International Journal of Electrical and Computer Engineering (IJECE)
Vol. 11, No. 4, August 2021, pp. 3434~3442
ISSN: 2088-8708, DOI: 10.11591/ijece.v11i4.pp3434-3442

An assistive model of obstacle detection based on deep learning: YOLOv3 for visually impaired people

Nachirat Rachburee, Wattana Punlumjeak
Department of Computer Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand

Article Info

Article history:

Received Jul 31, 2020
Revised Dec 22, 2020
Accepted Jan 19, 2021

Keywords:

Assistive model
Deep learning
Obstacle detection
Visually impaired
YOLOv3

ABSTRACT

The World Health Organization (WHO) reported in 2019 that at least 2.2 billion people were visual-impairment or blindness. The main problem of living for visually impaired people have been facing difficulties in moving even indoor or outdoor situations. Therefore, their lives are not safe and harmful. In this paper, we proposed an assistive application model based on deep learning: YOLOv3 with a Darknet-53 base network for visually impaired people on a smartphone. The Pascal VOC2007 and Pascal VOC2012 were used for the training set and used Pascal VOC2007 test set for validation. The assistive model was installed on a smartphone with an eSpeak synthesizer which generates the audio output to the user. The experimental result showed a high speed and also high detection accuracy. The proposed application with the help of technology will be an effective way to assist visually impaired people to interact with the surrounding environment in their daily life.

Corresponding Author:

Wattana Punlumjeak
Department of Computer Engineering
Rajamangala University of Technology Thanyaburi
39 Moo 1, Klong 6, Khlong Luang Pathum Thani 12110 Thailand
Email: wattana.p@en.rmutt.ac.th

1. INTRODUCTION

The visual-impairment or blindness people in the world today is at least 2.2 billion reported by the World Health Organization (WHO) on 8 October 2019. The definition of classification of diseases 11(2018) classifies vision impairment by visual acuity worse into two groups: distance, and near presenting vision impairment. The visual acuity worse presenting between 6/12 and 6/60 of distance is defined as vision impairment, whereas visual acuity worse presenting than 3/60 of distance is defined as blindness [1]. One difficulty in the daily life of the visually impaired or blind people is living an invisible life indoor or outdoor environment. Although, guide dogs or white cane still the most popular tool used for obstacle detectors to navigate but the visually impaired people cannot know what things or the name of the obstacles are. A large number of people who visually impaired or blinded have been realized for the researcher to find a technology or a solution to assist them in their daily life.

Object detection, image processing and machine learning are some of the popular topics and have become rapidly growing fields. Object detection is a computer technology that deals with detecting instances of semantic objects of a certain class such as humans, cars, dogs, or traffic signs in digital images and videos. Machine learning is a subset of application of artificial intelligence (AI) that subject in the scientific study of algorithms and statistical models which provides systems the ability to automatically learn and improve itself from experience without being explicitly programmed. Machine learning can be categorized into supervised,

semi-supervised or unsupervised. Classification is one of the supervised learning algorithms in machine learning category. The classification used in object detection is to classify the object into a certain class that has learned. Deep learning is part of a broader family of machine learning methods based on artificial neural networks that use multiple layers to progressively extract higher-level features from the raw input. In image processing, lower layers of neural networks may identify edges, while higher layers may identify the concepts relevant to object in that image such as humans, dogs, cats, or cars.

In this research, an efficient algorithm in machine learning is proposed. The PASCAL VOC2007 and the PASCAL VOC2012 data set is used to train the machine. The prototype of the system on the screen of the smartphone is developed to find the best assistive model for the visually impaired. The paper is organized as follows: after the introduction, literature review, and related works are presented in section 2, follow by the research method and proposed experiment in section 3, the result and discussion in section 4. Finally, in section 5 we provide conclusive remarks and our future work.

2. LITERATURE REVIEW AND RELATED WORK

In the past, one of the main tasks of machine learning was to classify things by creating classifiers that could classify whether the object in the image was a person or an animal (e.g. dog, cat) or any other object. In this era, most researchers had focused on finding and creating effective classifiers, from simple linear classifiers that combine features from linear combination until support vector machine (SVM) classifier used the kernel function to transform these features into mathematical kernel space. When research on the classification of things became saturated, researchers began to move on to more difficult and challenging problems, which were "detecting and classifying" what was an interesting object in the image. The paper [2] that had known as the pioneer of object recognition which used a convolutional neural network (CNN) with gradient-based learning to handwritten character recognition. The breakthrough in computer vision, the face detector system by Viola and Jones [3]. The main idea of this research has created a cascade classifier and combined it with AdaBoosting learning algorithm instead of using a one classifier. At that time, Viola and Jones research paper was considered state-of-the-art for object detection. Due to the limitations of the processors that are not fast enough, therefore CNN classifier was not received much attention. In the famous annual computer vision competition, imagenet large-scale visual recognition challenge in 2012: ILSVRC, Alex and his team presented a deep convolutional neural network architecture called AlexNet [4]. AlexNet showed the best performance in the competition. So, CNN is becoming more and more popular, and many CNN models and architecture had been improved from the previous AlexNet structure. After that, more research that adapted and fine-tuning from the previous architecture had been proposed e.g. VGG [5], GoogLeNet which its codenamed was Inception [6], Microsoft ResNet [7] and more.

The advent of region-based CNNs (R-CNN) which the authors purposed to solved object detection problems [8]. The R-CNN processes were split into two steps: region proposal step and the classification step. Region proposal step used selection search which proposed region of interest (ROI) and generated different 2000 regions then extract feature by CNN named AlexNet. Then, classified each region using linear SVMs in the classification step. The same author from R-CNN [9] improved the Fast R-CNN from the R-CNN in the problem of speed by used ROI pooling [10] through a ConvNet to extracted the feature and used a fully connected layer instead of SVM to classification or recognition. In 2016, Faster R-CNN was presented [11]. Faster R-CNN was improved from Fast R-CNN by replacing region proposals with region proposal network (RPN) after the last convolutional layer. Faster R-CNN had two outputs: a bounding-box offset for each candidate object and a class label of ROI. Mask R-CNN [12] extended Faster R-CNN by adding a mask branch to predict a segmentation mask on each ROI while the existing branch for classification and bounding box regression.

All the above of object detectors are the state of the art which based on a two-stage framework: The first stage generates region proposal to localize the object in the image, the second stage classifies the object. Despite the success of two-stage detectors, one-stage detectors are also applied. Single shot multibox detector (SSD) was presented in 2016 [13] which based on standard network architecture: VGG-16. The SSD produced a bounding boxes in different aspect ratios and score for each category of presence object.

Another famous one-stage detector is you only look once (YOLO) [14]. YOLO, a unified architecture which straightforward, simple and extremely fast in which the network architecture was inspired by GoogleNet, then is called Darknet. The network architecture has 24 convolutional layers working as feature extractors and 2 fully connected layers for the predictions in which the framework is trained on the ImageNet-1000 dataset. The YOLO architecture [14] showed in Figure 1. The YOLO used algorithms which based on regression where the process of detection, localization, and classification the object for the input image will take place in a single pass. The YOLO detection system process start by resizes the input image into 448 X 448 and then divided into S X S grid cell, then fed into the single convolutional network. Each

grid cell predicts B bounding box and confidence score which output of each bounding box consists of 5 prediction values: offset values (x, y, w, and h) where x and y are the coordinates of the object in the input image, w and h are the width and height of the object respectively for each of the bounding box, while the last prediction value is confidence score or class probabilities that given in terms of an IOU (intersection over union), which should have the object exist in the bounding box. The bounding box which has a high confidence score above the threshold value is selected and then used to locate the object within the image.

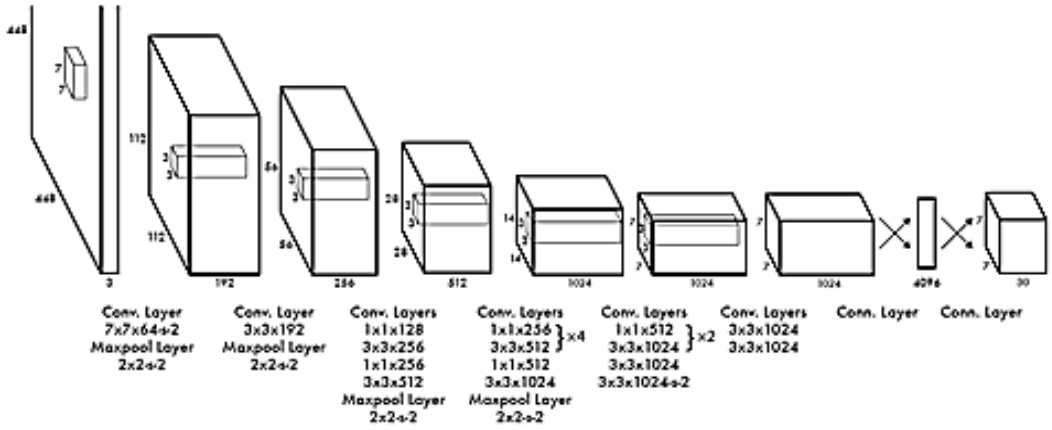


Figure 1. The YOLO architecture [14]

YOLO or named YOLOv1 has a limitation, that it could not find small objects in the image if they appeared as a cluster or group and difficult found in a generalization of objects if the image is different dimensions from the trained image. In December 2016, the second version of the YOLO has named YOLOv2 or YOLO9000 real-time framework for detection categories of the object more than 9000 categories have been published [15]. Mainly new thing was to introduce the anchor boxes which are designed for a given dataset by using k-means clustering to responsible to predict the bounding box. The architecture of YOLOv2 used the Darknet-19 architecture with 19 convolutional layers and 5 max-pooling layers and then a softmax layer for classification of the objects. YOLOv3: An incremental improvement has been published in April 2018 [16]. YOLOv3 has been improved from the previous version with a high accuracy of classifying the objects. To predict the score for the objects for each bounding box, YOLOv3 uses logistic regression and also used independent logistic classifiers for each box to predict the classes of the bounding box which may contain an object instead of softmax. YOLOv3 uses the Darknet-53 network [16] for feature extractor which has 53 convolutional layers which showed in Figure 2.

One of the challenging problems in the field of object detection and machine learning is assisting people who visually impaired. Many researchers proposed their work which aims to help visually impaired in daily life. Patient monitoring framework in telemedicine system was presented in different scenarios [17]. A greedy algorithm was developed to design cascade for applied real-time text detector for the visually impaired [18]. Arakeri *et al.* proposed a raspberry pi with NoIR camera to captures the readable material around the visually impaired and used a speech synthesis to generate sound in regional language [19]. Fink and Humayun [20], and more researcher [21, 22] presented the invention for the visually impaired. A digital camera mounted on the person's eye or head is used to take snapshots of an image on demand and provided to an image processing algorithm. Edge detection techniques are used to identify the object in the image and classified the known object by artificial neural networks that have been trained. The invention could determine the size, distance from another object and announced the computer-based voice synthesizer to describe the descriptive sentence for the blind. Tapu *et al.* [23] introduced a real-time obstacle detection and classification to assist visually impaired people in indoor and outdoor environments by handling a smartphone device with the help of a chest-mounted harness. The authors proposed the step of object detection by extract an image grid and using the multiscale Lucas Kanade algorithm to track the interested point. Then, estimate the motion classes with an agglomerative clustering technique to classify into clusters and refined them with the K-NN algorithm. The step of moving object classification incorporates the HOG descriptor into the bag of visual words (BoVW) retrieval framework. The results of the experiment in different environments achieved high accuracy rates and efficient to a blind person. The AI assistant through

an android mobile application for visually impaired was proposed by [24]. The group of researchers had focused their idea on image recognition, currency recognition, text recognition, chatbot and voice assistant by using voice command via interaction with the environment. Their application was developed on the google cloud platform which used cloud API libraries. Convolution neural networks for object detection systems for visually impaired or blind people were continually improved. Convolutional neural network and Haar cascade classifiers were compared by Shah *et al.* [25] to conduct a suitable algorithm to assist the visually impaired or blind person for a real-time scenario. The dataset used in the training process of CNN was COCO 2017. The experiment was conducted that CNN is more high accuracy to detect multiple objects than Haar cascade for real-time applications. Bianco *et al.* [26] presented a category-based image quality assessment named DeepBIQ that used to extract features from a CNN fine-tuned for image quality task. The group of researchers [27] used the single shot multibox detector (SSD) in their system to identified objects after a webcam captured a real-time scene and extracted. Raspberry Pi 3 was used as a prototype, and the audio-based detector was generated the detection information as sound to the connected headphone. The model was worked well although offline condition compared with fast R-CNN. An experiment of object detection and localization in the street environment model was proposed by [28]. The pre-trained model based on faster R-CNN was used with the COCO dataset while transfer learning was fine-tuned. The self-made dataset was acquired from the internet in different kinds include the object in the urban street. They concluded that faster R-CNN on a self-made dataset improved average accuracy and the fine-tuned network was effective.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure 2. Darknet-53 architecture [16]

In a single-stage object detection network, there are much interesting research and its applications: Detecting obstacles with the light field camera was proposed by [29]. YOLO, deep learning was used to classify objects in the image under the indoor environment into categories. The group of researchers had presented that the obstacle was accurately classified and was getting a high accuracy in size and their position. Human action recognition with YOLO object detection in the frame of the video was used by [30] with the LIRIS dataset. The proposed were presented effectively in terms of action label, confidence score, and localized action. YOLO and multi-task cascaded neural networking (MTCNN) structure were proposed by Rahman *et al.* [31] to implement an assistive model for visually impaired people on Raspberry Pi for object detection and facial recognition. The personalized dataset used in this proposed model consists of 3 positions: left, right, and center of the new face image and the name of that image act as a label of the data. The model achieved 6-7 FPS with a 63-80% accuracy rate for object detection while the accuracy rate of facial recognition achieved an 80-100%. To improve the accuracy of face detection, Garg *et al.* [32] used the YOLO framework on face detection dataset and benchmark (FDDB) dataset to train their proposed work. A model compared the execution time and the performance on two different machines. After fine-tuning all

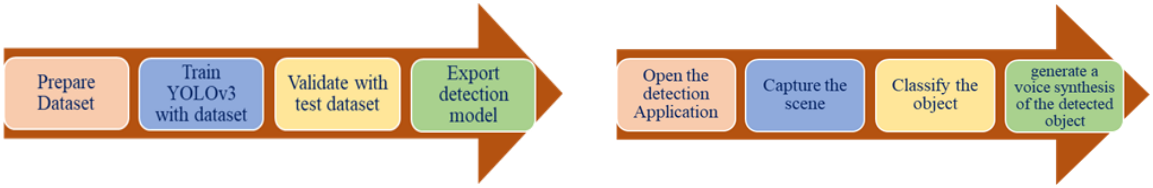
parameters and the suitable values of the proposed model, the accuracy was compared with Harr cascade and R-CNN algorithms. To improve the YOLOv2 structure for pedestrian detection, Lan *et al.* [33] proposed a YOLO-R structure which added three layers of pedestrian feature in front of the deep YOLOv2 network and also changed the passthrough layers to increase the ability of the network. The results of the proposed model had shown the high accuracy of pedestrian detection while the false rate and the miss rate was reduced, compared with the YOLOv2 network on the INRIA dataset. A real-time face detection model was proposed by Wang, and Jiachun [34] on the WIDER FACE dataset with the YOLO algorithm. The 20 various images size were selected from three datasets: Celeb Faces, FDDB, and WIDER FACE dataset to use for a testing phase in the proposed model. They had shown the high-speed rate of detection, reduced error rate and strong robustness of the YOLOv3 compared with traditional algorithms. To handle a real-time object detection for non-GPU computers, the group of researchers [35] was proposed the YOLO-LITE model which the best trial experiment was run on the COCO dataset. They had shown that YOLO-LITE was a faster, smaller, and more efficient model to detect the object compared with the state of the art model in a variety of devices. A pre-trained ssdlite_mobilenet_v2_coco_2018_05_09 model as a feature extractor was used for obstacle detection in sidewalks design and alert system for visually impaired people [36]. Raspberry Pi and Pi camera were used as hardware prototype and eSpeak was used as a speech synthesizer to represent the direction of the object by headphones. Whereas, the vibration sensor was activated when the object was detected and recognized. The application for visually impaired people for an android platform was proposed by [37]. The researchers had claimed that their application would be a virtual third eye for visually impaired. A suitable chest strap was designed to hold the phone [38]. Tiny-YOLO was used in their experiment and integrate with ARKit configuration to detect an object with augmented reality for iOS applications. The training period used Tiny YOLO with a Darknet base network on Turicreate engine and INRIA annotation for the Graz-02 dataset, while they tested the model by using a 100-random set of cars and bikes in different background, shape, and size. In their conclusion, the model could detect an object and overlayed 3D graphics at the location of an object in an effective way. [39] YOLOv3 algorithm was used to detect the five classes of a real-time object of traffic participants or road signalization in advanced driver assistance systems (ADAS). The proposed system evaluated on NVidia GeForce GTX 1060GPU by using the weights on the COCO pre-trained model and trained on the Berkley deep drive dataset. The effectiveness of the proposed model had shown in the variety of driving conditions.

3. RESEARCH METHOD

The proposed ideas of our work divided into two parts: Train the detection model and then developed the application. Our proposed method is shown as in a Figure 3(a) and (b). Our experiment model was done on Google Colaboratory: Colab 12 GB-RAM GPU. First, we prepared the data used by download the PASCAL VOC2007 and the PASCAL VOC2012 from The PASCAL Visual Object Classes Homepage. The two datasets have twenty object classes that have been selected are:

- Person: Person
- Animal: Bird, cat, cow, dog, horse, sheep
- Vehicle: Airplane, bicycle, boat, bus, car, motorbike, train
- Indoor: Bottle, chair, dining table, potted plant, sofa, television/monitor

The PASCAL VOC2007 have train/validation: 9,963 images containing 24,640 annotated objects and PASCAL VOC2012 have train/ validation: 11,530 images containing 27,450 ROI annotated objects and 6,929 segmentations. The two datasets have been split into 50% for training/validation and 50% for testing. In our experiment, we combined two dataset together as a dataset for training set and used PASCAL VOC2007 test set for data testing set. Second, train YOLOv3 by using Darknet on Colab with the dataset we prepared from the process above and then validate with testing dataset. The YOLOv3 structure shown as Figure 4.



(a) (b)

Figure 3. Proposed method (a) Train the detection model, and (b) The prototype of detection application

```
layer     filters    size              input                output
    0 conv     32  3 x 3 / 1   416 x 416 x   3   ->   416 x 416 x  32 0.299 BF
    1 conv     64  3 x 3 / 2   416 x 416 x  32   ->   208 x 208 x  64 1.595 BF
    2 conv     32  1 x 1 / 1   208 x 208 x  64   ->   208 x 208 x  32 0.177 BF
    3 conv     64  3 x 3 / 1   208 x 208 x  32   ->   208 x 208 x  64 1.595 BF
    4 Shortcut Layer: 1
    5 conv    128  3 x 3 / 2   208 x 208 x  64   ->   104 x 104 x 128 1.595 BF
    6 conv     64  1 x 1 / 1   104 x 104 x 128   ->   104 x 104 x  64 0.177 BF
    7 conv    128  3 x 3 / 1   104 x 104 x  64   ->   104 x 104 x 128 1.595 BF
    8 Shortcut Layer: 5
    9 conv     64  1 x 1 / 1   104 x 104 x 128   ->   104 x 104 x  64 0.177 BF
   10 conv    128  3 x 3 / 1   104 x 104 x  64   ->   104 x 104 x 128 1.595 BF

                                    ...

   99 conv    128  1 x 1 / 1    52 x  52 x 384   ->    52 x  52 x 128 0.266 BF
  100 conv    256  3 x 3 / 1    52 x  52 x 128   ->    52 x  52 x 256 1.595 BF
  101 conv    128  1 x 1 / 1    52 x  52 x 256   ->    52 x  52 x 128 0.177 BF
  102 conv    256  3 x 3 / 1    52 x  52 x 128   ->    52 x  52 x 256 1.595 BF
  103 conv    128  1 x 1 / 1    52 x  52 x 256   ->    52 x  52 x 128 0.177 BF
  104 conv    256  3 x 3 / 1    52 x  52 x 128   ->    52 x  52 x 256 1.595 BF
  105 conv     75  1 x 1 / 1    52 x  52 x 256   ->    52 x  52 x  75 0.104 BF
  106 yolo
Total BFLOPS 65.428
```

Figure 4. The YOLOv3 structure

4. RESULTS AND DISCUSSION

From our proposed methodology above, after training the YOLOv3 with the dataset prepared above, a detection model had been as an output. So, we export the model from Google COLAB to our local drive. Then, we developed a prototype of an application on a smartphone which installed the obstacle detection model. We designed the user interface (UI) in a simple way and used eSpeak as a function to generate the audio output. The eSpeak is open-source software that synthesizes text to speech in English and other languages. The example of the indoor and outdoor images we captured in the real-time view or in the real situation which mimics as an input to the obstacle detection application shown in Figure 5(a) and 5(b).

(a) (b)

Figure 5. The example of the image in the real situation

The captured image was then forwarded to the obstacle detection model to classify the object. After that, the system showed the class of the output it detected and generated a voice synthesis of the detected object to notify or assist the visually impaired or blind people to identify the object. The output of the system shown in Figure 6(a) and 6(b).

The prototype of the obstacle detection system on the screen of the smartphone shown in Figure 7. The experimental in real situation results based on YOLOv3 showed a high speed and also high detection accuracy in the real-time view. The proposed model of the obstacle detection system on a smartphone will be assisting visually impaired people about the surrounding environment.

(a) (b)

Figure 6. The output of the obstacle detection system

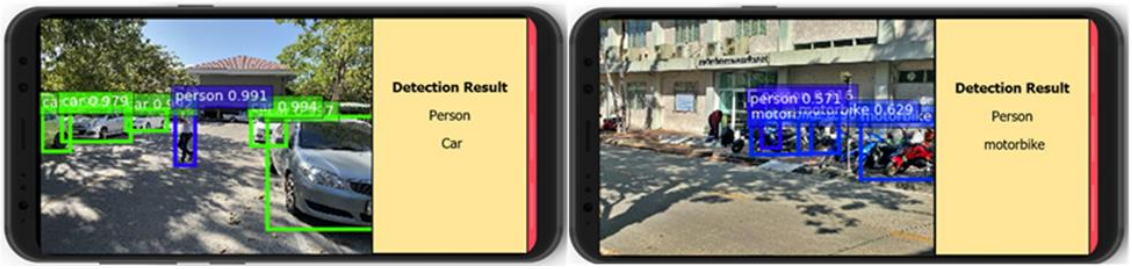


Figure 7. The output of the system prototype on the smartphone

5. CONCLUSION

In this paper, we have introduced a novel framework of an application on a smartphone for obstacle detection and classification which based on deep learning: YOLOv3. Our proposed application on a smartphone works in real-time to capture an image and forward it to the obstacle detection system. The experiment results prove the effectiveness of the system which not only able to show the output of the obstacle detected and can classify the name in the class of the obstacle but also can generate the audio output in their own languages. An application of obstacle detection and classification for visually impaired people will be a benefit in safety and comfort for a better quality of living in daily life. In our future works, we will study the distance between visually impaired people and obstacles. We plan to study a similar triangle, Euclidean distance, and other theories and then integrate it to improve the overall application.

REFERENCES

[1] World Health Organization, “Blindness and vision impairment,” [Online], Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. of the IEEE*, vol. 86, no. 11, 1998, pp. 2278-2324, doi: 10.1109/5.726791.

[3] Viola, P., and Jones, M. J., "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137-154, 2004.

[4] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84-90, May 2017, doi: 10.1145/3065386.

[5] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[6] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D *et al*., "Going deeper with convolutions, " in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[8] Girshick, R., Donahue, J., Darrell, T., and Malik, J., "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, " in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.
[9] Girshick, Ross, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440-1448.
[10] He, K., Zhang, X., Ren, S., and Sun, J., "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 37, no. 9, pp. 1904-1916, Sep. 2015, doi: 10.1109/TPAMI.2015.2389824.
[11] Ren, S., He, K., Girshick, R., and Sun, J., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.
[12] He, K., Gkioxari, G., Dollár, P., and Girshick, R., "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961-2969.
[13] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., and Berg, A. C., "Ssd: Single shot multibox detector," in *European conference on computer vision*, Springer, Cham, 2016, pp. 21-37.
[14] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
[15] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 6517-6525, doi: 10.1109/CVPR.2017.690.
[16] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv preprint arXiv:1804.02767*, 2018.
[17] Chakraborty, C., Gupta, B., and Ghosh, S. K., "A review on telemedicine-based WBAN framework for patient monitoring," in *2013 Telemedicine and e-Health*, vol. 19 no. 8, pp. 619-626, 2013.
[18] Xiangrong Chen and A. L. Yuille, "A Time-Efficient Cascade for Real-Time Object Detection: With applications for the visually impaired," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*, San Diego, CA, USA, vol. 3, 2005, pp. 28-28, doi: 10.1109/CVPR.2005.399.
[19] Arakeri, M. P., Keerthana, N. S., Madhura, M., Sankar, A., Munnavar, T., "Assistive Technology for the Visually Impaired Using Computer Vision," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Bangalore, pp. 1725-1730, 2018, doi: 10.1109/ICACCI.2018.8554625.
[20] Fink, W., and Humayun, M., "Digital object recognition audio-assistant for the visually impaired," U.S. Patent Application 11/030,678, Sep. 2005.
[21] Guevarra, E. C., Camama, M. I. R., and Cruzado, G. V., "Development of Guiding Cane with Voice Notification for Visually Impaired individuals," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 1, pp. 104-112, 2018.
[22] Jaejoon Kim, "Application on character recognition system on road sign for visually impaired: case study approach and future," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 1, pp. 778-785, 2020.
[23] Tapu, R., Mocanu, B., Bursuc, A., and Zaharia, T., "A Smartphone-Based Obstacle Detection and Classification System for Assisting Visually Impaired People," in *2013 IEEE International Conference on Computer Vision Workshops*, Sydney, Australia, 2013, pp. 444-451, doi: 10.1109/ICCVW.2013.65.
[24] Felix, S. M., Kumar, S., and Veeramuthu, A., "A Smart Personal AI Assistant for Visually Impaired People," in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, 2018, pp. 1245-1250, doi: 10.1109/ICOEI.2018.8553750.
[25] Shah, S., Bandariya, J., Jain, G., Ghevariya, M., and Dastoor, S., "CNN based Auto-Assistance System as a Boon for Directing Visually Impaired Person," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 2019, pp. 235-240, doi: 10.1109/ICOEI.2019.8862699.
[26] Bianco, S., Celona, L., Napoletano, P., and Schettini, R., "On the use of deep learning for blind image quality assessment," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 355-362, Feb. 2018, doi: 10.1007/s11760-017-1166-8.
[27] Wong, Y. C., Lai, J. A., Ranjit, S. S. S., Syafeeza, A. R., and Hamid, N. A., "Convolutional Neural Network for Object Detection System for Blind People," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 11, no. 2, p. 6, 2019.
[28] Cai, W., Li, J., Xie, Z., Zhao, T., and Kang, L. U., "Street Object Detection Based on Faster R-CNN," In 2018 37th Chinese Control Conference (CCC), 2018, pp. 9500-9503.
[29] Zhang, R., Yang, Y., Wang, W., Zeng, L., Chen, J., and McGrath, S., "An Algorithm for Obstacle Detection based on YOLO and Light Filed Camera," in *2018 12th International Conference on Sensing Technology (ICST)*, Limerick, pp. 223-226, 2018, doi: 10.1109/ICSensT.2018.8603600.
[30] Shinde, S., Kothari, A., and Gupta, V., "YOLO based Human Action Recognition and Localization," *Procedia Computer Science*, vol. 133, pp. 831-838, 2018, doi: 10.1016/j.procs.2018.07.112.
[31] Rahman, F., Ritun, I. J., Farhin, N., and Uddin, J., "An assistive model for visually impaired people using YOLO and MTCNN," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy - ICCSP '19,* Kuala Lumpur, Malaysia, pp. 225-230, 2019, doi: 10.1145/3309074.3309114.
[32] Garg, D., Goel, P., Pandya, S., Ganatra, A., and Kotecha, K., "A Deep Learning Approach for Face Detection using YOLO," in *2018 IEEE Punecon*, Pune, India, pp. 1-4, 2018, doi: 10.1109/PUNECON.2018.8745376.

[33] Lan, W., Dang, J., Wang, Y., and Wang, S., "Pedestrian Detection Based on YOLO Network Model," in *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, Changchun, 2018, pp. 1547–1551, doi: 10.1109/ICMA.2018.8484698.
[34] W. Yang and Z. Jiachun, "Real-time face detection based on YOLO," in *2018 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII)*, Jeju, 2018, pp. 221-224, doi: 10.1109/ICKII.2018.8569109.
[35] Huang, R., Pedoeem, J., and Chen, C., "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018, pp. 2503-2510, doi: 10.1109/BigData.2018.8621865.
[36] Pehlivan, S., Unay, M., and Akan, A., "Designing an Obstacle Detection and Alerting System for Visually Impaired People on Sidewalks," in *2019 Medical Technologies Congress (TIPTEKNO)*, Izmir, Turkey, 2019, pp. 1-4, doi: 10.1109/TIPTEKNO.2019.8895181.
[37] S. Tosun and E. Karaarslan, "Real-Time Object Detection Application for Visually Impaired People: Third Eye," in *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, Malatya, Turkey, 2018, pp. 1-6, doi: 10.1109/IDAP.2018.8620773.
[38] S. Mahurkar, "Integrating YOLO Object Detection with Augmented Reality for iOS Apps," in *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, New York City, NY, USA, 2018, pp. 585-589, doi: 10.1109/UEMCON.2018.8796579.
[39] Ćorović, A., Ilić, V., Đurić, S., Marijan, M., and Pavković, B., "The Real-Time Detection of Traffic Participants Using YOLO Algorithm," in *2018 26th Telecommunications Forum (TELFOR)*, Belgrade, pp. 1-4, 2018, doi: 10.1109/TELFOR.2018.8611986.

BIOGRAPHIES OF AUTHORS

Nachirat Rachburee is a lecturer at Department of Computer Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand. His research interests include Data Mining, Big data analytics, Deep Learning, Neural Networks and Predictive analytics.

Wattna Punlumjeak is a lecturer at Department of Computer Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand. Her research interests include Data Mining, Big data analytics, Deep Learning, Neural Networks and Predictive analytics.