A Nighttime Vehicle Detection Method Based on YOLO v3

Yan Miao
College of Communication Engineering
Jilin University
Changchun, China
miaoyan17@mails.jlu.edu.cn

Fu Liu
College of Communication Engineering
Jilin University
Changchun, China
liufu@jlu.edu.cn

Tao Hou
College of Communication Engineering
Jilin University
Changchun, China
ht_happy@jlu.edu.cn

Lu Liu
College of Communication Engineering
Jilin University
Changchun, China
liulu17@mails.jlu.edu.cn

Yun Liu
College of Communication Engineering
Jilin University
Changchun, China
liuyun313@jlu.edu.cn

2020 Chinese Automation Congress (CAC) | 978-1-7281-7687-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/CAC51589.2020.9326819

***Abstract*—For the past several decades, the number of vehicles has been increased for many times, leading to a lot of traffic issues. The nighttime accident rate is much higher than that in daytime because of the weak light and the uneven distribution of nighttime brightness. Thus, in this paper an effective nighttime vehicle detection approach is designed. First, the original nighttime images were enhanced by an optimal MSR algorithm. Then, a pretrained YOLO v3 network was selected and fine-tuned by the enhanced images. Finally, the detection network was used to detect vehicles from the nighttime images and outperformed two widely used object detection methods, namely the Faster R-CNN and SSD, on the precision and detection efficiency. The average precision of the proposed method reaches 93.66%, which is 6.14% and 3.21% higher than that of the Faster R-CNN and SSD, respectively.**

Keywords—nighttime, image enhancement, vehicles detection, YOLO

I. INTRODUCTION

For the past several decades, the amount of vehicles has been raised year by year. By the end of 2019, the number of vehicles in China exceeded 348 million[1]. Our life becomes much more convenient with the increasing quantities of vehicles. However, at the same time, more vehicles bring a lot of traffic problems. It has been reported that more than 100 thousand people in China die in traffic accidents every year, and the accident rate at night is 1.5 times than that at daytime[2]. Therefore, an effective nighttime vehicle detection method is of vital importance to be designed.

Since many problems in condition of night environments such as the poor sight distance of drivers, less references, scene complexity, nighttime vehicle detection is more complex and harder than that in daytime. Existing detection methods for nighttime vehicle are roughly categorized into three aspects: motion based, vehicle lamp detection based, and deep neural network based methods. The motion based method usually selects a threshold to extract the moving regions in images by the pixel differences with their adjacent frames of video[3]. It is not so much robust due to the uneven distribution of brightness in the nighttime road images. Guo[4] and Kosaka[5] both proposed lamp detection based vehicle detection mechanisms to detect vehicles in nighttime scenarios. They segmented vehicle lamps and filtered out reflections to calculate how much similarities and symmetries of light against the vehicle lamps. The two vehicle lamp detection based methods may not work well when the lighting environment in the image is complex. More lighting information from the background will lead to misclassifications in some degree. Deep learning methods (Fast R-CNN[6], Faster R-CNN[7], R-FCN[8]) are generally two-stage methods. They used several convolutional neural networks (CNN) to extract related region proposals. Then determined which region proposals were more contributing to detecting vehicles. These two-stage methods usually have a high computation consuming, which are not reliable for the real-time detection. In order to make an acceleration, one-stage deep learning methods, such as YOLO v1-v3[9-11], were proposed by calculating the categories and location information directly by a single CNN. They utilized multi-scale features fusion instead of region proposals so that they had a lighter neural network, which will both reduce the computational consuming and ensure the accuracy of detection.

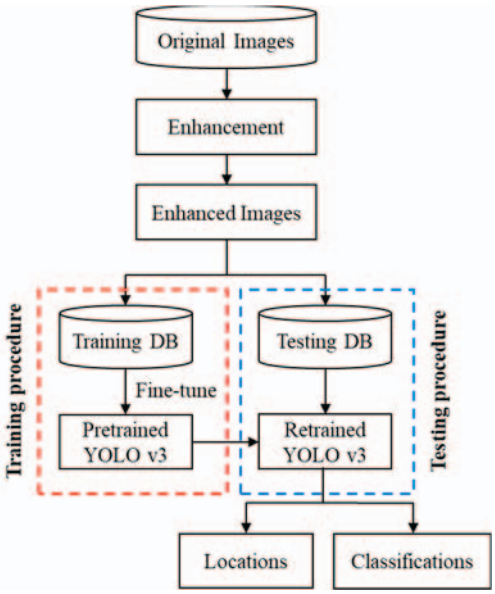


Fig. 1. The workflow of our nighttime vehicle detection method.

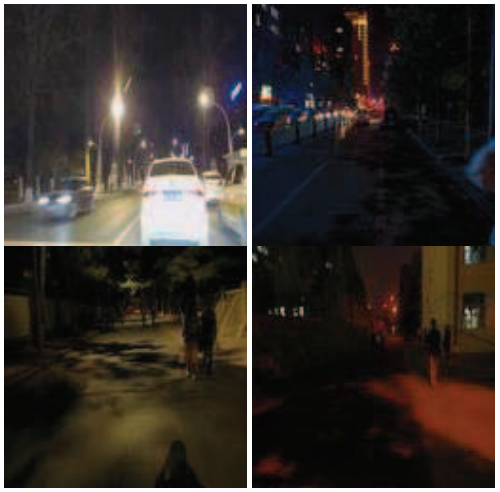
As the development of the YOLO from v1 to v3, YOLO v3 has become one of the most popular object detection methods. It has the ability of detecting relatively small objects from complex background with less computational consuming. So in this paper, a pre-trained YOLO v3 was utilized to construct the nighttime vehicles detection model. First, the collected nighttime images were enhanced by an

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on May 16,2021 at 16:45:53 UTC from IEEE Xplore. Restrictions apply.

optimal MSR algorithm to even the brightness and improve the sharpness and detail information of images. Then, a pretrained YOLO v3 network was selected and fine-tuned by the enhanced images. After several iterations of training, the model was utilized to extract the features of images. Finally, the network was used to detect vehicles from the nighttime images.

II. METHODS

The workflow of our nighttime vehicle detection approach is shown in Fig. 1. First, the original nighttime images were processed by an optimal MSR algorithm to improve the brightness and detail information of images. Then, a pretrained YOLO v3 network was fine-tuned by the training dataset from enhanced images. After several iterations of training, the retrained model was utilized to extract the features of images. Finally, the network was used to detect vehicles from the nighttime images in the testing dataset by outputting locations and classifications of the detected objects.



(a) Samples of original images.

(b) Samples of corresponding enhanced images.

Fig. 2. Samples of original and corresponding enhanced images.

A. Imang enhancement by an optimal MSR algorithm

In order to enhance the quality of the nighttime road images and the visibility and contrast of the vehicles in the images, an optimal MSR (Multi Scale Retinex) algorithm[12] was applied before CNN model, which alleviated the uneven brightness of images and improved the sharpness and detail information of images.

First, an image with the RGB color space was transformed from to the YUV color space. Secondly, the reciprocal of the minimum perceptible difference Just Noticeable Distortion (JND)[13] was used as the coefficient of incident image in MSR algorithm[14] to construct an optimized MSR algorithm, and the brightness of Y channel was adjusted adaptively by using this optimization algorithm, and the U and V channels were adjusted according to the changing proportion to generate a fused image. Then the fused image was fused with the original image by 1:1. Finally, the adaptive histogram CLAHE[15] with limited contrast was used to enhance the image contrast, and the final enhanced image was obtained.

All 1,760 nighttime images were enhanced by the optimal MSR algorithm. Samples of original and corresponding enhanced images were shown in Fig. 2. It is clear to be seen that the difference of brightness between the bright parts and the dark parts becomes smaller after enhancement, which may be more conducive to vehicles detection at night.

B. YOLO v3 based nighttime vehicles detection model

1) Darknet-53

YOLO v3 network (Fig. 3) is structured by 53 convolutional layers with some 3×3 and 1×1 filters, among which some residual connections namely shortcuts are added, namely Darknet-53. The architecture is more generally used than Darknet-19[10] and more significant than other residual networks, such as ResNet-101 and ResNet-152[16]. The convolutional layers with a stride of 2 are utilized between blocks to subsample feature maps. These bottleneck layers play the same role as the pooling layers in conventional CNN[17-22]. The last three feature maps with different scales are fused together so that detecting multi-scale objects is avalible.

2) Prediction of the Bounding Box

YOLO v3 uses dimension clusters to make predictions for bounding boxes as anchor boxes as what has been down in YOLO v2[10]. The network predicts five parameters, m_x, m_y, m_w, m_h, m_o, for each bounding box, which are used to predict the center and width of the border and the confidence level, respectively. The predictions correspond to:

$$box_x = \sigma(m_x) + d_x \tag{1}$$

$$box_y = \sigma(m_y) + d_y \tag{2}$$

$$box_w = q_w e^{m_w} \tag{3}$$

$$box_h = q_h e^{m_h} \tag{4}$$

where (d_x, d_y) represents the distance of the cell in the image; (q_w, q_h) shows the width and height of the bounding box prior; σ is the sigmoid function. Fig. 4 shows the bounding boxes that have both location prediction and dimension priors. The center coordinates of the box are determined by a sigmoid function. The accumulation of the squared error loss is calculated during training. Then during logistic regression, the network generates a score representing the correct probability of each bounding box.

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on May 16,2021 at 16:45:53 UTC from IEEE Xplore. Restrictions apply.

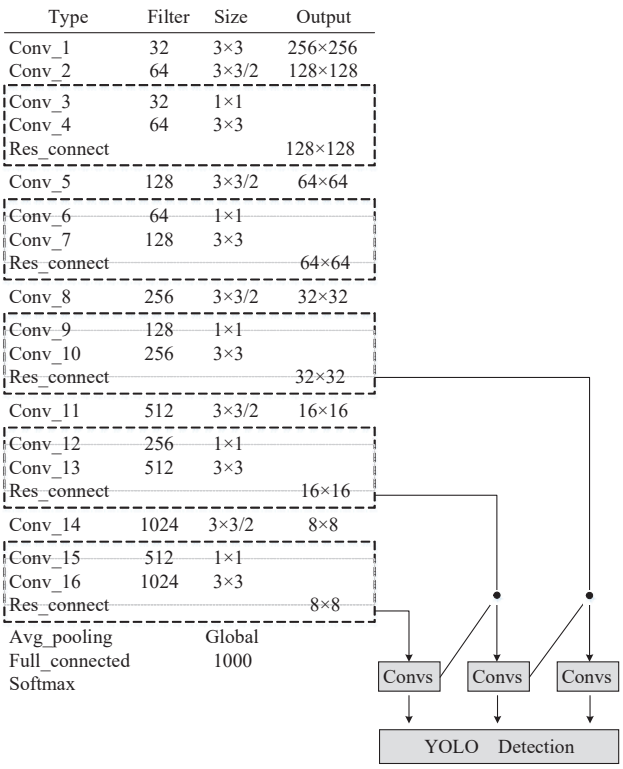


Fig. 3. Construction of YOLO v3 network.

3) Prediction

Every box calculates the probable categories in the bounding box through multi-label classification. Instead of a softmax that may be unnecessary for good performance, logistic classifiers of independency are used directly. Binary cross-entropy loss is calculated to predict the category while training.

III. Experiments and Results

We used Logitech C920 camera to collect 880 images containing vehicles in real nighttime scenes (Fig. 5). All of these images were mirrored horizontally to enlarge our dataset twice. The enlarged dataset was separated into two subsets randomly. The training set contains 1232 images. The testing set has 528 images. Then all of the 1760 images were labeled by LabelImg, an image annotation tool by Python. Fig. 6 shows a labeled image. We utilized Python 3.6.5 with Tensorflow 1.3.0. The CPU was Intel Xeon E5-2689. Nvidia Geforce GTX 1080Ti was used as GPU to accelerate the detection procedure.

Different sizes of bounding box priors determine different sizes of objects. In this paper, the sizes of bounding box priors were chosen as (4,8), (6,16), (10,10), (8,31), (13,20), (22,16), (22,30), (13,51), (36,42), (25,89), (54,66), (83,95), (57,155), (116,156), respectively. 'Darknet53.conv.74' was loaded as the pre-trained model. The active function was chosen as the Leaky-ReLU[23] with a slope of 0.01. In the progress of fine-tuning, momentum[24] with learning rates of 0.001 and 0.0001 before and after 15,000 iterations, respectively, was utilized to minimize the average loss. The training process will be terminated when the average loss does not decrease any more or oscillated between a certain value to avoid over fitting. The average loss in the process of training was shown in Fig. 7. When the model was trained for 35,000 iterations, the current average loss was stable at about 0.15. So we terminated the training strategy at 35,000 iterations. It is worth mentioning that the average loss decreased obviously at around 15,000 iterations. Thus we plotted the training accuracy from 2,000 to 35,000 iterations in Fig. 8. The training accuracy at 15,000 iterations was higher than the others, so the model trained by 15,000 iterations was utilized to detect nighttime vehicles from test dataset.

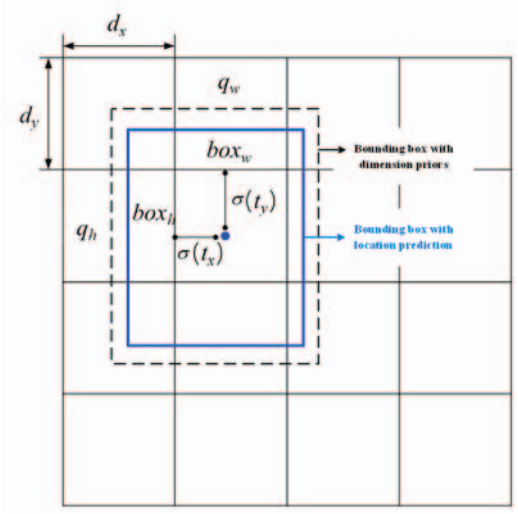


Fig. 4. Bounding boxes with dimension priors and location prediction.

Fig. 5. Samples of nighttime vehicles images.

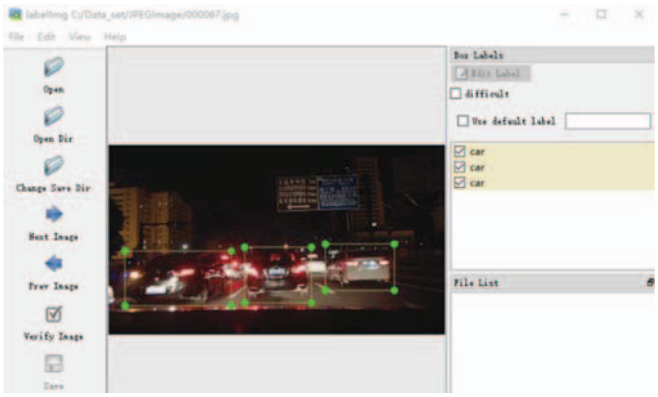


Fig. 6. An example of one labeled image.

The experimental results of the proposed method are shown in Fig. 9 (the detection confidence threshold was

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on May 16,2021 at 16:45:53 UTC from IEEE Xplore. Restrictions apply.

chosen as 0.8). Faster R-CNN[7] and SSD[25] are used to detect vehicles from testing set for comparison. For large and close vehicles, all of the three methods had a good performance. However, when it came to the vehicles with small target, vehicles with fuzzy contour information, vehicles with missing contour information, and vehicles parked side by side on the road, our method could detect a lot of more vehicles than the comparison methods. When there were overlapped vehicles in images, the proposed method still worked well. This may benefit from the fusion of multi-scale feature maps.

The average precision (AP) and frame per second (FPS) were calculated to further evaluate the performance of the three methods. AP represents the ratio of the accuracy of objects with category A on each image to the number of images containing category A. FPS represents the number of images that can be processed by the program per second under the same hardware condition, which reflects the detection efficiency of the method. The criteria is comparison in TABLE I. The AP and FPS of our method surpassed the Faster R-CNN and SSD, reaching 93.66% and 30.03, respectively. The proposed method achieved 6.14% and 3.21% higher AP and 20.26 and 4.09 higher FPS than the Faster R-CNN and SSD, respectively.

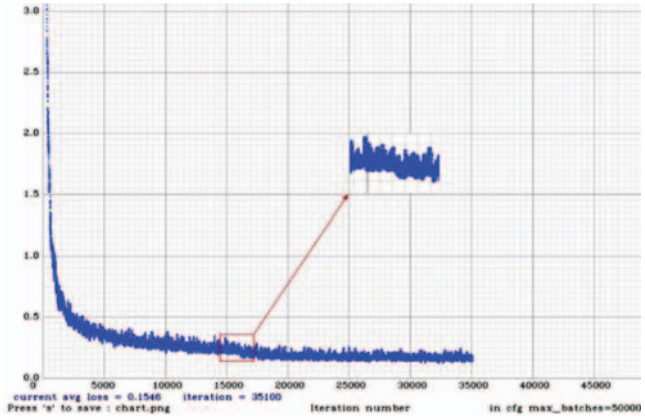


Fig. 7. The average loss in the process of training.

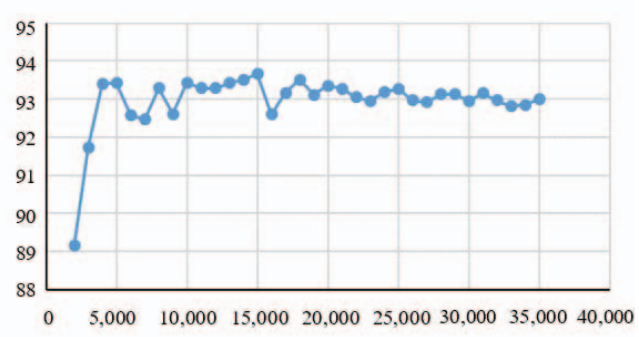
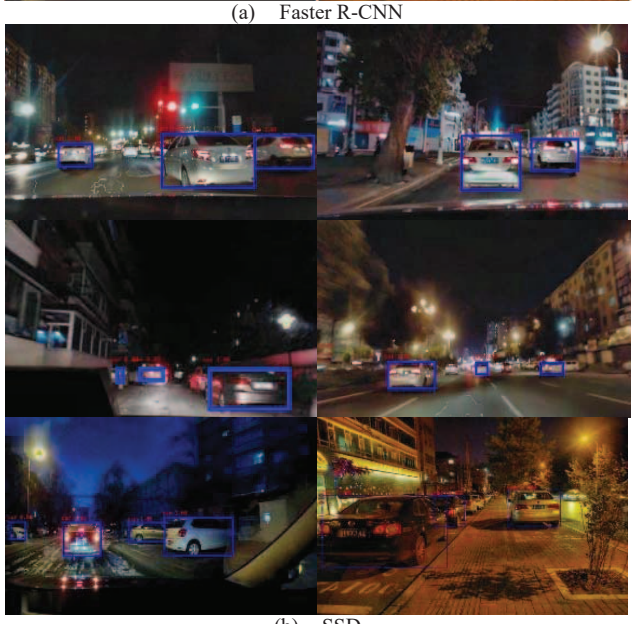


Fig. 8. The training accuracy from 2,000 to 35,000 iterations.

TABLE I. THE CRITERIA COMPARISON OF THREE METHODS

Methods	Criteria	
	***AP*/%**	***FPS***
Faster R-CNN	87.52	9.77
SSD	90.45	25.94
Proposed method	**93.66**	**30.03**

(a) Faster R-CNN



(b) SSD

(c) Proposed method

Fig. 9. Performance of nighttime vehicles detection.

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on May 16,2021 at 16:45:53 UTC from IEEE Xplore. Restrictions apply.

IV. CONCLUSION

In this paper, a YOLO v3 based nighttime vehicle detection method is proposed. All images were enhanced by an optimal MSR algorithm to alleviate the uneven brightness and to improve the sharpness and detail information. We established a nighttime vehicles dataset to train a pre-trained YOLO v3 network. Evaluated on the testing dataset, the proposed method detect more vehicles, and achieving a higher AP and FPS values than the Faster R-CNN and SSD.

ACKNOWLEDGMENT

The work in the paper has been supported by the National Natural Science Foundation of China (61503151) and the Natural Science Foundation of Jilin Province (20160520100JH). The corresponding author is Liu Yun (e-mail: liuyun313@jlu.edu.cn).

REFERENCES

[1] Dong X, Pang Y, Wen J, "Fast efficient algorithm for enhancement of low lighting video", ACM SIGGRAPH 2010 Posters (ACM), 2010.

[2] Li J, Li S Z, Pan Q, et al., "Illumination and motion-based video enhancement for night surveillance", Joint IEEE International Workshop on Visual Surveillance & Performance Evaluation of Tracking & Surveillance, 2005.

[3] Luo Min, Liu Dongbo, Wen Haoxuan, et al., "A New Vehicle Moving Detection Method Based on Background Difference and Frame Difference", Journal of Hunan Institute of Engineering (Natural Science Edition), 2019, Vol. 29, pp. 58-61.

[4] Guo J M, Hsia C H, Wong K S, et al., "Nighttime Vehicle Lamp Detection and Tracking With Adaptive Mask Training", IEEE Transactions on Vehicular Technology, 2016, Vol.65, pp. 4023-4032.

[5] Kosaka N, Ohashi G, "Vision-Based Nighttime Vehicle Detection Using CenSurE and SVM", IEEE Transactions on Intelligent Transportation Systems, 2015, Vol.16, pp. 1-10.

[6] R. Girshick, "Fast R-CNN", 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448.

[7] Ren S, He K, Girshick R, et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, Vol.39.

[8] Jifeng Dai, Yi Li, Kaiming He, et al., "R-FCN: Object Detection via Region-based Fully Convolutional Networks", 2016, arXiv [cs.CV]: https://arxiv.org/abs/1605.06409.

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, "You Only Look Once: Unified, Real-Time Object Detection", 2015, arXiv [cs.CV]: https://arxiv.org/abs/1506.02640.

[10] J. Redmon, A. Farhadi, "Yolo9000: Better, faster, stronger", In Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517–6525.

[11] Joseph Redmon, Ali Farhadi, "YOLOv3: An Incremental Improvement", 2018, arXiv [cs.CV]: https://arxiv.org/abs/1804.02767.

[12] Liu Fu, Liu Lu, Hou Tao, et al., "Night road image enhancement method based on optimized MSR", Journal of Jilin University (Engineering and Technology Edition): 1-8[2020-08-29]. https://doi.org/10.13229/j.cnki.jdxbgxb20190835.

[13] X.K. Yang, W.S. Ling, Z.K. Lu, et al., "Just noticeable distortion model and its applications in video coding", Signal Processing: Image Communication, 2005, Vol.20, pp. 662-680.

[14] Wen Wang, Bo Li, Jin Zheng, et al., "A fast Multi-Scale Retinex algorithm for color image enhancement", International Conference on Wavelet Analysis and Pattern Recognition, 2008, pp. 80-85.

[15] Reza A M, "Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement", Journal of Vlsi Signal Processing Systems for Signal Image & Video Technology, 2004, Vol.38, pp. 35-44.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, "Deep Residual Learning for Image Recognition", 2015, arXiv [cs.CV]: https://arxiv.org/abs/1512.03385.

[17] Krizhevsky A., Sutskever I., Hinton G.E, "Imagenet classification with deep convolutional neural networks", In Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012, pp. 1106–1114.

[18] Simonyan K., Zisserman A, "Very Deep Convolutional Networks for Large-Scale Image Recognition", 2014, arXiv[cs.CV]: https://arxiv.org/abs/1409.1556.

[19] Szegedy C., Liu W., Jia Y., et al., "Going deeper with convolutions", In Proceedings of the IEEE Conference on Computer Vision and PatternRecognition, 2015, pp. 1–9.

[20] Szegedy C., Ioffe S., Vanhoucke V, "Inception-v4, inception-resnet and the impact of residual connections on learning", 2016, arXiv[cs.CV]: https://arxiv.org/abs/1602.07261.

[21] Szegedy C., Vanhoucke V., Ioffe S., et al., "Rethinking the inception architecture for computer vision", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[22] Huang G., Liu Z., van der Maaten L., et al., "Densely connected convolutional networks", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[23] Maas A. L., Hannun A. Y., Ng A. Y, "Rectifier nonlinearities improve neural network acoustic models", In in ICML Workshop on Deep Learning for Audio, Speech and Language Processing, 2013.

[24] Ning Qian, "On the momentum term in gradient descent learning algorithms", Neural networks: the official journal of the International Neural Network Society, 1999, Vol12, pp. 145–151.

[25] Liu W, Dragomir Anguelov, Dumitru Erhan et al., "SSD: Single Shot MultiBox Detector", European Conference on Computer Vision (ECCV), 2016, pp. 9905.

Authorized licensed use limited to: UNIVERSITY OF NEW MEXICO. Downloaded on May 16,2021 at 16:45:53 UTC from IEEE Xplore. Restrictions apply.