ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 21-04-06
36

j.inst.Korean.electr.electron.eng.Vol.25,No.4,619~624,December 2021
https://doi.org/10.7471/ikeee.2021.25.4.619

Vehicle Orientation Detection Using CNN

Nguyen Huu Thang*, Jaemin Kim**★

Abstract

Vehicle orientation detection is a challenging task because the orientations of vehicles can vary in a wide range in captured images. The existing methods for oriented vehicle detection require too much computation time to be applied to a real-time system. We propose Rotate YOLO, which has a set of anchor boxes with multiple scales, ratios, and angles to predict bounding boxes. For estimating the orientation angle, we applied angle-related IoU with CIoU loss to solve the underivable problem from the calculation of SkewIoU. Evaluation results on three public datasets DLR Munich, VEDAI and UCAS-AOD demonstrate the efficiency of our approach.

Key words : Vehicle Orientation, Vehicle Detection, Real-Time, Convolutional Neural Network(CNN)

Ⅰ. Introduction

Vehicle detection is one of the key parts of an intelligent traffic management system. Within a dense traffic scene, both location and orientation of vehicles contribute significantly to maintaining the accuracy of the tracking system. Vehicle orientation detection with rotate bounding boxes can reduce the influence of the background image and provide the angle attribute to further processing such as tracking.

1. Object Detection Methods

The current object detecting methods can be divided into two categories: two-stage object detectors (fast R-CNN [1], faster R-CNN [2]), and single-stage object detectors single-shot mulch-box detector (SSD [3]), You Only Look Once (YOLO [4])). With a simple pipeline, the single-stage networks object detectors are very fast, while the two-stage networks are slow but have better accuracy.

Since Redmon et al. [5] introduced the first version of YOLO, it attracted lots of attention because of its feasibility to apply object detection in real-time. From the third version, the most notable feature of YOLO is that it makes detections at three different scales. Bochkovskiy et al. [6] performed a series of experiments with many of the most advanced innovation ideas of computer vision for each part of YOLOv3 the architecture to create YOLOv4. By applying the most advanced optimization methods (Mosaic augmentation, CIoU loss, Cross-stage partial connections (CSP), Mish activation, etc.), YOLOv4 is proved to be incredibly fast and still has high accuracy as shown in Fig. 1.

* Master's Degree Student, School of Electronic and Electrical Engineering, Hongik University
** Professor, School of Electronic and Electrical Engineering, Hongik University
★ Corresponding author
E-mail : jmkim@hongik.ac.kr, Tel : 82-2-320-1634

※ Acknowledgment
This work was supported by Hongik University Research Fund.
Manuscript received Nov. 24, 2021; revised Dec. 5, 2021; accepted Dec. 6, 2021.

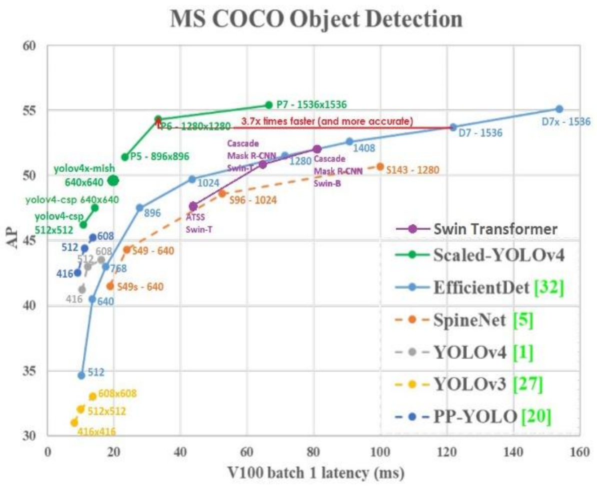


Fig. 1. Comparison of the Scaled YOLOv4 and other state-of-the-art object detectors [6].

Even though these methods achieved impressive results on different benchmarks their predicted bounding box output mostly aligned with the horizontal axis and ignored the orientation of objects. A horizontally aligned bounding box can cover not only object areas but also the background. Therefore it can confuse the network in complicated circumstances. Moreover, it is a difficult task to separate the objects with a horizontal bounding box when objects are closely attached in the 45-degree direction.

2. Object Orientation Detection

Yang et al. [7] employed a refinement stage to the RetinaNet to refine the bounding box and added a feature refinement module (FRM) to reconstruct the feature map. The FRM module uses the feature interpolation to reconstruct the feature map pixel-wise and achieve feature alignment. But their approximate SkewIoU loss calculation can be indifferentiable because objects can have various orientations appearance in images. Furong Shi et al. [8] proposed vehicle, center, scale, and orientation prediction-based detector (VCSOP) including four modules for multitask learning. By dividing the ResNet-50 into 5 stages, the localization information from the lower layer and the global semantic information in higher layer features would be fused for further use of vehicle feature prediction. This strategy can significantly improve the performance of object detection but consume lots of time for post-processing. Furthermore, their range of angle from 0-90 degrees does not reflect the diversity of the vehicle orientations in images.

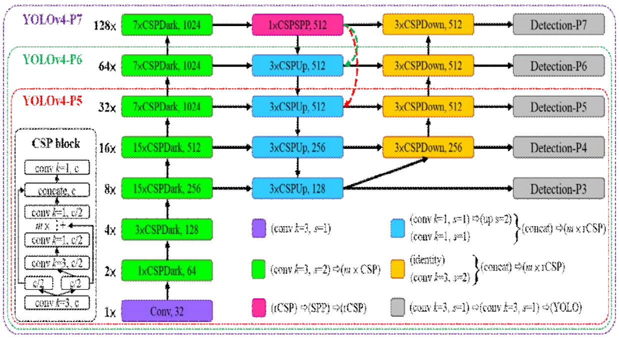


Fig. 2. Scaled YOLOv4 architecture.

3. The Proposed Method

The contributions of our approach are summarized as follows:

1. A real-time vehicle orientation detection is proposed, which can predict both location and orientation of vehicles in images.

2. Angle parameter is added to the traditional anchor boxes of Scaled YOLOv4 to make rotated anchors. These rotated anchors will assist the network to learn all the arbitrarily orientation.

3. For rotated objects detection, we applied angle-related IoU (ArIoU) [9] to solve the underivable problem from the calculation of SkewIoU [7].

Ⅱ. The Proposed Method

1. Rotate YOLOv4

The proposed vehicle orientation detector is based on Scaled YOLOv4 [4], denoted as Rotate YOLOv4. Fig. 2 illustrates the network structure of our method. YOLO splits the input image into S x S (7 x 7 by default) grid cell. These cells predict parameters of bounding boxes and probabilities of classes. Finally, YOLO applies

Non-Maximum Suppression (NMS) to remove the overlap bounding boxes.

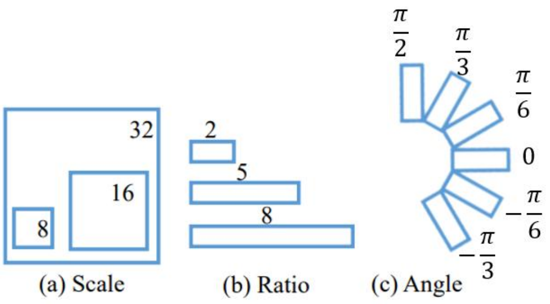
For YOLOv4-based rotation detection, the vehicle orientation is represented by a rotated bounding box with five parameters (t_x,t_y,t_w,t_h,θ). If the cell is offset from the top left corner of the image by (c_x,c_y) and the bounding box prior has width, height, and angle, then the predictions correspond to:

$$\begin{aligned} b_x &= \sigma(t_x)+c_x \\ b_y &= \sigma(t_y)+c_y \\ b_w &= p_w e^{t_w} \\ b_h &= p_h e^{t_h} \\ \theta &= p_\theta - \theta_a \end{aligned} \tag{1}$$

The width is set to the long side of the bounding box, and the height is set to the short side. θ is the acute angle between the horizontal axis and the long side, ranging from $(-\frac{\pi}{2},\frac{\pi}{2})$.

2. Rotated Anchor

The horizontally aligned anchors used in YOLOv4 [6] are not suitable for vehicle orientation detection. Therefore, we extend the existing anchor. First, we add an angle parameter to control the orientation. We use six different directions to cover the vehicle orientation. We keep the ratios (1:2, 1:5, 1:8) and scales of 8, 16, and 32 as YOLOv4 for the anchor box. The anchor box strategy is described in Fig. 3.



(a) Scale (b) Ratio (c) Angle

Fig. 3. Anchor strategy used in our approach.

3. Loss function

Inspired by the work of R3Det [7], we want to apply their loss function into our architecture. However, in their work, Yang et al. [7] explained that the smooth L1 loss function is not suitable for rotation detection, while SkewIoU is too sensitive for objects with large aspect ratios. Therefore, we cannot use the SkewIoU directly as the regression loss function. To solve that problem, we applied the angle-related IoU (ArIoU) proposed by Lei Liu et al [9] in our loss function because it is simpler and faster. We also adapted the CIoU, which is applied in YOLOv4 to have an effective regression loss function. The final loss function is described as below:

$$\begin{aligned} L &= \frac{\lambda_{coord}}{N}\sum_{n=1}^{N} obj_n \frac{L_{reg}(t',t)}{|L_{reg}(t',t)|}|f(ArIoU)| \\ L_{reg}(t',t) &= L_{smooth-L1}(t'_\theta,t_\theta) - CIoU(t',t) \end{aligned} \tag{2}$$

where N indicates the number of anchors, obj_n is a binary value (obj_n=1 for the foreground and obj_n=0 for the background), t' represents the predicted offset vectors, t denotes the targets vector of the ground-truth. $ArIoU$ denotes the overlap of the prediction box and ground-truth. The hyper-parameter λ controls the trade-off and is set to 1 by default. Compared to the traditional regression loss, the new regression loss can be divided into two parts, $\frac{L_{reg}(t',t)}{|L_{reg}(t',t)|}$ determines the direction of gradient propagation (a unit vector), which is an important part to ensure that the loss function is derivable. $|f(ArIoU)|$ is responsible for adjusting the loss value (magnitude of gradient), so it can enforce the detector to learn the right angle. Through such combination, the loss function is derivable, while its size is highly consistent.

Ⅲ. Experiments

1. Datasets

We evaluate our vehicle orientation detection framework on three public datasets: DLR Munich Vehicle dataset [10], Vehicle Detection in Aerial

Imagery (VEDAI) dataset [11], and UCAS-AOD dataset [12].

VEDAI dataset is acquired over Utah, U.S., and contains various backgrounds such as agrarian, rural, and urban areas. Moreover, the VEDAI dataset has two different resolutions and is divided into two parts: VEDAI512 and VEDAI1024. VEDAI512 comprises the downscaled images of VEDAI1024. The images are divided into ten folds for cross-validation. Each fold contains approximately the same number of vehicles. In our experiments, only VEDAI512 (512x512) was used.

DLR Munich vehicle dataset contains 20 aerial images sized 5616 x 3744 pixels. In the experiments, we only used the first 10 images for training and comparing methods.

UCAS-AOD has 1510 images of resolution 1280 x 659, with two categories. This dataset can be challenging due to the large aspect ratio with arbitrary orientation. In our experiment, we randomly chose 960 images for training, and 83 images for testing.

When comparing the performance of the proposed method with the existing methods, the performance values of the other methods are those published in previous papers. The experimental conditions of the proposed method were set to be as similar as possible to the experimental conditions used in the previous papers.

The goal of our method concentrates on detecting vehicles in images, so we reduced the classes in the two datasets VEDAI and DLR Munich by putting other object categories into class “Others”, and only keep the vehicles classes (i.e., car, truck, bus…)

2. Evaluation

To evaluate the performance of our approach, the popular evaluation criteria are applied, namely, precision, recall, F1 score, and average precision (AP). The precision and recall metrics measure the ratio of correctly identified vehicles to totally detected vehicles and to actual vehicles, respectively. The comprehensive performance of precision and recall is evaluated using the F1 score. The AP metric is measured by the area under the precision-recall curve. The higher the F1 score and AP, the better the performance.

Table 1. Performance comparison between different methods on DLR Vehicle Aerial dataset.

Methods	Recall	Precision	F1 score	Time Per image
ACF[13]	69.30%	86.80%	0.77	4.40s
Faster R-CNN[2]	68.74%	88.95%	0.78	3.84s
Oriented SSD[14]	78.84%	89.20%	0.82	5.17s
RRPN[15]	82.58%	85.88%	0.84	4.11s
VCSOP[8]	**86.52%**	**94.62%**	0.90	2.82s
Rotate YOLOv4	84.52%	91.34%	**0.93**	**1.41s**

The experiments in this paper are developed with Pytorch. We train 1000 epochs for each dataset on a single NVIDIA GTX 1080i GPU. The initial learning rate is 1×10^(-4). We employ Adam as the optimizer for the network.

Table 1. shows the performance of the proposed Rotate YOLOv4 and five other methods. Rotate YOLOv4 does not achieve the best performance in terms of recall and precision, but the best performance in terms of F1 score. Above all, the time per image of our method is significantly smaller than other methods.

Table 2. Results of different methods on the VEDAI512 dataset. The best mean average precision (mAP) is highlighted in bold.

Methods	SVM+LBP [16]	SVM+HOG31+LBP [16]	RRPN [15]	VCSOP [12]	Rotate YOLOv4
mAP	64.3	75.0	81.2	88.5	**92.5**

The images in the VEDAI dataset do not have dense traffic as in the DLR Munich vehicle. But it is more challenging due to the various clutter backgrounds and many disturbance factors. The proposed Rotate YOLOv4 achieves the best mean AP (mAP). By using YOLOv4 backbone, our detector is also much faster with 0.03s processing time for 512x512 images.

Table 3. Results of different methods on the UCAS-AOD dataset.

Method	mAP	Plane	Car
YOLOv2[5]	87.90	96.60	79.20
YOLOv4[6]	92.30	97.60	90.70
R3Det	96.17	98.20	94.14
Rotate YOLOv4	97.21	**98.75**	**97.12**

Table 3 shows the comparison of published methods on UCAS-ADO dataset, our results are better than R3Det [7].

With better accuracy and fast speed in this dataset, our Rotate YOLOv4 proved to be an efficient method for detecting vehicle orientation in real-time.

IV. Conclusion

We have presented a real-time vehicle orientation detector based on YOLOv4, namely Rotate YOLOv4 for rotating objects in dense distribution and clutter backgrounds. The detection results of our network are arbitrarily oriented rectangles, which can describe the vehicles in the traffic more precisely. For more accurate rotation estimation, we applied angle-related IoU (ArIoU) to help the network learn the right angle. Both the qualitative and quantitative results of the experiments proved that our approach can be reliable in various backgrounds. More importantly, our method can be applied in a real-time system.

References

[1] R. Girshick, “Fast R-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp.1440-1448, 2015.

[2] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.39, no.6, pp.1137-1149, 2017. DOI: 10.1109/TPAMI.2016.2577031

[3] W. Liu et al., “SSD: Single shot multibox detector,” in *Proc. Eur. Conf. Comput. Vis.*, pp.21-37, 2016.

[4] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao, “Scaled-YOLOv4: Scaling Cross Stage Partial Network,” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2021.

[5] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognition (CVPR)*, pp.7263-7271, 2017.

[6] Alexey Bochkovskiy, Chien-Yao Wang, and HongYuan Mark Liao, “YOLOv4: Optimal speed and accuracy of object detection,” arXiv preprint arXiv:2004.10934, 2020.

[7] Yang, Xue and Yan, Junchi and Feng, Ziming and He, Tao, “R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognition (CVPR)*, 2019.

[8] Furong Shi, Tong Zhang, and Tao Zhang, “Orientation-Aware Vehicle Detection in Aerial Images via an Anchor-Free Object Detection Approach,” *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*, Vol. 59, No.6, 2021. DOI: 10.1109/TGRS.2020.3011418

[9] Lei Liu, Zongxu Pan, Bin Lei, “Learning a Rotation Invariant Detector with Rotatable Bounding box,” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[10] Earth Observation Center, Public Datasets: https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-12760

[11] Vehicle Detection in Aerial Imagery (VEDAI): https://downloads.greyc.fr/vedai/

[12] UCAS-AOD dataset:

https://github.com/ming71/UCAS-AOD-benchmark

[13] J. Ma et al., “Arbitrary-oriented scene text detection via rotation proposals,” *IEEE Trans. Multimedia*, vol.20, no.11, pp.3111-3122, 2018.

[14] T. Tang, S. Zhou, Z. Deng, L. Lei, and H. Zou, “Arbitrary-oriented vehicle detection in aerial imagery with single convolutional neural networks,” *Remote Sens.*, vol.9, no.11, p.1170, 2017.

[15] P. Dollar, R. Appel, S. Belongie, and P. Perona, “Fast feature pyramids for object detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.36, no.8, pp.1532-1545, 2014.

[16] S. Razakarivony and F. Jurie, “Vehicle detection in aerial imagery: A small target detection benchmark,” *J. Vis. Commun. Image Represent.*, vol.34, pp. 187-203, 2016.

BIOGRAPHY

Nguyen Huu Thang (Member)

2019 : BS degree in Electrical and Communication Engineering, Danang University of Technology.
2020~ : Master’s Student in School of Electronics and Electrical Engineering, Hongik University

Jaemin Kim (Member)

1984 : BS degree in Electrical Engineering, Seoul National University.
1994 : PhD degree in ECSE Department, Rensselaer Polytechnic Institute.
2001~ : Professor in School of Electronic and Electrical Engineering, Hongik University.