [Online]. Available: https://www.itf-oecd.org/IRTAD/. [Accessed Jan. 5, 2021]

[4] C. Zhao and B. Chen, "Real-Time Pedestrian Detection Based on Improved YOLO Model," 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2019, pp. 25-28, doi: 10.1109/IHMSC.2019.10101

[5] S. Shinde, A. Kothari, V. Gupta, "YOLO based Human Action Recognition and Localization" 2018, Procedia Computer Science, Volume 133, pp. 831-838, doi: 10.1016/j.procs.2018.07.112