

國立清華大學 電機工程學系

實作專題研究摘要

*Design of a Drone Detection and  
Tracking System Based on Machine  
Learning*

基於機器學習之無人機辨識與追蹤系  
統設計

專題領域： 通訊系統

組別： A529

指導教授： 黃衍介

姓名： 鄭絲予

研究期間： 113年 7月 1日至 114年 5月 1日止，共 10 個月

## **Abstract**

The goal of this project is to design a system capable of detecting and tracking drone in real time, based on the YOLOv11 image detection model. A two-axis gimbal is used to support a USB camera mounted on top, which captures live video input. The system utilizes this input to perform object detection and executes tracking functions accordingly.

The system is developed using Google Colab and PyTorch. A dataset containing over 10,000 drone images was used for training. After running a large number of epochs, I analyzed the loss metrics and evaluated the model's precision using real drone flight footage.

After the training is completed, the best-performing weight file (best.pt) is loaded into PyCharm, and the laptop is used to perform real-time inference. The system automatically outputs control commands to drive the gimbal, allowing it to rotate horizontally or vertically to follow the drone's position. This ensures that the drone remains within the camera's field of view, achieving a dynamic tracking effect.

After training for nearly 500 epochs, the model demonstrated a reasonably good level of detection accuracy. However, there is still room for improvement in the overall system architecture. During development, I also attempted to integrate a Raspberry Pi as the primary image processing unit, but due to limited computational resources and challenges in setting up the environment, the implementation was not successful. In addition, I experimented with replacing PyTorch with Darknet as the training framework, but was unable to achieve the desired results. These limitations represent challenges that I hope to overcome in the future. Nevertheless, the core objectives of drone detection and tracking have been successfully achieved in this project.

## 摘要

本專題的目的在設計無人飛行器的目標偵測與追蹤系統，主要是基於 YOLOv11 模型的影像辨識，利用一個雙軸雲台以及上面搭載的 USB 鏡頭達成及時影像的擷取、辨識與追蹤功能。

此系統的實作環境為 Google Colab 與 Pytorch，使用已具備的超過 10000 張的無人機影像資料庫去做訓練基礎，並分析經歷過多個 epochs 出來的 loss 值分析，並用實際拍攝的無人機飛行影片驗證模型準確度。

訓練完成後，會將整體訓練中表現最佳的權重 best.pt 匯入 PyCharm 中，以筆電本身去做及時推論，並利用 python 程式去判斷目標具體的座標位置，自動輸出對應的控制指令至 Arduino，驅動雲台左右或上下移動去跟進無人機的位置，使其不要超出視角，達到一個動態追蹤的效果。

訓練了將近 500 個 epochs 可以發現模型有還不錯的辨識程度，但整體架構還有可以改進的地方，途中亦嘗試結合 Raspberry Pi 作為影像處理主要的系統，但受限於運算資源還有環境的建設，沒有順利完成應用，也試過使用 Darknet 取代 Pytorch，但沒有成功實現，這些都是未來希望可以克服的，不過整體而言，已經達成無人機辨識及追蹤的目的。

# 1. Introduction

## 1.1 研究動機

隨著 AI 技術的快速發展，機器透過神經網路的架構模擬人類大腦的運作方式，逐漸具備類似人類思考與判斷的能力。其中，影像辨識技術在多項領域中已展現高度應用潛力，特別是在目標偵測與即時追蹤方面表現亮眼。另一方面，無人機在近年來廣泛應用於空拍攝影、國防偵查、環境監控等場景，具有極高的實用價值與發展潛力。若能結合 AI 影像辨識技術與雲台控制系統，實現無人機的自動追蹤，將可提升系統整體的自動化程度，並拓展其實際應用層面。因此，本研究希望以人工智慧為基礎，開發一套具備目標辨識與追蹤能力的簡易實驗平台，作為未來擴充應用的基礎。

## 1.2 研究目的

本研究旨在透過 YOLO (You Only Look Once) 深度學習演算法訓練出可辨識無人機的影像辨識模型，並進一步結合 Arduino 微控制器與伺服雲台實現目標追蹤功能。整體系統期望能夠即時偵測並鎖定畫面中的無人機目標，透過控制訊號驅動雲台進行自動追隨。最終可建立一套具備基本智慧辨識與反應能力的應用平台，為未來在智慧監控、目標導引與自動追蹤等領域的應用提供技術基礎。

## 2. Research Methodology

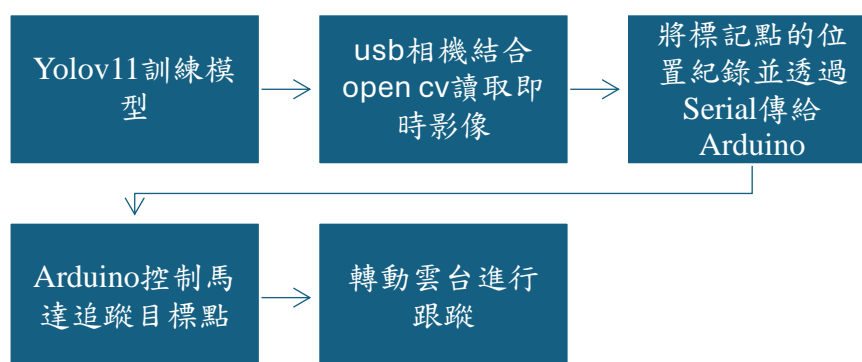
在這個計畫中，我使用 Google Colab，由於電腦效能不夠，因此使用 Google 提供的 GPU 去訓練，首先，在 Google Colab 上設置 PyTorch 訓練環境，並將無人機的圖片資料集上傳雲端，再從 Google Colab 掛載雲端即可將需要的資料都上傳。接著，要安裝必要的模組 Ultralytics，這樣就能直接使用 Ultralytics 官方的 YOLO 及時偵測模型，不需要安裝多個複雜的依賴像是 torch、opencv 等，接著便可以輸入指令開始訓練。

每一批 epochs 訓練完，都會檢測 loss 值變化，包含目標框回歸誤差(box loss)、類別分類誤差(cls loss)、物件信心分數誤差(obj loss)，如果發現 Loss 停滯甚至上升，要檢查資料以免過度訓練，也會監控過程適當的增加幾張圖片，並使用 CVAT 進行標籤，將這些資料合併原本的資料集下去訓練。

為實現即時追蹤，我使用了一個雙軸雲台控制系統(雲台由實驗室提供)，主要由兩個伺服馬達控制，一顆負責水平的旋轉，一顆負責俯仰的角度變化，接著將伺服馬達透過 PWM 角為連接至 Arduino 板，在將 Arduino 板接上筆電，即可在筆電的 PyCharm 跑訓練時發揮追蹤的功能。

完成模型與雲台控制整合後，進行實地的測試，使用 USB 攝影機取得即時影像，並以先前用 YOLOv11 訓練出的 best.pt 推論，接著計算無人機在畫面中的相對位置，轉換成角度變換再傳送指令給 Arduino 控制雲台追蹤無人機。若在實作中發現有追蹤角度不穩或者識別不良的情況再從程式上或訓練資料中進行調整改進。

運作流程：

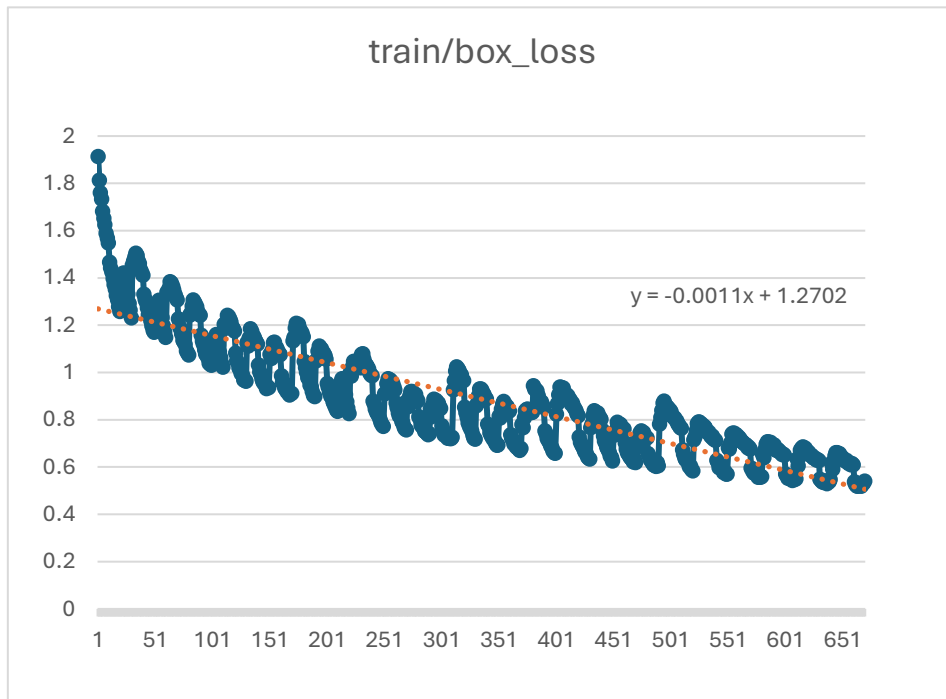


(Fig. 1、追蹤系統運作流程)

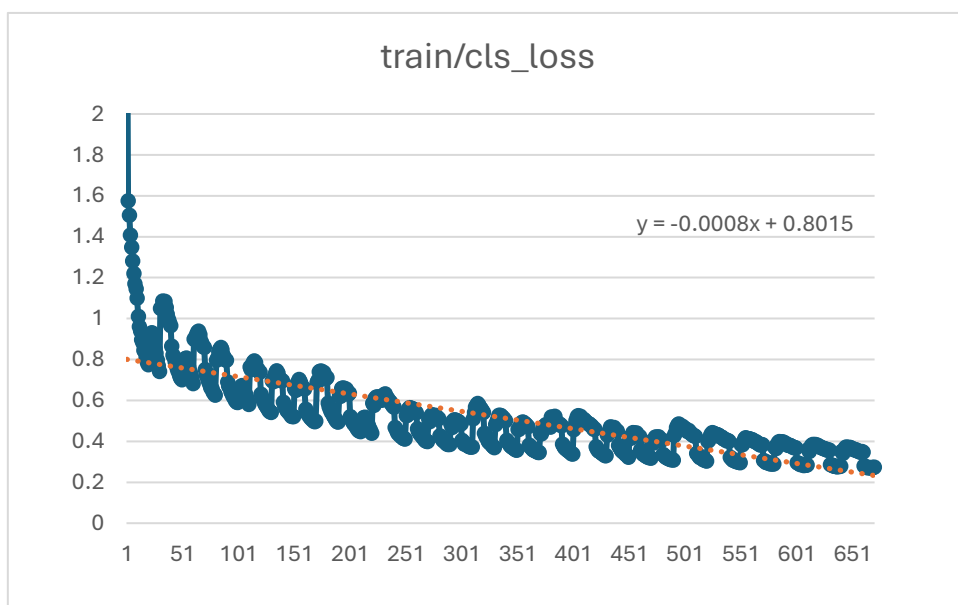
### 3. Experimental Results

#### 3.1 機器學習訓練

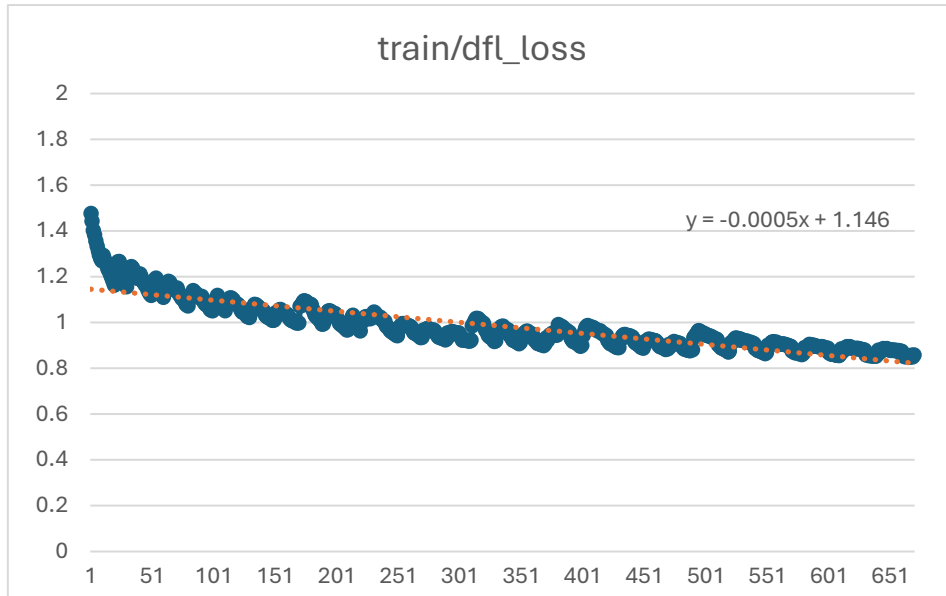
以 Google Colab 和 YOLO 架構訓練了 670 個 epochs 的結果



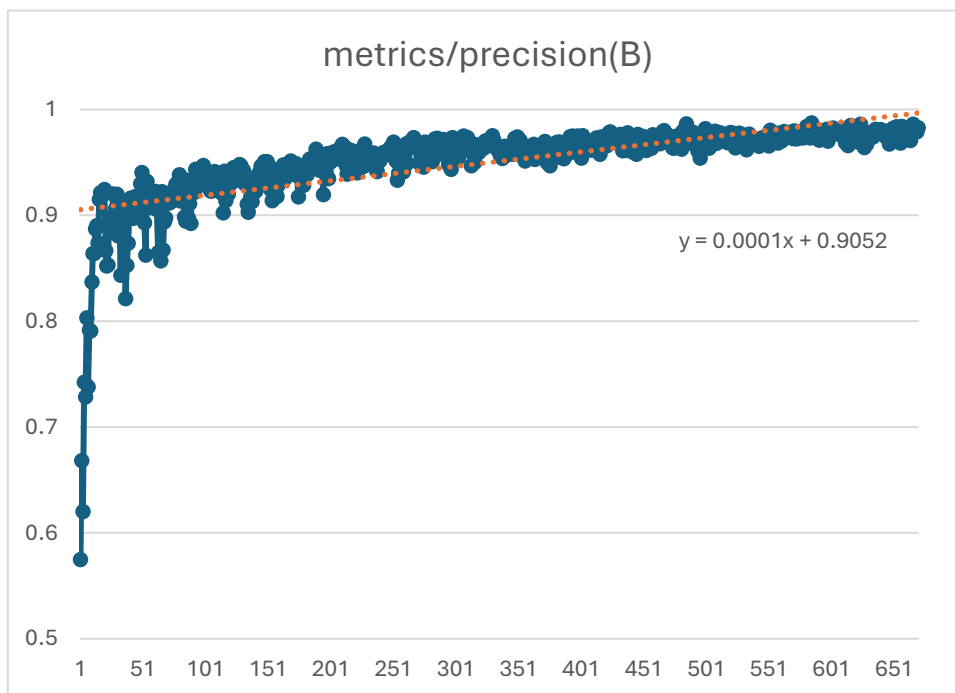
(Fig. 2、670 個 epochs 的 train/box\_loss)



(Fig. 3、670 個 epochs 的 train/clc\_loss)



(Fig. 4、670 個 epochs 的 train/df\_l\_loss)



(Fig. 5、470 個 epochs 的 metrics/precision(B))

可以看到，loss 是有在穩定下降的，這四張圖表分別代表：

1. Fig. 2, box loss(Box Regression Loss) 邊界框回歸損失：

這是關於模型的邊界框與真實框之間的重疊程度，例如無人機如果位於畫面正中間，但模型框畫在左下角，box loss 的值可能就會很高，他計算主要是採用 CIoU 的損失公式，當  $L_{CIoU}$  越接近 0，則代表訓練結果會越好，從訓練的圖表來看，box loss 從 1.9 下降到 0.5 左右，雖然整體而言這個數值還不夠低，但實測表現還算可以，因此可接受。

2. Fig. 3, cls loss(Classification Loss) 分類損失：

主要是用以衡量模型對物件「類別」的預測和實際標籤的差異，像是如果此物體的 label 要是 drone，預測到是 helicopter 或其他就會顯示錯誤，其運用的計算公式是 BCE(Binary Cross Entropy with Logits)，訓練期間，cls\_loss 從初始值 2.69975 穩定下降至約 0.3，顯示模型在分類任務上有明顯的收斂。然而，訓練資料最初包含 airplane、drone 與 helicopter 等三類標籤，但實際應用場景中我們僅關注 drone 的識別。因此，雖然 cls\_loss 是整體模型表現的一環，但其在本任務中的參考價值相對有限，因為在只保留單一類別的情況下，分類錯誤的機會極低。

3. Fig. 4, dfl loss(Distribution Focal Loss)：

dfl loss 的用途在於改善邊界框預測的準度，他會把一個連續值轉換成離散區間的機率分布，從上方圖表的 dfl 曲線可以判斷，loss 的值從 1.45 穩定下降至 0.85 左右，儘管曲線的下降程度看起來不及另外兩個 loss 曲線，但模型在邊界框的座標預測仍持續改善，穩定收斂的趨勢和最終便是的成效也顯示模型具備足夠的邊界位置辨識能力。

4. Fig. 4, precision：

precision 就很好理解了，主要就是判斷正確偵測到物體和錯誤偵測物體的比值，precision 越高就會代表模型預測越準確，而從上面的分布圖可以看出來 precision 逐漸從 0.58 上升到 0.96 以上，準確度相當不錯，而且穩定上升也顯示了模型有持續進步。

## 4.2 Arduino 雲台控制

撰寫程式使雲台能跟著標註框的位置平滑的移動，寫了一些對應的程式，加上非線性

放大與角度平滑參數去優化雲台的移動效果。由於辨識中發現的一些問題，針對過程中可能偵測到非目標物體的情況，設置一些門檻和延遲條件，避免雲台會因為誤判而移動頻繁，提升整體系統的穩定性與追蹤準確度。擷取的追蹤辨識圖片：



(Fig. 6、辨識與追蹤無人機的結果)

實地演示雲台會因為無人機的位置水平或垂直的轉動，使上方的相機視角改變，持續追蹤到無人機的蹤影。

## 4. Conclusion

在本專題的訓練過程中，模型透過大量的影像學習無人機的外型特徵，最終能達到不錯的辨識準確率。然而在實作上仍面臨許多挑戰，包括 USB 鏡頭無法自動對焦，導致遠距離時畫面模糊，辨識效果受限；室內燈光不足也會影響模型擷取影像的能力。此外，影像從 USB 鏡頭傳輸至筆電後再顯示於螢幕上存在明顯延遲，雖不致影響基本辨識，但若進一步應用至即時回應系統（如觸發雷射指向）會造成困擾。為改善上述問題，未來可考慮更高畫質且延遲更低的 webcam 或進行系統端延遲優化。

此外，本專題亦嘗試以 Darknet 框架訓練相同資料集，但模型難以收斂、出現 loss 爆炸現象，顯示其穩定性與彈性不如 Ultralytics 的 YOLO 系列。另一項挑戰為嘗試將 Raspberry Pi 納入影像前端處理流程，但因模型容量與運算資源受限，導致系統不穩，未能成功實現。未來可考慮對模型進行壓縮或量化（如剪枝），以提升部署靈活性。系統也將進一步拓展，於雲台上加裝雷射模組，實現「物體飛到哪，雷射就指到哪」的追蹤功能，進而應用於安防、導引或互動展示等場景，擴展實用性與創新性。

## 5. Review and Reflections

這次做專題的經歷是在求學過程中難得一次自發性完成的成果，而且從搜索資料到實作都須依靠自己，無法像修課一樣由任課教師一步步帶著你走，餵養你知識。在做這項專題前，我僅僅是知道機器學習在蓬勃發展，還未深入了解其背後的知識理論，這次自己實際拿資料操作後才發現 AI 的進展已經非常成熟，丟一大堆圖片進去，具備自己學習思考的能力，簡直就像人一樣。此外，在修習專題期間，我也培養了很多動手做的技能，像是繪製立體圖以供 3D 列印機製作。從實驗室中我學到了很多，非常感謝黃衍介教授、實驗室學長還有儀科中心專家的協助，讓我最後能順利完成這份專題報告

## 6. Reference

- [1] 林聖傑，「具人工智慧及影像視覺辨識之無人飛行器開發與應用」，碩士論文，資訊工程系，正修科技大學，高雄，台灣，2022 年 6 月。
- [2] 池宇非，「使用機器學習的快速人臉辨識系統」，碩士論文，電機工程學系，國立中央大學，桃園，台灣，2024 年 1 月。
- [3] Edge Electronics, “How to Train YOLO Object Detection Models in Google Colab (YOLO11, YOLOv8, YOLOv5),” *YouTube*, Apr. 2024. [Online]. Available: <https://www.youtube.com/watch?v=r0RspilG260&t=269s>
- [4] 林芝羽，「機器學習與深度學習人臉辨識差異之研究」，碩士論文，資訊管理系碩士班，龍華科技大學，新北，台灣，2022 年 1 月。
- [5] Ahmed Mohsen, **Drone Detection - new-peksv Dataset**, Roboflow Universe. Available online: <https://universe.roboflow.com/ahmedmohsen/drone-detection-new-peksv/dataset/5> (accessed on 3 May 2025).