

國立政治大學資訊科學系  
Department of Computer Sciences  
National Chengchi University

碩士論文

Master Thesis

基於語意分割之電纜線偵測

Power Line Detection

Based on Semantic Segmentation

指導教授：廖文宏 博士

研究生：游晉瑞 撰

中華民國 一百一十 年 四 月

April 2021

## 摘要

電腦視覺的領域中，語意分割是一項非常重要的技術，目前被廣泛應用於無人載具、自動駕駛、場景理解與避障等任務。隨著深度學習技術的進步，語意分割已取得一定的成果，然而針對電纜線檢測，仍有許多待改善的空間。現今電塔及電纜線的開源資料有限，本研究主要對目前兩大開源的電纜線資料集進行電纜線檢查的語意分割實驗，分別為 Ground Truth of Powerline Dataset 及 PLD-UVA 資料集，並將兩資料集之 Ground Truth 重新標記，使模型能夠精準的獲得 Ground Truth 資訊。

近年來學者已注意到將不同感測器的資料融合，如熱感測器或深度感測器的資料，可以協助提升光學影像模型的準確度，其中 RTFNet 利用兩個編碼器將熱影像特徵融合到光學影像中，但此架構並未考慮到也可以將光學影像的特徵融合到熱影像中，達到互相幫助的效果。藉此，本研究以 RTFNet 為基礎提出 Dual Segmentation Model，希望透過邊緣強化來加強電纜線的光學影像模型，讓語意層面的資訊得以互補，進而提升分割模型的準確度，使其超越現今電纜線檢測表現相當優異的 LS-Net 之結果。本論文提出之 dual segmentation 模型，其 Precision 以 0.7919 與 LS-Net 的 0.8004 並駕齊驅，Recall 以 0.7710 超越 LS-Net 之 0.5368，總結以上兩者所表現的 F-score 得出 0.7753，超越 LS-Net 的 0.5940 將近 0.2 之多，體現出本研究所提出模型之可靠性。

關鍵字: 電腦視覺、語意分割、深度學習

## Abstract

In the field of computer vision, semantic segmentation is a technique widely employed in tasks such as autonomous driving, scene understanding and obstacle avoidance. With the advancement of deep learning technology, the performance of semantic segmentation has progressed rapidly. Yet there is still much room for improvement in the detection of power lines. Currently, open-source datasets of electric towers and power lines are quite limited. This research mainly conducts the semantic segmentation experiment of power lines inspection using two open-source datasets, namely, the Ground Truth of Power lines dataset and the PLD-UVA dataset. We re-labeled the above two datasets to better locate the region occupied by power lines.

Researchers have reported the benefits of fusing data from different sensors, such as thermal sensors or depth sensors, to enhance the accuracy of optical image models. Among them, RTFNet utilizes two encoders to fuse thermal image features. However, this architecture fails to consider the characteristics of the optical image to promote mutual assistance. This research proposes the dual segmentation model (DSM) based on RTFNet. We hope to strengthen the optical image model of the power lines through edge enhancement, so that the semantic-level information can be complemented, and the accuracy of the segmentation model can be improved. Experimental results indicate that dual segmentation model outperforms the LS-Net model. Specifically, the precision of our model (0.7919) is comparable with that of LS-Net (0.8004). The recall (0.7710) surpasses that of LS-Net (0.5368). F-score of the DSM model is 0.7753, exceeding that of LS-Net (0.5940) of by nearly 0.2, validating the superiority of the proposed approach.

Key words : computer vision , semantic segmentation , deep learning

# 目錄

表目錄 .....	5
圖目錄 .....	6
<b>第一章 緒論</b> .....	<b>9</b>
1.1 研究背景與動機 .....	9
1.2 研究目的 .....	11
1.3 論文架構 .....	11
<b>第二章 背景與相關研究</b> .....	<b>12</b>
2.1 深度學習的架構與介紹 .....	12
2.2 卷積神經網路概述 .....	15
2.2.1 全連接與局部連接 .....	16
2.2.2 池化層 .....	17
2.2.3 ResNet .....	18
2.2.4 U-Net .....	18
2.2.5 語意分割 .....	20
2.3 電纜線檢測相關研究 .....	21
2.3.1 LS-Net .....	21
2.3.2 Hough transform .....	23
2.3.3 傳統基於梯度的方法 .....	24
2.3.4 基於深度學習的方法 .....	25
2.3.5 Encoder-Decoder 架構 .....	26
2.3.6 RTFNet .....	26
2.3.7 低階特徵 Domain Adaptation (SEAdIN) .....	28
2.3.8 高階特徵自注意力機制 Strip Pooling Module .....	29
2.4 小結 .....	30
<b>第三章 電纜線偵測與資料集介紹</b> .....	<b>31</b>
3.1 Canny Edge Detector .....	31
3.2 Ground Truth of Powerline Dataset .....	33

3.3 PLD-UVA.....	34
第四章 研究方法與實驗結果分析.....	37
4.1 電纜線影像語意分割結果.....	37
4.2 Dual Segmentation Model 模型架構.....	49
4.3 Ablation Study.....	51
4.4 Denoise.....	54
4.5 小結.....	56
第五章 結論與未來展望.....	57
5.1 研究結論.....	57
5.2 未來展望.....	58
參考文獻.....	60



## 表目錄

表 4-1. U-Net 語意分割模型訓練結果(訓練樣本數:測試樣本數=8:2).....	38
表 4-2. U-Net 語意分割模型訓練結果(訓練樣本數:測試樣本數=9:1).....	39
表 4-3. PSPNet 語意分割模型訓練結果(訓練樣本數:測試樣本數=8:2).....	39
表 4-4. PSPNet 語意分割模型訓練結果(訓練樣本數:測試樣本數=9:1).....	40
表 4-5. U-Net 重標 ground truth 後結果 (訓練樣本數:測試樣本數=8:2).....	41
表 4-6. U-Net 重標 ground truth 後結果(訓練樣本數:測試樣本數=9:1).....	42
表 4-7. PSPNet 重標 ground truth 後結果(訓練樣本數:測試樣本數=8:2).....	42
表 4-8. PSPNet 重標 ground truth 後結果(訓練樣本數:測試樣本數=9:1).....	43
表 4-9. U-Net 預測 PLD 訓練結果(訓練樣本數:測試樣本數=8:2).....	45
表 4-10. PSPNet 預測 PLD 訓練結果(訓練樣本數:測試樣本數=8:2).....	45
表 4-11. U-Net 混合模型預測結果(訓練樣本數:測試樣本數=8:2).....	48
表 4-12. PSPNet 混合模型預測結果(訓練樣本數:測試樣本數=8:2).....	49
表 4-13. Dual Segmentation Model 結果(訓練樣本數:測試樣本數=8:2)...	52
表 4-14. Dual Segmentation Model 結果(訓練樣本數:測試樣本數=8:2)...	53
表 4-15. Dual Segmentation Model 結果(訓練樣本數:測試樣本數=9:1)...	53
表 4-16. Dual Segmentation Model 結果(訓練樣本數:測試樣本數=9:1)...	54
表 4-17. 單一 Encoder Segmentation Model 對 GTD 進行預測結果.....	55

## 圖目錄

圖 1-1. 傳統人力電纜巡檢.....	10
圖 1-2. 使用直升機進行電纜巡檢.....	10
圖 2-1. 神經元.....	13
圖 2-2. 類神經網路.....	13
圖 2-3. ILSVRC 歷年最低之 Top-5 錯誤率.....	14
圖 2-4. LeNet-5 架構.....	15
圖 2-5. VGG 網路架構.....	16
圖 2-6.全連接(左)與局部連接(右).....	17
圖 2-7. 最大池化層與平均池化層.....	17
圖 2-8. Residual block.....	18
圖 2-9 U-Net 網路架構.....	19
圖 2-10. 語意分割辨識示意圖.....	20
圖 2-11.語意分割模型訓練效果 1.....	21
圖 2-12. 語意分割模型訓練效果 2.....	21
圖 2-13 PBR 繪製電纜線的合成圖像.....	22
圖 2-14. LS-Net 架構圖.....	22
圖 2-15. LS-Net 對於各種 grid 組合之實驗結果.....	23
圖 2-16. 影像空間和參數空間之映射.....	23
圖 2-17.雷登轉換之幾何示意圖.....	25
圖 2-18.基於深度學習的電纜線流程圖.....	25
圖 2-19 Encoder-Decoder 架構.....	26
圖 2-20. RTFNet 模型架構圖.....	27
圖 2-21. 邊緣檢測範例 1.....	28
圖 2-22. 邊緣檢測範例 2.....	28

圖 2-23. SEAdIN Module 示意圖.....	29
圖 2-24. Strip Pooling Module.....	30
圖 3-1. 可見光之原圖及已標記圖片 .....	33
圖 3-2. 紅外光之原圖及已標記圖片.....	34
圖 3-3. 電纜線分布於城市市區場景( PLDU ) .....	34
圖 3-4. 電纜線分布於城山區場景( PLDM ) .....	35
圖 3-5. 各場景之 Ground Truth .....	35
圖 3-6. 填滿電纜線之 Ground Truth .....	36
圖 4-1. 電纜線語意分割結果 1.....	38
圖 4-2. 電纜線語意分割結果 2.....	38
圖 4-3. 重新標記 GTD 之 Ground Truth 之語意分割結果 1.....	40
圖 4-4. 重新標記 GTD 之 Ground Truth 之語意分割結果 2.....	41
圖 4-5 U-Net 預測 PLDU 原始資料的預測結果 1.....	44
圖 4-6 U-Net 預測 PLDU 原始資料的預測結果 2.....	44
圖 4-7 將 PLD 之 Ground Truth 填滿.....	44
圖 4-8.U-Net 預測 PLD 標記過後預測結果 1.....	46
圖 4-9.U-Net 預測 PLD 標記過後預測結果 2.....	46
圖 4-10. 混合模型預測 Ground Truth Powerline Dataset1.....	47
圖 4-11. 混合模型預測 Ground Truth Powerline Dataset2.....	47
圖 4-12. 混合模型預測 PLD-UVA Dataset1.....	47
圖 4-13. 混合模型預測 PLD-UVA Dataset2.....	48
圖 4-14. Dual Segmentation Model 架構示意圖.....	50
圖 4-15.DSM Model 預測 GTD 之結果 1.....	51
圖 4-16.DSM Model 預測 GTD 之結果 2.....	52
圖 4-17.DSM Model 預測 GTD 之結果 3.....	52



圖 4-18 DSM 預測 Ground Truth Powerline Dataset 1.....54  
圖 4-19 DSM 預測 Ground Truth Powerline Dataset 2.....55  
圖 5-1 現有電塔及電纜線資料集一覽.....58



# 第一章 緒論

## 1.1 研究背景與動機

隨著科技日新月異的發展，近年來人工智慧(Artificial Intelligence)受到學術和業界的廣泛關注，世界各國均高度重視人工智慧的相關研究，研究電腦如何模擬或實現人類的學習行為，以獲取新的知識或技能，重新組織學習到的知識結構使其不斷改善自身的效能，是人工智慧的技術核心，人工智慧在許多領域皆有重大的發展，包括機器學習(Machine Learning, ML)、深度學習(Deep Learning, DL)、知識圖譜(Knowledge Graph)、自然語言處理(Natural Language Processing, NLP)、電腦視覺(Computer Vision)、人機互動(Human Computer Interaction)、生物特徵識別(Biometrics)、虛擬實境(Virtual Reality, VR)、擴增實境(Augmented Reality, AR)和混合實境(Mixed Reality, MR)等相關技術，更有許多 AI 已經超越了人類。

2016 年，代表人類與 AlphaGo 展開五番棋勝負的韓國圍棋棋王李世石說道「即使我登上世界第一，也有無法擊敗的對象。」現今人工智慧的圍棋實力普遍超越頂尖棋士，而最近的 AlphaGo Zero 更是超越了之前的版本，不需要人類資料，透過跟自己對戰，在 72 小時內就能超越上述和李世石對戰的 AlphaGo (AlphaGO Lee)，以 100:0 的成績超越了 AlphaGo Lee 的實力，並在 40 天內超過了所有之前的版本。深度學習源於多層神經網路的機器學習法，透過層層處理，將大量的訊號抽取有用的特徵將其轉為有用的資訊並解決問題。

在此發展趨勢下，深度神經網路用於無人載具的應用成為近來熱門研究的議題，一個完整的無人機智慧系統由四個功能模組組成：認知模組負責情境意識和決策的控制、運動模組提供運動規劃和鏡頭視角路徑規劃、感知模組開發影像分析和識別算法、介面模組包含資料視覺化及人機互動介面，所有模型都有助於空間的建構和任務的執行。結合人工智慧、深度學習、機器學習、電腦視覺等領域的研發能力，開發出適用於空中之無人機智慧系統。

隨著無人機製造成本逐漸降低，各種針對民生應用的無人機服務蓬勃發展，傳統機器人無法跨過的地方，旋翼無人機都可以從空中輕易抵達。過去高空高壓電纜巡檢方式從純人力巡檢，如圖 1-1 所示，轉變為人力巡檢加上載人直升機巡檢的「協同發展」，如圖 1-2.所示。相較於傳統高度依賴人工方法，透過無人機空拍巡檢能降低時間與金錢成本十倍以上，因此近年無人機日漸成為高空高壓電纜巡檢重要的工具。



圖 1-1. 傳統人力電纜巡檢



圖 1-2.使用直升機進行電纜巡檢

## 1.2 研究目的

無人機的操作有著一定的難度，近年電力巡檢單位都需特別成立航拍小組或小隊專門學習如何操作無人機並進行拍攝，而拍攝到的電纜照片大多都是透過人眼去判斷外觀是否變形，一架無人機還是需要配置一個操縱者和觀察員，並無法有效降低人力成本，因此本論文希望能發展偵測電纜線異常之系統，提供電纜線檢測人員進行快速且安全的高空檢測手段。

本論文具體貢獻如下：

1. 本論文探索 FCN、U-Net、PSPNet、PAN 等泛用的語意分割模型於電纜線資料集，並記錄實驗過程進行分析與討論，藉以選取最佳的解決方案組合。
2. 本論文提出 Dual Segmentation Model，使用 RTFNet 為開發基礎，在低階特徵利用 Domain Adaptation 做資訊融合，同時在高階特徵採用注意力機制建構空間上的關係，藉由上述兩個 Encoder 做結合，使資訊得以互補，進而提升分割模型的準確度。
3. 將深度神經網路架構應用於高空電纜線檢測任務，建立高辨識精準度之自動化系統，研究之成果將可大幅降低電纜線檢測所需之人力資源與時間。

## 1.3 論文架構

本論文分成五個章節，內容分述如下，第一章為緒論，主要描述研究背景、動機與目的。第二章為技術背景與相關研究，在此章將介紹近年來發展快速的深度學習網路架構，接著探討電纜線之影像語意分割模型等相關研究。第三章將詳細說明邊緣檢測之方法與電纜線資料集介紹。第四章研究方法與結果，主要說明本論文所提出語意分割模型和測試不同編碼器之效果。第五章為結論與未來展望。

## 第二章 背景與相關研究

本章首先回顧深度學習網路的發展背景與架構，並介紹語意分割之概念及框架。接著說明近年來指標性之語意分割模型，最後探討語意分割相關研究成果，並延伸探究其在本研究之可用性。

### 2.1 深度學習的架構與介紹

深度學習是基於深層類神經網路(Deep neural network)對資料進行特徵學習所建構出的機器學習模型，能透過層層的處理將大量的訊號轉成有用的資訊並解決問題，與傳統機器學習演算法最大的差異在於，以往的特徵需要經由人為選擇，而深度學習則是在訓練過程中從中學出這些特徵進而讓訓練結果更加完善。在1943年神經生理學家 Warren McCulloch 以及數學家 Walter Pitts 首次提出了神經元的數學模型，即 MCP 神經元(McCulloch-Pitts Neurons) 的數學模型，開啟了「神經計算」的時代。

1956年夏天在達特茅斯學院舉辦的達特茅斯會議。舉辦了名為「人工智慧夏季研討會」(Summer Project on Artificial Intelligence)的會議上第一次正式提出了人工智慧這個概念。1958年，心理學家 Frank Rosenblatt 提出了感知器(Perception)的概念，感知器模型是模仿生物神經元的計算模型。在這樣的架構中，每個神經元代表著對應輸入資料的權重(Weights)加權加總後再透過激勵函數(Activation function)來獲得最後的輸出結果，如圖 2-1 所示。而類神經網路則是將神經元一層一層的連起來，可分為輸入層(Input layer)、輸出層(Output layer)和中間的隱藏層(Hidden layers)，如圖 2-2 所示。每一層的神經元皆有連結且擁有各自的權重來計算訊號的加權，而深層類神經網路其實就是擁有多層隱藏層的類神經網路，我們給網路一個輸入，透過隱藏層中神經元的層層運算後，得到一個

輸出，再把這個輸出值和我們事先預設的答案相比，用這個差距來回修正各神經元的權重值，使其提高網路輸出的準確率。

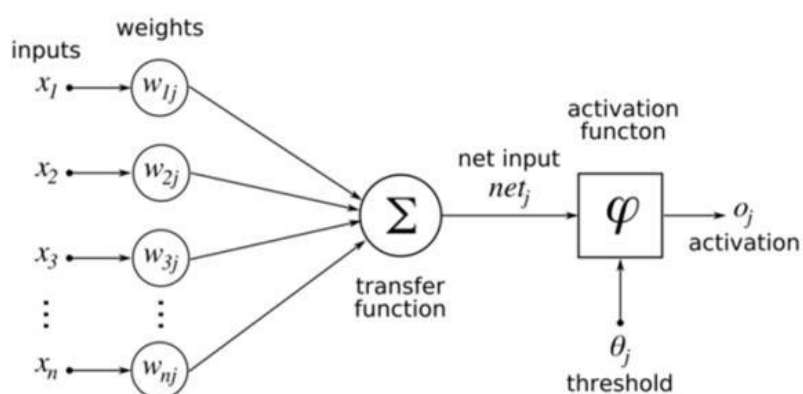


圖 2-1. 神經元

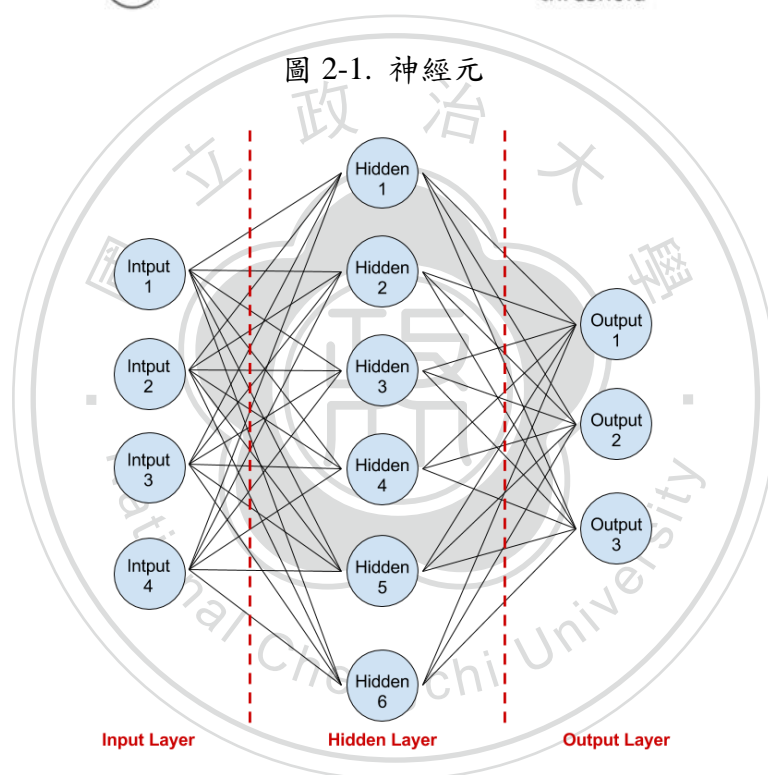


圖 2-2. 類神經網路

深度學習的研究需要取得大量的資料來訓練深度學習模型，電腦硬體的運算能力需求要求也越來越高，近年來網路及電腦硬體，尤其是 GPU 與平行運算技術的發展，深度學習再次受到關注，在 2010 年，深度學習在圖像領域中一個公認的評判資料集 MNIST[1]手寫數字的資料集上得到了突破性的發展，得到了僅有 0.3%的錯誤率。同年，舉辦了第一屆 ILSVRC(ImageNet Large-Scale Visual Recognition Challenge)，使用的資料集是 ImageNet[2]資料集的子集合，共 140 萬

張圖像，將 120 萬張圖片作為訓練集，5 萬張當驗證集，15 萬張當測試集，共 1000 個類別，此次比賽當中的第一名利用了支持向量機(Support vector machine, SVM)的方法，辨識分類的 TOP-5 錯誤率為 28% [3]。

2012 年的 ILSVR2012 中，由 Krizhevsky 等人提出了 AlexNet [4] 展現出圖像辨識上驚人的表現，不僅贏過先前的 SVM，更達到了僅 15.3% 的 Top-5 錯誤率，比起第二名的傳統方法整整低 10% 以上。在 ILSVRC 2014 中，Google 團隊設計出 GoogLeNet [5]，透過多個卷積核擷取影像不同尺寸的資訊，最後進行融合，可以得到影像更好的表徵，這就是 Inception Module 的核心思想，Top-5 錯誤率為 6.7%，接近了人眼視覺的表現(約 5~10% 錯誤率)。而在 ILSVRC2017[6]，Momenta 胡杰團隊(WMW)提出一個新的網路架構，Squeeze-and-Excitation Networks (簡稱 SENet)[7]，一舉取得最後一屆 ILSVRC 的冠軍，SENet 增加額外的卷積核權重層根據 loss 去學習特徵權重，使得有效的特徵圖(Feature map)權重大，無效或是效果小的特徵圖權重小，透過這樣的方法訓練模型達到更好的效果，Top-5 錯誤率降低至 2.25%，甚至超過人眼的表現。如圖 2-3 所示，綜觀歷年 ILSVRC Top-5 錯誤率不免看出深度學習在影像辨識上豐碩的成果，其背後的成功都與兩個不可或缺的要素有關：卷積神經網路以及龐大的資料。



圖 2-3. ILSVRC 歷年最低之 Top-5 錯誤率

## 2.2 卷積神經網路概述

類神經網路是一種模仿生物神經網路，尤其是大腦的結構和功能的數學模型或計算模型，而在 1962 年，Hubel 和 Wiesel 研究貓的視覺神經時，發現貓的視覺皮層對訊息的抽取式一層一層的結構，每個視覺神經元只負責處理局部視覺信號，有的神經元對簡單的特徵敏感，有的則對複雜的特徵敏感，而處理複雜圖像特徵的神經元輸入並不是來自原始圖像，而是來自處理簡單圖像特徵的神經元的輸出，所以視覺神經元是有層次的，此一想法影響了神經網路架構，不必然都是全連接，高層神經元會將低層神經元檢測到的訊息整理出更複雜的特徵，接著交給更高層的神經元處理。

LeCun 等學者從這項研究獲得啟發，並在 1989 年發表全球第一個 CNN 框架 LeNet-5 [8] 便是使用了多層卷積層與池化層的經典案例，被譽為是卷積神經網路的「Hello Word」，架構圖如圖 2-4。此架構在 MNIST 手寫辨識上達到了 99% 以上的準確度。多層卷積與池化的概念亦在更新的網路如 VGG [9]、ResNet [10] 與 DenseNet [11] 等架構中使用。如圖 2-5 為 VGG 網路的架構。

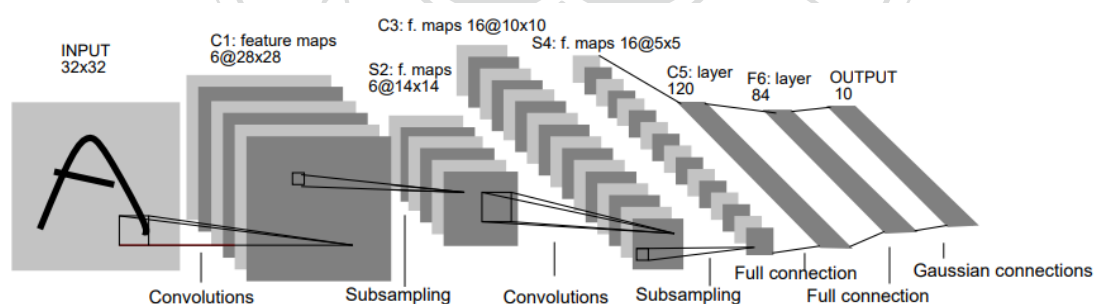


圖 2-4. LeNet-5 架構[8]



ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 <b>LRN</b>	conv3-64 <b>conv3-64</b>	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 <b>conv3-128</b>	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 <b>conv1-256</b>	conv3-256 conv3-256 <b>conv3-256</b>	conv3-256 conv3-256 conv3-256 <b>conv3-256</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

圖 2-5. VGG 網路架構

幾個 CNN 重要的概念如下：

### 2.2.1 全連接與局部連接

局部連接(Local Connectivity)也稱為局部感知或稀疏連接(Sparse connectivity)，每個神經元不再和上一層神經元相連接，而是指和一小部分神經元連接，卷積核的大小遠比輸入大小還要小，CNN 的神經元用此一概念，每個神經元只需要對整張圖片的局部區域進行感知，對於局部周圍的像素關係較為緊密，距離較遠的像素關聯性則較低。當輸入的影像為 1000 x 1000 假使第一層隱藏層就有  $1 \times 10^6$  個，如果使用權連接層，如圖 2-6(左)，則會產生  $1000 \times 1000 \times 10^6 = 10^{12}$  個參數，而圖 2-6(右)是 CNN 透過 10 x 10 的卷積核與局部圖像相連接，這樣參數量則變成  $10 \times 10 \times 10^6 = 10^8$ ，減少了大量的參數量，使網路的訓練更加的有效率。

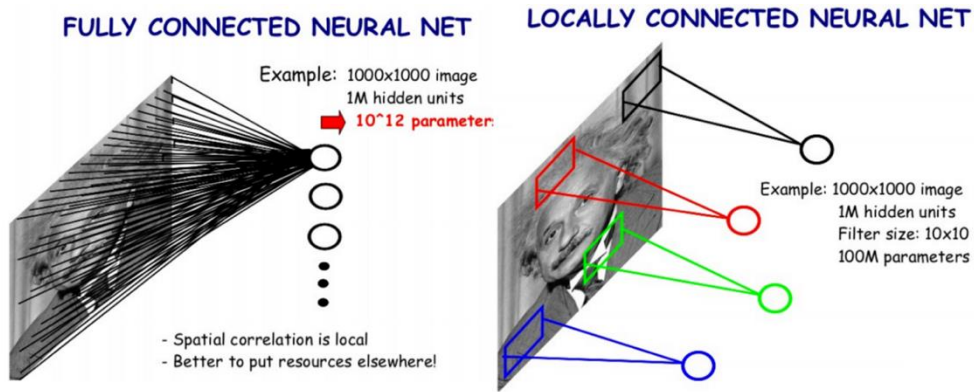


圖 2-6. 全連接(左)與局部連接(右)

## 2.2.2 池化層

另一個與 CNN 密不可分的機制則是池化(Pooling)，目的是用來減少輸入項量維度，提高運算速度，是一種降採樣，作用方式為將輸入圖分成數個區域，並在每個區域中選出一個代表值。常見的方式有最大池化(Max pooling)與平均池化(Average pooling)兩種，如圖 2-7.為一個步幅(Stride)為 2，核大小(Kernel size)為 2x2，前者在每個區域中選出最大的值，後者則選出區域的平均值，在實驗中最大池化層被證明往往獲得較好的表現。

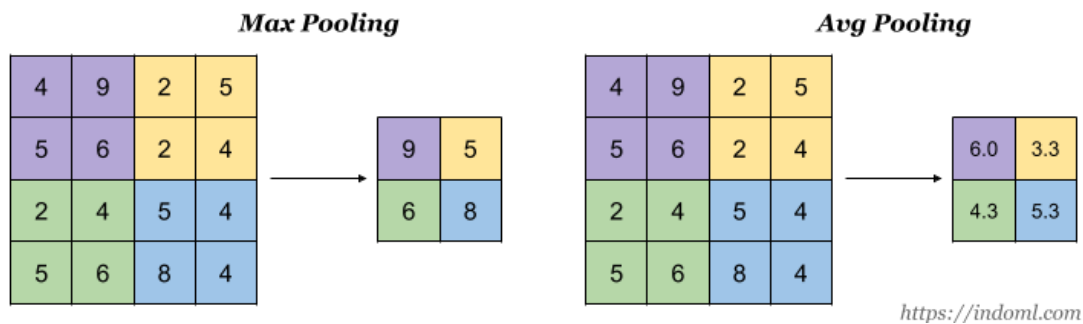
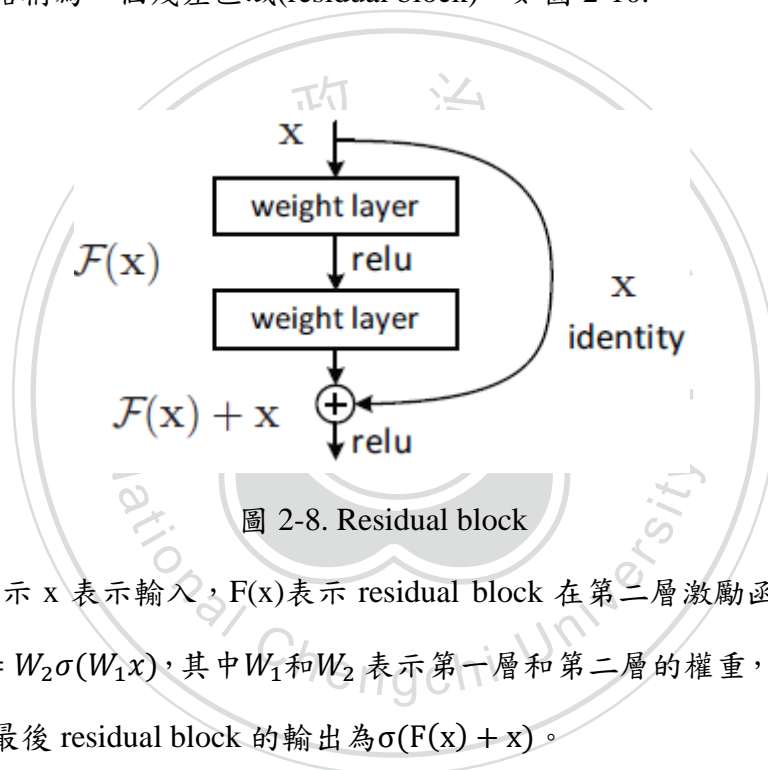


圖 2-7. 最大池化層與平均池化層

### 2.2.3 ResNet

2015 年由微軟亞洲研究院何凱明團隊所提出的一種網路結構，以 3.75% 的 TOP-5 錯誤率獲得當時 ILSVRC 大賽冠軍(如圖 2-3.)，作者根據瑞士學者 Schmidhuber 提出的 Highway Network [12]設計了深度殘差學習框架(Deep Residual Learning Framework)，Highway Network 其相當於改變每一層的激勵函數，使的上一層的特徵圖訊息有一定的機率直接傳到下一層，就像在神經網路中加入一條捷徑(shortcut)，變得更加容易被優化，包含一個 shortcut connection 的幾層神經網路稱為一個殘差區域(residual block)，如圖 2-10。

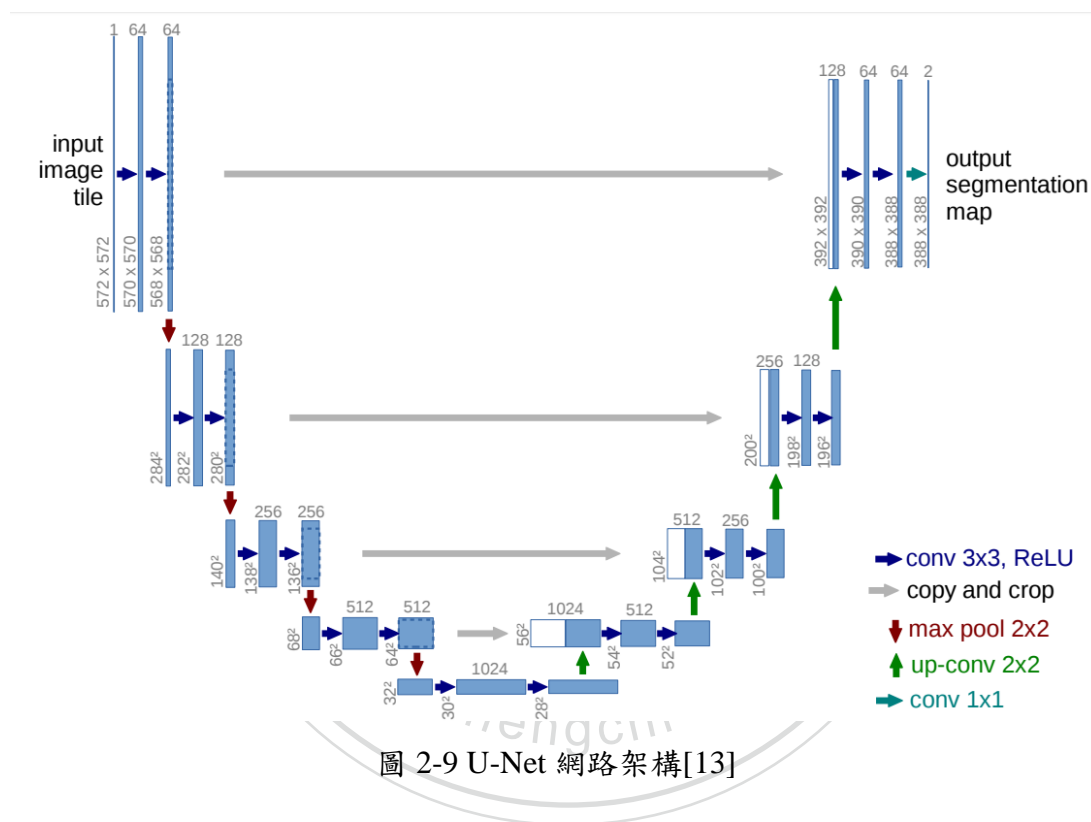


如圖 2-10.所示  $x$  表示輸入， $F(x)$ 表示 residual block 在第二層激勵函數之前的輸入，即 $F(x) = W_2\sigma(W_1x)$ ，其中 $W_1$ 和 $W_2$ 表示第一層和第二層的權重， $\sigma$ 表示 ReLU 激勵函數。最後 residual block 的輸出為 $\sigma(F(x) + x)$ 。

### 2.2.4 U-Net

U-Net 於 2015 年由 Ronneberger [13]等人所提出，在圖像分割中，需要將特徵圖(Feature maps)轉換成一個向量，再從這個向量重建圖像，U-Net 在常規的卷積網路後面新增連續的層，前半部的下取樣其作用是特徵提取，它包括兩個  $3 \times 3$  卷積的重複應用，每個卷積後面是一個 ReLU 激勵函數和  $2 \times 2$  的最大池化層(Max pooling)，每個下取樣步驟中的特徵通道數量都增加一倍，後半部是上取樣其作

用是擴展特徵維度，路徑中的每一步都包含一個向上取樣的特徵圖，然後接 2x2 卷積，該卷積將特徵通道數量減半，與收縮路徑中的特徵圖相應進行連線，以達到與下取樣對應區塊相同大小，如同編碼-解碼器結構，由於此網路架構類似於大寫的英文字母 U 如圖 2-14.，故名 U-Net，其貢獻在於，在上取樣和深入網路的過程中，將下取樣中的高分辨率特徵與上取樣特徵連結起來以便在後續的卷積過程中更好地學習特徵。



Ronneberger 等人透過在最後的特徵圖上 soft-max 的輸出結果計算交叉熵損失函數，以此來做為整個網路的能量函數，其表達式如下公式 2-1 所示：

$$p_k(x) = \frac{\exp(a_k(x))}{\sum_{k=1}^K (\exp(a_k(x)))} \quad (\text{式 2-1.})$$

$a_k(x)$  表示在位置  $x$  之像素在特徵圖中的第  $k$  層激活值(activation)，其中

$x \in \Omega$ ,  $\Omega \subset Z^2$ , 即  $x$  屬於空間  $\Omega$ , 而空間  $\Omega$  是整數集合  $Z^2$  的子集,  $K$  為像素點的類別總數,  $p_k(x)$  為近似最大函數。

透過預先計算權重圖來獲得每一個像素在損失函數中的權重, 這種方法補償了訓練數據每類像素的不同頻率, 對邊界進行更多的加權操作使網路能夠學習相鄰實例間間隔, 其權重的計算方法如下公式 2-2. 所示:

$$W(x) = W_C(x) + W_0 \cdot \exp\left(-\frac{(d_1(x)+d_2(x))^2}{2\sigma^2}\right) \quad (\text{式 2-2.})$$

其中  $W_C$  是權重圖來平衡像素的頻率,  $d_1$  表示最近單元邊界距離,  $d_2$  表示到第二最近單元的邊界距離。

## 2.2.5 語意分割

圖像語意分割其目的就是使機器自動將圖像中對每個像素分配到某個對象類別, 分割出多個對象區域, 並識別其內容, 如圖 2-8, 不僅要識別出各個顏色的類別物件, 還要標示出對象的邊界。因此, 與分類目的不同, 相關的模型要具有像素層級的預測能力。

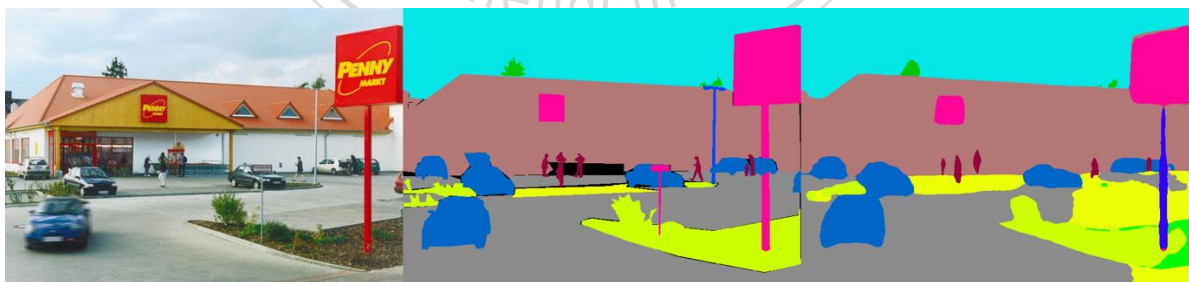


圖 2-10. 語意分割辨識示意圖

2019 年 Yakubovskiy 時一手打造包含四種模型架構且每個架構有三十種可用的編譯器, 持續更新至今, 已有七種模型架構, 和五十七種可用編譯器, 所有編譯器皆具有預先訓練的權重, 可使訓練更快且更好的收斂。作者使用 CamVid

數據集來進行街景的分割，建構模型時，使用具指標性的 U-Net 架構並使用 se\_resnext50\_32x4d 預訓練編碼器，訓練結果即使車體部分被遮擋，或是在黑夜中，模型都能辨識出來如圖 2-9、圖 2-10。此研究的實驗效果相當優秀，因此本論文將會使用此模型架構來進行接下來的實驗。



圖 2-11. 語意分割模型訓練效果 1



圖 2-12. 語意分割模型訓練效果 2

## 2.3 電纜線檢測相關研究

隨著自動檢測電纜線的需求不斷增長，近年來已經開發了許多電纜線檢測的方法，許多研究人員專注於分析 LiDAR 數據，從而提供電纜線周圍環境的高精度 3D 點雲數據，但是這些方法成本過高且不實用，因此，我們主要集中研究圖像的方法上，基於圖像的電纜線檢測方法可分為兩類，傳統基於梯度的方法及基於深度學習的方法，以下將介紹相關代表性的手法。

### 2.3.1 LS-Net

Nguyen 等人[14]在 2020 發表了一個網路架構 LS-Net，由於開源電纜線資料集很有限，他們和 Nordic Media Lab (NMLab)合作，利用物理渲染的方式(PBR)

繪製電纜線的合成圖像，如圖 2-13，但在論文中並沒有提及他們利用多少資料量來作為訓練資料，而測試資料則是利用整個 Ground Truth of Powerline Dataset(GTD)為測試資料集。



圖 2-13 PBR 繪製電纜線的合成圖像

LS-Net 除了主要 grid，其利用和 SSD 及 YOLO 一樣的 grid 偵測電纜線之外，還透過另外三個 grid，分別為 horizontal grid、vertical grid 及 central grid，透過這些 grid 來檢測不同方向間隙，從而消除線中的不連續性，如圖 2-14，此篇論文中同時使用四個 grid 的效果是最好的，如圖 2-15。

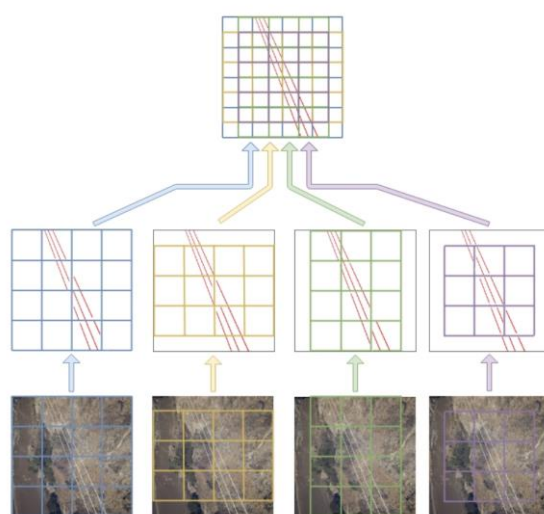


圖 2-14. LS-Net 架構圖

Method	APR	ARR	F1 Score
1 Grid			
LS-Net-1-M	<b>0.8312</b>	0.3791	0.4847
2 Grids			
LS-Net-2-MH	0.8174	0.4717	0.5540
LS-Net-2-MV	0.8173	0.4703	0.5533
LS-Net-2-MC	0.8165	0.4776	0.5574
3 Grids			
LS-Net-3-MHC	0.8080	0.5121	0.5792
LS-Net-3-MVC	0.8080	0.5117	0.5791
LS-Net-3-MVH	0.8064	0.5165	0.5826
4 Grids			
<b>LS-Net-4-MHVC</b>	0.8004	<b>0.5368</b>	<b>0.5940</b>

圖 2-15. LS-Net 對於各種 grid 組合之實驗結果其 M,H,V,C 分別代表 main, horizontal, vertical 和 central 之 grid

### 2.3.2 Hough transform

Hough transform (霍夫轉換)於 1962 年由 Hough[15]首次提出，其後於 1972 年由 Duda 和 Hart 推廣使用[16]，是圖像處理中從圖像檢測幾何形狀的基本方法之一，也是經典的直線檢測方法，其基本原理在於利用點與線的對偶性，直線偵測的霍夫轉換將影像空間中的線轉換為參數空間中的點，如圖 2-16，透過這樣的方法，將直線檢測問題轉換為參數空間中的峰值檢測問題。

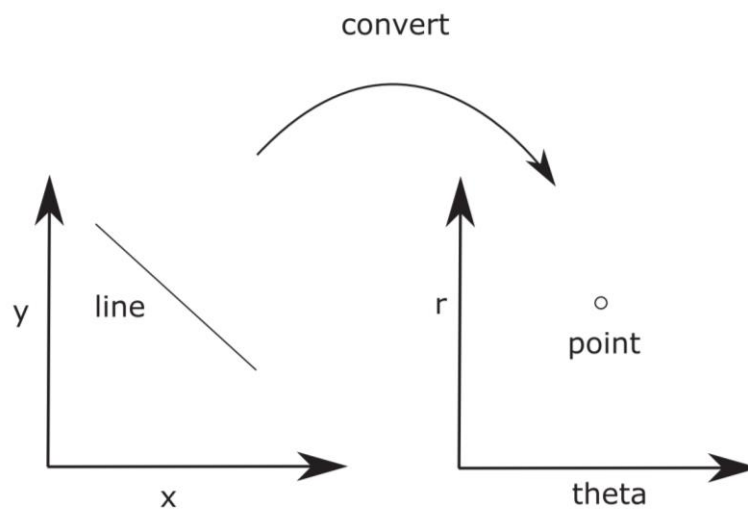


圖 2-16. 影像空間和參數空間之映射



Hough transform 把每一個影像點映射至多重的參數點，其使用一個累加器 (Accumulator) 來記錄每一組參數點出現的次數，出現頻率最高的一組參數，就是影像空間上最具代表性的一直線。

本論文透過 Hough transform 找尋圖像中的電纜線並標記出來，作為我們的 Ground truth 來和預測結果做比較，從而計算其 loss，並從中改進，使電纜線檢測任務的預測結果越來越好。

### 2.3.3 傳統基於梯度的方法

這些方法主要關注梯度、亮度、紋理和其他低級的局部特徵，根據電纜線是圖像強度最低且彼此平行的直線或多項式曲線的條件下專門設計，先使用 Canny 和 Sobel 等邊緣檢測器將電纜線的潛在像素與背景區分，然後利用霍夫轉換來檢測線段。Li 等人提出利用霍夫轉換檢測電纜線，由於霍夫轉換會檢測任何線性物體，無法單獨識別出電纜線，所以他們基於 K-means 集群分析(K-means Clustering)，對所有檢測到的線進行聚合，以識別感興趣的電纜線。

Yan 等人[17]採用雷登轉換(Radon Transform)提取線段，如圖 2-17，然後採用分組方法和卡爾曼濾波器將線段連接成一直線，而 Chen 等人[18]提出了一個改進 Yan 等人使用 Radon 轉換的版本，叢集雷登轉換(Cluster Radon Transform, CRT)，用於從高分辨遙測影像中檢測電纜線的線性特徵。儘管上述方法討論了圖像中電纜線的特性，但其中的算法僅關注直線的檢測，未考慮到圖像邊緣和其他電纜線相似的可模糊線性特徵。

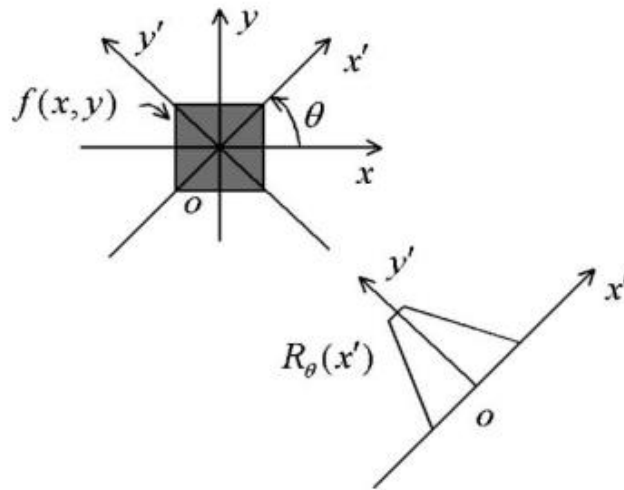


圖 2-17. 雷登轉換之幾何示意圖

### 2.3.4 基於深度學習的方法

隨著深度學習迅速發展，電腦視覺中的各種任務發展到更高的水平，在邊緣檢測中，從 UC Berkeley 的標準來看[19] [20]，基於 CNN 的方法得出很好的效果[21] [22] [23] [24]，儘管電纜線檢測的任務和自然圖像中的國家/地區檢測不同，但這些方法具有學習多尺度特徵和感知全局訊息的強大能力，其特性還是可以用來電纜線的檢測，這些基於深度學習的偵測方法通常包含幾個階段，包括前處理、特徵提取和分類，如圖 2-9。

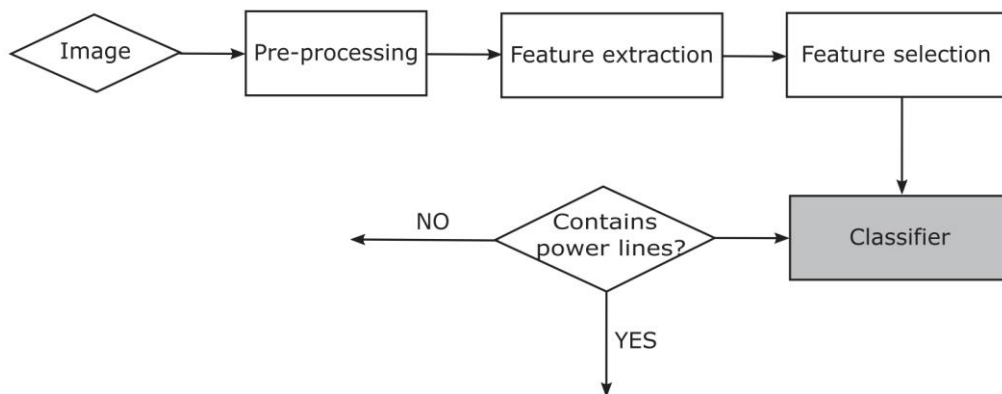


圖 2-18. 基於深度學習方法的電纜線偵測流程圖

Madaan 等人[25]利用擴張卷積(dilated convolution)的電纜線檢測網路架構，Chen 等人[26]則使用不同架構的擴張卷積設計多個網路，並在實際環境和合成數據集上進行評估，以找到最佳網路，上述兩個方法[25][26]皆利用語意分割取得新的進展，與傳統方法相比大大提高其效果，並且在 NVIDIA Jetson TX2 上非常有效，但是，如果不使用先驗的訊息和電纜線的結構，輸出的干擾仍然相當顯著。

### 2.3.5 Encoder-Decoder 架構

在圖像切割模型中，Encoder-Decoder 是一種有效的架構，Encoder 進行特徵擷取，Decoder 則是將特徵圖的解析度還原。以 Fully Convolutional Network [27] 為例如圖 2-19 所示，將一些分類模型 AlexNet、VGGNet、GoogLeNet [5] 的 Fully Connected Layer 替換成 Fully Convolutional Layer 在圖像分割的速度與結果上均有進步。

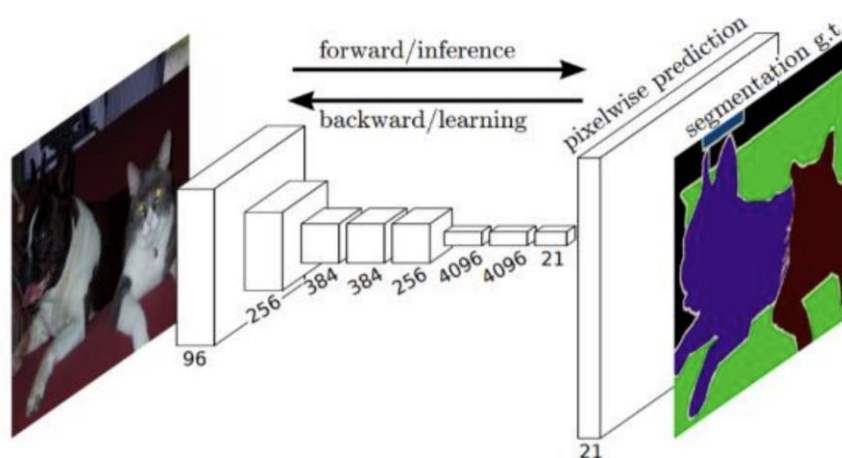


圖 2-19 Encoder-Decoder 架構

### 2.3.6 RTFNet

電纜線檢測任務需要龐大的資料集，但是目前可用的電纜線資料集非常有限，上述利用 Yetgin 資料集所進行的實驗外，在辨識其他電纜線的圖片時效果不彰，

我們想透過其他模型架構來構成多個通道的模型編碼器，本研究將利用 Sun 等人所提出的編碼-解碼器架構 RTFNet，其利用兩個編碼器 ResNet 分別對 RGB 和 Thermal 影像進行特徵的擷取，架構如圖 2-20.所示，在編碼階段透過其特徵的累加來實現，而解碼階段，提出 Upception 網路將特徵的解析度逐漸回復原始圖像的解析度，採用 Encoder-Decoder 架構，將光學影像與熱影像透過兩個 Encoder 進行特徵擷取，Decoder 還原影像尺寸並且將熱影像 Encoder 的特徵圖加到對應的光學影像 Encoder 中，這樣的融合方法對模型的切割結果有所改善。

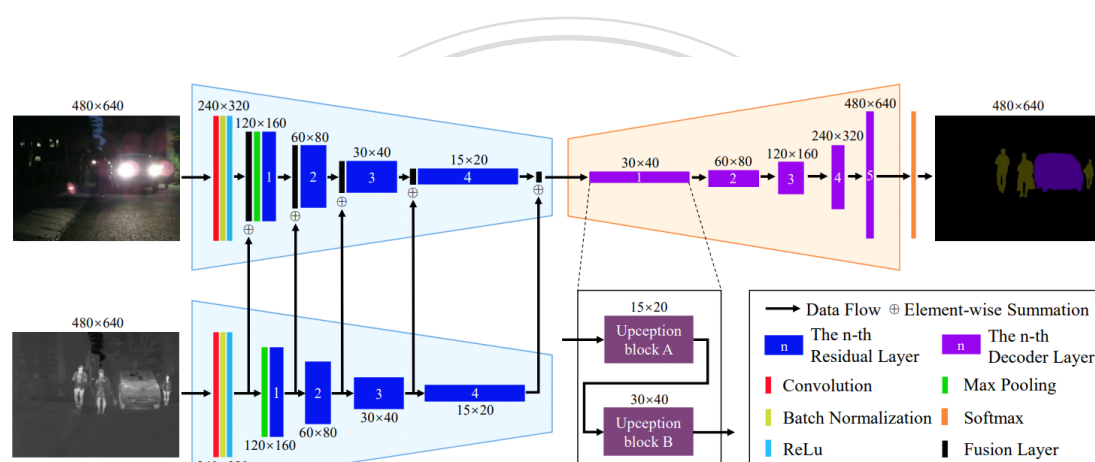


圖 2-20. RTFNet 模型架構圖

然而，此架構並未考慮除了將熱影像的特徵融合到光學影像特徵外，也可以將光學影像的特徵融合到熱影像中，達到互相幫助的效果。本研究將利用其他邊緣檢測方法如：Canny、Sobel 等等方法使其邊緣輪廓更加明顯，如圖 2-21.及 2-22.所示，並改善 RTFNet 之架構並未考慮將熱影像的特徵融合到光學影像特徵，我們先將邊緣強化圖像先經過 SEAdaIN module 再融合到光學影像，同時也將光學影像的特徵透過 SEAdaIN module 融合到邊緣強化的影像中，達到互助的效果如圖 2-23.所示，提供模型更完善的資訊來提高模型辨識的準確度。

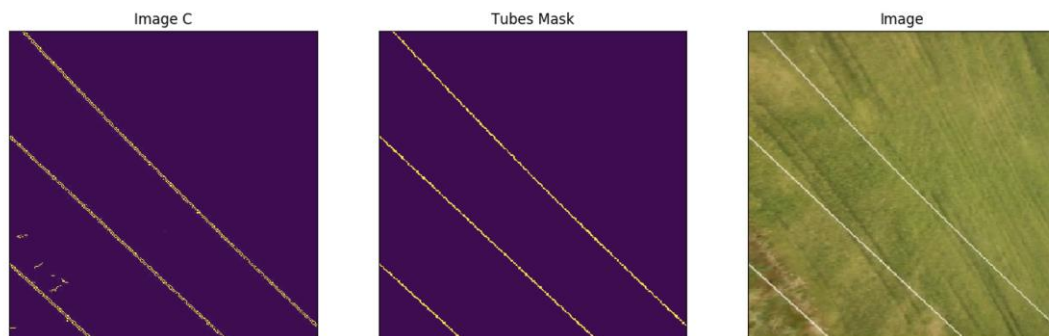


圖 2-21.邊緣檢測範例 1 (左：Canny 圖像，中：原標記圖像 右：原圖)

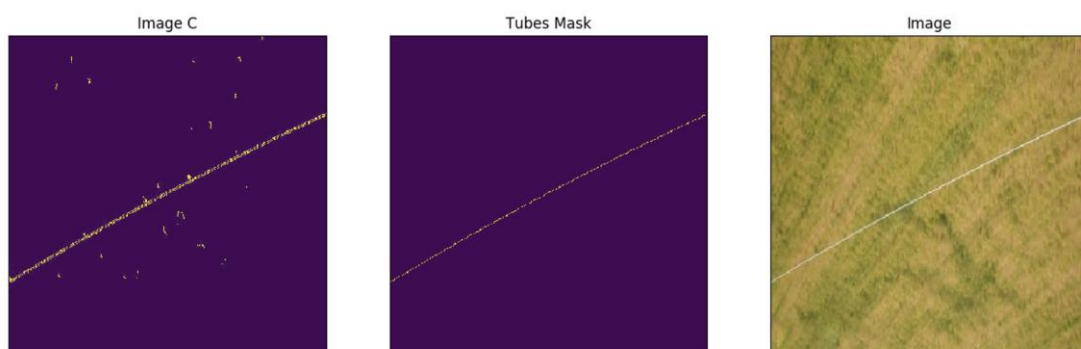


圖 2-22. 邊緣檢測範例 2 (左：Canny 圖像，中：原標記圖像 右：原圖)

### 2.3.7 低階特徵 Domain Adaptation (SEAdIN)

此模組如圖 2-23，其整合 SELayer [33] 的概念與風格轉換的做法，將來源影像的特徵資訊透過此模組轉換為目標影像的特徵資訊的值域，避免來源特徵的資訊過度影響目標影像特徵資訊的值域，同時達到輔助的效果。

注意力機制(Attention mechanism)指的是對 Layer 輸出的特徵圖進行加權，透過該機制可增加特徵表達的能力，以提升下游任務的準確度。主要方式有 spatial-wise attention 與 channel-wise attention，其對不同維度的特徵圖做選擇性的強化。為了避免 spatial-wise attention 耗費過多的運算資源，因此我們的解決方案使用 channel-wise attention，針對特徵圖(Feature Map)進行選擇性的加權。具體來說，我們採用的是 Squeeze-and-Excitation Layer[33]，該方法屬於自注意力(self-

attention)機制，會判斷特徵圖重要與否，透過重新分配特徵圖的權重，藉此聚焦於重要的特徵圖。

自適應實例正規化(Adaptive Instance Normalization)[34]是由 Huang 與 Belongie 提出，起初是被用來做風格的轉換，透過轉換 source domain 上每個特徵圖統計資訊至 target domain，使來源影像的風格特徵能被遷移至其他影像上，近年來其變形架構也被應用在領域自適應(domain adaptation)的視覺任務。在此解決方案中，我們將注意力機制加權後的特徵圖做為 AdaIN 的輸入，藉由一個交叉的結構，將相同尺度的邊緣強化影像分支的特徵圖轉化至可見光分支；在可見光分支同樣執行該遷移架構，使之輔助邊緣強化影像分支。我們將該融合架構稱為 SE-AdaIN(如圖 2-23.)，並將其用於多尺度的融合，從低階的紋理特徵到高階的語意特徵，透過 SE-AdaIN 協助不同資訊進行互補與增強，進而使準確度提升。

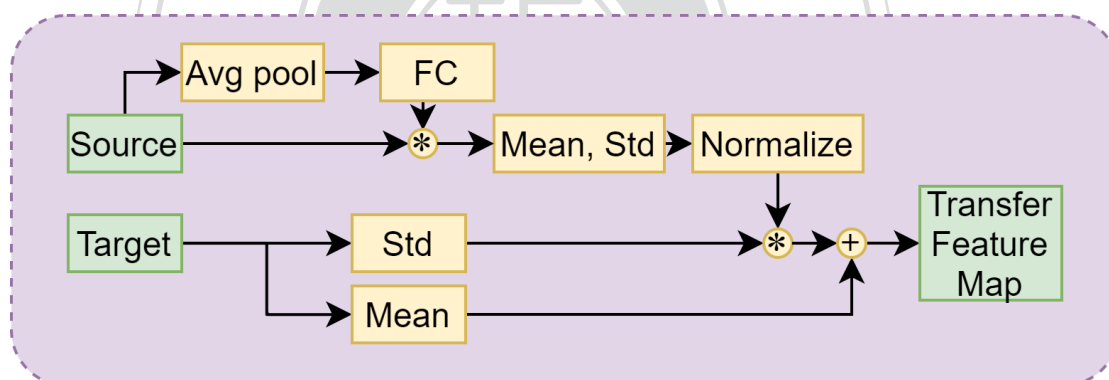


圖 2-23. SEAdIN Module 示意圖

### 2.3.8 高階特徵自注意力機制 Strip Pooling Module

在 2.3.7 中有說明，因為 spatial-wise attention 的計算成本相當高，所以我們並不採用該機制，但模型的長距離語意資訊會影響語意分割的辨識準確度，因此在本研究中我們選擇導入 Strip pooling module [35]，一種相當輕量的模組架構，透過強化多尺度、縱向與橫向的資訊來捕捉長距離依賴資訊，以提升模型的效果。具體來說，Strip pooling 相較於一般的 Spatial Pooling，

使用長條狀的  $1 \times N$  或  $N \times 1$  的池化 (Pooling) 與一維的卷積運算，利用極少的參數量建立長距離的特徵依賴關係。同時，為了使特徵表達更具判別力，採用輕量化的 Pyramid Pooling Module (PPM) 去捕捉局部的特徵，不同 kernel-size 的池化運算能建立不同位置的關聯。我們藉由將該 Strip pooling module 模組應用至 Encoder 的最後一層，提升模型在捕捉語意上下文時，同時建構特徵圖不同位置的短距離與長距離關係，架構圖如圖 2-24 所示。

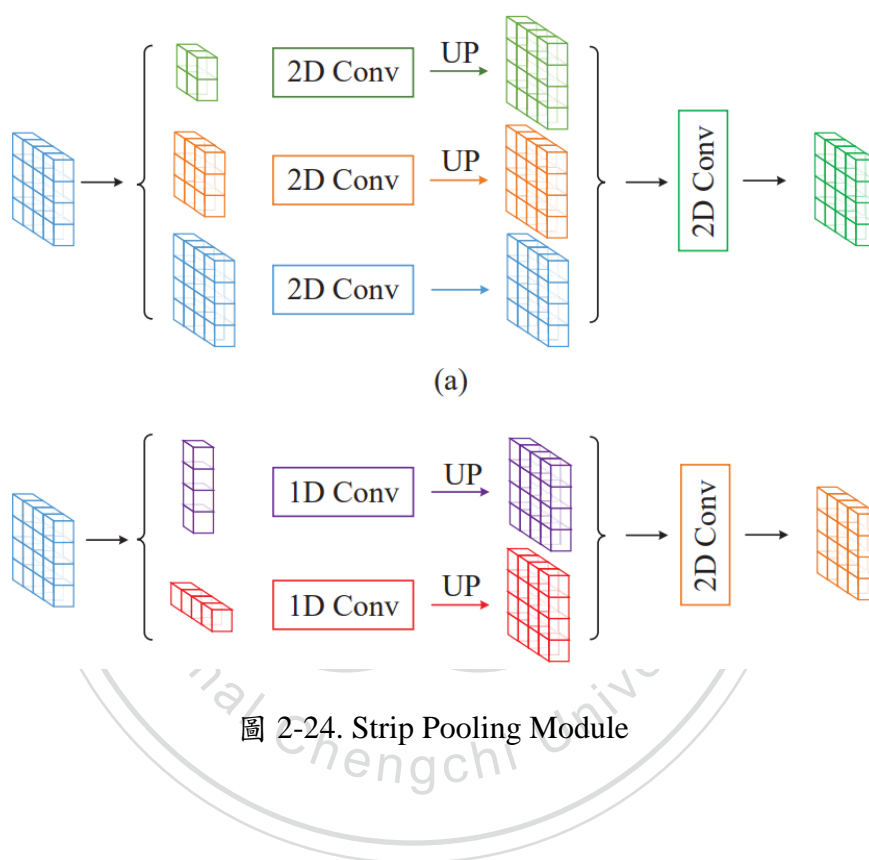


圖 2-24. Strip Pooling Module

## 2.4 小結

本論文透過使用 RTFNet 為開發基礎改善傳統電纜線檢測的任務，透過 SE-AdaIN 協助不同資訊進行互補與增強，並在語意分割模型中探索 U-Net 之模型架構，也嘗試將邊緣強化之影像加入語意分割模型之效果，使電纜線檢測的任務能夠透過深度學習的方式，有效節省人力成本，高效且精準地完成電纜線檢測任務。

## 第三章 電纜線偵測與資料集介紹

為了使資料集的資料更利於語意分割模型訓練，我們利用基於傳統梯度的邊緣檢測器將圖像做前處理，使語意分割模型能更準確的標註出電纜線，藉此幫助模型完成電纜線檢測任務。在本章節中，將詳細說明邊緣檢測之方法與電纜線資料集介紹。

### 3.1 Canny Edge Detector

Canny Edge Detector 是一種邊緣偵測運算方法，它採用多級算法偵測圖像中的邊緣。它是由 Canny 在 1986 年開發[28]，其中提出了邊緣偵測的計算理論，並解釋其工作的計算原理。

Canny Edge Detection 演算法分為五個步驟：

步驟一：降噪

由於所有邊緣偵測結果都容易受到圖像噪聲的影響，因此必須濾除雜訊防止偵測錯誤，應用高斯濾波器對圖像進行平滑處理，將高斯濾波核（Gaussian filter kernel）與圖像進行卷積運算，以減少邊緣偵測器上明顯雜訊的影響。高斯濾波核的大小取決於預期的模糊效果，核的尺寸越大，偵測器對雜訊的靈敏度就越低，大小為 $(2k+1) \times (2k+1)$ 的高斯濾波核方程式為以下公式 3-1.所示：

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right); 1 \leq i, j \leq (2k+1) \quad (\text{式 3-1.})$$

步驟二：梯度計算

由於圖像中的邊緣可能為各種方向，因此 Canny 演算法使用四個 filter 來偵測模糊圖像中的水平、垂直以及對角線之邊緣，偵測運算符號水平方向( $G_x$ )垂直方向( $G_y$ )，邊緣強度及方向的計算如下公式 3-2 所示：



$$G = \sqrt{G_x^2 + G_y^2}$$

$$\theta = \arctan\left(\frac{G_x}{G_y}\right) \quad (\text{式 3-2.})$$

其  $\arctan$  為反正切函數可求得角度，為了簡化計算我們將求得之角度分為垂直、水平、及兩對角線( $0^\circ, 45.32^\circ, 90.32^\circ, 135.32^\circ$ ) 當角度落在  $0\sim 22.5$  或  $157.5\sim 180$  度及分類為 0 度；角度落在  $22.5\sim 67.5$  及分類為 45 度；角度落在  $67.5\sim 112.5$  及分類為 90 度；角度落在  $112.5\sim 157.5$  度及分類為 135 度。

### 步驟三：非最大抑值

通常邊緣出現在梯度強度高的地方，根據步驟二求出的梯度強度及方向，依序比對圖像中所有像素梯度，找尋邊緣方向上具有最大值的像素，否則將梯度強度設為 0，若像素梯度方向為 0 度則需要再比較其左、右的像素梯度強度；梯度方向為 45 度則要再比較其右上、左下的像素；梯度方向為 90 度則要再比較其上、下的像素；梯度方向為 135 度則要再比較其左上、右下的像素梯度強度。

### 步驟四：雙閾值

將閾值分為高閾值( $T_H$ )和低閾值( $T_L$ )，根據閾值判斷此像素是否為邊緣，判斷的依據有三個如下：

1. 該像素梯度  $< T_L$ ，則該像素不為邊緣
2. 該像素梯度  $> T_H$ ，則該像素為邊緣
3. 該像素介於高低閾值之間，且周圍八個相鄰像素中，存在一個像素的梯度強度  $> T_H$ ，則該像素為邊緣

### 步驟五：滯後邊緣跟蹤

因高強度邊緣的像素是從圖像真實邊緣中擷取，所以其應包含在最終邊緣圖像中，反觀弱強度邊緣像素，因其可以從真實邊緣擷取，也可雜訊/顏色變化中提取。為

了獲取準確結果，應該過濾某些原因引起的弱強度邊緣，透過查看弱強度像素及其鄰近的 8 個相鄰像素，弱強度邊緣若與真實邊緣和高強度邊緣相連通，可由此判斷是否為真實邊緣。

### 3.2 Ground Truth of Powerline Dataset

儘管有關飛機安全的電纜線檢測文獻很豐富[29]，但大多數工作都是透過很少的訓練圖像來進行優化及訓練，有些甚至使用人工合成的圖像，其背後原因為獲取真實的電纜線資料需要龐大的人力，考量到這一需求， Yetgin 等人與土耳其電力公司(TEIAS)合作[30]，從實際飛機上的影片來收集資料，獲得此具有價值的資料集供本研究做實驗，如圖 3-1.及圖 3-2.。

在此資料集中總共包含 400 張紅外光資料(IR)和 400 張可見光資料(VL)，其中各有 200 張有電纜線(TV)，200 張無電纜線(TY)，並將圖片縮放至 512\*512 的大小。IR 資料集中包含電纜線的 IR 圖像、已標記的標準答案(Ground Truth)及兩者圖像之疊加圖像，VL 資料集包含電纜線的 VL 圖像、Ground Truth 及兩者圖像之疊加圖像。IR 及 VL 皆包含常規場景也包含複雜背景之場景。

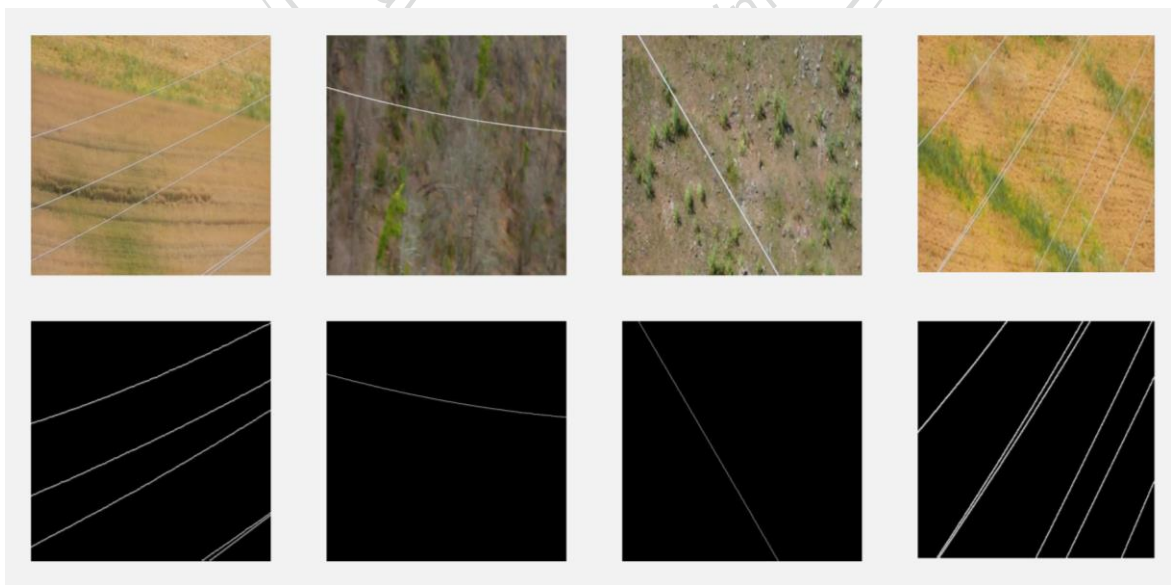


圖 3-1.可見光之原圖及已標記圖片

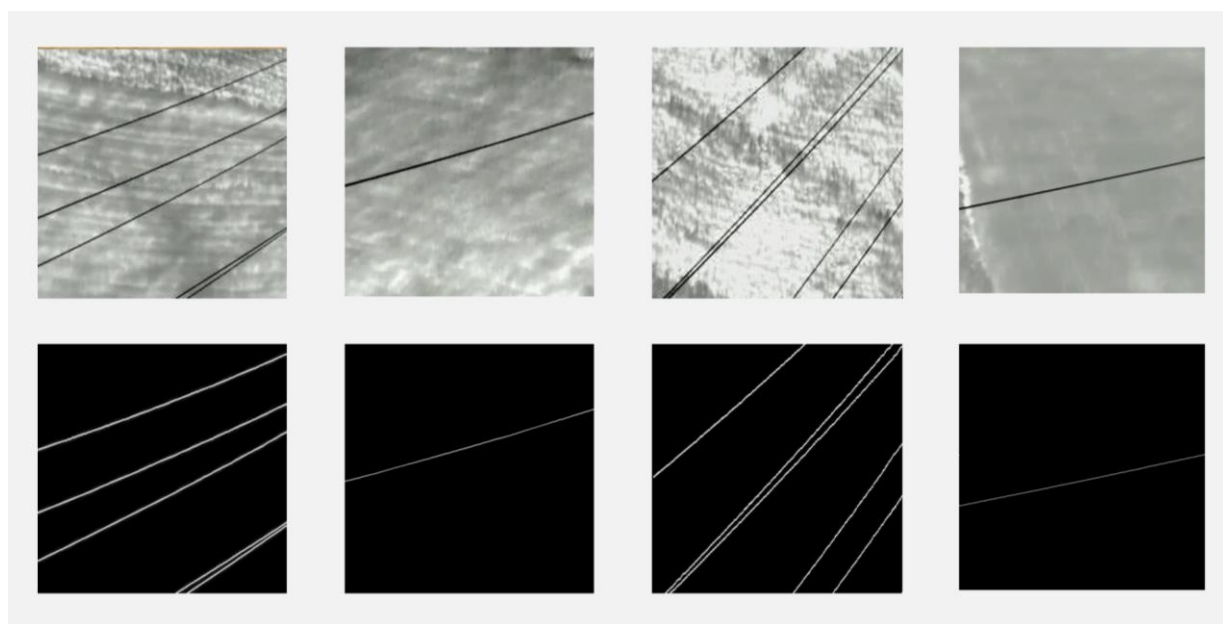


圖 3-2.紅外光之原圖及已標記圖片

### 3.3 PLD-UVA

PLD-UVA[31]為無人機電纜線巡檢的資料集，其包含了城市場景(Power line dataset of urban scene, PLDU)如圖 3-3.和山區場景(Power line dataset of mountain scene, PLDM)如圖 3-4.，以及各場景之 Ground Truth 如圖 3-5.，這些圖片包含複雜的背景和具有不同功能的電纜線，所有圖像都使用手動標籤，也執行了包括旋轉、翻轉和大小的調整可供使用。



圖 3-3. 電纜線分布於城市區場景(PLDU)



圖 3-4. 電纜線分布於山區場景(PLDM)



圖 3-5. 各場景之 Ground Truth

因為其 Ground Truth 電纜線的標記部分只有邊緣標記，中間線材的部分沒有標記到，使模型獲得的 Training Data 和 Ground Truth 訊息有所出入，導致模型準確度不佳，我們將線材完整填滿，如圖 3-6 所示，使其提高實驗的準確度。

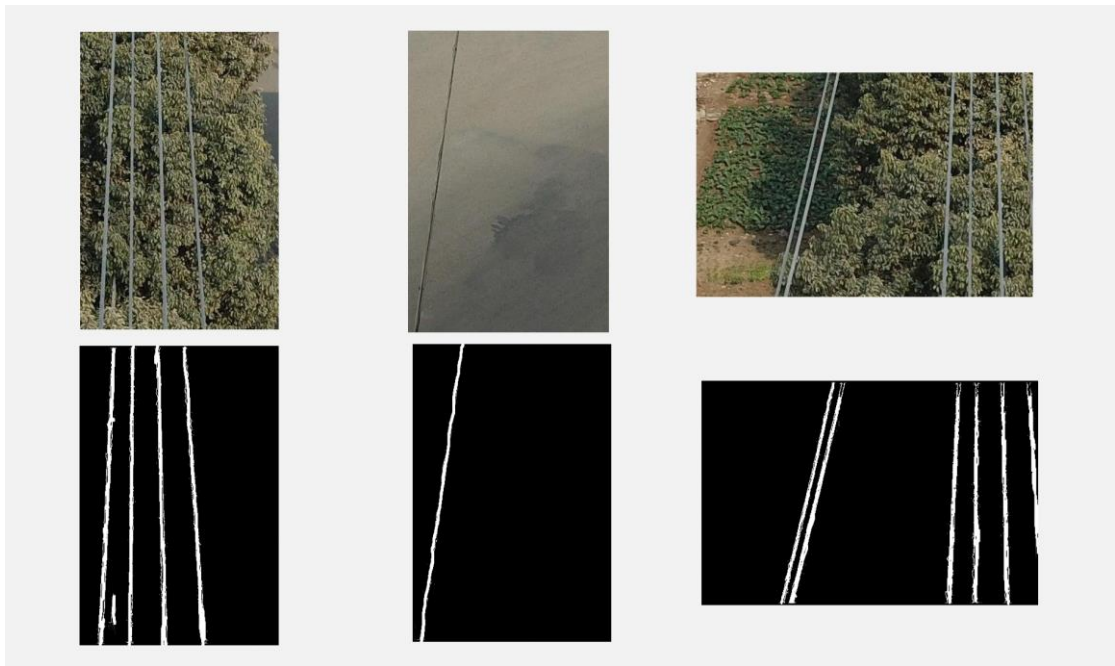


圖 3-6 填滿電纜線之 Ground Truth



## 第四章 研究方法與實驗結果分析

本論文使用 Yetgin 等人所收集之資料集及 PLD-UVA 進行研究[31]，Yetgin 等人資料集中，包含 400 張紅外光資料(IR)和 400 張可見光資料(VL)，其中各有 200 張有電纜線(TV)，200 張無電纜線(TY)，並將圖片縮放至 512\*512 的大小。本論文實驗使用其可見光並含有電纜線之資料作為訓練的資料集，Yetgin 等人於 2019 年也提出使用深度卷積網路訓練此資料及進行電纜線的辨識，他們將資料集分割 10% 作為測試資料，我們依循其比例將 180 張歸為訓練資料集和 20 張的測試資料，也另外嘗試分割 20% 作為測試資料並將兩組實驗各自進行交叉驗證 (Cross Validation)，而 Nguyen 等人[14]於 2020 所發表的 LS-Net 即利用 GTD 資料做為測試集進行語意分割實驗，本論文希望透過新的模型將其辨識結果提升。

首先我們將 GTD 和 PLD 分別應用於單一 Segmentation Model 中觀察其結果，於 PLD 資料集表現相對優異，但 GTD 中表現不如 LS-Net，於是我們提出 Dual Segmentation Model 應用於 GTD 本論文中，透過兩個 Encoder 將邊緣強化影像及光學影像融合，使得模型資訊得以互補，進而提升分割模型的準確度，使其超越 LS-Net 之表現。

### 4.1 電纜線影像語意分割結果

將資料集應用於電纜線之語意分割模型，語意切割會針對影像中每一個像素點進行分類，透過理解像素所屬之類別資訊，可以更容易了解影像中的鄰近區域關係資訊，本論文嘗試不同語意分割模型之架構，如：PSPNet 及 U-Net，從中發現由於電纜線與背景比例差異過大，導致現有方法在定位精準度的效果仍待改進，因此本論文測試不同編碼器，如：SE ResNeXt50 32x4d、EfficientNet-b5，其中加入 Channel attention 的 SE ResNeXt50 32x4d 搭配 U-Net 模型效果最佳，如圖 4-1. 及 4-2，實驗結果如表 4-1 至 4-4 所示(採訓練樣本數:測試樣本數為 8:2 與 9:1 分別測試，最終結果均經過交叉驗證)。

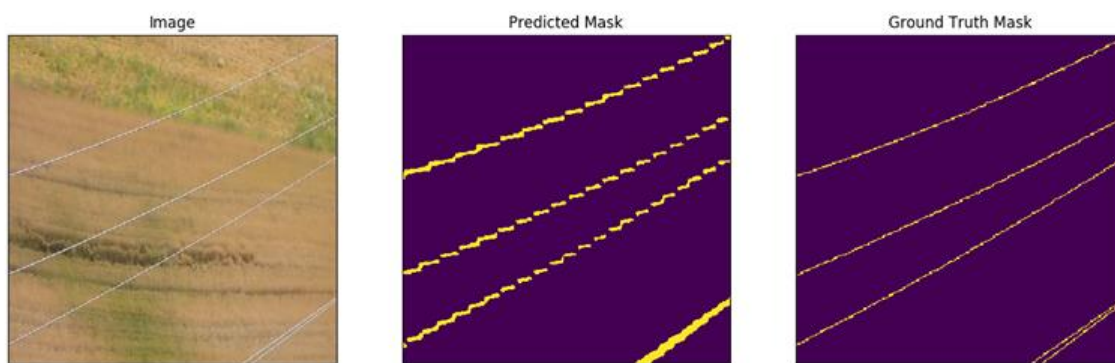


圖 4-1.電纜線語意分割結果 1

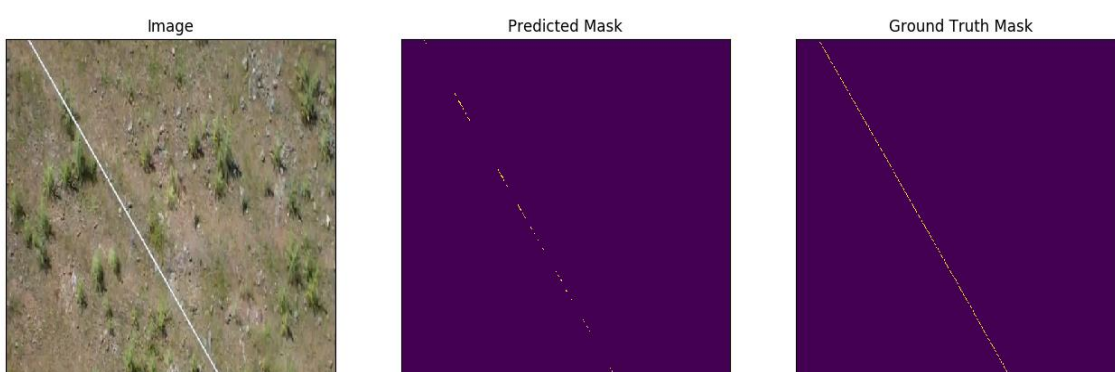


圖 4-2.電纜線語意分割結果 2

Architecture	Framework		
	U-Net		LS-Net
Encoder	SE ResNeXt50 32x4d	EfficientNet-b5	-
Train : Test	8 : 2		-
Params	26M	28M	-
IOU score	0.4912	0.4851	-
Precision	0.5467	0.5761	0.8004
Recall	0.8044	0.8315	0.5368
F-score	0.6371	0.6291	0.5940

表 4-1. U-Net 語意分割模型訓練結果 (訓練樣本數:測試樣本數=8:2)

Architecture	Framework		
	U-Net		LS-Net
Encoder	SE ResNeXt50 32x4d	EfficientNet-b5	-
Train : Test	9 : 1		-
Params	26M	28M	-
IOU score	0.5089	0.4987	-
Precision	0.5742	0.5906	0.8004
Recall	0.8261	0.8492	0.5368
F-score	0.6318	0.6451	0.5940

表 4-2. U-Net 語意分割模型訓練結果(訓練樣本數:測試樣本數=9:1)

Architecture	Framework		
	PSPNet		LS-Net
Encoder	SE ResNeXt50 32x4d	EfficientNet-b5	-
Train : Test	8 : 2		-
Params	26M	28M	-
IOU score	0.2547	0.2218	-
Precision	0.2916	0.2719	0.8004
Recall	0.6918	0.7561	0.5368
F-score	0.4064	0.3695	0.5940

表 4-3.PSPNet 語意分割模型訓練結果(訓練樣本數:測試樣本數=8:2)



Architecture	Frameworks		
	PSPNet		LS-Net
Encoder	SE ResNeXt50 32x4d	EfficientNet-b5	-
Train : Test	9 : 1		-
Params	26M	28M	-
IOU score	0.2817	0.2436	-
Precision	0.3142	0.2814	0.8004
Recall	0.7175	0.7761	0.5368
F-score	0.4128	0.3948	0.5940

表 4-4.PSPNet 語意分割模型訓練結果(訓練樣本數:測試樣本數=9:1)

觀察上述實驗結果後，我們發現 GTD 中多數 Ground Truth 資料標記尚可，但某些 Ground Truth 的標記並不完全，其線段不連續，導致某些資料預測出來的結果非常不佳，如圖 4-2，反映在數據上即 Recall 夠高但其 Precision 太低，從而拖累了整體的 F-score，為了改善這個問題，我們便將 GTD 之 Ground Truth 進行重新標記，將線段標記完整如圖 4-3.及圖 4-4.，並觀察其實驗結果獲得了不錯的改善，使得 F-score 超越了 LS-Net 的表現，雖然 F-score 及預測的結果都有相當的改善，但在 Precision 部分和 LS-Net 仍有差距，如表 4-5.至表 4-8。

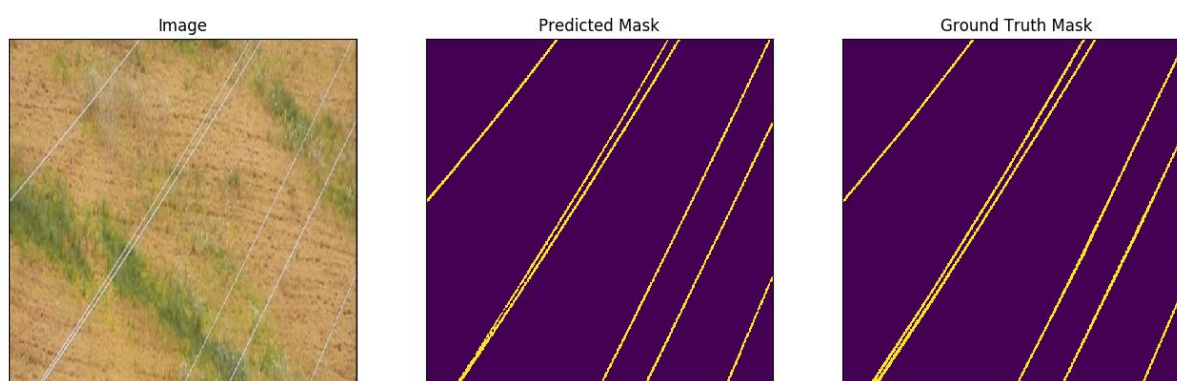


圖 4-3. 重新標記 GTD 之 Ground Truth 之語意分割結果 1

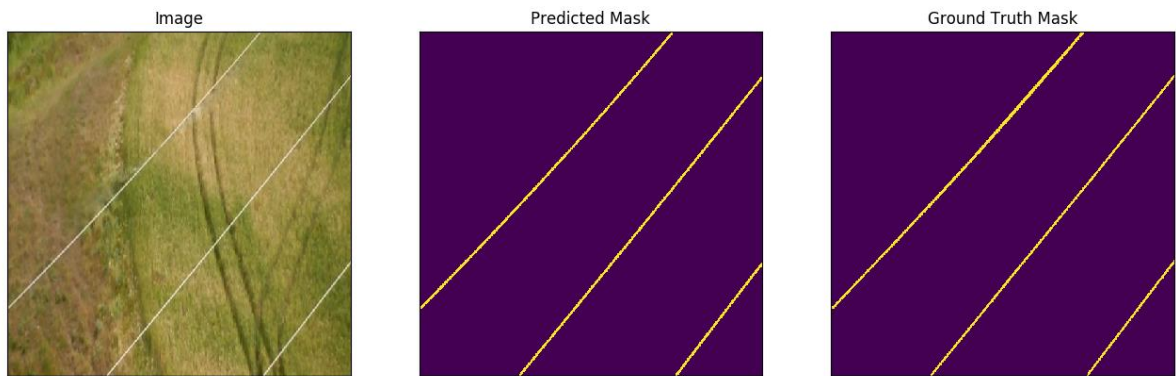


圖 4-4. 重新標記 GTD 之 Ground Truth 之語意分割結果 2

Architecture	Framework		
	U-Net		LS-Net
Encoder	SE ResNeXt50 32x4d	EfficientNet-b5	-
Train : Test	8 : 2		-
Params	26M	28M	-
IOU score	0.5053	0.5184	-
Precision	0.5967	0.5991	0.8004
Recall	0.7791	0.8048	0.5368
F-score	0.6719	0.6695	0.5940

表 4-5. U-Net 重標 ground truth 後 GTD 單一 segmentation model 結果

(訓練樣本數:測試樣本數=8:2)

Architecture	Framework		
	U-Net		LS-Net
Encoder	SE ResNeXt50 32x4d	EfficientNet-b5	-
Train : Test	9 : 1		-
Params	26M	28M	-
IOU score	0.5551	0.5669	-
Precision	0.6217	0.6079	0.8004
Recall	0.8282	0.8183	0.5368
F-score	0.6987	0.6988	0.5940

表 4-6. U-Net 重標 ground truth 後 GTD 單一 segmentation model 結果  
(訓練樣本數:測試樣本數=9:1)

Architecture	Frameworks		
	PSPNet		LS-Net
Encoder	SE ResNeXt50 32x4d	EfficientNet-b5	-
Train : Test	8 : 2		-
Params	26M	28M	-
IOU score	0.4392	0.4581	-
Precision	0.6240	0.5376	0.8004
Recall	0.6843	0.7964	0.5368
F-score	0.6517	0.6581	0.5940

表 4-7. PSPNet 重標 ground truth 後 GTD 單一 segmentation model 結果  
(訓練樣本數:測試樣本數=8:2)

Architecture	Frameworks		
	PSPNet		LS-Net
Encoder	SE ResNeXt50 32x4d	EfficientNet-b5	-
Train : Test	9 : 1		-
Params	26M	28M	-
IOU score	0.4816	0.5023	-
Precision	0.6572	0.5496	0.8004
Recall	0.7349	0.8014	0.5368
F-score	0.6747	0.6817	0.5940

表 4-8. PSPNet 重標 ground truth 後 GTD 單一 segmentation model 結果  
(訓練樣本數:測試樣本數=9:1)

我們一開始先將 PLD-UVA 中市區的原始資料集 (PLDU) 直接應用於上述語意分割模型，且使用的架構及編碼器也相同，但其表現結果不如預期，如圖 4-5 及圖 4-6 所示，此原始資料的 Ground Truth 只有將電纜線的輪廓標示出來，所以我們試著將其 Ground Truth 電纜線輪廓進行填滿，如圖 4-7，其效果有顯著的提升。後續加入山區資料集 (PLDM) 一起將整個資料集 (PLD) 進行實驗，如表 4-9. 及表 4-10. (訓練樣本數:測試樣本數按照原資料集為 8:2)，模型的預測結果如圖 4-8. 及 4-9.

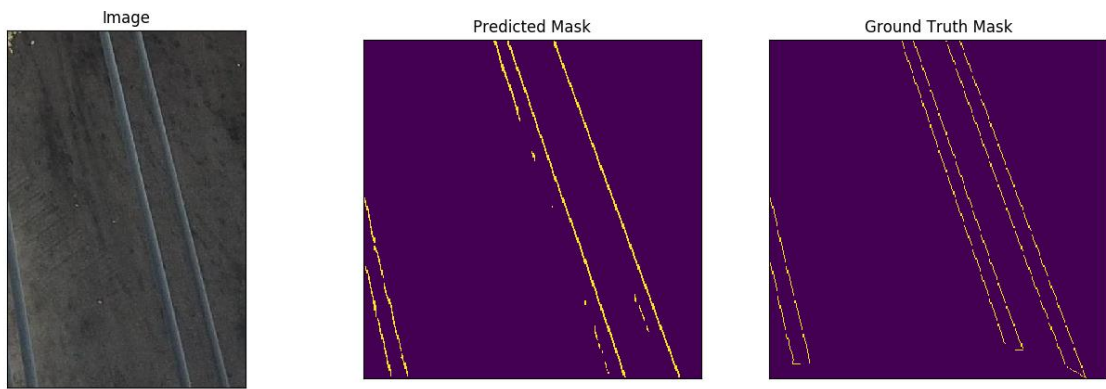


圖 4-5 U-Net 預測 PLDU 原始資料的預測結果 1

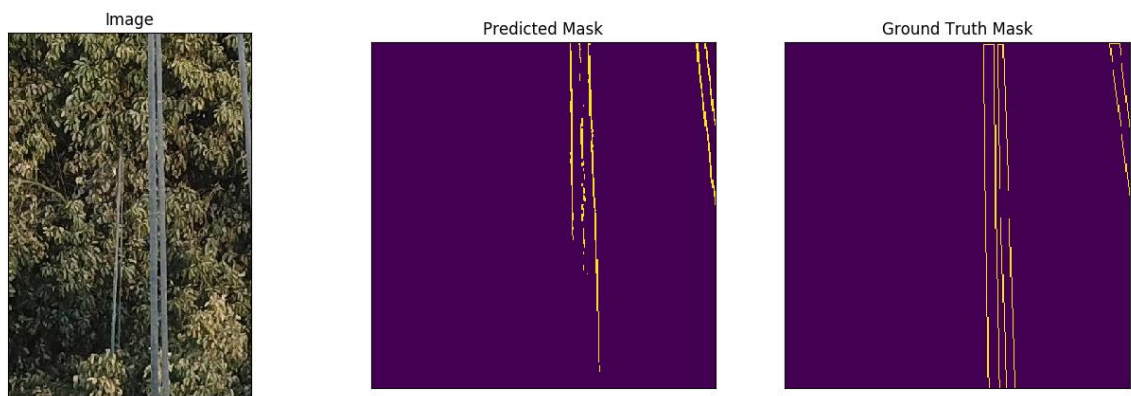


圖 4-6. U-Net 預測 PLDU 原始資料的預測結果 2

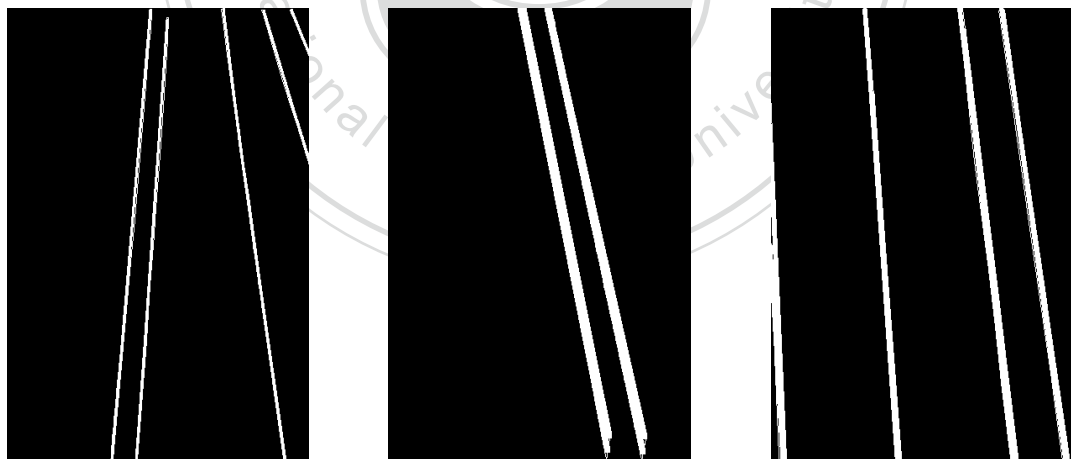


圖 4-7. 將 PLD 之 Ground Truth 填滿

Architecture	Framework	
	U-Net	
Encoder	SE ResNeXt50 32x4d	EfficientNet-b5
Params	28M	28M
IOU score	0.7893	0.7324
precision	0.8624	0.8798
recall	0.9054	0.8229
F-score	0.8806	0.8369

表 4-9. U-Net 預測 PLD 資料集填滿標記過後訓練結果

(訓練樣本數:測試樣本數=8:2)

Architecture	Framework	
	PSPNet	
Encoder	SE ResNeXt50 32x4d	EfficientNet-b5
Params	28M	28M
IOU score	0.7893	0.7324
precision	0.8624	0.8798
recall	0.9054	0.8229
F-score	0.8806	0.8369

表 4-10. PSPNet 預測 PLD 資料集填滿標記過後訓練結果

(訓練樣本數:測試樣本數=8:2)

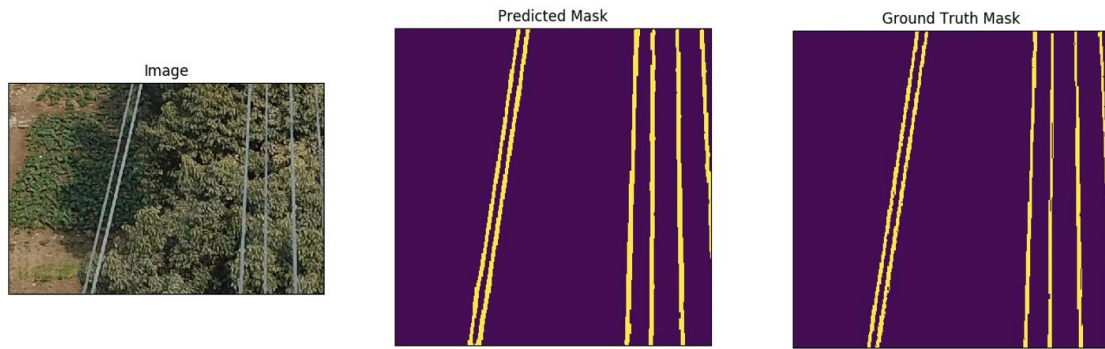


圖 4-8.U-Net 預測 PLD 標記過後預測結果 1

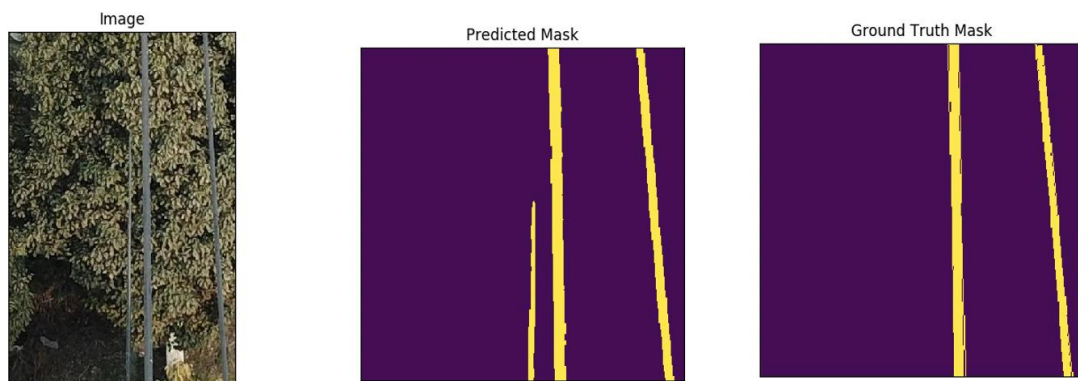


圖 4-9. U-Net 預測 PLD 標記過後預測結果 2

我們也將上述兩個資料集混合在一起訓練進行預測，觀察其結果會如何呈現，可以發現在 Ground Truth Powerline Dataset(GTD)在電纜線較明顯的圖像中辨識的效果還不錯，但若其電纜線之間隔太相近，會將兩條線填滿，且預測結果整體要比 Ground Truth 要來的粗，辨識效果較差，如圖 4-10.及 4-11.，推測是在 PLD 底下因其輪廓填充極其明顯，且線段粗很多所導致，但整體 PLD 辨識效果顯著，如下圖 4-12.及 4-13.，混合模型的預測結果如表 4-11.及 4-12.所示(訓練樣本數: 測試樣本數為 8:2)。不管採用哪一種語意分割模型，GTD 的表現都不如預期，即使有足夠高的 Recall，但預測出來的 Precision 太低，導致 F-score 過低，仍須改善。

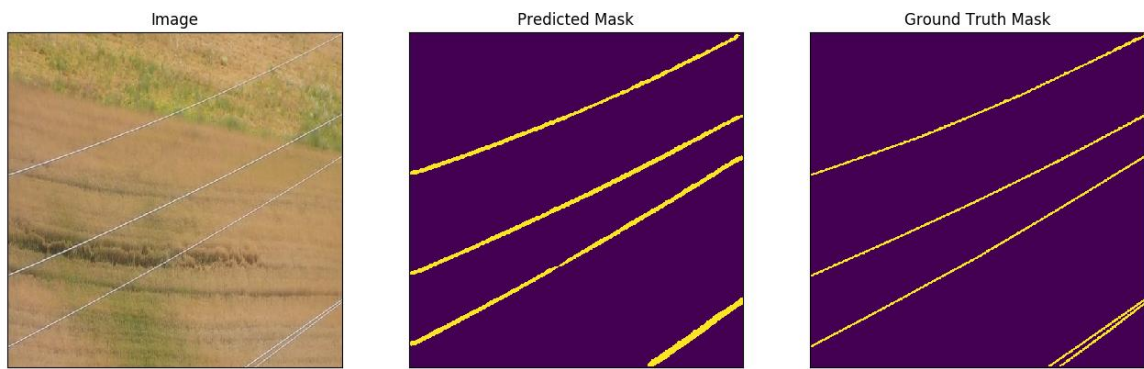


圖 4-10. 混合模型預測 Ground Truth Powerline Dataset1

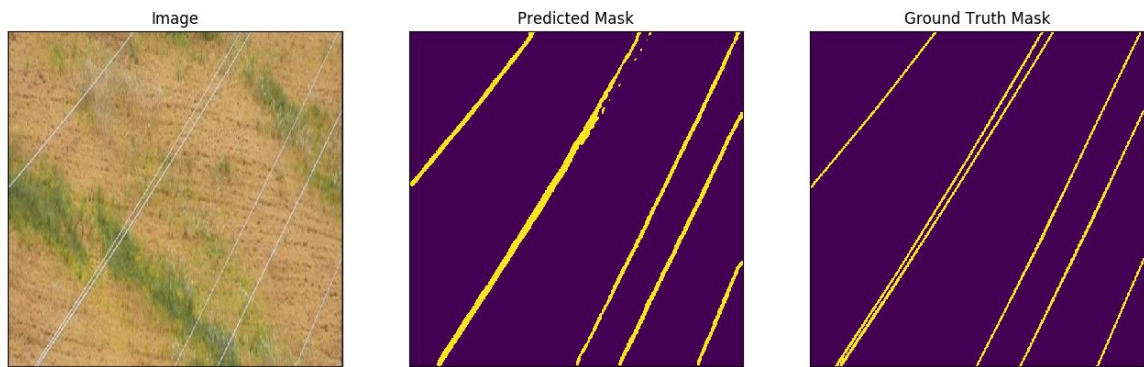


圖 4-11. 混合模型預測 Ground Truth Powerline Dataset2

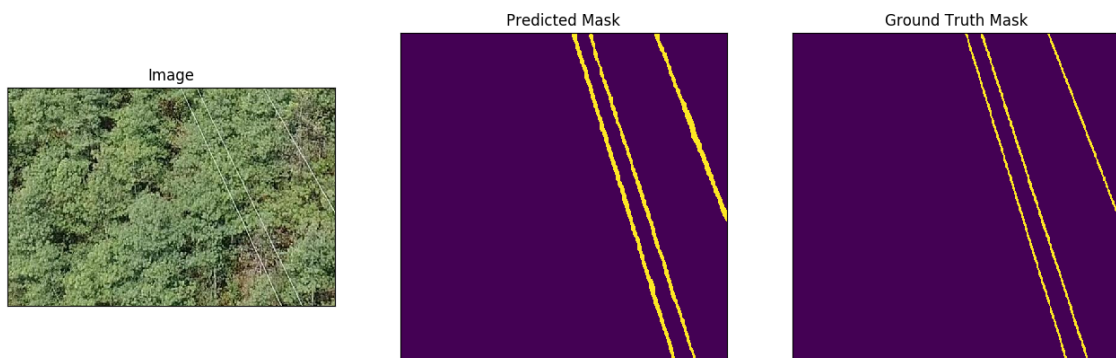


圖 4-12. 混合模型預測 PLD-UVA Dataset



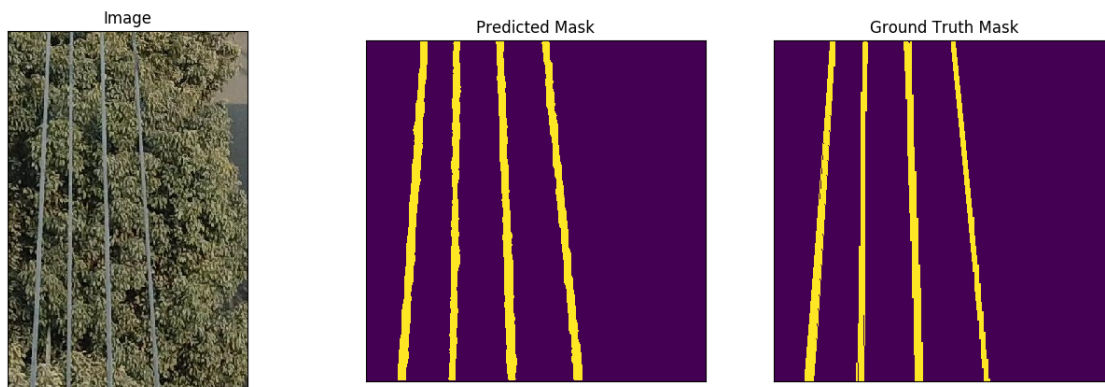


圖 4-13. 混合模型預測 PLD-UVA Dataset2

Architecture	Framework	
	U-Net	
Encoder	SE ResNeXt50 32x4d	
Dataset	Ground Truth Powerline Dataset	PLD
IOU score	0.4729	0.6216
precision	0.5087	0.7206
recall	0.8788	0.8356
F-score	0.6336	0.7455

表 4-11. 混合訓練的模型對兩個資料集進行預測結果

(訓練樣本數:測試樣本數=8:2)

Architecture	Framework	
	PSPNet	
Encoder	SE ResNeXt50 32x4d	
Dataset	Ground Truth Powerline Dataset	PLD
IOU score	0.3421	0.5897
precision	0.4584	0.7077
recall	0.6940	0.7934
F-score	0.5032	0.7284

表 4-12. 混合訓練的模型對兩個資料集進行預測結果

(訓練樣本數:測試樣本數=8:2)

根據混合訓練結果，在 PLD 和 Ground Truth Powerline Dataset 中，由於極細的纜線上輪廓較不明顯，在混合訓練中沒辦法藉由 PLD 的高準確率來改善 GTD 的準確率，以至於辨識效果仍然較差，為解決此一情況，本研究希望藉由將極細的圖像透過其他邊緣檢測的方式使其輪廓更加明顯，使模型更加容易取得資訊進而提升其辨識準確度。

## 4.2 Dual Segmentation 模型架構

本模型使用 RTFNet [32]為基礎，在低階特徵利用 Domain Adaptation 做資訊融合，同時在高階特徵採用注意力機制建構空間上的關係，藉由結合上述兩個 Encoder，讓語意層面的資訊得以互補，進而提升分割模型的準確度。

加入新的元件後必然使整體運算的複雜度變高，我們在 GeForce RTX 2080 上測試，模型的執行速度為 9.1FPS，相較於傳統人力巡檢電纜線，成果可大幅降低其所需之人力資源與時間。

Dual Segmentation 模型架構如圖 4-19，使用 Resnet-50 作為 Backbone，建構類似於 FCN 的架構。Encoder 部分會有兩個，一個針對光學影像提取特徵，一個針對邊緣強化影像擷取特徵。每個 Encoder 一開始會先進行 Convolution、Batch Normalize、ReLU 的步驟，並將 Feature Map 導入 SEAdaIN Module 後，與另一個 Encoder 的 Feature Map 做加法，讓光學影像與邊緣強化影像在提取特徵時所得到的 Feature Map 資訊可以交錯融合，達到輔助的效果。繼續做 Max Pooling 降低解析度後，輸入至 Resnet 中的第一層，並且一樣做資訊的融合，重複做第二層至第四層，得到 Encoder 部分最終的 Feature Map。在 Encoder 與 Decoder 中間，我們使用 Strip Pooling Module，強化高階的語意特徵，並將結果繼續透過 Decoder Layer 將解析度放大回原本的影像尺寸，最終輸出切割結果。我們使用 Pytorch、CUDA 與 cuDNN 套件進行訓練，使用的 Batch size 為 2，架構中 Resnet-50 的 Backbone 參數使用預訓練值，其他參數則是用 Xavier 初始化。

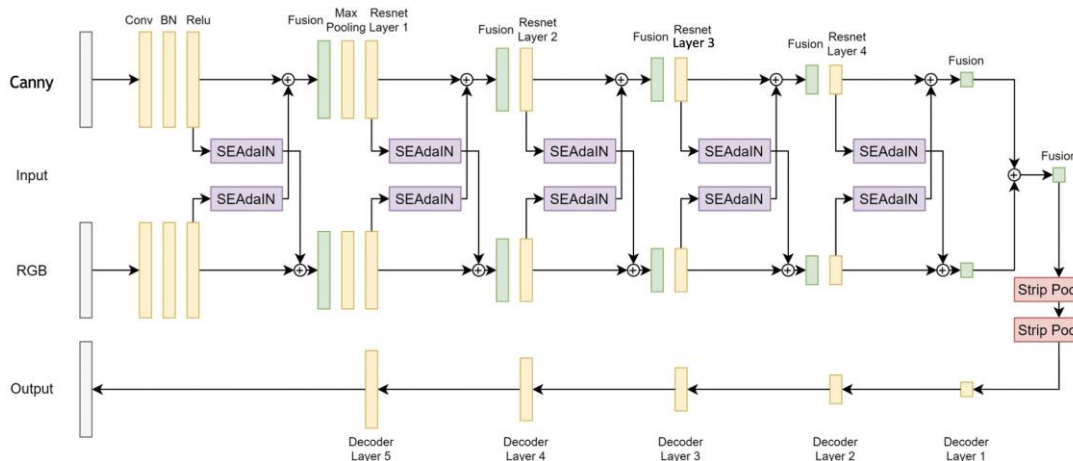


圖 4-14. Dual Segmentation Model 架構示意圖

### 4.3 Ablation Study

在上述對 GTD 的實驗中，雖然重新標記資料之後的 F-score 有些許提升，但其 Precision 相比 LS-Net 還是有很大的不足，為此我們將利用 Dual Segmentation Model (DSM) 以改善 Precision 低下的問題，我們分別對 DSM 的架構做 Ablation experiment，將其分以下五種狀況進行分析：

1. 兩個 Encoder 皆做 SEAdaIN，最後也做 Strip Pooling(原模型)
2. 兩個 Encoder 皆不做 SEAdaIN，最後也不做 Strip Pooling
3. 兩個 Encoder 皆不做 SEAdaIN，只在最後做 Strip Pooling
4. RGB 影像之 Encoder 做 SEAdaIN，最後不做 Strip Pooling
5. RGB 影像 Encoder 做 SEAdaIN，最後也做 Strip Pooling

F-score 最高的模型預測結果如圖 4-20. 4-21 及 4-22 所示，實驗結果如表 4-13. 至 4-16，其結果相比單一 Segmentation Model 中的 Precision 有大幅的提升，在五種組合中，原模型雖然沒有最高的 Precision，但有最佳的 Recall 且整體數據表現較為平均，並優於 LS-Net。

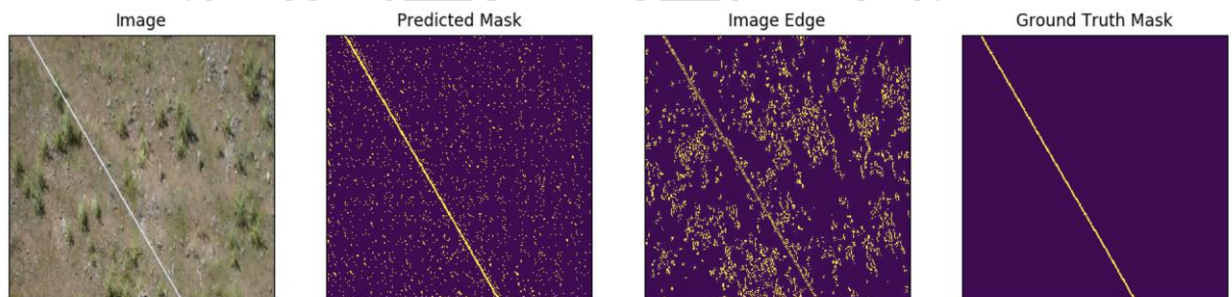


圖 4-15. DSM 預測 GTD 之結果 1

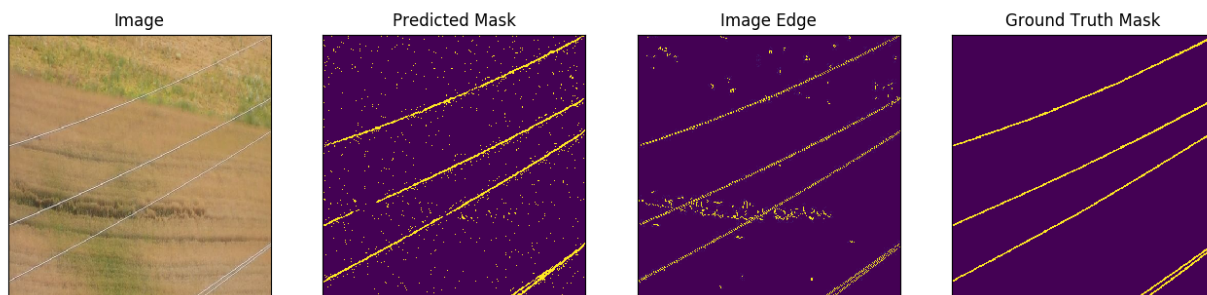


圖 4-16. DSM 預測 GTD 之結果 2

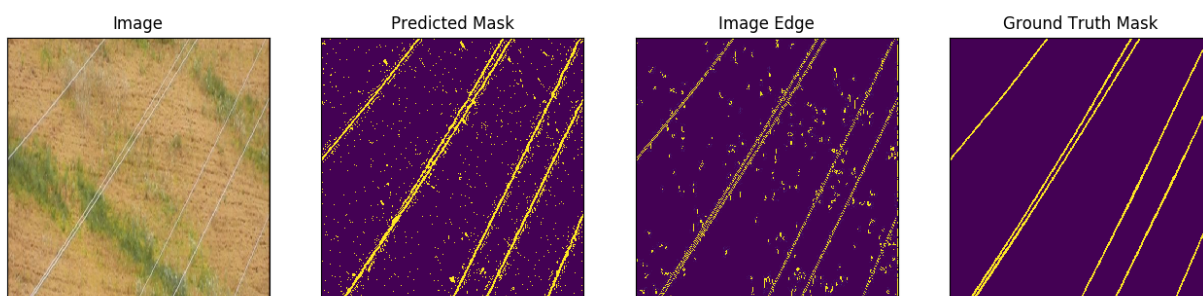


圖 4-17. DSM 預測 GTD 之結果 3

Architecture	Framework			
	Dual Segmentation Model			LS-Net
Encoder	Resnet50			-
Train : Test	8:2			
Module	原模型	(X)SEAdaIN (X)Strip Pooling	(X)SEAdaIN (O)Strip Pooling	-
IOU score	0.5808	0.5768	0.6024	-
Precision	0.7432	0.8454	0.8179	0.8004
Recall	0.7139	0.6273	0.6912	0.5368
F-score	0.7128	0.7179	0.7338	0.5940

表 4-13. 不同設定之 Dual Segmentation Model 對 GTD 進行預測結果

(訓練樣本數:測試樣本數=8:2)

Architecture	Framework		
	Dual Segmentation Model		LS-Net
Encoder	Resnet50		-
Train : Test	8:2		-
Module	(RGB)SEAdaIN (X)Strip Pooling	(RGB)SEAdaIN (O)Strip Pooling	-
IOU score	0.5438	0.5969	-
Precision	0.8324	0.8524	0.8004
Recall	0.5787	0.6591	0.5368
F-score	0.6802	0.7363	0.5940

表 4-14. 不同設定之 Dual Segmentation Model 對 GTD 進行預測結果

(訓練樣本數:測試樣本數=8:2)

Architecture	Framework			
	Dual Segmentation Model			LS-Net
Encoder	Resnet50			-
Train : Test	9:1			-
Module	原模型	(X)SEAdaIN (X)Strip Pooling	(X)SEAdaIN (O)Strip Pooling	-
IOU score	0.5893	0.5804	0.603	-
Precision	0.7657	0.8509	0.8251	0.8004
Recall	0.7251	0.6344	0.6991	0.5368
F-score	0.7157	0.7258	0.7342	0.5940

表 4-15. 不同設定之 Dual Segmentation Model 對 GTD 進行預測結果

(訓練樣本數:測試樣本數=9:1)

Architecture	Framework		
	Dual Segmentation Model		LS-Net
Encoder	Resnet50		-
Train : Test	9:1		-
Module	(RGB)SEAdaIN (X)Strip Pooling	(RGB)SEAdaIN (O)Strip Pooling	-
IOU score	0.5573	0.5438	-
Precision	0.8424	0.8324	0.8004
Recall	0.5995	0.5787	0.5368
F-score	0.701	0.6802	0.5940

表 4-16. 不同設定之 Dual Segmentation Model 對 GTD 進行預測結果

(訓練樣本數:測試樣本數=9:1)

#### 4.4 Denoise

上述實驗結果顯示 DSM 相較於單一 Segmentation Model，其 Precision 已經有大幅提升，但觀察其預測出來的 Predicted Mask，可以發現圖中有許多的噪點，接下來我們的想法是將 Dual Segmentation Model 預測出來的 Predicted Mask 當作前述單一 Segmentation Model 中的輸入重新進行訓練，觀察預測出來的 Predicted Mask 如預期的消除了原先的噪點，如圖 4-18.及 4-19.，整體數據則列於表 4-17.。

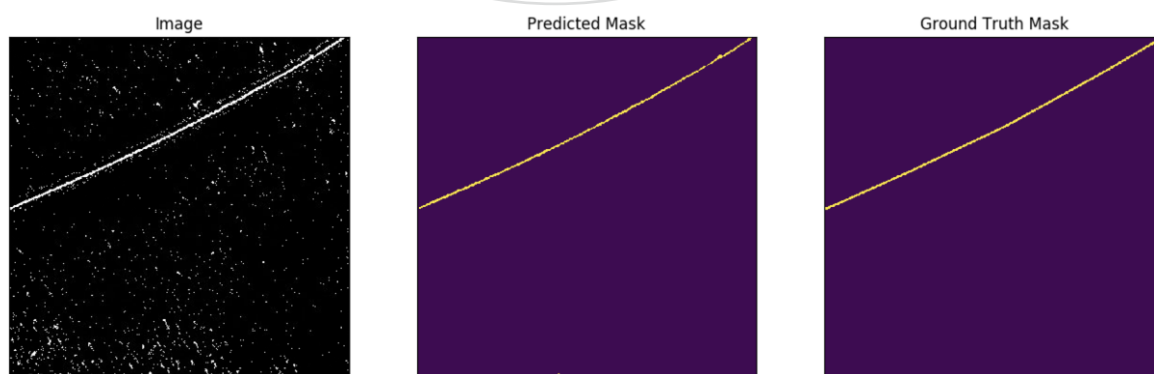


圖 4-18 DSM 預測 Ground Truth Powerline Dataset 1

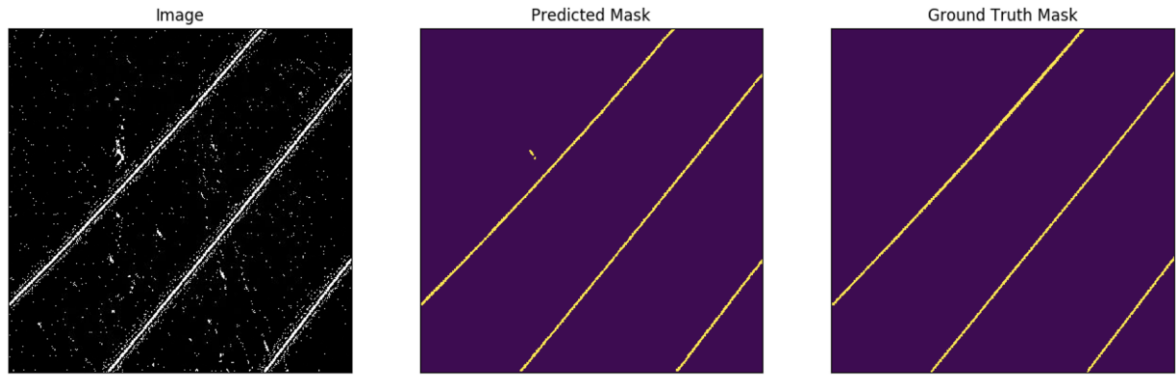


圖 4-19 DSM 預測 Ground Truth Powerline Dataset 2

Architecture	Framework		
	U-Net		LS-Net
Encoder	SE ResNeXt50 32x4d		-
Train : Test	8 : 2	9 : 1	-
IOU score	0.6534	0.6574	-
Precision	0.7861	0.8012	0.8004
Recall	0.7709	0.7719	0.5368
F-score	0.7714	0.7831	0.5940

表 4-17. 單一 Encoder Segmentation Model 對 GTD 進行預測結果

先前在 GTD 資料集之下的實驗使用 Channel attention 的 SE ResNeXt50 32x4d 搭配 U-Net 模型效果最佳，而在我們將 Dual Segmentation 模型的 Predicted Mask 當作其 Input 後，整體結果都獲得了提升，一樣維持著較高的 Recall，並改善了原先 Precision 至近乎 LS-Net 的高準確率表現，且整體 F-score 高出了 LS-Net 許多。



## 4.5 小結

在缺乏無人機電纜線數據集的情況下，本研究主要對目前開源的兩大電纜線資料集進行電纜線檢查的語意分割實驗，透過 Dual Segmentation Model 來強化單一 Segmentation Model 的訓練結果，並在 GTD 資料集訓練出來的模型，透過 Dual Segmentation Model 中兩通道編碼器提高模型的 Precision 進而提升了整體 F-Score，超越了 LS-Net 中利用 GTD 資料集預測他們自行生成資料的結果，而在 PLD 資料集上，透過我們重新標記資料後，使電纜線輪廓更加明顯進而使單一 Segmentation Model 上各方面實驗數據上有不錯的表現。



## 第五章 結論與未來展望

本研究基於電腦視覺提出基於語意分割之電纜線檢測，我們使用 Dual Segmentation Model，相較於傳統基於梯度的方法，本論文將 Dual Segmentation Model 應用於傳統的電纜線檢測，探索其效果，使電纜線檢測的任務能夠透過深度學習的方式，有效節省人力成本，高效且精準地完成電纜線檢測任務。

### 5.1 研究結論

本研究所提出的 Dual Segmentation Model 提供現今電纜線檢測一個很好的選擇，其優點及貢獻下列將分點論述：

1. 目前現有基於深度學模型檢測的方法中，LS-Net 表現較為優異，本研究以其為標準，最後實驗結果優於 LS-Net。
2. 在相似架構的語意分割模型 RTFNet 中，我們改善其未考慮 RGB 影像融合到另一個通道的 Encoder，導入 SEAdaIN Module 後，與另一個 Encoder 的 Feature Map 做加法，讓光學影像與邊緣強化影像在擷取特徵時所得到的 Feature Map 資訊可以交錯融合，達到輔助的效果。
3. PLD 資料集聚焦於 Detection 問題，本研究將其 Ground Truth 進行重新標記後應用於 Segmentation 任務，改善原先 Ground Truth 的實驗表現並獲得更佳的準確率。
4. GTD 資料集中的 Ground Truth 我們也將其重新標記，改善了原先 Ground Truth 不連續的標記，並應用於本研究中，獲得優秀的準確率。

## 5.2 未來展望

針對目前本研究所提出的方法與實驗之結果，仍有幾點可以改進及延伸：

### 1. 資料庫擴增：

現今電塔及電纜線的開源資料有限，Rabab Abdelfattah 等人[29]也有提及這些資料集中開源的僅有 GTD 及 PLD 如圖 5-1，本研究主要對目前這兩大開源的電纜線資料集進行電纜線檢查的語意分割實驗，雖然 GTD 資料集中光學影像和紅外線光影像的資料量各有 2000 張，但經過標記可作為我們與語意分割模型之實驗的資料僅有 200 張，獲取真實的資料集得來不意，往後若有其他實用的開源電纜線資料集，便能加入本研究中進行嘗試，當可增加模型整體的可靠度。

Table 1: Related Datasets.

Target	Dataset	Public	Image#(Pos.)	Image Size	Annotation Type	Syn.	Manual
TTs	[23]	No	3,200(1,600)	64×128	bounding box	No	No
TTs	[24]	No	28,674	6,048×4,032	bounding box	No	No
TTs	[26]	No	600	1,280×720	Binary Mask	No	No
PLs	[20]	Yes	4,000(2,000)	128×128	Binary Classif.	No	No
PLs	[21]	Yes	573	540×360	Binary Mask	No	Yes
PLs	[21]	Yes	287	540×360	Binary Mask	No	Yes
PLs	[27]	No	3,568	5,12×512	Binary Mask	No	Yes
PLs	[28]	No	718,000	-	Binary Mask	Yes	No
PLs	[29]	No	67,000	480×640	Binary Mask	Yes	No
Both	[22]	Yes	1,290	Various	Class Label	No	No
<b>Both</b>	<b>TTPLA</b>	<b>Yes</b>	<b>1,100</b>	<b>3,840×2,160</b>	<b>Instance Seg.</b>	<b>No</b>	<b>No</b>

圖 5-1.現有電塔及電纜線資料集一覽 [29]

### 2. 在雜亂的背景中檢測電纜線：

我們的模型已經可以解決在 GTD 真實資料集中複雜的背景，但如果遇到 PLD 中更加複雜的背景，在生成邊緣強化影像時，便增加其線段輪廓辨識的難度，導致整體辨識不如單純的 Segmentation Model，如何克服諸如市區以及山區場景的複雜背景，也是未來需要處理的課題。

### 3. 輕量化模型：

對於要取代傳統人力檢測電纜線，我們勢必要將 Dual Segmentation Model 能夠在無人機上發揮，但由於目前無人機上硬體上的限制，沒有較高的計算能力，必需輕量化我們的模型，未來才能夠獲得更佳廣泛的應用。



## 參考文獻

- [1] Yann LeCun, Corinna Cortes, Christopher J.C. Burges. THE MNIST DATABASE of handwritten digits. <http://yann.lecun.com/exdb/mnist/>, last visited on Dec 2018
- [2] ImageNet. <http://www.image-net.org/>
- [3] Yuanqing Lin, Fengjun Lv, Shenghuo Zhu, Ming Yang, Timothee Cour, Kai Yu, Liangliang Cao, Thomas Huang. Large-scale image classification: Fast feature extraction and SVM training. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1689-1696, 2011.
- [4] Deng, Jia, et al. "Imagenet large scale visual recognition competition 2012 (ILSVRC2012)." See [net. org/challenges/LSVRC](http://www.image-net.org/challenges/LSVRC) (2012).
- [5] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. Going Deeper with Convolutions. arXiv:1409.4842v1, 2014.
- [6] Park, E., et al. "ILSVRC-2017." URL [http://www. image-net. org/challenges/LSVRC/2017](http://www.image-net.org/challenges/LSVRC/2017) (2017).
- [7] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." arXiv preprint arXiv:1709.01507 7 (2017).
- [8] LeCun, Yann, et al. "Gradient-based learning applied to document. recognition." (1989)
- [9] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
- [10] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

- [11] Huang, Gao, et al. "Densely connected convolutional networks." CVPR. Vol. 1. No. 2. 2017.
- [12] Rupesh Kumar Srivastava, Klaus Greff, Jürgen Schmidhuber "Highway Network" arXiv:1505.00387 [cs.LG]
- [13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597v1 [cs.CV] 18 May 2015
- [14] Van Nhan Nguyen\*†, Robert Jenssen\* , and Davide Roverso† \*The UiT. Machine. Learning Group, UiT The Arctic University of Norway, 9019 Tromsø, Norway †Analytics Department, eSmart Systems, 1783 Halden, Norway "LS-Net: Fast Single-Shot Line-Segment Detector" arXiv:1912.09532v2 [cs.CV] 24 Jan 2020
- [15] P.V.C. Hough, Machine Analysis of Bubble Chamber Pictures, Proc. Int. Conf. High Energy Accelerators and Instrumentation, 1959.
- [16] Duda, R. O. and P. E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures," Comm. ACM, Vol. 15, pp. 11–15 (January, 1972).
- [17] Chen, Y.; Li, Y.; Zhang, H. Automatic power line extraction from High-resolution remote sensing imagery based on an improved radon transform. Pattern Recognit. 2016, 49, 174–186
- [18] Yunping Chen , Yang Li , Huixiong Zhang , Ling Tong , Yongxing Cao , Zhihang Xue "Automatic power line extraction from high resolution remote sensing imagery based on an improved Radon transform" <https://doi.org/10.1016/j.patcog.2015.07.004>
- [19] Arbeláez, P.; Maire, M.; Fowlkes, C.; Malik, J. From contours to regions: An empirical evaluation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009; pp. 2294–2301.

- [20] Arbeláez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2011, 33, 898–916.
- [21] Shen, W.; Wang, X.; Wang, Y.; Xiang, B. DeepContour: A deep convolutional feature learned by positive sharing loss for contour detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 7–12 June 2015; pp. 3982–3991.
- [22] Bertasius, G.; Shi, J.; Torresani, L. DeepEdge: A multi-scale bifurcated deep network for top-down contour detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 7–12 June 2015; pp. 4380–4389.
- [23] Yang, J.; Price, B.; Cohen, S. Object Contour Detection with a Fully Convolutional Encoder-Decoder Network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 193–202.
- [24] Maninis, K.K.; Pont-Tuset, J.; Arbeláez, P. Convolutional oriented boundaries: From image segmentation to high-level tasks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2018, 40, 819–833.
- [25] Madaan, R.; Maturana, D.; Scherer, S. Wire detection using synthetic data and dilated convolutional networks for unmanned aerial vehicles. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, BC, Canada, 24–28 September 2017; pp. 3487–3494.
- [26] Xingchen Zhang, Gang Xiao, Shanghai Jiao Tong University, Shanghai, China Ke Gong, East China Normal University, Shanghai, China Ping Ye, Junhao Zhao, Shanghai Jiao Tong University, Shanghai, China “Power Line Detection

for Aircraft Safety Based on Image Processing Techniques: Advances and Recommendations”

- [27] Long, J., E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
- [28] J. Canny, "A Computational Approach to Edge Detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679-698, Nov. 1986, doi: 10.1109/TPAMI.1986.4767851.
- [29] Rabab Abdelfattah, Xiaofeng Wang, and Song Wang Department of Electrical Engineering, University of South Carolina, USA Department of Computer Science and Engineering, University of South Carolina, USA , " TTPLA: An Aerial-Image Dataset for Detection and Segmentation of Transmission Towers and Power Lines "
- [30] Yetgin, Ömer Emre; Gerek, Ömer Nezir (2019), "Ground Truth of Powerline. Dataset (Infrared-IR and Visible Light-VL)", Mendeley Data, [v9http://dx.doi.org/10.17632/twzp8xcccw.9](http://dx.doi.org/10.17632/twzp8xcccw.9)
- [31] PLD-UVA Dataset :<https://github.com/SnorkerHeng/PLD-UAV>
- [32] Sun, Y., Zuo, W., & Liu, M. (2019). *RTFNet: RGB-Thermal Fusion Network for Semantic Segmentation of Urban Scenes. IEEE Robotics and Automation Letters*, 1–1. doi:10.1109/lra.2019.2904733
- [33] Hu, J., L. Shen, and G. Sun. Squeeze-and-excitation networks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018



- [34] Huang, X. and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. in Proceedings of the IEEE International Conference on Computer Vision. 2017.
- [35] Hou, Q., et al. Strip Pooling: Rethinking Spatial Pooling for Scene Parsing. 2020. arXiv:2003.13328.

