Spatial Pyramid Pooling and Adaptively Feature Fusion based Yolov3 for Traffic Sign Detection

Shimin Xiong
School of Computer Science, Northeast Electric Power University, China
2202000683@neepu.edu.cn

Bin Li
School of Computer Science, Northeast Electric Power University, China
corresponding author:
libinjlu5765114@163.com

Shiao Zhu
School of Computer Science, Northeast Electric Power University, China
shiao_zhu0411@163.com

Dongfei Cui
School of Computer Science, Northeast Electric Power University, China
Yong1ling@163.com

Xiaonan Song
School of Computer Science, Northeast Electric Power University, China
wangdongxiajlnu@163.com

Abstract: *Traffic sign detection is a key part of intelligent assisted driving, but also a challenging task due to the small size and different scales of objects in foreground and closed range. In this paper, we propose a new traffic sign detection scheme: Spatial Pyramid Pooling and Adaptively Spatial Feature Fusion based Yolov3 (SPP and ASFF-Yolov3). In order to integrate the target detail features and environment context features in the feature extraction stage of Yolov3 network, the Spatial Pyramid Pooling module is introduced into the pyramid network of Yolov3. Additionally, Adaptively Spatial Feature Fusion module is added to the target detection phase of the pyramid network of Yolov3 to avoid the interference of different scale features with the process of gradient calculation. Experimental results show the effectiveness of the proposed SPP and ASFF-Yolov3 network, which achieves better detection results than the original Yolov3 network. It can archive real-time inference speed despite inferior to the original Yolov3 network. The proposed scheme will add an option to the solutions of traffic sign detection with real-time inference speed and effective detection results.*

Keywords: *Traffic sign detection, yolov3, spatial pyramid pooling, adaptively spatial feature fusion.*

Received July 26, 2021; accepted February 6, 2022
https://doi.org/10.34028/iajit/20/4/5

1. Introduction

Intelligent assisted driving technology is committed to building a safe, green and comfortable transportation system. As a key part of intelligent driving assistance technology, traffic sign detection technology needs to accurately and real-time locate and classify the traffic signs in the scene when the vehicle is running at high speed [22]. Traffic sign detection is vital to driving safety, as well as 3D object detection [20], stress level detection [7] and other safety approaches.

Traffic sign detection methods can be roughly divided into two categories: traditional feature based [3, 4, 8, 21] method and deep learning based method. The former method is mainly divided into two steps: feature extraction (using Histogram of Oriented Gradient (HOG), Scale Invariant Feature Transform (SIFT), Key lines [2], etc.,) and classification (into two categories: target and not target). However, the traditional features of images are low-level visual features, which are often designed for certain objects. Therefore, the traditional method suffers from poor robustness and unsatisfying accuracy when facing detection tasks with complex background.

The deep learning based method can be divided into two categories: candidate region based approaches and end-to-end approaches. The detection process of these candidate region based approaches [5, 15, 25] is divided into two stages. In the first stage, the candidate region is generated from the input image, and then the target in the candidate region is corrected and classified. The candidate region based approaches have high detection accuracy, but those approaches have high computational complexity and cannot meet the needs of real-time detection. The end-to-end detection approaches [18, 24] send the image to be detected into the network, and then the location coordinates and the probability of the category of the target in the image can be obtained. The end-to-end detection approaches do not need to generate candidate regions, so they can meet the real-time requirements. Among the end-to-end target detection networks, the object detection method based on Yolo [12, 13] attracted researchers due to the superiority of real-time processing and high accuracy. At present, many studies are based on Yolov3 network to complete the task of traffic sign detection. Zhang *et al*. [22] Add the augmented path and data augmentation module to the Yolov3 network, and obtain the traffic sign detection

effect with faster speed and better accuracy than the traditional method. Zhang *et al*. [23], utilized a modified YOLOv2 to design their real-time traffic sign detection method. Sichkar and Kolyubin [16] proposed an end-to-end traffic sign detection algorithm based on Yolov3. In [1], different Yolo architectures are used as baseline detectors, and several preprocessing methods are proposed to meet the real-time requirements of traffic signs.

Traffic sign detection networks [9, 11, 14, 17, 19], including Yolov3, have pyramid network structures. When detecting traffic signs, this network structure will face two challenges: small in size, and different scales of objects in foreground and closed range. The former may cause the detection network to miss the detailed features in the detection scene, and then lead to the missed detection of traffic signs. The latter may lead to the problem that in the pyramid structure of the detection network, the corresponding features of traffic signs of different scales conflict with each other and interfere with gradient back propagation. In order to recognize traffic signs more accurately, many studies [9, 11, 14, 17, 19] focus on how to better use deep neural network to extract the features of traffic signs. Because the proportion of traffic signs in the detection scene is small, the characteristics of traffic signs are the details of the scene. Only when the detail features in the scene are combined with the overall environmental context features of the scene can it be more helpful for the successful detection of traffic signs. Moreover, the traffic sign detection methods based on pyramid detection network [9, 11, 14, 17, 19] did not consider the problem of different scale traffic sign features interfering with the network gradient back propagation process. In order to solve the above two challenges in the process of traffic sign detection, in this paper, a new traffic sign detection scheme: Spatial Pyramid Pooling and Adaptively Feature Fusion based Yolov3(SPP and ASFF-Yolov3) is proposed. Our main contributions are as follows:

1. Three Spatial Pyramid Pooling modules are added in the feature extraction stage of Yolov3 network to integrate the detailed features of traffic signs and the global environment semantic information of traffic signs.
2. Three Adaptively Spatial Feature Fusion modules are added in the target detection stage of Yolov3 network to solve the problem that different scale features affect gradient calculation in the training process.
3. Experiments on Tsinghua-Tencent 100K traffic sign database show that the proposed traffic sign detection method based on SPP&ASFF-Yolov3 is effective and feasible.

The rest of this paper is arranged as follows. Section 2 gives a briefly introduce about the structure of the Yolov3, Spatial Pyramid Pooling and the Adaptively Spatial Feature Fusion. Section 3 introduces the proposed SPP and ASFF-Yolov3 traffic sign approach in details. In section 4, we evaluate the SPP&ASFF-Yolov3 on Tsinghua-Tencent 100K traffic sign detection database, and section 5 concludes the paper.

2. The Network Structure of Yolov3

Figure 1 illustrates the overall work flow of our proposed network. SPP modules are introduced into the feature extraction network, and there are three scales in the detection network with ASFF modules integrated.

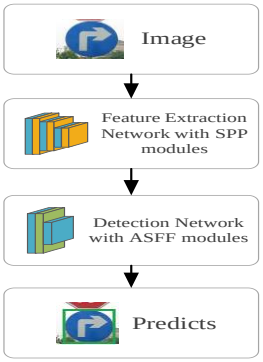


Figure 1. Work flow of SPP&ASFF-Yolov3.

The network structure of Yolov3 can be roughly divided into two parts: a bottom-up feature extraction network (the part surrounded by yellow dotted line in Figure 2) and a top-down multi-scale detection network (the part surrounded by purple dotted line in Figure 2). The feature extraction network with depth of 75 layers consists of a DBL module and five residual blocks: res1, res2, res8, res8, res4. DBL is the most basic part of Yolov3, including convolution layer, Batch Normalization (BN) operation and Leaky ReLU activation function.

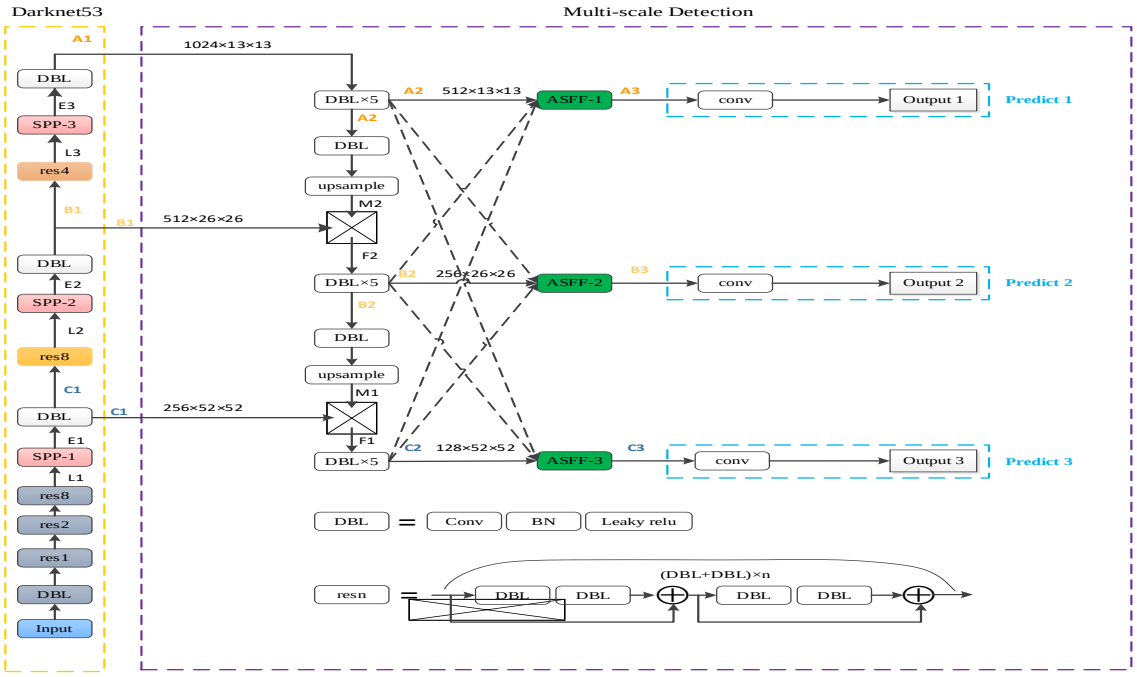


Figure 2. Network structure of SPP&ASFF-Yolov3.

The feature extraction network can be divided into three stages: the first stage is the gray DBL module and the following three residual modules, the second stage is the yellow res8 module and the third stage is the orange res4 module. The feature maps generated in the three stages are C1, B1, and A1. The feature map generated in each stage not only propagates upward to continue to extract higher-level features, but also propagates directly to the corresponding scale of the detection network. From the dimensions of feature map C1, B1, and A1 shown in Figure 2, we can see that Yolov3 network is a network with pyramid structure. The original Yolov3 network did not contain the dotted line in Figure, three white DBL rectangles, SPP and ASFF modules in Figure 2.

3. SPP&ASFF-Yolov3

3.1. SPP Modules In Yolov3 Network

Due to the small proportion of traffic signs in the scene, the traffic signs related features extracted by the feature extraction algorithm are the detailed features of the scene. Only when the detail features are combined with the context reflecting the overall scene of the image, can they have better feature expression ability and play a better role.

Spatial pyramid pooling network was proposed by He *et al*. [6] in 2015. In the previous convolutional neural network, the size of the input image is fixed, so it is necessary to cut or stretch the input image. The SPP was originally designed to solve the impact of clipping and stretching image on feature extraction. The SPP can make convolutional neural network accept any size of input image.

If the input image is replaced by feature map, Spatial Pyramid Pooling (SPP) can complete the task of fusing different scale features.

We introduce SPP module into the feature extraction network of Yolov3, and name it as spatial pyramid potential Yolov3 (SPP-Yolov3). The positions of the three SPP modules in Yolov3 network are shown in the three pink rectangles in Figure 2. In Figure 2, the input passes through a DBL module (i.e., the module for conv2d+BN+leaky operation) and three residual modules (res1, res2, and res8 shown in the gray filling block in the figure) and then outputs the feature map L1. The L1 enters the SPP-1 module, and the output feature map of the SPP-1 network module is E1. The number of channels of the feature map E1 is adjusted by a DBL module, and the adjusted feature map is C1. The trend of the C1 is divided into upward and right branches. In the right branch, feature graph C1 is fused with top-down high-level semantic feature M1 and output feature map F1. After the feature map F1 passes through the DBL×5 module, the DBL×5 module outputs the feature map C2, and then the feature map C2 is input into the branch of predict 3 in the multi-scale detection network for the minimum scale traffic sign detection.

Table 1. detailed parameters of three SPP modules.

	Dimensions of K. M, N and input L	The size of convolution kernel		
		k×k	m×m	n×n
SPP-1	52×52×256	43×43	47×47	51×51
SPP-2	26×26×512	17×17	21×21	25×25
SPP-3	13×13×1024	5×5	9×9	13×13

In the upward branch, C1 passes through the res8 (yellow filling block), and then res8 outputs the feature map L2. L2 enters SPP-2 module. The output feature

map of SPP-2 module is E2. DBL module adjusts the channel number of E2, and the adjusted feature map is B1. The trend of B1 is also divided into two branches: up and right.

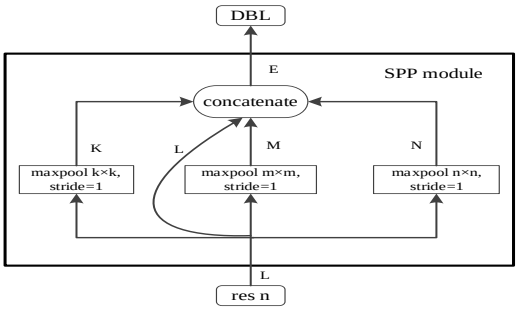


Figure 3. Network structure of SPP module.

In the right branch, B1 is fused with the top-down high-level semantic feature M2, and the fused feature map is F2. F2 passes through DBL×5 module, and the output of DBL×5 module is B2. B2 is input into predict2 branch of multi-scale detection network to detect medium-scale traffic signs. In the upward branch, B1 passes through the residual module res4 (orange filling block), and the output of res4 is the feature map L3. L3 enters SPP-3 module, and the output of SPP-3 module is E3. DBL module adjusts the channel number of E3, and the adjusted characteristic diagram is A1. A1 passes through DBL×5 module, and the output of DBL×5 module is feature map A2. Then A2 is input into predict1 branch of multi-scale detection network for large-scale traffic sign detection.

The structures of SPP-1, SPP-2 and SPP-3 are shown in Figure 3. It can be seen from Figure 3 that before the SPP module is the residual module, and after the SPP module is the DBL module. In Figure 3, the input of SPP module is L. The SPP module uses three scale convolution kernels to perform the maximum pooling operation with stride of 1 and padding and obtains K. M and N with the same size as the input L. The K. M and N correspond to the detail features and overall features of different scales in the detection scene respectively. Finally, K. M, N, and L are concatenate in SPP module and get the final output E. The detailed parameters of SPP-1, SPP-2 and SPP-3 are shown in Table 1.

3.2. ASFF Modules in Yolov3 Network

Figure 4 is a schematic diagram of the impact of multi-scale targets on the detection network. The blue parallelogram in the lower half of Figure 4 represents the scene to be detected. The upper part of Figure 4 is a multi-scale detection network with a pyramid structure. There are two targets A and B in the scene. The size of target A is larger than that of target B. Suppose that level 1 of detection network is responsible for detecting target A, and level 3 of detection network is responsible for detecting target B. The location of target A is a positive sample for level 1, but it may be a negative sample for levels 2 and 3. Therefore, in the process of back propagation, the gradient information about the location of target A may be both positive and negative samples. The inconsistency of gradient information at the same location in the scene on different scale networks will interfere with gradient calculation.

Liu *et al*. [10] proposed an Adaptively Spatial Feature Fusion algorithm. The ASFF solves the interference of multi-scale targets on gradient calculation by making the network corresponding to each level learn how to filter out the conflict information of other levels in space adaptively. The ASFF module can eliminate the interference information generated by negative samples in the process of gradient back propagation.

As shown in the three green squares in Figure 2, three ASFF modules are added to the three scale detection networks, namely ASFF-1, ASFF-2 and ASFF-3. ASFF-3 module is at the third level of detection network. In Formula 1, the ASFF-3 module adjusts the feature map of the other two scales to the same scale as the feature map of the third level, and splices them with the feature map of the third level. The dotted line in Figure 2 indicates that the ASFF module of each level accepts feature map from other levels.

$$\mathrm{ASFF}-3 = X^{1\to 3}\otimes\alpha^3 + X^{2\to 3}\otimes\beta^3 + X^{3\to 3}\otimes\gamma^3 \qquad (1)$$

In Equation (1), X^1, X^2 and X^3 represent the feature map from level 1, level 2 and level 3 respectively. $X^{1\to 3}$ means to adjust the size of the feature map of level 1 to be the same as level 3, $X^{2\to 3}$ meansto adjust the size of the feature map of level 2 to be the same as level 3, and $X^{3\to 3}$ means to adjust the size of the feature map of level 3 to be the same as level 3, that is, the size of the feature map of level 3 remains unchanged. In the process of feature fusion, first multiply the adjusted features by their corresponding weight parameters α^3, β^3, and γ^3 and then add the multiplied results to obtain new fusion features. The new fusion features are the output feature map of ASFF-3 network module. Weight parameter α^3, β^3 and γ^3 are parameters that can be learned through back propagation.

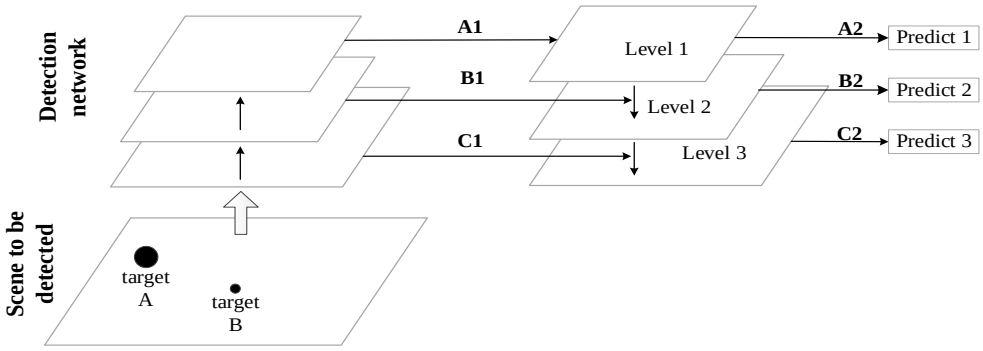


Figure 4. Impact of multi-scale target on detection network.

4. Experiment

There are two most popular traffic sign databases: German Traffic Sign Detection Benchmark (GTSDB) and Tsinghua-Tencent 100K (TT100K) traffic sign detection database. The TT100K database contains natural weather factors (such as fog, cloudy and rainy weather conditions), partial occlusion and significant changes in illumination and perspective. These characteristics make TT100K closer to the real scene than GTSDB. TT100K traffic sign data set is divided into training set and test set. The training set contains 6105 images and the test set contains 3065 images. The dataset is divided into three categories, named indicative, prohibitive and warning. In TT100K traffic sign data set, the resolution of the image is 2048×2048. Implementation Details: SPP&ASFF-Yolov3 is trained using the Stochastic Gradient Descent (SGD) optimizer, with batch size of 32 and learning rate 0.0001 for 300 epochs. The operating system version used in this paper is Ubuntu 18.04.6, CUDA 10.1, and the model of GPU is NVIDIA Tesla T4×4.

4.1. Traffic Sign Detection on TT100K Datasets

In this experiment, the performance of SPP and ASFF-Yolov3 is evaluated by Average Precision (AP) and mean Average Precision (mAP). AP is the average precision value of each category, which can be calculated by 2. The mAP is the average of the average precision values of all categories, which can be calculated by 3.

$$AP = \int_0^1 P(R)dR \tag{2}$$

$$mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q} \tag{3}$$

$$P = \frac{TP}{TP + FP} = \frac{TP}{N} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

Where TP is true positives, FP is false positives, and R is recall. ASFF&SPP-YOLOv3, SPP-YOLOv3, original YOLOv3, traffic sign detection network based on Faster R-CNN and traffic sign detection network based on RetinaNet on TT100K dataset.

From the results in Table 2, we can draw the following conclusions:

Table 2. Performance of different detection algorithms on TT100K dataset.

Method	indicative	prohibitory	warning	mAP
Faster R-CNN	64.8	66.8	60.4	64.0
RetinaNet	63.3	64.7	67.5	65.2
YOLOv3	77.3	68.0	77.4	74.2
SPP-YOLOv3	78.9	68.9	78.6	75.5
Cascaded network	**80.2**	69.2	79.9	76.4
ASFF&SPP-YOLOv3	79.8	**69.9**	**80.1**	**76.6**

1. The proposed ASFF and SPP-YOLOv3 achieves the highest AP value and the highest mAP value in three traffic sign categories of TT100K dataset, which shows the effectiveness of the proposed method.
2. The AP value of SPP-YOLOv3 in three traffic sign categories of TT100K is about 1% higher than that of YOLOv3. The mAP value of SPP-YOLOv3 is 1.3% higher than that of YOLOv3. This shows the effectiveness of SPP module in the network. The AP and mAP values of ASFF and SPP-YOLOv3 are similar to those of Cascade network.
3. The AP value and mAP value of the network with ASFF module (ASFF and SPP-YOLOv3) are higher than SPP-YOLOv3. This shows the effectiveness of ASFF module in the network.

4.2. Inference Time

As shown in Table 3, YOLOv3 has the fastest reasoning speed. Due to the addition of SPP and ASFF modules, the number of parameters in the network increases and the inference speed decreases. However, the inference speed of the proposed ASFF and SPP-YOLOv3 can reach 30 FPS, which still meets the real-time requirements. Although the mAP value of cascaded network is similar to our ASFF and SPP-YOLOv3, its inference speed is slower than our network in the table.

Table 3. Number of parameters and inference time of different detection algorithms.

Method	parameter	FPS
Faster R-CNN	—	7.4
RetinaNet	—	22.1
YOLOv3	61534432	**50.8**
SPP-YOLOv3	63896288	32.3
Cascaded network	84987968	28
ASFF&SPP-YOLOv3	72473592	30

4.3. Pictures of traffic sign detection results

Detection results of ASFF and SPP-YOLOv3 are shown in Figure 5. As shown in Figure 5, the traffic signs in the scene account for a small proportion of the whole scene. The two traffic signs in Figure 5-a) are of the same size, but the traffic signs Figure 5-c) are of different sizes. The traffic signs in Figure 5-b) are disturbed by deformation and uneven illumination. Although there are many challenges for the successful detection of traffic signs in the scene shown in Figure 5, the proposed ASFF and SPP-YOLOv3 algorithm accurately detects all traffic signs in the scene.

a) Traffic signs of the same size.

b) Deformation and uneven illumination.

c) Traffic signs of different sizes.

Figure 5. Detection results of ASFF&SPP-YOLOv3 on TT100K dataset.

5. Conclusions

This paper presents the Spatial Pyramid Pooling and Adaptively Spatial Feature Fusion based Yolov3 (SPP and ASFF-Yolov3) network and applies it to traffic sign detection. We conducted experiments on TT100K dataset and compared the proposed SPP&ASFF-Yolov3 with the state of the art methods, which shows the proposed network can archive better mAP than Faster R-CNN, RetinaNet and the original YOLOv3 network, with 12.6%, 11.4% and 2.4 increase respectively. Despite the decrease of inference speed compared to YOLOv3 network, the inference speed is still reasonable and much faster than candidate region-based methods (1.5-3 times faster), and our proposed network can meet the real-time performance of traffic sign detection.

Acknowledgment

This work was supported by the Science and Technology Development Plan Project of Jilin Province.

References

[1] Avramović A., Sluga D., Tabernik D., Skočaj D., Stojnić V., and Ilc N., "Neural-Network-Based Traffic Sign Detection and Recognition in High-Definition Images Using Region Focusing and Parallelization," *IEEE Access*, vol. 8, pp. 189855-189868, 2020. DOI: 10.1109/ACCESS.2020.3031191

[2] Dhia H. and Ghani R., "A Proposed Method for Scale Drawing Calculating Depending on Line Detector and Length Detector," *Iraqi Journal for Computer Science and Mathematics*, vol. 2, no. 2, pp. 6-17, 2021. DOI: https://doi.org/10.52866/ijcsm.2021.02.02.002

[3] Dinh V., Lee Y., Choi H., and Jeon M., "Real-Time Traffic Sign Recognition," *in Proceedings of IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, JeJu, pp. 206-212, 2018. DOI: 10.1109/ICCE-ASIA.2018.8552110

[4] Greenhalgh J. and Mirmehdi M., "Real-Time Detection and Recognition of Road Traffic Signs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1498-1506, 2012. DOI: 10.1109/TITS.2012.2208909

[5] Han C., Gao G., and Zhang Y., "Real-Time Small Traffic Sign Detection with Revised Faster-RCNN," *Multimedia Tools and Applications*, vol. 78, no. 4, pp. 13263-13278, 2019. DOI:10.1007/s11042-018-6428-0

[6] He K., Zhang X., Ren S., and Sun J., "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904-1916, 2015. DOI: 10.1109/TPAMI.2015.2389824

[7] Karimi M., Khandaghi Z., and Shahipour M., "Designing an Intelligent System to Detect Stress Levels During Driving," *The International Arab Journal of Information Technology*, vol. 19, no. 1, pp. 81-89, 2022. https://doi.org/10.34028/iajit/19/1/10

[8] Liang M., Yuan M., Hu X., Li J., and Liu H., "Traffic Sign Detection by ROI Extraction and Histogram Features-Based Recognition," *The International Joint Conference on Neural Networks*, Dallas, pp. 1-8, 2013. DOI: 10.1109/IJCNN.2013.6706810

[9] Liu L., Wang Y., Li K., and Li J., "Focus First:

Coarse-To-Fine Traffic Sign Detection with Stepwise Learning," *IEEE Access*, vol. 8, pp. 171170-171183, 2020. 10.1109/ACCESS.2020.3024583

[10] Liu S., Huang D., and Wang Y., "Learning Spatial Fusion for Single-Shot Object Detection," *arXiv preprint arXiv:1911.09516*, 2019. https://doi.org/10.48550/arXiv.1911.09516

[11] Nguyen H., "Fast Traffic Sign Detection Approach Based on Lightweight Network and Multilayer Proposal Network," *Journal of Sensors*, vol. 2020, pp. 1-1, 2020. DOI:10.1155/2020/8844348

[12] Redmon J. and Farhadi A., "YOLO9000: Better, Faster, Stronger," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, pp. 7263-7271, 2017. DOI:10.1109/CVPR.2017.690

[13] Redmon J. and Farhadi A., "Yolov3: An Incremental Improvement," *arXiv Preprint arXiv:1804.02767*, 2018. http://arxiv.org/abs/1804.02767

[14] Shan H. and Zhu W., "A Small Traffic Sign Detection Algorithm Based on Modified Ssd," *IOP Conference Series: Materials Science and Engineering*, vol. 646, no. 1, pp. 012006, 2019. DOI 10.1088/1757-899X/646/1/012006

[15] Shufang Z., Qinyu W, Tong Z., and Yuhong L., "Detection and Classification of Small Traffic Signs Based on Cascade Network," *Chinese Journal of Electronics*, vol. 30, no. 4, pp. 719-726, 2021. https://doi.org/10.1049/cje.2021.05.014

[16] Sichkar V. and Kolyubin S., "Real Time Detection and Classification of Traffic Signs Based on Yolo Version 3 Algorithm," *Journal Scientific and Technical of Information Technologies, Mechanics and Optics*, vol. 20, no. 3, pp. 418-424, 2020. 10.17586/2226-1494-2020-20-3-418-424

[17] Sun C., Ai Y., Wang S., and Zhang W., "Dense-Refinedet for Traffic Sign Detection and Classification," *Sensors*, vol. 20, no. 22, pp. 6570, 2020. https://doi.org/10.3390/s20226570

[18] Tian Z., Shen C., Chen H., and He T., "Fcos: Fully Convolutional One-Stage Object Detection," *in Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9627-9636, 2019.

[19] William M., Zaki P., Soliman B., Alexsan K., Mansour M., El-Moursy M., and Khalil K., "Traffic Signs Detection and Recognition System Using Deep Learning," *in Proceedings of 9th International Conference on Intelligent Computing and Information Systems*, Cairo, pp. 160-166, 2019. DOI: 10.1109/ICICIS46948.2019.9014763

[20] Xiong S., Li B., and Zhu S., "DCGNN: A Single-Stage 3D Object Detection Network Based on Density Clustering and Graph Neural Network," *Complex and Intelligent Systems*, pp. 1-10, 2022. DOI:10.1007/s40747-022-00926-z

[21] Yang Y., Luo H., Xu H., and Wu F., "Towards Real-Time Traffic Sign Detection and Classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 7, pp. 2022-2031, 2015. DOI: 10.1109/TITS.2015.2482461

[22] Zhang H., Qin L., Li J., Guo Y., Zhou Y., Zhang J., and Xu Z., "Real-Time Detection Method for Small Traffic Signs Based on Yolov3," *IEEE Access*, vol. 8, pp. 64145-64156, 2020. DOI: 10.1109/ACCESS.2020.2984554

[23] Zhang J., Huang M., Jin X., and Li X., "A Real-Time Chinese Traffic Sign Detection Algorithm

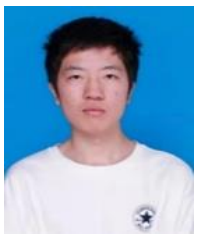
[24] Based on Modified Yolov2," *Algorithms*, vol. 10, no. 4, p. 127, 2017. DOI:10.3390/a10040127

[25] Zhou X., Zhuo J., and Krahenbuhl P., "Bottom-Up Object Detection by Grouping Extreme and Center Points," *in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 850-859, 2019.

[26] Zhu Y., Liao M., Yang M., and Liu W., "Cascaded Segmentation-Detection Networks for Text-Based Traffic Sign Detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 209-219, 2017. DOI:10.1109/TITS.2017.2768827

Shimin Xiong received his bachelor's degree from the Northeast Electric Power University in 2020. He is currently a graduate student with the School of Computer Science at Northeast Electric Power University. His research interests include pattern recognition, image processing and deep learning.

Bin Li received his M.S. and PhD degrees from the School of Computer Science and Technology at Jilin University in China in 2011 and 2015, respectively. He is currently an Associate Professor with the School of Computer Science at Northeast Electric Power University. His research interests include image processing, computer vision, and pattern recognition.



Shiao Zhu is now an undergraduate in Northeast Electric Power University. His research interests are image processing, deep neural networks.

Dongfei Cui is now an undergraduate in Northeast Electric Power University. His research interests are image processing, deep neural networks.

Xiaonan Song received her bachelor's degree from the Anyang Institute of Technology 2017. She is currently a graduate student with the School of Computer Science at Northeast Electric Power University. Her research interests include computer vision, image processing.