

# 靜 宜 大 學

## 資 訊 工 程 學 系

### 畢 業 專 題 成 果 報 告 書

利用深度學習於機械手臂煎荷包蛋系統

學 生：

資工四 B    411018052    張閔翔

資工四 B    411018230    陳宜煒

資工四 B    411023007    林冠豪

指導教授：劉志俊    教授

西 元 二 〇 二 四 年 十 二 月

# 利用深度學習於機械手臂煎荷包蛋系統

學生：張閔翔、陳宜煒、林冠豪

指導教授：劉志俊

## 靜宜大學資訊工程學系

### 摘要

本專題研製之基於深度學習的智慧煎蛋系統，以改善下廚耗費的時間與門檻。隨著生活節奏的加快，家庭烹飪頻率下降，導致了外食需求增加。然而，外食成本較高且難以長期維持健康飲食習慣，這也使得簡化家庭烹飪的需求日益增加。此外，行動不便者如高齡者或身體有障礙的人士，若想進廚房烹飪將會是場挑戰，因此數位廚房的發展不僅能幫助提升烹飪效率，也能實現無障礙廚房的目標，對於推動社會的平等和便利化生活有重大意義。

為了有效減少家庭烹飪時間，提高飲食健康的品質，並幫助行動不便者也能獲得烹飪的機會。本專題透過物件偵測技術實現對烹飪過程中關鍵物件的識別，例如雞蛋、鍋鏟、平底鍋等，最終結合機械手臂完成煎蛋過程中的一系列動作，亦同時探討多種 YOLO 系列物件偵測技術中不同版本之效能差異，找出最適合本專題需求的最佳解法。不僅證明了深度學習技術在數位廚房領域的可行性和潛力，並對於推動數位廚房的普及具有重要的啟發意義，幫助數位廚房的發展向前邁進一步。

## 致謝

本專題的完成離不開許多人的幫助，必須要感謝的人很多，特此致以誠摯的感謝。

首先衷心感謝指導教授劉志俊教授的悉心指導與鼓勵，在整個研究過程中提供了耐心而詳細的指導與建議，尤其是在研究背景與系統架構方面的分析，帶給我們相當多的啟發，使我們獲益匪淺，對本專題的完成具有重要意義。此外，承蒙口試老師林耀鈴教授與戴自強教授提供許多寶貴的建議和指正，謹致以最深的謝意。

其次，感謝專題團隊的所有組員，在面對問題時皆能積極討論並分析出最佳結果，在每次的修正錯誤中互相砥礪成長，以及感謝實驗室所有成員提供的協助，不管是在設備上或是經驗的分享上，都提供了莫大的幫助。

同時，感謝上銀科技股份有限公司提供儀器支持，機械手臂作為本專題系統不可缺少的部分，有你們良好的設備，使我們實驗更加順利。

最後，感謝雙親與家人還有周遭的朋友，在整個專題的過程中給予無盡的支持與鼓勵，使我們能全力以赴的面對問題。衷心感謝他們無所求的付出。

# 目錄

中文摘要 .....	i
致謝 .....	ii
目錄 .....	iii
表目錄 .....	v
圖目錄 .....	vi
第一章 緒論 .....	1
第二章 專題內容與進行方法 .....	2
2.1 動機與目的 .....	2
2.2 專題相關現有系統回顧與優缺點分析 .....	2
2.2.1 機器人餐廳的應用與客戶體驗分析 .....	2
2.2.2 食物數據集的建構與分析 .....	3
2.2.3 食物識別系統的應用與改進 .....	4
2.2.4 高度自動化廚房與自主烹飪系統的發展 .....	4
2.3 專題進度規劃與進行方法說明 .....	5
第三章 專題流程與架構 .....	7
第四章 專題成果介紹 .....	9

4.1 實驗環境.....	9
4.2 資料集介紹.....	10
4.2.1 資料採集.....	10
4.2.2 影像邊框標記.....	11
4.3 實驗結果.....	12
4.3.1 模型評估指標.....	12
4.3.2 實驗一.....	15
4.3.3 實驗二.....	16
4.3.4 實驗三.....	17
4.3.5 實驗四.....	18
4.3.6 綜合評估.....	19
第五章 專題學習歷程介紹.....	21
5.1 專題相關軟體學習介紹(YOLOv8 安裝).....	21
5.2 專題製作過程遭遇的問題與解決方法.....	25
第六章 結論與未來展望.....	27
參考文獻.....	28

## 表目錄

表 1：實驗環境.....	9
表 2：YOLO 系列物件偵測技術對於雞蛋的辨識效能 .....	16
表 3：YOLO 系列物件偵測技術對於平底鍋的辨識效能 .....	16
表 4：YOLO 系列物件偵測技術對於鍋底的辨識效能 .....	17
表 5：YOLO 系列物件偵測技術對於鍋鏟的辨識效能 .....	18
表 6：YOLO 系列物件偵測技術對於握把的辨識效能 .....	18
表 7：YOLO 系列物件偵測技術對於煎蛋的辨識效能 .....	19
表 8：YOLO 系列物件偵測技術對於全類別平均的辨識效能 .....	19

## 圖目錄

圖 1：專題進行方法說明圖.....	5
圖 2：實驗流程圖.....	6
圖 3：系統架構圖.....	7
圖 4：機械手臂運作流程範例.....	7
圖 5：實驗影像資料集(a)自行拍攝(b)網路蒐集(c)生成式 AI 產生(d)相關資料庫蒐集.....	11
圖 6：鍋鏟人工邊框標記範例.....	12
圖 7：YOLOv11 全類別平均混淆矩陣.....	20
圖 8：確認 GPU 環境與 CUDA 版本.....	21
圖 9：下載 PyCharm.....	22
圖 10：安裝 PyTorch 套件.....	23
圖 11：在終端輸入安裝 PyTorch 指令.....	23
圖 12：複製 YOLO 專案.....	24
圖 13：安裝成功結果.....	24

# 第一章 緒論

廚房是食物的產地，更是飲食與建築的交會處，隨著時代與科技的演進，廚房已經衍伸出許多不同的形態與樣貌，其中「數位廚房(Digital Kitchen)」的概念逐漸成為未來廚房發展的關鍵。2005 年，麻省理工學院（Massachusetts Institute of Technology, MIT）的 Bonanni 等人[1]率先提出數位廚房的概念，旨在透過嵌入廚房空間、設備和器具等環境技術，將廚房中的活動變得更加便利與有效率，並同時降低人力成本。數位廚房是創造一個能夠智慧化管理烹飪過程、減少人力投入的空間，從而提升生活品質，這一概念不僅僅是技術的進步，更是對現代生活方式的一種響應。

而在本國，根據行政院主計總處統計[2]，國人的餐飲及旅館支出占家庭消費總支出在這 40 年來節節攀升，從 1976 年的 2.48%至 2019 年的 12.8%，成長了 5.2 倍。這一數據顯示，隨著生活節奏的加快，越來越多的人選擇外食，以節省時間和精力。在這樣的社會背景下，數位廚房的發展顯然將成為提升國人生活品質的關鍵，透過智慧化的廚房系統，烹飪將變得更加簡單、高效，這不僅滿足了現代人對便利生活的追求，也在提升便利性的同時，實現生活品質的提升。



## 第二章 專題內容與進行方法

### 2.1 動機與目的

近年來，全球人工智慧技術蓬勃發展，機械手臂開發也逐漸成熟。然而，儘管有大量配備機械手臂的廚房設備出現，但大多數家庭仍無法在日常烹飪中使用[3]。這種現象可能源於設備價格高昂，或是現有技術偏於自動化，僅針對特定任務為目的，無法靈活應對家庭廚房中多變且小規模的烹飪需求。因此，本專題將利用深度學習之物件偵測技術結合機械手臂開發一套智慧烹飪系統，使國人在能負擔成本的狀況下，減少自行下廚耗費的時間並降低在家料理的門檻，能夠有效提供更實惠又便捷且舒適的烹調過程。此外，本研究也有助於解決人口高齡化、和勞動力短缺等問題，智慧烹飪系統不僅能在一般家庭中協助作業，亦能替烹飪能力有限或行動不便的人提供幫助，讓他們能從中獲得各種烹調經驗與獨立感。

### 2.2 專題相關現有系統回顧與優缺點分析

#### 2.2.1 機器人餐廳的應用與客戶體驗分析

Emily Ma, Yafang Bao, Leijun Huang, Danni Wang, and Misun (Sunny) Kim 探討了機器人餐廳在美國和中國的應用[4]，並比較其在產品層次理論(PLT)[5]和體驗經濟模型[6]上的影響。研究問題聚焦於機器人技術如何在不同產品層次(核心、基本、期望、增強)上影響顧客體驗。方法包括兩部分：

第一部分分析供應側如何應用機器人技術於不同層次，第二部分則通過神秘顧客調查和社交媒體評論，研究顧客對機器人餐廳的體驗差異。實驗數據來自美國的 Spyce 和 Creator，以及中國的海底撈和 Robot.He。主要發現指出，美國餐廳主要在核心和基本層次應用機器人技術，而中國餐廳則涵蓋更多層次，包括預期和增強層次，顯示出不同文化背景下顧客對機器人技術的接受度及體驗差異。研究結果不僅擴展了 PLT 和體驗經濟模型在機器人餐廳中的應用，也提供了跨文化比較的新視角，有助於理解機器人技術如何塑造餐飲服務的顧客體驗。

### 2.2.2 食物數據集的建構與分析

資料集方面，我們需要更多的資料以及研究方向，我們了解 2017 年據 Xin Chen, Yu Zhu, Hua Zhou, Liang Diao, Dongyan Wang 等人所知，當時不存在成熟的中餐辨識系統，原因是缺乏了資料集的量與品質，因此他們建立了包含 208 個類別與 185628 張的食物照片的名為“ChineseFoodNet” [7]的超大規模中國菜肴圖像資料集，並且公開供學術研究，其中具我們所需的“荷包蛋”類別，他們也一併提出“TastyNet”兩步數據融合方法，使得多種 CNN 模型在該資料集上的表現顯著提升，達到 81.43%和 81.55%的 top-1 準確率。另一案例是 Lukas Bossard, Matthieu Guillaumin, Luc Van Gool 為了要自動識別食物圖像，利用隨機森林(RF)的方法來挖掘能夠區分不同食物類別

的區域[8]，並透過將補丁對齊至超像素來提升挖掘與分類的效率，為了驗證此方法的性能，此模型在團隊引入的一個包含 101 個食物類別，共 101000 張圖像的“Food-101”資料集的平均準確率為 50.76%，優於改進型費舍爾向量(IFV)和現有的區別性部分挖掘算法 11.88%和 8.13%，在 MIT-Indoor 數據集上也表現優異。

### 2.2.3 食物識別系統的應用與改進

接下來是已經應用於實際生活的案例，Yoshiyuki Kawano, Keiji Yanai 提出了一個實時移動食物識別系統[9]，旨在解決用戶記錄飲食習慣的繁瑣問題，並提供食物的卡路里和營養信息，該系統讓用戶通過手機拍攝食物圖像，並在屏幕上繪製邊界框來標記食物位置，系統使用 GrubCut 算法對食物區域進行分割，提取顏色直方圖和 SURF 特徵，並利用線性支持向量機(SVM)和快速  $X^2$  內核進行分類；實驗資料集包含 50 種食物類別，系統在前五個候選分類中的準確率達到 81.55%。用戶研究顯示，該系統相比傳統的手動食物記錄方式更受歡迎，有效提升了記錄效率和準確性。

### 2.2.4 高度自動化廚房與自主烹飪系統的發展

由於工業製造走向自動化，其後又逐漸邁向民生化。Yummy Operations Robot Initiative (YORI)為了實現高度自動化的烹飪過程，開發了一種自主烹飪系統[10]，利用模組化的機器人廚房和雙臂本體感操控器，實現高度自動

化的烹飪過程。為了配合雙臂機械手與烹飪環境的侷限性，透過複雜的模型和演算法來解決工作空間設計中固有的最佳化問題來實現[11]，這些工具和器具利用本體感受執行器的雙臂機械手臂操作，可以實現快速而精確的運動，有效減輕烹飪中不可避免的影響。

## 2.3 專題進度規劃與進行方法說明

智慧烹飪系統不僅改善了整個廚房的工作流程，更將成為數位廚房的重大進步。然而，考慮到整個智慧烹飪系統的複雜性，需分成多個部分進行階段性開發，本專題以煎蛋為主要範例，以雞蛋辨識與鍋鏟握持等幾個關鍵動作為首要目標，詳細方法說明如下圖 1 所示。

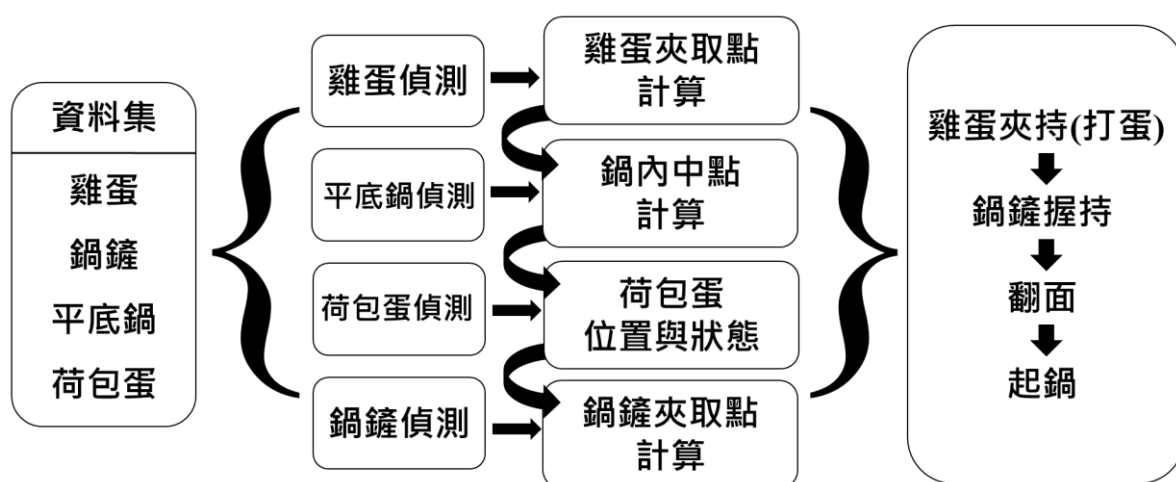


圖 1：專題進行方法說明圖

深度學習之物件偵測技術為本系統核心，因此本專題將著重於物件偵測技術的分析，並設計了四項實驗分別對 YOLO 系列最新版本進行分析，分別是雞蛋偵測、平底鍋與鍋底偵測、鍋鏟與握把偵測還有荷包蛋偵測，透過各個模型效能評估指標進行比較，挑選出智慧煎蛋系統之物件偵測最佳解，實驗流程如圖 2 所示。

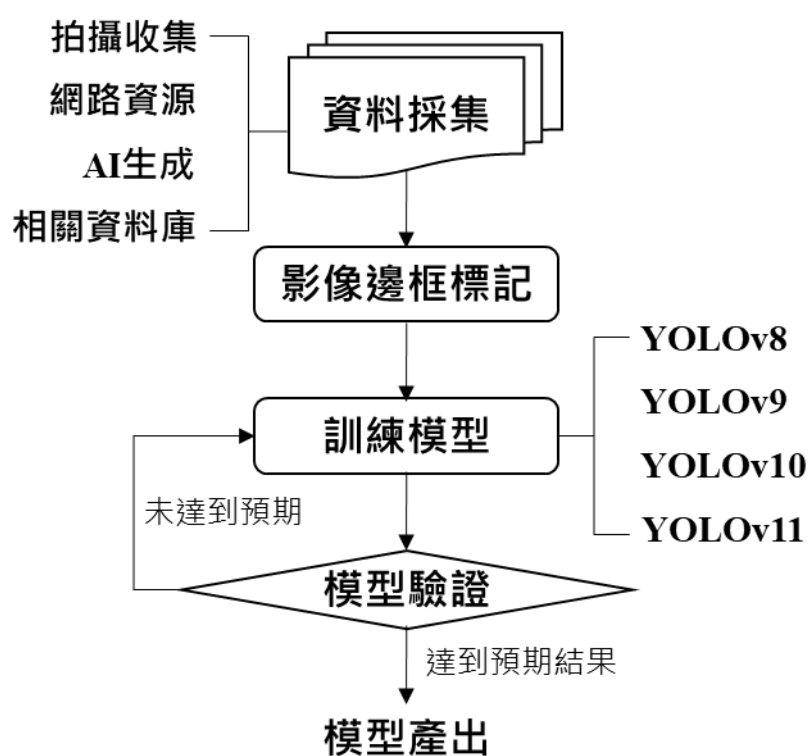


圖 2：實驗流程圖

### 第三章 專題流程與架構

本研究所提出的智慧煎蛋系統之系統架構如圖 3 所示，機械手臂運作流程範例如圖 4 所示，主要方法與步驟如下：

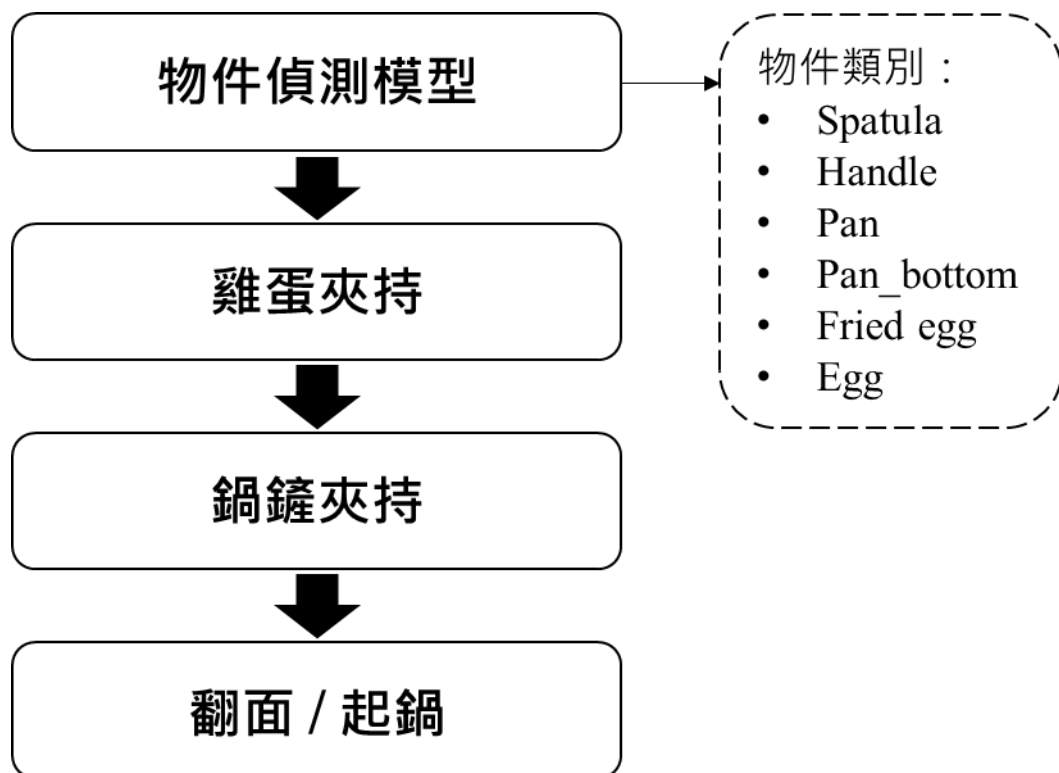


圖 3：系統架構圖

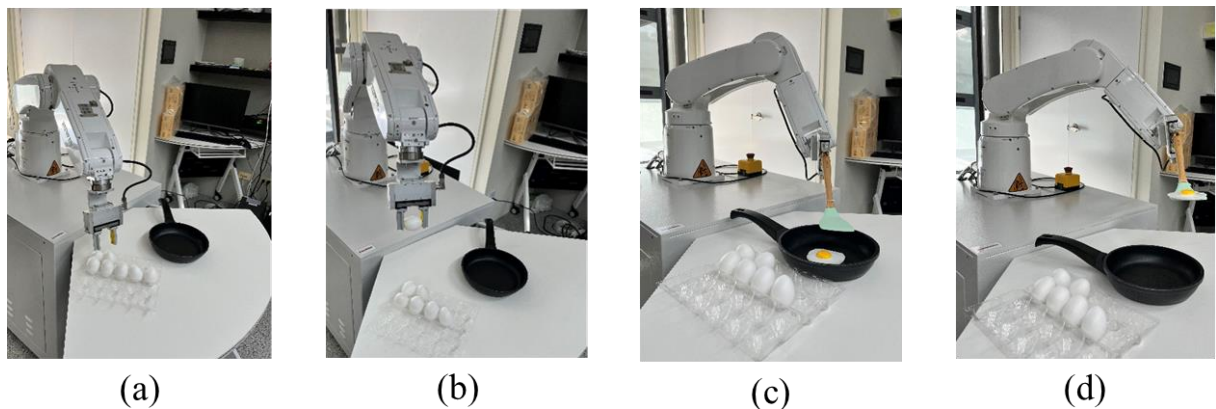


圖 4：機械手臂運作流程範例

- 物件偵測模型：如圖 2(a)所示，本研究採用四種最新 YOLO 系列物件偵測模型 YOLOv8[12]、YOLOv9[13]、YOLOv10[14]與 YOLOv11[15]進行 Spatula、Pan、Fried egg 和 Egg 各類別之物件偵測模型訓練，並且比較四種物件偵測技術何者效能最佳。
- 雞蛋夾持：如圖 2(b) 所示，利用機械手臂夾具上方攝影機所拍下的影像，透過已訓練的 YOLO 模型進行物件偵測，框出雞蛋位置並計算其座標中點，即為雞蛋夾持點。
- 鍋鏟夾持：如圖 2(c) 所示，利用已訓練的模型進行物件偵測，框出鍋鏟與鍋鏟握把位置，透過兩者重疊面積之比例可得知是否為同一物件，並計算鍋鏟握把標記框之矩形對角線交點，其為鍋鏟握把之中心位置，亦為鍋鏟夾持點。
- 翻面或起鍋：流程範例如圖 2(d)。依據上述步驟所得各物件之抓握位置，並根據鍋內煎蛋狀態，機械手臂將進行相對應的動作，以完成整個智慧煎蛋流程。

## 第四章 專題成果介紹

### 4.1 實驗環境

實驗環境如表 1 所示，開發環境為 PyCharm Professional 2024.1.3；程式環境為 Python 3.12；深度學習環境為 PyTorch；物件偵測方式包含 YOLOv8、YOLOv9、YOLOv10 與 YOLOv11。電腦硬體為 i7-12700KF 3.60GHz CPU、32GB RAM、RTX 4070 Ti SUPER GPU；軟體工具包含 LabelImg 影像標註工具 (<https://github.com/HumanSignal/labelImg>)、Stable Diffusion 與 Image Creator 影像生成工具。

表 1：實驗環境

CPU	Intel i7-12700KF 3.60GHz
GPU	NVIDIA GeForce RTX 4070 Ti SUPER
開發環境	PyCharm Professional 2024.1.3
Python 版本	3.12
物件偵測方式	YOLOv8、YOLOv9、YOLOv10、YOLOv11



影像標註工具	LabelImg ( <a href="https://github.com/HumanSignal/labelImg">https://github.com/HumanSignal/labelImg</a> )
生成式 AI	Stable Diffusion、Image Creator

## 4.2 資料集介紹

本研究的實驗資料集處理包含資料採集與邊框標記，詳細說明如下：

### 4.2.1 資料採集

本研究所使用的影像資料集類別共有 6 種分別是 Spatula、Handle、Pan、Pan\_bottom、Fried egg 和 Egg，總數有 840 張，其中 210 張為自行拍攝收集，210 張為由網路上蒐集而來，還有 210 張由 Stable Diffusion 與 Image Creator 等生成式 AI 產生，最後 210 張是由相關資料庫蒐集而來，如圖 3 所示。



(a)

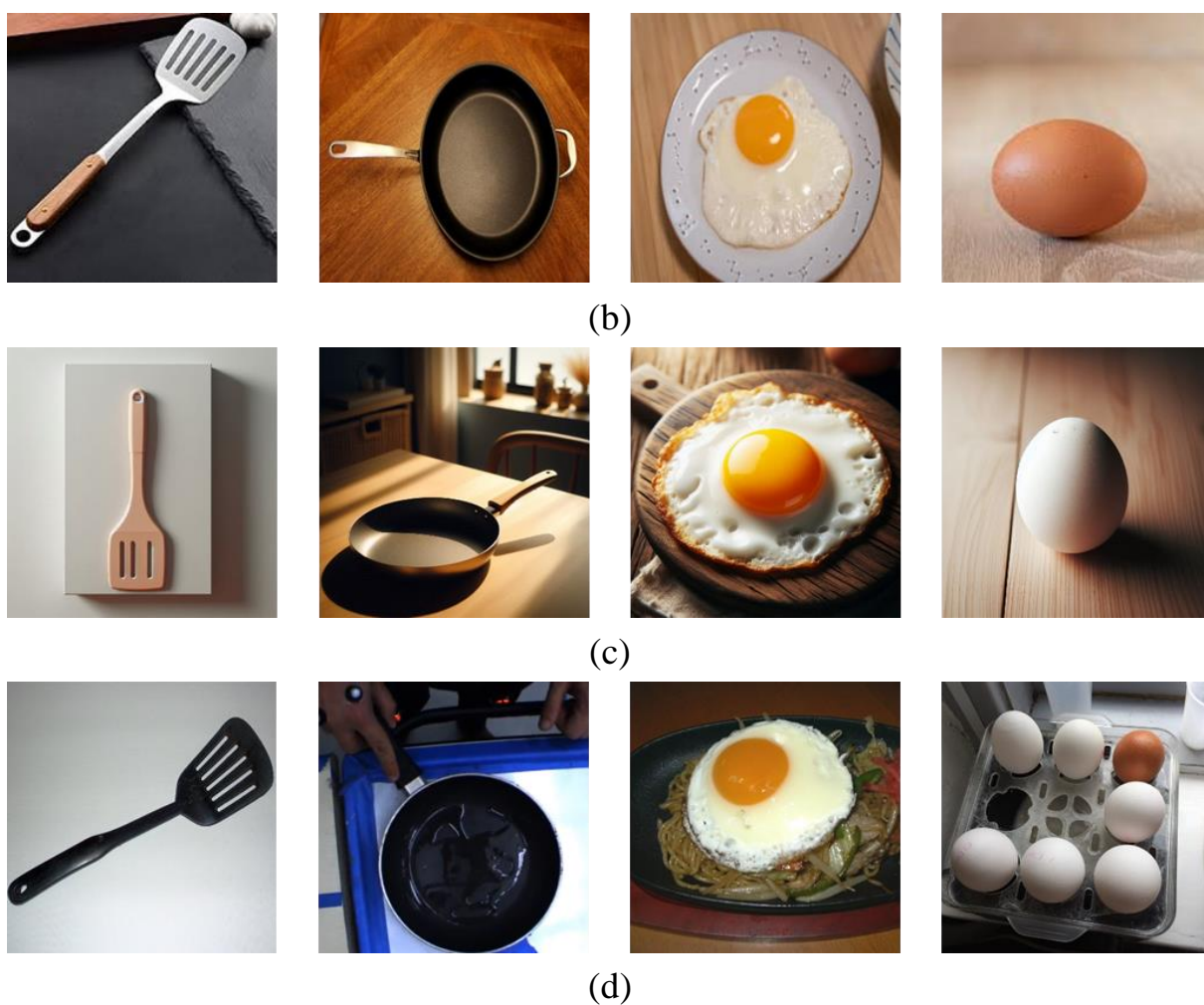


圖 5：實驗影像資料集(a)自行拍攝(b)網路蒐集(c)生成式 AI 產生(d)相關資料庫蒐集

#### 4.2.2 影像邊框標記

本研究以人工進行邊框標記，採用 LabelImg 軟體標註工具 (<https://github.com/HumanSignal/labelImg>)並以 YOLO 格式儲存個標記框位置，以提供後續物件偵測模型訓練，如圖 4 所示。

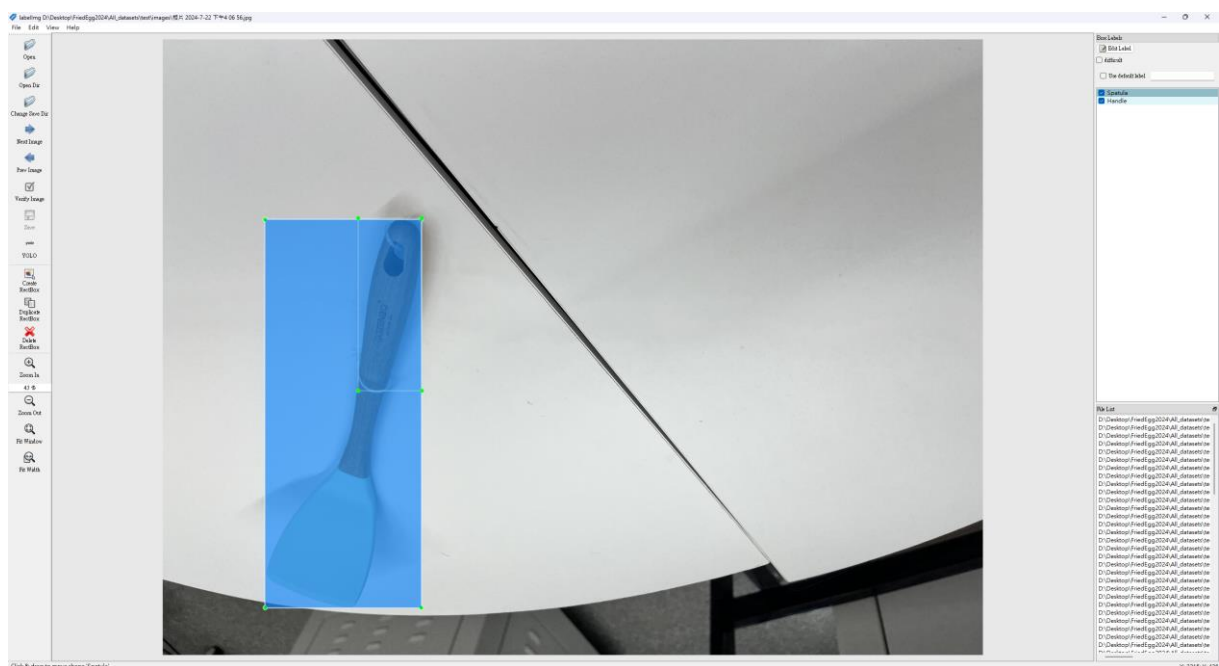


圖 6：鍋鏟人工邊框標記範例

## 4.3 實驗結果

本專題利用共 840 張影像的資料集進行模型訓練，將資料劃分為 600 張訓練資料集、120 張驗證資料集和 120 張測試資料集，進行共 4 種實驗分別是分別是雞蛋偵測、平底鍋與鍋底偵測、鍋鏟與握把偵測還有荷包蛋偵測，類別包含 Spatula、Handle、Pan、Pan\_bottom、Fried egg 以及 Egg 共 6 種，每項實驗都將使用 YOLO 系列中 YOLOv8、YOLOv9、YOLOv10 與 YOLOv11 進行比較並選出何者最佳，模型評估指標與各項實驗結果如下說明：

### 4.3.1 模型評估指標

在實驗中，我們對以下目標進行偵測(雞蛋、荷包蛋、平底鍋、鍋鏟)，

會得到真陽性(True Positive)、偽陽性(False Positive)、真陰性(True Negative)和偽陰性(False Negative)，其定義如下：

- 真陽性(True Positive)：模型能夠正確地檢測到圖片中的雞蛋、荷包蛋、平底鍋、鍋鏟、鍋鏟把手與平底鍋鍋底，假如實際上有 10 個雞蛋，模型正確檢測到了 8 個，那麼 TP 數量就是 8。
- 偽陽性(False Positive)：模型檢測到不存在的目標物體，假如模型將一個不是雞蛋的物體誤檢測為雞蛋的次數，如果有 3 次誤檢測，那麼 FP 數量就是 3。
- 真陰性(True Negative)：模型正確地識別出圖片中沒有目標物體的數量。這在物體檢測中不太常見，因為一般都是有目標物體或無目標物體的情況。
- 偽陰性(False Negative)：模型未能檢測到實際存在的目標物體，假如實際上有 10 個雞蛋，但模型只檢測到了 8 個，有 2 個未檢測到，那麼 FN 數量就是 2。

得到 TP、FP、TN 和 FN 後我們會依此來計算以下三個指標：精確率(Precision)、召回率(Recall)、F1 分數(F1 Score)；交並比(IoU, Intersection over Union)、精確率-召回率曲線(Precision-Recall Curve)以及平均準確率 mAP@0.5 以上共六種指標來對我們的效能進行評估，以下是各項指標的定

義：

- 精確率(Precision)：是衡量分類模型預測結果準確性的一個指標。

$$Precision = \frac{TP}{(TP+FP)} \quad (1)$$

- 召回率(Recall)：也稱為靈敏度(Sensitivity)或真陽性率(True Positive Rate)，是衡量分類模型能夠正確識別所有正樣本的能力的指標。它定義為在所有實際為正類(Positive)的樣本中，模型正確預測為正類的比例。

$$Recall = \frac{TP}{(TP + FN)} \quad (2)$$

- F1 分數(F1 Score)：是一種用來評估二分類模型性能的指標。它是精確率(Precision)和召回率(Recall)的調和平均數，用來綜合考量模型的準確性和完整性。

$$F1 = \frac{2TP}{(2TP + FP + FN)} \quad (3)$$

- 交並比(IoU，Intersection over Union)：兩個邊界框之間的重疊程度。

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (4)$$

- 精確率-召回率曲線(Precision-Recall Curve)：該曲線[16]描繪了在不同分類閾值下模型的精確率(Precision)與召回率(Recall)之間的權衡關係。隨著閾值的變化，精確率和召回率會發生變化：通常情況下，降低分類閾值會增加召回率，但可能會降低精確率；相反，提高分類閾值會增加精確率，但可能會降低召回率。精確率-召回率曲線的理想狀態是位於右上角，表示模型能夠在高精確率的同時也具有高召回率。該曲線的面積

(Area Under the Curve, AUC)越大，表示模型性能越好，能夠更好地平衡精確率和召回率，特別是在處理正負樣本數量嚴重不平衡的情況下。

- 平均準確率(mAP@0.5)：mAP@0.5[17]指在 IoU(Intersection over Union) 閾值為 0.5 時計算出的平均準確率。當模型預測的邊界框與真實邊界框的 IoU 大於或等於 0.5 時，預測被視為正確。在計算 mAP@0.5 時，首先對每個類別計算其平均準確率(AP, Average Precision)，即該類別的精確率-召回率曲線下的面積，然後對所有類別的 AP 取平均值。mAP@0.5 能夠綜合反映模型在所有類別上的整體檢測性能，值越高表示模型在多個類別檢測中具有更好的準確性和召回率平衡。

#### 4.3.2 實驗一

測試 YOLOv8、YOLOv9、YOLOv10 與 YOLOv11 之物件偵測技術對於生雞蛋的辨識效能。資料集總共為 140 張，平均來自網路資源、生成式 AI、自行蒐集和相關資料庫 4 種不同來源。雞蛋偵測的實驗結果如表 2 所示，四種 YOLO 物件偵測技術對雞蛋整體雖然不及煎蛋的高度準確，不過效能依舊高水準，其精確率皆有 0.687 以上、召回率皆有 0.846 以上且 F1 分數皆有 0.788 以上，其中平均準確率 mAP@0.5 以 YOLOv11 的 0.916 為最佳。

表 2：YOLO 系列物件偵測技術對於雞蛋的辨識效能

機蛋偵測模型	YOLOv8	YOLOv9	YOLOv10	YOLOv11
Precision	0.687	0.839	0.762	0.753
Recall	0.923	0.846	0.923	0.942
F1	0.788	0.842	0.835	0.837
mAP@0.5	0.902	0.901	0.872	0.916

### 4.3.3 實驗二

測試 YOLOv8、YOLOv9、YOLOv10 與 YOLOv11 之物件偵測技術對於平底鍋以及鍋底的辨識效能。資料集總共為 280 張，平均來自網路資源、生成式 AI、自行蒐集和相關資料庫 4 種不同來源。平底鍋偵測的實驗結果如表 3 所示，四種 YOLO 物件偵測技術對平底鍋的精確率皆有 0.855 以上、召回率皆有 0.675 以上且 F1 分數皆有 0.772 以上，其中平均準確率 mAP@0.5 以 YOLOv9 的 0.935 最佳。鍋底偵測的實驗結果如表 4 所示，精確率皆有 0.881 以上、召回率皆有 0.825 以上且 F1 分數皆有 0.868 以上，其中平均準確率 mAP@0.5 以 YOLOv8 的 0.914 最佳。綜上所述，雖然整體效能需面對到召回率略低的問題。

表 3：YOLO 系列物件偵測技術對於平底鍋的辨識效能

平底鍋 偵測模型	YOLOv8	YOLOv9	YOLOv10	YOLOv11
-------------	--------	--------	---------	---------

Precision	0.855	0.958	0.902	0.93
Recall	0.9	0.85	0.675	0.875
F1	0.877	0.901	0.772	0.902
mAP@0.5	0.915	0.935	0.85	0.921

表 4：YOLO 系列物件偵測技術對於鍋底的辨識效能

鍋底偵測模型	YOLOv8	YOLOv9	YOLOv10	YOLOv11
Precision	0.881	0.916	0.918	0.938
Recall	0.875	0.825	0.875	0.875
F1	0.878	0.868	0.896	0.905
mAP@0.5	0.914	0.886	0.889	0.905

#### 4.3.4 實驗三

測試 YOLOv8、YOLOv9、YOLOv10 與 YOLOv11 之物件偵測技術對於鍋鏟以及握把的辨識效能。資料集總共為 280 張，平均來自網路資源、生成式 AI、自行蒐集和相關資料庫 4 種不同來源。鍋鏟偵測的實驗結果如表 5 所示，四種 YOLO 物件偵測技術對鍋鏟的精確率皆有 0.798 以上、F1 分數皆有 0.734 以上不過召回率最高卻只有 0.769，平均準確率 mAP@0.5 以 YOLOv8 的 0.798 最佳。握把偵測的實驗結果如表 6 所示，精確率皆有 0.817 以上，但召回率最高只有 0.61 以及 F1 分數最高只有 0.725，平均準確率 mAP@0.5 以 YOLOv9 的 0.725 最佳。綜上所述，四者之間並無巨大差異，精確率皆有不錯的表現，但在召回率上卻不盡人意，是未來需要加以改善的部分。



表 5：YOLO 系列物件偵測技術對於鍋鏟的辨識效能

鍋鏟偵測模型	YOLOv8	YOLOv9	YOLOv10	YOLOv11
Precision	0.798	0.898	0.92	0.967
Recall	0.769	0.683	0.61	0.704
F1	0.783	0.776	0.734	0.815
mAP@0.5	0.798	0.754	0.727	0.783

表 6：YOLO 系列物件偵測技術對於握把的辨識效能

握把偵測模型	YOLOv8	YOLOv9	YOLOv10	YOLOv11
Precision	0.892	1	0.817	0.939
Recall	0.61	0.557	0.366	0.585
F1	0.725	0.715	0.506	0.721
mAP@0.5	0.641	0.725	0.56	0.662

#### 4.3.5 實驗四

測試 YOLOv8、YOLOv9、YOLOv10 與 YOLOv11 之物件偵測技術對於煎蛋的辨識效能。資料集總共為 140 張，平均來自網路資源、生成式 AI、自行蒐集和相關資料庫 4 種不同來源。平底鍋偵測的實驗結果如表 7 所示，四種 YOLO 物件偵測技術對煎蛋皆有相當高的準確率，其精確率皆有 0.873 以上、召回率皆有 0.913 以上且 F1 分數皆有 0.913 以上，其中平均準確 mAP@0.5 以 YOLOv11 的 0.993 為最佳。

表 7：YOLO 系列物件偵測技術對於煎蛋的辨識效能

煎蛋偵測模型	YOLOv8	YOLOv9	YOLOv10	YOLOv11
Precision	0.873	0.974	0.958	1
Recall	0.957	0.913	0.988	0.957
F1	0.913	0.943	0.973	0.978
mAP@0.5	0.987	0.987	0.992	0.993

#### 4.3.6綜合評估

全資料集總數為 840 張，其中包含 280 張鍋鏟影像、280 張平底鍋影像、140 張生雞蛋影像以及 140 張荷包蛋影像，平均來自網路資源、生成式 AI、自行蒐集和相關資料庫 4 種不同來源。全類別平均結果如表 8 所示，四種 YOLO 物件偵測技術全類別平均的精確率皆有 0.831 以上、召回率皆有 0.739 以上且 F1 分數皆有 0.803 以上，其中平均準確率 mAP@0.5 以 YOLOv9 的 0.865 最佳，不過綜合評比而言 YOLOv11 的平均辨識效能仍略勝一籌。

表 8：YOLO 系列物件偵測技術對於全類別平均的辨識效能

全類別平均	YOLOv8	YOLOv9	YOLOv10	YOLOv11
Precision	0.831	0.931	0.879	0.921
Recall	0.839	0.779	0.739	0.823
F1	0.835	0.848	0.803	0.869
mAP@0.5	0.859	0.865	0.815	0.863

因此，本專題以 YOLOv11 之全類別平均混淆矩陣評估模型的改進方向。如圖 5 所示，鍋鏟與握把的召回率較低是本專題模型改進的重要問題，針對實驗準確率與召回率問題，本研究將以擴大資料集或對訓練影像進行前處理等方式，測試模型辨識效率能否更加提升。

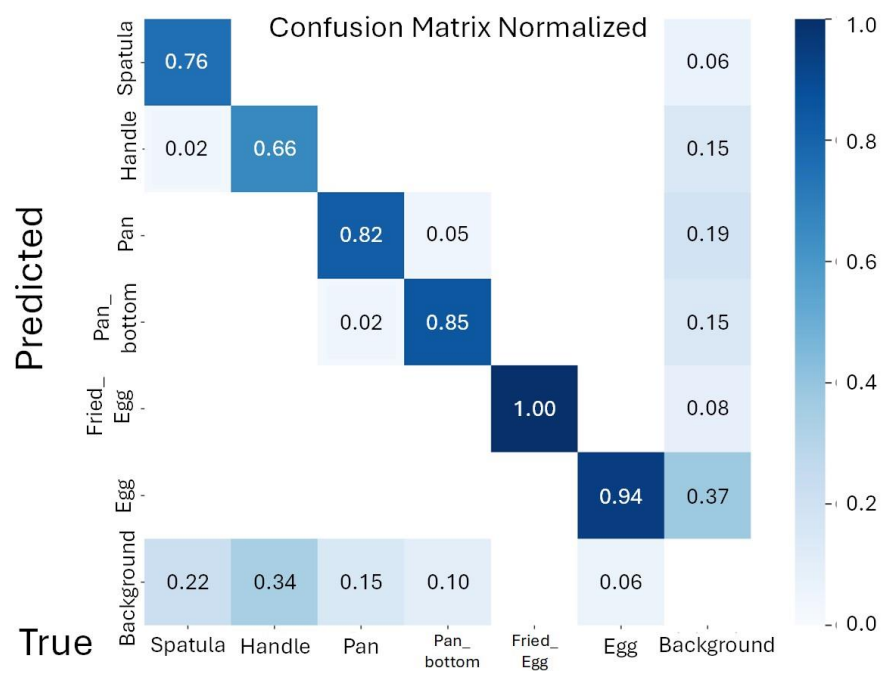


圖 7：YOLOv11 全類別平均混淆矩陣



PyCharm 專業版，若無帳號請自行創建帳號，下載網址

(<https://www.jetbrains.com/pycharm/download/>)。

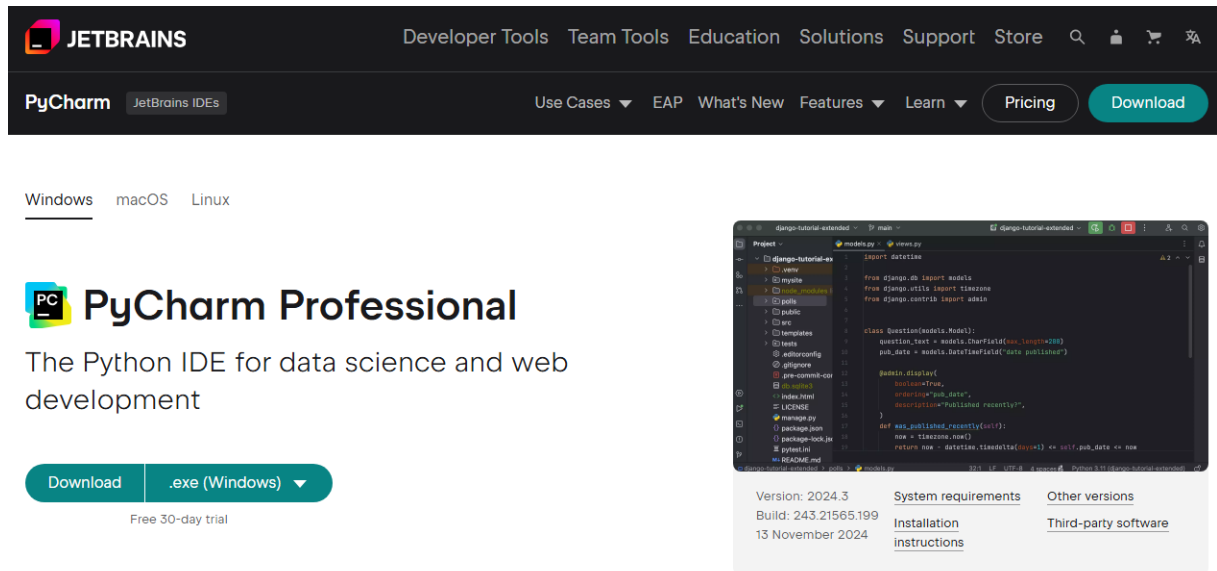


圖 9：下載 PyCharm

3. 新建專案並安裝 PyTorch 套件：請至以下網址選擇並安裝合適的 PyTorch 版本(<https://pytorch.org/>)，並將指令複製到專案下的終端。

Select your preferences and run the install command. Stable represents the most currently tested and supported version of PyTorch. This should be suitable for many users. Preview is available if you want the latest, not fully tested and supported, builds that are generated nightly. Please ensure that you have **met the prerequisites below (e.g., numpy)**, depending on your package manager. Anaconda is our recommended package manager since it installs all dependencies. You can also [install previous versions of PyTorch](#). Note that LibTorch is only available for C++.

PyTorch Build	Stable (2.0.1)		Preview (Nightly)	
Your OS	Linux	Mac	Windows	
Package	Conda	Pip	LibTorch	Source
Language	Python		C++ / Java	
Compute Platform	CUDA 11.7	CUDA 11.8	ROCm 5.4.2	CPU
Run this Command:	<pre>pip3 install torch torchvision torchaudio --index-url https://download.pytorch.org/whl/cu118</pre>			

圖 10：安裝 PyTorch 套件

```

9  print(f'Hi, {name}') # Press Ctrl+F8 to toggle the breakpoint.
10
11
12  # Press the green button in the gutter to run the script.
13  if __name__ == '__main__':
14      print_hi('PyCharm')
15
base of pip available: 22.3.1 -> 23.1.2
run: python.exe -m pip install --upgrade pip
mProjects\AnimalDetection_YOLOv8> pip3 install torch torchvision torchaudio --index-url https://download.pytorch.org/whl/cu118

```

圖 11：在終端輸入安裝 PyTorch 指令

4. 複製 YOLO 專案：在 Terminal(終端)打上以下指令“pip install ultralytics”。

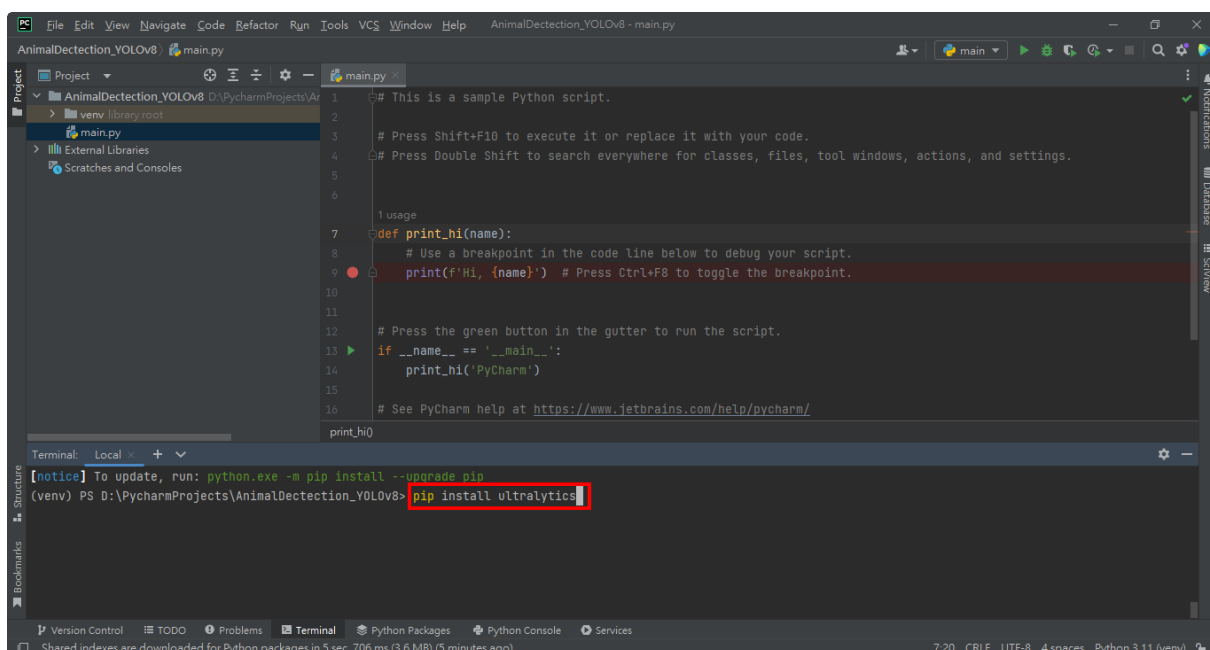


圖 12：複製 YOLO 專案

5. 執行測試：在終端輸入 (`yolo predict model=yolov8n.pt source='https://ultralytics.com/images/bus.jpg'`)，執行後在 runs 資料夾中出現如圖 13 畫面即為成功。

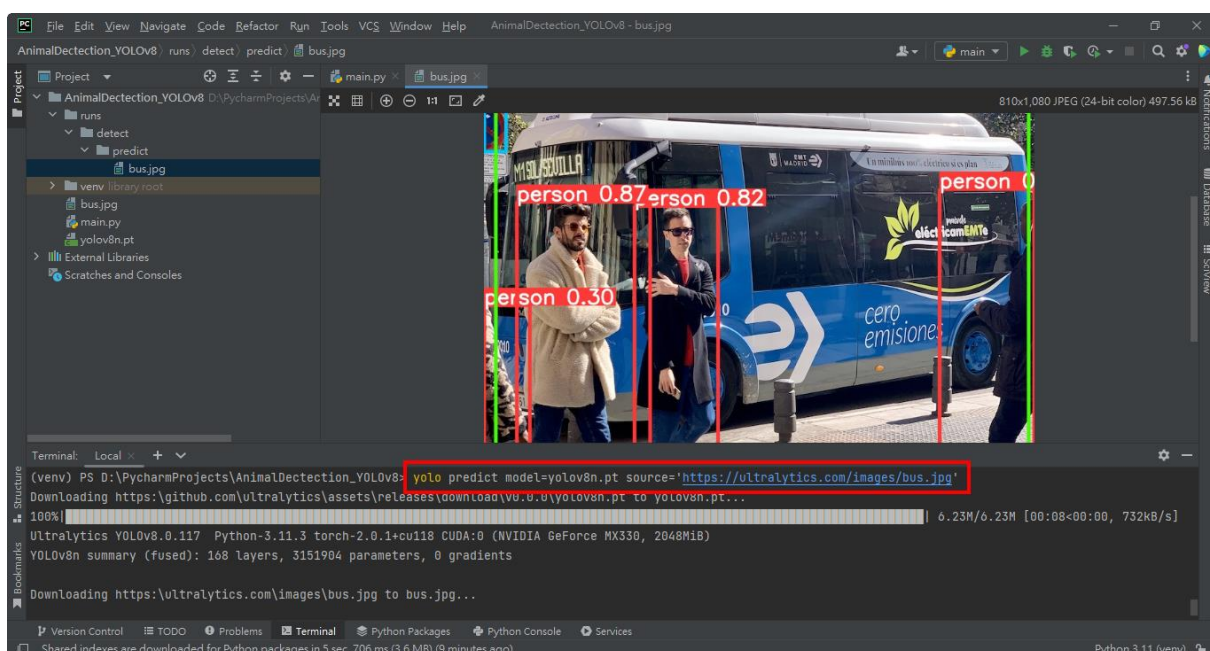


圖 13：安裝成功結果

## 5.2 專題製作過程遭遇的問題與解決方法

在本研究中，我們的目標是利用 YOLO 模型與攝影鏡頭來辨識煎蛋過程中所需的關鍵物件，包括雞蛋、鍋鏟、鍋鏟柄、平底鍋、以及鍋底的荷包蛋，並結合機械手臂的自動化程序完成煎蛋的全流程。在實現這一目標的過程中，我們面臨了許多實際操作與技術層面的挑戰，以下將逐一分析這些問題與我們採取的解決方法。

### 6. 問題一：數據集不足與標註精度

在開始訓練 YOLO 模型時，我們發現針對特定的廚房場景與物件（如鍋鏟、雞蛋、鍋底荷包蛋）的公開數據集十分有限，而自行收集數據又面臨拍攝角度、光線條件的多樣化需求。此外，手動標註物件的過程非常繁瑣且容易出現標註誤差。

我們決定自行建立數據集，拍攝了不同光線、角度與背景下的場景圖片，以涵蓋物件的多樣性。同時，採用了半自動化標註工具，如 LabelImg，來加速標註過程。為了提高標註的準確性，我們對標註結果進行了多次人工校對。

### 7. 問題二：模型訓練與精度不足

在初期訓練 YOLO 模型時，模型的精度表現不如預期，尤其是在辨識小型物件（如鍋鏟柄）與複雜背景下的雞蛋時，常出現誤判或漏判現象。



針對此問題，我們首先優化了數據集的平衡性，確保每類物件的樣本數足夠均衡。其次，調整了 YOLO 模型的錨框（anchor box）參數，使其更適合我們的物件大小特徵。

## 第六章 結論與未來展望

本研究採用最新 YOLO 系列之物件偵測技術 YOLOv8、YOLOv9、YOLOv10 和 YOLOv11，針對廚房物品辨識問題進行比較，根據實驗結果顯示，四者對於智慧煎蛋系統的辨識效能介於伯仲之間，不過以綜合評比而言 YOLOv11 略勝一籌，其 mAP@0.5 對鍋鏟為 0.739、握把為 0.64、平底鍋為 0.965、鍋底為 0.852、煎蛋為 0.995 以及雞蛋為 0.866。雖然初步的實驗結果仍有進步空間，不過也證實了將深度學習結合機械手臂之智慧煎蛋系統的可行性。

在未來，本研究將著重於優化系統與提升辨識效能，將擴大影像資料集或對訓練資料集做影像前處理以減輕資料集資訊複雜度，測試面對大量資料集模型辨識效能表現是否能有顯著提升。此外，我們將擴增熟度辨識模型的功能，使模型能辨識不同狀態的煎蛋，讓系統不局限於煎蛋而能完成各種蛋料理如炒蛋、太陽蛋或歐姆蛋等，更進一步擴增系統功能的多樣性。

## 參考文獻

- [1] Ku Azir, K. N. F., Baber, C., & Jusoh, M. (2015). Cooking guide: Direct and indirect form of interaction in the digital kitchen environment.
- [2] 陳楚恬，2021，「天天老是在外一剖析外食趨勢」，《政策研究指標資料庫》，國家實驗研究院科技政策研究與資訊中心，  
<https://pride.stpi.narl.org.tw/index/topic>
- [3] Ban, P., Desale, S., Barge, R., & Chavan, P. (2020). Intelligent Robotic Arm. In ITM Web of Conferences (Vol. 32, p. 01005). EDP Sciences.
- [4] Ma, E., Bao, Y., Huang, L., Wang, D., & Kim, M. (2023). When a robot makes your dinner: a comparative analysis of product level and customer experience between the US and Chinese robotic restaurants. *Cornell Hospitality Quarterly*, 64(2), 184-211.
- [5] Riaz, W., & Tanveer, A. (2012). Marketing mix, not branding. *Asian Journal of Business and Management Sciences*, 1(11), 43-52.
- [6] Pine, B. J., & Gilmore, J. H. (1998). Welcome to the experience economy (Vol. 76, No. 4, pp. 97-105). Cambridge, MA, USA: Harvard Business Review Press.
- [7] Chen, X., Zhu, Y., Zhou, H., Diao, L., & Wang, D. (2017). ChineseFoodNet: A large-scale image dataset for Chinese food recognition. *arXiv preprint arXiv:1705.02743*. Zhang, Y., Deng, L., Zhu, H., Wang, W., Ren, Z., Zhou, Q., ... & Wang, S. (2023). Deep learning in food category recognition. *Information Fusion*, 98, 101859.
- [8] Bossard, L., Guillaumin, M., & Van Gool, L. (2014). Food-101—mining discriminative components with random forests. In *Computer vision—ECCV 2014: 13th European conference, Zurich, Switzerland, September 6-12, 2014, proceedings, part VI 13* (pp. 446-461). Springer International Publishing.
- [9] Kawano, Y., & Yanai, K. (2013). Real-time mobile food recognition system. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1-7).
- [10] Noh, D., Nam, H., Gillespie, K., Liu, Y., & Hong, D. (2024). YORI: Autonomous Cooking System Utilizing a Modular Robotic Kitchen and a Dual-Arm Proprioceptive Manipulator. *arXiv preprint arXiv:2405.11094*.
- [11] Zhang, J., & Fang, X. (2017). Challenges and key technologies in robotic cell

- layout design and optimization. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 231(15), 2912-2924.
- [12]Wang, G., Chen, Y., An, P., Hong, H., Hu, J., & Huang, T. (2023). UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors*, 23(16), 7190.
  - [13]Wang, C. Y., Yeh, I. H., & Liao, H. Y. M. (2024). Yolov9: Learning what you want to learn using programmable gradient information. *arXiv preprint arXiv:2402.13616*.
  - [14]Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*.
  - [15]Khanam, R., & Hussain, M. (2024). YOLOv11: An Overview of the Key Architectural Enhancements. *arXiv preprint arXiv:2410.17725*.
  - [16]Davis, J., & Goadrich, M. (2006, June). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning* (pp. 233-240).
  - [17]Olorunshola, O. E., Irhebhude, M. E., & Ewwiekpaefe, A. E. (2023). A comparative study of YOLOv5 and YOLOv7 object detection algorithms. *Journal of Computing and Social Informatics*, 2(1), 1-12.

靜宜大學

資訊工程學系

利用機械手臂煎蛋系統  
深度學習於

西元二〇二四年十二月