國立臺灣師範大學科技與工程學院圖文傳播學系

碩士論文

應用深度學習演算法之海報文字區域檢測實驗

An experiment with application of deep learning algorithm to detect texts area for poster

研 究 生：盧　聖　侃

指導教授：張晏榕　博士

中華民國 111 年 6 月

摘要

近年來，數位化的廣泛應用也促使了互聯網的發展。伴隨著互聯網技術日新月異，大量的社交媒體和其他應用程式不斷推陳出新，數位圖像已然成為社會中一種主要的資訊獲取來源。在當今資訊量爆炸的社會裡，海報作為生活中最常見的資訊傳達媒介，成為生活中處處可見的藝術表現方式並充斥在現代人的生活當中。若能提出一個檢測方法來辨識海報中的文字區域，不僅能提取海報文字區域作為後續分析的資訊，也能使海報在網路中的更容易被使用者檢索。隨著深度學習的興起，越來越多研究者利用深度學習來完成影像分析及物件檢測。而其中，Mask R-CNN 與 Yolov4 分別代表著 two-stage 與 one-stage 的目標檢測方法，無論是在物件的瑕疵檢測、人臉的偵測、交通路況的偵測等領域都有很好的研究結果。然而，以上大多都是檢測自然場景物件，較少應用在平面設計的領域之中。基此，為了提取海報圖像的文字區域，本研究將訓練 Mask R-CNN 與 Yolov4 兩個檢測方法，分別來對海報圖像文本進行檢測。實驗結果顯示，Mask R-CNN 檢測文字區域的 mAP_{50} 可達 79.0%；Yolov4 檢測文字區域的 mAP_{50} 也高達 85.1%。意味著兩個目標檢測方法都可在海報版面中，定位出海報中文字區域，提供未來作為文字辨識的數據。而對比 Mask R-CNN 與 Yolov4 兩種演算法的輸出結果後，發現 Yolov4 可以更準確地檢測文字區域，並且較不受海報因色彩、文字大小、文字間隔等設計因素影響到檢測結果。

關鍵字：海報版面、深度學習、Mask R-CNN、Yolov4

Abstract

In poster design, designers often simplify and artisticize the information, quickly capture the audience's attention. The text in the poster must be brief and clear to the audience at a glance. If a detection method can be proposed to identify the text area in the poster, it can not only extract the text area of the poster as information for subsequent analysis, but also make the poster on the Internet easier to be retrieved by users.

With the progress of deep learning and the improvement of computer hardware equipment, many researchers also use deep learning to complete image analysis and object detection. Among many object detection methods, Mask R-CNN and Yolov4 represent the two-stage and one-stage object detection methods respectively. Both of them have relatively outstanding performance in accuracy and computational efficiency. It can also be observed in real life that many researchers use this method to solve many problems, such as object defects detection, face detection, and traffic condition detection. However, most of the methods above detect objects in natural scenes, and are less used in the field of graphic design. In order to understand the ability of deep learning in poster layout analysis. In this study, two detection methods, Mask R-CNN and Yolov4, will be trained to detect poster image text respectively. The experimental results show that the mAP_{50} of Mask R-CNN can reach 79.0%; the mAP_{50} of Yolov4 can also be as high as 85.1%. It means that both object detection methods can be able to locate the text area in the poster layout, and provide data for text recognition in the future.

Keywords:, Poster Layout, Deep Learning, Mask R-CNN, YoloV4.

目錄

第壹章、緒論

第一節、研究背景與動機

近年來，數位化的廣泛應用也促使了互聯網的發展。伴隨著互聯網技術日新月異，大量的社交媒體和其他應用程式不斷推陳出新，數位圖像已然成為社會中一種主要的資訊獲取來源。在當今資訊量爆炸的社會裡，海報圖像作為生活中最常見的資訊傳達媒介，是生活中處處可見的藝術表現方式並充斥在現代人的生活當中。海報設計是視覺語言的集中應用與表現形式，品牌標誌、標準字、標準色和其他要素的組合，能夠直接而生動的塑造品牌形象，增強品牌的感召力、感染力和傳播力（李嫣、胡清，2021）。在海報圖像的設計中，設計師將文字、圖像、色彩等元素靈活運用，將資訊簡單化、藝術化，以達到更快速的捕捉受眾眼球（殷建、鄭童，2021）。

文本定位是文本識別的前提，也是文本提取的關鍵（晉瑾、平西建、張濤、陳明貴，2007）。可以將圖像文本大致分為兩類：人工文本（artificial text）及場景文本（scene text）。人工文本是指通過圖像處理工具對圖像進行編輯，對圖像進行人工標註的文本。這類文本的內容意義性極強，對圖像的檢索有重要的作用（楊捷、劉進鋒，2018）。而海報圖像中的資訊大多是設計者為了傳遞資訊並在圖像中進行編輯的文本，因此做為人工文本而言，其所富含的意義在實際上是相當豐富且具有價值。海報圖像中的文本包含豐富、明確的資訊，如果這些文本能被自動提取，對於圖像文本的語意理解、索引和檢索是非常有價值的。而透過深度學習打造的辨識模型，具有精準的文字辨別能力，能夠偵測圖片中的表格與文字，可將所有圖檔轉化成有用的數據資料。將提取出的海報文本和網路使用者所搜尋的關鍵字進行比對，可以用於文件自動審查、電視與雜誌輿情監控等服務，使圖像或影像中的資訊更容易的曝光在大眾眼前。

在現今科技的不斷創新之下，在 1929 年開始有研究學者提出 OCR（Optical Character Recognition）方式，對文本資料的圖像檔案進行分析辨識處理，取得文字區域資訊。不過此種方法對 OCR 的識別正確率依賴性高、容錯性低，而且對文件檔案的圖像質量要求也較高，因此有了更多的改進文本分析的方法出現。在文檔版面分析方面，胡芝蘭、林行剛和嚴洪（2006）提出了一種基於二值化文檔圖像分層密度特徵檢索方法。對圖像進行預處理並提取文本特徵區域，透過文本特徵區域檢測出圖像像素的空間分布信息，

以此找出文字區域。不過對於彩色圖像或是具有複雜背景的版面，在圖像分割上還是有一定的限制。2008 年，徐銳義、吳煒、何小海和楊玉科則採用基於二分投影遞歸算法版面分析並用於名片的文本提取及分割，在傳統的投影算法上加以改進，處理速度加快，可靠性更高。但名片大部分的版面設計上都較整齊固定，若是在版面背景較複雜的環境下，分割效果也會有一定影響。針對印刷版面的藏文文檔，陳圓圓、王維蘭、劉華明、蔡正琦和趙鵬海(2021)以中小學藏文教材文檔圖像為例，對藏文文本圖像進行預處理，利用 ARLSA 算法定位版面元素，進行連通域分析，分割版面，分類文本及非文字區域。在以往的研究中，可以發現到傳統的版面分析方法有著效率低、對複雜版面的處理效果不佳等問題（陳璇、賀建軍、李厚杰、武林秀，2019)。而近年來，深度學習的快速發展也帶給許多研究學者在版面分析中更多的可能性，李翌昕、鄒亞君和馬盡文（2019）就提出了基於特徵提取和機器學習分類模型的方法，在提取版面圖像特徵後，將特徵圖像丟進 CNN（卷積神經網路）進行文字與圖像的分類。Roullet, Fredrick, Gauch and Vennarucci(2019)也為了創建一個強大的數據集來識別龐貝藝術及古蹟，對歷史古蹟圖像書卷進行圖像提取、存檔、分析和分類所有圖像數據。

深度學習的快速發展衍伸了許多目標檢測方法，其中分為一階段（One-Stage）與二階段（two-Stage）兩種分類。顧名思義，one-stage 的目標檢測方法中的目標定位與目標分類同時進行；而 two-stage 的目標檢測方法則會分開來進行。因此 one-Stage 的檢測速度較 two-stage 快，但在檢測精確度方面通常比起 two-stage 則會較為遜色。在 two-stage 方面，Mask R-CNN 是 He, Gkioxari., Dollár, and Girshick（2017）所發表的目標檢測方法，能夠實現目標檢測和實例分割兩個任務。在許多的研究實驗結果也顯示出，利用 Mask R-CNN 對圖像進行分類及區域提取，可以在複雜的背景環境下，有不錯的分割效果。而在 2020 年提出的 Yolov4（Bochkovskiy, Wang, & Liao, 2020）也是目前熱門的 one-stage 目標檢測分法，其 FPS 及 AP(average precision)都不遜色於 two-stage 方法，在現實生活中也經常用於道路的汽車及行人檢測。

自 2012 以來，在神經網路模型的表現能力和 GPU 的計算能力放面取得重大進展。可以將深度學習的成功歸因於:「進階的網路模型的出現」、「演算能力的加強」、「大規模標記數據類別的可用性」(Sun, Shrivastava, Singh, & Gupta, 2017)。深度學習已證明具有學習圖像特徵的出色能力，並已廣泛用於目標檢測。但在過去大部分的研究中，卻發現

較少有利用深度學習來做為海報文本檢測的工具，因此本研究將嘗試使用 Mask R-CNN 與 Yolov4 兩種目標檢測方法，針對海報圖像中的文本進行辨識訓練，並提取海報文字區域的定位資訊。

第二節、研究目的與問題

根據前述的研究背景與動機，為了使海報圖像中的文本更容易被網路索引及提取。因此本次研究的目的將著重於海報文字區域的目標檢測，引用深度學習的概念，提取海報文字區域位置。從前人的研究中可以發現深度學習具備著學習圖像特徵的出色表現，而 Mask R-CNN 與 Yolov4 也是許多研究者用來做目標檢測的工具，並且在其他領域研究上都具有優秀的檢測水準。因此本研究嘗試使用兩個方法作為研究工具，將研究目的定為探討深度學習對於海報圖像中之文字區域檢測準確度。

根據研究目的所提出之研究問題如下：

一、深度學習對於海報文字區域檢測之準確度為何？

二、不同演算法對於海報文字區域檢測的差異？

第三節、研究範圍與限制

一、研究範圍

深度學習的範圍相當廣泛，根據研究的目的可以有多樣化的運用方式，像是醫學、農業、交通、影像等，都是可以利用深度學習的領域範圍，本研究將以 Mask R-CNN 與 Yolov4 作為主要研究工具，其他目標檢測方法將不列入本研究範圍內。

本次研究主要目的在於提取海報圖像中的文字區域位置，提供海報文本資訊內容並使其更容易在網路中作為索引資訊。利用 Mask R-CNN 與 Yolov4 找出版面中的文字區域，因此其他類型的海報設計元素將不列入本次研究的檢測範圍。

二、研究限制

本研究所使用的深度學習目標檢測工具為 Mask R-CNN 與 Yolov4，並分別利用兩種方法提取海報圖像文字區域。基於研究範圍，本研究的研究結果與結論無法含括至其他方法的成效。

第貳章、文獻探討

第一節、海報圖像版面

一、海報版面設計

版面設計是指在規定的平面範圍內，將文字、圖像再加上色塊、線條、點等設計要素進行有創意的剪輯、組合、編排，以表達設計者的設計意圖（張勁，2014）。在一定程度上對視覺的傳達具有重要的影響，甚至可將視覺傳達視為版面設計的一部份，掌握版面的設計知識及理論，可以提高對於文字、圖像等處理能力和版面的製作能力。而版面設計主要的設計內容就是對於文字與圖像的設計編排，並且向受眾傳達相關訊息，在設計過程中透過藝術形式賦予文字及圖像美感，提升其視覺效果（張芳，2016）。平面設計遍布於生活的每個角落，透過形式各異的排版風格吸引著人們的關注。版面設計在平面設計中發揮著重要的作用,提高了版面的視覺衝擊力。在設計時不僅應該注重美的展現,還需要提高視覺傳遞的速度。充分發揮版面設計的作用，引導讀者的視線，並透過藝術型態表現（李寧，2020）。在大數據及互聯網蓬勃發展的時代背景下，人們的視覺體驗不斷攀升，各式各樣的訊息衝擊著讀者的視覺及情緒，讀者大多不願意花費太多時間來閱讀整個版面內容，因此版面設計逐漸朝向圖文並重的發展模式，對於版面的內容進行合理編輯，加強設計的創新能力，使版面更加豐富多彩。文字與圖像為版面中最主要的設計元素，在版面設計中佔了很大成份的比例，其字體的設計與大小標題的調整可使整體更具有層次分明的效果；而圖像的大小、形狀和位置，也可使版面起到均衡的作用，使得整個版面適當得體、主次分明、輕快舒適（馬雲峰，2016）。版面設計設計後所呈現出的效果會直接對受眾產生影響，好的版面能夠吸引更多受眾，反之，則可能會失去大量客群。版面設計的原理並沒有固定的定律或是公式，設計的風格會隨著不斷實踐的過程中變化，一個好的版面設計，是令讀者感受不到設計的存在，但卻能突出主題，提高訊息傳遞的效果（張勁，2014）。

二、海報設計要素

在進行版面設計時，必須要確保好一幅作品所要呈現的主旨，並了解讀者受眾的視覺需求，因此我們可以從文字與圖像這兩個版面的構成要素著手。文字作為版面設計最重要也是最基礎的組成元素，是負責傳達訊息的主要載體。對於文字內容進行適當的編排及配置，可以使整體版面更具美感及形式感，從而引導讀者對編輯內容及訊息進行閱讀（張芳，2016）。李寧（2020）提到文字不僅彰顯了藝術設計的視覺效果，也是文字情感的傳遞要素，在版面設計中主要是關注文字的大小、字體、藝術形式、位置、所占面積等，文字設計的內容要貼合文章主旨，與內容相輔相成，增加受眾興趣。荊奕君（2018）也提到版面設計中要注意字體及大小的一致性，讓讀者感受版面的整潔與統一，大小標題可利用醒目加粗體的字樣進行編輯，方便讀者快速獲取資訊。圖像相較於文字在敘述過程中更加形象、生動、具體，能夠充分表達出要傳達的訊息，是人類所使用最早的訊息傳達方法（張芳，2016）。圖像可以直觀傳遞訊息，形象的表達情感。巧妙的運用圖像並注意其組合方式，可保證版面的結構層次和情感傳遞效果並增強版面的活力（李寧，2020）。海報中最佳的設計理念是多圖像、少文字。隨著訊息量暴增的時代，設計者往往會通過設計與主旨相符的圖像來達到更直觀有衝擊力的傳播效果以吸引受眾（殷建、鄭童，2021）。在新媒體崛起的世代中，每天都有大量的訊息灌輸到每個人的生活，人們不願意再花費過多的時間進行閱讀。因此適當的插入圖像可以讓讀者在短時間內獲取更多資訊，讓不同年齡層都能輕鬆閱讀，而圖文的結合更能夠吸引讀者的注意力（荊奕君，2018）。

三、海報圖像文本

海報圖像中的文本識別不同於文檔圖像中的文本識別，文檔圖像中的文本一般大多是白底黑字，背景顏色也會較為單一，因而在文檔文本的識別已經是許多研究者研究的領域，是較為成熟且達到實用的要求。相對於文檔圖像文本，海報圖像中的文本背景或是文字本身色調及樣式都較為複雜且多樣，沒有固定的格式。海報中的文本是傳達訊息的重要載體，設計者根據海報的主題內涵來設計字體，發揮創意並凸顯文字個性與語言效果，提升視覺表現（陳慧姝，2019）。為了引導讀者對海報圖像內容與資訊進行閱讀，可以對文字進行適當的排版，不但會使版面更具美感和形式感，也非常符合人們的審美觀點，方便於讀者接受訊息（張芳，2016）。

Van Dalen, Gubbels, Engel, and Mfenyana 在 2002 年對海報設計提出了幾個要點：

1.文本標題或關鍵信息需足夠清晰且容易抓住讀者目光，應避免呈現過多資訊反而失去焦點，資訊應盡可能有限。

2.盡量避免使用色彩豐富的背景，海報雖需要看起來漂亮且吸引人，但不應使讀者分散閱讀資訊的注意力，良好的顏色搭配會是最好的。

3.文本字體盡量統一，同一張海報最多使用到兩種不相同的字體樣式。

4.文字的大小應足夠使全年齡的讀者都容易閱讀，縮小文字雖可將塞入大量資訊，但也會造成讀者閱讀負擔，文字之間應避免過長的空間，此外，將文本對齊會使讀者在閱讀時感到舒適。

5.將大約 50%的海報版面空間留白。

Keely（2004）也強調若要設計一張易於閱讀的海報，文本字體的大小應謹慎的選擇。比起使用深色背景搭配淺色文本的海報，深色文本配上淺色背景的海報設計也相對使讀者更舒適的閱讀（Keely, 2004; Nemcek, Johnson, & Anderson, 2009）。此外，相較於居中

對齊的文本，靠左對齊的文本反而容易閱讀，因為居中對齊的文本常留下參差不齊的邊緣，增添眼睛疲勞（Berg, & Hicks, 2017）。

由上述文獻中可以發現，文本若要吸引讀者注意，整體海報文本的文字字體、文字大小、顏色對比、背景色調、文字間隔及對齊方式，都是一張海報的成功要素。而海報中，文本大多會結合圖像進行設計，透過文字與圖像的搭配吸引讀者的目光，因此海報圖像通常不會是以白底黑字的形式呈現給讀者閱讀，豐富的色彩結構甚至可能影響到目標檢測方法的檢測結果。

第二節、深度學習演算法 Mask R-CNN

Mask R-CNN（Mask Region based Convolutional Neural Network）是 He, Gkioxari, Dollár & Girshick（2017）發表的目標實例分割（Instance Segmentation）。基於 Faster R-CNN（Ren, He, Girshick, & Sun, 2015）的演算框架進行擴展，使用深度殘差神經網路（ResNet)取代原本的 VGG 網路並搭配 FPN(特徵金字塔網路，Feature Pyramid Network）用以加強圖像特徵提取的能力。引入 RoI Align 取代 Faster R-CNN 中的 RoI Pooling，改善了原本檢測框偏移的問題，提升了整體模型的準確率，並多增加了一條遮罩預測(Mask ）任務的分支，用於預測目標物件在圖像上的遮罩區域（Mask）範圍。目前，AlexNet、ZF、VGG、GoogleNet 和 ResNet 等網路模型為深度神經網絡的主要模型。深度殘差網路（Residual Network, ResNet）為 2016 年 He, Zhang, Ren and Sun 提出的神經網路，其中加深神經網絡的深度雖然可能導致更高的精度，但模型訓練和檢測速度將會降低且導致更高的訓練誤差。而殘差神經網路（ResNet）的結構不會新增模型參數，並有效地緩解了梯度消失和訓練退化的困難，提高了模型的收斂效能(Yu, K. Zhang, Tang, & D. Zhang, 2019)。Mask R-CNN 的演算框架中主要同時完成目標檢測、目標分類和目標分割任務。隨著深度學習的發展，Mask R-CNN 被提出作為一種實例分割演算法，在影像分割和目標識別方面表現出良好的效能。圖 2-1 為原文 Mask R-CNN（He et al , 2017）擷取的網路架構圖。

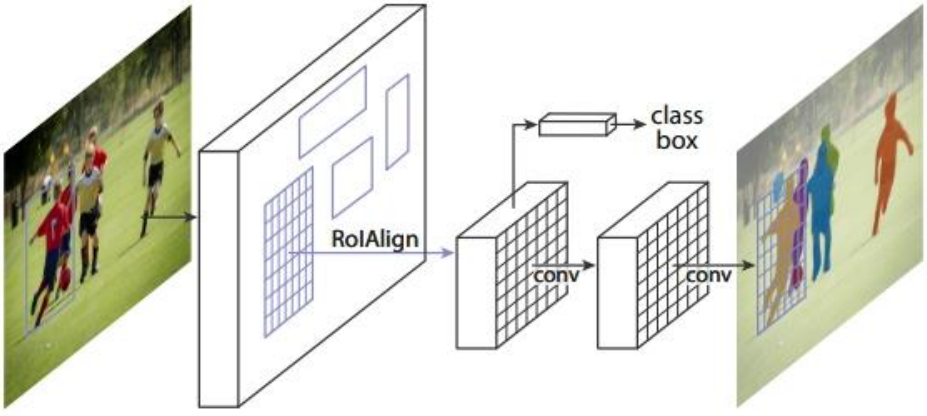


圖 2-1：原文 Mask R-CNN 論文中擷取的網路架構圖

一、語義分割與實例分割

電腦視覺（Computer Vision，CV）並不是一個孤立的領域，其與深度學習的關係是由淺推理的過程中透過更多的研究方法來達到更精細推理的一個自然步驟。語義分割（Semantic Segmentation）為電腦視覺(Computer Vision，CV)領域的一個重要應用，是實現更精細的影像識別任務的過程，針對圖像中各個像素點做推斷及分類，並把相同類別的物件劃分在同個區域。而實例分割更是語義分割進一步的演進（Garcia-Garcia, Orts-Escolano, Oprea, Villena-Martinez, Martinez-Gonzalez, & Garcia-Rodriguez, 2018）。實例分割（Instance Segmentation）是結合了目標檢測與語義分割的影像辨識技術，相對於目標檢測，實例分割可以精確到目標的邊緣；相對於語義分割，實例分割可以標註出圖像上同一物件但不同個體。圖 2-2 為語義分割與實例分割的示意圖。其中，語義分割是將場景中物件包括背景都切割出來並將相同物件劃分到一起；實例分割除了分割物件外，不同的實例也會賦予不同顏色。

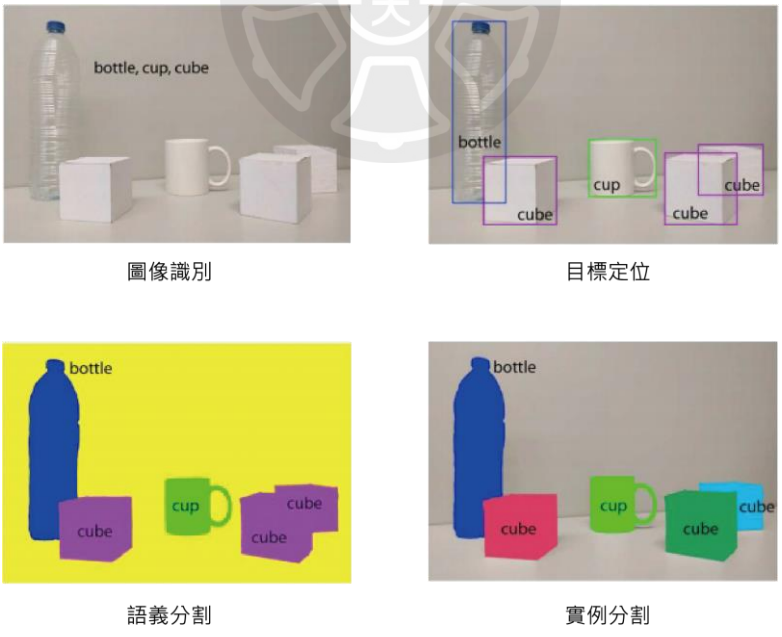


圖 2-2：語義分割與實例分割

影像實例分割已經成為機器視覺研究中一個相對重要、複雜且富有挑戰性的領域。以預測物件類別標籤和物件實例遮罩為目標，對不同影像中的不同實例物件的類別進行

定義。實例分割的發展很大程度上幫助機器人、自動駕駛、監控等領域。隨著深度學習的發展及卷積神經網路（CNN）的提出，影像分割的方法相繼而出，其影像分割的效果精度也迅速的提高，較為著名的方法有 R-CNN、Fast R-CNN、Faster R-CNN、Mask R-CNN 等（Hafiz & Bhat, 2020）。

R-CNN（Regions with Convolutional Neural Networks）為 2014 年 Girshick 提出的目標檢測方法，首先會通過 Selective Search（Uijlings, Van De Sande, Gevers, & Smeulders, 2013）的運算，提取約 2000 個的待檢測區域（Region Proposal），之後輸入到卷積神經網路（CNN）進行特徵擷取，再將特徵圖像送入各類別的 SVM（Support Vector Machine，支持向量機）分類器進行目標分類，最後對目標定位的邊框進行回歸。不過 R-CNN 的缺點為耗費時間的運算過程，三個模組（Selective Search、CNN、SVM）須分別訓練且訓練時對於存儲空間的消耗巨大。因此，2015 年 Girshick 再度提出 Fast R-CNN 來解決並優化了先前 R-CNN 運算效能的問題。除了 Selective Search 模組，其餘部分被整合在一起訓練，取代分別訓練的缺點，只採用一個 CNN 網路對輸入圖像進行特徵提取，並利用 ROI Pooling 固定特徵圖像的尺寸，接著輸入全連接層進行目標分類及目標邊框回歸。

Fast R-CNN 雖然優化了 R-CNN 的缺點，但 Selective Search 模組的運算依然相當耗時。有鑑於此，Ren et al.（2016）在基於 Fast R-CNN 的基礎上進行優化，衍生了 Faster R-CNN 的演算框架，取代 Selective Search 冗長的運算模組，加入了 RPN（Region Proposal Network）來獲取特徵圖像中的 ROI（Region of Interests），並一樣利用 ROI Pooling 調整特徵圖像尺寸，作為之後目標分類及邊框回歸的輸入，優化過後的準確率也大幅地提升。

Mask R-CNN 如同前一節所述，繼承先前 Faster R-CNN 架構的影像實例分割演算法，是電腦視覺領域中的集大成者。除了目標檢測之外，更新增了一條遮罩（Mask）回歸，在目標識別準確率及演算速度都較先前的網路架構更為優秀，且更容易進行模型訓練（Hafiz & Bhat, 2020）。

二、Mask R-CNN 網路架構

Mask R-CNN 繼承 Faster R-CNN 的框架皆採用 Two-State 的網路結構，包含了兩個階段進行圖像的解析。第一階段的任務是負責目標圖像的特徵提取，圖像匯入模型後，首先通過基於殘差神經網路（ResNet）與 FPN（Lin, Dollár, Girshick, He, Hariharan, & Belongie, 2017）搭配而成的主幹網路生成特徵圖像，而緊接著的 RPN（Region Proposal Network，特徵提取網路）會接收經由殘差神經網路（ResNet）與 FPN 輸出的特徵圖像作為輸入，對圖像中的每一個 ROI（Region of Interests）進行分類、定位，並輸出一系列的目標物件候選建議框。第二階段將前一階段的生成的特徵圖像及候選建議框匯入到 ROI Align 進行更進一步的特徵校正，使每個 ROI 生成固定尺寸的特徵圖，接著透過全連接層(FC)完成目標的分類任務和目標檢測框回歸，並新增一條分支以全卷積層(FCN)進行遮罩（Mask）回歸任務。

Mask R-CNN 最大的特色就是使用 ROI Align 取代 Faster R-CNN 中的 ROI Pooling，兩者在模型中都是接在 RPN 之後的階層。在 Faster R-CNN（Ren et al., 2016）中，特徵圖像經過 RPN 的運算得到一系列的目標候選區域(Proposals)，所提取出來的 ROI 區域塊大小都不一致，但是接在後面的網路要求固定大小的輸入，所以在輸入到之後的全連接層（FC）及全卷積層（FCN）前，會先透過 ROI Pooling 層把大小不一的候選區塊調整到一致的大小。ROI Pooling 在調整計算 ROI 的過程中，採用的是取整數值的方式，但這樣的計算方式可能遺失掉很多數據，導致提取 ROI 的位置出現誤差。雖然對於目標的分類任務來說影響不大，但如果要在像素中精確地找出目標的遮罩區域（Mask），這樣的計算方式就會有很大的負面影響（林泊宏，2018）。

ROI Align 是在 Mask R-CNN 中引入的方法。在 He et al.（2017）的研究中，發現到在像素級別的語義分割演算時，ROI Pooling 所帶來的誤差會被放大，所以引入 ROI Align 來消除前一個方法所帶來的計算誤差。ROI Align 避免了取整數值的方式提取 ROI 特徵位置，改使用雙線性插值法來計算每個 ROI 位置的精準值，在最後的實驗結果發現，使

用 ROI Align 對於遮罩面積的預測有著顯著的改進效果。圖 2-3 為 ROI Align 的說明圖示。

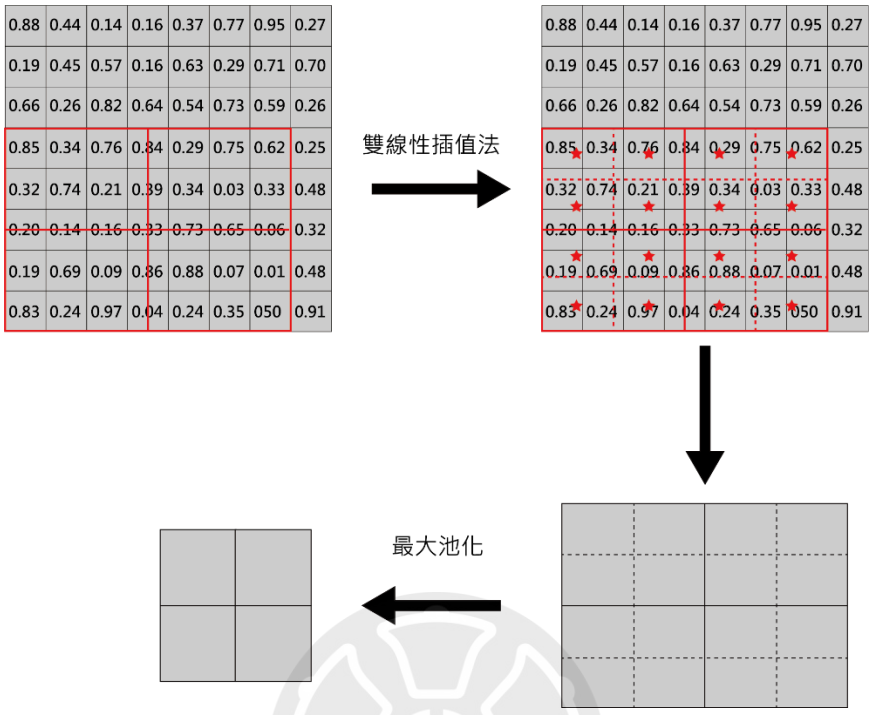


圖 2-3：ROI Align 說明圖示

圖示中為經過卷積運算後生成的特徵圖像，特徵圖像中的框框表示 ROI 的位置，而最終假設經由 ROI Align 後的輸出為 2×2 的特徵圖。整個運算過程不取整數值，保留浮點數信息。在每個單元中，利用雙線性插值算法找出四個座標位置（圖 2-3 中星星的位置），再藉由最大池化操作取出四個座標位置最大的數值，最後輸出一個 2×2 的特徵圖。

第三節、深度學習演算法 Yolov4

早期在 1929 年就有研究學者利用 OCR 對文字圖像檔案進行分析辨識處理，在圖檔版面取得文本資訊。不過此種方法對於文檔圖像的質量要求較高、容錯性低，通常在格式較良好的文檔上會取得較高的準確率（Berg, & Hicks, 2017）。然而海報圖像並沒有一個固定的格式，其複雜的背景、文本布局與字體變化讓文本辨識面臨更大的挑戰。隨著科技的進展，促使了基於深度學習的文本特徵提取研究。其獲取的特徵可以很好的表現目標的信息，因此深度學習和大數據結合的方法逐漸成為該領域的趨勢（楊飛，2016；郭芬紅、謝立艷、熊昌鎮，2018）。

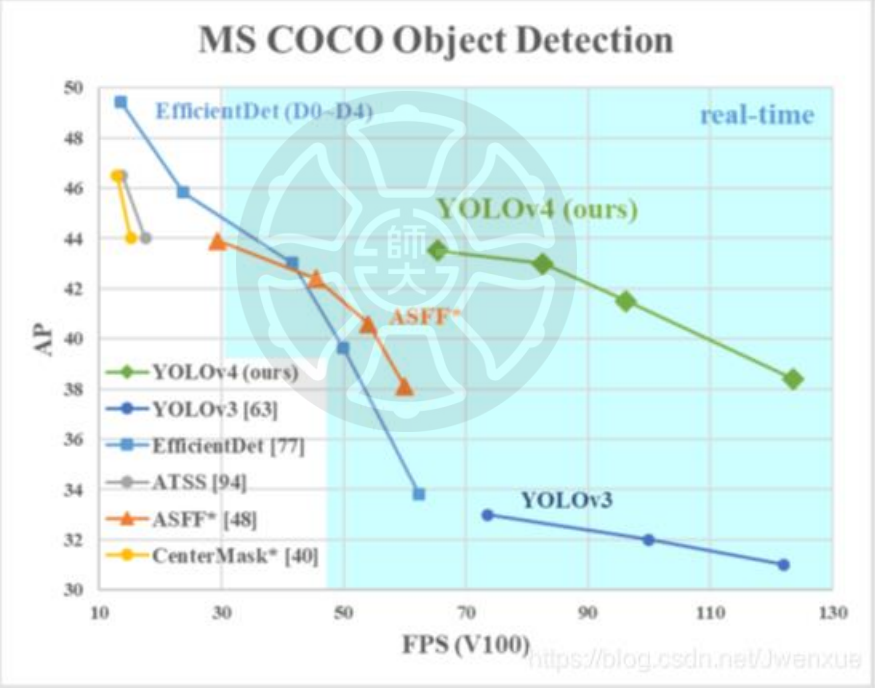


圖 3-1：Yolov4 在 MS COCO 數據集上與其他模型之比較圖（Bochkovskiy et al, 2020）

一、one-stage 目標檢測

Yolov4（Bochkovskiy, Wang, & Liao, 2020）是目前熱門的目標檢測方法之一，屬於 One-Stage 的 Object Detection 演算法，但其 FPS 及 AP(average precision)都不遜色於 Two-Stage 演算法。由 Yolov3（Redmon, & Farhadi, 2018）衍伸而來的神經網路模型，在保證速度的同時，增加了模型檢測精確度，也降低硬體使用的要求，上圖 1 為 Yolov4 在 MS COCO 數據集上與其他模型的比較（Bochkovskiy et al., 2020）。Yolov4 的網絡架構由輸入（input）、骨幹（backbone）、頸部（neck）和頭部（head）所組成。

二、Yolov4 網路架構

Yolov4 基於前三個版本上做了些改變，主要體現在以下四個部分：

1.輸入：對輸入圖像使用馬賽克（Mosaic）數據增強進行訓練。

2.骨幹：骨幹網路 CSPDarknet53 是在原有的 DarkNet-53 基礎上加上 CSPNet（Wang, Liao, Wu, Chen, Hsieh, & Yeh, 2020）概念。增強了 CNN 網絡的學習能力並保證準確率。

3.頸部：插入 SPP（He, Zhang, Ren, & Sun, 2015）、PANet（Liu, Qi, Qin, Shi, & Jia, 2018）結構來擴大感受野與進行圖像特徵融合。

4.頭部：沿用 Yolov3 的頭部，獲取特徵圖像並做出預測。

Yolov4 在輸入端採用 Mosaic 數據增強方法，通過隨機縮放、裁剪、排列四張圖片來增強數據，極大地豐富了檢測數據集，增加了小目標的數量，提高了網絡的檢測效果。在骨幹網路部分，Yolov4 基於 Yolov3 主網路 Darknet53 和 CSPNet 構建了 CSPDarknet53 網絡。CSPDarknet53 不僅增強了 CNN 網絡的學習能力，也保證了模型檢測準確率，降低使用 GPU 記憶體成本。為了更好地提取目標的特徵，Yolov4 網絡在主網絡和輸出層之間插入了 SPP 模塊和 FPN+PANet。通過使用 SPP 模塊，可以擴大主幹特徵接受範圍，並且可以顯著分離最重要的上下文特徵。FPN+PANet 結構聚合了不同主網路層中不同

檢測層的參數，進一步提高了網絡的特徵提取能力。主要功能是將提取的特徵信息轉化為坐標、類別等信息。Yolov4 並不是全新的網路架構，而是將現有的研究整合，包含許多的創新思路，與其他網路架構相比，其檢測速度及準確率都有突出的表現。

第參章、研究方法

本研究的研究方法大致可以分為兩大部分：（一）對目標檢測模型進行訓練；（二）檢測海報圖像中的文字區域。

第一部分主要是海報資料蒐集及標註圖像，並利用整理好的訓量資料對檢測模型進行訓練。第二部分則是將訓練好的模型對海報版面進行測試及分析，檢測出海報版面中的文字區域。

第一節、資料來源與敘述

海報的圖像資料來源為國立臺灣師範大學圖文傳播學系大學生課程上製作的個人海報作業。基於分類檢測出海報圖像文字區塊的研究目的，學生所製作的個人海報富含豐富資訊，符合本次研究的需求。總共蒐集到 200 張的海報圖像。圖 3-1 為蒐集到的海報圖像範例，且海報圖像中富含海報設計者欲傳遞的資訊文本內容。為保護海報圖像中的當事人的個資，故在圖像中將當事人的個人圖像及姓名進行模糊及馬賽克處理。

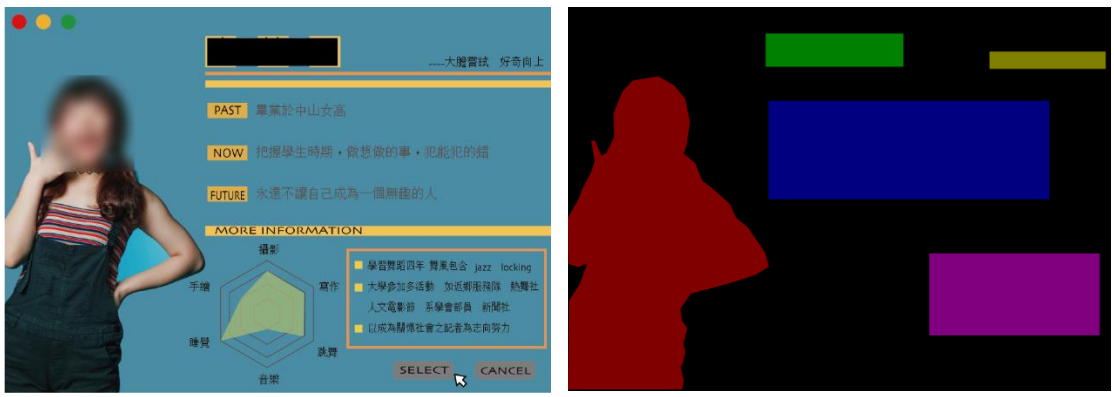
圖 3-1：海報圖像範例

第二節、圖像預處理

目前並沒有公開的海報圖像數據集來完成海報圖像之文本分析的任務，因此本研究為了準備適合 Mask R-CNN 與 Yolov4 之數據集，分別使用 labelme 與 labelImg 來做圖像文本標註。此外，硬體設備及環境方面，硬體設備使用的是 Ubuntu 20.04 作業系統以及 8G 的 NVIDIA GeForce GTX 1650Ti 顯示卡。演算法環境方面，Mask R-CNN 的環境為 Tensorflow-gpu 1.13.1 配合 Keras 2.2.4 版本，搭載 CUDA Toolkit 10.0 和 CUDNN 7.4.2。Yolov4 則搭載 CUDA 11.0.3 和 CUDNN 8.0.5。

一、labelme

為對應 Mask R-CNN 之輸入圖像格式，使用 labelme 圖像標註工具來針對先前蒐集到的海報圖像進行標註，建構了屬於 Mask R-CNN 之海報圖像數據集。經過圖像標註後，海報圖像從原始 png 格式轉為 json 格式的標註文件。逐張圖像標註完成後，完成海報圖像數據集，後續將會利用此數據集進行模型訓練，圖 3-2（a）為海報原始圖像，圖 3-2（b）為經過標註完成後的海報圖像。為保護海報圖像中的當事人，故在圖像中將當事人的個人圖像及姓名進行模糊及馬賽克處理。



（a）　　（b）

圖 3-2：labelme 圖像標註範例

二、labelImg

為對應 Yolov4 之輸入圖像格式，本研究使用 LabelImg 圖像標註工具來針對先前蒐集到的海報圖像進行文本標註，建立海報圖像數據集。定義圖像中文本的 bounding box 與類別標籤，標註完成後會輸出對應的 txt 標註文件檔。為保護海報圖像當事人，故在文中會對圖像中的個資做遮擋的處理，如下圖 3-3 所示：

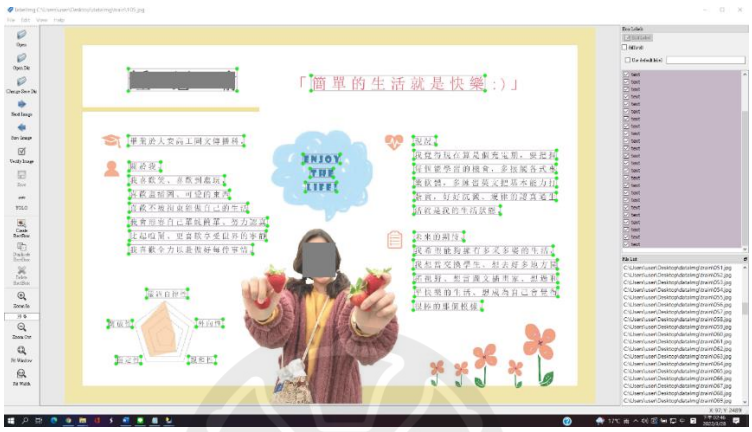


圖 3-3：labelImg 圖像標註範例

第三節、模型訓練

一、Mask R-CNN 網路結構

本研究之 Mask R-CNN 的架構是以 ResNet-101 作為主幹網路，並加入特徵金字塔網路（Feature Pyramid Network，FPN）以減少影像中物件尺度不一致帶來的影響。圖像在進入 Mask R-CNN 模型後，主要分為兩個階段（two-stage）。第一個階段會先將海報圖像送入到主幹網路，由 RestNet-101 網路和 FPN 網路所構成的卷積神經網路中提取圖像特徵（Feature Map）。將卷積神經網路生成的特徵圖像輸入到 RPN 中並依靠可滑動窗口建立生成目標候選區域，RPN 會在特徵圖像上指定 ROI（Region of Interests）的位置。第二個階段則是將 RPN 所輸出的 ROI 送入 RoI Align 層進行像素校正和更進一步的特徵提取。從 RoI Align 層得到的特徵圖像會再接著送入 FC（全連接層）和 FCN（全卷積

神經網路層），前者對 ROI 進行類別和邊框的預測，後者則是將 ROI 的像素進行分類並得到遮罩（Mask）預測。本研究基於 Mask R-CNN 所構建的訓練結構如以下圖 3-3。

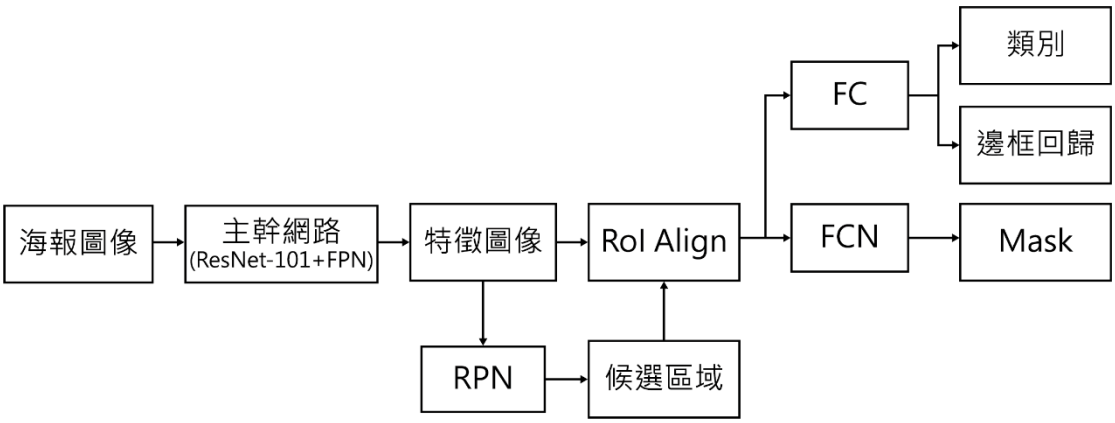


圖 3-3：Mask R-CNN 網路結構

二、Yolov4 網路結構

做為 one-stage 代表的 Yolov4，整體架構不像 Mask R-CNN 那麼複雜。將事先標住過的海報圖像數據集作為 Yolov4 的輸入樣本，圖像輸入後會先在 CSPDarknet53 中初步提取圖像特徵，接著在 SPP 與 PANet 中對於在 backbone 所提取的特徵圖像（Feature Map）進行整合，最後再藉由 head 對目標定位及做出分類預測，整個過程並不像 Mask R-CNN 需要做端到端的訓練及辨識，因此在訓練的時間上會來得比較快，下圖 3-4 為 Yolov4 整體訓練結構。

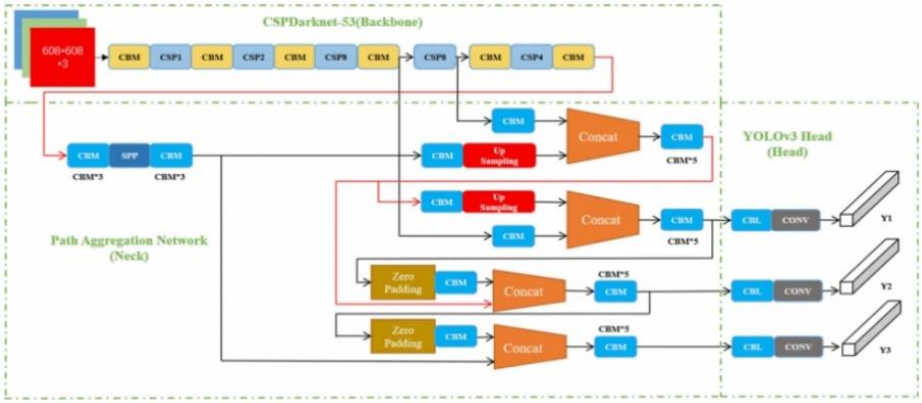


圖 3-4：Yolov4 網路結構

第四節、評量方式

本研究實驗採用基於不同 IOU（Intersection over Union）域值的平均準確率 AP（Average Precision）對每張海報圖像版面的分析結果進行評估。採用 mAP（mean Average Precision）對海報圖像文本辨識性能進行評估。以下章節將會說明本次研究會用到的評量工具。

一、混淆矩陣

鑑別模型的檢測成效並不是一件容易的事情，因此需要一些指標來判斷模型的好或壞，並作為驗證模型的依據。混淆矩陣（Confusion Matrix）是用來評估模型好壞常見的方法，藉由混淆矩陣的概念可以衍生計算出對模型的衡量指標，具體指標會在以下章節進行說明。圖 3-5 為混淆矩陣的圖像範例。

	預測 True	預測 False
實際 True	真陽性 True Positive，TP	偽陰性 False Nagative，FN
實際 False	偽陽性 False Positive，FP	真陰性 True Nagative，TN

圖 3-5：混淆矩陣

二、衍生評量方式

（一）IOU（Intersection over Union）

IOU 是目標檢測研究中常見的數值。在進行物件檢測之前，通常必須先以人工進行圖像標註，並將標註的結果成為 ground-truth，也就是標準答案。而經過深度學習辨識所產生的預測結果，則稱為 predicted。通常，由模型產生的預測結果越貼合標準答案，模型的檢測效果就越完美。以下為計算 IOU 數值的公式：

$$\text{IOU} = \frac{\text{area of overlap(交集範圍)}}{\text{area of union(聯集範圍)}}$$

IOU 等於交集範圍除以聯集的範圍，數值會介於 0～1 之間。數值越接近 1 代表預測結果越接近標準答案，而通常會以 IOU 大於 0.5 來做為判斷辨識率的門檻，並做下一步的計算。

（二）平均精確率 AP（Average Precision）

在目標檢測中，平均精確率 AP 是用來判斷單一類別的平均準確率，在討論平均精確率前，必須先了解計算平均精確率中的各項數值，像是精確率 Precision 及召回率 Recall。

精確率 Precision 是指所有被模型預測為正確情形的個數中，確實為正確的比例，也就是模型的判斷精準度究竟有多少。藉由混淆矩陣的概念，可將精確率的公式定為：

$$\text{精確率（Precision）} = \frac{TP}{TP + FP}$$

召回率 Recall 則是指在所有實際為正確的個數中，模型預測能召回多少正確個數的比例。藉由混淆矩陣的概念，可將召回率的公式定為：

$$\text{召回率（Recall）} = \frac{TP}{TP + FN}$$

前述兩項評估數值通常會以圖像中的單一類別來做評估，通常所需進行測試的圖像，就單一類別可能就會準備數十甚至數百張以上的圖像訓練數據集。而 AP 數值就是評估網路模型對於單一類別的平均精確率，數值越高，模型的檢測成果越好。以 Precision 為縱軸和 Recall 為橫軸，兩項數值可繪製出 PR 曲線（Precision-Recall curve），PR 曲線下所計算出的面積即為平均精確率 AP。以本研究而言，需考慮的類別只有一項，為海報圖像中的文字區域。因此針對文字區域類別所得到的平均精確率 AP，即為總平均精確率 mAP（mean Average Precision），數值越高，模型對於文字區域的檢測能力越準確。

第肆章、研究結果

海報圖像數據集由大學設計課學生所製作的個人海報所蒐集而來，總共蒐集 200 張海報圖像電子檔，並藉由 Labelme 及 LabelImg 兩個圖像標註工具在圖像中標註類別與文本，最後將整理好的海報圖像數據集分為 90%的訓練集與 10%的驗證集，並分別送入 Mask R-CNN 與 Yolov4 兩個模型進行訓練。需辨識的類別只有文本（text）一種，訓練的迭代次數定為 4000 回合，每迭代 1000 回合會保存一次訓練權重。訓練時間方面，Mask R-CNN 約為 19 小時、Yolov4 約為 8 小時。在本次實驗中，訓練文本檢測的模型方面，one-stage（Yolov4）的方法在訓練的效率上來的比 two-stage（Mask R-CNN）約快 2.5 倍。

一、實驗結果與數據

表 1：模型評估指標結果統計（均為判斷 text 之 mAP）

	mAP_{50}	mAP_{75}
Mask R-CNN	0.790	0.513
Yolov4	0.851	0.630

實驗採取不同 IOU 閾值的 mAP 作為模型的評估方式，在海報圖像文本辨識結果中，得到的評價指標結果見表 1。IOU 是目標檢測研究中常見的數值，為模型預測框與實際 ground-truth 的範圍比值。其中 mAP_{50} 為基於 IOU＝0.50 所計算出來的；mAP_{75} 則是基於 IOU＝0.75 計算得出得數據。

從表 1 中可以觀察兩個方法對於文字區域所計算出的平均準確率，Mask R-CNN 的 mAP_{50} 為 79%、mAP_{75} 為 51%，Mask R-CNN 模型對於文字區域的檢測能力在基於 IOU=0.50 的情況下，準確率會達到 0.790 的數值，而基於 IOU=0.75 的文本檢測準確率則會達到 0.513 的數值。相對於 Mask R-CNN，Yolov4 的 mAP_{50} 為 85%、mAP_{75} 為 63%，Yolov4 在 IOU=0.50 的情況下，準確率數值高達 0.851，而在基於 IOU=0.75 的情況下，

依舊有高達 0.630 的數值。從兩個檢測方法的實驗數據顯示兩者都具有在複雜背景色調的情況下，正確的辨識出文字區域目標。另外，也發現到 Yolov4 對於文字區域的檢測準確率較 Mask R-CNN 高。

二、實驗分析與討論

Mask R-CNN 在做文本檢測時，會給予文字區域一個檢測框以及預測遮罩，並且為每個不同的文字區域加上不同顏色及類別標籤，將海報圖像中類別相同但不同實例的文字區域進行區分。在文獻探討中有提到 Mask R-CNN 為端到端（two-stage）的檢測方法，分為兩段式的檢測過程也導致訓練時常會相對較高，圖 4-1 為海報圖像在經由 Mask R-CNN 檢測後所得出的輸出結果，可以觀察到在圖 4-1 中，每一個文字區域都有一組檢測框與預測遮罩以及實例的類別標籤，且每個預測結果的文本實例顏色也都不相同。

而 Yolov4 在做文本檢測時，跟 Mask R-CNN 檢測結果唯一的不同就在於輸出結果不會有預測的遮罩，且也無法將不同的區分出海報圖像中不同的文本實例，Yolov4 的輸出結果會給予每一個文字區域一個類別標籤與預測的檢測框。圖 4-2 為海報圖像在經由 Yolov4 檢測後所得出的輸出結果，圖 4-2 中可以觀察到輸出結果相對單純，沒有不同顏色的預測遮罩，將海報圖像中的每一個文字區域以預測框的方式標示出來。

圖 4-1：Mask R-CNN 輸出結果

圖 4-2：Yolov4 輸出結果

在兩個模型測試過程中，發現兩者對於海報圖像的文字區域辨識能力都有一定的水準，絕大多數的海報文本都能被辨識並加上檢測框。在除了正常的文本檢測樣本之外，被訓練用來做文本檢測的 Mask R-CNN 與 Yolov4 在做測試的時候也有產生檢測不完全的情形。在觀察從兩個目標檢測模型輸出的結果後，將主要的可能因素歸納為四種：（一）文字區域較小之文本檢測；（二）與背景色調接近之文本檢測；（三）多行文字區域之文本檢測；（四）重複性辨識之文本檢測。以下將就四個因素進行分析及模型之間的比較，表 2 為兩方法之間的比較分析。

表 2：兩方法之間的檢測比較

	Mask R-CNN	Yolov4
文字區域較小之文本檢測	弱	**強**
與背景色調接近之文本檢測	弱	**強**
多行文字區域之文本檢測	弱	**強**
重複性辨識之文本檢測	多	**少**

（一）文字區域較小之文本檢測

海報圖像若須呈現較豐富的資訊內容，通常會以壓縮文本大小來使內容配置在有限的海報版面空間中。對於 Mask R-CNN 與 Yolov4 來說，文本若太小會導致在檢測中忽略較小的文字區域而出現漏檢的情形。舉例來說，圖 4-3（a）以及圖 4-3（b）都可以觀察到當檢測資訊量較豐富的海報圖像時，檢測結果都會有將較小文字區域忽略漏檢的狀況產生。仔細比較兩張圖的檢測結果也會發現 Mask R-CNN 的漏檢情況來的比 Yolov4 更多，在檢測其他張海報圖像也有類似的情形發生，因此本研究認為 Yolov4 在對於海報圖像中較小的文字區域有著較高的檢測準確率。

一般來說，在設計海報圖像時，關鍵的資訊需要足夠清晰且容易抓住讀者目光，過多資訊反而容易使讀者失去焦點，欲傳達的資訊量應該盡可能在有限的空間內來作呈現。但如果海報圖像中配置較豐富的資訊量時，對於文本大小的設計也會相對重要，文本太小或壓縮文本空間使得文字區域太過擁擠不只會導致目標檢測模型辨識的錯誤率提高，也有可能造成讀者閱讀時的負擔。

（a）Mask R-CNN　　（b）Yolov4

圖 4-3：文字區域較小的檢測比較

（二）與背景色調接近之文本檢測

Mask R-CNN 在檢測文本與背景色彩對比度較低的海報圖像時，會有無法有效識別文字區域邊界的問題，導致輸出的檢測框無法正確判斷個別文字區域，反而將不同文字區域視為同一個文字區域，或甚至沒有檢測到文字區域。而 Yolov4 在檢測這類對比度較低的海報圖像時，雖然也會有漏檢的情形，但對檢測文字區域邊界則是有較好的檢測能力。對比圖 4-4（a）、圖 4-4（b）兩張輸出海報圖像，在與背景色調接近之文本檢測的能力上，Yolov4 有著比 Mask R-CNN 還有好的檢測能力。

經過實驗檢測後，海報圖像的背景色調與文本色調若過於接近時，會導致 Mask R-CNN 與 Yolov4 在檢測文字區域時產生檢測不完整的情形。因此在設計海報圖像時，雖然需要有漂亮的設計來吸引讀者的目光，但應該盡量避免使用色彩豐富的背景，過多的色彩或是對比度不足的色彩搭配反而會導致檢測時出現缺漏的情形，無法有效提取海報圖像中的文本資訊。讀者閱讀此類的海報圖像也容易分散閱讀文本資訊的注意力。

（a）Mask R-CNN　　（b）Yolov4

圖 4-4：與背景色調接近之文本檢測比較

（三）多行文字區域之文本檢測

海報圖像中的文字區域通常都不會有制式的設計格式，設計者的自由發揮使得海報文本透過具有美感的方式呈現在讀者眼前。因此每一個文字區域都有不一樣的大小空間，有的文字區域只有一行文字，像是標題、標語等；有的則是標題之下的資訊內容，而資訊內容通常不會有固定字數，因此資訊內容的文字區域通常多會以多行文本的方式呈現。圖 4-5（a）、圖 4-5（b）為 Mask R-CNN 與 Yolov4 對於單行文本及多行文本的檢測比較。

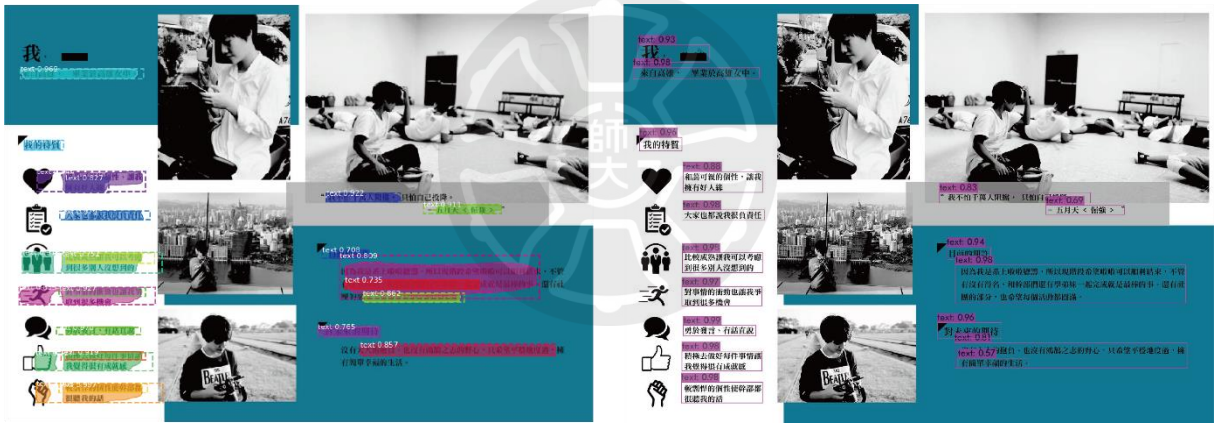
Mask R-CNN 對於辨識單行文字區域有較為出色的能力，但在檢測多行文字區域時，常會有辨識不出結果的情形，在檢測結果會看到部分多行文字區域無法被 Mask R-CNN 辨識。Yolov4 對於單行文字區域或多行文字區域的檢測能力相比 Mask R-CNN 會發現幾乎每個文字區域都有被辨識給予辨識框。在辨識及劃分文字區域的能力上，Yolov4 有著比較出色的檢測能力。

（a）Mask R-CNN　　（b）Yolov4

圖 4-5：多行文字區域之文本檢測比較

（四）重複性辨識之文本檢測

在前一段提到海報圖像中的資訊內容大多是以多行文字區域的方式呈現，Mask R-CNN 在檢測多行文本時。除了會出現檢測不完整的情形之外，也會有重複性辨識的問題，如圖 4-6（a）所示。Mask R-CNN 會在檢測圖像後，會額外在檢測目標加上預測的遮罩，將目標區域用顏色標記顯示。但文字區域沒有一定的樣式，會根據文本內容或設計方式有所調整。因此在檢測多行的文字區域時，可能會因為文本行與行之間的間距太近導致出現重複性的遮罩預測。在區分單行文本及多行文本的能力是 Mask R-CNN 較為不足的地方。反觀 Yolov4 在同一張海報圖像的檢測結果，在區分單行文本及多行文本的能力明顯比 Mask R-CNN 高，可以將海報圖像的文本分段給予檢測結果，如圖 4-6(b) 所示。



（a）Mask R-CNN　　（b）Yolov4

圖 4-6：重複性辨識之文本檢測比較

第伍章、結論與建議

在網路發達的世代中，數位圖像已經是作為資訊傳播的重要工具。而海報圖像當然也不例外會被作為傳播的手段，也是生活中最常見的資訊傳達媒介，在我們生活中處處可見，充斥在現代人的生活當中。既然透過網路進行資訊傳播，若能檢索海報圖像的文字區域，將會使資訊傳播的範圍更加擴大。文本檢測也一直是目標檢測及文字辨識的熱門領域，本研究透過結合深度學習的目標檢測方法，試圖檢測出海報圖像中的文字區域，作為往後研究海報文字辨識的資訊，將實驗結果統整為以下兩小節研究結論與研究建議：

一、研究結論

（一）深度學習演算法具備檢測海報文字區域的能力

實驗數據顯示，Mask R-CNN 與 Yolov4 都有著對海報文字區域良好的檢測表現，mAP_{50} 分別都高達 79.0%及 85.1%。證明結合深度學習的目標檢測方法除了可以利用在自然物景的場合，也可以檢測多樣化海報圖像的文字區域。海報圖像經由 Mask R-CNN 和 Yolov4 的檢測過後，皆會在海報圖像中標註文字區域位置，給予檢測框以及定位資訊。

（二）Yolov4 相對於 Mask R-CNN 有更穩定的檢測結果

從兩種檢測模型所檢測出來的平均準確率（mAP）數據雖然差距不大，但實際檢測的輸出結果卻有明顯的差距。在前一章實驗結果討論中，發現到 Yolov4 比較不受海報圖像因設計缺陷所造成的因素而影響到辨識結果，能夠清楚地檢測海報圖像的文字區域。在與 Mask R-CNN 經過實驗結果的比較後，在文本與海報背景色彩對比度較不明顯時、檢測文本內容較多的海報圖像時，或是檢測海報圖像中較小的文本時，Yolov4 都有較好的檢測結果。而 Mask R-CNN 在檢測目標時，會以遮罩的方式進行目標物件標註，對實際目標的像素定位會有更精細的計算。但海報圖像的文字區域通常是不固定大小的目標物件，沒有一定的格式或形式。尤其文本上下行間距會隨著設計者的喜好而有不同的比

例，使得檢測過程無法有個較明確的文字區域提供 Mask R-CNN 做辨識的學習。因此在檢測文字區域時，會造成 Mask R-CNN 的誤判，在大部分的輸出結果中，都有無法正確辨識多行文本與單行文本的問題。

二、研究建議

本研究所蒐集的海報圖像類型為學生製作之個人海報圖像，準備了海報圖像數據集共 200 張電子檔，因此目標檢測模型對於此類型海報圖像的文字區域會有較高的檢測結果。但海報圖像類型多元且豐富，為因應各類型海報圖像的文字區域檢測，建議可以在往後研究中增加更多類型的海報圖像，藉此豐富海報圖像數據集以提供目標檢測模型做訓練，對各種類型的海報圖像有全面的檢測效果。此外，設計者在設計海報的時候，可參考目標檢測方法辨識時會出現的問題。注意海報文字區域的排版及海報色調的搭配，因應各種可能使檢測出錯的因素作設計的調整，使海報圖像在經目標檢測方法檢測後有更正確的輸出結果。

海報圖像若要能在網路中能夠被快速的檢索，需要的是檢測速度較快且準確的目標檢測模型。比較 Mask R-CNN 以及 Yolov4 的 FPS（影格率），Mask R-CNN 只有 18.04 的 FPS，而 Yolov4 的 FPS 則高達 35.11，在檢測速度上明顯較快。在綜合兩種方法的檢測表現及檢測速度後，本研究認為 Yolov4 有著對海報圖像文本更好的檢測能力，未來可以利用 Yolov4 針對海報圖像文字區域的定位資訊進行進一步的文字識別，實現對海報資訊的索引及檢索，增加海報圖像在網路的曝光率。本研究之研究工具為 Mask R-CNN 與 Yolov4 兩種不同目標檢測方法，實驗結果為基於此兩種方法而獲得研究數據及成果，進而探討目標檢測方法對於海報圖像的文字區域檢測能力。實際上已有許多研究學者發表過結合深度學習的目標檢測方法，未來亦可以嘗試其他目標檢測方法來作為檢測海報圖像文字區域的工具，比較各種方法的檢測效果。

參考文獻

Ahmed, B., & Gulliver, T. A. (2020). Image splicing detection using mask-rcnn. *Signal, Image and Video Processing*, 1-8.

Berg, J., & Hicks, R. (2017). Successful design and delivery of a professional poster. *Journal of the American Association of Nurse Practitioners*, *29*(8), 461-469.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, *37*(9), 1904-1916.

Keely, B. (2004). Planning and creating effective scientific posters. *Journal of Continuing Education in Nursing, 35*(4), 182–185.

Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8759-8768).

Liu, Y., Chen, Z., Xu, C., Liu, T., & Guo, X. (2022). Driver Fatigue Detection Algorithm Based on Improved Yolov4. *World Scientific Research Journal*, *8*(1), 58-63.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision* (pp. 843-852).

Van Dalen, J., Gubbels, H., Engel, C., & Mfenyana, K. (2002). Effective poster design. *Education for Health*, *15*(1), 79-84.

Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 390-391).

Wang, Y., Wang, L., Jiang, Y., & Li, T. (2020). Detection of self-build data set based on YOLOv4 network. In *2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)* (pp. 640-642).

Ye, Q., & Doermann, D. (2014). Text detection and recognition in imagery: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *37*(7), 1480-1500.

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International journal of computer vision*, *59*(2), 167-181.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., & Garcia-Rodriguez, J. (2018). A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing*, *70*, 41-65.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).

Hafiz, A. M., & Bhat, G. M. (2020). A survey on instance segmentation: state of the art. *International journal of multimedia information retrieval*, 1-19.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Jiao, L., & Zhao, J. (2019). A survey on the new generation of deep learning in image processing. *IEEE Access*, *7*, 172231-172263.

Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., & Qu, R. (2019). A survey of deep learning-based object detection. *IEEE access*, *7*, 128837-128868.

Jobin, K. V., Mondal, A., & Jawahar, C. V. (2019, September). DocFigure: A dataset for scientific document figure classification. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)* (Vol. 1, pp. 74-79). IEEE.

Li, Y., Zhang, J., Gao, P., Jiang, L., & Chen, M. (2018, June). Grab cut image segmentation based on image region. In *2018 IEEE 3rd international conference on image, vision and computing (ICIVC)* (pp. 311-315). IEEE.

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).

Qin, J., & Zhang, Y. (2021). Design of Protein Crystal Detection System Based on Mask RCNN. *World Scientific Research Journal*, *7*(5), 85-88.

Rajan, V., & Stiehl, H. S. (2019, September). Making DIA Accessible to Non-Experts: Designing a Visual Programming Language for Document Image Analysis. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)* (Vol. 3, pp. 23-27). IEEE.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*, 91-99.

Rother, C., Kolmogorov, V., & Blake, A. (2004). " GrabCut" interactive foreground extraction using iterated graph cuts. *ACM transactions on graphics (TOG)*, *23*(3), 309-314.

Roullet, C., Fredrick, D., Gauch, J., & Vennarucci, R. (2019). An automated technique to recognize and extract images from scanned archaeological documents. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)* (Vol. 1, pp. 20-25). IEEE.

Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE Access*, *7*, 53040-53065.

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision* (pp. 843-852).

Terauchi, A., Mori, N., & Ueno, M. (2019, September). Analysis Based on Distributed Representations of Various Parts Images in Four-Scene Comics Story Dataset. In *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)* (Vol. 1, pp. 50-55). IEEE.

Tian, Y., Yang, G., Wang, Z., Li, E., & Liang, Z. (2020). Instance segmentation of apple flowers using the improved mask R–CNN model. *Biosystems Engineering*, *193*, 264-278.

Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. *International journal of computer vision*, *104*(2), 154-171.

Xu, L., Fei, M., Zhou, W., & Yang, A. (2018, December). Face expression recognition based on convolutional neural network. In *2018 Australian & New Zealand Control Conference (ANZCC)* (pp. 115-118). IEEE.

Yu, C., Fan, X., Hu, Z., Xia, X., Zhao, Y., Li, R., & Bai, Y. (2020). Segmentation and measurement scheme for fish morphological features based on Mask R-CNN. *Information Processing in Agriculture*, *7*(4), 523-534.

Yu, Y., Zhang, K., Yang, L., & Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture*, *163*, 104846.

Zhang, Q., Chang, X., & Bian, S. B. (2020). Vehicle-damage-detection segmentation algorithm based on improved mask RCNN. *IEEE Access*, *8*, 6997-7004.

Zhang, S., Pan, X., Cui, Y., Zhao, X., & Liu, L. (2019). Learning affective video features for facial expression recognition via hybrid deep learning. *IEEE Access*, *7*, 32297-32304.

于藕（2019）。探討圖像處理軟件在平面設計中的應用。**科學技術創新，(34)**，108-109。

王瑋瓊（2013）。後現代語境下的海報版面設計研究（碩士論文）。取自 https://tra.oversea.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201401&filename=1014148714.nh。

田萱、王亮、丁琪（2019）。基于深度學習的圖像語義分割方法綜述。**軟件學報，(02)**，440-468。doi:10.13328/j.cnki.jos.005659。

李寧（2020）。探究版式設計在平面設計中的作用。**科技經濟導刊 28(26)**，62-63。

李佳薇（2022）。2021 海報設計流行趨勢探究。**現代商貿工業，(7)**，194-195。

李翌昕、鄒亞君、馬盡文（2019）。基於特徵提取和機器學習的文檔區塊圖像分類算法。**信號處理，35(5)**，747-757。

李嫣、胡清（2021）。北京大學生電影節海報設計發展評析。**藝術教育，23**，203-206。

林季穎（2018）。基於深度學習之人臉特徵辨識與應用（碩士論文）。取自華藝線上圖書館。

林泊宏（2018）。基於遮罩區域卷積類神經網路之木節偵測暨分類演算法（碩士論文）。取自華藝線上圖書館。

俞越東（2019）。淺議展覽版面設計的四項基本原則。**文化創新比較研究(02)**，65+69。

胡芝蘭、林行剛、嚴洪（2006）。基於分層密度特徵的文檔圖像檢索。**清華大學學報，46(7)**，1231-1234。

徐津（2005）。報紙版面設計中的人機工程學問題。**包裝工程(04)**，201-204。

徐銳義、吳煒、何小海、楊玉科（2008）。中文商務名片版面分割研究。**四川大學學報：自然科學版，45(2)**，331-335。

晉瑾、平西建、張濤、陳明貴（2007）。圖像中的文本定位技術研究綜述。**計算機應用研究，24(06)**，8-11。

殷建、鄭童（2021）。平面海報設計在信息傳達設計中的研究。**西部皮革，7**，126-129。

荊奕君（2018）。新媒體技術對平面設計的影響與發展。**新聞戰線，(16)**，168-169。

袁子意（2018）。運用深度神經網絡實現驗證碼識別（碩士論文）。取自華藝線上圖書館。

馬勇（2021）。新媒體時代報紙美術編輯設計創新。**新聞傳播，(03)**，119-120。

馬雲峰（2016）。圖文搭配在版面設計中的視覺效果分析。**新媒體研究，20(2)**，43-44。

張芳（2016）。版面設計的要素與視覺傳達。**新媒體研究，18(2)**，178-179。

張勁（2014）。平面媒體編輯版面設計基礎。**美術大觀，6**，110。

張旋（2021）。平面海報設計中圖形符號的視覺傳達探討。**美與時代(上)，(04)**，74-76。doi:10.16129/j.cnki.mysds.2021.04.026。

曹偉（2017）。電影海報細節設計與電影文本內容的敘事性互動。**現代視聽，(5)**，58-61。

郭芬紅、謝立艷、熊昌鎮（2018）。自然場景圖像文字檢測研究綜述。**計算機應用，38(S1)**，173-178。

陳圓圓、王维蘭、劉華明、蔡正琦、趙鵬海（2021）。基於自適應游程平滑算法的藏文文檔圖像版面分割與描述。**中國學術期刊，58(14)**。

陳慧姝（2019）。電影海報中的視覺傳達表現。**包裝工程，40(12)**，313-318。

陳璇、賀建軍、李厚杰、武林秀（2019）。基於 Mask R-CNN 的滿文文檔版面分析。**大連民族大學學報，21(3)**，240-245。

傅隆生、馮亞利、Elkamil Tola、劉智豪、李瑞、崔永杰（2018）基于卷積神經網絡的田間多簇獼猴桃圖像識別方法。**農業工程學報(02)**，205-211。

曾建澔、林孟緯、謝佳蒨、吳志泓（2019）。基於智慧影像分析模式之道路積淹水自動辨識系統。**TANET2019 臺灣網際網路研討會**，92-96。

賀瑩（2015）。平面媒體編輯版面設計基礎研究。**品牌，4**，181。

楊利娜（2021）。色彩構成原理在海報設計中的表現。**西部皮革(08)**，40-41。

楊飛（2016）。自然場景圖像中的文字檢測綜述。**電子設計工程，24(24)**，165-168。

楊捷、劉進鋒（2018）。利用 CTPN 檢測電影海報中的文本信息。**電腦知識與技術，14(25)**，213-215。

劉成林（2019）。文檔圖像識別技術回顧與展望。**數據與計算發展前沿(06)**，17-25。

致謝

本論文幸蒙恩師張教授晏榕這三年來之悉心指導，不僅傳授學生理論上的知識，也培養學生嚴謹與往後踏入社會的處事態度。感謝老師在論文撰寫的過程中，全心全意的投入與付出，並犧牲許多假日的時間指導學生論文的方向與細節。在老師的指導下，我才能順利將影像辨識結合到圖文傳播的領域，在此謹向恩師致上最高的敬意！口試期間，承蒙口試委員周遵儒教授與林玲遠教授之細心指正，並給予許多寶貴的建議，使得本研究能趨於完善，在此致上真誠的感謝。

研究所這三年當中，由於沒有長久的資工背景與紮實的程式基礎，研究影像辨識的過程實屬煎熬，曾經一度想放棄。感謝宥銓、彥融、瑞謙、育璟與育名這些大學朋友們的技術分享與激勵，才能使我可以在不斷的失敗中有再站起來的勇氣。

研究期間時常有心情煩悶、枯燥的時候，特別感謝博揚、廷恩、煜婷、昱華及凱莘，在我最脆弱的時候願意給我鼓勵，並經常帶我出去踏青放鬆心情，若沒有這些適時的放鬆，相信我絕對無法完成這項研究。

最後，我想感謝同屆的研究所同學們峻瑋、雅云、聿祈、芷馨以及靖雯，大家一起每天熬夜研究及討論，讓我受益良多。也非常感謝圖傳所全體同學一路上的支持與鼓勵，與你們創造許多美好的回憶，豐富了我的研究所生涯。當然，也要感謝我的父母親鼓勵我繼續讀研究所，才有今天完成論文的機會，在此對所有幫助過我的朋友及親戚們致上我最深的謝意。