COURSEWORK2 – OBJECT LOCALISATION (PART 2: INDIVIDUAL REFLECTION ESSAY)

CMT307 - Applied Machine Learning

Group G9

Tianbai Peng

C1981034

1. Introduction

Nowadays, object detection techniques have become more and more popular for distinguishing objects in images and videos. For target detection, the most popular target detection algorithm framework is RCNN, that is Region - Convolutional Neural Networks. As far as the RCNN algorithm is concerned, it has evolved for many generations, from the initial RCNN to the Fast RCNN that uses regression, to the Faster RCNN with a high recall rate .Here the Mask R-CNN which is very similar to Faster R-CNN is a very good solution which provides us with a method for object detection and instance segmentation. We can see that Mask R-CNN integrates many previous excellent research results and it has been widely accepted in industry and in academia.

In this project, we focus mainly on the theory and implementation of Mask R-CNN for object detection. We start by performing dataset analysis and end with outlining the numerical results made by our trained model. We believe the edibility and accuracy of our model meets all the requirements of a desired object detection model.

Personally, in this object detection project, I actively participated in all aspects of the program development process, mainly in data preprocessing, and part of the implementation of Mask R-CNN Model. In this process, I not only learned how to analyze the data set, but also had a certain understanding of The PASCAL Visual Object Classes Challenge, and learned how those maestros used deep learning technology to fit the model with a large number of pictures in a high efficiency. On the other hand, thorough developing programs, I not only improved my python programming skills, but also learned more about convolutional neural networks, which laid the foundation for me to enter the algorithm industry in the future. This essay will focus on describing our project from a personal perspective, while discuss my contribution to the group report and to the overall group work.

2. Contribution

At the beginning, the original job I was allocated is the work of implementation. But after our serious thinking we decide not follow it completely because we think it is very hard to work in this way. Thus, each member involved a bit of each original work and finally finish this project together. In this way, I did a lot of small things for the team, such as communication, search information and writing code, etc., but here I will list the mostly job I did for our project:

1. involve implement the Mask R-CNN training code
2. preprocessing the dataset (convert annotation file)
3. read some articles for group and help write group report

2.1. Implementation

The main thing that I did is joining a part of implement of model train. Training the Voc2012 Dataset is not very easy with the Mask_RCNN Framework at the beginning, thus I start by carefully reading it's README.md file and learn from balloon.py sample which covers the process starting from annotating images to training to using the results in a sample application. In summary, to train the model on Voc2012 dataset by Mask_RCNN Framework we need to extend two classes: Config and Dataset.

Config class contains the default configuration. I create a class called mrcnnConfig extend it and define some important attributes like config name, number of classes, number of training steps per epoch and some learning information.

Dataset provides a consistent way to work with any dataset. I create a class called people extend it and it allows us to load Voc2012 datasets for training without having to change the code of the model. In this class I added some object class of image and finally add each image data with corresponding annotation.

Firstly, I coded these and train the model in the local device:

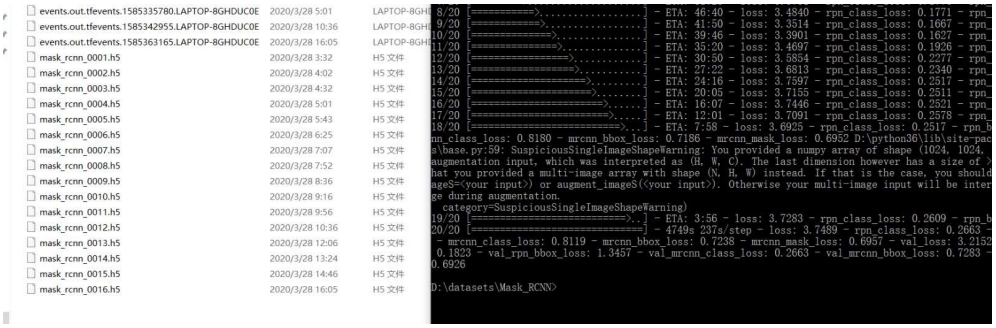

Figure 1 – Mask–RCNN model training result

But the accuracy was not very good, the loss is around 3. After our discuss, Alexandre had also managed to get a Masked RCNN model to work and he decided to run it on the google colab, after he had tried every kind of model parameter finally, we got loss:0.75 which is much better.

Alexandre really has a deep understanding of CNN. I learned a lot from his code when we were developing. Whether it is the choice of optimizer, continued fitting on appropriate weights, or the construction of Neural Network Architecture in M-RCNN, all of those things can greatly improve the accuracy of the trained model. I think the most important thing I learned from him is that the field of neural network is really where dragons hide and tigers crouch! There are too many factors could improve the accuracy of the model, and each of them is worth for us spend much time to research and learn. Because of this, we have to calm down and learn more.

2.2 preprocessing the dataset

Before we start training the model, it's very necessary to preprocess the VOC2012 dataset, however we cannot analyze the data-set until the contents be extracted from annotation files. After analysis, we found that these annotations are .xml files. Each xml file contains many tags indicating various information in the picture, such as filename, size, height, width, and other objects information. At the beginning my opinion is convert these xml file to json because this is more in line with the specifications. Hence, I wrote a script for converting those xml annotations to a json file, which according to the form of ViA. The principle is traversing each tag in the .xml file and extracting the key information then to dump them into a new json file. the structure of this json file looks like below:

Figure 2 – convert of annotation file

But in the end Alexandre decided to convert the .xml file to a .csv file, because after our careful consideration, we think that using the pandas module to read the .csv file would be more easily, and the structure of the final .csv file is roughly the same as the previous json file.

2.3 help write group report

I also participated in writing of the group report. The parts that I wrote are introduction, literature review, and conclusion. I write the draft first, and then other members can revise them freely. In the process of writing the report through consultation and cooperation with other group members, I not only improved my writing skills, but also strengthened the friendship between the teams

And before starting to write a report, I read several useful articles, which is mainly about Deep Neural Networks, Faster R-CNN and Mask R-CNN, these articles are very helpful for literature review, and also offered us some good guidance for our program development.

3. Conclusion

This is a very good deep learning project which provide me a chance to understand the basic knowledge about Mask-RCNN and basic thing about dataset analysis and model training. At the same time, it has also greatly enhanced my teamwork ability. Thorough doing the project, I consider that there still be a lot of points need me to master, especially the data descriptive analysis part and still need to think how to improve the model accuracy and understanding every kind of calculations deeper. To this end, my next step is to continue learning from this project and explore more unknown information from some classic deep learning models for getting a algorithms job offer in future.