

利用小物件增強注意力機制偵測海洋廢棄物

彭浩言01157063，陳奕翰01157149

指導老師：蔡宇軒

本研究已被2025 Seventh International Symposium on Computer, Consumer and Control (IS3C) 收錄並發表
(DOI: 10.1109/IS3C65361.2025.11131032)

研究動機

- 根據聯合國永續發展目標的第六項著重在淨水與衛生，其中海洋漂浮垃圾的自動檢測是一項迫切的研究課題，不僅有助於減輕生態破壞，亦能提升航行安全。傳統人工檢測效率低且成本高，因而亟需基於高空影像與電腦視覺技術的自動化解決方案。
- 2021 年發表的 FloW-Img (Cheng 等人，2021) 資料集提供了高解析度的影像，捕捉了多樣海洋環境下的漂浮垃圾，促進了專為此任務設計的物件偵測模型的發展。

研究方法

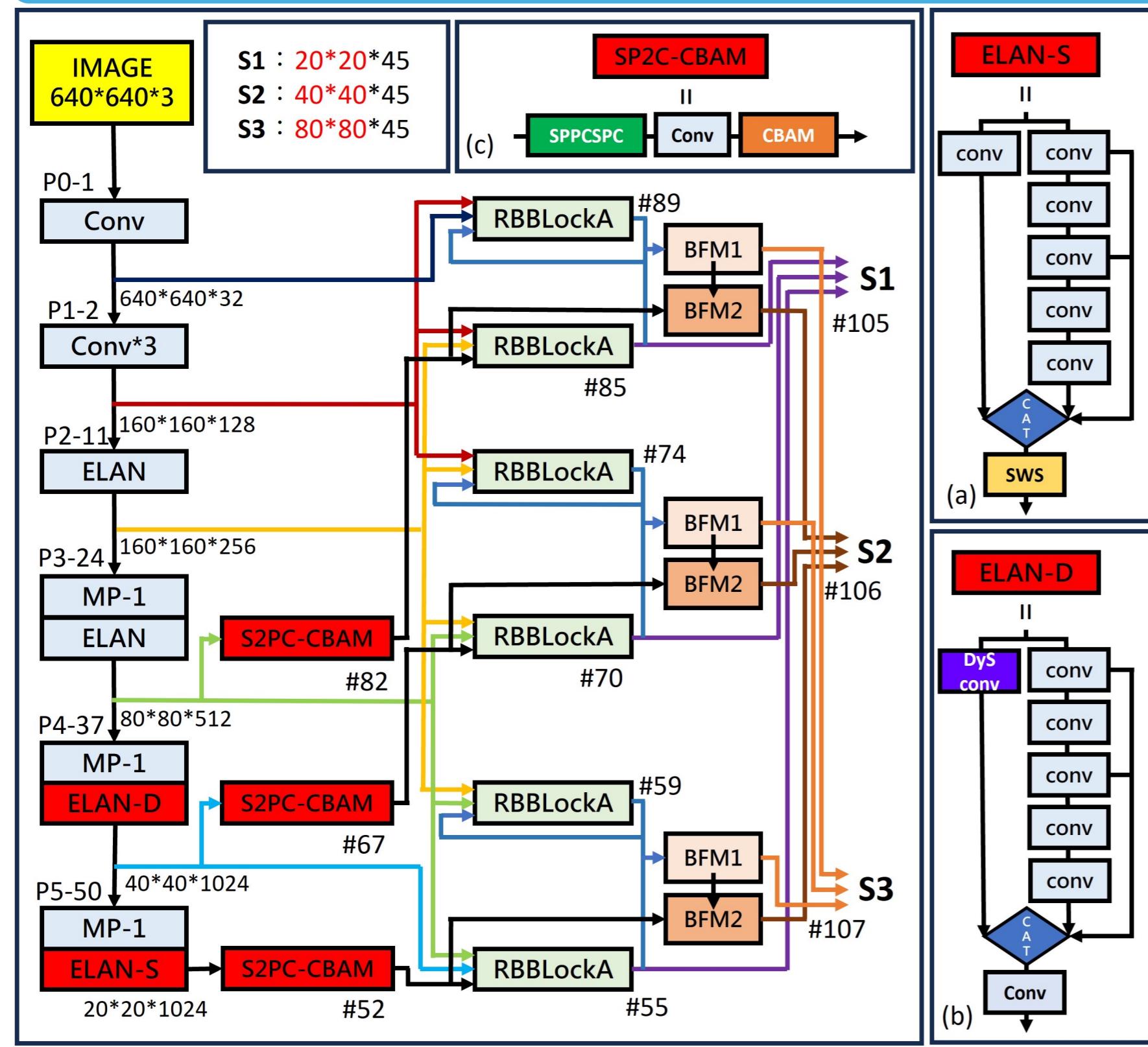


圖1 本研究提出的神經網路結構

各模組化增強策略的概述如下：

a) 本研究提出的改良版SimAM注意力機制

Algorithm 1 Our proposed attention mechanism

```

Require: Feature map  $x \in R^{N \times C \times H \times W}$ , partitions  $i, j \in \{0, 1\}$ , coefficient  $\lambda$ 
1: // Spatial Partitioning
2: for  $i = 0$  to  $1$  do
3:   for  $j = 0$  to  $1$  do
4:      $Blocks_{(i,j)} \leftarrow x[:, :, \frac{iH}{2} : \frac{(i+1)H}{2}, \frac{jW}{2} : \frac{(j+1)W}{2}]$ 
5:   end for
6: end for
7:
8: // Apply SimAM on Each Block
9: for  $i = 0$  to  $1$  do
10:   for  $j = 0$  to  $1$  do
11:      $\bar{B}_{(i,j)} \leftarrow \text{mean}(Blocks_{(i,j)})$ 
12:      $\mathcal{D}_{(i,j)} \leftarrow (Blocks_{(i,j)} - \bar{B}_{(i,j)})^2$ 
13:      $\mathcal{V}_{(i,j)} \leftarrow \frac{1}{N} \sum_{n=1}^N \mathcal{D}_{(i,j)}$ 
14:      $\mathcal{E}_{(i,j)} \leftarrow \frac{\mathcal{D}_{(i,j)}}{4(\mathcal{V}_{(i,j)} + \lambda)} + 0.5$ 
15:      $\mathcal{Y}_{(i,j)} \leftarrow Blocks_{(i,j)} \times \sigma(\mathcal{E}_{(i,j)})$ 
16:   end for
17: end for
18:
19: // Reconstruct Spatially Weighted Feature Map
20:  $\mathcal{F}^* \leftarrow \begin{bmatrix} \mathcal{Y}_{(0,0)} & \mathcal{Y}_{(0,1)} \\ \mathcal{Y}_{(1,0)} & \mathcal{Y}_{(1,1)} \end{bmatrix}$ 
21: return  $\mathcal{F}^*$ 

```

b) 應用Dynamic Convolution

c) 導入CBAM (Convolutional Block Attention Module)

訓練結果



圖2 Ground Truth

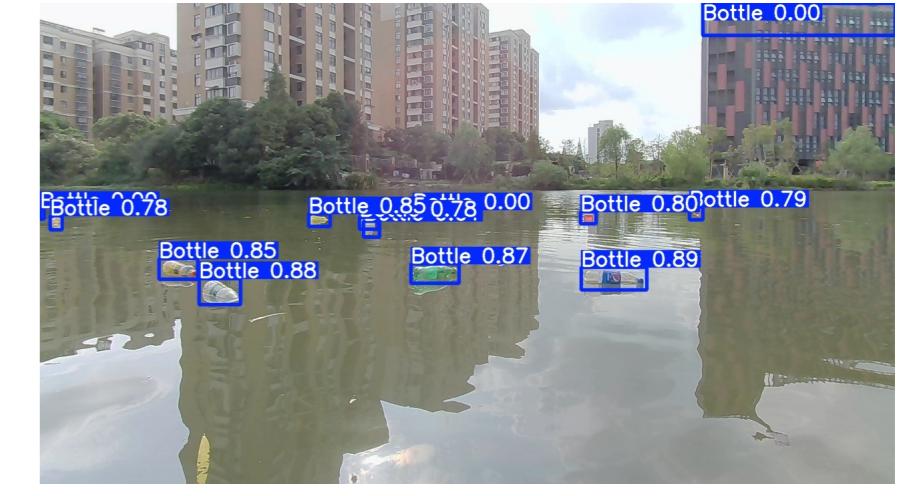


圖3 Yolov11



圖4 Yolov12



圖5 PRB-NET



圖6 Ours

Model	Precision	Recall	mAP50	mAP50-95	F1
Yolov11	0.838	0.809	0.858	0.429	0.82
Yolov12	0.858	0.799	0.855	0.422	0.83
PRB-NET	0.839	0.844	0.869	0.418	0.84
Ours	0.86	0.837	0.867	0.403	0.85

表1 FloW-Img 測試集上的偵測性能比較

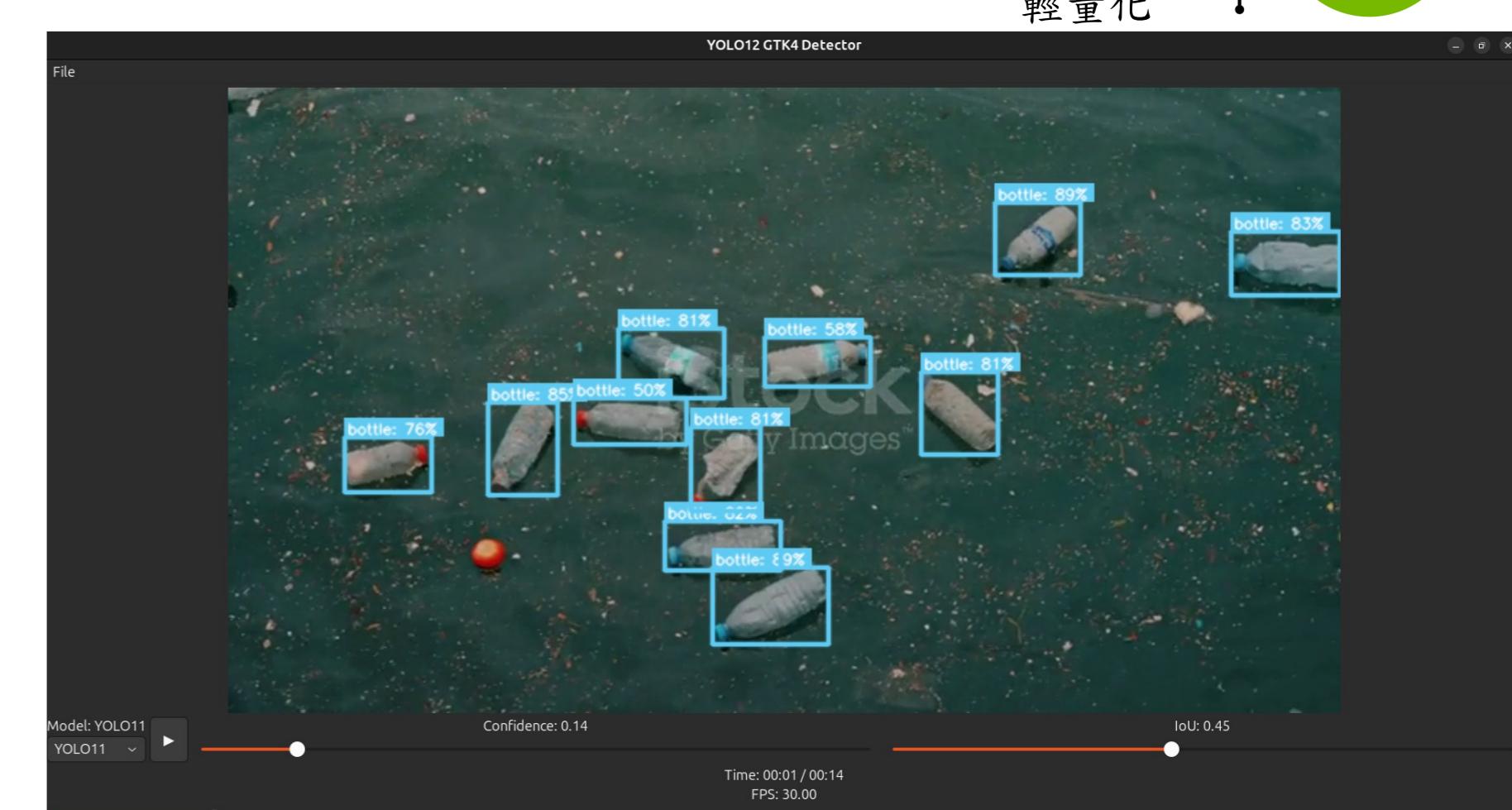


圖7 在 C++ 環境中使用匯出的 ONNX 小物件模型逐幀偵測影片中的海廢

Model (FP16)	Throughput	Latency (mean)	GPU compute time (mean)
Yolov11	697.96 qps	2.20117 ms	1.42969 ms
Yolov12	219.732 qps	5.29559 ms	4.52393 ms
PRB-NET	89.7195 qps	11.9482 ms	11.1121 ms
Ours	55.1739 qps	18.9216 ms	18.0791 ms

表2 在TensorRT上的跑分 (RTX 4070)