

Digital Image Processing - Final Project : Water Segmentation (Group 20)

楊哲旻¹(413511003), 黃云潔²(312581023)

國立陽明交通大學 電機工程學系博士班¹

國立陽明交通大學 智能系統研究所碩士班²

1. 前言 (Introduction)

近年來，隨著遙感技術(Remote Sensing)與影像處理技術的迅速發展，利用高解析度影像進行水體監測已成為可能。無論是來自衛星、無人機，還是其他傳感器的影像，這些數據都提供了豐富的空間信息，幫助研究者更準確地觀測地表水的分布。然而，在處理這些影像數據時，如何自動且準確地分割水體範圍，仍然是一項充滿挑戰的技術難題。傳統的影像處理方法(如基於閾值的分割)然在簡單場景中表現良好，但在面對更複雜的環境條件(如光照變化、河岸植被干擾或濁水環境)時，其性能往往受到限制。為了解決上述這些問題，B. Bovcon 和 M. Kristan 在過去的文獻中提出了 Water Segmentation and Refinement (WaSR) Network 架構，該方法採用了基於空洞卷積的 ResNet101 作為編碼器，並結合包含慣性測量單元(Inertial Measurement Unit, IMU)資訊的解碼器，大幅提升了分割的穩健性 [1]。此外，為了滿足無人水面車輛在即時水體檢測中的自動化應用需求並保障其安全性，研究者還提出了一種輕量級的全卷積神經網路架構 [2]。因此，如何有效應用深度學習等先進技術，進一步提升水體分割的準確性和適應性，已成為一個重要且亟待解決的研究課題。

2. 資料與方法 (Materials and Methods)

本研究架構如圖 1 所示，原始影像與遮罩經過資料前處理後，分為 k 折分別進行影像分割模型訓練，以交叉驗證的並交比指標比較每個模型。因此，本章節分為三部分為 2.1 資料前處理、2.2 深度學習方法與 2.3 傳統影像處理方法，其

中 2.2 深度學習方法包含三個影像分割模型、損失函數、環境與超參數以及性能指標。最後在 2.3 傳統影像處理方法中，我們也使用基於多範圍 HSV 的影像處理方法進行水的分割。

2.1 資料前處理 (Data Pre-processing)

本研究資料集含有 80 張進行訓練與驗證集，20 張作為測試集，由於影像張數過少，所以我們使用水平翻轉、隨機方正裁剪方式增量，如表 1 所示。由於考慮水中倒影可能鏡像緣故，而沒有採用垂直翻轉與旋轉。最終資料集增量於原圖數量的 5 倍，總共 400 張的訓練與驗證集。在遮罩中的像素白色為水，其餘黑色為非水，增量方法也同樣用於遮罩中。

在裁減方法上，裁剪後的寬 W_c 與高 H_c 是一定與原始影像中的固定範圍比例大小 $[min_{ratio}, max_{ratio}]$ ，即 r 為原圖與裁減後大小的寬 W 與高 H 等比例縮放之尺度因子，以確保不會裁剪出影像大小過小(像素過少)導致沒有任何資訊的影像區域，以及寬高不相等比例導致變形，如公式 1。本研究 min_{ratio} 設為 0.5， max_{ratio} 設為 0.8。

$$W_c = r \times W, H_c = r \times H, r \in [min_{ratio}, max_{ratio}] \quad (1)$$

在資料集中計算水像素與總像素數量的比例後發現，最高比例為 99.08%(57.png)，最低為 5.78%(33.png)。儘管裁減後的影像包含了大面積的像素，且裁減範圍內的像素完全由水或非水組成(如表 1 中的第二個例子)，我們認為這些影像仍能提供模型學習的參考資訊，因此並未對其進行範圍限制或移除處理。

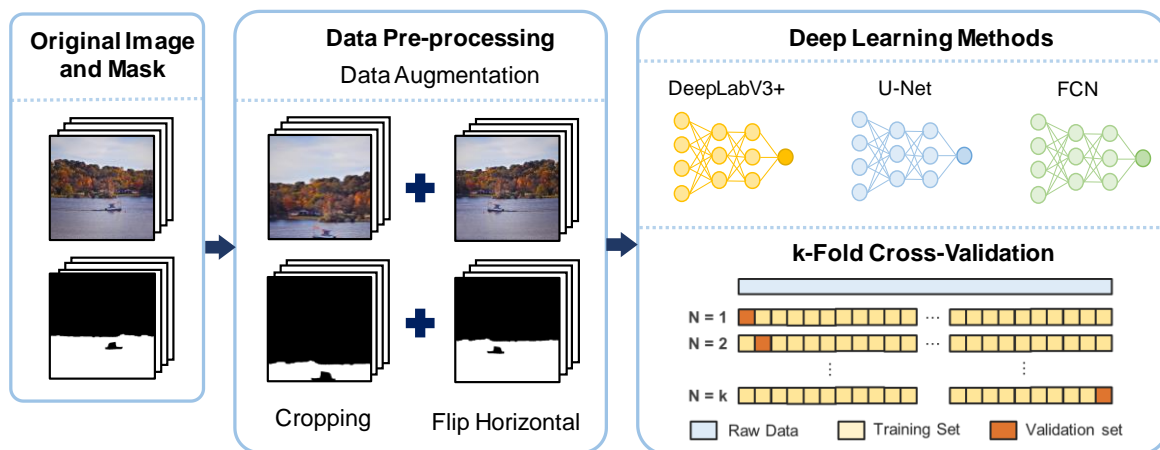







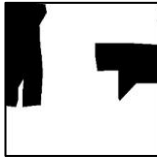


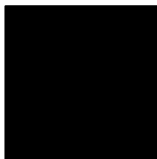



圖 1: 本研究水的影像分割之架構圖

表 1: 影像增量方法

	Original Image	Cropped Image	Cropped and Flip Horizontal Image
Image			
			
Mask			
			

2.2 深度學習方法 (Deep Learning Methods)

2.2.1 DeepLabV3+

DeepLabV3+ [3] 的解碼器 (Decoder) 利用空洞卷積 (Atrous Convolution) 以擴大感受野 (Receptive Field)，從而更有效地捕捉圖像中的全局信息。結合多尺度特徵金字塔池化 (Atrous Spatial Pyramid Pooling) 的設計，透過合併多個平行

的不同尺度的空洞卷積，提升模型對不同尺度上下文信息的感知。而在編碼器 (Encoder) 可以選擇常見的分類網路，可移除除於後全連階層來當骨架，是作為提取影像重要特徵之用。本研究使用 ResNet50 [4] 為編碼器的骨架，DeepLab V3+ with ResNet50 架構如圖 2 所示。

2.2.2 U-Net

U-Net [5] 的網路結構呈現 U 形，其中編碼器是將影像中重要特徵用降維的方式逐步提取出來，而解碼器逐步將編碼器的特徵圖還原至相同大小。解碼器中的每一層大小依序與相對應的編碼器層相同，並透過跳躍連結 (Skip Connection)，將每一層的輸出與對應的編碼器層的輸出串接在一起，以避免失去重要的重建信息。本研究同樣使用 ResNet50 為編碼器的骨架，U-Net with ResNet50 架構如圖 3 所示。

2.2.3 FCN

全卷積網路 (Fully Convolutional Network, FCN) [6] 有分不同版本，它們區別在解碼器的設計上。第一種為 FCN-32s 將編碼器輸出的特徵圖大小以 32 倍還原為原始圖像的大小。其次為 FCN-16s 則在 FCN-32s 基礎上，將其的特徵圖放大兩倍，合併資訊，使特徵圖大小以 16 倍還原為原始圖像的大小。最終 FCN-8s 也依此類推，獲得比 FCN-32s 和 FCN-16s 更細微的分割結果，但計算過程相較於長。本研究因不考慮即時性檢測，所以採用為 FCN-8s 架構，以獲得較細節的水分割性能。而在編碼器 (Encoder) 同樣是作為提取影像重要特徵之用。本研究同樣使用 ResNet50 為編碼器的骨架，FCN with ResNet50 架構如圖 4 所示。

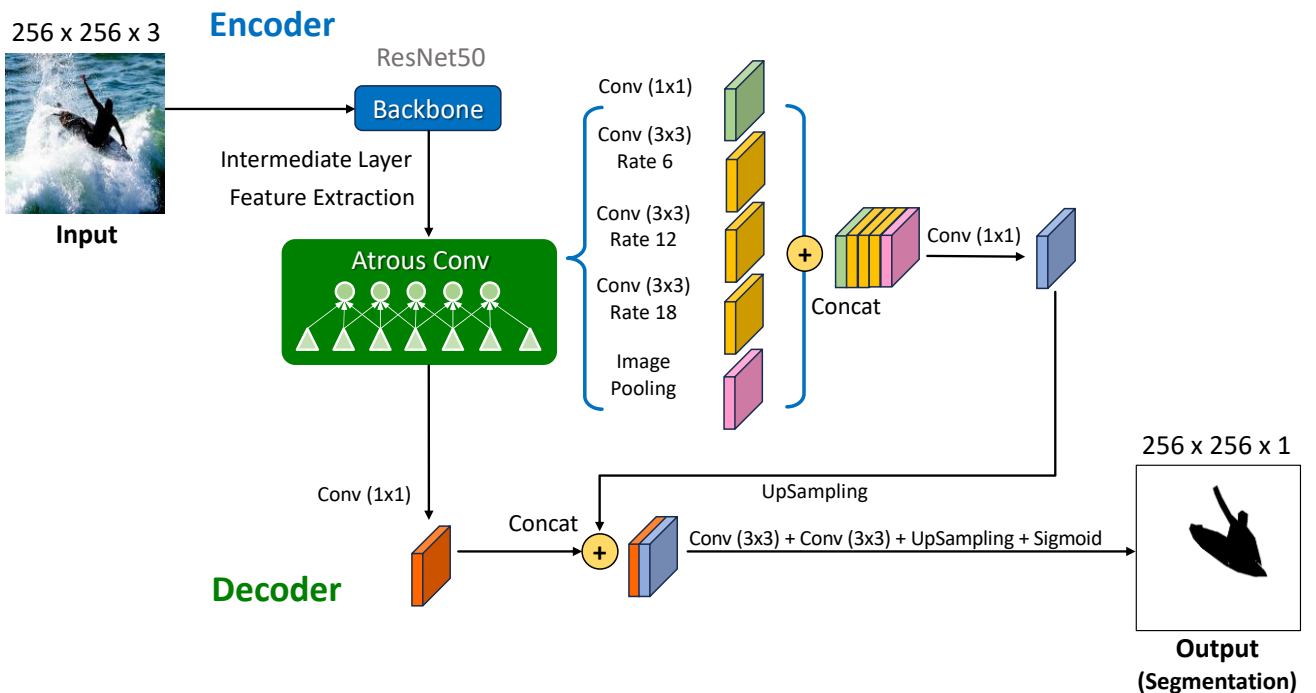


圖 2: 本研究 DeepLabV3+ 的影像分割架構

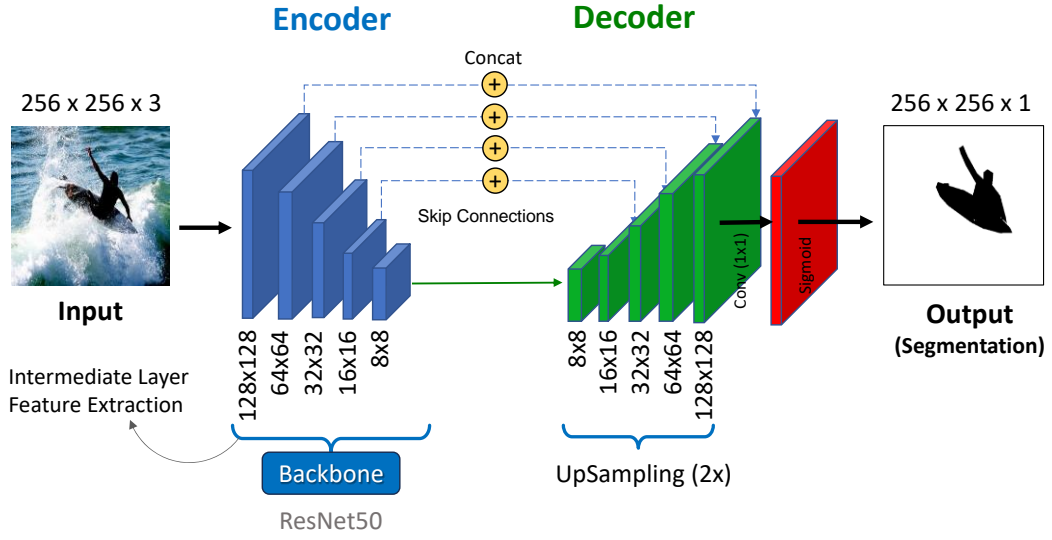


圖 3: 本研究 U-Net 的影像分割架構

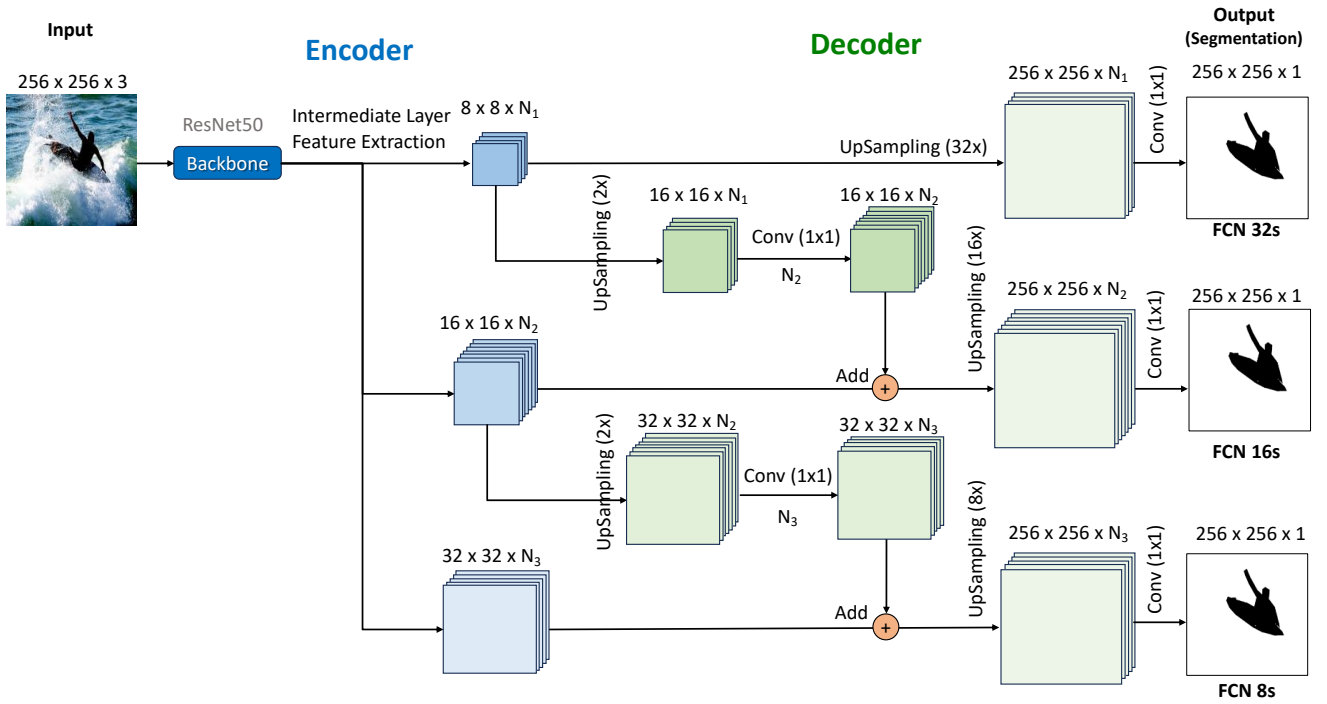


圖 4: 本研究 FCN 的影像分割架構

2.2.4 損失函數 (Loss Function)

在影像分割的損失函數的選擇，我們使用 BCE Loss 和 Dice Loss，如公式 2-3 所示，其中 N 表示樣本數量， y 表示真實標籤， \hat{y} 表示模型的預測概率， ε 為平滑係數，本研究 ε 設為 1。這兩個損失函數分別用於評估像素級別的預測結果，和預測結果與實際分割影像之間的相似性。訓練過程的目的是最小化損失 BCE Loss 和 Dice Loss 之總和，以使模型能夠產生準確的分割結果。

$$BCE\ Loss = -\frac{1}{N} \sum_{i=1}^N [y_i \times \log(\hat{y}_i) + (1 - y_i) \times \log(1 - \hat{y}_i)] \quad (2)$$

$$Dice\ Loss = 1 - \frac{2 \times \sum_{i=1}^N (\hat{y}_i \times y_i) + \varepsilon}{\sum_{i=1}^N \hat{y}_i + \sum_{i=1}^N y_i + \varepsilon} \quad (3)$$

2.2.5 環境與超參數

本研究使用 Google Colab [7] 提供的 T4 GPU 進行模型訓練，每個模型均採用 Adam 最佳化演算法 [8]，學習速率設為 0.001，Batch Size 設定為 32，訓練 Epochs 為 300。我們引入了「ModelCheckpoint 回調函數」，在訓練過程中保存驗證集交集比 (Intersection over Union, IoU) 表現最佳的模型，確保所保存的模型在所有 Epochs 中對應的 Loss 最低。

2.2.6 性能指標 (Performance Metrics)

本研究使用五折交叉驗證 (5-fold Cross-Validation, 5-Fold CV) 評估每個模型的性能，即訓練與驗證比例為 4:1。增量影像僅包含於所有 5 個子集的訓練與驗證集，並不包含測試集中。

影像分割的性能指標上，影像中的實際水的像素預測正確個數為真陽性(True Positive, TP)，預測錯誤個數為假陰性(False Negative, FN)，而影像中的非水像素預測正確個數為真陰性(True Negatives, TN)，預測錯誤為個數為假陽性(False Positives, FP)。而 IoU 定義如下為實際與預測兩個像素的交集與聯集之比值，定義如公式 4，其中 GT 為實際標籤(Ground Truth)的遮罩，而 PRED 為預測的結果。

$$IoU = \frac{GT \cap PRED}{GT \cup PRED} = \frac{TP}{TP + FP + FN} \quad (4)$$

2.3 傳統影像處理方法 (Conventional Method)

本研究提出了一種基於 HSV 顏色空間的多範圍遮罩法。該方法針對偏白、偏綠及偏藍的水域顏色進行特徵定義，並結合影像前處理與形態學操作以提升分割準確性。




由於影像可能受亮度和對比度的影響，為提升色彩辨識的準確性，因此我們先使用線性變換(Linear Transformation, LT)提升影像亮度和對比度。再將影像從 BGR 色空間轉換至 HSV，並對飽和度通道進行正規化來提升顏色分離效果，然後再使用雙邊濾波(Bilateral Filter, BF) [9]平滑影像以同時保留邊緣細節。我們為了分割不同水域顏色，定義三種 HSV 範圍包含偏白水域 (Whitish Water Region, WWR)、偏綠水域 (Greenish Water Region, GWR)與偏藍水域 (Bluish Water Region, BWR)，如公式 5-7，色彩區域範圍如表 2。

$$WWR = 0 \leq H \leq 180 \text{ and } 0 \leq S \leq 30 \text{ and } 200 \leq V \leq 255 \quad (5)$$

$$GWR = 40 \leq H \leq 80 \text{ and } 50 \leq S \leq 255 \text{ and } 50 \leq V \leq 255 \quad (6)$$

$$BWR = 90 \leq H \leq 140 \text{ and } 50 \leq S \leq 255 \text{ and } 50 \leq V \leq 255 \quad (7)$$

表 2: 三種水域的色彩區域範圍

Water Region	Color Range
WWR	
GWR	
BWR	

經雙邊濾波器處理的影像再劃分為固定大小的區塊 (128×128 像素)，逐區塊計算遮罩，並基於每個範圍的 HSV

閾值進行分割。分塊後的遮罩合併為全影像遮罩，但由於遮罩可能包含雜訊，為提升結果一致性，我們使用形態學操作進行清理，其中閉運算(Closing)來填補空洞，開運算(Opening)來移除小型雜訊點，最終獲得水的分割結果，如圖 5。

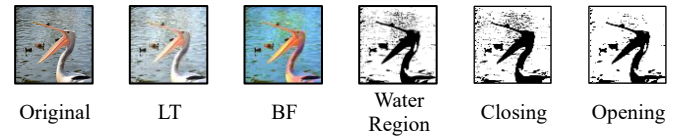
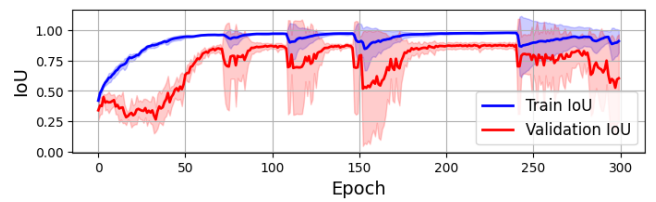


圖 5: 本研究傳統影像處理方法的每個流程之結果圖

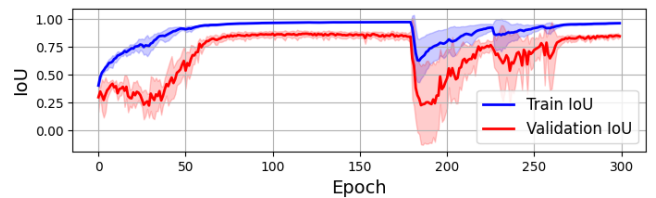
3. 結果

3.1 驗證集

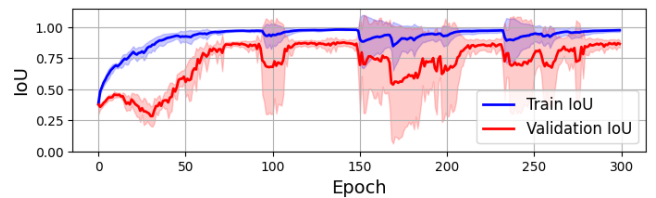
圖 6 顯示本研究中三個影像分割模型架構在 5-Fold CV 訓練過程中的 IoU 學習曲線(Learning Curve)，並展示其平均值與標準差。藍色曲線代表訓練集，紅色曲線代表驗證集，確保這三個架構在每次訓練中都能達到完整的收斂。表 3 比較了三個模型在驗證集上的表現，結果顯示 DeepLabV3+ with ResNet50 達到最高 IoU 值 89.06%，但傳統方法在整個訓練與驗證集的 IoU 值為 43.92%。



(A) DeepLabV3+ with ResNet50



(B) U-Net with ResNet50



(C) FCN with ResNet50

圖 6: 本研究的影像分割模型訓練過程中的 IoU 學習曲線

表 3: 本研究使用的影像分割模型之驗證集損失值與 IoU

Methods	DeepLabV3+ with ResNet50		U-Net with ResNet50		FCN-8s with ResNet50	
	Loss	IoU(%)	Loss	IoU(%)	Loss	IoU(%)
5-Fold CV						
Fold 1	0.40	86.84	0.53	84.23	0.54	86.73
Fold 2	0.34	90.00	0.30	88.70	0.35	90.12
Fold 3	0.40	89.71	0.38	88.63	0.33	89.71
Fold 4	0.26	91.31	0.27	89.78	0.32	91.07
Fold 5	0.40	87.42	0.39	86.64	0.39	86.01
Average	0.36	89.06	0.37	87.60	0.39	88.73

表 4: 驗證集的影像分割結果


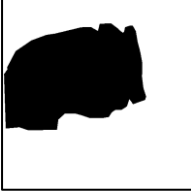





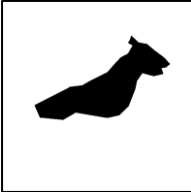













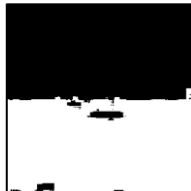





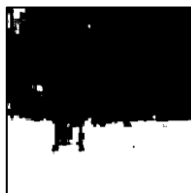



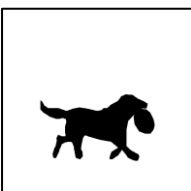
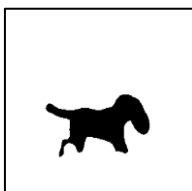
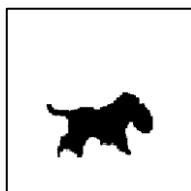
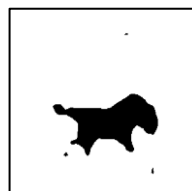

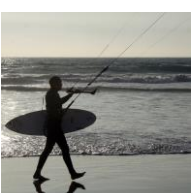









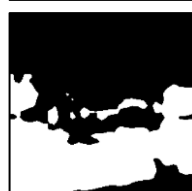







No	Image	Mask	DeepLabV3+ with ResNet50	U-Net with ResNet50	FCN-8s with ResNet50	Conventional Method
3						
14						
22						
24						
54						
59						
62						
63						
75						

表 4 展示了各最佳模型在驗證集上的影像分割結果。可以觀察到，DeepLabV3+ with ResNet50 的邊緣細節更為精細且連貫；U-Net with ResNet50 的邊緣則呈現較明顯的鋸齒狀；而 FCN-8s with ResNet50 則會在部分水區出現小片狀的錯誤預測為水。在傳統影像處理方法中，由於天空顏色與水色過於相近，容易將天空誤判為水；此外，水的渾濁以及物體在水中的倒影也常被錯判為非水，因此整體效果表現不佳。

3.2 測試集

表 5 為 5-Fold 所訓練的模型來驗證在測試集的性能，結果顯示 DeepLabV3+ with ResNet50 達到最高 IoU 值 50.37%，而在傳統方法在測試集的 IoU 值為 36.01%。

表 6 展示了各最佳模型在測試集中的影像分割結果。由於測試集中包含許多影像特徵與訓練集不同，例如垂直俯瞰的水面照片，導致大多數模型難以精準完成分割。然而，傳統影像處理方法對於俯瞰水面的照片影響較小，但其效果容易受到環境的複雜變動影響，表現存在較大的局限性。

表 5: 本研究使用的影像分割模型之測試集 IoU (%)

Methods	DeepLabV3+ with ResNet50	U-Net with ResNet50	FCN-8s with ResNet50
Fold 1	54.42	52.33	51.34
Fold 2	48.21	48.00	47.14
Fold 3	50.18	46.62	49.80
Fold 4	49.80	50.27	49.59
Fold 5	49.25	51.01	50.43
Average	50.37	49.65	49.66

4. 結論與未來展望

本研究中我們使用三個先進的影像分割模型：DeepLabV3+、U-Net 和 FCN，其中以 DeepLabV3+ 的性能最佳。另外，在骨幹網路的選擇上，我們嘗試 VGG、MobileNetV2 和 ResNet50。結果顯示，VGG 模型的參數量較大，但分割性能不如 ResNet50；而 MobileNetV2 雖然輕量化，但由於資料集規模小且多樣性不足，導致模型無法完全收斂，因此最終選擇 ResNet50 作為骨幹網路。此外，我們也嘗試使用 ImageNet 比賽權重進行遷移式學習(這為額外測試，未納入正式結果與程式碼中)，結果每個模型均有顯著性能提升，這進一步說明資料的豐富性對深度學習模型的影響至關重要，有大量的資料集所訓練下的權重是有助於進行權重微調進行其他檢測任務。在傳統影像處理方面，我們採用了基於 HSV 色彩空間的遮罩方法來分割水域，但分割效果易受倒影、水體渾濁與天空影響，表現不佳。可能需要更複雜的條件限制或更靈活的調整來改善水域分割的效果。

在我們的測試集與驗證集中可以觀察到，大部分模型在細節處理上仍有不足。由於期末專題時間有限，我們未能完成原計畫中的兩階段方法：第一階段進行深度學習模型的粗略分割，第二階段對初步結果細化，例如使用條件隨機場(Conditional Random Fields, CRF) [10]進行後處理，以補充修正分割結果，改善邊界或小結構的準確性。

未來，隨著資料量的累積與技術的進步，影像分割的精確度和細節處理能力有望進一步提升。我們相信，這些改進將為水域分割的應用提供更加穩健且高效的解決方案，並為相關領域的研究帶來新的價值。

參考文獻

[1] B. Bovcon and M. Kristan, “WaSR—A Water Segmentation and Refinement Maritime Obstacle Detection Network,” *IEEE Transactions on Cybernetics*, vol. 52, no. 12, December 2022.

[2] J. Taipalmaa, N. Passalis, H. Zhang, M. Gabbouj, and J. Raitoharju, “High-Resolution Water Segmentation for Autonomous Unmanned Surface Vehicles: A Novel Dataset and Evaluation,” *IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, Pittsburgh, PA, USA, October 2019.

[3] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation,” *European Conference on Computer Vision (ECCV)*, pp. 833-851, October 2018.

[4] S. Targ, D. Almeida, and K. Lyman, “Resnet in Resnet: Generalizing Residual Architectures,” *International Conference on Learning Representations (ICLR)*, San Juan, Puerto Rico, May 2016.

[5] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234-241, November 2015.

[6] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, June 2015.






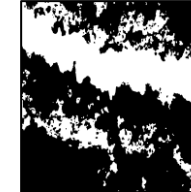








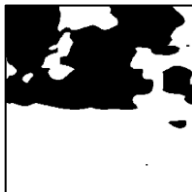


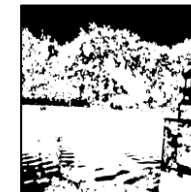






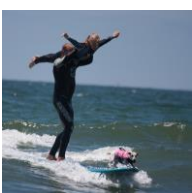











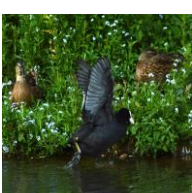


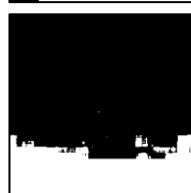


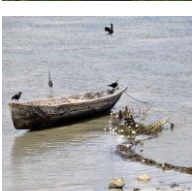
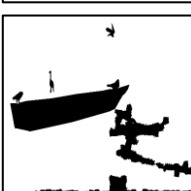
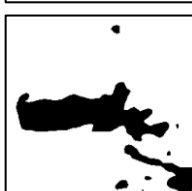
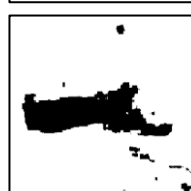
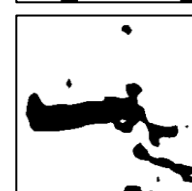
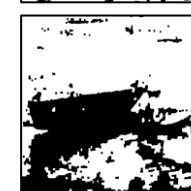
[7] E. Bisong, “Google Colab,” Building Machine Learning and Deep Learning Models on Google Cloud Platform, pp. 59–64, 2019.

[8] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *3rd International Conference on Learning Representations (ICLR)*, San Diego, USA, 2014.

[9] C. Tomasi and R. Manduchi, “Bilateral Filtering for Gray and Color Images,” *6th International Conference on Computer Vision, Bombay, India*, January 1998.

[10] N. Plath, M. Toussaint, S. Nakajima, “Multi-Class Image Segmentation Using Conditional Random Fields and Global Classification,” *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 817–824, June 2009.

表 6: 測試集の影像分割結果

No	Image	Mask	DeepLabV3+ with ResNet50	U-Net with ResNet50	FCN-8s with ResNet50	Conventional Method
1						
2						
3						
4						
8						
9						
14						
16						
18	