ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0062

Improved Traffic Sign Detection Algorithm in Complex Environment

Qiyuan Xiao

School of Shanghai Maritime University, Shanghai 200000, China.

2637515159@qq.com

Abstract

The recognition of traffic signs plays an important role in the safe driving of vehicles, especially in the presence of light changes and occlusions, the automatic recognition of traffic signs with high accuracy and good real-time performance needs to be solved urgently. Among the deep learning algorithms, YOLOV3 and Faster-RCNN have achieved excellent target detection performance. Compared with traditional feature detection and recognition methods, the use of YOLOV3 algorithm helps to solve the problem of small target detection, traffic sign recognition under the conditions of light changes, partial occlusion, etc. It is the main way to improve the performance of automatic driving and unmanned driving in the future.

Keywords

Traffic Sign Detection; Traffic Sign Recognition; YOLOV3; Faster-RCNN; Small Target Detection; Automatic Driving.

1. Introduction

In the field of intelligent transportation [1] traffic sign recognition aims to provide vehicles with accurate traffic sign information, and then provide services for applications such as autonomous driving [2]. For autonomous vehicles, it is necessary to accurately and timely identify traffic signs to ensure that they comply with road laws. Now the research on traffic sign recognition [3] has made good progress, but there are many problems in related algorithms. Traditional detection algorithms achieve detection by extracting target features. These features cannot obtain satisfactory results in terms of speed and accuracy. And with the sustained and rapid development of the social economy, the number of vehicles per capita has risen rapidly, and the defects of the traditional transportation industry have become increasingly prominent, including frequent traffic accidents, road traffic congestion, and serious environmental pollution. Therefore, the intelligentization of vehicles is imminent. The primary task of vehicle intelligence is to let the vehicle understand the surrounding scene and know the "what" of the surrounding target, that is, to realize the target detection in the road environment.

Since the advent of AlexNet [4] in 2012, deep convolutional neural networks have made important achievements in tasks such as image processing and natural language processing. In order to enhance the nonlinear fitting ability of the convolutional neural network, the network is designed to be wider and deeper, requiring a lot of storage space and computing resources, which makes CNN not suitable for most hardware devices with limited computing power (such as embedded devices). Applicable. Therefore, it is very important to compress large-scale networks while ensuring minimal loss of accuracy, while obtaining a lightweight model that occupies storage, forward inference calculations, and energy consumption. The main methods include fine model design, pruning, and low Rank decomposition, quantification and knowledge distillation, etc.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0062

Since the composition of CNN contains artificially set training parameter modules such as convolutional layer and fully connected layer, this local optimal hyperparameter tuned out by repeated experiments through experience cannot represent the "real needs" of the network. According to the new double U-shaped deviation-variance risk curve of deep neural networks: when the network parameters are not enough, as the complexity increases, the risk becomes a U-shaped curve; when the parameters are redundant, the more complex the model, the lower the risk. Regardless of resource constraints, networks with more parameters and more complex models perform better. This is exactly the development status of CNN. However, the complex model over parameters does not weigh the relationship between cost and performance, and the model has redundancy. Pruning just hopes to obtain a small parameterized lightweight model at the first U-shaped minimum point of the double U-shaped curve from the over-parameterized complex model by pruning the redundant modules in the network.

In recent years, algorithms based on deep learning [5] have made major breakthroughs in computer vision. These algorithms can be divided into two categories: region-based detection methods (such as R-CNN[6], Fast R-CNN[7], Faster R-CNN[8], etc.) and regression-based detection methods (such as YOLO[9], YOLOV2[10], YOLOV3[11], SSD[12], etc.). The target detection algorithm through deep learning [13] is significantly better than traditional detection methods. But for traffic sign detection [14], there is still a lot of room for improvement in the accuracy of deep learning algorithms.

2. Development status and research status at home and abroad

The most researched topic in the field of image processing is image recognition and segmentation. So far they are still the focus of many researchers, thinking that the existing research results cannot solve many existing problems.

Image recognition has gone through three stages of development, namely: text recognition [15], digital image processing and recognition [16] and object recognition [17]. Text recognition first began in 1950. It recognizes letters, numbers and symbols. From printed text recognition to handwritten text recognition are all within its capabilities. Object recognition mainly refers to the perception and cognition of objects and the environment in the three-dimensional world, which is the category of advanced computer vision [18]. It is used in various industries and robotics fields. Poor adaptability is one of the shortcomings of current image recognition technology. Because the image is contaminated by strong noise [19] or because the image itself has large defects, we usually cannot obtain the desired result.

The mathematical nature of the image recognition problem belongs to the mapping problem from the pattern space to the category space [20]. There are currently three main image recognition methods, namely: statistical pattern recognition [21], structural pattern recognition [22] and fuzzy pattern recognition [23]. Image segmentation [24] is a key technology in image processing. Thousands of theories have been proposed so far.

In addition, there are many image segmentation methods, including threshold segmentation method [25], edge detection method [26], region extraction method [27] and so on. As early as 1965, someone proposed an edge detection operator [28], which resulted in many classic edge detection algorithms. Image segmentation methods usually need to combine some specific theories, methods and tools, such as image segmentation based on mathematical morphology, image segmentation based on wavelet transform [29], image segmentation based on genetic algorithm [30], etc.

2.1 Current status of yolo algorithm research

YOLO is the abbreviation of You Only Look Once, which means that the neural network only needs to look at the picture once to output the result. YOLO released a total of four versions, of which YOLOv1 laid the foundation for the entire series, and the following series are improvements based on the first version, just to improve performance.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0062

YOLOv1:

Problem background: The analogy two-stage method such as R-CNN divides the detection problem into two parts, first generates the candidate region (region proposal), and then uses the classifier to classify the region, resulting in multi-stage training and difficult optimization.

Innovation point: Regarding detection as a regression problem, using a network output location and category to implement a unified system, which is one-stage from the perspective of detection.

YOLOv2

Problem background: YOLOv1 detection performance is low

The current detection task is limited by the label of the data set (the data set must have a label or be assigned a label through classification). However, the label detection image is much more expensive than the label classification image, so the detection data and the classification data are not on the same scale.

Innovation point: For the first problem, use some methods to improve the performance of YOLOv1 and get YOLOv2. In response to the second question, a combination method of ImageNet and COCO data sets and a joint training method are proposed. The model obtained after training YOLOv2 is called YOLO9000.

YOLOv3

Problem background: YOLOv3 was not proposed to solve any problem, the whole paper is actually a technical report. YOLOv3 has made some small improvements on the basis of YOLOv2. The article is not long and the core idea is similar to YOLOv2 and YOLO9000.

Model improvements: 1. Bounding box prediction: The positioning task uses the anchor box method to predict the bounding box. YOLOv3 uses logistic regression to predict an objectness score for each bounding box. The score is based on the overlap between the predicted box and the object. If the overlap of a certain box is higher than other boxes, its score is 1, ignore those boxes that are not the best and have an overlap greater than a certain threshold (0.5). 2. Category prediction: Like YOLOv2, YOLOv3 still adopts multi-label classification. 3. Multi-scale prediction. 4. Use the new network Darknet-53 to extract features.

3. Research ideas, methods and technical routes

3.1 Ideas

This paper improves the YOLOV3 network structure and proposes a traffic sign detection method based on deep fusion of multi-scale feature maps, which can improve the recognition accuracy of small targets. Then the kmeans clustering algorithm is replaced with k-means++ algorithm, and the darknet53 network is replaced with resnet101 with stronger feature extraction ability to improve its detection accuracy in complex environments. The introduction of GIOULoss can solve the structure that YOLOV3 original IOU cannot directly optimize. Then after pruning, it becomes a lightweight network model, and its calculation is improved to increase the speed

3.2 Method

This method is a multi-target detection algorithm that directly detects the location and category of urban road traffic signs. In view of the problem of small target missed detection in YOLOV3 algorithm, the feature map output by YOLOV3 is down-sampled 32 times, 16 times, 8 times, and 4 times, respectively, and a feature pyramid of 4 scale convolutional layers can be obtained [31], namely 13 ×13, 26 × 26, 52 × 52, 104 × 104, and then double upsampling, the convolutional layer of the network is multi-scale and deep fusion, which effectively solves the problem of too small traffic sign targets. Then use the K-means++ clustering algorithm [32] to optimize the K-means clustering algorithm, and replace darknet53 with resnet101 with stronger feature extraction ability. This step can effectively solve the difficulty of target detection in a complex road environment. Then introduce GIOULoss, which can solve the structure that YOLOV3 original IOU cannot directly optimize. After

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0062

detecting and designing a new network structure, it reduces the amount of parameters and enhances the reuse of features and non-overlapping parts. Finally, perform sparse training on the trained model. After training, obtain a network model with a scaling factor in the BN layer. According to the scaling ratio, reduce the number of convolution kernel network channels to complete the model compression, and then fine-tune the pruned model. A lightweight YOLOV3 network model with small parameters, small calculations and constant accuracy is obtained.

3.3 Technical route

3.3.1. Multi-scale feature fusion network algorithm design

In order to improve the performance of YOLOV3 algorithm for small target detection, the improved YOLOV3 uses 4 scales to detect targets of different sizes, and uses 8 times down-sampled feature maps to detect small targets. In order to obtain more fine-grained feature and location information from small targets, we explore the feature map that uses 4 times downsampling in the original network because it contains more fine-grained small target information. In order to connect the shallow featuremap to the deep featuremap, the featuremap is upsampled and then downsampled, and connected to the output of the second residual block in YOLOV3. In this way, a down-sampled feature fusion target detection layer is established [33], which is used to detect small targets. The 3 detection scales in YOLOV3 have been extended to 4 detection scales. Therefore, the improved YOLOV3 has better detection performance for targets of different sizes.

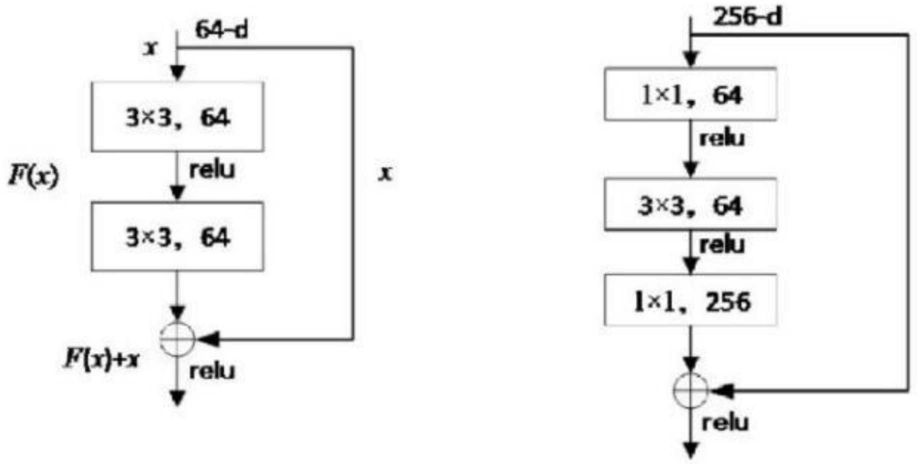
3.3.2. Improved GIOU Loss is integrated into YOLOV3

The disadvantage of using IOU for loss is that when the two target frames do not overlap, the IOU value is 0. If you directly use IOU=0 as the loss, the gradient will not be returned, so there is nothing for training. Significantly, in actual situations, this 0 is very important. It proves that the distance between the two boxes is relatively far. In order to solve the shortcomings of IOU, this paper proposes the method of using GIOU. The specific process of the algorithm is as follows: For the two boxes A and B, first find the smallest bounding box to cover the A and B areas; secondly, calculate the smallest box C to remove A , The ratio of the area outside B to the total area in C, and finally the value of GIOU is IOU minus this ratio. GIOU is the algorithm of two frame measurement criteria.

3.3.3. K-means++ clustering algorithm a priori box calculation method

YOLOV3 uses the K-means clustering algorithm for the data set, which can efficiently and automatically find the appropriate initial a priori box instead of manually selecting the a priori box. In this paper, the samples generated by the K-means++ clustering algorithm can reflect the distribution of the samples in each data set, making it easier for the network to obtain better prediction results. K-means++ clustering algorithm is used on the Tsinghua-Tencent100k traffic sign dataset. Twelve a priori boxes were selected for comprehensive consideration.

3.3.4. YOLOv3 based on residual network feature extraction



ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0062

Deepening the convolutional neural network can enhance the feature extraction ability, but at the same time the phenomenon of gradient disappearance will be more obvious, and the training effect of the network will decrease instead, and the residual network allows the network to deepen as much as possible, and its structure is shown in the figure. Therefore, this article uses resnet101[34] to replace the original darknet53 to improve the detection performance of yolov3. In addition to the normal output of the convolutional layer, the residual module has a branch that connects the input directly to the output. The output and the output of the convolution are arithmetic added to get the final output, as shown in Figure 4, x is as shown The input of the structure, F(x) is the output of the convolution branch, and H(x) is the output of the entire structure. The residual structure artificially creates the identity map, which makes the entire structure converge in the direction of the identity map, ensuring that the final error rate will not get worse due to the increasing depth.

3.3.5. YOLO network integrated pruning algorithm steps based on sensitivity

By sparsening the layers with more convolution kernels and then performing integrated pruning, you can avoid the problem that the integrated pruning method cannot handle the redundant layers and cannot distinguish the importance of the convolution kernels. The generalization ability of pruning is better, and it also has Conducive to obtaining a uniform model and higher pruning accuracy. The algorithm flow is as follows:

Input: VOC data set, cfg configuration file and initialization weight of YOLO original network, expected pruning ratio K%

Output: pruned YOLO model cfg configuration file and corresponding weight

1. Pre-train the YOLO model.

2. Determine the pruning layer, and use the sparse scaling factor method for sparse training for the redundant pruning layer of the YOLO model.

3. Use the integrated pruning algorithm to sort the convolution kernels of each pruning layer by importance using three convolution kernel significance judgment criteria, and add the three sorts of each convolution kernel to obtain the convolution kernel Importance score.

4. Test the original accuracy on the verification set, cut out the K% convolution kernel with low importance of each pruning layer, test the new accuracy on the verification set, subtract the new accuracy from the original accuracy, and divide by the cut The number of convolution kernels is the contribution of each convolution kernel that is cut off to the accuracy.

5. Solve the system of equations and get the number of convolution kernels that should be pruned for each pruning layer.

6. For each pruning layer in the model, the number of convolution kernels should be reduced layer by layer to obtain a simplified model.

7. Fine-tuning the simplified model to restore the model accuracy.

8. If the number of convolution kernel pruning does not reach the expected value, return 4.

4. Innovation

The data collected on the Internet has complex road environments, smaller targets for traffic signs, occlusion, light and other factors that will affect the results of the experiment. The target detection accuracy obtained at the end of the experiment needs to be improved.

First, use the traffic sign detection method of deep fusion of multi-scale feature maps, and then use the sensitivity integrated pruning algorithm. Compared with the model obtained by the network slimming method, the model obtained has fewer parameters, smaller model, less calculation, and speed Faster and higher accuracy. Then optimize the k-means clustering method, and replace the original darknet53 with resnet101 with stronger feature extraction ability, and then introduce GIOULoss to solve the structure that YOLOV3's original IOU cannot directly optimize. Finally, what

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0062

we got was a model with higher accuracy for small target detection, smaller size, faster speed, and better portability.

References

[1] Wang Strontium, Zhou Yan. Investigation on the status quo of China's intelligent transportation[J]. Transportation Enterprise Management, 2017, 32(04): 12-14.

[2] A. Mogelmose, M. M. Trivedi and T. B. Moeslund, "Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey," in IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 4, pp. 1484-1497, Dec. 2012.

[3] Wang Gangyi. Research on Traffic Sign Detection and Classification Algorithm [D]. Harbin: Harbin Institute of Technology, 2013.

[4] N. Dwivedi, D. K. Singh and D. S. Kushwaha, "Weapon Classification using Deep Convolutional Neural Network," 2019 IEEE Conference on Information and Communication Technology, Allahabad, India, 2019, pp. 1-5.

[5] LECUN Y,BINGO Y,HINTON G.Deep learning[J].Nature,2015,521:436-444.

[6] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 580-587.

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

[8] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017.

[9] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788.

[10] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6517-6525.

[11] YOLOv3:An Incremental Improvement[C/OL]. REDMON J,FARHADI A. https://arxiv. org/pdf/ 1804. 02767.pdf. 2018.

[12] Liu W. et al. (2016) SSD: Single Shot MultiBox Detector. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016.

[13] Object Detection in20 Years:A Survey. ZOU Z X,SHI Z W,GUO Y H, et al. https://arxiv. org/pdf/ 1905. 05055. pdf . 2019.

[14] He Kaihua. Traffic sign recognition based on target detection network[J]. Software Engineering, 2020, 23(10): 19-22.

[15] X. Ding, L. Chen and T. Wu, "Character Independent Font Recognition on a Single Chinese Character," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 2, pp. 195-204, Feb. 2007.

[16] Liu Xinghan, Xu Bin. Research on digital image recognition technology in the era of big data [J]. Digital World, 2020(10): 81-82.

[17] Qu Liang, Xu Xing, Wang Wei. Research on Multi-Pose Object Recognition Technology Based on Key Points of Edge AI Mobile Video[J]. Information Technology and Informatization, 2019(12):205-207.

[18] Jiang Xingyu. Research on assisted driving algorithm based on computer vision [D]. Qinghai Normal University, 2020.

[19] Gao Tao, He Xiuli. Color image Poisson denoising based on wavelet frame [J]. Journal of Nanjing University of Information Technology (Natural Science Edition), 2020, 12(05):614-618.

[20] Xiao Chengying, Yang Mingge. Non-emptiness and lower semicontinuity of effective solution mapping for parameter vector optimization problems in Banach space[J]. Journal of Southwest University (Natural Science Edition), 2020, 42(09): 95-100 .

[21] Webb A. Statistical pattern recognition[M]. Publishing House of Electronics Industry, 2004.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0062

[22] Tan Jianrong, Yue Xiaoli, Lu Guodong. Basic principles and methods of graphic similarity and its application in structural pattern recognition [J]. Chinese Journal of Computers, 2002(09):959-967.

[23] Chen Shouyu. Fuzzy pattern recognition cross-iterative model and convergence [J]. Journal of Dalian University of Technology, 2001, 41(003):264-267.

[24] Y. Li, J. Zhang, P. Gao, L. Jiang and M. Chen, "Grab Cut Image Segmentation Based on Image Region," 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, 2018, pp. 311-315.

[25] Long Jianwu, Shen Xuanjing, Chen Haipeng. Adaptive minimum error threshold segmentation algorithm[J]. Acta Automatica Sinica, 2012(07):1134-1144.

[26] Wei Weibo, Rui Xiaoting. Research on image edge detection method[J]. Computer Engineering and Applications, 2006, 42(030): 88-91.

[27] F. Xu, S. Chen, Y. Fan and C. Wang, "Region Feature Extraction Based on Improved Regularization Method in SAR Image," IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, 2008, pp. III - 1147-III - 1149.

[28] J. Lu, H. Pan and Y. Xia, "The weld image edge-detection algorithm combined with Canny operator and mathematical morphology," Proceedings of the 32nd Chinese Control Conference, Xi'an, 2013, pp. 4467-4470.

[29] M. Polyakova, V. Krylov and N. Volkova, "The methods of image segmentation based on distributions and wavelet transform," 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, 2016, pp. 243-247.

[30] Gao Weizeng, Zhang Baojian, Chen Fugui, et al. Cutting path optimization based on genetic algorithm[J]. Journal of Southwest Jiaotong University, 2005, 40(4).

[31] X. Gao et al., "An End-to-End Neural Network for Road Extraction From Remote Sensing Imagery by Multiple Feature Pyramid Network," in IEEE Access, vol. 6, pp. 39401-39414, 2018.

[32] Cui Dandan. Research and improvement of K-Means clustering algorithm [D]. Anhui University.

[33] Shu Xinyin, Wang Ping. Smart car application detection segmentation technology based on dynamic feature fusion [J]. Computer Engineering and Design, 2020, 41(10): 2838-2842.

[34] Sai Sundar K V, Bonta L R, Kumar R B A, et al. Evaluating Training Time of Inception-v3 and Resnet-50,101 Models using TensorFlow across CPU and GPU[J]. 2018:1964-1968.