HW4

JIE-YI LI
Minn Wu School of Computing
National Cheng Kung University
Tainan City, Taiwan
nn6114069@gs.ncku.edu.tw

Abstract – The goal of this assignment is to design and implement a deep learning model capable of performing both object detection and semantic segmentation simultaneously, using two distinct datasets. Traditionally, instance or panoptic segmentation can be used to learn both object detection and segmentation tasks from the same image. However, unlike these traditional methods, this assignment requires the use of two separate datasets: one for object detection and another for semantic segmentation.

Keywords - Object detection, Semantic segmentation, Dual-task deep learning

I. INTRODUCTION

A. *Object detection*

物件檢測（Object Detection）的目標是在影像或影片中檢測和識別出物件的位置和類別。目前對於許多應用領域，如自動駕駛、監控系統、物件識別等都有廣泛的應用。通常，Object Detection model需要在影像中標示出物體的邊界框（Bounding Box），同時給出每個邊界框對應的類別。檢測通常可以分為兩個主要步驟：區Region Proposal和Object Classification and Localization。

在Region Proposal階段會生成可能包含物件的候選區域，這些候選區域通常是不同大小和形狀的矩形區域，並且被認為可能包含目標物件。

在Object Classification and Localization階段會對每個候選區域進行分類和定位，判斷該區域是否包含物件，並預測該物件的類別和位置。

近年來，深度學習在Object Detection上取得了很多重大突破，特別是CNN的方法。許多深度學習模型，如Faster R-CNN、YOLO（You Only Look Once）和SSD（Single Shot MultiBox Detector）等，都是Object Detection中的SOTA model。

B. *Semantic segmentation*

語義分割（Semantic Segmentation）的目標在於將影像中的每個像素分類為不同的類別。與物件檢測不同之處是語義分割不僅要檢測物件的存在，還要將每個像素分配到相應的類別，實現對影像的像素級別的分類和分割，並能夠捕捉到物體的形狀、邊界和空間關係。這對於許多電腦視覺應用非常重要，例如自動駕駛、圖像分割、地圖製作等。

語義分割通常使用像素級的分類器來實現，它將影像中的每個像素視為獨立的樣本，並將其分類為不同的類別。深度學習模型同樣在語義分割任務中取得很大的進展，特別是基於CNN的方法。如U-Net、FCN（Fully Convolutional Network）、DeepLab等，都已經在許多語義分割dataset上取得了優異的性能。

II. METHODOLOGY

A. *Data*

使用了兩個dataset分別用於Object detection和Semantic segmentation。其中VOC2007是PASCAL VOC（Pattern Analysis, Statistical Modeling and Computational Learning Visual Object Classes）挑戰的一部分，VOC2007包含20個不同的物件類別，如人、車、飛機、狗、貓等。VOC2007的圖像都有一個唯一的ID，並存在JPEG格式的影像檔案中。而在本次作業中VOC2007也被用於Object detection。

另一個ADE20K資料集則在本次作業中用於Semantic segmentation。ADE20K將影像中的每個像素分類為150個不同的語義類別，包含來自不同場景的2,000個高解析度影像，這些影像涵蓋了多種不同的場景類型，包括室內和室外環境、城市和鄉村地區、道路和建築物等。ADE20K中的語義類別包括各種物體、場景元素和背景類別，如人、動物、車輛、植物、家具、建築物等。

B. *Model*

預計在Object detection部分使用Faster R-CNN；在Semantic segmentation 部分使用DeepLabV3+

III. CONCLUSION

A. 針對VOC2007和ADE20K的影像部分已讀為array，label的部分VOC2007已處理完畢
B. 其他部分不及完成

REFERENCES

[1] 上課講義
[2] Chatgpt 3.5