

Assignment 4 - Object detection&Semantic segmentaion *

Xiang Yu-Ye

Mechanical Engineering
National Cheng Kung University
Tainan City East Dist. Daxue Rd., Taiwan
N16114445@gs.ncku.edu.tw

Abstract -使用 VOC 2007 及 ADE20K 兩個 Dataset 作為 Object detection&Semantic segmentation 兩個 task 的 Train data，本次使用 Mask R-CNN 並用 ResNet50 作為 backbone。

Index Terms – Deep Learning、Object detection、Semantic segmentation、Mask R-CNN

I. INTRODUCTION

Object detection 與 Semantic segmentation 是 Downstream Tasks 熱門的研究項目。Object detection 為將想要偵測到的物件用矩形框框起來，常見的 model 架構有 Faster R-CNN[1]、FPN[2]。Semantic segmentation 則是將想要的物件用顏色標註起來，常見的 model 架構有 U-Net[3]、FCN[4]、Deeplabv3[5]。而本次作業要做的是使用一個模型結合 Object detection&Semantic segmentation，Mask R-CNN[6] 就是一個可以做 instance segmentation 的 model，但由於本作業不是要做 semantic segmentation，故會更改 model 中的 roi_head 來做我們的 task，Fig. 1. 為 semantic segmentation 及 instance segmentation 的差別。

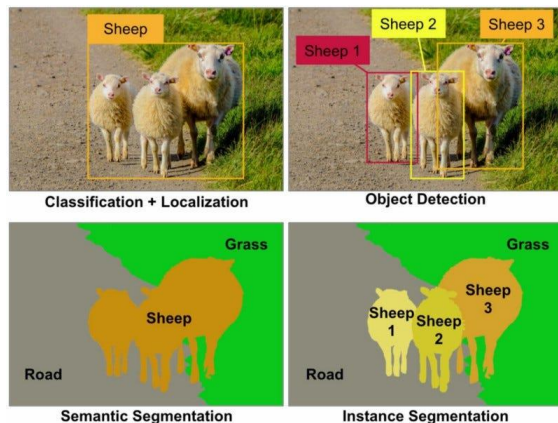


Fig. 1 Object Detection&Semantic Segmentation&Instance Segmentation

II. METHOD

A. MASK R-CNN

Fig. 2. 為 Mask R-CNN 的架構圖，此架構是改造 Faster R-CNN，主要是將 ROI Pool 改成 ROI Align，將原 ROI Pooling 取整數的部分忽略來增加精度，並增加了預測 Mask 的一個 output。

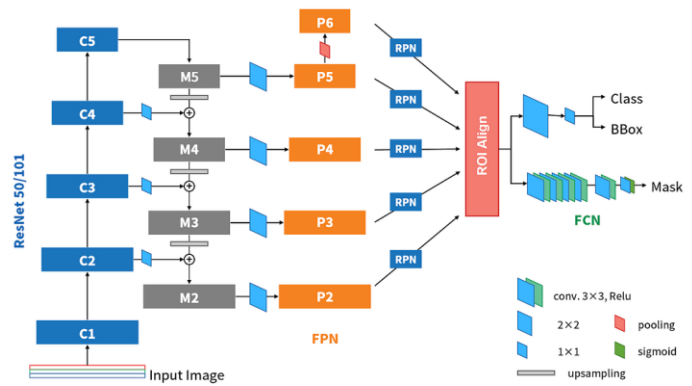


Fig. 2. Mask R-CNN Architecture

B. VOC 2007

VOC 2007 為 PASCAL 競賽對於視覺分類及偵測的一個基準測試，共 9963 張圖片，20 個類別，本作業拿來做為訓練 Object detection 的 Dataset。

C. ADE20K

ADE20K 專門做為 Segmentation 訓練的 dataset，只使用 validation dataset 的 2000 張圖片、150 個類別做訓練。

III. CONCLUSION

在寫 code 的過程中一直遇到就算上網查也無法解決的問題，model 的部分使用了 pytorch 的 maskrcnn_resnet50_fpn 的函式，用在 COCO 上預訓練的 weight。讀取 VOC 2007 為 pytorch 預訓練的 maskrcnn_resnet50_fpn 可以訓練的 dataloader 就一直卡關了，再來是 ADE20K，到最後還是不太知道怎麼把 github 上提供的 code 改成 model 需要的 dataloader，對於 model 的建構及訓練，我的想法是將 VOC_2007 用於訓練 Box 的 Output，而 ADE20K 用於訓練 Mask，但花很多時間研究還是寫不出來。最後想說老師我盡力了，但藉由這個作業，第一次使用 pytorch，感受了與 Tensorflow 不一樣的地方，且了解 Object Detection&Semantic Segmentation 的相關論文與方法，對於沒接觸過的這兩個 Task 有個初步了解。

REFERENCES

- [1] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, "Faster R-CNN", last revised 6 Jan 2016 (this version, v3)]
- [2] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie, "Feature Pyramid Networks for Object Detection", last revised 19 Apr 2017 (this version, v2)].

- [3] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", Submitted on 18 May 2015
- [4] Jonathan Long, Evan Shelhamer, Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation", last revised 8 Mar 2015 (this version, v2)
- [5] Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation", ast revised 5 Dec 2017 (this version, v3)
- [6] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, "Mask R-CNN", last revised 24 Jan 2018 (this version, v3)