

OAC-NYCU-111-008 ( 研究報告 )

# 以 AI 為基礎之白海豚族群調查系統開發

正式報告

海洋委員會補助研究

中華民國 111 年 10 月

「本研究報告僅供海洋委員會施政參考，並不代表該會政策，該會保留採用與否之權利。」

OAC-NYCJ-111-008

以AI

為基礎之白海豚族群調查系統開發

正式報告

海洋委員會

OAC-NYCU-111-008 ( 研究報告 )

# 以 AI 為基礎之白海豚族群調查系統開發

正式報告

學校：國立陽明交通大學

指導教授：謝君偉

學生：劉冠麟

研究期程：中華民國 111 年 5 月至 111 年 12 月

研究經費：新臺幣 7 萬元

海洋委員會補助研究

中華民國 111 年 10 月

「本研究報告僅供海洋委員會施政參考，並不代表該會政策，該會保留採用與否之權利。」

# 目 次

表 次.....	2
圖 次.....	3
摘 要.....	5
第一章 前言 .....	7
第一節 研究緣起與問題背景.....	7
第二節 現況分析 .....	7
第三節 研究目的及研究重點.....	10
第四節 預期目標 .....	11
第二章 研究方法與過程 .....	12
第一節 影像資料搜集 .....	12
第二節 影像資料標註 .....	15
第三節 模型訓練方法 .....	16
第三章 結果與討論 .....	21
第四章 結論 .....	24
參考資料 .....	25

## 表 次

表 一 MS COCO 資料集與其他 SOTA 方法的比較 .....	20
-------------------------------------	----

## 圖次

圖 一 神經網路發展趨勢圖 .....	8
圖 二 Max pooling 運算結果圖 .....	9
圖 三 影像案例 ( 一 ) .....	12
圖 四 影像案例 ( 二 ) .....	13
圖 五 影像案例 ( 三 ) .....	13
圖 六 影像案例 ( 四 ) .....	14
圖 七 影像案例 ( 五 ) .....	14
圖 八 影像標記說明 .....	15
圖 九 標注後匯出之 txt 檔 .....	15
圖 十 PRB-FPN 架構圖 ( 一 ) .....	17
圖 十一 PRB-FPN 架構圖 ( 二 ) .....	17
圖 十二 Bottom-Up Fusion Module (BFM) .....	18
圖 十三 YOLO v4 與 PRB-FPN 於極小物件上的差異 .....	18
圖 十四 YOLOv4 和 PRB-FPN 之比較圖 .....	19
圖 十五 即時偵測結果 ( 一 ) .....	21
圖 十六 即時偵測結果 ( 二 ) .....	21
圖 十七 即時偵測結果 ( 三 ) .....	22

圖 十八 即時偵測結果 ( 四 ) .....	22
圖 十九 即時偵測結果 ( 五 ) .....	23
圖 二十 即時偵測結果 ( 六 ) .....	23

# 摘要

關鍵詞：白海豚偵測、深度學習、PRB-FPN

## 一、研究緣起

臺灣西部沿海的白海豚族群分布範圍與西部航運、填海造陸、興建海堤、工業工程、離岸風電等經濟、工業之重點發展及沿近海漁業活動區域重疊，所受到的人為衝擊相當大，而對此類數量極低的小族群，不可預期性的擾動更進一步提高其滅絕的風險。因此我們希望能夠運用我們對於 AI 人工智慧以及電腦視覺的技術去幫助白海豚之生態研究，讓相關團隊能夠更精準且確實的去進行白海豚族群保育及復育。

## 二、目的

在本研究計畫中，我們開發一套專門為白海豚的物件偵測系統，並為了增加其準確度，會出海並藉由空拍機去蒐集更多的白海豚影像，並以此當作資料集去訓練深度學習模型，並當此系統達到一定的準確度後，能夠實際應用於台中港北堤架設監控設備，進行長時間有效率的監測，並以此進行白海豚之族群分析。

## 三、研究方法與過程

我們首先出海以空拍機蒐集影像，從影片中分割出有白海豚出現之影像並進行標註，透過本實驗室提出的 PRB-FPN 神經網絡架構去進行白海豚的偵測，即時且準確的抓出影像中的白海豚，並加以定位。

## 四、重要發現

在本系統訓練完善後，目前大部分能夠清楚偵測出白海豚之蹤跡，我們從中隨機挑選影片進行測試，其平均即時正確率為 82.31%



## 五、結論與展望

台灣白海豚之數量已經日漸減少，但是我們對於白海豚之蹤跡仍然不太清楚，因此希望能夠藉由本套系統增加我們對白海豚族群蹤跡的了解，並能以此去進行妥善的保育與復育規劃。

# 第一章 前言

## 第一節 研究緣起與問題背景

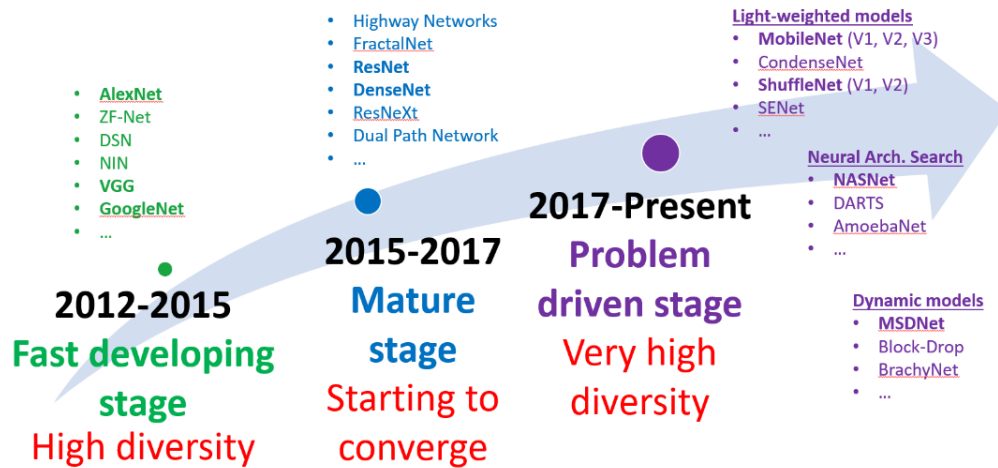
在海洋保育署有提到，臺灣西部沿海的白海豚族群分布範圍與西部航運、填海造陸、興建海堤、工業工程、離岸風電等經濟、工業之重點發展及沿近海漁業活動區域重疊，所受到的人為衝擊相當大，而對此類數量極低的小族群，不可預期性的擾動更進一步提高其滅絕的風險。海洋保育署彙整相關文獻資料，歸納影響白海豚族群其生存威脅因子為6大類：棲地劣化、海洋污染、食源減少、漁業混獲、海上活動及水下噪音等。

為此海洋保育署也成立了白海豚保育小組，並進行監測研究、棲地維護、人為衝擊管制、保育教育宣導與在地參與，而在其中最為重要的就是要去了解白海豚之族群分布與出現的蹤跡。

目前的白海豚監測方式是以人力目擊，其方法除了需要耗費大量人力之外，當觀測人疲勞時，相當容易錯失極少出現的白海豚蹤跡，也因此我們需要藉由機器去做監測，能夠達到 24 小時不間斷的監測，避免錯失任何一隻白海豚之出現。

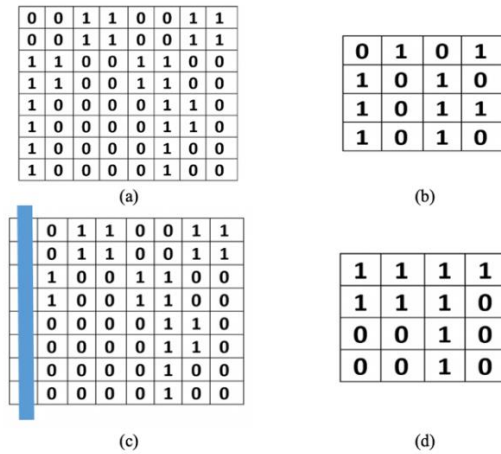
## 第二節 現況分析

深度學習從神經網路演進而來，大概從 2012 開始興起，到目前為止我們可以將發展階段分為三個時期，如圖一，2012 到 2015 為快速發展期，百花齊放，有各種深度學習網路架構被提出來，例如 VGG、AlexNet、GoogleNet 等等，這些網路架構的準度還不夠，接下來進入成熟期，2015 到 2017 年的階段，網路架構（backbone）往更深的方向進行，往提高準度方向進行，代表性的網路架構是 ResNet、DenseNet、DualPathNetwork 等等，這些網路架構的確提升了準度，但代價是犧牲了速度。



圖一 神經網路發展趨勢圖

到了 2017 後，進入成熟期，這時主要以 Problem-driven 為主，一個方向是輕量化，另一個方向是自動的網路設計（Neural Architecture Searching），因此在本計畫中我們也將朝網路架構輕量化進行，同時不只在速度的改善，同時也改進偵測的準確度。如前面所述，為了擷取有效的語意特徵，每一個 backbone 都須做 pooling 的運算，事實上 pooling 運算不是一種 shift-invariant 的運算，以圖二(a)為例，經由 MaxPooling，我們可以得到如圖二(b)的結果，如果圖二(a)往左 shift 一個像素（因為每次拍攝同一個物件不可能都會拍在同一個位置），得到如圖二(c)的結果，圖二(c)經過 MaxPooling 我們可以得到圖二(d)的結果，明顯的圖二(b)和圖二(d)結果不一樣，這意味著物件稍微有位移一下，pooling 會造成不一樣的特徵，也會造成物件偵測準確度的下降，特別是小物件偵測，此計畫將修改深度學習網路架構，來解決此問題。



圖一 Max pooling 運算結果圖

(a) : 原圖 ;(b) : pooling 結果

對於白海豚之偵測，為了讓拍攝廣度提升而提升空拍拍攝高度時，白海豚的大小常常會小於 20x20，在這種情況下，目前有名的物件偵測器方法皆無法有效偵測，這些偵測器為了提高準確度，都會用比較深的 backbone 網路來抽取特徵（如：darknet 53、ResNet、DenseNet 等），這些 backbone 的層數都很高，為了有效擷取物件的語意資料（semantics），通常都含有五層的 pooling（包含 convolution with stride 2），每一次 pooling 都會造成解析度在 X 或 Y 變一半，經過這五層降解析度的作用，原本大小為 32x32 的物件，通過神經網路後只會變成一個 pixels，那也意味著如果一個物件的尺度小於 32x32，就會缺乏語意資料，而造成無法偵測，因此可以得知現在有名的物件偵測器對偵測白海豚仍具有一定的困難與挑戰。2020 ICCV 會議，在 COCO 資料庫比賽中（LPIRC 2019），為了增加小物件的偵測準確度，往往會造成大物件偵測不好，反過來也是，因此在這比賽中，冠軍得主說他們忽視小物件的偵測，全力改善大物件與中物件的偵測準確度因而得到最高的準確度，此計畫將發展新的輕量化網路架構，並嘗試解決 pooling 的一些問題，不只速度的提升，且能對小物件的偵測準確度大大的提升。

因此在本研究中我們使用自行開發的 PRB-FPN，此模型架構也發表於影像處理的最頂級期刊 IEEE transactions on Image Processing，Impact factor = 10.865 (2021, Nov.)，此模型有一個平行特徵金字塔徹底解決上述提到的問題，我們創建了一個並行雙融合結構來並行融合三層特徵圖以生成三個預測，對應到不同尺度的物件，能保留更深和更淺的特徵，並能達到快速高準確度的極小物件檢測。

### 第三節 研究目的及研究重點

在本研究計畫中，我們開發一套專門為白海豚的物件偵測系統，並為了增加其準確度，會出海並藉由空拍機去蒐集更多的白海豚影像，以此當作資料集去訓練深度學習模型，並當此系統達到一定的準確度後，將其應用於海洋保育署之白海豚監測工作，運用定點設備去拍攝白海豚之蹤跡，並藉由本套白海豚即時偵測系統去偵測，取代以往人力辨識之方式，大大增進白海豚辨識之效率。

而在物件偵測模型的挑選中，我們並非選擇日前大家所使用的 YOLO 系列，我們選用本實驗室研發的 PRB-FPN 網路架構去進行白海豚偵測作業，改善了目前市面上之架構在小物件偵測上的問題，本架構能夠更加完善的去偵測出影像中的小物件，讓我們在監測白海豚的時候，在不失準確度的情況下，拍攝畫面能夠更廣，我們所需要的監測設備也就更少。

#### 第四節 預期目標

本研究計畫的主軸為物件偵測，而為了能夠符合白海豚生態監測之需求，我們必須提升準確度，期許能夠達到平均準確度 80%，並使模型輕量化，能夠即時偵測白海豚，記錄取白海豚出現之蹤跡。而在我們投入實驗過後，經過實測，平均準確度為 82.31%，在影片上也能即時偵測，達到了我們預期的目標。

台灣白海豚之數量已經日漸減少，但是我們對於白海豚之蹤跡仍然不太清楚，希望能夠藉由本套系統增加我們對白海豚族群蹤跡的了解，並以此去進行妥善的白海豚保育與復育規劃。

## 第二章 研究方法與過程

### 第一節 影像資料搜集

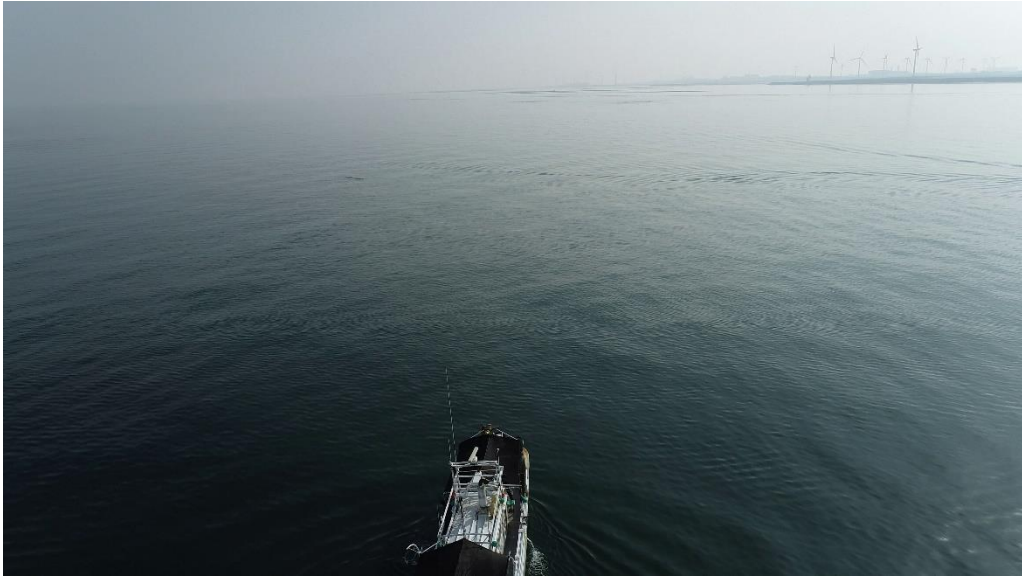
本研究計畫至目前為止共蒐集了 2273 張圖像資料，透過本團隊出海以空拍機拍攝影片，以及使用海保署與台電所提供的白海豚影片。在完成影片收集後，我們將蒐集之影片分出有白海豚出現之片段，並將其切為一張張圖片，並逐一從中挑出無白海豚之圖片並捨棄，剩餘的作為用來訓練神經網路模型的資料。

以下為本計畫研究中收集的影像範例。

以圖二及圖三為例，此兩張影像為有斜角的空拍影像，此兩張影像之白海豚物件相當小，圖三之兩隻白海豚出現於圖片之左上角，而圖四之兩隻白海豚正要潛入水中，在我們的標註作業中也有紀錄起此種正要潛入水中之白海豚，希望能夠將水中(淺層)之白海豚也偵測出來。

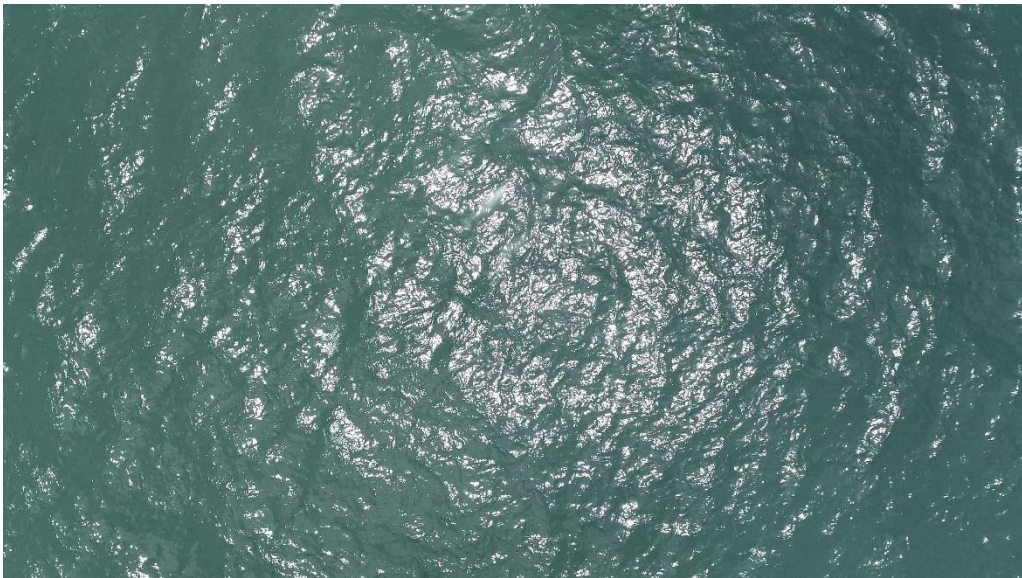


圖二 影像案例 (一)



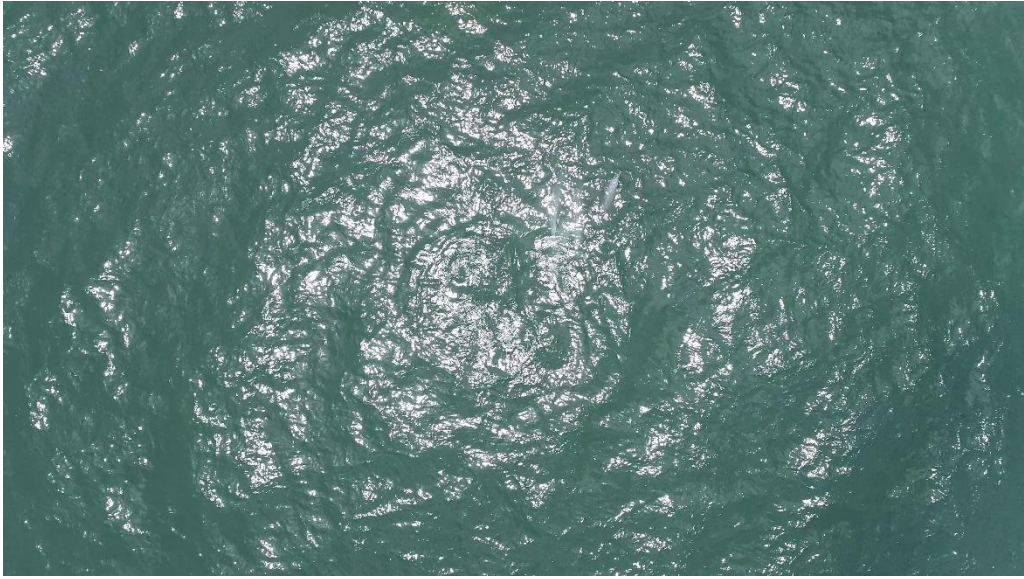
圖三 影像案例 ( 二 )

圖四、圖五及圖六為第二份空拍影像，拍攝角度為垂直海面，雖說此類影像白海豚出現之大小相對較大，但如圖五、圖六所示，海面反光相當嚴重，容易將身軀顏色為白色的白海豚搞混，而在圖七中反光情況較為輕微，因此此圖之白海豚雖在海中，但由於物件相對明顯，此圖在我們的資料集中為簡單的例子。

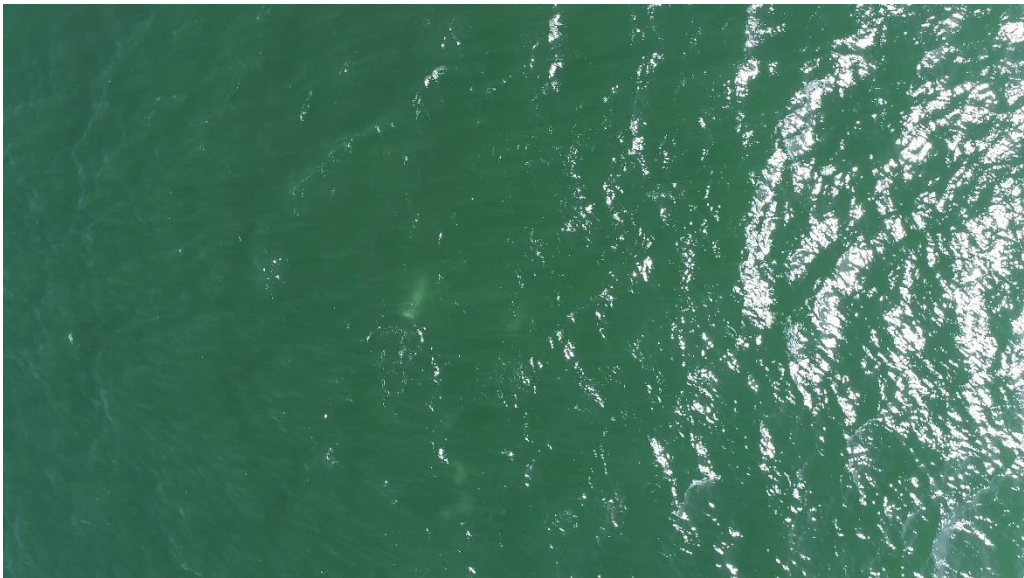


圖四 影像案例 ( 三 )





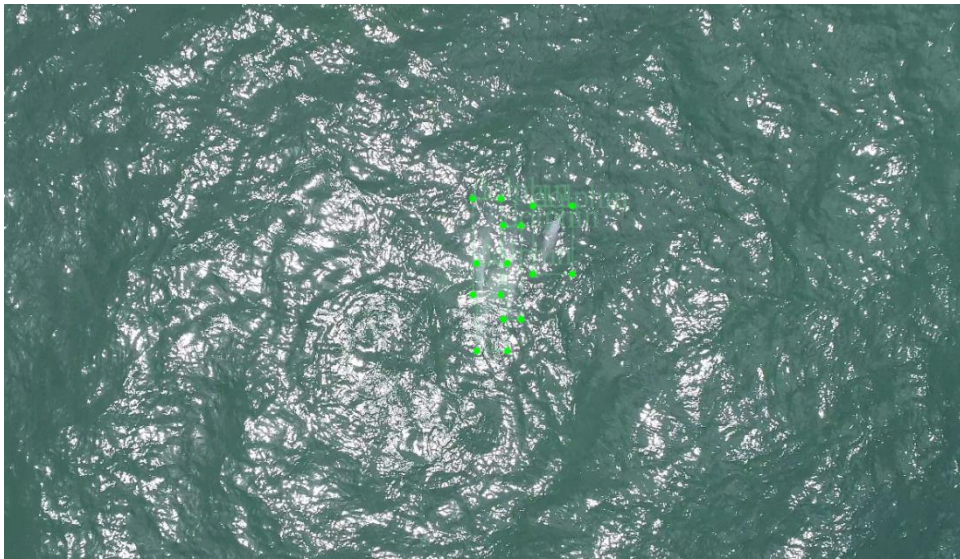
圖五 影像案例 ( 四 )



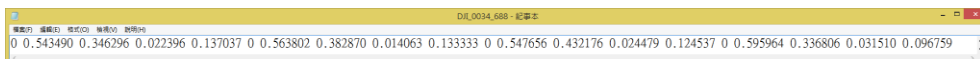
圖六 影像案例 ( 五 )

## 第二節 影像資料標註

在影像資料標註的工作項目中，我們使用 LabelImg 來做為標註工具，在標註目標位置及名稱時，LabelImg 會將這些資訊儲存在 txt 檔裡，方便日後訓練模型時讀取。以圖八為例，當中綠色方框為人工畫出的標記框，在此圖中出現了四隻白海豚，其中有三隻在水內，一之探出頭，只要我們能夠辨識出為白海豚之物件，都會將其進行標註，無論在水內還是水上，也因此才能讓我們的偵測模型能夠偵測到水下的白海豚。在完成標註作業後，我們將這些資訊轉換為如圖九一樣的 txt 檔，供訓練模型讀取。



圖七 影像標記



圖八 標註後匯出之 txt 檔

### 第三節 模型訓練方法

在完成資料蒐集與標注後，我們運用深度學習的技巧，選用實驗室自行研發的 PRB-FPN 神經網絡架構去進行訓練，以此達到高準確度且即時偵測的效果，以下為我們訓練環境，以及模型選擇之考量。

訓練環境：

作業系統：Ubuntu 18.04.2 LTS

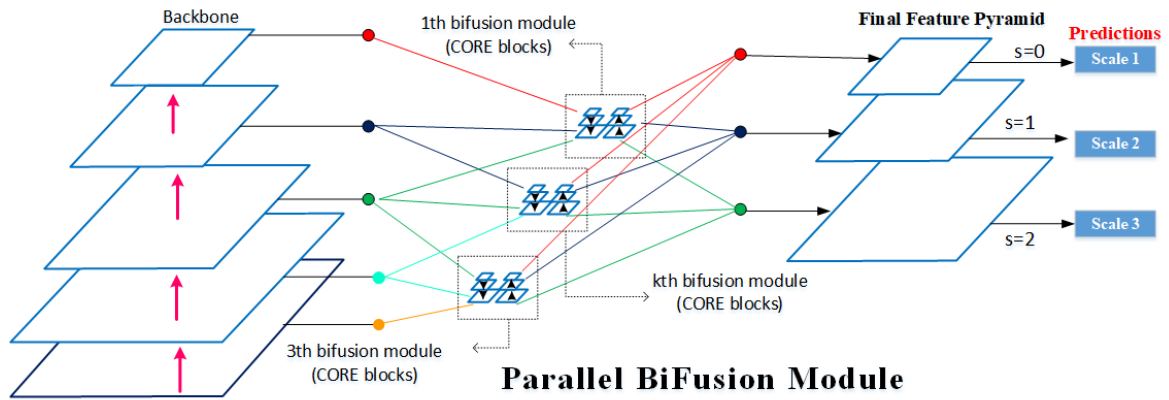
GPU：Geforce RTX 1080 ti

CUDA version：10.2

在模型選擇方面，我們選用本團隊開發之 PRB-FPN (Parallel Residual Bi-Fusion Feature Pyramid Network) 對於現今常用之 YOLOv4，本架構能夠更加準確地去偵測出極小物件，並在不同大小的物件同時出現時，能夠同時準確偵測出不同尺度之物件。

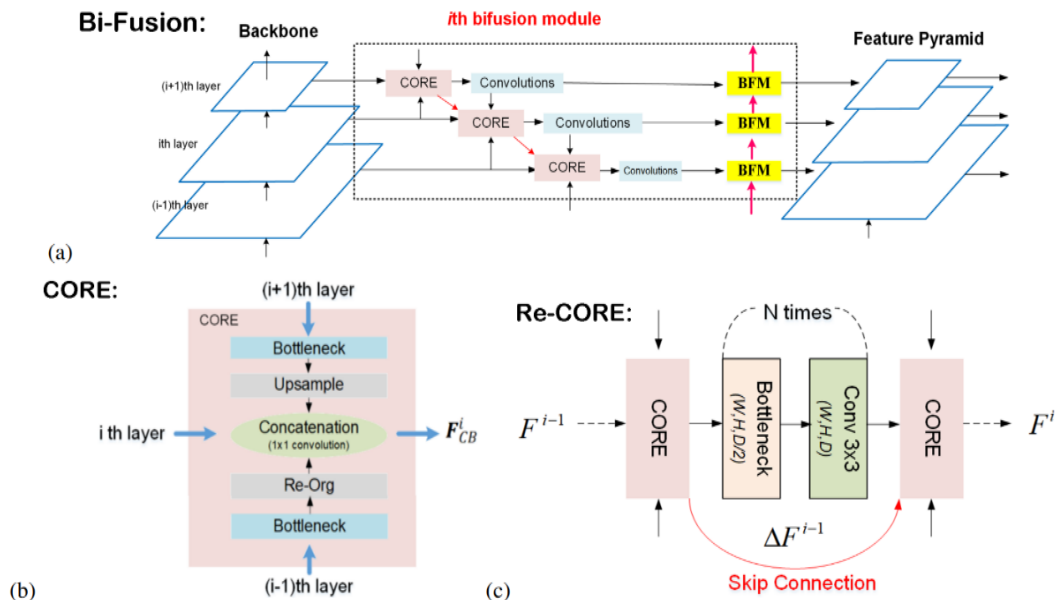
特徵金字塔 (Feature Pyramid, FP) 為一種特徵結構，將特徵圖弄成不同的解析度，從 top 到 down 的方式將特徵資料往下傳給淺層的特徵層來增加物件偵測的準確度，如圖十，YOLO v3 跟 SSD 都有引進此 FP 架構來大幅提升物件偵測的準確率，這種 Feature Pyramid 架構對大物件有不錯的效果，但因為要有效擷取語意特徵，絕大部分的 backbone 都需要有五層 pooling，因為在這 FP 架構下，如果物件小於 32x32，在最後一層此物件會消失不見，就無法將物件的語意資料往下，從 top 到 down 的方式傳到下一層，因此 FP 架構對小物件效果就沒有那麼好，對此我們提出 BiFusion 的架構，藉由混和上下層的特徵，經由純化得到最佳的語意資訊，能有效地偵測出小物件，

我們開發的這種 PRB-FPN 架構具有並行設計和多項改進，除了可以保留更深層和更淺層的特徵，並能達到快速準確的單次目標檢測，並創建了並行雙融合結構來並行融合三層特徵圖以生成三個預測同時繪製地圖。



圖十 PRB-FPN架構圖(一)

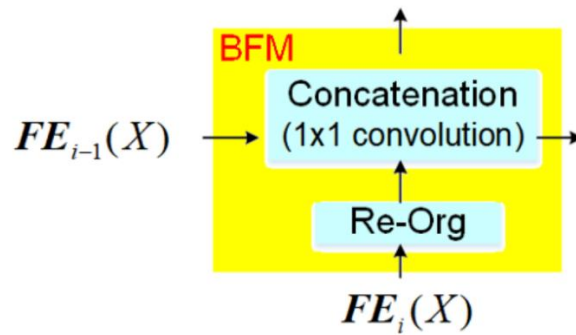
如下圖十一(a)，在本架構中我們設計一套由下而上融合特徵的 Bifusion module，形成平行 bifusion FP(特徵金字塔)，其中我們藉由 CORE，如圖十一(b)，去融合特徵，提供一個從下而上的特徵融合途徑，使讓特徵金字塔可以雙向融合特徵，並取得較低層遺失的特徵。並設計一套全新的 Re-CORE，如圖十一(c)，以 ResNet 的概念做發想，我們在原先的 CORE 中加入 skip connection 技巧，讓模型能夠輕鬆訓練到淺層與深層的特徵。



圖十一 PRB-FPN架構圖(二)

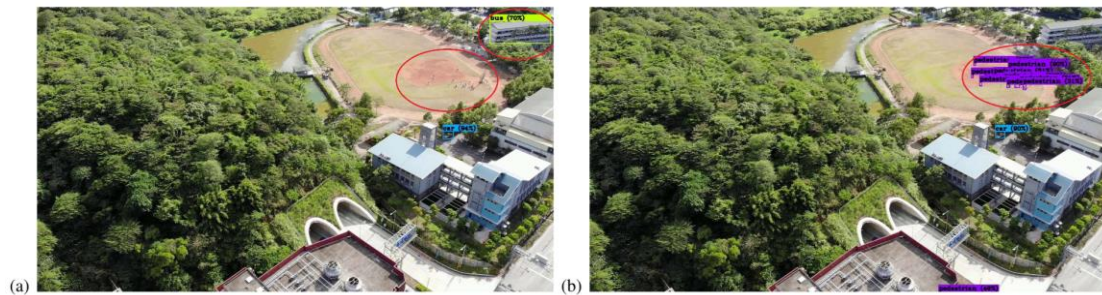


而為了能夠一次以高精度偵測大小物件，我們也設計了一套 Bottom-Up Fusion Module(BFM)將上下層特徵平行融合，如下圖十二，假設輸入影像為  $X$ ， $FE_l(X)$ 代表第  $l$ 層的特徵，我們嘗試保留上一次的特徵 $FE_{l-1}(X)$ 及目前層 $FE_l(X)$ 的特徵，混合新的特徵，遞迴處理每一層，就可以有效地偵測小物件。



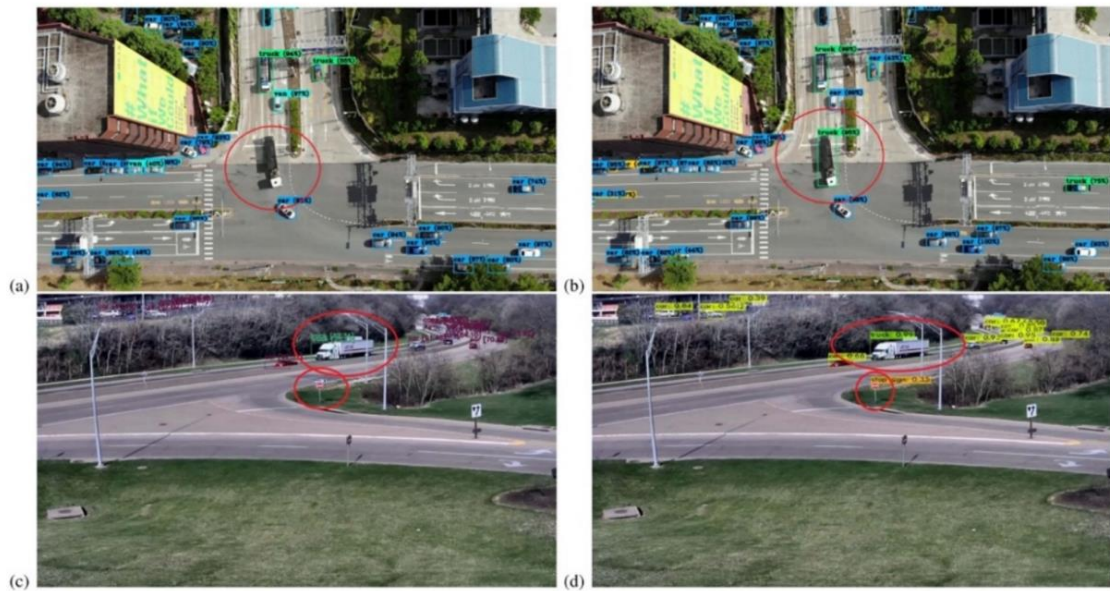
圖十二 Bottom-Up Fusion Module(BFM)

目前物件偵測的相關應用有很大部份開發於YOLO系列上，但在極小物件的實際應用中，YOLO系列的方法表現有限，下圖十三為我們的技術與YOLO v4在小物件上辨識結果的差異。



圖十三 YOLO v4與 PRB-FPN 於極小物件上的差異

圖十四顯示了 YOLOv4 和 PRB-FPN 之間目標檢測的比較，YOLOv4 在同時檢測大物體和小物體時不能有很好的表現，因為 YOLOv4 放大輸入圖像以檢測小物體的方法通常無法檢測大物體並且效率並不好，在圖十四(a)中，貨運卡車被 YOLOv4 遺漏，但我們的 PRB-FPN 成功檢測，如圖十四(b)所示，此外，YOLOv4 經常將一個大對象檢測為幾個小對象，如圖十四(c)所示，一輛卡車被檢測為兩輛汽車，PRB-FPN 可在圖十四(d)中成功檢測到它，另外 YOLOv4 在圖十四(c)中未偵測成功的停車標誌，在我們的 PRB-FPN 中圖十四(d)能順利偵測出來。



圖十四 YOLOv4 和 PRB-FPN 之比較圖

下表(表一)為在實驗設備 NVIDIA VOLTA V100 下，我們的 PRB-FPN 方法在 MS COCO 資料集上與其他 SOTA 方法的比較。AP 為 Average Precision，可以看做整個模型的表現，分數越高表示模型能力越佳，而 AP 後的不同數字及英文字母，表示對偵測框所採用的不同閾值及偵測框的面積範圍。

表一 MS COCO 資料集與其他 SOTA 方法的比較

Method	Backbone	Input size	FPS	AP	AP50	AP75	APS	APM	APL
YOLOv4 [13]	CSPDarknet-53 [47]	512x512	83	43	64.9	46.5	24.3	46.1	55.2
EfficientDet-D0 [11]	Efficient-B0 [48]	512x512	97.0	33.8	52.2	35.8	12	38.3	51.2
EfficientDet-D1 [11]	Efficient-B1 [48]	640x640	74.0	39.6	58.6	42.3	17.9	44.3	56
EfficientDet-D2 [11]	Efficient-B2 [48]	768x768	57.0	43	62.3	46.2	22.5	47	58.4
EfficientDet-D3 [11]	Efficient-B3 [48]	896x896	36.0	47.5	66.2	51.5	27.9	51.4	62.0
SM-NAS: E2 [50]		800x600	25.3	40	58.2	43.4	21.1	42.4	51.7
SM-NAS: E3 [50]		800x600	19.7	42.8	61.2	46.5	23.5	45.5	55.6
SM-NAS: E5 [50]		1333x800	9.3	45.9	64.6	49.6	27.1	49.0	58.0
NAS-FPN [12]	ResNet-50 [5]	640	24.4	39.9					
NAS-FPN [12]	ReNet-50 [5]	1024	12.7	44.2					
ATSS [51]	ResNet-101 [5]	800x	17.5	43.6	62.1	47.4	26.1	47	53.6
ATSS [51]	ReNet-101 [5]	800x	13.7	46.3	64.7	50.4	27.7	49.8	58.4
YOLOv7 [52]	E-ELAN [52]	640x640		51.4	69.7	55.9	31.8	55.5	65.0
RB-FPN [Ours]	CSPDarknet-53 [47]	512x512	76.9	45.1	67.2	48.2	27.1	48.5	57
PRB-FPN [Ours]	CSPDarknet-53 [47]	800x800	37.5	48.9	69.5	55.9	30.8	55.9	60.2
PRB-FPN-CSP* [Ours]	CSPDarknet-53 [47]	640x640		51.8	70.0	56.7	32.6	55.5	64.6
PRB-FPN* [Ours]	E-ELAN [52]	640x640	28.5	52.5	70.4	57.2	33.4	56.2	65.8
PRB-FPN-E6E* [Ours]	E-ELAN [52]	1024x1024		55.1	72.2	60.6	37.0	59.2	68.0
PRB-FPN-W6* [Ours]	E-ELAN [52]	1024x1024	-	-	-	-	-	-	-

### 第三章 結果與討論

本研究以自行出海以空拍機蒐集之影像，以及海保署與台電提供之影像做為資料集進行模型訓練，訓練完畢後，隨機選取影片進行偵測能夠大部分準確且即時偵測出白海豚，其平均偵測準確度為 82.31%。

圖十五為白海豚較小且密集的偵測結果。



圖十五 即時偵測結果 (一)

圖十六及圖十七為清楚且較大的白海豚偵測結果。



圖十六 即時偵測結果 (二)





圖十七 即時偵測結果 ( 三 )

圖十八與圖十九為清楚較大且密集的偵測結果。



圖十八 即時偵測結果 ( 四 )



圖十九 即時偵測結果 ( 五 )

圖二十為極小的白海豚偵測結果。



圖二十 即時偵測結果 ( 六 )

## 第四章 結論

白海豚目前在台灣的數量日漸減少，為了保育與復育白海豚，海洋保育署花費許多資源投入在白海豚計畫當中，其中包含監測、維護、管制、宣導，而我們團隊主要研究電腦視覺與 AI 深度學習，藉由我們的技術，以 AI 模型去進行白海豚偵測，能夠大大增進白海豚監測作業之效率，也能讓平時架設之監測設備能夠加入偵測系統，即時去偵測白海豚之蹤跡，讓白海豚的保育相關研究能夠更加順利。

目前物件偵測仍然為電腦視覺重要的一環，而其普遍使用之架構為 YOLO 系列之神經網絡架構，但我們選用自行研發之 PRB-FPN 神經網絡架構，改善以往極小物件偵測效果不好的問題，因此當白海豚之監測系統納入本套白海豚偵測系統後，能夠以更加遠(高)的距離去做白海豚監測，讓監測設備更加節省，監測範圍也能更廣。

最後在進行完相關研究後，本套系統之平均偵測準確度達到 82.31%，且能夠在影片上即時偵測，讓未來應用之單位能夠即時偵測白海豚，並即時回報給相關單位去做追蹤，也能夠藉此更加了解白海豚之族群習慣。

白海豚保育與復育計畫牽涉甚廣，與西岸人們的生活習慣與政府經濟發展規劃息息相關，在規畫相關政策時要考量白海豚之族群生態之前，我們必須要先對白海豚有足夠的了解，才能避免浪費不必要的資源，而要對白海豚有足夠的了解就需要進行大量且長期的監測，藉由大量的數據去歸納出白海豚的活動範圍與時間，但若是以出海人力監測的方式除了監測時間過短之外，也需要大量耗費人力且無效率，因此我們這套以 AI 模型去進行白海豚偵測的系統，能夠應用在白海豚之監測作業，只需要架設好監視器後，並配置適當的設備，就能夠即時且準確的進行白海豚偵測，達到預期之長時間且準確有效率的白海豚監測作業，拓增我們對白海豚的了解，也能更加知道如何去對其保育與復育，在政策設計上也能去避免影響到白海豚之生態，共同維護台灣週遭的海洋生物多樣性。

## 參考資料

- [1] Ping-Yang Chen, Jun-Wei Hsieh, Chien-Yao Wang, Hong-Yuan Mark Liao, “Recursive Hybrid Fusion Pyramid Network for Real-Time Small Object Detection on Embedded Devices,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2020, pp. 402-403.
- [2] Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, I-Hau Yeh, “CSPNet: A New Backbone That Can Enhance Learning Capability of CNN,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2020, pp. 390-391.
- [3] Ping-Yang Chen, Jun-Wei Hsieh, Munkhjargal Gochoo, Ming-Ching Chang, Chien-Yao Wang, and Hong-Yuan Mark Liao, “DRONE-BASED VEHICLE FLOW ESTIMATION AND ITS APPLICATION TO TRAFFIC CONFLICT HOTSPOT DETECTION AT INTERSECTIONS,” IEEE conference on image processing, 2020.
- [4] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma. Single-image crowd counting via multi-column convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 589 – 597, 2016
- [5] Mark Marsden, Kevin McGuinness, Suzanne Little, Noel E. O'Connor ResnetCrowd A Residual Deep Learning Architecture for Crowd Counting, Violent Behaviour Detection and Crowd Density Level Classification. In Proceedings of the AVSS 2017
- [6] Kai Kang, Chen Change Loy, and Xiaogang Wang. Deeply Learned Attributes for Crowded Scene Understanding. In Computer Vision and Pattern Recognition, 2015
- [7] Z. Lin, L.S. Davis, D. Doermann, and D. DeMenthon, “Hierarchical Part-Template Matching for Human Detection and Segmentation,” International Conference on Computer Vision, pp. 1-8, Oct. 2007.

- [8] B. Wu and R. Nevatia, "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors," International Conference on Computer Vision, Vol. 1, pp. 90-97, Oct. 2005.
- [9] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.886-893, June 2005.
- [10] Y. Ding and J. Xiao, "Contextual Boost for Pedestrian Detection," IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2895-2902, June 2012.
- [11] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," In Computer Vision and Pattern Recognition, pp. 1-8, 2008.
- [12] P. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," IEEE transactions on pattern analysis and machine intelligence, vol.32, no.9, pp.1627-1645, 2010.
- [13] G. Shu, A. Dehghan, and M. Shah, "Improving an Object Detector and Extracting Regions using Superpixels," IEEE Conference on Computer Vision and Pattern Recognition, pp. 3721-3727, June 2013.
- [14] R. Benenson, M. Mathias, R. Timofte, and L. V. Gool, "Pedestrian detection at 100 frames per second," IEEE Conference on Computer Vision and Pattern Recognition, pp. 2903-2910, June 2012.
- [15] R. Girshick, F. Iandola, T. Darrell, and J. Malik, "Deformable part models are convolutional neural networks," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.437-446, 2015.
- [16] A. Gaszczak, T. P. Breckon and J. Han, "Real-time people and vehicle detection from UAV imagery," IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2011.

- [17] G. J. Burghouts, A. W. M. van Eekeren, J. Dijk, “Focus-of-attention for human activity recognition from UAVs,” *Proceedings of the SPIE*, vol. 9249, pp.12, Oct. 7, 2014.
- [18] Fast crowd density estimation with convolutional neural networks. *Engineering Applications of Artificial Intelligence*, 2015.
- [19] Weixin Li, Vijay Mahadevan, and Nuno Vasconcelos. Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.
- [20] Kawak, D., Kim, D., Liu, R., Nath, B., Iftode, L., “DoppelDriver: counterfactual actual travel times for alternative routes,” *IEEE PerCom*, pp. 178 – 185, 2015
- [21] H. Aly, A. Basalamah, and M. Youssef, “Lanequest: An accurate and energy-efficient lane detection system,” *IEEE PerCom*, pp. 163 – 171, 2015.
- [22] C. Chen, D. Zhang, B. Guo, X. Ma, G. Pan, and Z. Wu, “Trippanner: Personalized trip planning leveraging heterogeneous crowdsourced digital footprints,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1259 – 1273, Jun., 2015.
- [23] Y. T. Chuang, C. W. Yi, Y. C. Tseng, C. S. Nian, and C. H. Ching, “Discovering phase timing information of traffic light systems by stopgo shockwaves,” *IEEE Trans. Mobile Comput.*, vol. 14, no. 1, pp. 58 – 71, Jan. 2015.
- [24] Fayazi, S.A. and Vahidi, A., “Crowdsourcing Phase and Timing of Pre-Timed Traffic Signals in the Presence of Queues: Algorithms and Back-End System Architecture,” *IEEE Trans. Intelligent Transportation Systems* , vol. 17, no. 3, pp. 870 – 881, Mar. 2016.
- [25] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv:1409.1556*, 2014.
- [26] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," In *CVPR*, 2014.

- [27] S. Ren, et al., “Faster r-cnn: Towards real-time object detection with region proposal networks,” In NIPS, 2015.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” In ECCV, 2014.
- [29] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask r-cnn,” In ICCV, 2017.
- [30] T.Y. Lin, et al., “Feature pyra-mid networks for object detection,” In CVPR, 2017.
- [31] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” arXiv:1804.02767, 2018.
- [32] W. Liu, et al., “SSD: Single shot multibox detector,” In ECCV, 2016.
- [33] J. Redmon and A. Farhadi. “Yolo9000: better, faster, stronger,” arXiv:1612.08242, 2016.
- [34] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” In CVPR 2016.
- [35] D. Scherer, A. Muller, and S. Behnke, “Evaluation of pooling operations in convolutional architectures for object recognition,” In ICANN, 2010.
- [36] T. Wang, et al. “Learning Rich Features at High-Speed for Single-Shot Object Detection,” In ICCV 2019.
- [37] S.-W. Kim, H.-K. Kook, J.-Y. Sun, M.-C. Kang, and S.-J. Ko, “Parallel Feature Pyramid Network for Object Detection,” In ECCV, 2018.
- [38] S. Y. Woo, et al., “Gated bidirectional feature pyramid network for accurate one-shot detection,” vol. 30, pp.543-555, machine vision and applications, 2019.
- [39] P. Sermanet, et al., “Overfeat: Integrated recognition, localization and detection using convolutional networks,” In ICLR, 2014.
- [40] J. Redmon, et al., “You only look once: Unified, real-time object detection,” In CVPR, 2015.

- [41] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal Loss for Dense Object Detection,” In ICCV, 2017.
- [42] R. Zhang, “Making Convolutional Networks Shift-Invariant Again,” In ICLR 2019.
- [43] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, “Single-Shot Refinement Neural Network for Object Detection,” In CVPR, 2018.
- [44] R. J. Wang, X. Li, C. X. Ling, “Pelee: A Real-Time Object Detection System on Mobile Devices,” In NIPS, 2018.
- [45] G. Huang, Z. Liu, L. van der Maaten, L., K.Q. Weinberger, “Densely connected convolutional networks,” In CVPR, 2017.
- [46] A. Khosla, N. Jayadevaprakash, B. Yao, and F.-F. Li, “Novel dataset for fine-grained image categorization: Stanford dogs,” In Proc. CVPR Workshop on Fine-Grained Visual Categorization, volume 2, page 1, 2011.
- [47] C.Y. Wang, et al., “CSPNet: A New Backbone that can Enhance Learning Capability of CNN,” arXiv e-prints arXiv:1911.11929 (Nov 2019).
- [48] H. Law and J. Deng, “CornerNet: Detecting Objects as Paired Keypoints,” In ECCV, 2018.
- [49] X. Wu, et al., “Single-Shot Bidirectional Pyramid Networks for High Quality Object Detection,” AAAI 2018.
- [50] Z. Tian, C. Shen, H. Chen, and T. He, “Fcos: Fully convolutional one-stage object detection,” In CVPR 2019.
- [51] C. Zhu, Y. He, and M. Savvides, “Feature selective anchor-free module for single-shot object detection,” In CVPR 2019.
- [52] J. Cao, Y. Pang, J. Han, and X. Li, “Hierarchical Shot Detector,” In ICCV 2019.
- [53] Q. Zhao, et al., “M2Det: A Single-Shot Object Detector based on Multi-Level Feature Pyramid Network,” AAAI 2019.



- [54] E. Bochinski, T. Senst and T. Sikora, “Extending IOU Based Multi-Object Tracking by Visual Information,” In AVSS 2018
- [55] A. Bewley, Z. Ge, L.Ott, F. Ramos and B. Upcroft, “Simple online and real-time tracking,” IEEE International Conference on Image Processing, 2016.
- [56] C. Dicle, M. Sznaiier, and O. Camps, “The way they move: Tracking multiple targets with similar appearance,” In International Conference on Computer Vision, 2013.
- [57] S. H. Rezatofighi, A. Milan, Z. Zhang, A. Dick, Q. Shi, and I. Reid. “Joint Probabilistic Data Association Revisited,” In International Conference on Computer Vision, 2015.
- [58] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, “Multiple Hypothesis Tracking Revisited. In International Conference on Computer Vision,” 2015.
- [59] D. Reid, “An Algorithm for Tracking Multiple Targets,” Automatic Control, Vol. 24, pp. 843 – 854, 1979.