**6. Literature review / Related work**

**Introduction**

The aim of our project was to create a convolutional neural network model which could detect and locate certain objects in images. A common paradigm to address this problem is to train object detectors which operate on a sub-image and apply these detectors in an exhaustive manner across all locations and scales, suggests by Christian Szegedy, Alexander Toshev and Dumitru Erhan(2013). Therefore, before start the project, it’s the high time for us to focus some important articles which closely relevant to the object detection problem to find out which detector is best to train and use. The first one which is the most important one describes the method of developing a Mask r-cnn by introducing how to getting started and some step by step detection which from Matterport’s Mask r-cnn implementation (2017). And Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun(2015) tells a good introduction about the theory of Faster r-cnn towards real-time object detection which give us some ideas about the project. The third provides an example of the use of Mask r-cnn for real-world applications which is from Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross B. Girshick(2017). At the same time, Google ’s Inception series models also perform well on this task. For example，Chollet François(2016) gives the method to develop the Inception model, which can be called as Xception model. All of these articles are talking about different viewpoints around convolutional neural network and how can neural network work for our VOC2012, which is worthy of referencing in our group project.

**Deep Neural Networks for Object Detection**

To do object Detect we may need to think about using DNN. Using deep learning and DNN for target detection is unlike traditional shallow neural networks, deep neural networks can learn more complex models, not only for classification, but also for accurate positioning .You can predict a bbox in a given image, and more precisely, it will also generate a binary mask in the box, which is proposed by Christian Szegedy, Alexander Toshev and Dumitru Erhan(2013). Therefore, for the problem of target detection, it is necessary to use a deep neural network. And in past experiments, deep neural networks have achieved good results in the VOC challenge, Christian Szegedy, Alexander Toshev and Dumitru Erhan has shown this approach has excellent performance of VOC2011, and they also trained on the larger VOC2012 training set, finally show state-of-the art performance on most of the models.

**Advantage of the Mask R-CNN**

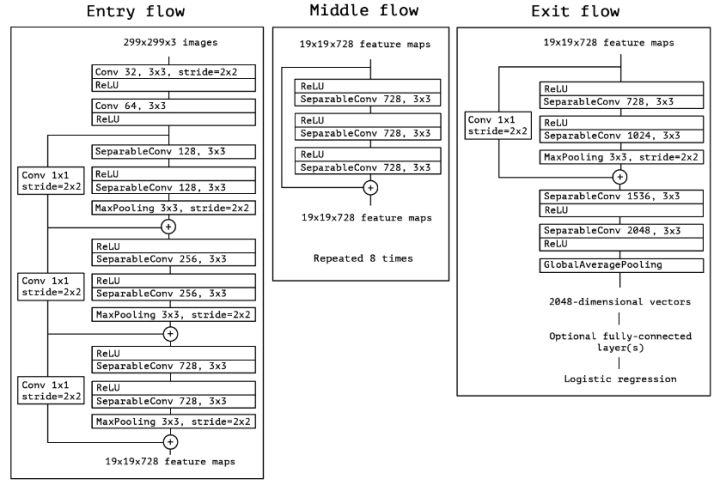
It has been determined that deep neural networks can be used well for object detection. However, this problem is not so easy to solve. The size of the object varies widely, the angle of the object is placed, the pose is uncertain, and can appear anywhere in the picture, and moreover, objects can also be of multiple categories. There are too many methods based on DNN for solving these problems, what kind of object detection algorithm is perfect for our VOC2012 dataset? Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross B. Girshick(2017) think Mask R-CNN adopts the same two-stage procedure, with an identical first stage (which is RPN). In the second stage, in parallel to predicting the class and box offset, the multi-task learning of both the bounding box detection and object segmentation using masks, the precision of object detection is improved. whose performance made Mask R-CNN is an excellent choice for performing object segmentation. What’s more it is very similar to Faster R-CNN(Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun(2015)), in the field of Faster R-CNN, Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun has done much experiments on PASCAL VOC , it shows high performance and low cost on each layers from their table and figure. In both of their studies, they show that Mask R-CNN is very fast, with better speed/accuracy tradeoffs achieved.

**Implementation of Mask R-CNN**

Their article does not explain in detail how to implement mask r-cnn. Matterport(2017) implements a Mask R-CNN framework for Object Detection and Segmentation, which provide us a lot of visualizations and allow running the model step by step to inspect the output at each point. Matterport also give some example projects that made by extension Mask R-CNN model with other datasets. Our project is mainly based on the Mask R-CNN library

by Matterport. But at the same time, it is inseparable from the theory from Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross B. Girshick (2017)

**Advantage of the Xception**



From this architecture, we can see that Xception absorbs the residual connection of ResNet and the global average pooling of the network in the network. The special one in the picture is SeparableConv, which is called depth separable convolution. In simple terms, it is to convolve each input channel separately, and then use 1x1 convolution to project the output of each channel. The overall design of Xception is very simple, elegant, and easy to modify, and the entire model is very lightweight, so it has great application potential.( Franc¸ois Chollet, 2016)

**Implementation of Xception**

For example, at the end of 2017, Face ++ ’s Light-Head R-CNN, after replacing ResNet-101 with Xception, although the performance has decreased, but the speed has been further optimized, the speed exceeds YOLO and SSD models known for high speed YOLO and SSD, but this aspect is mainly due to other designs of Light-Head R-CNN). For example, in the Light-Head R-CNN of Face ++ at the end of 2017, after replacing ResNet-101 with Xception, although the performance has decreased, the speed has been further optimized, and the speed exceeds the models known for high speed such as YOLO and SSD. At the same time, the performance is of course Also surpassed YOLO and SSD.( Zeming Li.et.al, 2017)