International Core Journal of Engineering Volume 8 Issue 8, 2022
ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0089

Mask Detection based on Faster R-CNN

Jiafu Xiang

School of computer Science, Yangtze University, Jingzhou 434023, China

Abstract

In recent years, with the improvement of hardware computing power and the innovative development of artificial intelligence algorithms, deep learning algorithms are widely used in target detection. In view of the shortcomings of the existing manual way to check the wearing condition of personnel masks, a fast RCNN algorithm based on deep learning is proposed to realize the real-time detection of the wearing condition of masks. The algorithm first normalizes the data set, then connects the data to the fasterrcnn network for iterative training, and saves the optimal weight data as the test set. The experimental results show that the proposed algorithm has high detection accuracy and strong real-time performance, and can meet the needs of practical use.

Keywords

Faster R-CNN; Mask Detection; Deep Learning.

1. Introduction

In recent years, artificial intelligence technology, especially the related technology in the field of deep learning, has been applied to various industries. From the actual application effect of existing products, the effect of computer vision and natural language processing algorithms is remarkable, which has great advantages over traditional algorithms and machine learning algorithms. As one of the applications of computer vision algorithms, target detection algorithms are also used in aerospace detection, traffic safety, industrial equipment product detection and so on. Compared with traditional target detection and recognition methods, vision algorithms have better effects in multi-target, large-scale, small target, multi overlap and so on.

As the COVID-19 is spreading all over the world, endangering the lives and property of 7billion people, China requires that people must wear masks when taking public transport and gathering places. For the problem of whether personnel wear masks, the existing method is to manually check and equip inspectors at fixed entrances and exits. This method is not all-weather, and it is easy to miss inspection in areas with large population flow, which brings great potential safety hazards to epidemic prevention and control.

Deep learning is more and more widely used in the field of target detection. At present, the mainstream models have good results in the speed and accuracy of detection. At present, deep learning target detection algorithms are mainly divided into single-stage and two-stage detection. Single stage target detection algorithms Yolo [1], SSD [2], etc. these algorithms transform the classification problem into regression problem. Compared with two-stage algorithms, they do not need candidate boxes to directly judge the location and category of targets. Two stage target detection algorithms fast r-cnn [3], fast r-cnn [4], etc. these algorithms form candidate boxes and then make classification judgment. Taking the single-stage target detection algorithm Yolo as an example, it has the advantages of simple structure, fast speed and strong universality. It has also been widely concerned and promoted by researchers in practical research.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0089

2. Faster R-CNN Network Model

In 2012, based on Alex net, r-cnn transferred Alex net's ability in image classification to pascalvoc target, and realized the detection of regional targets by using the region proposal method. Fast r-cnn was published in nips 2015. Fast r-cnn was based on r-cnn and used deep convolution network to effectively classify. Compared with RCNN, fast r-cnn not only improves the training and testing speed, but also improves the detection accuracy. Compared with the slow speed of r-cnn test training, its advantages are mainly to overcome the shortcomings of slow extraction of candidate feature regions and the need for additional space for training to preserve the extracted features.

Fast r-cnn is the third article of regional convolution neural network. It is proposed to solve the shortcomings of long-time and slow candidate region extraction method and separation of target detection network. The general framework of the whole fast r-cnn still follows the basic structure of fast r-cnn, which uses a new technical means, namely regional recommendation network, in the region, The extraction of candidate regions is organically combined with the target detection network of fast r-cnn to realize target detection in the same network.

In terms of performance, fast r-cnn uses a very deep vgg16 network, which is 9 times faster than r-cnn and 213 times faster when tested. It has achieved a higher map on Pascal VOC 2012. At the same time, compared with sppnet, which is 24 times faster than R CNN, fast r-cnn trains vgg16 3 times faster, tests 10 times faster, and its detection results are more accurate.

2.1 Anchor

Anchor mechanism is the core in RPN. It is the essence of RPN network to generate candidate region box by sliding window, but it does not operate directly on the input original image, but shares convolution features with the last convolution layer of convolution neural network, that is, the features extracted from convolution layer are taken as the input of RPN network, and the candidate region is directly generated by sliding window. The specific method is as follows: take the feature map generated by the last layer of fast RCNN feature extraction network convolution layer as the input of RPN network, and perform convolution operation on the feature map with convolution kernel with window size of 3 * 3 and step size of 1. When the convolution kernel of 3 * 3 slides to each position of the feature map, the mapping point of the center of the current sliding window in the original map is called the anchor, and anchors of different sizes and aspect ratios are generated with the anchor as the center. In fast RCNN, in order to meet the multi-scale characteristics of the target, three convolution kernels with size feature maps of 128256512 and three convolution kernels with aspect ratios of 1:1, 1:2 and 2:1 are used, When RPN performs convolution operation, each sliding corresponds to 9 anchors on the original figure. When the number of channels of the feature map is 256, the 256 features generated after each 3 * 3 convolution operation of RPN are used by 9 anchors for position regression and category judgment. After ranking all output boxes with category confidence, select the output boxes with the highest confidence as candidate boxes.

2.2 NMS

NMS is a classical algorithm in target detection post-processing. It was first proposed by neubeck to remove the duplicate prediction frame of the two-stage target detection algorithm and save the best prediction frame. NMS algorithm first filters the prediction frames whose confidence is less than the threshold, then continuously performs the intersection and comparison operation with the prediction frame with the maximum classification confidence with other prediction frames, filters the prediction frames whose IOU value is greater than the preset intersection and comparison threshold, and finds the local optimal prediction frame in the form of iteration.

2.3 Model Structure

Fast r-cnn is mainly composed of four parts: backbone feature extraction network, RPN regional detection network, ROI pooling layer, classification and regression network. The feature extraction network includes a series of convolution and pooling operations. The classical network model vgg16

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0089

is generally used, and the weight parameters of the convolution layer are shared by RPN and fast RCNN, which is helpful to accelerate the iteration of the network. Based on the multi-scale anchor introduced by the network model, the RPN network term generates the regional candidate box. It classifies and judges whether the anchors belong to the target or background through softmax, and uses BBR to predict the anchors, so as to obtain the accurate position of the candidate box and use it for subsequent target recognition and detection. The ROI network integrates the information of the convolution layer feature map and the candidate box, maps the coordinates of the candidate box in the input image to the last layer feature map, and pools the corresponding areas in the feature map to obtain a fixed size (7) × seven × 7) The output pooling results are connected to the full connection layer behind. For the classified regression network, the full connection layer is followed by two sub connection layers: classification layer and regression layer. The classification layer is used to judge the category of the feature map, and the regression layer predicts the exact location of the target through BBR.

3. Experiment and Results

3.1 Experimental Data Set and Experimental Environment

The experiment uses the mask images in the public data set website kaggle, including the image data of 800 personnel masks. Make training and test data sets. The data sets include three categories: with_mask, without_mask and mask_worn_incorrect.

Convert the dataset to Pascal VOC format. The test set is divided into 10%. The data set division is shown in Table 1. The experimental environment uses the latest operating system of windows 10 and geforce rtx2070super graphics card for operation.

Table 1. Data set division

data set	number
dataset size	800
training set	700
test set	100

3.2 Experimental Result

In the network model training stage, the iterative batch setting size is 10 and the attenuation coefficient is 0.5 0005, the total number of iterations is 50, and the initial learning rate is set to 0 001, when the number of iterations reaches 50 respectively. Loss function total_ Loss reached 0.1228, val_ Loss reached 0.1380. Mean precision, that is, the sum of the average precision of all categories divided by the average value of the average precision of the categories in the data set. When the model is iterated to 50 times, the mean precision is close to 0.88.

After the model training, input the test data set into the model, and the test results are shown in Figure 6. The blue target box in the figure indicates to wear the mask correctly; The Yellow target box indicates wearing a mask; The pink target box indicates that the mask is not worn according to the standard specifications. The value on the target box represents the confidence of each category label. From the test results of the algorithm, it can be seen that the algorithm can well distinguish three kinds of mask wearing situations, and missed detection occurs under multi-target conditions. Limited by the performance of my graphics card, the time required for more than 50 iterations may take more than ten days, so I temporarily conducted 50 iterations of training.

4. Conclusion

The model has high accuracy in the training data. From the input prediction pictures, the model has a good effect on the mask detection of low population density, but it can not be comprehensive in the face of high-density population, especially the missing detection of small-size faces. In the subsequent

ISSN: 2414-1895 DOI: 10.6919/ICJE.202208_8(8).0089

improvement, it can be effectively improved by increasing the training data of mask and the number of training iterations.

References

[1] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR), Las Vegas, NV, Jun 27-30, 2016. Piscataway, NJ: IEEE, 2016: 779-788.

[2] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//European conference on computer vision(ECCV), Amsterdam, Holland, Oct 8-16, 2016. Berlin:Springer, 2016: 21-37.

[3] GIRSHICK R. Fast RCNN[C]: Proceedings of the IEEE international conference on computer vision(ICCV), San tiago, Chile, Dec 7-13, 2015. Piscataway, NJ: IEEE , 2015: 1440-1448.

[4] RENS, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.