Research and Implementation
of Anti-occlusion Algorithm for Vehicle
Detection in Video Data

Yongqi Wu[1(✉)], Zhichao Zhou[1], Lan Yao[1], Minghe Yu[2],
and Yongming Yan[3]

[1] College of Computer Science and Engineering, Northeastern University,
Shenyang 110819, China
`wuyongqi@stumail.neu.edu.cn`
[2] College of Software, Northeastern University, Shenyang 110819, China
[3] Department of Video Big Data, DIXN Technology Co., Ltd.,
Shenyang 110004, China

Abstract. Object detection is an important branch of image processing and computer vision, which has become a hot research issue in recent years. Accurate target detection in a video is the foundation of intelligent surveillance system. Since the background scenario is dynamic and even especially complicated when vehicles occlude, the target detection accuracy is declined. Therefore, based on the bounding box regression algorithm, this paper constructs adjacent punishment mechanism to make the bounding box clear off other objects. The proposal weak confidence suppression is leveraged for the robustness of the detector when occlusion happens. Experiments show that the proposed method outperforms traditional methods on three different datasets.

Keywords: Object detection · Vehicle occlusion · Bounding box regression · Confidence suppression

1 Introduction

Vehicle detection, as an important part of intelligent transportation system, has always been a research hotspot to detect mobile vehicles in a video. It is the foundation of vehicle tracking, vehicle type identification, flow statistics, video velometer and other technologies. However, in practical applications, scenes often change dynamically. For example, if two vehicles are similar in exterior, it is likely to blur them. This detection result need to be further revised by the non-maximum suppression algorithm [1], otherwise the prediction box on the occluded vehicle is probably suppressed by the one on the adjacent vehicle causing a detection loss. Our work will improve the detection accuracy under occlusion circumstance.

The main contributions of the work are:

- Propose a bounding box regression algorithm that introduces the adjacent punishment mechanism, so that the bounding box frames its specified target and being clear off from adjacent targets.

- A proposal weak confidence suppression algorithm is proposed to reduce the probability with which the proposal is suppressed by adjacent targets.
- Experiments are conducted on PASCAL VOC, UA-DETRAC and Highway-Vehicle datasets and show the efficiency of the proposed algorithm.

2   Related Work

Most of the early object detection algorithms are based on manual features. Malisiewicz [2] trains each sample through the SVM classifier based on the HOG feature. Felzenszwalb [3] proposes to use deformable part model to detect multi-class objects. In recent years, with the rise of deep learning, convolution neural network (CNN) has brought new ideas for object detection [4–10]. Networks such as SPP-net, R-FCN, and GoogLeNet [11–13] can be used for object detection.

For vehicle detection, many researchers propose methods with efficiency and acceptable accuracy [14–16]. However, they have not considered occlusion. When the traffic flow on the road increases, the vehicles will occlude each other, which seriously affects the detection accuracy of vehicles if the above method are applied.

This paper improves the bounding box regression algorithm to make the proposal of the vehicle separated from adjacent targets, and proposes a proposal weak confidence suppression algorithm to avoid the proposal being mistakenly suppressed by adjacent targets, so as to improve the detection accuracy with occlusion.
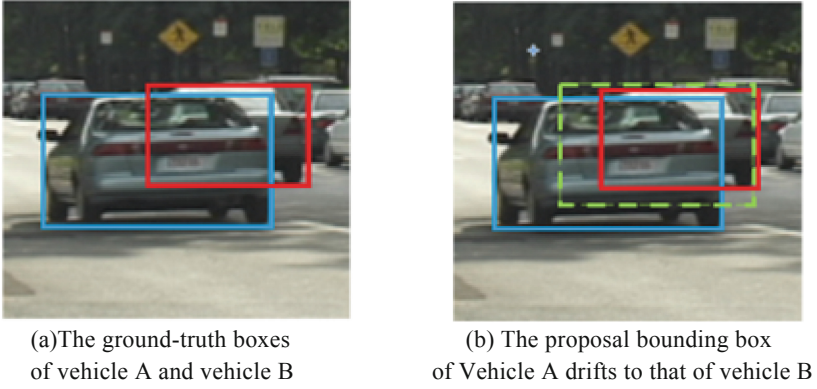
3   Implementation Method

3.1   Bounding Box Regression and Its Improvement

Bounding box regression is used in object detection methods such as RCNN and Fast R-CNN to revise the position of the proposal and make it close to the designated target. Let P and G be the original proposal bounding box and true box represented by their center coordinates and width and height. Bounding box regression is to find a map from P to $\hat{G}$, while $\hat{G}$ is approximate to the ground-truth box and defined as:

$$f(P) = (\hat{G}) \quad \text{s.t.} \quad dist(G, \hat{G}) < dist(G, P) \tag{1}$$

However, this regression only makes the proposal as close as possible to its target, without considering the influence of adjacent objects. When occlusion occurs, as shown in Fig. 1(a), the small box is ground-truth of A and the big box is ground-truth of B. When A is partially occluded by B, A's proposal is likely to be misaligned due to the similarity of A and B as shown in Fig. 1(b). The dotted line box is A's proposal and it may drift as B's.

To solve this problem, the adjacent punishment mechanism is applied as the bounding box regression algorithm. In the detector training process, each proposal will not only approach its ground-truth and define it as positive term L_P, but also clear off from the ground-truth of adjacent objects and define it as negative term L_N. By

(a)The ground-truth boxes
of vehicle A and vehicle B

(b) The proposal bounding box
of Vehicle A drifts to that of vehicle B

Fig. 1. The proposal bounding boxes of the occluded vehicle

introducing the repulsive effect of adjacent objects on the proposal bounding box, the detection accuracy can be improved by avoiding the proposal drifting to the near similar objects when the target is occluded. The regression calculation is defined as:

$$L = L_P + L_N \tag{2}$$

The positive term L_P and negative term L_N are defined as:

$$L_p(p, g) = \sum_{i \in \{x,y,w,h\}} smooth_{L_1}(p_i - g_i) \tag{3}$$

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases} \tag{4}$$

Since $IoU(p, G_i) \in [0, 1]$, the $smooth_{L_1}$ function is modified as follows:

$$L_N(p, G_n) = \sum_{i \in n} smooth_{L_1^1}(IoU(p, G_i)) \tag{5}$$

$$smooth_{L_1^1}(x) = \begin{cases} (0.5 + x)^2 - 0.25 & x \leq 0.5 \\ \frac{|\ln(1-x)|}{2-x} - 0.5 & x > 0.5, \end{cases} \tag{6}$$

L_P is used to narrow the gap between proposals and ground-truth boxes, L_N is used for repulsive effect on the proposal bounding box. G_n represents the ground-truth boxes set of all objects except the target, and $IoU(p, G_i)$ represents the IoU between the proposal and the ground-truth box.

3.2   Proposal Weak Confidence Suppression Algorithm

After the bounding box regression, a large number of proposals are generated near the target. In order to eliminate the false, the non-maximum suppression (NMS) algorithm is used to remove redundant proposals by the overlapping area (IoU), and reserve the proposal with the highest confidence for each target.

NMS algorithm selects the highest scored box b_m from the proposal set B, then removes the proposal whose IoU value with b_m is greater than the threshold N_t from B, and repeats until B is empty. However, there are some drawbacks in NMS. When dense occlusion occurs, the artificial N_t will matter the detection accuracy. If N_t is too large, it causes false detection. Contrarily, it loses detection.

Therefore, we improve NMS and propose a proposal weak confidence suppression algorithm. Instead of deleting proposals from the set B, the confidence S_i of the proposal b_i, which is determined by the polarity of IoU values of b_i and b_m is introduced to avoid false suppression and reduce the impact of N_t. Assuming that U_p represents the IoU between b_i and b_m, if U_p is greater than the threshold N_t, the confidence S_i of b_i is multiplied by the confidence attenuation coefficient α:

$$\alpha = -U_p * \ln(1 + U_p) + 1 \tag{7}$$

Its function is to reduce the confidence S_i. Conversely, if U_p is less than or equal to the N_t, the confidence S_i will not be changed. Repeat the process until set B is empty. Finally, the prediction box set D and the confidence set S_d is output.

4   Experiments and Results Analysis

We use two open source image datasets: the PASCAL VOC dataset and the UA-DETRAC dataset [17]. The PASCAL VOC dataset is used for PASCAL VOC Challenge Competition with a total of 1,659 vehicle images. The UA-DETRAC dataset is mainly taken in Beijing and Tianjin with 6,250 vehicle images. We select the peak time video of Shanghai-Hangzhou-Ningbo Expressway, frame it to obtain images, and employ 20 students to annotate images manually. Finally, the image and annotation information are constructed into Highway-Vehicle dataset, which contains 12,800 vehicle images.

In order to verify the performances, Faster R-CNN is applied as the detector and the VGG-16 network in Faster R-CNN is replaced by ResNet-101 network with stronger feature extraction. The related parameters are set as follows: the training learning rate is 0.001, the attenuation step size is 30000, the attenuation coefficient is 0.1, the training batch size is 128, and the detection confidence threshold is 0.5.

For the evaluation, if the IoU between the prediction box and its ground-truth box is greater than 0.5, it is a correct detection, otherwise it is a false detection. When the annotated vehicle is not detected, it is a detection loss. Using three image datasets, Faster R-CNN and its reformative method is verified by five-fold cross validation under the different IoU thresholds. The average precision (AP) of these two methods are shown in Table 1.

Table 1. Vehicle detection AP of two methods on three datasets

N_t	Method	PASCAL VOC	UA-DETRAC	Highway-Vehicle
0.2	Faster R-CNN	0.7353	0.6611	0.6134
	Proposed method	0.7587	0.6827	0.6447
0.3	Faster R-CNN	0.7521	0.6773	0.6353
	Proposed method	0.7763	0.7011	0.6724
0.4	Faster R-CNN	0.7603	0.6817	0.6407
	Proposed method	0.7833	0.7083	0.6752
0.5	Faster R-CNN	0.7581	0.6855	0.6378
	Proposed method	0.7805	0.6972	0.6718
0.6	Faster R-CNN	0.7413	0.6715	0.6232
	Proposed method	0.7705	0.6876	0.6633
0.7	Faster R-CNN	0.7154	0.6573	0.5946
	Proposed method	0.7428	0.6647	0.6285

The experimental results show that the detection performance on three image datasets is improved when the bounding box regression algorithm with adjacent punishment mechanism and the proposal weak confidence suppression algorithm is introduced. When the IoU threshold is 0.4, the proposed method yields the best performance. It is shown that the proposed method can effectively improve the detection accuracy and improve the accuracy and stability of the detector under the occlusion condition.

5 Conclusion

Vehicles have always been an important target for object detection. Due to the complexity of the actual road environment, such as vehicle occlusion, the accuracy and stability of object detection algorithms have been challenged. Compared with the traditional object detection algorithm, the proposed algorithm makes the proposal close to its designated target while clear from other nearby targets, and reduces the probability that the proposal is mistakenly suppressed by adjacent targets, so as to improve the detection performance of the detector. The experimental results show that the reformative Faster R-CNN is more efficient in vehicle detection.

References

1. Neubeck, A., Gool, L.: Efficient non-maximum suppression. In: International Conference on Pattern Recognition, vol. 3, pp. 850–855. IEEE, Hong Kong (2006)
2. Malisiewicz, T., Gupta, A., Efros, A.: Ensemble of exemplar-SVMs for object detection and beyond. In: IEEE International Conference on Computer Vision 2011, ICCV, vol. 1, no. 2. IEEE, Barcelona (2011)

3. Felzenszwalb, P., Girshick, R., Mcallester, D.A.: Visual object detection with deformable part models. Commun. ACM **56**(9), 97–105 (2010)
4. Zhao, H., Xia, S., Zhao, J., Zhu, D., Yao, R., Niu, Q.: Pareto-based many-objective convolutional neural networks. In: Meng, X., Li, R., Wang, K., Niu, B., Wang, X., Zhao, G. (eds.) WISA 2018. LNCS, vol. 11242, pp. 3–14. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02934-0_1
5. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 2012, NIPS, pp. 1097–1105. IEEE, Lake Tahoe (2012)
6. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2014, CVPR, pp. 580–587. IEEE, Columbus (2014)
7. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision 2015, ICCV, pp. 1440–1448. IEEE, Santiago (2015)
8. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems 2015, NIPS, pp. 91–99. IEEE, Montreal (2015)
9. Cai, Z., Fan, Q., Feris, Rogerio S., Vasconcelos, N.: A unified multi-scale deep convolutional neural network for fast object detection. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 354–370. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_22
10. Lin, T., Girshick, R.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, CVPR, pp. 2117–2125. IEEE, Hawaii (2017)
11. He, K., Zhang, X., Ren, S.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1904–1916 (2014)
12. Dai, J., Li, Y., He K.: R-FCN: object detection via region-based fully convolutional networks. In: Advances in Neural Information Processing Systems 2016, NIPS, pp. 379–387. IEEE, Barcelona (2016)
13. Szegedy, C., Liu, W., Jia, Y.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015, CVPR, pp. 1–9. IEEE, Boston (2015)
14. Fan, Q., Brown, L., Smith, J.: A closer look at Faster R-CNN for vehicle detection. In: IEEE Intelligent Vehicles Symposium 2016, vol. IV, pp. 124–129. IEEE, Gothenburg (2016)
15. Song, H., Zhang, X., Zheng, B., Yan, T.: Vehicle detection based on deep learning in complex scene. Appl. Res. Comput. **35**(04), 1270–1273 (2018)
16. Lee, W., Pae, D., Kim, D.: A vehicle detection using selective multi-stage features in convolutional neural networks. In: International Conference on Control, Automation and Systems 2017, ICCAS, pp. 1–3. IEEE, Singapore (2017)
17. Wen, L., Du, D., Cai, Z.: UA-DETRAC: a new benchmark and protocol for multi-object detection and tracking. arXiv preprint arXiv:1511.04136 (2015)