0 Introduction

Nowadays, object detection technique has become more and more popular for distinguishing objects in image and video, which is the most effective solution for the ImageNet Object Localization Challenge. Object detection in images is usually split up into two sections, classification, which concerns itself with identifying what each object in an image is and localization which is the task if locating the objects in the image. In this project we focus mainly on the localization aspect, however, our Mask R-CNN models used will perform both parts of object detection. Object detection is really a challenging because it requires the correct detection of all objects in a picture while also need segment each instance very precisely. Here Mask R-CNN is a very good solution which provides us with a method for object detection and instance segmentation, which is built of faster R-CNN, yet constructing the mask branch properly is critical for good results , its feature extraction uses the ResNet-FPN architecture, and an additional Mask prediction branch is added. we can see that Mask R-CNN integrates many previous excellent research results and it has been widely accepted by Industry and academia. Therefore, it is necessary for us to learn and practice Mask R-CNN in this era.

In this project, we focus on the basic use of Mask R-CNN after carefully analyze this task and pre-process the dataset, then train and test the model, and finally present a high accuracy model and precise visual results. Our models can also run very quickly on a CPU. We believe the flexibility and accuracy of our model meets all requirements of

the work and will benefit and ease future research on object detection.