



國立成功大學工程科學系

專題成果報告

Department of Engineering Science

National Cheng Kung University of Engineering

Practicum of Engineering Science

運用深度學習建置夜間路面坑洞辨識模型之探討與實作

學生：林以諾、鄭仲勛

指導教授：陳牧言 博士

中華民國 112 年 9 月

目 錄

第一章	緒論-----	1
第一節	研究動機-----	1
第二節	研究目的-----	1
第二章	文獻探討-----	3
第一節	Faster-RCNN-----	3
第二節	RetinaNet-----	4
第三章	研究方法-----	6
第四章	實驗設計和績效評估-----	8
第一節	實驗環境-----	8
第二節	資料集說明-----	9
第三節	參數設定-----	9
第四節	實驗結果-----	11
第五章	結論-----	15
第一節	研究貢獻-----	15
第二節	研究限制-----	15
參考文獻	-----	16

第一章 緒論

第一節 研究動機

在深度學習領域中，影像辨識已成為一個極為熱門的應用範疇。現今，越來越多領域正在積極採用影像辨識技術，其中包括汽車和醫學等多個領域。透過影像辨識，汽車能夠偵測路面上的坑洞，進而自動執行煞車或避開障礙，而在醫學領域，它能夠用於識別超音波圖像等應用。

在本專題研究中，目標是運用影像辨識技術，實現能夠準確辨識路面坑洞的解決方案。通過學習如何在黑夜中辨識路面坑洞，希望能夠建立一個系統，只需輸入具有坑洞的照片，即可精確識別坑洞的位置。期望是未來能夠在汽車上應用這項技術，使汽車能夠根據路面坑洞的情況及時調整其避震器和其他相關設定，提高行車的舒適性和安全性。

第二節 研究目的

本專題將整合多項相關技術，包括「深度學習訓練中的影像辨識」、「資料集的預處理和擴增」、「資料集的切割」以及「圖片的標註」等，以實現能夠準確識別圖片中坑洞位置的成果。

並利用上述技術，研究以下主題：

- 訓練出能夠判斷夜間路面坑洞的學習模型：

研究訓練數據之處理過程、訓練模型結構和訓練過程。

- 使用訓練好的模型，辨認非資料集中的路面情況：

驗證模型泛化能力，避免模型應用範圍侷限在訓練樣本標籤內，導致未來應用上需花費大量成本蒐集樣本。

- 根據研究結果，探討one-stage和two-stage之使用情境

透過閱讀相關文獻和實際實驗結果，評估不同模型間的適用場景。

本專題希望藉由以上技術的結合和研究，協助日後在各領域的實作及應用，如自駕車、政府路面坑洞檢修系統、盲人輔助裝置等等，以達成安全且實用的智慧運用。

第二章 文獻探討

這次的專題主要會使用到二項技術，分別是Faster-RCNN、RetinaNet。

第一節 Faster-RCNN

Faster-RCNN相較於早期的RCNN，引進了RPN(Region Proposals Network)的技術，突破了過往因區域建議的生成(region proposal)和模型預測步驟分開的關係，將兩者整合進了網路架構中，讓檢測的速度比以往快上不少，且mAP值也沒有因此降低，以下詳細講述Faster-RCNN的流程[1]：

1. 卷積神經網路(Convolutional Neural Network, CNN)

輸入圖像後會首先經過卷積神經網路進行特徵提取，會得到feature map，特徵圖保留了圖像的抽象特徵，可以用來進行物體檢測和其他相關任務。

2. RPN

RPN通過比對錨點（anchor）和實際物體位置的偏移量（offsets）來判定某個區域是否可能包含物體。接著根據這些機率對候選錨點進行排序，通常會選擇機率最高的前300個anchor box用於後續處理步驟。

3. ROI Pooling

RPN生成的候選區域通常為不同尺寸和大小的比例，需將不同尺寸的區域提議映射到相同大小的特徵，以便進行後續的物體檢測任務。這樣的操作有助於保持區域的空間結構並提取有關物體的特徵。

4. 分類器(Detector)

將固定大小的特徵圖物件分類，並將物件類別機率和定位訊息產生出來，在疆域測出的Bounding Box(BBox)和實際區域進行回歸計算，對其進行最佳化。

5. NMS(Non-maximum suppression)

比較候選區域間的重疊度並過濾掉多餘的BoundingBox(BBox)，讓物件檢測的結果準確值更高。

第二節 RetinaNet

RetinaNet為一種密集檢測器，主要用於檢測由Facebook AI Research所發表的論文"Focal Loss for Dense Object Detection" [2] 中的 Focal Loss之效果。

Focal Loss為RetinaNet中的核心技術，用於解決進行目標檢測任務時遇到的類別不平衡所導致模型在訓練時更關注於出現頻率高的類別，而忽略了出現頻率低的類別之情況。其引入了「調節因子」並透過調節因子與交叉熵損失相乘，使在處理難以分類的樣本時，損失值更加集中，進而使模型更專注於解決類別不平衡的問題，提高少數類別的分類精度。

RetinaNet的架構為：

1. 殘差網路(Residual Network)

Residual Network(ResNet)廣泛的運用在許多架構上，是一種強大的特徵萃取器；透過將欲學習的目標函數 $H(x)$ 與 x 相減得到殘差，並優化學習這些殘差，相較之下比學習整個 $H(x)$ 更容易成功，並透過快捷連接使結果近似 $H(x)$ ，達到簡化的目的。

2. 特徵金字塔(Feature Pyramid Network)

在 RetinaNet 裡，使用 FPN 架構，建立了多層金字塔特徵結構，每層有不同尺度的特徵。這讓模型可以更好地處理不同大小的物體。RetinaNet 進一步優化了 FPN，讓每個偵測層能運用這些不同尺度的特徵，更有效地偵測不同大小的物體，提升了偵測效果。

3. 類別子網路(Class Subnet)

這種模型架構用於物體偵測。輸入特徵圖有256通道，每個位置預測 K 個錨的類別機率。使用 4 層 3×3 卷積網絡提取特徵，然後經過類別預測層生成每個錨的機率。這些機率通過 Sigmoid 函數處理，用於判定每個類別的存在。

4. 框子網路(Box Subnet)

這子網路用於預測物體框位置修正，類似類別子網路。每位置輸出4A個線性值，對應錨的數量。它修正預測框相對於實際框的位置，提高偵測的精確度。

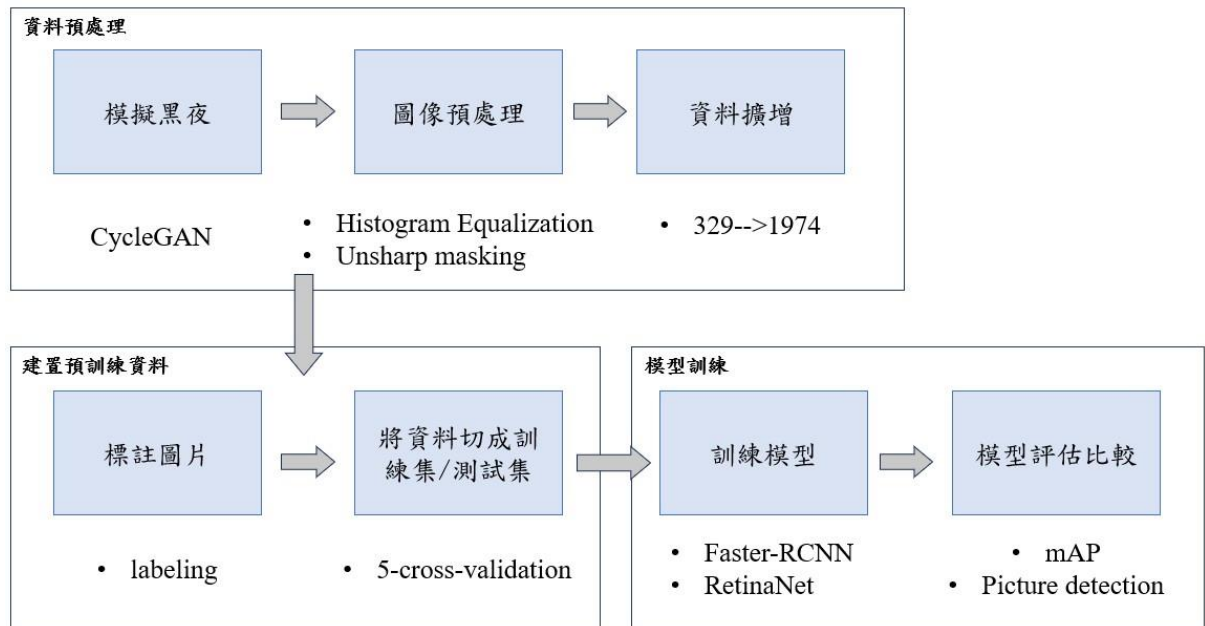
5. Anchors(錨點)

根據金字塔層次調整anchor大小，有多個不同比例和縮放的anchors，每層共9個anchors。

每個anchor都有分類和框回歸目標，根據IOU標記前景和背景，框回歸目標計算anchor與實際框的偏移。這樣的設計使得模型能夠準確地偵測不同尺寸和形狀的物體。

第三章 研究方法

下圖為本次專題之實作流程：



圖一、本專題之研究流程

1. 資料預處理

本專題使用了Kaggle數據平台提供的Pothole Image Dataset，其中包含329張白天路面坑洞的照片。然而，本專題的研究重點是在黑夜中辨識路面坑洞，因此，

- 由於網路上的坑洞照片，以白天為主，為了蒐集大量資料以利模型訓練，本專題首先利用CycleGAN將這些照片轉換成模擬黑夜環境的狀況。
- 為了提高後續辨識過程的順暢度和準確性，本專題採用了Histogram Equalization技術，以減輕照片的過曝問題。接著，進一步使用Unsharp Masking技術來增強坑洞的邊緣，使坑洞的特徵更加突出。
- 為了改善模型的性能，本專題採取了資料擴增的方法，包括翻轉和旋轉，將資料集擴充至1974張照片。

2. 建置預訓練資料

- d. 在開始訓練之前，對照片進行了手動標註，框出了坑洞的位置，以建立後續訓練和模型評估的基準。
- e. 最後，本專題透過5折交叉驗證，將資料集分為訓練集和測試集，比例為8:2，以確保模型的穩健性和準確性。

3. 模型訓練

- f. 為了深入研究在物件偵測任務中使用不同架構的差異，本專題選擇了RetinaNet（one-stage object detection）和Faster-RCNN（two-stage object detection）兩種模型作為本專題的訓練模型。

本專題旨在比較這兩種不同的物件偵測架構，以評估它們在任務中的性能差異。本專題使用Python語言在PyCharm集成開發環境中撰寫了程式碼，使用相同的訓練資料集進行了實驗，並選擇了相同的ResNet-50骨幹網絡以確保兩個模型的基準環境一致。這將有助於了解哪種方法最適合解決物件偵測問題，是one-stage還是two-stage的方法。

- g. 在模型訓練完成後，將進行照片測試的階段。本專題將使用兩個評估指標來評估模型的性能，分別是mAP（平均精確度均值）值和非Kaggle提供的坑洞照片進行實際預測。

首先，使用mAP值來評估模型在測試集上的性能。mAP是一個常用的物件偵測性能指標，它考慮了模型的精確度和召回率，通常用來評估物件偵測算法的準確性。較高的mAP值表示模型在偵測物件方面表現較好。

其次，下載非Kaggle提供的坑洞照片，進行模型的實際預測。這將有助於了解模型在實際場景中的表現，特別關注坑洞的複雜性、單一坑洞辨識的準確度以及在白天狀況下的表現。

經過以上步驟，可以深入探討這兩個評估指標之間的關聯性，並評估模型的整體性能。這些實驗結果將有助於更好地理解模型的優勢和限制，並可能提供改進模型的方向或調整訓練策略的建議。最終，本專題的目標是提高坑洞偵測的準確性和實用性。

第四章 實驗設計和績效評估

第一節 實驗環境

本專題之實驗環境主要是在自身筆電並透過Pycharm進行，用來訓練模型之GPU為NVIDIA GeForce GTX 1050，虛擬機之RAM為15.9GB，磁碟空間為301GB。詳細硬體規格如表一。

硬體名稱	硬體規格
GPU	NVIDIA GeForce GTX 1050
RAM	15.9GB
Disk	301GB

表一、本專題使用之硬體規格

程式撰寫的部分，本專題所使用之語言為Python，使用keras在tensorflow之上來建立深度學習模型，裡頭透過pillow來進行圖像的前置處理，透過matplotlib將完成後的結果視覺化，詳細套件版本如表二。

套件名稱	套件版本
Python	3.6.4
TensorFlow	2.2.0
matplotlib	3.5.3
keras	2.7.0

Pillow	8.4.0
--------	-------

表二、本專題使用之套件版本

第二節 資料集說明

本專題使用了Kaggle數據平台提供的Pothehole Image Dataset，並加以使用CycleGAN工具將這些照片轉換成夜晚的模擬版本。為了提高模型的性能，對圖片進行了預處理，包括使用Histogram Equalization技術降低曝光問題，以及Unsharp Masking技術增強坑洞的特徵。為了擴充訓練資料，本專題對圖片進行了翻轉和縮放，從而將原始資料集中的329張圖片增加到1974張。最後，手動標記了坑洞的位置，並將資料集分成訓練集和測試集，比例為8:2，以確保模型的有效訓練和評估。

第三節 參數設定

1. Faster-RCNN模型訓練參數設定[3]：

backbone	ResNet50
anchor size	[128,256,512]
freeze epoch	50
total epoch	100
batch size	4(in freeze) 2(in unfreeze)

Optimizer	Adam (initial Rate:1e-4)
loss	LossHistory

Faster-RCNN預測參數設定：

backbone	resnet50
confidence	0.5
nms_iou	0.3
anchors_size	[128, 256, 512]

Faster-RCNN訓練過程參數：

權重的更新會在「解凍」時調整，此使被允許進行微調更新，本專題設定每五個epoch保存一次權值，訓練一開始(設定前50個epoch)採以RESNET50為主幹特徵提取網絡，此時為凍結，以確保一開始是穩定的進行訓練，後續的Epoch再採取解凍，解動前後batch_size也從4降低為2，調整特徵值以完成所需工作，其中學習率透過optimizer來自動調整

2. RetinaNet模型訓練參數設定[4]：

backbone	ResNet50
anchor size	[32,64,128,256,512]
epoch	100
batch size	25
steps	1263

Optimizer	clipnorm(default=0.001)
loss	Focal Loss
alpha in Focal Loss	0.25
gamma in Focal Loss	2.0

經由論文作者測試Focal Loss效果後，發現alpha為0.25及gamma為2.0時，雖然整體loss總量減少，但有更多的注意力放到了數量較少的類別上；故在此直接引用論文作者之Focal Loss參數。

第四節 實驗結果

以下針對Faster-RCNN(two-stage object detection)、RetinaNet(one-stage object detection)的訓練結果相互比較。

1. 非kaggle照片集實際預測結果比較

以下的圖左邊皆為RetinaNet(藍色框)，右邊皆為Faster-RCNN(紅色框)

- a. 實際將訓練後的模型拿來預測網路上找到的各種夜間坑洞圖片，可以發現大部份狀況下都能準確預測，如下圖所示，但像圖中右側燈光明顯不足的情況下，模型在判斷時仍然較為吃力，無法準確判斷出坑洞。



- b. 當考慮到坑洞數量較多的情況時，通常指的是在影像中存在大量小型或密集的目標物體，例如，許多小型物件或細節需要被檢測和識別。在這種情況下，Faster-RCNN的物體檢測性能通常較佳，因為它能夠更有效地處理多個候選區域，並且在複雜場景中提供更精確的物體定位和識別。透過下圖的比較就可以明顯看出其差異。



- c. 本專題普遍觀察到當目標物體相對簡單的圖像中，RetinaNet呈現出更低的誤報率，而同時RetinaNet在mAP值的表現較為優異，進而驗證高mAP值通常代表著一個物體偵測模型在綜合表現方面相對優秀，它能夠有效地減少誤報率，提高對實際目標的準確性。

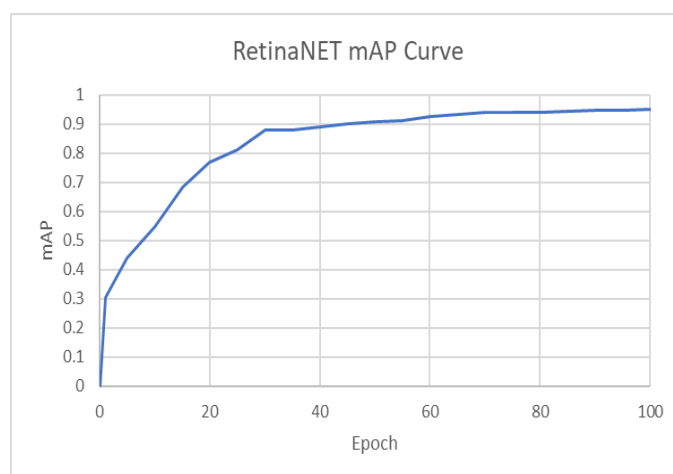


2. mAP比較

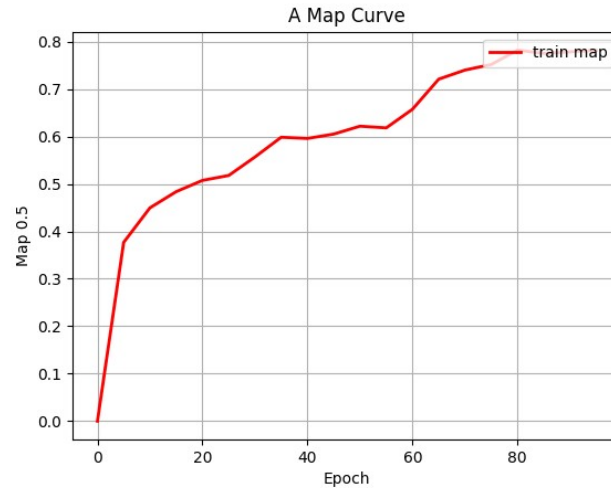
mAP(mean Average Precision)是一個綜合性能指標，結合了精確度（Precision）和召回率（Recall）。當mAP值高時，通常表示模型能夠實現高精確度和高召回率。高精確度表示模型對其檢測結果的信心很高，並且很少誤判。高召回率表示模型能夠找到大多數實際存在的目標物體，不會遺漏該判斷的目標物件。

Epoch數	RetinaNet	Faster-RCNN
5	0.3036	0.3768
50	0.8780	0.6216
100	0.9508	0.7815

表三、兩模型選出當5、50、100個epoch數時為代表，所呈現之mAP值



圖二、RetinaNet mAP curve



圖三、Faster-RCNN mAP curve

從訓練後的圖表(圖二、圖三)中可以明顯看出，在這次的研究中，RetinaNet在訓練速度和最終的mAP值方面表現出更出色的成績。

儘管根據論文，one-stage物件偵測模型(RetinaNet)通常將物件偵測和物件定位合併成一個單一的過程，這種方法具有更快的訓練速度。相對地，two-stage物件偵測模型(Faster-RCNN)將物件偵測分為兩個階段：首先生成候選區域，然後進行物件分類和定位。這種方法通常在精確性方面表現更優，但它的推理速度較慢。

然而，猜測造成此次RetinaNet在最終mAP值表現優於Faster-RCNN的原因可能與其使用的Focal Loss有關，這使得其在一階段物件檢測模型中的表現不像以往一樣遠遜於二階段模型，也是RetinaNet在one-stage object detection中脫穎而出。

第五章 結論

第一節 研究貢獻

- 針對夜間坑洞辨識之訓練模型，本專題之研究成果精確度普遍為佳，對坑洞圖像之處理、模型結構以及訓練過程可以給予參考。
- 避免了訓練資料限制，利用同為黑夜路面坑洞的圖片，卻非訓練資料的圖像進行路面坑洞辨識，辨識效果與辨識訓練資料之成果接近，驗證了模型高度的泛化能力，能夠有效的適應新樣本。
- 透過任務需求，選擇適合的預測模型結構。當圖像簡單且需要快速獲得訓練結果的話，可以選擇one-stage模型；而當圖像較為複雜需要精確判別的圖形，可以選擇two-stage模型。
- 夜間坑洞辨識模型可以應用在汽機車上，使駕駛者在低光環境中能準確辨別道路上的坑洞，減少事故的風險；同時也能同步調整避震器，使車輛能在更穩定的狀態下通過坑洞，進而使底盤、輪胎相關零件受到較小的損壞亦能提高駕駛的舒適度。
- 夜間坑洞辨識技術可以協助城市當局更迅速地辨識和修復道路上的坑洞，這有助於提升城市基礎設施的維護效率。該技術的模型能夠精確記錄坑洞的位置，並將相關信息傳遞給相關的當地政府機構，使其能夠更快速地做出應對措施，以改善城市道路的狀態和安全性。

第二節 研究限制

在這次的專題研究中，大部分情況下坑洞的預測算法相當成功。然而，這項研究僅限於靜態圖像的應用，無法用於影片上的坑洞檢測。這為未來與其他技術的整合提供了改進的空間。此外，由於某些圖像情境受到多種因素的干擾，以及用於訓練的坑洞資料集有限，這導致模型仍然無法完全準確地識別夜間坑洞等。因此，本專題認為在這方面仍然存在改進的潛力，例如尋找更適合訓練的資料集、將其他常見路面情境納入分類考慮，以及研究更精確識別坑洞影像的演算法等等。期望未來能實現更多感知模型，以提高坑洞識別的能力，無論是應用於自駕車或其他領域上的結合，都能為此做出貢獻。

參考文獻

- [1]Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun (Jan 2016), Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks
- [2] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár(Aug 2017), Focal Loss for Dense Object Detection
- [3]Faster R-CNN：在TensorFlow 2中實現的兩階段目標檢測模型（2022），檢自https://github.com/bubbliiiing/faster-rcnn-tf2?fbclid=IwAR23Ngxpg20wq7Uqeh97LrTlj9IJTsm7sMSCe7c_jnE3vCIEnvz8Tybydshk#%E8%AF%84%E4%BC%B0%E6%AD%A5%E9%A%A4
- [4]Keras-RetinaNet訓練自己的數據詳細教程
https://blog.csdn.net/sinat_35512245/article/details/88188365