利用機器學習辨識手部與虛擬控制系統/利用深度學習辨識手部並進行猜拳

指導教授： 余松年 教授　學生：喻奕程、黃程榆

摘要

本專題利用 MediaPipe 框架中的手部姿態辨識技術，結合機器學習的方法，實現了基於手勢的精準虛擬滑鼠操控。系統能夠在多維度上有效解碼使用者手勢，支持精確的滑鼠位置控制、模擬點擊、滾輪捲動等功能。實現高效且無接觸的交互操作，進一步提升人機界面的流暢度與實用性。

與此同時，我們還利用深度學習實現手勢辨識，辨識剪刀、石頭、布三種不同類型的手勢。選擇MobileNetV2作為基礎模型，經過數據增強以增加數據多樣性、類別權重調整以解決類別不平衡問題，並應用了凍結、學習率調整及早停法等優化技術以確保模型訓練的穩定性與泛化能力。此外。以強化學習的方式作為猜拳的邏輯，經由不斷的跟電腦交手，實時對戰來持續學習，並根據勝負平情況動態更新Q表格，以優化電腦的策略選擇。

背景

MediaPipe是以圖形化的數據流（Dataflow Graph）為核心架構，支援多模組的即時處理。Hand Tracking模組通過精確捕捉手部21個關鍵點的位置，提供手掌和手指的即時3D位置資訊。

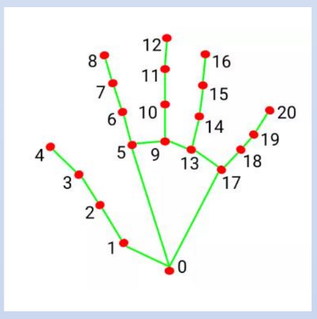

圖1、偵測手掌的21個節點

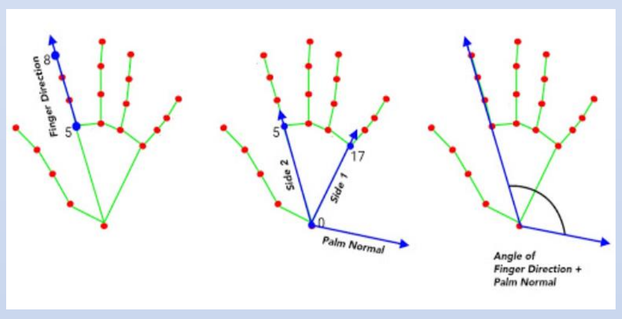

圖2、節點如何計算角度

這些關鍵點構成了骨架模型，使得手部動作能夠精確地被解析和跟踪。

而PyAutoGUI是一個圖形化用戶界面（GUI）自動化控制框架，其核心在於提供高層次的GUI操作自動化功能，通過程式化指令來模擬並精準執行用戶在圖形界面上的行為。它支援低層圖形事件的生成，同時也提供了對操作流程的序列化控制，以數據驅動方式構建自動化工作流。並支援動態參數調整（如移動速度、加速度），以實現更平滑的滑鼠移動過程。

MobileNetV2是基於卷積神經網路（CNN）的深度模型，其核心引入了“逆殘差結構”和“深度可分離卷積”。逆殘差結構讓模型在低維度數據學習後能保留更多重要特徵，減少記憶體占用；而深度可分離卷積則將卷積操作分解成獨立的深度卷積和點卷積，從而大幅減少計算量並提高運行速度。

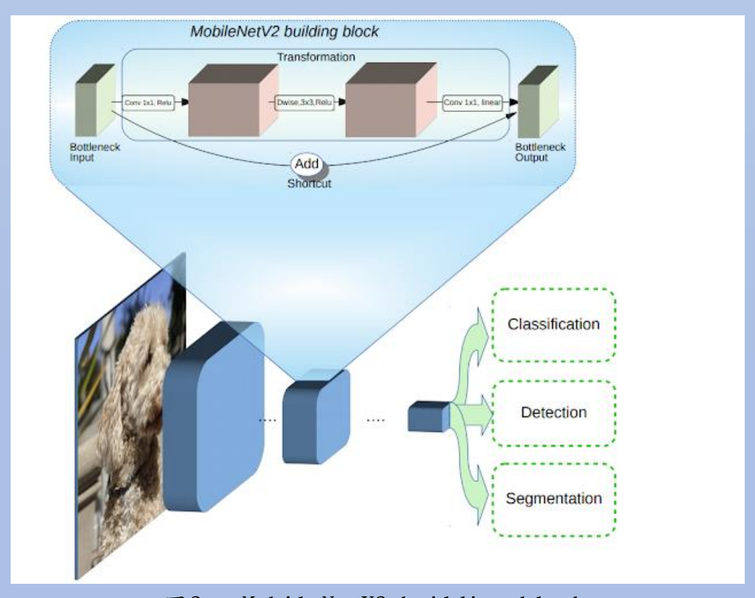

圖3、 MobileNetV2 building block

強化學習（Reinforcement Learning, RL）是一種機器學習方法，旨在讓代理（agent）通過試驗和錯誤來學習如何在環境中做出決策。Q-learning 是一種無模型的強化學習演算法，通過學習動作的價值來指導代理的行為。它使用 Q 表（Q-table）來記錄每個狀態-動作對的預期回報，並不斷更新以優化策略。

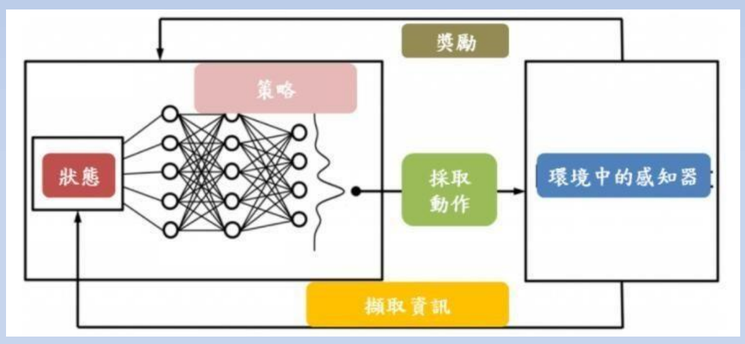

圖4、強化學習流程圖

研究方法

使用HandLandmark模型取得手掌關節的關鍵節點位置，針對這些節點資料，我們計算相鄰節點間的角度和距離，以確保能夠準確地解析出特定手勢的細微變化。

為每一種手勢進行角度測量。針對其他手指的彎曲或伸展情況則建立了條件判斷，利用PyAutoGUI將識別到的手勢映射為滑鼠和滾輪控制指令。將食指的指尖座標映射到螢幕的對應座標上，達成滑鼠移動；根據特定手勢，如「拳頭」的判斷，當雙手握拳時透過pygetwindow最小化當前活躍視窗。

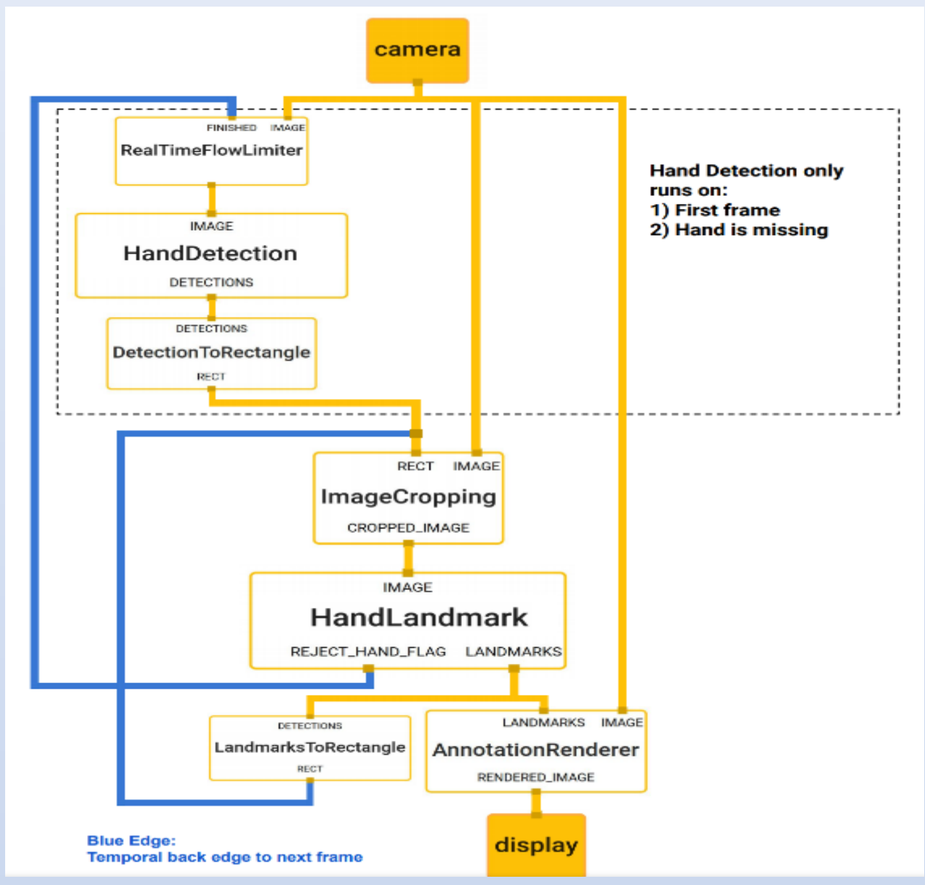

圖5、手部辨識與坐標點計算的過程

使用簡單背景的資料集，以 MobileNetV2 為基礎模型。資料經 OpenCV 處理、歸一化，並以 80% 訓練集、10% 驗證及測試集比例劃分。透過 ImageDataGenerator 增強數據多樣性，構建模型後先凍結再微調最後 20 層，以改善適應性，並利用類別權重及回調函數防止過度擬合。最終選取表現最佳的模型用於手勢識別。

研究成果

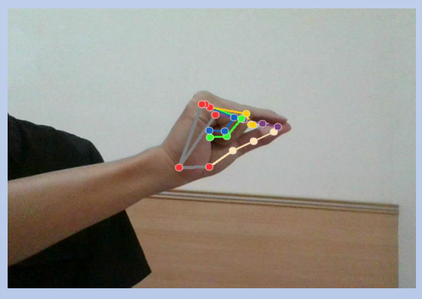

圖6、操控滑鼠的手勢點擊

圖7、操控滑鼠的手勢縮小視窗

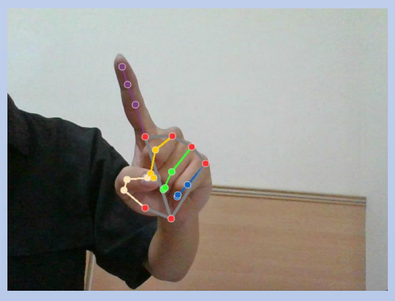

圖8、操控滑鼠的手勢移動

動作名稱/數據	平均反應時間(毫秒)	靈敏度(像素/秒)
move mouse	101.14	5.70
click	107.49	17.94
scroll up	101.81	0.20
scroll down	101.19	-0.20
minimize window	61.42	39.38

圖9、操控滑鼠的各個手勢數據

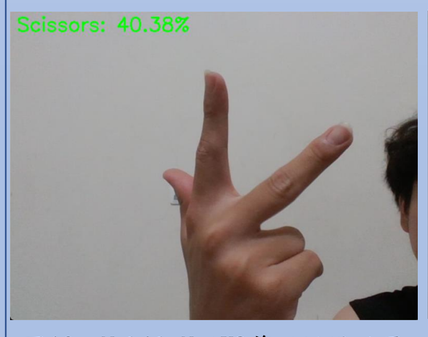

圖10、MobileNetV2剪刀辨識效果

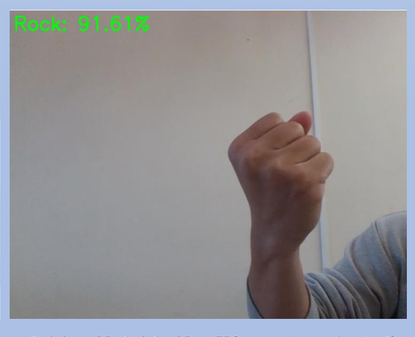

圖11、MobileNetV2石頭辨識效果

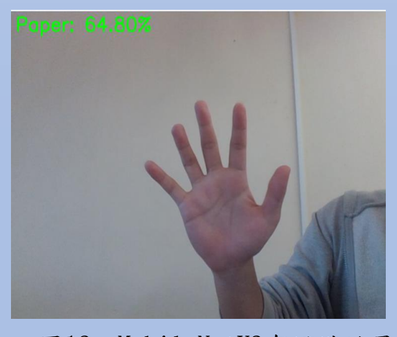

圖12、MobileNetV2布辨識效果

圖13、Training and Validation Loss

圖14、Training and Validation Accuracy

結論

我們的手勢辨識系統能有效追蹤手部動作，提供快速反應，適用於無觸控的人機互動應用，如模擬滑鼠操作、視窗控制等。系統結構包含手部檢測與標記。使用深度學習模型達成手勢分類時，雖然系統的反應速度良好，但不同操作的準確性仍有差異，模型有過擬合的徵兆，特定手勢的精確度有待提升，以確保在實際應用中的穩定性和精確度。