Computer Life (CPL) ISSN: 1819-4818
DELIVERING QUALITY SCIENCE TO THE WORLD

An Overview on R-CNN Algorithm and Applications

Hong Zhang [1,*], Xinyu Chen [2], Xueyao Jiang [2]

[1]The College of Information Science and Technology, Chengdu University of Technology, Chengdu, Sichuan, China

[2]The College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Chengdu, Sichuan, China

---

Abstract: The major issue of Computer Vision is to parse the information that can be understood by the computer from the image. In this very field of Computer Vision, there are three main levels of understanding images, which is Classification, Object Detection, and Segmentation. The introduction of deep learning solves the problem of target detection that cannot be well dealt with by traditional methods. In this paper, we introduced the main application directions in the field of object detection, then we started from R-CNN, to introduce the current mainstream Fast R-CNN, Faster R-CNN, and Mask R-CNN in the field of object detection.

Keywords: Computer Vision, Object Detection, Deep Learning, R-CNN.

---

1. INTRODUCTION

Convolutional neural network (CNN) has been widely used in the field of computer vision in recent years due to its strong feature extraction ability. In 1998, Yann LeCun et al. proposed LeNet-5 network structure[1], which enables the convolutional neural network to train end-to-end.

Object detection has long been a fundamental problem in computer vision, and it has stalled since around 2010. Since the publication of R-CNN in 2014[2], the situation of object detection has changed from the original traditional manual feature extraction method to feature extraction based on Convolutional Neural Network.

First, the core problems to be solved by object detection are:

1. The target may appear anywhere in the image.
2. Goals come in all sizes.
3. Targets can come in many different shapes.

Object detection based on the convolutional neural network has surpassed the traditional object detection methods and has become the mainstream method of object detection. According to the usage of the convolutional neural network, object detection based on the convolutional neural network can be divided into two categories: category-based object detection based on convolutional neural

network and regressive object detection based on convolutional neural network. This article will focus on the first category.

The object detection algorithm based on a convolutional neural network can also be called a two-stage detection algorithm. Traditional object detection methods include preprocessing, window sliding, feature extraction, feature selection, feature classification, post-processing, etc. While the convolutional neural network itself has the functions of feature extraction, feature selection, and feature classification. Therefore, the candidate regions generated by each sliding window can be directly binary classified by the convolutional neural network to determine whether it is the object to be detected.

Compared with the six steps of traditional object detection, the taxonomic neural network object detection has only three steps: window slippage, generation of candidate regions (Region Proposals), classification of candidate region images, and post-processing.

These researches mainly focused on how to improve the feature extraction ability, feature selection ability, and feature classification ability of the convolutional neural network to improve the accuracy of image recognition.

Typical representatives of such algorithms are R-CNN system algorithms based on the regional proposal, such as R-CNN, Fast R-CNN[3], Faster R-CNN[4], and Mask r-CNN[5].

2. APPLICATIONS

Object detection has been applicated in many fields. Among them, face detection[6], pedestrian detection[7], vehicle detection[8] are widely studied.

2.1 Face Detection

The key task of face detection is the relative position of the face in the image. The output information is not only the coordinate information of the face matrix in the image, but also may contain information such as Angle.

In the field of computer vision, face detection is a hot topic that has been deeply studied. In the process of face recognition, face detection is the first and most important step. It has great application value in man-machine interaction, security, witness comparison and so on.

2.2 Pedestrian Detection

Pedestrian detection is a typical target detection problem, which plays an important role in video surveillance, pedestrian flow statistics, and automatic driving.

The goal of pedestrian detection is to accurately find all the pedestrians in the image or video frame and define them with rectangular boxes.

In self-driving technology, intelligent video surveillance, traffic statistics such as applications, the pedestrian detection has great research value, different from general object detection at the same time due to the human body, typically do not have a specific position, shape and color, and its appearance by dress, observation Angle of view, and keep out the influence of factors such as light, in the field of computer vision, the pedestrian detection is a challenging research topic.

2.3 Vehicle Detection

Vehicle detection plays a very important role in the intelligent traffic system, traffic video surveillance, and automatic driving.

In the vehicle detection and identification system, vehicle detection technology is usually required for vehicle flow statistics and automatic analysis of vehicle violations. In automatic driving, the first problem to be solved is to determine the location of vehicles around to ensure driving safety.

3. MODELS

3.1 RCNN

The application of the convolutional neural network greatly improves the effect of object detection[9]. The sliding window has been used in the traditional object detection algorithm, the detection occurred each move of the sliding window. The information on adjacent windows is highly overlapped, and the detection speed is slow.

When R-CNN uses Selective Search to select the candidate area for detection, which reduces the information redundancy and improves the detection speed. Then the convolutional neural network was used to extract features from these candidate regions.

There are four steps in R-CNN algorithm:

1. Regional Proposal: 1,000 to 2,000 candidate regions are generated from an image by the Selective Search algorithm.
2. Feature extraction: CNN was used to extract the features of each candidate region
3. Classification: Put the features of each candidate region into the classifier SVM to obtain the classification result of the candidate region
4. Regression: The eigenvectors of the candidate region are put into FCN to obtain the information of coordinates of the corresponding position

3.2 Fast RCNN

Fast R-CNN is an improvement based on R-CNN. In Fast R-CNN, the author points out several shortcomings of R-CNN:

1. The training requires several processes, such as parameter adjustment, SVM establishment and boundary regression.
2. In the process of SVM training and boundary regression, the features obtained from the candidate box of the image must be written to disk, which takes up a lot of memory and GPU.
3. During testing, features need to be extracted from each test image candidate box. Detection using VGG16 requires an average of 47s per image.

To address these shortcomings, the authors of R-CNN proposed a new Fast R-CNN architecture: Firstly, input the images and multiple regions of interest (RoI) into the fully convolutional network. After the polling layer, each RoI is formed into a fixed-size feature map and then mapped to feature vectors through fully connected layers. Each RoI of the network has two output vectors: Softmax probability and each type of bounding box regression offset.

Compared with R-CNN, Fast R-CNN takes the whole image as the input, which eliminates the overlap and redundancy between the candidate boxes and makes feature extraction faster. Meanwhile,

taking the RoI feature vector as an input to the two fully connected layers eliminates the need for disk storage features and is also faster.

3.3 Faster RCNN

Before Faster R-CNN, we used a series of heuristic algorithms to generate candidate regions based on Low-Level features. There are two problems:

1. The credibility of the generated RoI is low. Generating a large number of invalid areas will result in a waste of computing power, and less generating areas will miss important information.
2. The algorithm for generating candidate regions runs on the CPU, and the training process is on the GPU. Cross-structure interaction will inevitably lose efficiency

To solve these two problems, Ren et al. proposed a concept of Region Proposal Networks, which uses neural networks to learn by themselves to generate candidate regions.

This method solved the two problems above at the same time. The neural network is now able to learn more high-level, and abstract features and the reliability of the candidate regions is greatly improved. RPNs and RoI Pooling share the previous Convolutional Neural Network-Embed RPNs into the original network, and the original network and RPNs are predicted together, which greatly reduces the number of parameters and the prediction time.

At the same time, Faster R-CNN introduces the concept of Anchor in RPN. Each sliding window position in the feature map will generate k anchors, and then judge whether the image covered by the anchor is the foreground or the background, and at the same time return to the fine position of the Bbox, the predicted Bbox More precise.

3.4 Mask RCNN

Faster R-CNN has achieved very good performance in object detection, and Mask R-CNN goes one step further: get pixel-level detection results. For each target object, not only its bounding box is given, but also whether each pixel in the bounding box belongs to the object is marked. Mask R-CNN turns RoI Pooling into RoI Align and then outputs one more branch to predict whether each pixel belongs to the mask of the target object.

In 2017, Kaiming He et al. proposed Mask R-CNN and won the ICCV2017 Best Paper Award. The author points out that Faster R-CNN rounds the size of the feature map when doing down-sampling and RoI Pooling. This approach has basically no effect on the classification task, but it will have a certain impact on the detection task and semantic segmentation. The precision impact of pixel-level tasks is even more serious.

For this reason, the author does not use rounding operations but fills in pixels at non-integer positions through bilinear differences. This makes there is no position error when the downstream feature map is mapped to the upstream, which not only improves the target detection effect but also enables the algorithm to meet the accuracy requirements of the semantic segmentation task.

4. SUMMARY

This article mainly summarizes the main applications in the field of object detection and classic algorithm models. In terms of applications, it mainly includes face detection, pedestrian detection,

and vehicle detection. In terms of models, this article focuses on selecting four models of the R-CNN series, namely R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN.

Based on object detection, researchers successfully designed the instance segmentation Mask RCNN. RCNN is the cornerstone of object detection in recent years. Many researchers are analyzing and improving R-CNN. We look forward to seeing more excellent models in this promising field.

REFERENCES

[1] Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), pp.2278-2324.

[2] Girshick, R., Donahue, J., Darrell, T. and Malik, J., 2014. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition,.

[3] Girshick, R., 2015. Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV),.

[4] Ren, S., He, K., Girshick, R. and Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), pp.1137-1149.

[5] He, K., Gkioxari, G., Dollar, P. and Girshick, R., 2017. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV),.

[6] Verschae, R. and Ruiz-del-Solar, J., 2015. Object Detection: Current and Future Directions. Frontiers in Robotics and AI, 2.

[7] Liu, T. and Stathaki, T., 2018. Faster R-CNN for Robust Pedestrian Detection Using Semantic Segmentation Network. Frontiers in Neurorobotics, 12.

[8] Song, H., Liang, H., Li, H., Dai, Z. and Yun, X., 2019. Vision-based vehicle detection and counting system using deep learning in highway scenes. European Transport Research Review, 11(1).

[9] Fang, Lu. and He, Hang., 2018. Survey of object detection algorithms. Computer Engineering and Applications, 2018, 54(13), pp.11-18.