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 YOLOv7-based Radian Object Detection

指導教授：蘇志文

研 究 生：林冠良

中華民國 112 年 O 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](#_Toc133337600)

[Abstract II](#_Toc133337601)

[致謝 III](#_Toc133337602)

[第一章 緒論 1](#_Toc133337603)

[1.1 研究動機 1](#_Toc133337604)

[1.2 論文架構 2](#_Toc133337605)

[第二章 相關文獻 3](#_Toc133337606)

[第三章 研究方法 4](#_Toc133337607)

[第四章 實驗結果與分析 5](#_Toc133337608)

[4.1 實驗環境 5](#_Toc133337609)

[4.2 實驗資料 5](#_Toc133337610)

[參考文獻 6](#_Toc133337611)

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，包含單／二階段的物體檢測模型架構、有／無錨框之檢測方法在機制上的差異性以及優缺點以及基於迴歸／分類的角度檢測方法等。
* 第三章節闡述本論文之研究方法，並且詳細說明如何加入旋轉角度之偵測頭、替新增的角度參數加入損失函數以及最後增加旋轉框ＮＭＳ的後處理方法。
* 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

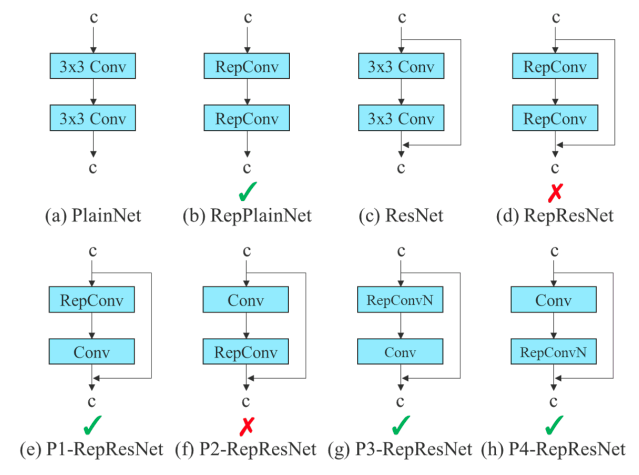
第三章 研究方法

於此章節，我們將詳細說明本論文所提出之方法，在角度偵測的方面，大致上分成迴歸與分類兩種處理方式，存在著彼此的優缺點，為了不使模型在預測層上過於厚重，我們採用迴歸的方式去處理角度的偵測，並選用本身已經有優異表現的YOLOv7，額外在預測層中加入旋轉角度的偵測頭，使模型多具備預測物體角度的能力。

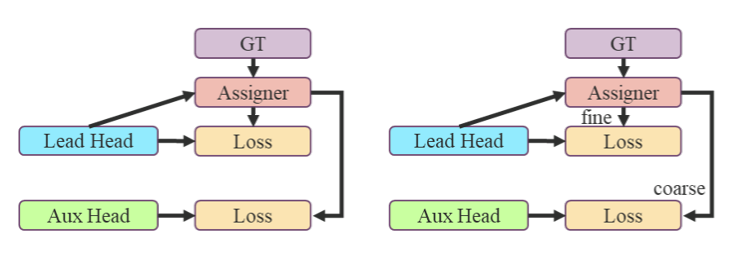
3.1 以YOLOv7為基礎

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：

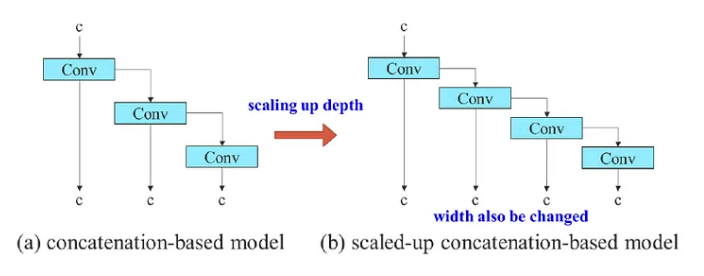
3.1.1 新的重參數化方法（Re-parameterized model）

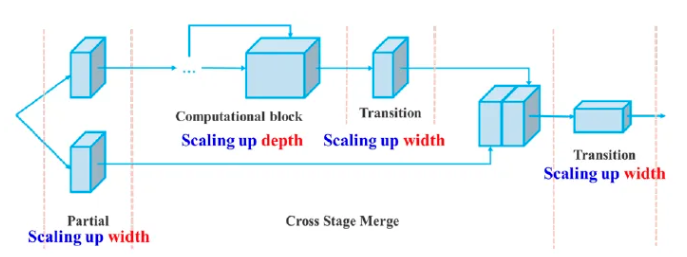
作者等人發現RepConv中的Identity連結會破壞ResNet中的殘差（Residual）與DenseNet中的串連（Concatenation），提出不使用Identity連結的改良版RepConv（RepConvN）來設計網路的架構。

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。

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Computational block的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Computational block輸出通道的變化量，並以該變化量對Transition layer的寬度進行相對應變化量之縮放，如下圖。



3.2

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