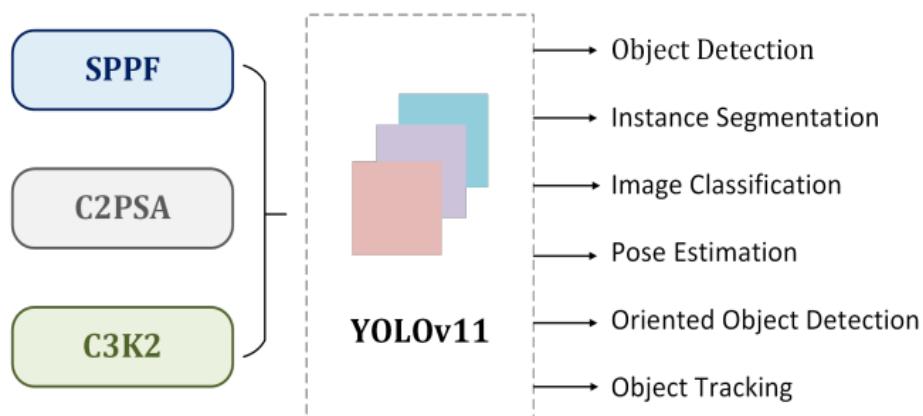


電腦視覺實務與深度學習作業一

班級:電機碩一、姓名:林郁庭、學號:114318058、指導老師:郭天穎

1 模型架構

本次作業[1]我選用 yolo11 模型[2]，且 yolo11 還有分 n、s、m、l、x 等多個不同運算量大小的模型。另外 yolo11 也支持不同電腦視覺任務，包含物件偵測、分割、分類和追蹤等功能。而這次作業需要物件偵測的功能，並且我主要選用 l 的模型進行實驗。



圖一、YOLO11 主要模塊

Yolo 系列的架構大部分皆有分 backbone、neck 和 head 等三個部分。

1.1 Backbone

Backbone 就是特徵萃取器的部分，目標是要從影像中提取重點區域，主要包含 SPPF 和 C2PSA 等兩個主要的模塊。SPPF 中文全稱快速金字塔池化，負責快速降維。C2PSA 則是引入了空間注意力機制，能讓模型更專注於重要區域。

1.2 Neck

Neck 的部分主要負責在不同尺度下處理 feature map，並將其送入 head 部分進行預測，其中 c3k2 模組取代了過去的 c2f 模組，為模型提供了更快更

有效率的整體表現。

c3k2 模組用兩個小型的卷積層取代掉一個大型的卷積層，以達到更快的特徵萃取。

1.3 Head

Head 的部分負責生成出預測結果，包括分類和物件偵測等。輸出 `labels` 和 `bounding box`。

2 實驗細節

2.1 訓練策略

本次實驗我採用訓練-微調兩階段訓練測略，在訓練階段時將會用隨機初始化權重從頭訓練。微調階段將會加入 `feature extractor` 在 ImageNet 上預訓練權重加強訓練，並啟用更強的資料增強策略(2.2)。

2.2 前處理和資料增強

本次實驗將訓練樣本切成訓練集和驗證集。訓練集共有 **1139** 張(90%)，驗證集共有 **127** 張(10%)，以方便在訓練時監督。

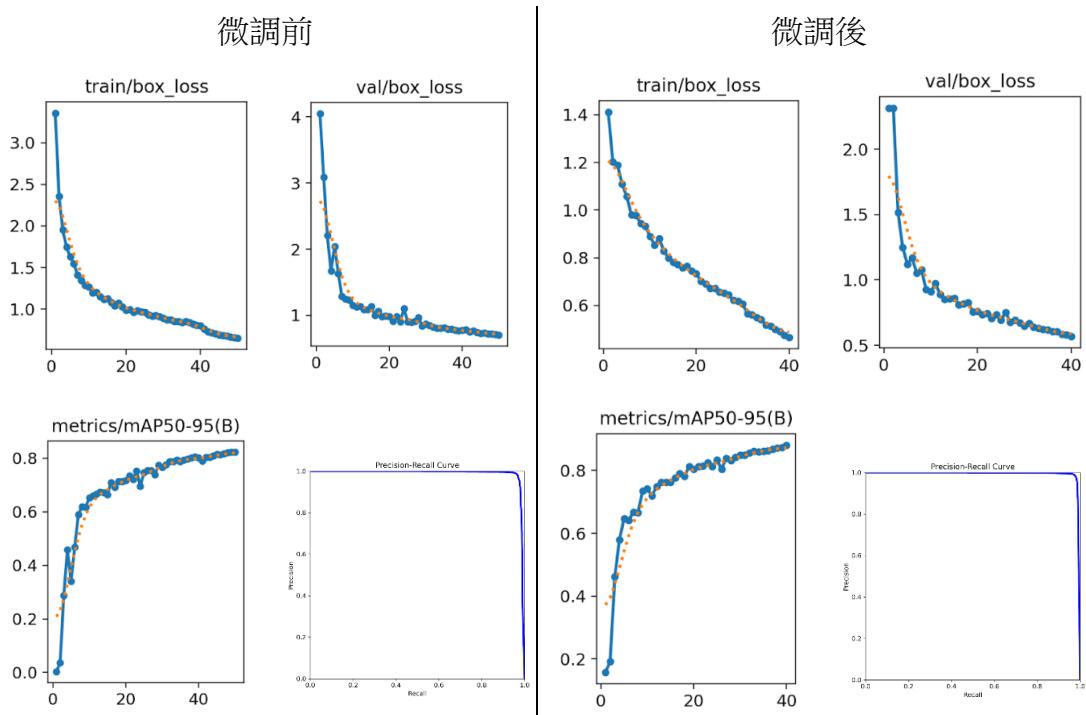
本次實驗同樣也有對訓練資料採用了資料增強的前處理策略，目的是希望模型可以學習到更多元化的資料分布。本文採取的包括**隨機旋轉 10 度**、**平移**、**左右翻轉**、**縮放**和 **Mosaic** 等。**Mosaic** 的原理是將隨機四張圖片拼接，以增加資料多樣性和模型魯棒性，不過因為 **Mosaic** 是較強的資料增強方法，所以本文只會將 **Mosaic** 和縮放用在微調階段，也就是在訓練階段模型已經學習到基本資訊後，避免模型從一開始就無法學習到任何模式。

2.3 超參數和損失函數

本次實驗採用 **AdamW 優化器**，並設定學習率 **0.002**。損失函數為 `box_loss`、`cls_loss` 和 `dlf_loss` 等。但本次作業是做單物件偵測，所以主要觀察 **box_loss**，也就是關注預測框和 `ground truth` 差距有多大。

3 量化分析

本次實驗我在 colab 上用 T4 訓練，訓練階段 **50 epochs** 加上微調階段 **40epochs**。如圖二所示，最後微調階段的 `box_loss` 停留在 **0.4652**，`mAP50:95` 停留在 **0.88**。



圖二、本次實驗的曲線圖

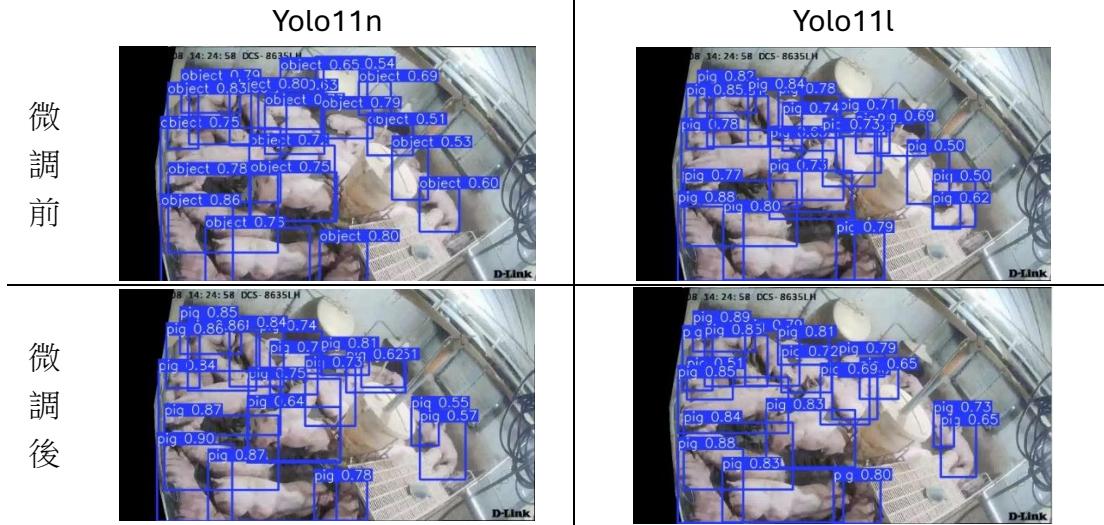
架構	訓練週期	mAP50:95	分數
Yolo11n	100+30	0.79	0.24
Yolo11l	50+30	0.86	0.29
Yolo11l	50+40	0.88	0.32

表一、(註:僅有最後一個實驗有開啟較強的資料增強)

架構	微調前		微調後	
	mAP50:95	分數	mAP50:95	分數
Yolo11n	0.782	0.22	0.79	0.24
Yolo11l	0.823	0.28	0.88	0.32

表二

除了本次實驗外，我也針對了 YOLO11n 模型做了實驗，並針對有無微調的結果做出比較。表一的實驗結果顯示，l 架構的表現明顯皆比 n 架構還好。也顯示了在微調階段除了長一點的訓練週期外，開啟較強的資料增強能幫助表現提升。另外表二顯示了不管是 n 或 l 架構，使用微調皆可以提升模型的表現。圖三也顯示了模型在微調前比微調後還要更容易框到不是目標的物件，且信心值普遍比微調後低。這也證明了微調在本次實驗的重要性。



圖三、實際測試圖

4 結論

綜合上述實驗結果可知，無論是 Yolo11n 或 Yolo11l 架構，微調皆能有效提升模型的表現，尤其是在目標辨識的準確性上有明顯進步。微調後不僅提升了模型的信心值，也減少了誤判非目標物件的情況。此外，適當地延長訓練週期及加強資料增強策略，亦能進一步優化模型效能。這些發現證明，微調在提升深度學習模型表現上扮演著關鍵角色，未來相關研究可持續針對微調方法與資料增強技術進行深入探討，以追求更高的辨識準確率。

5 引用

- [1] CVPDL.Fall.2025. [NTU] CVPDL 2025 HW1. Kaggle, 2025,
[https://kaggle.com/competitions/ntu-cvpdl-2025-hw-1.](https://kaggle.com/competitions/ntu-cvpdl-2025-hw-1)
- [2] Khanam, Rahima, and Muhammad Hussain. "Yolov11: An overview of the key architectural enhancements." *arXiv preprint arXiv:2410.17725* (2024).