中原大學

中

原

大

學

資訊工程學系

碩士

學位論文

以Y

O

L

O

v

7

為基礎之旋轉物件辨識方法

林冠良

中

華

民

國

112

年

月

資訊工程學系

碩士學位論文

以YOLOv7為基礎之旋轉物件辨識方法

A Oriented-Object Detector Based On YOLOv7

指導教授：蘇志文

研 究 生：林冠良

中華民國 112 年 6 月

**摘要**

近年來，物件偵測技術在空拍影像應用中扮演了越來越重要的角色。然而，現有的物件偵測模型通常只能偵測物件的位置和類別，而無法辨識物件的角度。為了解決這個問題，我們對YOLOv7模型進行了修改，以實現對物件角度的準確偵測。這些修改包括增加了一個旋轉檢測層，用於檢測物件的旋轉角度，並在損失函數中加入了旋轉角度的損失項，以進一步提高模型的準確性。此外，我們還加入了旋轉的NMS，以確保偵測出的物件框框能夠更好地貼合物件。

為了驗證修改後的YOLOv7模型的效果，我們使用了HRSC、DOTA1.0和我們自己的十字路口車流資料集進行了實驗。實驗結果表明，修改後的YOLOv7模型在這些資料集上都取得了不錯的成果，並且在物件角度偵測方面優於現有的物件偵測模型。這驗證了我們的修改可以有效地提高物件偵測模型的準確性，並有望未來在實際應用中發揮更大的作用。

**Abstract**

In recent years, object detection technology has played an increasingly important role in aerial image applications. However, existing object detection models can only detect the position and category of objects, but not the orientation of objects. To address this issue, we modified the YOLOv7 model to achieve accurate detection of object orientation. These modifications included adding a rotation detection layer to detect the rotation angle of objects, and adding a loss term for the rotation angle in the loss function to further improve the model's accuracy. In addition, we also added rotational non-maximum suppression (NMS) to ensure that the detected bounding boxes fit the objects more accurately.

To verify the effectiveness of the modified YOLOv7 model, we conducted experiments using three datasets: HRSC, DOTA1.0, and our own crossroad traffic dataset. The experimental results showed that the modified YOLOv7 model performed better on all three datasets and outperformed existing object detection models in detecting object orientation. This indicates that our modifications can effectively improve the accuracy of object detection models and have the potential to have a greater impact in future practical applications, including traffic management and surveillance in crossroad scenarios.

**致謝**

**目次**

[摘要 I](#_Toc134349774)

[Abstract II](#_Toc134349775)

[致謝 III](#_Toc134349776)

[目次 IV](#_Toc134349777)

[第一章 緒論 1](#_Toc134349778)

[1.1 研究動機 1](#_Toc134349779)

[1.2 論文架構 2](#_Toc134349780)

[第二章 相關文獻 3](#_Toc134349781)

[第三章 研究方法 4](#_Toc134349782)

[3.1 物體偵測 5](#_Toc134349783)

[3.1.1 新的重參數化方法（Re-parameterized model） 5](#_Toc134349784)

[3.1.2 新的動態標籤分配策略（Dynamic label assignment strategy） 6](#_Toc134349785)

[3.1.3 模型縮放與擴展（Compound model scaling） 6](#_Toc134349786)

[3.1.4 更快的速度與更高的準確率 7](#_Toc134349787)

[3.2 角度偵測 9](#_Toc134349788)

[3.2.1 旋轉邊界框之角度定義 9](#_Toc134349789)

[3.2.2 角度偵測頭 10](#_Toc134349790)

[3.3 數據增強 12](#_Toc134349791)

[3.3.1馬賽克（Mosaic） 12](#_Toc134349792)

[3.3.2翻轉（Flip up-down, Flip left-right） 12](#_Toc134349793)

[3.3.3旋轉（Rotation） 12](#_Toc134349794)

[3.4 基於YOLOv7之空拍單階段旋轉邊界框物體偵測器 13](#_Toc134349795)

[第四章 實驗結果與分析 14](#_Toc134349796)

[4.1 實驗環境 14](#_Toc134349797)

[4.2 實驗資料 14](#_Toc134349798)

[4.3 實驗結果 15](#_Toc134349799)

[第五章 結論與未來方向 16](#_Toc134349800)

[參考文獻 17](#_Toc134349801)

**第一章 緒論**

**1.1 研究動機**

隨著無人機的普遍化，物件辨識也開始大量應用於遙測航拍影像的辨識上，雖然說在前幾年的經驗中，水平邊界框對於密集排列與長寬比變化大的物體辨識上，稍顯無力，使得大部分的研究者開始去研究以旋轉擬合框為基礎的偵測器模型，在本篇論文中，因受到YOLOv7強大的性能啟發，希望能以YOLOv7為基礎，在原本成熟的模型上，以旋轉擬合框加入角度的偵測，去研究如何對空拍影像中的物件進行角度偵測是一個具有重要意義的課題。

針對這個問題，我們對YOLOv7模型進行了修改，以實現對物件角度的準確偵測。我們增加了一個旋轉檢測層，用於檢測物件的旋轉角度，並在損失函數中加入了旋轉角度的損失項，以進一步提高模型的準確性。此外，我們還加入了旋轉的NMS，以確保偵測出的物件框框能夠更好地貼合物件。

為了驗證我們提出的方法的有效性，我們採納了HRSC和DOTA1.0這兩個具有代表性的資料集，同時為了多方驗證，也使用了本實驗室有的十字路口空拍車流資料集進行實驗。實驗結果顯示，我們修改後的YOLOv7模型可以準確地偵測空拍影像中的物件角度，並在各種複雜場景中取得了較好的效果。因此，我們相信我們的研究可以為空拍影像中的物件偵測問題提供一個新的解決方案，有望在未來的實際應用中發揮更大的作用。

**1.2 論文架構**

在本論文中，共分為五個章節。

* 第一個章節為緒論，其中包括研究動機以及論文架構。
* 第二章節將會介紹本論文相關之文獻，。
* 第三章節闡述本論文之研究方法，說明本論文使用之基礎架構的優點、說明如何加入旋轉角度之偵測頭、與使用的數據增強方式、替新增的角度參數加入損失函數，最後統整本論文提出之模型架構。
* 第四章節的部分展示本論文提出之方法所得到的實驗結果，並與不同的方法的實驗結果進行比較與分析。
* 第五章節作為總結並依據實驗結果提出相對應之結論，進一步提出未來的發展方向。

**第二章 相關文獻**

**第三章 研究方法**

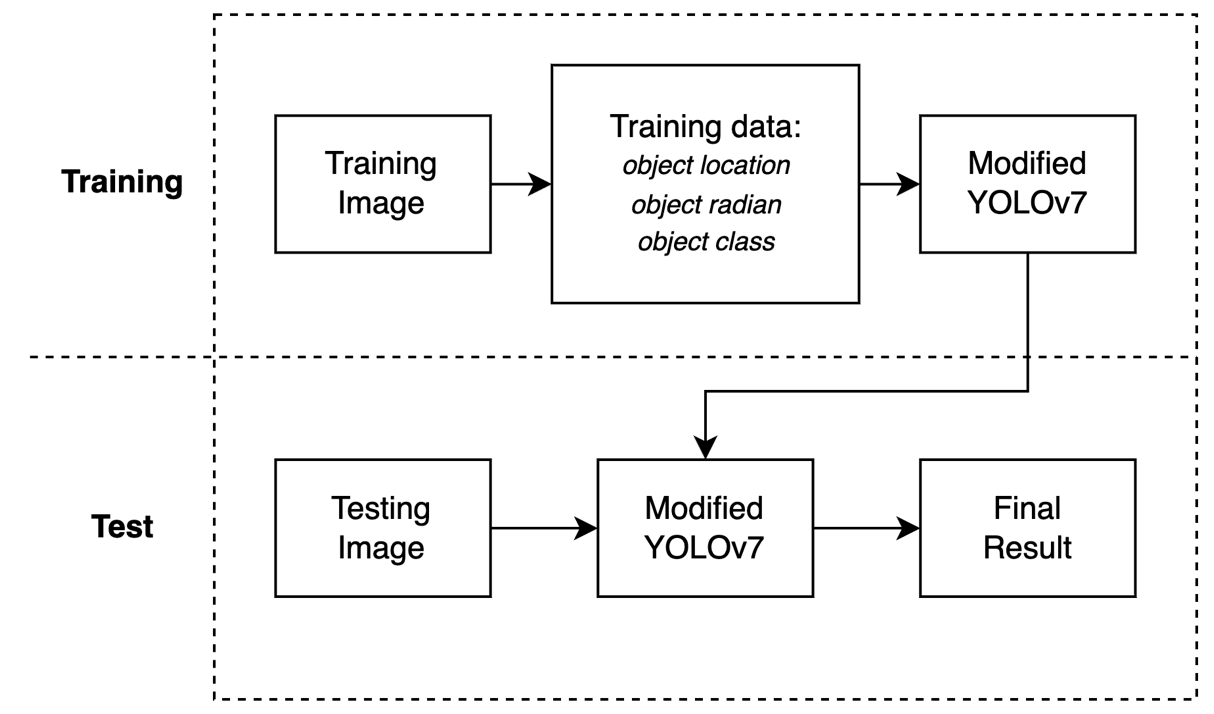
於此章節，我們將詳細說明本論文所使用之基礎模型以及其優點，並接著介紹我們提出之修改方法，為了應付多類別以及多物體的偵測任務，我們選用本身在物件偵測上已有優異表現的YOLOv7，並額外在模型末端之預測層中加入旋轉角度的偵測頭，使模型額外具備預測物體角度的能力；而在角度偵測的方面，為了不使模型在預測層上過於厚重，我們則是採用直接迴歸的方式去預測物體的角度。

圖 本方法訓練及測試階段流程示意圖

**3.1 物體偵測**

2022年由Chien-Yao Wang等人提出YOLOv7，為YOLO系列提出了多種修改的方法，提高其準確性、速度以及穩健性，主要修改的部分包括三類：

3.1.1 新的重參數化方法（Re-parameterized model）

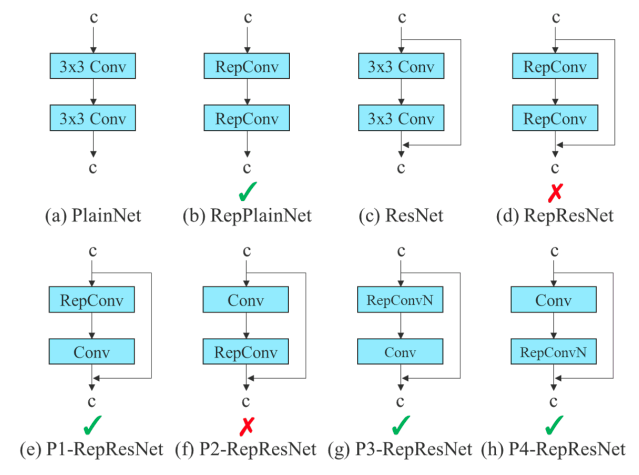
作者等人發現RepConv中的Identity連結會破壞ResNet中的殘差（Residual）與DenseNet中的串連（Concatenation），提出不使用Identity連結的改良版RepConv（RepConvN）來設計網路的架構。

圖 YOLOv7說明Identity連結破壞殘差與串連架構

3.1.2 新的動態標籤分配策略（Dynamic label assignment strategy）

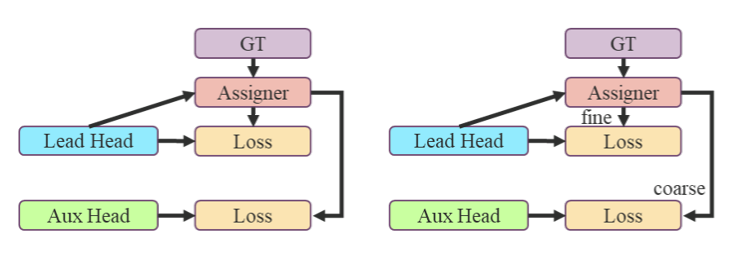
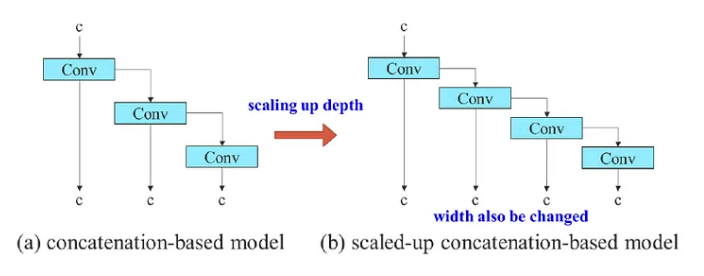
作者討論了在分類上提出兩種新的輔助頭跟主導頭分配軟標籤的做法，如圖，（一）Lead head guided label assigner：由於主導頭有較強的學習能力，因此讓主導頭預測結果與物體真實標籤最佳化運算得出軟標籤，再將此軟標籤作為輔助頭與主導頭的真實標籤進行學習，讓較淺的輔助頭直接學習主導頭所學到的資訊，而主導頭則是更關注於未學到的殘差資訊；（二）Coarse-to-fine lead head guided label assigner：此方法是將上述方法一最佳化運算後得出的軟標籤再分為兩類Coarse label以及Fine label，在Coarse label上放寬對正樣本的限制，讓YOLOv7中更多網格視為正樣本，並將此標籤用於學習能力較差的輔助頭上，而後者Fine label則用於主導頭，此外，在優化的任務上優先優化輔助頭的召回率，而主導頭的輸出會從輔助頭中高召回綠的結果中篩選出高準確率的作為模型最後輸出。

圖 主導頭與輔助頭之動態標籤分配示意圖

3.1.3 模型縮放與擴展（Compound model scaling）

在之前有提及模型縮放的論文中，如EfficientNet與Scaled-YOLOv4各自基於PlainNet與ResNet 類型的架構上進行縮放時，每層之間的輸入輸出量並不會改變，因此能夠將圖像大小、模型層數與通道數量這些常見的縮放因子與參數量、計算量之間的影響獨立分析。作者進一步分析分析在串連基礎的模型下並不能直接套用上述兩種模型的縮放方法，由於對於深度進行縮放將影響到後續計算模組（Computational block）的過渡層（Transition layer）的輸入，如下圖所示；此外也提出了相對應的解決方法，在深度進行縮放完後，計算計算模組（Computational block）輸出通道的變化量，並以該變化量對過渡層（Transition layer）的寬度進行相對應變化量之縮放，如下圖。

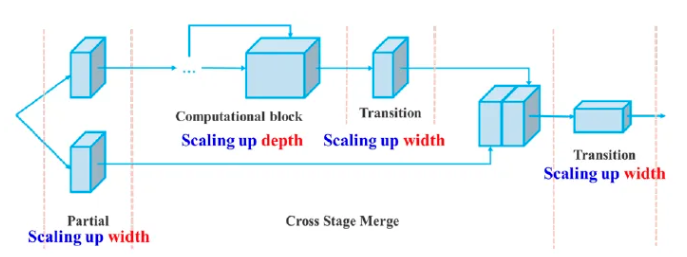
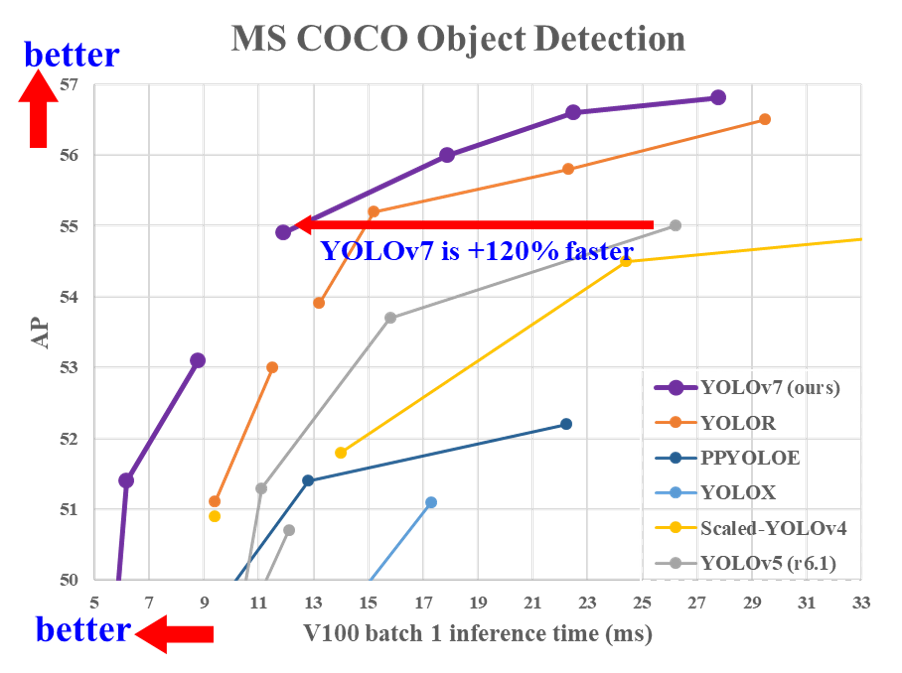
圖 Computational block縮放後導致Transition layer輸入通道寬度縮放示意圖

圖 YOLOv7所提出之模型模組縮放解決方法

3.1.4 更快的速度與更高的準確率

因為以上的修改以及優化使得YOLOv7與之前的最頂尖（State of the Art, SOTA）的即時物件偵測模型相比降低了40%參數量、50%每秒浮點運算次數（FLOPs），並有更快的推理運算速度及準確率。此外，所有模型皆非轉移學習，如圖，YOLOv7對比其他YOLO系列在精度以及速度上都得到了不小的提升，可說是擁有非常穩健的表現。

圖 YOLOv7與其他YOLO系列之速度與精度比較

**3.2 角度偵測**

3.2.1 旋轉邊界框之角度定義

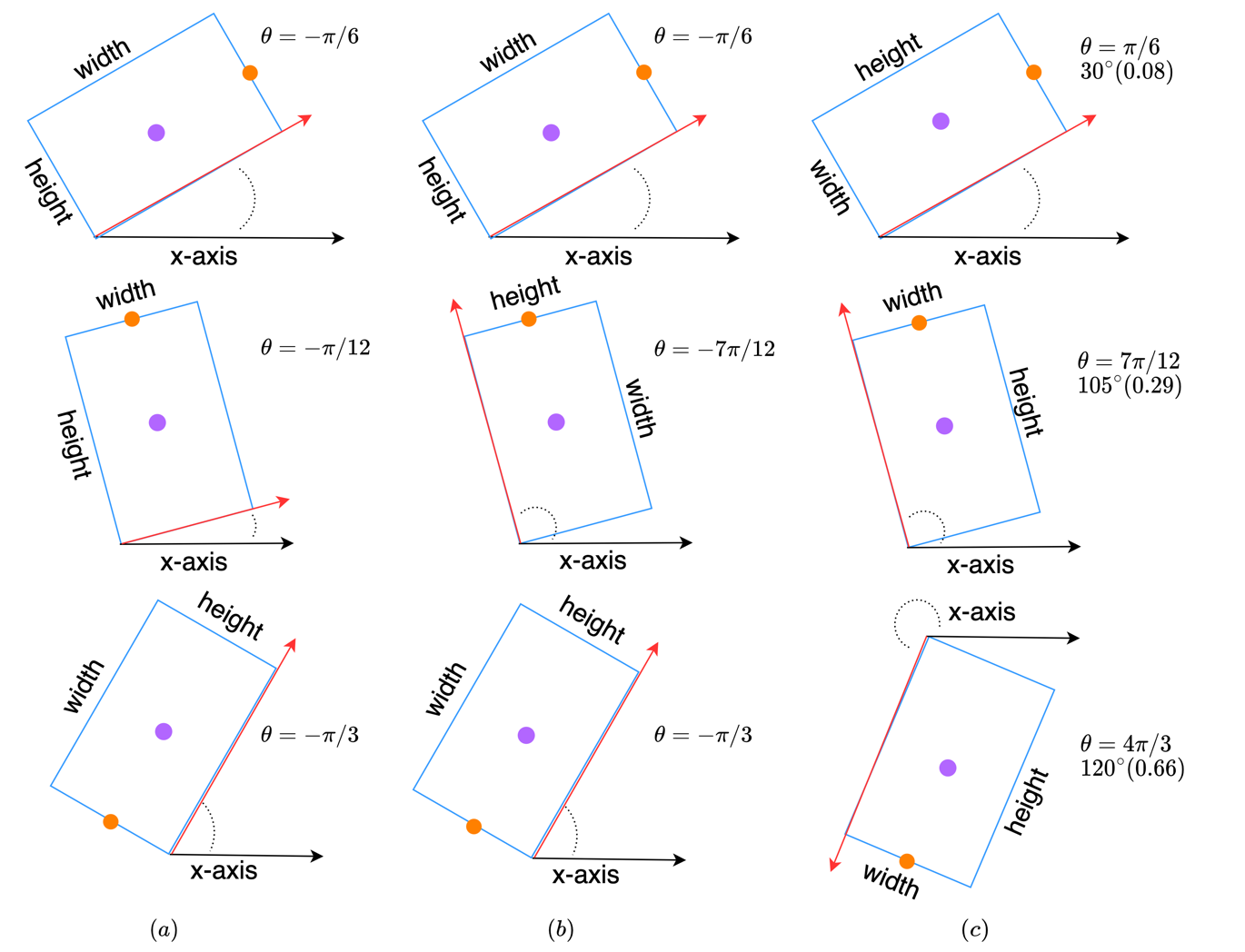
除了我們自有的十字路口空拍資料集外，HRSC2016與DOTA都有各自的標籤格式，為了使這些資料集能順利輸入進我們所修改的模型訓練，我們將資料集原先物體的四個角點的座標格式轉換為中心點座標（x, y）、寬度（w）、高度（h）與角度（θ）等五個參數的YOLO格式，在後面三者的值域定義上又因認知上的不同，分為OpenCV表示法與長邊表示法，上述兩種表示法雖然可以明確表達物體的角度，但為了能表示像十字路口車流等動態物體方向性，我們在角度（θ）的值域上提出如圖，以X軸正向為0，逆時針為正，最大值360度為1的標準化角度表示法，以達到在預測時及時看出物體的方向性。

圖 不同物件角度之定義與本論文對角度之定義

3.2.2 角度偵測頭

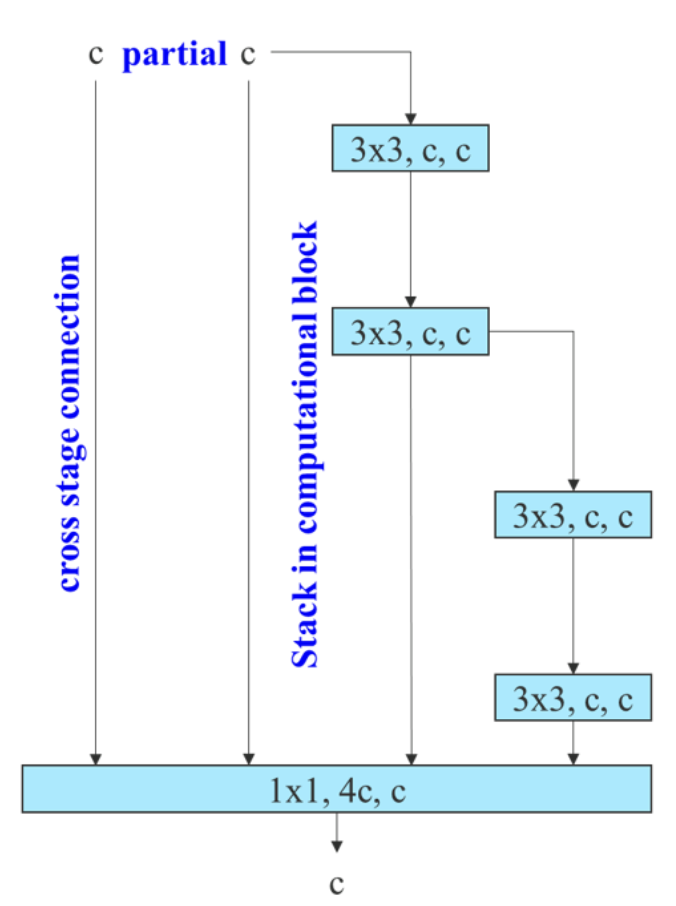
YOLOv7本身在最後的偵測頭部分，原先只有迴歸物件中心點（x, y）、物件長寬（w, h）四個值，而YOLOv7有別於以往的YOLO系列模型，加入了新的網路架構ELAN，對模型的深度進行縮放，如圖，同時為了維持深度縮放完的輸入輸出的大小，計算深度縮放的變化量並對寬度進行同樣的處理，進而控制最短與最長梯度路徑，使更深的網絡可以有效地學習和收斂。

圖 YOLOv7提出之ELAN架構

而本論文所提出之額外的角度數值預測上，我們在模型最後的預測層中另外加入了一個角度分支偵測頭，去另外對於物體角度進行預測，如圖，在模型最後預測輸出中，三種尺度下的特徵圖中，每個網格將負責預測一個邊界框，並附帶（物體類別數 C+4+1+1）個參數，如圖，而邊界框的位置由框的中心點（x, y）與框的寬高（w, h）所組成，並包含該邊界框所匡列的物體之置信度與分類，最後還有我們所加入之角度預測值，代表物體的旋轉角度。

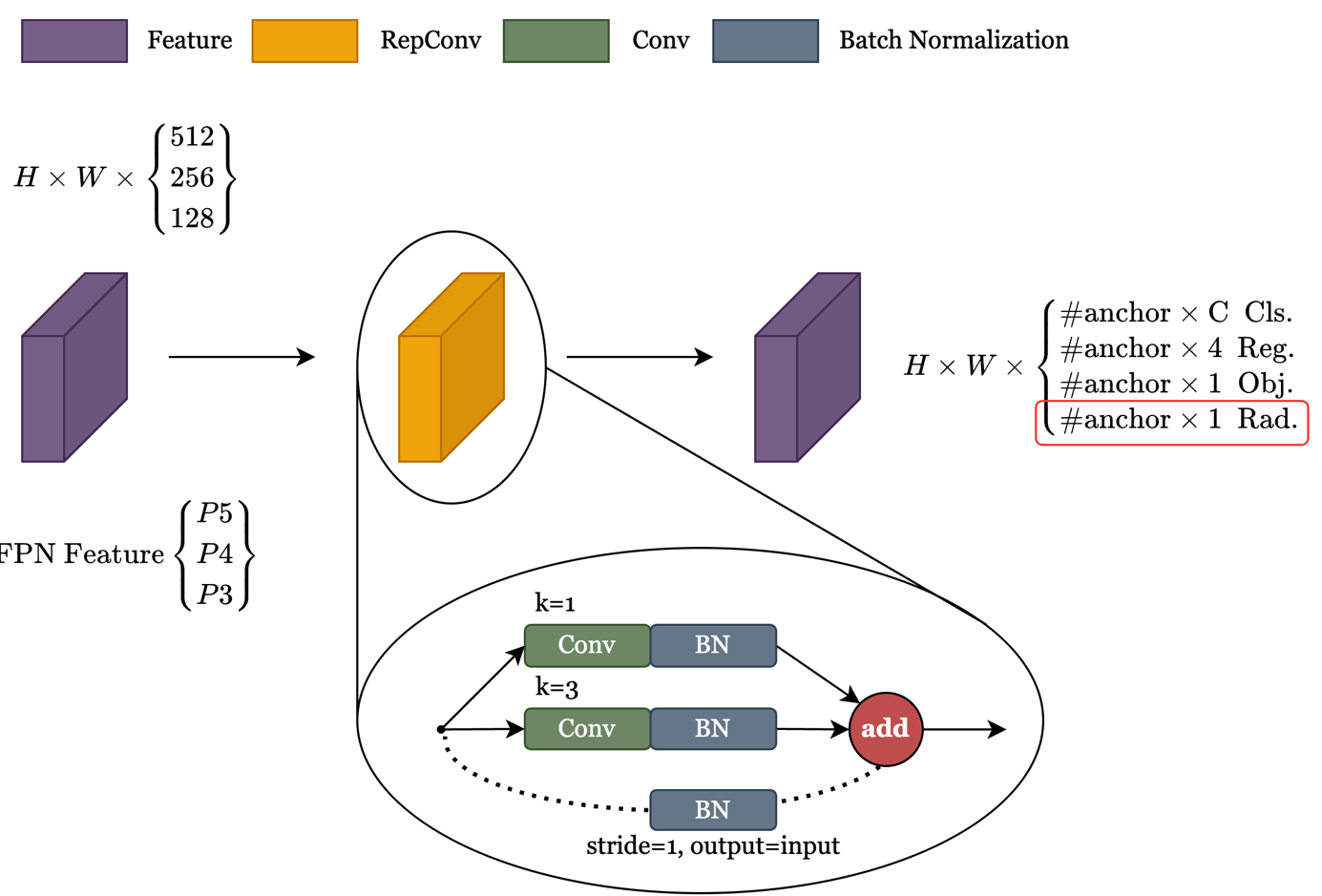
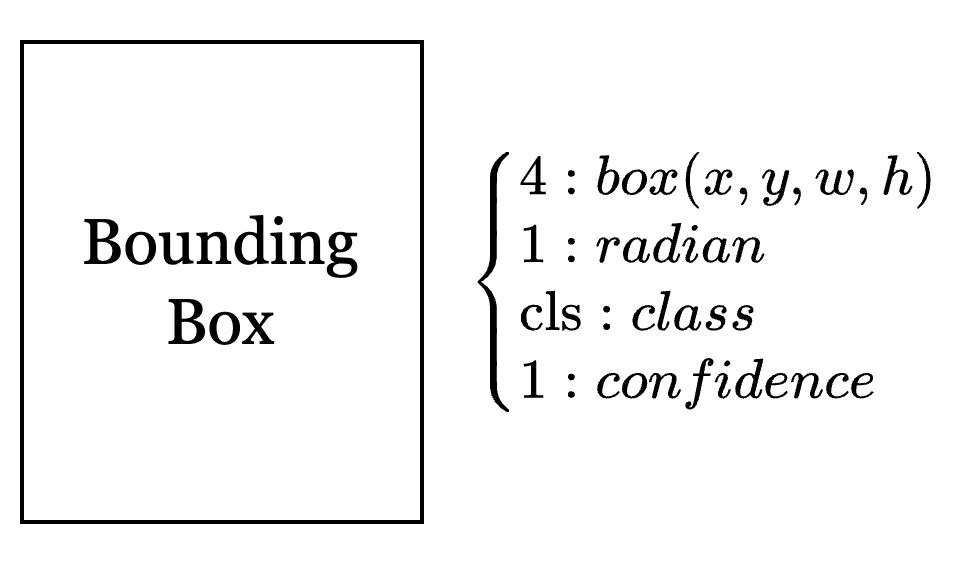
圖 YOLOv7偵測頭以及我們所加入之角度偵測頭

圖 邊界框預測資訊

**3.3 數據增強**

3.3.1馬賽克（Mosaic Data Augmentation）

馬賽克數據增強由Alexey Bochkoviskiy等人於2020年提出，參考自2019年Sangdoo Yun等人提出之CutMix數據增強，在兩張拼貼的影像基礎上，如下圖採取四張影像進行拼貼，並進行隨機的縮放、剪裁、平移與傾斜等處理過後，生成圖，圖像雖機處理完後，再針對影像標註檔進行相對應的更新，如上述所說，在對資料集經過多重隨機加工後，大幅提升資料集的豐富性與數量；值得一提的是，在做縮放的同時也會將小物件放大，使得這些物件得以被模型學習，進而提升網路模型的性能。不過此數據增強本身建立於水平邊界框的基礎，對影像與標註檔的轉換，都是以水平邊界框的座標進行處理，而本篇論文的資料集均是採用旋轉擬合框，存在物件偏斜角度，所以我們也針對此部分進行程式的修改，使的旋轉物體在進行馬賽克數據增強過後，影像本身與標註檔依舊能維持對應的座標。

3.3.2翻轉（Flip up-down, Flip left-right Data Augmentation）

針對上段馬賽克數據增強過後的影像，將再進行隨機的水平翻轉（圖）或垂直翻轉（圖），也因為旋轉邊界框跟水平邊界框在數值上轉換方式不同，我們也針對角度的數值上，在數據增強後，對角度值本身進行相對應的更新。

3.3.3旋轉（Rotation Data Augmentation）

在旋轉物體的資料增強上，我們也對了訓練資料集的影像進行任意旋轉，如圖，將原先的物件中點（x, y），寬（w），高（h），角度（radian）轉換為物件的四個相對應角點，再透過OpenCV的函式以旋轉角度取的四個新的角點座標，最後在轉換回原先的標註檔格式，取的新的物件終點、寬高與角度。

**3.4 基於YOLOv7之空拍單階段旋轉邊界框物體偵測器**

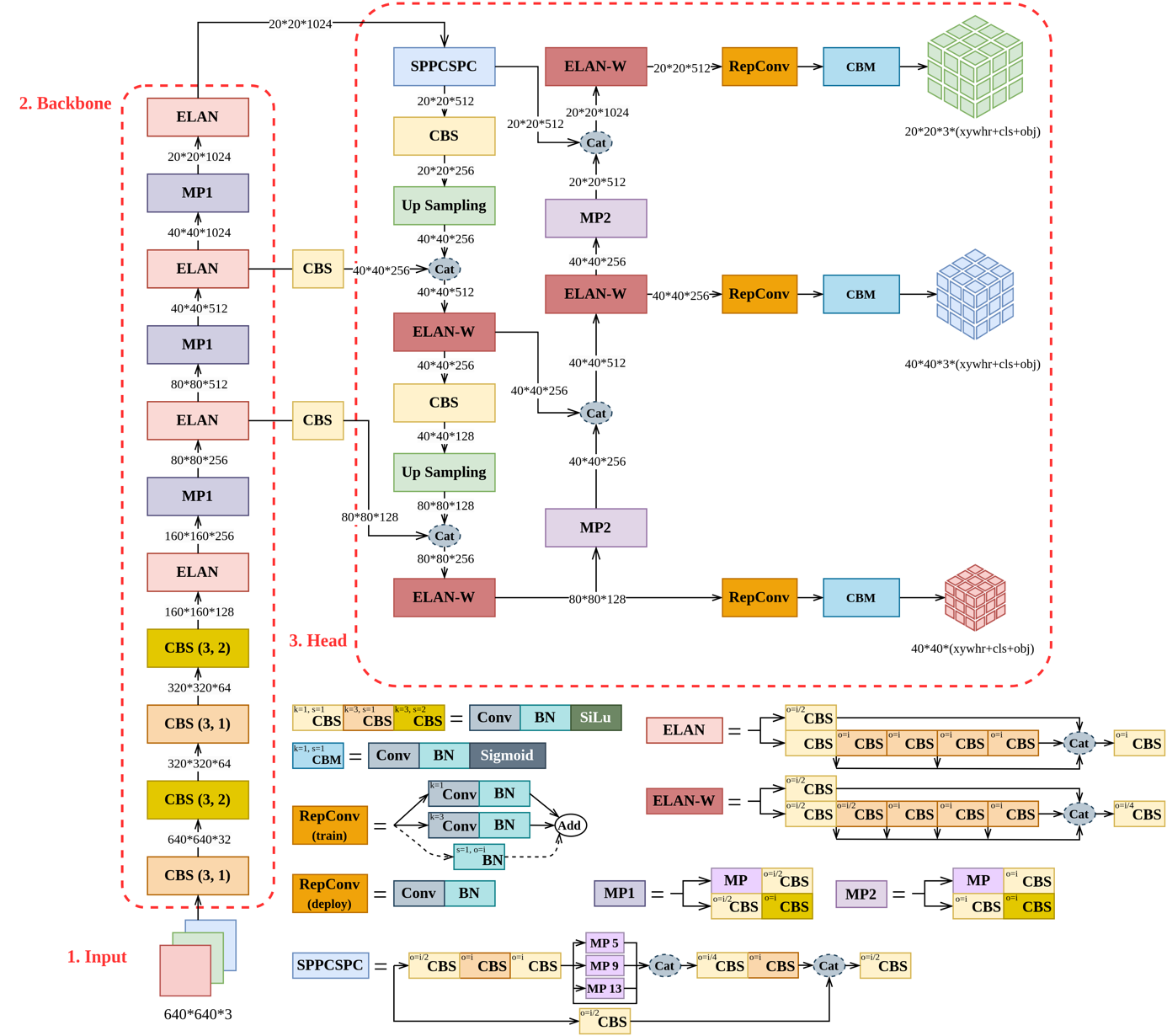
本篇論文旨在基於YOLOv7此一無錨框單階段物體偵測方法，使模型輸出具備預測物體角度的能力。由於YOLOv7本身訓練採用的為常見的水平邊界框，但為了滿足具有方向性的空拍影像旋轉邊界框之需求，將水平邊界框轉為旋轉擬合框進而減少模型學習到的冗余資訊，於以上的基礎下，我們提出基於YOLOv7模型的旋轉擬合框物體偵測方法，以所有的資料集訓練模型時，我們將訓練資料輸入進模型時的解析度定為640×640，並如下圖，經過骨幹網路的特徵提取後，取得三種不同尺度的特徵圖，之後將上層特徵圖的特徵上採樣融合至下層，並在原先的特徵金字塔網路（FPN）上引入路徑聚合（Path Aggregation），對下層特徵圖進行下取樣融合至上層，以此結合下層精確的物體位置和上層豐富的語意特徵等特性，藉以提升整體模型的性能。

圖 本論文所使用之YOLOv7網路架構圖

最後提及每個神經網路架構中極其重要的角色：損失函數（Loss Function），左右模型的收斂速度與梯度方向的正確與否，而我們在物件邊界框上使用的是交聯比損失函數（Intersection over Union Loss, IoULoss）；物體置信度與物體分類則是使用包含Sigmoid數字標準化的二元交叉熵損失函數（Binary Cross Entropy Loss, BCELoss），分別為底下的式（1）、式（2）與式（3），二元交叉熵損失函數大部分用來評估二分類的問題；在角度值的損失計算上，我們選用平滑平均絕對誤差函數（Smooth L1 Loss, SL1Loss），在平均絕對誤差與平均方差函數的基礎下，讓離群值的梯度影響力縮小，避免梯度爆炸，也使得梯度在原點可導，修正平均絕對誤差在原點之不平滑問題。

( 1 )

( 2 )

( 3 )

( 4 )

( 5 )

第四章 實驗結果與分析

**4.1 實驗環境**

本實驗使用一個Docker的Container作為實驗環境，硬體上Host主機為搭載AMD Ryzen 9 5950X 3.4GHz 16核 ，RAM為128G，顯卡為一張NVIDIA GeForce RTX 3090 24G，Container創建時設定為CPU6核，RAM 32G：軟體上Python版本為3.10.9，PyTorch版本為2.0.0，CUDA版本為11.4。

**4.2 實驗資料**

本論文在訓練以及驗證上使用公開的HRSC2016[]以及DOTA[]兩大具代表性的空拍資料集，並另外加入實驗室獨有的台灣縣市十字路口空拍車流的資料集增加比對性，HRSC影像大小介於300x300到1500x900像素，物件類別都是大型船隻；DOTA則是介於800x800到20,000x20,000像素，包含了各種長寬比、方向以及形狀的物體，分為15個常見類別。在全數2,806張影像中一共包含188,282個物體。（放資料可視化的結果）；實驗室獨有的十字路口資料集則是1080p的畫質（1920x1080），類別包含道路上常見的交通工具類別。

另外在DOTA資料集上，因為解析度過大，若直接輸入模型進行訓練，勢必會超越顯示卡記憶體的容量；所以事先針對資料集進行內插的前處理動作，切割為1024x1024，水平與垂直方向重疊部分取200像素，確保有一定的重疊區域，最後在輸入進模型做訓練時，都會將訓練圖像縮放至640x640進行正式訓練。

**4.3 實驗結果**

**第五章 結論與未來方向**

雖然說傳統的旋轉邊界框偵測器

**參考文獻**