Proceedings of 2018 IEEE
International Conference on Mechatronics and Automation
August 5 - 8, Changchun, China

Pedestrian Detection Based on YOLO Network Model

Wenbo Lan, Jianwu Dang, Yangping Wang and Song Wang

Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphics & Image Processing

LANZHOU JIAOTONG UNIVERSITY

Lanzhou,Gansu Province,China

hilanwenbo@163.com

***Abstract* - After going through the deep network, there will be some loss of pedestrian information, which will cause the disappearance of gradients, causing inaccurate pedestrian detection.This paper improves the network structure of YOLO algorithm and proposes a new network structure YOLO-R.First, three Passthrough layers were added to the original YOLO network. The Passthrough layer consists of the Route layer and the Reorg layer. Its role is to connect the shallow layer pedestrian features to the deep layer pedestrian features and link the high and low resolution pedestrian features.The role of the Route layer is to pass the pedestrian characteristic information of the specified layer to the current layer, and then use the Reorg layer to reorganize the feature map so that the currently-introduced Route layer feature can be matched with the feature map of the next layer.The three Passthrough layers added in this algorithm can well transfer the network's shallow pedestrian fine-grained features to the deep network, enabling the network to better learn shallow pedestrian feature information.This paper also changes the layer number of the Passthrough layer connection in the original YOLO algorithm from Layer 16 to Layer 12 to increase the ability of the network to extract the information of the shallow pedestrian features.The improvement was tested on the INRIA pedestrian dataset. The experimental results show that this method can effectively improve the detection accuracy of pedestrians, while reducing the false detection rate and the missed detection rate,and the detection speed can reach 25 frames per second.**

Index Terms - YOLO Network Model; Pedestrian Detection; Passthrough Layer; Route Layer; Reorg Layer.

I. INTRODUCTION

Pedestrian detection is the basis of pedestrian target information identification and pedestrian behavior analysis. Pedestrian detection is also the basis for later identification and processing.At present,the most classic pedestrian detection method is the HOG+SVM pedestrian detection proposed by Dalal [1] et al,and has almost perfect performance on the MIT pedestrian data set.Felzenszwalb et al,proposes an improved DPM algorithm [2], DPM algorithm is a component-based detection method and has a strong robustness to the deformation of the target. Dollar et al. proposed the integral channel features [3] and the aggregated channel features [4] to integrate gradient histogram, LUV, and gradient magnitude features to obtain better performance of pedestrian features.

With the development of the deep neural network, the deep learning model has also been widely used in pedestrian detection.Deep learning has great advantages over traditional target detection.The traditional method is to manually extract features and require experts in related fields to manually design and process them through years of accumulation and experience.The method of deep learning can learn the features of difference in response data through a large amount of data and is more representative.The deep learning model simulates the human brain's visual perception system.It extracts features directly from the original image, and the features are passed through the layer by layer to obtain the high-dimensional information of the image, making it a great success in the field of computer vision.Ouyang [5] and others proposed a component-based detection method that is robust to the deformation of the target.Tian[6] and others proposed using deep learning combined with Part Model to obtain a Deep Parts to solve the occlusion problem in pedestrian detection.Anelia [7] and others proposed a new real-time pedestrian detection method, which uses the accuracy of deep neural networks and the efficiency of cascaded classifiers for pedestrian detection to achieve high accuracy.The current excellent target detection models include:R-CNN [8],SPP-Net [9],Fast-RCNN [10],Faster-RCNN [11],SSD [12], YOLO [13] and ResNet [14].

The YOLO network model excels in real-time detection of targets. Therefore, the YOLO network model will be used for pedestrian detection and the network will be improved to improve its detection capabilities. The main improvement of this paper is to add 3 passthrough layers in front of the deep network, and change the number of Passthrough layer connections in the original YOLO network structure from Layer 16 to Layer 12 to form a new YOLO-R network model. In this experiment, the INRIA pedestrian detection data set was used to test. Experiments show that the improved algorithm improves the accuracy of pedestrian detection while reducing the rate of missed detection and false detection and the detection speed reaches 25 frames/s.

II. PEDESTRIAN DETECTION PROCESS

The YOLO network model uses the target detection as a regression problem for a spatially separate target box and its category confidence.A single neural network can directly predict the confidence of the target box and the category from the entire image.The pedestrian detection process of the YOLO network model is shown in Fig. 1:

978-1-5386-6075-1/18/$31.00 ©2018 IEEE

Fig. 1 YOLO network pedestrian detection process

1) First divide the image into a grid of SxS.If the pedestrian is in a grid, the grid is responsible for detecting the pedestrian.Each grid predicts B detection boxes and the confidence of these detection boxes,The number of detection frames for each picture is SxSxB.
2) Each detection box has 5 predicted values (X, Y, W, H, Conf).Where X,Y is the offset of the center of the prediction box relative to the cell boundary, and W and H are the ratios of the predicted box width to the entire image, and Conf represents the confidence of the detection box.
3) Each grid predicts the pedestrian's conditional probability Pr(class|object), provided that the known grid contains pedestrians.
4) At the time of detection,the conditional probability is multiplied by the predictive value of different detection box confidences to obtain the confidence score for each detection box pedestrian category.These pedestrian types also include the probability of pedestrians appearing in the detection frame and check the match between the box and the pedestrian goal.The confidence score formula as in (1):

$$Conf(class) = \Pr(class) * IOU_{\Pr ed}^{Truth} . \quad (1)$$

In formula (1): $\Pr(class)$ represents the probability of pedestrians appearing in the grid, and $IOU_{\Pr ed}^{Truth}$ indicates the overlapping ratio of the area between the predict box and the ground truth box.Pred indicates the area of the prediction box, and Truth indicates the area of the ground truth box.The larger the IOU,the higher the accuracy of pedestrian detection.The final output vector of each picture through the network is SxSxBx[X,Y,W,H,Conf,Conf(class)].

III. NETWORK STRUCTURE

In this paper, the YOLO v2 [14] network is used as the prototype. Three Passthrough layers have been added to the original YOLO v2 network, and the number of layers connected to the Route layer in the Passthrough layer in the original YOLO v2 algorithm has been improved from the 16th layer to the 12th layer. This constitutes a new YOLO-R network structure, which allows the improved network to further improve pedestrian detection accuracy. The structure diagram is shown in Table I.

In the YOLO-R network,a Passthrough layer is added before the 21st layer, It is composed of a Route layer (19th layer) and a Reorg layer (20th layer),combine the features of the maxpooling shallow layer features in the 11th layer with the deep layer features in the 22nd layer.The Passthrough layer was added before the 25th layer and consisted of a Route layer (23rd layer) and a Reorg layer (24th layer),combine the features of the maxpooling shallow layer features in the 11th layer with the deep layer features in the 25th layer.The Passthrough layer was added before the 30th layer and consisted of the Route layer (28th layer) and the Reorg layer (29th layer),combine the features of the maxpooling shallow layer features in the 11th layer with the deep layer features in the 30th layer.And the original YOLO network structure in the 31rd layer of the Route layer from the original 16th to the 12th layer, in order to extract more shallow layer feature information and the deep layer information to do fusion.

IV. NETWORK TRAINING

Based on the open source neural network framework Darknet [15] and the YOLO-R network structure as a model, pedestrian detectors are trained.In order to speed up the training and prevent over-fitting, the selected impulse constant is 0.9 and the momentum attenuation coefficient is 0.0005.The learning rate adopts a multi-step strategy. In order to reduce the training time, the convolutional network is initialized with the network parameters trained by the Darknet19 [15] network model.

A. DATA SET

The data set used in this paper is the INRIA data set, which is the most commonly used static pedestrian detection data set. It provides the original pictures and the corresponding annotation files. The training set has 614 positive samples (including 2416 pedestrians). The test set has 288 positive samples (including 1126 pedestrians). In the picture, most of the human body is in a standing position with a height of more than 100 pixels, and the picture has a high definition.

V. EXPERIMENTAL RESULTS AND ANALYSIS

The software and hardware platform used in this experiment is shown in Table II.

TABLE I
YOLO-R NETWORK STRUCTURE

Layer	Type	Filters	Size/Stride	Input	Output
0	Conv	32	3x3/1	416x416x3	416x416x32
1	Maxpool		2x2/2	416x416x32	208x208x32
2	Conv	64	3x3/1	208x208x32	208x208x64
3	Maxpool		2x2/2	208x208x64	104x104x64
4	Conv	128	3x3/1	104x104x64	104x104x128
5	Conv	64	1x1/1	104x104x128	104x104x64
6	Conv	128	3x3/1	104x104x64	104x104x128
7	Maxpool		2x2/2	104x104x128	52x52x128
8	Conv	256	3x3/1	52x52x128	52x52x256
9	Conv	128	1x1/1	52x52x256	52x52x128
10	Conv	256	3x3/1	52x52x128	52x52x256
11	Maxpool		2x2/2	52x52x256	26x26x256
12	Conv	512	3x3/1	26x26x256	26x26x512
13	Conv	256	1x1/1	26x26x512	26x26x256
14	Conv	512	3x3/1	26x26x256	26x26x512
15	Conv	256	1x1/1	26x26x512	26x26x256
16	Conv	512	3x3/1	26x26x256	26x26x512
17	Maxpool		2x2/2	26x26x512	13x13x512
18	Conv	1024	3x3/1	13x13x512	13x13x1024
19	Route	11			
20	Reorg		/2	26x26x256	13x13x1024
21	Conv	512	1x1/1	13x13x1024	13x13x512
22	Conv	1024	3x3/1	13x13x512	13x13x1024
23	Route	11			
24	Reorg		/2	26x26x256	13x13x1024
25	Conv	512	1x1/1	13x13x1024	13x13x512
26	Conv	1024	3x3/1	13x13x512	13x13x1024
27	Conv	1024	3x3/1	13x13x1024	13x13x1024
28	Route	11			
29	Reorg		/2	26x26x256	13x13x1024
30	Conv	1024	3x3/1	13x13x1024	13x13x1024
31	Route	12			
32	Reorg		/2	26x26x512	13x13x2048
33	Route	32 30			
34	Conv	1024	3x3/1	13x13x3072	13x13x1024
35	Conv	30	1x1/1	13x13x1024	13x13x30
36	Detection				

Table II
EXPERIMENTAL SOFTWARE AND HARDWARE PLATFORM

Hardware Platform		Software Platform	
CPU	IntelCore i7-6700 3.4GHz	Operating System	Ubuntu16.04 64-bit
GPU	GTX 1050Ti	Deep Learning Framework	Darknet
RAM	32G	Programming Language	C

This paper makes comparison results of the YOLO v2 and YOLO-R network models on the test set of the INRIA data set. The experimental results are shown in Table III. It can be seen from Table III that the YOLO-R network model is superior to the original YOLO v2 network model in terms of Precision, Recall and IOU.

Table III
COMPARISON OF THE RESULTS OF THE TWO ALGORITHMS ON THE INRIA TEST SET

	YOLO v2	YOLO-R
Precision	97.37%	98.56%
Recall	89.33%	91.21%
IOU	74.46%	76.18%

Evaluation index in formula (2) and (3).

$$\Pr ecision = \frac{TP}{TP + FN} . \quad (2)$$

$$\mathrm{Re}call = \frac{TP}{TP + FP} . \quad (3)$$

In the formula (2) and (3), TP, FP, and FN respectively represent the number of samples that correctly recognized pedestrians as pedestrians, the number of samples that identified non-pedestrians as pedestrians, and the number of samples that identified pedestrians as non-pedestrians.

Fig. 2 shows the some experimental results

Fig. 2 The experimental results of some images. (a) the result of YOLO v2,there is pedestrian missing phenomenon in the picture. (b) the result of YOLO-R,missing pedestrian detected in the picture. (c) the result of YOLO v2,there is pedestrian obstruction phenomenon in the picture. (d) the result of YOLO-R, obstructed pedestrian detected in the picture. (e) the result of YOLO v2,there is pedestrian false detection in the picture. (f) the result of YOLO-R, there is no pedestrian false detection in the picture.

Fig. 3 shows the comparison of the Loss curves of the two algorithms in the training process.The formula for the Loss function as in (4):

$$L\left(\hat{y},y\right)=-\left(y\cdot\log\left(\hat{y}\right)+\left(1-y\right)\cdot\log\left(1-\hat{y}\right)\right). \qquad (4)$$

Where y is the true value of the model and $\hat{y}$ is the predicted value of the model.This paper selects the model that has been trained to 45,000 times as the final weight model for pedestrian detection.

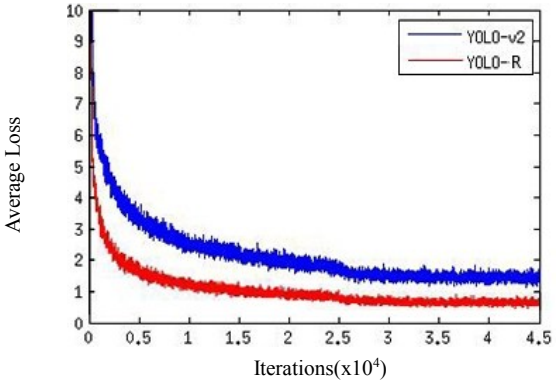

Fig. 3 Comparison of Loss curves of YOLO v2 and YOLO-R

In order to verify the performance requirements of the new passthrough layer for pedestrian detection in this paper, LAMR[16] (Log-Average miss RATE) was selected as the evaluation index.LAMR reflects the relationship between false negative per image (FPPI) and missed rate.The lower the FPPI, the better the performance of the pedestrian detection method.This paper selected the INRIA data set and tested it.When FPPI=0.1,The missed rate of YOLO-R network model was 10.05% lower than 11.29% of YOLO v2 network model.For comparison, this paper also compares with some popular pedestrian detection algorithms [17-20].The results are shown in Fig. 4.

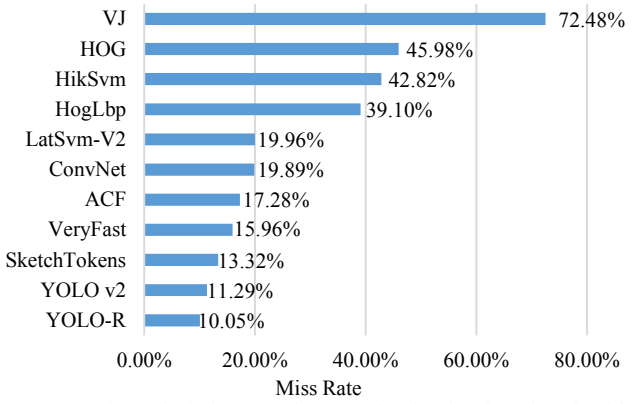

Fig. 4 Comparison of missing detection rate of pedestrian detection algorithm

VI. CONCLUSION

This paper adopts the deep learning method to improve the network structure of YOLO v2 and obtains the YOLO-R network model. It has achieved good results in pedestrian detection. In this paper, three Passthrough layers are added to the YOLO v2 network to extract the shallow layer pedestrian features, and the shallow layer features extracted from the Route layer of the original algorithm are improved from the 16th layer to the 12th layer, combine shallow layer features with deep layer features to extract more fine-grained features.Experiments show that this algorithm improves the accuracy of pedestrian detection. The number of detection frames can reach 25 frames/s, basically meeting the requirements of real-time performance. In the future, we intend to integrate more pedestrian context features to improve the accuracy of pedestrian detection.

ACKNOWLEDGMENT

This work was supported by the Science and Technology Project of Gansu Colleges and Universities (Grant No. 2017D-08).

REFERENCES

[1] Dalal,N., Triggs, B.:Histograms of oriented gradients for human detection. In:CVPR.(2005).
[2] Felzenszwalb, P., McAllester,D., Ramanan, D.A.:discriminatively trained, multiscale,deformable part model.In:CVPR.(2008).
[3] Dollár P, Tu Z, Perona P, et al. Integral Channel Features[C]//British Machine Vision Conference, BMVC,2009:1-11.
[4] Dollar P, Appel R, Belongie S, et al. Fast Feature Pyramids for Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2014,36(8):1532-45.
[5] Ouyang, W., Wang, X. : Joint deep learning for pedestrian detection. In : ICCV. (2013).
[6] Tian Y, Luo P, Wang X, et al. Pedestrian detection aided by deep learning semantic tasks[J].2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2015:5079-5087.
[7] Anelia Angelova,Alex Krizhevsky,Vincent Vanhoucke, et al. Real-Time Pedestrian Detection With Deep Network Cascades. In:BMVC.(2015).
[8] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[J]. Computer Science,2013:580-587.
[9] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2014, 37(9):1904-16.
[10] Girshick R. Fast R-CNN[C]//IEEE International Conference on Computer Vision.IEEE,2015:1440-1448.
[11] Ren S,He K,Girshick R,et al.Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015:1-1.
[12] L. Wei, A. Dragomir: SSD: Single Shot MultiBox Detector.arXiv preprint arXiv:1512.02325v5,2016.
[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi.You only look once: Unified, real-time object detection. arXiv preprint arXiv:1506.02640, 2015.
[14] Redmon J, Farhadi A. YOLO9000:Better, Faster, Stronger.arXiv preprint arXiv:1612.08242v1,2016.
[15] Redmon J. Darknet: Open source neural networks in c.http://pjreddie.com/darknet/, 2013–2016.
[16] Wojek C, Dollar P, Schiele B, et al. Pedestrian Detection: An Evaluation of the State of the Art[J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012,34(4):743-761.
[17] Benenson R, Mathias M, Timofte R,LV Gool:Pedestrian detection at 100 frames per second. Computer Vision &Pattern Recognition, 2012, 157(10):2903-2910.
[18] Chen G, Ding Y, Xiao J, et al. Detection Evolution with Multi-order Contextual Co-occurrence[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013:1798-1805.
[19] Zhang S, Benenson R,Omran M,et al.How Far are We from Solving Pedestrian Detection? arXiv preprint arXiv:1602.01237v2,2016.
[20] Sermanet P, Kavukcuoglu K, Chintala S, et al. Pedestrian detection with unsupervised multi-stage feature learning [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC:IEEE Computer Society, 013:3626-3633.