A Review of Object Detection Based on Convolutional Neural Networks and Deep Learning

MingYuan Wang[1] and Watis Leelapatra[2]

[1,2]Department of Computer Engineering, Khon Kaen University, Thailand
E-mail: wangmingyuan@kkumail.com, watis@kku.ac.th

Received: January 26, 2022 / Revised: June 6, 2022 / Accepted: June 17, 2022

***Abstract*—Object detection, as one of the three main tasks of computer vision, is of great importance for the development of artificial intelligence in the future. The rapid advancement of convolutional neural networks (CNNs) and deep learning have provided a broader arena for object detection. From traditional methods to state-of-the-art algorithms, numerous innovative technologies and methods have been proposed. This paper reviews the one-stage and two-stage object detection algorithms and compares their advantages and shortcomings from various aspects. Some applications in real life, such as self-driving, weeding robots, and face recognition are illustrated in this paper. Finally, current issues and future studies direction prospect.**

***Index Terms*—Object Detection, Deep Learning, Computer Vision**

I. Introduction

Object detection technology is a fundamental part of computer vision research and essential for the development of artificial intelligence. The core task is to process the image data, using various feature extraction and classification algorithms to obtain useful semantic and positional information, and ultimately to figure out the useful features in the image quickly, accurately, and reliably. But, the process of capturing images is inevitably affected by angle, light, equipment, and shading, which can cause distortion and noise results. To overcome and solve such problems, a great number of researchers have explored and proposed different algorithms from various aspects.

The basic process of a traditional algorithm consists mainly of region proposal selector, feature extraction, and classifier. Analysis of images that are stored in a computer, since images can be viewed mathematically as a data set with one or more matrices. The selector uses sliding windows of different sizes to slide sequentially from left to right and top to bottom on the image to find the region proposal. Then, feature extraction is performed on the region proposal. In traditional algorithms, Scale Invariant Feature Transform (SIFT) [1] and Histogram of Oriented Gradient (HOG) [2] are commonly used for feature extraction. The role of classifiers is to classify and identify the extracted features to achieve the purpose of object detection, and typical classifiers commonly used are Support Vector Machines (SVM) [3] and Deformable Part Model (DPM) [4].

Most state-of-the-art algorithms are based on convolutional neural networks (CNNs) and deep learning. The rapid development of CNNs and the Graphics Processing Units (GPU) have opened up new research directions for object detection. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is one of the most popular and prestigious academic competitions in the field of machine vision in recent years. AlexNet [5] won the 2012 ILSVRC champion and was a milestone in computer vision for CNNs. Various structures were subsequently introduced, such as VGG [6], GoogLeNet [7], SPPnet [8], ResNet [9]. For different network structures, object detection algorithms can be divided into two categories, one is two-stage algorithms based on region proposals, such as R-CNN [10], Fast R-CNN [11], and Faster R-CNN [12]. The other is a regression-based end-to-end one-stage algorithm. The YOLO [13] and SSD [14] algorithms have performed very well. Both have their advantages and drawbacks. The one-stage algorithm is faster than the two-stage and has a higher Frames Per Second (FPS) in video detection, which can be applied to real-life situations, such as a self-driving car. The two-stage algorithm, on the other hand, is slower but has higher accuracy. So, it can be applied to situations where high accuracy is required.

II. Assessment Measures and Datasets

Popular measures used to evaluate object detection algorithms are Precision (P), Recall (R), F1 score, Average Precision (AP), Mean Average Precision (mAP), FPS, and the amount of parameters in the network. To better explain the above measures, we need to introduce the concept of Intersection over Union (IoU). IoU is a ratio of the area where the prediction box intersects the real box and the area where the two are combined, it reflects the deviation of the prediction result from the truth, the value is between 0 and 1, the closer to 1 the better the result. Thus, at a certain threshold level we can calculate the values of TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) according to the confusion matrix. The P shows the accuracy of detection, while the R shows the proportion of detected objects to the total number of true objects. The F1 is the summed average of the P and R. The values obtained at a specified threshold are rather lopsided in multiple classification tasks, so we calculate the corresponding P and R values of different thresholds and then plot the Precision-Recall curve (P-R curve). The area enclosed by the P-R curve with x and y axes is AP, which ranges from 0 to 1. The larger the AP, the better the model. Some models will predict several categories, and mAP is the value obtained by summing all AP values and then averaging.

Most algorithms use the principle of gradient descent to optimize the loss function and find optimal weight value. In order to optimize the result in the right direction, we need to feed our network with true results during the training process, this is supervised learning. Datasets play a crucial role in this kind of process. Popular data sets include VOC2007, VOC2012, Microsoft COCO [15], ImageNet [16] and Open Image Challenge Object Detection (OICOD). The VOC dataset was introduced in the PASCAL VOC Competition, the most important versions are VOC2007 and VOC2012, and they are the most widely used in the field of object detection. They split all images into three parts, the training set, the validation set, and the test set. As shown in Table I, VOC2007 and VOC2012 both consist of a large number of images. Each image has its corresponding xml file, which records the location and category information of those objects within an image.

TABLE I
Comparison of VOC2007 and VOC2012 Datasets

List	Train		Validation		Test	
	Image	Object	Image	Object	Image	Object
VOC2007	2501	6301	2510	6307	4952	12032
VOC2012	5717	13609	5823	13841	11540	27450
Total	8218	19910	8333	20148	16492	39482

III. Two-Stage Algorithms

A. R-CNN Model

Regions with CNN features (R-CNN) is one of the classical algorithms for two-stage object detection. The reason why it is called a two-stage algorithm is that in the object detection process, firstly, through some approaches, such as Selective Search [17], Multi-scale combinatorial grouping [18], or Edge Boxes method [19] to extract region proposal and then regression prediction is performed on the region proposal to obtain the predicted object, the whole detection consists of two stages. Compared to traditional object detection methods, R-CNN [10], proposed by Girshick in 2014, is the first to use CNNs to feature extraction of suggested region proposals.

This method achieves 58.5% mAP on the VOC2007 dataset, an improvement of 24.2% compared to DPMv5. However, R-CNN also faces many problems, the search method used is very time-consuming, which is one of the reasons why the whole network is not very time-efficient. Multiple models in the network have to be trained separately, which not only requires a lot of space to store those models but also increases training time. To solve the problem of inconsistent region proposal size, R-CNN directly changes the size of the input feature map, which inevitably loses some data from the original image.

B. Fast R-CNN Model

In 2015, the introduction of Fast R-CNN [11] improved the network structure of R-CNN and combined with SPPnet to propose a region of interest (RoI) pooling layer. This approach avoids the problem of repeated training of multiple suggestion boxes in R-CNN, and after RoI pooling, the problem of the inconsistent size of the fully connected layer that is obtained by Singular Value Decomposition (SVD) can be solved. The backbone network uses VGG-16 which has a deeper layer network. Testing on the VOC2007 dataset obtained mAP of 70%, while R-CNN and SPPnet were 66% and 63.1% respectively. The training time was also reduced from 84 hours in R-CNN to 9.5 hours, which is about 9 times faster than R-CNN. The whole detection process of Fast R-CNN still suffers some weaknesses.

1) *Using selective search methods to generate a large number of suggested regions proposal, resulting in a long training and prediction time, which has not yet achieved the purpose of real-time detection.*
2) *Multiple fully connected layers are used at end of the network and are calculated separately, the weights are not shared, which increases the number of parameters.*

C. *Faster R-CNN Model*

Ren et al. found that region proposal selection is a computational bottleneck during a two-stage algorithm, and therefore proposed the Faster R-CNN [12] in 2017 to address this problem. A new Region Proposal Network (RPN) is proposed to replace the sliding window in Fast R-CNN to generate regions proposal. RPN generates k anchor frames (k defaults is 9) in the convolutional feature layer. Anchor frames of different scales predict different sizes objects, and the location and classification are determined by category regression and anchor frame regression. This is also the advantage of Faster R-CNN, which improves the detection speed of whole network. Multi-channel convolutional layer, RPN layer, RoI pooling layer, classification and regression layer are responsible for feature extraction, region proposal selection, preventing images cropping distortion and optimization of results in its network respectively. A comparison with Fast R-CNN tested on different backbone networks revealed that Faster-RCNN with VGG-16 obtained 73.2% of mAP on the VOC2007 and the detection speed is 5 FPS. Due to large amount of down sampling in the network, however, the final anchor is less accurate for small objects, and the network also has more fully connected layers and a huge number of parameters, while the best FPS obtained from the test still does not meet the requirements of real-time detection.

TABLE II
PERFORMANCE COMPARISON OF TWO-STAGE ALGORITHMS

Model	**Backbone**	**Speed/($f \cdot s^{-1}$)**	**mAP/% VOC2007**
R-CNN	AlexNet	0.03	58.5
Fast R-CNN	VGG-16	7	70
Faster R-CNN	VGG-16	5	73.2

IV. ONE-STAGE ALGORITHMS

A. *YOLO Model*

The YOLO method is one of the one-stage detection algorithms that has performed a number of modifications and improvements. Redmon et al. in 2016 innovated the idea of incorporating regression into the object detection task and proposed the YOLOv1 [13] model. A single convolutional network can predict the position of an object and its category from an image, while eliminating the need for regions proposal, resulting in a significant speed increase of up to 45 FPS. A smaller FAST YOLO network achieved 155 FPS, which was twice the speed of other real-time object detection models at the time.

The process of YOLOv1 is divided into three steps: (1) Reshape the input image to 448×448, while to prevent image distortion, filling in the edges around the image to maintain the original image ratio of width to height. (2) Feed images into the trained model for prediction, the Fully Connected Layer (FC) will output the predicted objects' position and category. (3) Finally, through the Non-Maximum Suppression (NMS) algorithm to filter the Bounding Box (bbox), select the best result, and plot bbox, category, and confidence on the image. The test result on the VOC2007 dataset was 63.4% mAP.

To compensate for some drawbacks in YOLOv1, Redmon et al. proposed YOLOv2 [20] model in 2017. The main improvement points include replacing the backbone network with DarkNet-19 and adding a batch normalization operation to each convolutional layer, allowing for as much reduction as possible in one direction during gradient descent, preventing the gradient from jumping back and forth. For multi-scale training, YOLOv2 changes the size of input images according to its own iteration process, thus meeting the requirements of multi-scale images for training. To some extent, these methods all benefit the detection speed and mAP, finally, with 40 FPS and 78.6% mAP on the VOC2007.

YOLOv3 [21], in 2018 Redmon et al. further optimized the network structure and loss function, etc., based on YOLOv2 to make the model more accurate and robust. In contrast to YOLOv2, the backbone of YOLOv3 is DarkNet-53, which has a deeper network structure and uses a residual structure to prevent gradient disappearance. The Feature Pyramid Networks (FPN) [22] structure is also used to extract multi-scale features from images, which effectively improves the detection accuracy of small objects.

Bochkovskiy et al. proposed YOLOv4 [23] model in 2020. YOLOv4 was designed to achieve goals of higher speed and accuracy. Various network optimization solutions are used to improve network performance, such as increasing receptive fields, data augmentation, and changing the loss function. The backbone network was replaced with CSPDarkNet-53, which was experimentally found to be more suitable for object detection networks. Multichannel fusion using Path Aggregation Network (PANet) discards the previous FPN to improve the receptive field of the input images.

B. *SSD Model*

SSD [14] was proposed by Liu et al. in 2016 as a new end-to-end and one-stage object detection model based on YOLOv1. To overcome the inaccurate problem of small objects in YOLOv1, the SSD network structure is based on the VGG network, replacing the fully connected layers with 6 different sizes CNNs. Small objects are predicted by the front convolutional network, while large objects are detected by the back convolutional network. Experiments have also demonstrated that mixed-scale feature images can effectively improve accuracy.

Also, data enhancement is essential and SSD improves whole network robustness by changing the scale size of images, with a result of 79.8% mAP on the VOC2007. Unlike YOLOv1 which uses a fully connected layer to output results, SSD removes the last fully connected layer and uses a convolutional network to output results directly, reducing the total amount of parameters in the network.

In 2017 Jeong et al. proposed the R-SSD [24] model by combining pooling with deconvolution rather than expanding the number of feature maps. This approach allows the entire network to be characterized by both low-and high-resolution feature maps. The result is an increase in detection accuracy and speed without increasing network parameters and a better score of 80.8% mAP on VOC2007.

Under the influence of SSD, the improved models DSSD [25] and FSSD [26] have also been proposed sequentially. DSSD changes backbone of SSD from VGG to ResNet-101 that is a residual structure and a deeper network, which can effectively improve the accuracy and prevent gradients from disappearing during backpropagation. The mAP reached 81.5% on the VOC2007 dataset, but DSSD detection speed was only 5.5 FPS, which is one of drawbacks of this model. Like YOLOv3, inspired and influenced by feature pyramid structure, Li et al. introduced FSSD, which up samples and down samples input images while fusing different layers to create a richer network. It was then tested on the VOC2007 and obtained results of 82.7% mAP and 65.8 FPS.

C. *RetinaNet Model*

Lin et al. 2017 first proposed that the extreme foreground-background class imbalance can also have an impact on the prediction results, and used a new focal loss function to address this problem, making the entire training process more focused on difficult samples, and subsequently proposed the RetinaNet [27] model. In RetinaNet, the ResNet network is first used as backbone network, then the feature pyramid structure is also used to obtain rich multi-scale feature maps, finally there are two sub-networks responsible for category regression and prediction box regression respectively. When compared to YOLOv2 and DSSD algorithms, RetinaNet-101 has a clear advantage, with a mAP of 17.5% higher than YOLOv2 and 6.1% higher than DSSD513 when tested on the COCO dataset.

TABLE III
PERFORMANCE COMPARISON OF OEN-STAGE ALGORITHMS

Model	Backbone	Speed/($f \cdot s^{-1}$)	mAP/%		
			VOC2007	VOC2012	COCO
YOLOv1	VGG-16	45	63.4	57.9	-
YOLOv2	Darknet-19	40	78.6	73.5	21.6
YOLOv3	Darknet-53	51	-	-	57.9
YOLOv4	CSPDarknet-53	23	-	-	43.5
SSD	VGG-16	19.3	79.8	78.5	28.8
R-SSD	VGG-16	16.6	80.8	-	-
DSSD513	ResNet-101	5.5	81.5	80.8	33.2
FSSD	VGGNet	65.8	82.7	-	-
RetinaNet	ResNeXt-101+FPN	5.4	-	-	40.8

V. PRACTICAL APPLICATIONS

A. *Self-Driving Cars*

The ability to effectively and quickly recognize objects while driving is a key technology and challenge for self-driving cars, and one-stage object detection algorithms based on CNNs offer the possibility to achieve this. Real-time scene analysis and detection during driving can provide path planning and active safety measures. Fast and compact models are also a trend for the development of object detection on automotive platforms. In 2022, Li et al. proposed an EfficientDet-Gs [28] model with better efficiency and fewer network parameters based on the EfficientDet [29]. Based on YOLOv3, Zhang et al. [30] applied adaptive feature fusion and DSC approaches to improve network performance, and devised a technique to help drivers by automatically detecting vehicle license plates.

B. *Weeding Robots*

Although herbicides can be effective to eliminate weeds in agriculture, they also have many harmful effects. New weed control methods such as mechanical weeding, laser weeding and flame weeding are all based on the object detection. In 2018, Sun et al. [31] improved AlexNet to detect crops and weeds by using atrous convolution and global pooling, finally achieved better results compared to the original AlexNet, with an accuracy of over 90%. Zhang et al. [32] applied three different backbone networks, VGG-16, ResNet-50 and ResNet-101, to the Faster R-CNN model for ablation experiments and successfully applied CNNs to the oilseed rape and weeds recognition, through using migration model and testing on the COCO dataset, achieving precision and recall of 83.90% and 81.30%, respectively.

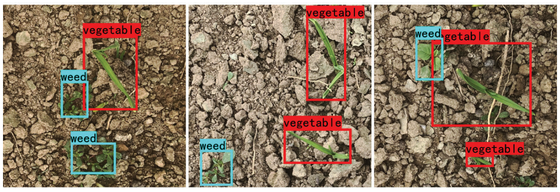

Fig. 1. Weeds detection results of weeding robot

C. *Face Recognition*

With the development of numerous detection algorithms, face recognition also is one of the great successful applications and provides many conveniences to your daily life. Face recognition can be used in a lot of industries, such as contactless payment, phone unlock, criminal investigation and pedestrian detection. As shown in Fig. 2, Jiang et al. [33] have found the sharing of convolutional layers would affect the effectiveness of face recognition using the Faster R-CNN. In 2021, Li et al. [34] proposed that illumination is an important factor for the performance of face recognition and established face datasets based on different light intensities, finally improved the detection accuracy to 98%.

Fig. 2. Face recognition results of Faster R-CNN

VI. Conclusion

This paper focuses on the evolution of object detection algorithms and their trends. Convolutional neural networks and deep learning algorithms stand out for their powerful feature extraction and excellent classification characteristics. Through a comprehensive overview to two kinds of object detection algorithms, we describe their concepts, network structure, and implementation methods, as well as evaluate their advantages and shortcomings, and then outline some applications in our daily lives. In general, the market for object detection is promising and potential. However, it also still faces many problems that need to solve such as:

1) Supervised learning requires a large number of data sets as support. It is only through tremendous training that the loss function can be optimally solved and weights can be optimized. However, due to confidentiality, capture difficulties and other reasons caused a lack of datasets in some industries.

2) Detection of small objects has always been a difficult problem, because the pixel data of small objects on the image is lacking, so training process cannot achieve a very satisfactory result. This is also one of the directions that need to be studied in the future.

3) For large datasets processing, only focusing on speed improvement is unrealistic since we should consider many factors, such as the parameter number, the performance of GPU, and the depth of the whole network. A reasonable strategy is to find out an acceptable balance between speed and economy according to our practical application.

References

[1] G. Bingtao, W. Xiaorui, C. Yujiao et al., "High-Accuracy Infrared Simulation Model Based on Establishing the Linear Relationship between the Outputs of Different Infrared Imaging Systems," *Infrared Physics & Technology*, vol. 69, pp. 155-163, Mar. 2015.

[2] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 886-893.

[3] S. R. Gunn, "Support Vector Machines for Classification and Regression," *ISIS Technical Report*, vol. 14, pp. 5-16, May. 1998.

[4] P. F. Felzenszwalb, R. B. Girshick, D. McAllester et al., "Object Detection with Discriminatively Trained Part-Based Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1627-1645, Sep. 2010.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1104, 2012.

[6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv Preprint arXiv:1409.1556*, vol. 1, pp. 1-14, Sep. 2014.

[7] C. Szegedy, W. Liu, Y. Jia et al., "Going Deeper with Convolutions," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1-9.

[8] K. He, X. Zhang, S. Ren et al., "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 1904-1916, Jan. 2015.

[9] K. He, X. Zhang, S. Ren et al., "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[10] R. Girshick, J. Donahue, T. Darrell et al., "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580-587.

[11] R. Girshick, "Fast r-cnn," in *Proc. IEEE International Conference on Computer Vision*, 2015, pp. 1440-1448.

[12] S. Ren, K. He, R. Girshick et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 1137-1149, Jun. 2015.

[13] J. Redmon, S. Divvala, R. Girshick et al., "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779-788.

[14] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: Single Shot Multibox Detector," *European Conference on Computer Vision*, 2016, pp. 21-37.

[15] T. Lin, M. Maire, S. Belongie et al., "Microsoft COCO: Common Objects in Context," in *Proc. European Conference on Computer Vision*, 2014, pp. 740-755.

[16] J. Deng, W. Dong, R. Socher et al., "ImageNet: A Large-Scale Hierarchical Image Database," *IEEE*, pp. 248-255, Jun. 2009.

[17] J. R. Uijlings, K. E. Van De Sande, T. Gevers et al., "Selective Search for Object Recognition," *International Journal of Computer Vision*, vol. 104, pp. 154-171, Sep. 2013.

[18] P. Arbeláez, J. Pont-Tuset, J. T. Barron et al., "Multiscale Combinatorial Grouping," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 328-335.

[19] C. L. Zitnick and P. Dollár, "Edge Boxes: Locating Object Proposals from Edges," in *Proc. European Conference on Computer Vision*, 2014, pp. 391-405.

[20] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7263-7271.

[21] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv Preprint arXiv:1804.02767*, pp. 1-6, Apr. 2018.

[22] T. Lin, P. Dollár, R. Girshick et al., "Feature Pyramid Networks for Object Detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117-2125.

[23] A. Bochkovskiy, C. Wang, and H. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv Preprint arXiv:2004.10934*, pp. 1-17, Apr. 2020.

[24] J. Jeong, H. Park, and N. Kwak, "Enhancement of SSD by Concatenating Feature Maps for Object Detection," *arXiv Preprint Arxiv:1705.09587*, pp. 1-12, May. 2017.

[25] C. Fu, W. Liu, A. Ranga et al., "DSSD: Deconvolutional Single Shot Detector," *arXiv Preprint arXiv:1701.06659*, Jan. 2017.

[26] Z. Li and F. Zhou, "FSSD: Feature Fusion Single Shot Multibox Detector," *arXiv Preprint arXiv:1712.00960*, Dec. 2017.

[27] T. Lin, P. Goyal, R. Girshick et al., "Focal Loss for Dense Object Detection," in *Proc. IEEE International Conference on Computer Vision*, 2017, pp. 2980-2988.

[28] Y. LI, X. Zhang, M. Zhang et al., "Object Detection in Autonomous Driving Scene Based on Improved Efficient Diet," *Computer Engineering and Applications*, vol. 58, pp. 183-191, Mar. 2022.

[29] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and Efficient Object Detection," in *Proc. EE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10781-10790.

[30] J. Zhang, Y. Xiong, X. Sun et al., "License Plate Detection Using Siamese Feature Pyramid and Cascaded Positioning," *Computer Engineering and Applications*, vol. 12, no. 12, pp. 1-15, Nov. 2021.

[31] J. Sun, X. He, W. Tan et al., "Recognition of Crop Seedling and Weed Recognition Based on Dilated Convolution and Global Pooling In CNN," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 34, pp. 159-165, 2018.

[32] L. Zhang, X. Jin, L. Fu et al., "Recognition Method for Weeds in Rapeseed Field Based on Faster R-Cnn Deep Network," in *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, 2020, pp. 304-312.

[33] H. Jiang and L. Erik, "Face Detection with The Faster R-CNN," in *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, 2017, pp. 650-657.

[34] W. Li, Y. Liu, and G. Xing, "Illumination Analysis of Deep Face Recognition," *Journal of Computer-Aided Design & Computer Graphics*, vol. 34, pp. 74-83, Nov. 2021.

MingYuan Wang received a B.Eng. degree in Mechanical Design, Manufacturing, and Automation from Henan University of Science and Technology, Henan, China, in 2013. Past work experience as an acting product designer at YTO Group Corporation from 2014 to 2020. He currently is a graduate student at the Department of Computer Engineering, Khon Kaen University, Thailand. His current research interests include computer vision and its applications to other science and engineering areas.

Watis Leelapatra received a B.Eng in Computer Engineering from Khon Kaen University, Khon Kaen, Thailand, an M.S. degree in Computer Science from the Case Western Reserve University, Cleveland, USA, and a D.Eng. degree in Computer Science from Asian Institute of Technology, respectively. He is concurrently a lecturer at the Department of Computer Engineering, Khon Kaen University.