International Core Journal of Engineering Volume 7 Issue 12, 2021
ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0046

Vehicle identification algorithm based on UAV lightweight YOLOV5 snow background

Yuanhang Shang [a] and Gang Wu [b, *]

School of Information Engineering, University of Tarim, Alal, Xinjiang, 843300, China

[a] coolmanfrog@qq.com, [*, b] Corresponding author e-mail: wgdem_lt@126.com

Abstract

Unmanned aerial vehicle (UAV) has the advantages of small size, wide field of vision and good flexibility. It has great advantages in accident treatment, traffic guidance and flow monitoring. China has a vast territory, long winter time in the northern region, and a large number of ice and snow cover. At present, there are few UAVs designed for snow background in northern China. In order to improve the accuracy of existing UAVs in snow background recognition, a vehicle recognition model based on YOLOV5 in snow weather is proposed. Through the lightweight YOLOV5 algorithm, the vehicle video recorded by UAV under the snow background is used to produce the data set. Through manual labeling, after image enhancement, the model is fed into the neural network training. The results show that the improved algorithm model can effectively identify vehicles based on UAV snow background.

Keywords

UAV. YOLOV5. Snow background. Vehicle identification. Lightweight.

1. Introduction

China's northern winter is long, up to six months of winter, most areas of snow cover time is long, covering a wide area. The existing UAV vehicle recognition model has low recognition accuracy for vehicles in snow background.

In recent years, with the continuous development of deep learning, many target detection algorithms based on convolutional neural networks are proposed. Convolutional neural networks can train and learn data independently, update parameters and obtain a more accurate model. Target detection algorithms based on convolutional neural networks can be divided into two categories. One is a two-step target detection algorithm, such as R-CNN, SPP-net, FastR-CNN, FasterR-CNN, MaskRCNN and so on[1]. These algorithms divide the target detection into two steps, namely RegionProposal, and send RegionProposalinto the network structure to extract features, and predict the location of the target and identify the category of the target.

The other is the one-step target detection algorithm, such as YOLOv1, YOLOv2, YOLOv3, YOLOv4, YOLOv5, etc[2-4]. This kind of algorithm does not need to generate RegionProposal, but directly divides the image into S × S grids, extracts features on the basis of grids, and obtains the location and category of the target by the algorithm. Two-step target detection algorithm generates about 2000 RegionProposals, which takes a lot of time. Its detection speed is relatively slow, but the accuracy is relatively high. Since the single-target detection algorithm extracts features on the basis of image meshing, its detection speed is fast, but the early YOLO algorithm will have some errors [5]. With the development of technology, anchorbox mechanism is proposed in YOLO. For example, there are nine anchorboxes with different sizes in YOLOv5, which can accurately detect targets at different scales and improve the detection accuracy [6].

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0046

2. The detection idea of YOLO

Traditional target detection method: sliding window classification method is time-consuming and complex, which needs to manually generate a large number of samples. The detection performance depends on the performance of the classifier and the performance of traversal. Therefore, yolov1 realizes the algorithm of directly fitting the coordinate position x, y, object width and height w, h, confidence. There is obvious defect coordinate position x, y can be arbitrary value, can also be meaningless negative value, reduce the training speed and prediction accuracy. Therefore, yolov2 is changed to the prediction of offset: the anchor mechanism is used for local prediction, that is, the coordinate positions x, y are predicted through the anchor, and then the width and height w, h are predicted based on the width and height of anchor [7]. The backbone of yolov2 is changed to darknet19, and the adjustment training diagram after different epochs is trained, namely Multi-ScaleTraining. The connection detection head cancels the connection mode of FC, and is changed to convolution layer. But yolov2 still has the problem of inaccurate prediction of small targets.

As the residual network resnet improves the backbone performance, yolov3 has a new performance improvement. Firstly, because resnet appears, the network can do deeper, backbone is changed to darknet53, and then the detection head is also changed to multi-scale. The nine scales obtained from the statistical data are down-sampled at 32 times, down-sampled at 16 times, and down-sampled at 8 times. Three detection heads are used to predict respectively [8]. When yolov4 changed to CSPDarknet53 in bakbone, some techniques were used to improve the performance. First, cutmix andmosaic were used for data enhancement, class label smoothing, and Mosaic was to merge four graphs, which played a role in label detection across the top and bottom. At the same time, the number of minibatch was increased by four times, and the same GPU memory training improved the performance. The loss function is changed to mish. The network structure adopts SPP, PAN, SAM network, DropBlcokCmBN. Yolov1-v3 is improved by yolov4 because the sigmod interval used for activation is (0, 1) open. Multiply 1.01 at sigmod [9].

Yolov5 is divided into Yolov5s, Yolov5m, Yolov5l and Yolov5 according to the network depth and the width of the feature map, and yolov5s is used as the use model in this paper [10]. The structure of Yolov5 is divided into input, backbone, Neck, prediction layer.

(1) Mosaic's data enhancement method is used at the input end. Four images are randomly called, and the random size and distribution are stacked to enrich the data, increase many small targets and improve the recognition ability of small objects. Four images can be calculated simultaneously, which is equivalent to increasing the size of Mini-batch and reducing the GPU memory consumption. Yolov5 can first set the anchor size by clustering, and then calculate the ahchor value in different training sets during each training. Then, the scaling mode of adaptive image size is used in the prediction, and the prediction speed is improved by reducing the black edge.

(2) Focus structure and CSPnet structure are mainly used on Backbone.

(3) FPN structure and PAN structure were used on Neck.

(4) GIOU _ Loss is used in the loss function.

3. YOLOV5 Training Process

The YOLO algorithm, like other neural networks, needs to be sent to the neural network model after labeling a large number of images to train the parameters of the neural network so that the trained neural network model can be better used to detect the target.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0046

3.1 Experimental Environment

Table 1. Experimental Environment

Operating system	Windows 10 64 bit operating system
CPU	Inter(R) Core(TM) 3.60 GHz
Run memory	32 GB
GPU	NVDIA GeForce GTX 2070S (8 GB)
Deep learning framework	Pytorch
CUDA version	11.0
Python version	3.7

3.2 Image Annotation

Classify the dataset into unwanted folders and mark the vehicle with LabelImg. The rectangle should only stick to the edge of the vehicle to prevent convergence during training.

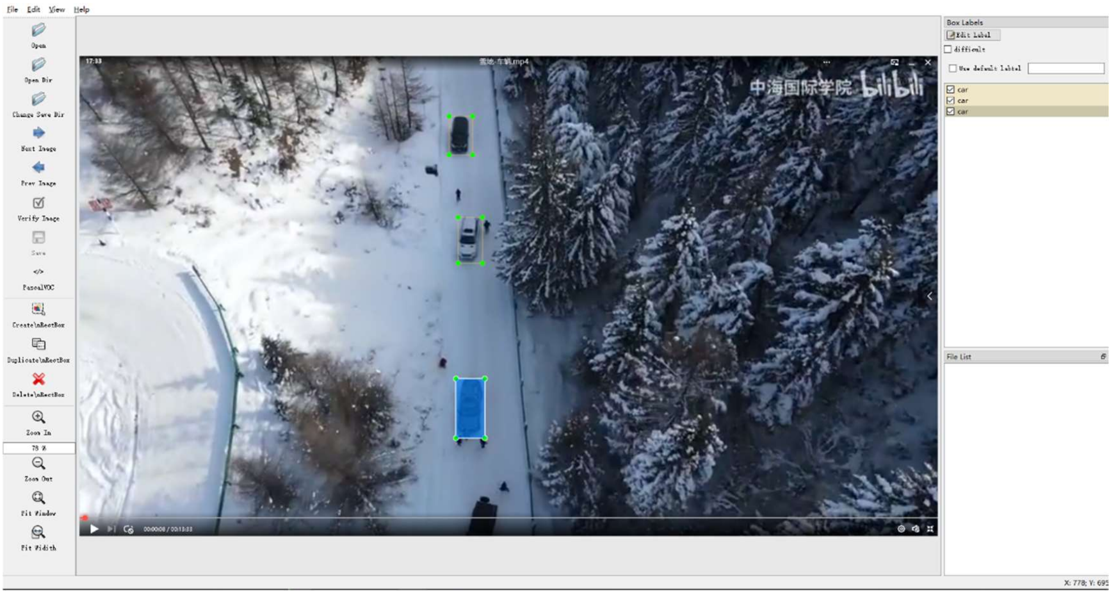

Figure 1. Vehicle Dataset Labeling Interface

3.3 Modify the configuration file

Download the pascalvoc dataset and manually annotate the image using Labelimg table box tool. The data set segmentation ratio is 5:3:2, the training set ratio is 50%, the test set ratio is 30%, and the verification set ratio is 20%. The parameters in yolov3. cfg are configured according to our computer. Classes are changed to 1. The filters under the first convolutional above each yolo are changed to 18. The voc. names file is opened and the content is changed to car. At the same time, voc. data are modified. The win system terminal is opened and the training command is input for training. During training, every 100 rounds of training will generate a weight file under backup. The obtained weight is copied to the corresponding folder. The terminal is opened and the image path to be tested is input. You can see the effect.

4. YOLO Evaluation Method

The IOU values of the detection bounding box and reference standard madness are calculated to judge the false positive and true cases in the results, and the true case is when IOU 30.5 ; when IOU < 0.5, it is a false positive ; when the value is 0, it is a false counterexample. The recall rate = TP / (TP + FN), the error rate = TP / (TP + FP), TP, FP and FN are the number of real cases, false positive cases and false negative cases, respectively .

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0046

5. Experimental results

The detection effect of vehicles without posture is shown in Figure 2-5. It can be seen from the figure that the YOLO algorithm has a good detection effect in the environment of occlusion between vehicles and shadow interference.

Figure 2. Vehicle top-view angle detection effect

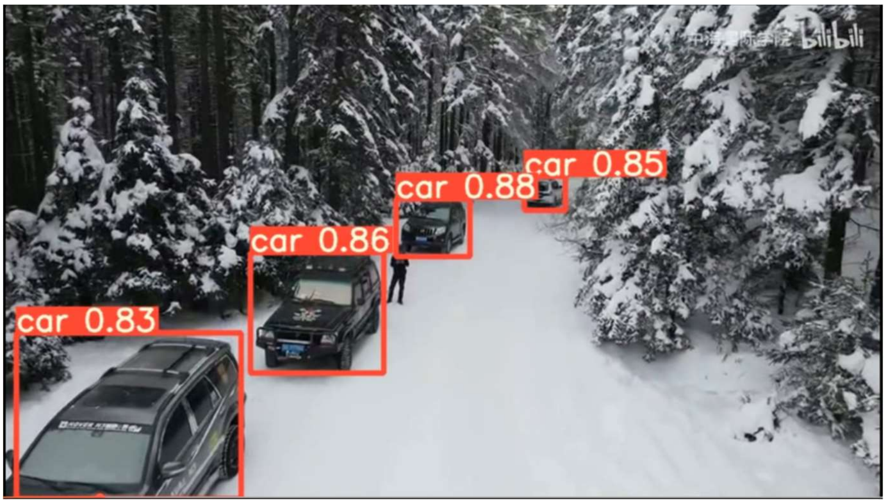

Figure 3. Detection effect above vehicle side

Figure 4. Vehicle frontal detection effect

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0046

Figure 5. Detection effect of multi-vehicle overlooking angle

6. Conclusion

In this paper, YOLO detection algorithm is used to realize vehicle detection in video, which meets the real-time requirements of vehicle recognition in snow background based on UAV, and also shows the ability of YOLO to deal with other visual tasks.

References

[1] SIVARAMAN S, TRIVEDI M M. Looking at vehicles onthe road: a survey of vision—based vehicle detection,tracking,and behavior analysis[J]. IEEE Transactions onIntelligent Transportation Systems, 2013,14 (4) : 1773-1795.

[2] ABDULRAHIM K, SALAM R A. Traffic surveillance: areview of vision based vehicle detection, recognition andtracking[J]. International Journal of Applied EngineeringResearch,2016,11(1) : 713-726.

[3] Redmon J,Farhadi A.YOLOv3:an incremental improvement[EB/OL].2018.

[4] Girshick,R. ,Donahue,J. ,Darrell,T. ,Malik,J. RichFeature Hierarchies for Accurate Object Detection and Se-mantic Segmentation[C]// Computer Vision and PatternRecognition(CVPR),2014 IEEE Conference on,2014.

[5] Ren S,He K,Girshick R,et al. Faster R-CNN:TowardsReal-Time Object Detection with Region Proposal Net-works[J]. IEEE Transactions on Patlern Analysis and Ma-chine Intelligence,2015,39(6):1137-1149.

[6] Redmon J,Divvala S,Girshick R,et al. You Only LookOnce:Unified,Real-Time Object Detection[C]// Pro-ceedings of IEEE Conference on Computer Vision and Pat-tern Recognition(CVPR),2016:779-788.

[7] Redmon J,Farhadi A. YOLO9000:Better,Faster,Stron-ger[J]. ArXir:1612. 08242,2016.

[8] Redmon J,Farhadi A. YOLOv3:An Incremental Improve-ment[J]. ArXir:1804. 02767,2018.

[9] SEO C J. The Study of Car Detection on the Highway Us-ing YOLO_v2 and UAVs[J]. Transaction of the Korean In-stitute of Electrical Engineers,2018,9(4):67-73.

[10] Bochkovskiy A, Wang C Y, Liao H Y Mark.YOLOv4: optimalspeed and accuracy of object detection [EB/OL].2020. 3-7.