

# 李濬屹 + 王建堯老師 - Ensemble Fusion for Small Object Detection

**出自**

**題目**

Ensemble Fusion for Small Object Detection

**Introduction**

**背景**

small object detection, 應用於監視器、自駕、航空資料分析

受限於低對比、少 feature、本身影像模糊以及bounding box 容易偏調

**動機**

不能使用 feature extraction 因為細節會被模糊

## 目的

用 Ensemble 的方法來利用每個 Model 的長處

從 Cascade RCNN [6] and CenterNet [7] 抽 model 出來，做成許多變體

在 inference phase 會使用 Slicing Aided Hyper Inference (SAHI) 與 test time augmentation (TTA)

再得到更多 variants

Slicing Aided Hyper Inference : 將大圖片切成多個有重疊的小塊區域，再放入 model 中找 bounding box，最後貼回大圖片再用 NMS 整合 bounding box

test time augmentation : inference 時會將 input 做 augmentation 例如，水平翻轉、旋轉、調整亮度等）一起送入 model 中預測，最後將所有預測結果**整合**起來（例如，取平均或投票），得到最終的預測結果

ensembling 演算法使用 Weighted Box Fusion method (WBF)

## 相關研究

方法一、使用現有的 object detection models  $\Rightarrow$  準確度仍然不夠

## 1. one-stage object detection models : 較快

ex : CenterNet  $\Rightarrow$  output 一張熱力圖，每個像素代表該位置是物體中心點的可能性。  
接著找到局部最大值並額外預測出物體的大小，就可以生成最終邊界框。特點是不用 NMS

## 2. two-stage : 較準

ex : Cascade R-CNN , 是 Faster R-CNN 的進階。Cascade R-CNN 使用多個、遞增的 IoU 閾值來解決單一 iou threshold 問題。

單一 iou threshold 問題：訓練時，我們通常會使用較低的 IoU 閾值（例如 0.5）來定義正樣本，如果閾值過高樣本會太少，會 overfit ;但在推論時，我們希望 IoU 值高，因此會造成 mismatch

所以 Cascade R-CNN 的

第一個 R-CNN 頭部使用一個較低的 IoU 閾值（例如 0.5）

第二個 R-CNN 頭部使用一個較高的 IoU 閾值（例如 0.6）

以此類推

方法二、專門做一個 SOD 的 Function

(a) 使用 low-level features 或用 super-resolution

(b) 修改 down sampling 的方法 or 採用 multi-scale feature fusion (ex: dilated convolutions or adjusting model necks 使小特徵依然能夠保留)

Model 分成三段：主幹（backbone）」，「頸部 (neck)」和「頭部 (head)」

主幹抽出高階特徵，頭部負責做最後預測，而 Neck 是中間橋梁

(c) 用 attention

(d) data augmentation 與 更改 ground truth 的表現方法

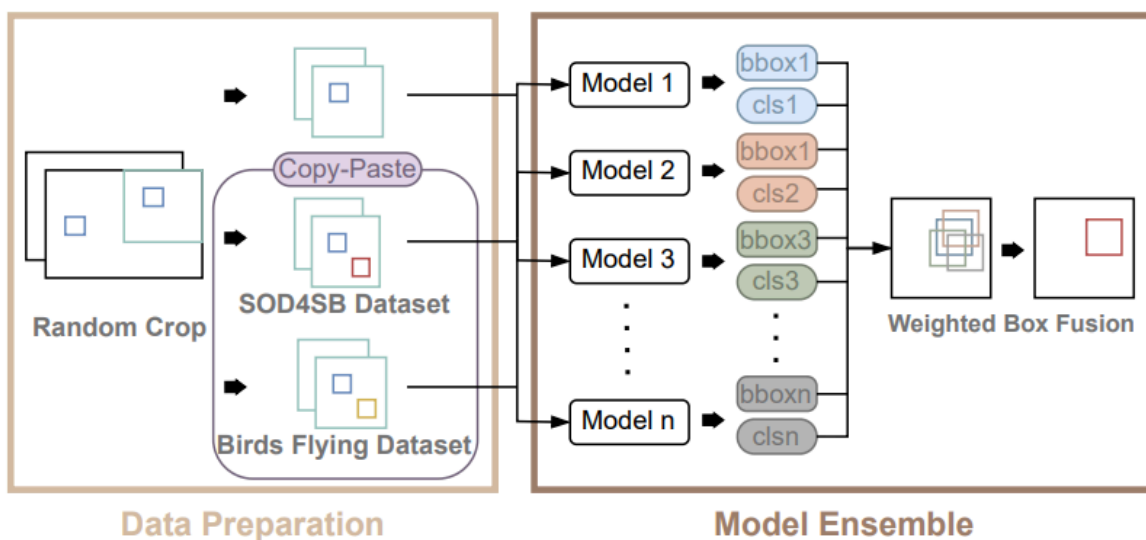
非極大值抑制（NMS）：找出最準確的 bounding box，可以使用 iou 或是 nwd 的方法

NWD：用 wessterin distance 來從多個框住同一個物體的 bounding box 中選出一個最好的 bounding box。因為 iou 在小 object 時可能因為都沒有重疊所以效果不好。

每一個 bounding box 都是一個 Gaussian Distribution，mean = 中心點，covariance matrix 用 w, h 去定義，接著算兩個 Gaussian Distribution 的 Wessterin Distance

Copy-Paste based augmentation：將小物體重複複製到其他的 image 中，增加小物體的出現率以及多變程度

## 方法



## Two Stage Training

### Data Preparation stage

從 SOD4SB Dataset 以及 Birds Flying Dataset 抽鳥圖片出來做 copy augmentation

### Model Ensemble Stage

WBF：根據每個模型預測的置信度分數，**加權平均**這些邊界框的位置和大小。消除多餘的邊界框，用 IOU Threshold 來 fuse bounding box，iou 不夠的 bounding box 會被當成另一個物體

NMS：從多個 Bounding Box 中選一

### Model Variant

用的 backbones 有 ResNet [10] and InternImage

InternImage 中有 DCNv3 operator (**Deformable Convolutional Networks**)，讓他的表現直達 transformer

## Experiments

## 可以學習的地方

# 問題