

智慧型人臉與行為辨識系統

沈宏勳、鄒博森、黃士熏、柯信汶、陳明僑*

摘要

近年來，人工智慧技術與雲端運算的蓬勃發展，已成為各行各業提升效率與創新能力的關鍵驅動力。智慧型人臉與行為辨識系統的應用範圍越來越廣泛，包括監控、健康管理及自動化流程等領域。本研究實作一套智慧型人臉與行為辨識系統，提供有效的解決方案來管理人員進出及行為記錄。本系統結合了 YOLO 模型的物件偵測技術、MediaPipe 骨架識別以及 AWS 雲端系統所提供的人臉辨識服務，實現了多層次的行為辨識系統。透過 MediaPipe 系統，能夠準確捕捉人體骨架資訊，並計算特定角度以進一步分析行為模式；而 AWS 人臉辨識功能則用於確保實驗室人員身份辨識的精確性，提供高效的進出記錄。這些技術的整合，使系統能以簡單、實用且具有擴展性的方式，達成高度自動化的行為分析。此外，本系統附有直觀的前端網頁介面，能展示經處理的影片，並提供統計圖表來呈現人員與行為的綜合數據分析；此介面讓使用者能快速檢視實驗室活動，進一步提升管理效率與安全性。

關鍵字：人臉辨識、行為辨識、雲端運算、人物追蹤。

沈宏勳，國立臺東大學資訊工程學系學生。Email: 11011271@gm.nttu.edu.tw

鄒博森，國立臺東大學資訊工程學系學生。Email: 11011236@gm.nttu.edu.tw

黃士熏，國立臺東大學資訊工程學系學生。Email: 11011248@gm.nttu.edu.tw

柯信汶，國立臺東大學應用數學系學生。Email: 11004132@gm.nttu.edu.tw

陳明僑(通訊作者)，國立臺東大學資訊工程學系助理教授。E-mail: joechen@nttu.edu.tw

Intelligent Face and Behavior Recognition System

Hung-Shiun Shen , Po-Sen Tsou , Shih-Hsun Huang , Xin-Wen Ke , Ming-Chiao Chen*

Abstract

In recent years, the rapid development of artificial intelligence and cloud computing has been a key driver of efficiency and innovation across various industries. Intelligent facial and behavior recognition systems have seen widespread applications in areas such as surveillance, health management, and automation. This paper presents the implementation of an intelligent facial and behavior recognition system designed to effectively manage personnel access and behavior records. The system integrates object detection using the YOLO model, skeleton recognition via MediaPipe, and facial recognition services provided by AWS Cloud, forming a multi-level behavior recognition system. Through MediaPipe, the system accurately captures skeletal information and calculates specific angles to analyze behavior patterns. Meanwhile, AWS facial recognition ensures precise identification of laboratory personnel, streamlining access control and record-keeping. By integrating these technologies, the system enables highly automated behavior analysis in a simple, practical, and scalable manner. Additionally, it features an intuitive web-based front-end that displays processed videos and statistical charts, providing a comprehensive analysis of personnel activities and behaviors. This interface enables users to efficiently review laboratory activities, further enhancing management effectiveness and security.

Keywords: Facial Recognition, Behavior Recognition, Cloud Computing, Human Tracking

Hung-Shiun Shen, Undergraduate Student, Department of Computer Science and Information Engineering, National Taitung University. E-mail: 11011271@gm.nttu.edu.tw

Po-Sen Tsou, Undergraduate Student, Department of Computer Science and Information Engineering, National Taitung University. E-mail: 11011236@gm.nttu.edu.tw

Shih-Hsun Huang, Undergraduate Student, Department of Computer Science and Information Engineering, National Taitung University. E-mail: 11011248@gm.nttu.edu.tw

Xin-Wen Ke, Undergraduate Student, Department of Applied Mathematics, National Taitung University. E-mail: 11004132@gm.nttu.edu.tw

Ming-Chiao Chen (Corresponding Author), Assistant Professor, Department of Computer Science and Information Engineering, National Taitung University. E-mail: joechen@nttu.edu.tw

壹、前言

隨著人工智慧技術的迅速發展，其應用範圍已廣泛涵蓋日常生活與產業領域，包括監控系統、自動化管理及智慧城市建設等。在實驗室環境中，如何有效監控人員進出及分析行為模式，成為提升管理效率與安全性的重要課題。然而，許多現有解決方案對硬體資源與技術門檻的需求較高，使得小型實驗室難以負擔，進而限制了技術的普及化應用。

本研究的動機源於希望為實驗室提供一套成本較低且易於部署的智慧型人臉與行為辨識系統。透過結合現成的人工智慧工具與雲端服務，例如 YOLO 的物件偵測技術、MediaPipe 骨架識別及 AWS 的人臉辨識服務，實現高效的行為與身份辨識功能。

為了讓管理者能即時了解實驗室的運作狀況，本系統設計了一個直觀的前端使用者介面，用於展示經處理的影像及統計分析數據。此設計不僅降低了技術使用門檻，也提升了系統的實用性與可擴展性。藉此研究，我們希望探索如何在資源有限的情況下，透過整合現有技術，實現一套符合實際需求的智慧型監控系統，並為小型機構的智慧管理提供具體參考。

貳、文獻探討

一、目標檢測與追蹤

Zhang Y.等人(Zhang et al., 2021)在其研究中提出了 ByteTrack，一種專注於提高多目標追蹤精度的算法。其核心是通過匈牙利演算法(Hungarian Algorithm)匹配檢測框與追蹤 ID，即在物體追蹤中，將追蹤框與物體關聯起來。實現跨幀目標的連續追蹤。該算法的核心創新在於對每個檢測框進行關聯，即便是那些具有較低信任度的檢測框也被納入考慮，從而大幅提升了追蹤的穩定性和準確性。特別是在擁擠場景中，ByteTrack 能有效區分重疊對象，為目標的長期追蹤提供了支持。這對本研究的目標追蹤模組有重要啟示，特別是在使用 MediaPipe 進行骨架提取時，軌跡的穩定性可以直接影響行為辨識的準確性。

Redmon J.等人(Redmon et al., 2015)則提出了 YOLO (You Only Look Once)演算法，這是一種將目標檢測視為迴歸問題的創新方法。YOLO 將整個圖像劃分為網格，直接預測每個網格中物體的位置與類別。與傳統 R-CNN 系列方法相比，YOLO 透過單一卷積神經網路 (CNN) 完成端到端學習，不需經過候選區域生成步驟，大幅提高了目標檢測的速度與效率。YOLO 的主要貢獻包括其簡單高效的結構和出色的實時檢測能力。例如，其全卷積網路 (Fully Convolutional Network, FCN) 能處理任意大小的圖像，並靈活應對不同尺寸的物體。此外，YOLO 的損失函數設計進一步提升了預測框的定位精度，這對於行為辨識中需要高精度骨架數據的應用場景尤為重要。

在本研究中，採用了 Huang A. 等人(Huang et al., 2023)提出的 YOLOv8 做為目標檢測模型，專為實時應用場景設計。相比於之前的版本，YOLOv8 在速度和準確性方面均有顯著提升，尤其是在複雜場景下對小型物體的檢測能力得到了加強。該模型的改進主要包括更高效的特徵提取網路和優化的損失函數，使其在推理速度上達到了與同類方法的領先水平。本研究中，YOLOv8 作為目標檢測模組的核心技術，不僅提供了穩定的檢測輸出，還為後續的追蹤和骨架提取提供了準確的輸入數據。

二、骨架提取

Bazarevsky V. 等人(Bazarevsky et al., 2020)提出的 BlazePose 架構，是 Mediapipe 框架中的一個組件，一種針對移動設備優化的人體姿態估計方法。隨著智能設備的普及，對於即時、精確的姿態檢測需求不斷增長，尤其是在健康監測、運動分析、虛擬現實（VR）等領域。然而，傳統的姿態估計方法大多依賴於高性能的伺服器端處理，並且要求較高的計算資源，這使得它們難以在移動設備上實現。因此，BlazePose 的提出便是為了解決這一挑戰，並提供一種高效、輕量級的即時人體姿態檢測解決方案。BlazePose 引入了多階段的推理架構，旨在平衡計算效率和準確性。該架構首先使用一個輕量級的姿勢區域檢測器來快速確定人體所在的位置，這樣可以大幅減少後續處理的計算負擔。接下來，BlazePose 利用深度神經網路進行關鍵點迴歸，精確地預測出人體的 33 個關鍵點，如圖 1，這比傳統的 17 點姿態估計提供了更高的細節，從而提升了姿態檢測的準確性。

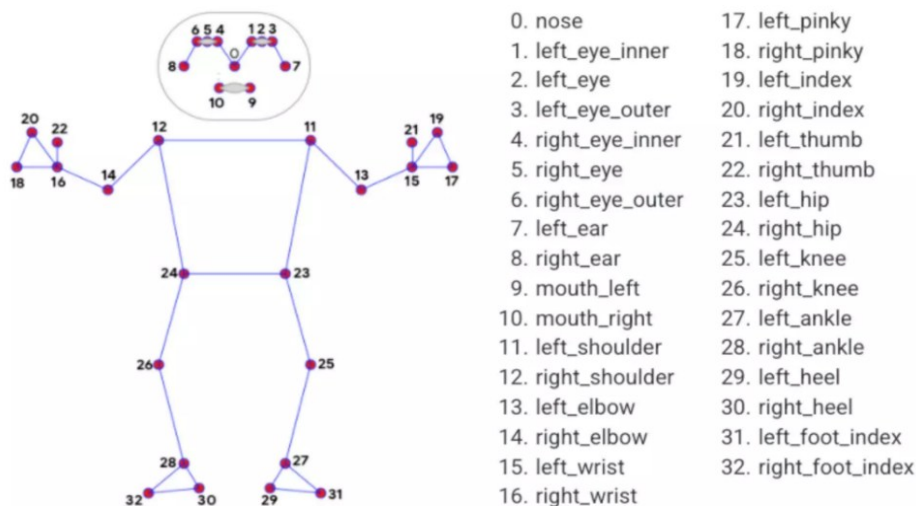


圖 1 BlazePose 骨架節點(Bazarevsky et al., 2020)

三、行為辨識方法

Schlegel K. 等人(Schlegel et al., 2024)提出了一種基於關節角度的行為辨識方法。該方法依托於人體骨架的結構，通過計算人體各關節之間的相對角度來提取動作特徵。

透過計算人體骨架中各關節之間的角度，這些角度特徵不僅可以準確反映出每個關節的運動範圍，還能夠捕捉到全身的動作模式。具體而言，這些角度是通過向量內積來計算的，這樣可以減少因攝影機視角變化而造成的影響，從而提高辨識準確性。人體行為往往具有強烈的時序性。為了有效捕捉行為中的時間依賴關係，作者選擇了長短期記憶網絡 (LSTM) 來建模時間序列特徵。LSTM 能夠捕捉長期依賴關係，對於行為辨識尤其重要，因為許多行為（如走路、跑步等）需要在時間上進行長時間的跟蹤和分析。在特徵提取後，作者使用多層感知機 (MLP) 作為分類器來區分不同的行為類型。該分類器對於提取出的角度特徵進行學習，並最終完成動作的識別。

Sengar S. 等人 (Sengar et al., 2024) 提出了一種基於 MediaPipe 的高效人體姿態估計方法，專注於解決實時應用中因快速運動和遮擋而引起的準確性與效率問題。他們開發了一種多層次卷積神經網路 (CNN) 模型，結合角度計算與空間關聯性，提升了人體標誌點的準確性與穩定性。此外，針對遮擋情境，他們採用先驗知識和時序信息推測關鍵點位置，改善了模型的穩健性。其研究著重在肩膀和肘部的關鍵點，姿勢估計管道計算三個關鍵點之間的角度，分別為肩膀、肘部和手腕。使用反正切函數計算角度，所得角度從弧度轉換為度數以便進一步分析。該研究的模型可基於人體的解剖結構，將未遮擋的關鍵點作為參考，推測遮擋關鍵點的可能位置。例如，若肩膀和肘部的關鍵點已知，則可以根據人體幾何推測手腕的位置。在計算效率方面，作者對模型進行輕量化優化，使其能夠在嵌入式和移動設備上以低延遲運行。實驗結果顯示，該方法在動態場景下的平均誤差比傳統方法降低 15%，並提升了處理速度。此研究在人體姿態估計的準確性方面取得了顯著的進步，與現有解決方案相比，交並比 (IoU) 指標提高了 20%。此外，這些增強功能還使處理時間減少了約 30%，證實了整合到 MediaPipe 框架中的最佳化的有效性。這些進步有助於在需要即時數據處理的領域提供更靈敏和更精確的應用程式。

參、研究方法

如圖 2 的系統架構圖所示，本研究將系統主要分四個子系統，包括：管理介面、人臉辨識、行為辨識以及資料管理，管理介面用於查看及查詢影片紀錄，人臉辨識與行為辨識部分將影片處理，並將結果存入資料庫。以下將對系統的功能和實施方法進行詳盡的描述。

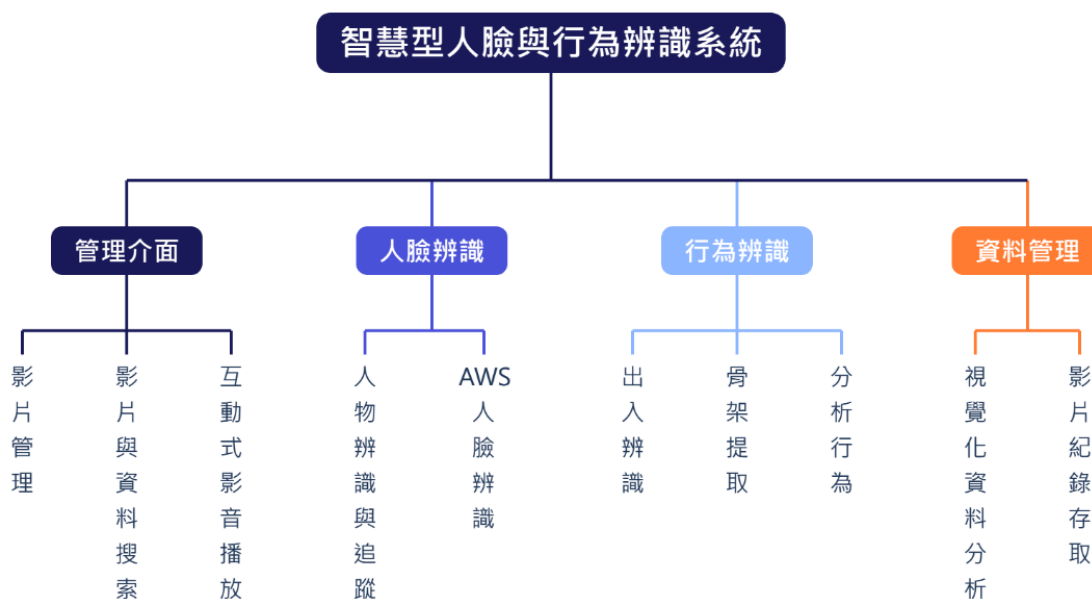


圖 2 系統架構圖

一、管理介面

本研究採用現代化網頁開發技術與方法論進行系統實作。在前端技術選擇上，採用 Alpine.js 作為核心框架，主要用於處理前端狀態管理和響應式更新，此框架的選擇基於其輕量級特性和優秀的性能表現。介面設計方面使用 Tailwind CSS 框架，通過其原子化的 CSS 類實現響應式設計，確保系統在不同設備上都能提供一致的使用體驗。

在後端技術架構上，選用 PHP 作為主要的伺服器端程式語言，透過實作 RESTful API 架構實現前後端的數據交互。數據存儲採用混合式架構，使用檔案系統管理影片資源，同時通過 JSON 格式存儲行為識別結果和系統配置信息。系統監控模組則採用即時數據收集和分析方法，實現對 CPU 使用率、記憶體使用情況和硬碟空間的動態監控。

功能實現方面，影片管理模組採用分層設計 pattern，實作包含檔案處理層、業務邏輯層和展示層。其中，檔案處理層負責處理影片上傳、存儲和格式驗證；業務邏輯層處理影片分類、標記和管理；展示層則提供直觀的使用者介面。行為識別模組採用事件驅動架構，通過非同步處理方式實現識別結果的即時更新和展示。為確保系統安全性，實作了完整的訪問控制機制，包括 Session 管理、權限控制和安全令牌驗證，同時對檔案上傳和存儲進行嚴格的安全控制。

系統優化方面採用多層次策略，包括前端資源延遲加載、後端快取機制實現、以及資料庫查詢優化等方法。通過這些技術的整合應用，成功建構了一個高效能、易維護且安全可靠的智慧型影像分析系統。此外，為確保系統的可維護性，實作了完整的日誌記錄和系統診斷機制，便於系統的持續維護和優化。

(一)影片管理

本研究設計了一套完整的影片管理機制，採用雙軌分類方法來處理影片資料。系統建立兩個主要資料夾，分別存放原始影片(videos/)和已處理影片(pre_videos/)，並提供狀態標記功能使影片可在這兩個分類間轉移。為確保系統安全性與效能，影片上傳機制整合了多項驗證功能，包含檔案類型驗證、安全性檢查，以及自動化的檔案權限設定。特別在檔案命名方面，採用時間戳記命名規則，有效提升檔案管理效率。

(二)影片與資料搜索

搜索功能的實作著重於多維度資料整合，建立了三種主要的搜索機制。首先是時間範圍搜索，支援精確至分鐘的時間區間篩選，可同時搜尋原始和已處理的影片檔案。其次在行為辨識搜索方面，系統整合了行為記錄檔中的 JSON 資料，能夠根據特定行為類型（如 Standing、Sitting 等）進行篩選。第三是人臉辨識搜索，除了支援已知人名的搜索外，也包含未知人員的篩選功能，並提供進出事件類型的分類搜索，搜索結果中更整合了辨識信心度資訊，提供更精確的資料參考。

(三)互動式影音播放

在播放器的實作上，我們整合了多項互動控制功能以提升使用體驗。播放控制除了基本的播放/暫停功能外，特別加入了 10 秒快進/倒退功能，並支援 0.5x 至 2.0x 的播放速度調整。為提高操作便利性，系統實作了鍵盤快捷鍵功能，包含空白鍵控制播放/暫停，以及左右方向鍵控制影片進度。在播放資訊顯示方面，即時顯示當前播放時間、總時長等資訊，並支援影片時間戳記的顯示。另外，也額外設計點選撥放功能，提供使用者點選影片行為發生時間點，並觀看該時間點的影片片段。特別考慮到跨裝置使用需求，播放器支援自適應縮放，確保在不同螢幕尺寸下都能維持良好的顯示效果。

二、人臉辨識

本系統實作了一個人臉辨識與追蹤方法，其流程如圖 3。首先，在人臉偵測上本系統採用了 YOLOv8n-face 模型作為人臉檢測器，它能夠快速準確地定位影片中偵測到的人臉區域。在偵測到人臉後，輸出一組矩形框，標識出圖像中存在的人臉位置，這些矩形框由兩組座標(x1, y1), (x2, y2)來表示，並將擷取出的人臉傳至 AWS Rekognition 進行人臉辨識，再將人臉辨識的結果視為唯一 ID，再使用 ByteTrack 進行持續追蹤。



圖 3 人臉辨識與追蹤流程

(一)人物辨識與追蹤

人物辨識與追蹤分別由 YOLOv8 和 ByteTrack 組成，這兩部分的整合實現了從目標檢測到多幀追蹤的完整流程，適合用於多目標監控與行為分析的應用場景。

YOLOv8 是系統中的物件檢測模組，其主要任務是檢測每一幀畫面中的所有物體，並對其進行分類與定位。在系統中，YOLOv8 被初始化為一個專門負責人物檢測的模型，並通過以下步驟完成檢測任務：首先，YOLOv8 以高效的深度學習演算法結構處理每幀畫面，產生檢測框及其對應的物件類別和置信度。為了確保結果的準確性，程式設置了置信度和 IoU (Intersection over Union) 的閾值，過濾掉低置信度或重疊過高的檢測框。接著，系統進一步過濾 YOLOv8 的輸出，僅保留類別 class_id 為 0 的人物目標，為後續的追蹤過程提供準確的輸入資料。

完成目標檢測後，檢測框的資料被傳遞至 ByteTrack，用於進行跨幀的目標追蹤。在每幀中，ByteTrack 會先處理高置信度的檢測框，將其與上一幀的追蹤結果進行匹配，確保目標的 ID 在多幀間保持一致。對於未能匹配的低置信度檢測框，ByteTrack 會進一步進行二次匹配，這使得演算法在目標出現短暫遮擋或低可見性的情況下仍能保持穩定的追蹤。並加入了平滑處理模組，進一步優化了 ByteTrack 的追蹤結果。檢測框的平滑處理減少了因畫面抖動或檢測器不穩定而導致的追蹤 ID 的變動，提升了整體追蹤的穩定性與準確性。最後，每個追蹤目標都被分配了一個唯一的 tracker_id，用於標記該目標的跨幀身份，這些 ID 與對應的檢測框一同輸出，並進行可視化標註。

(二)AWS 人臉辨識

在人臉辨識方面，嘗試使用穩定性較高的 AWS 雲端運算服務來取代 YOLOv8 模型，且由於攝影機位於較高的位置，導致畫面中擷取到的人臉解析度較不理想，經過多次測試後，使用 AWS Rekognition 的辨識準確率略高於使用單一 YOLOv8 模型的效果，因此最終採用 AWS Rekognition 來做人臉辨識。首先在 Rekognition 中創建一個集合，將每個想要辨識的人取大約 5~10 張包含正面以及上下左右約 30° 角的照片，並且將同一人以相同的標籤(External ImageId)做標記，最後將這些照片加入集合中。系統根據上述使用 YOLO 偵測到的人臉矩形座標，使用 numpy 將其切割成包含人臉的小圖像，將這些圖像以 JPEG 格式的字節流資料發送給

AWS Rekognition 進行辨識，Rekognition 會與集合中的人臉做比對，若有符合的人臉將會回傳對應的 External ImageId 和信心程度。

三、行為辨識

本研究的行為辨識系統，使用 MediaPipe Pose 進行骨架提取，結合幾何特徵分析，實現了人體行為辨識。本系統能有效辨識人類的基本行為模式，包括站立、坐下及行走等狀態。

(一) 出入辨識

在進出實驗室判斷中，系統根據物體移動時座標的改變判斷移動方向，並在門邊設立一條可觸發的虛擬檢測線，如圖 4，依據不同移動方向有不同觸發方式，當物體向左移時，以物體左邊界觸發檢測線視為進入，當物體向右移時，以物體右邊界觸發檢測線視為離開，並且在觸發加入了防誤觸機制，避免有人在檢測線中快速來回進出造成誤判。



圖 4 虛擬檢測線

(二) 骨架提取

骨架檢測部分，採用了 MediaPipe Pose 解決方案，然而 MediaPipe 預設只能在一個畫面中提取一個人的骨架，為使 MediaPipe 可以進行多人骨架提取，本系統在骨架提取上結合 YOLO11s 模型進行人物提取。並設置為最高複雜度模式以提升定位精確度。本系統將 YOLO11s 檢測與追蹤信任度設定為 0.7，結合關鍵點平滑處理，為後續行為分析提供高質量的骨架數據基礎。另外系統特別關注膝、臀關鍵節點的可見度，並設置 0.6 的最低可見信任度，確保姿態分析的可靠性。

(三) 行為分類

行為分析算法部分，本系統定義了三種行為，分別為坐下、站立、行走，利用 YOLO11s 模型提取人物框，再使用 MediaPipe 進行骨架提取。最後使用演算法進行行為分類，本演算法的核心基於 MediaPipe 提供的關鍵點檢測，這些關鍵

點包括肩膀、臀部、膝蓋和腳踝等部位。透過計算這些點的相對位置與角度，演算法能夠判斷當前姿勢的特徵。行為辨識流程如圖 5。

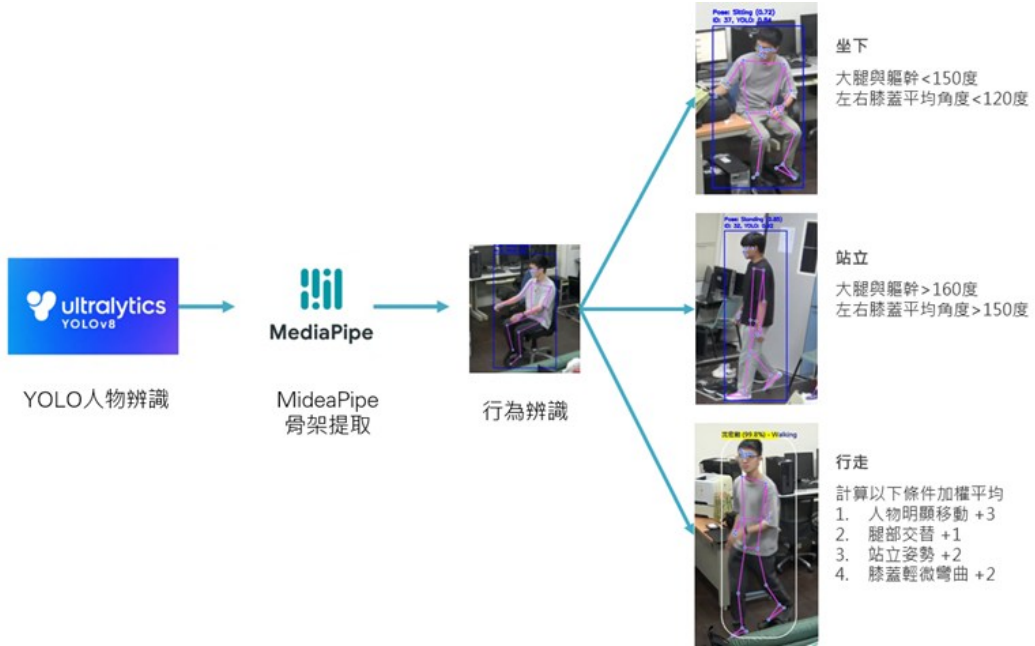


圖 5 行為辨識流程圖

為了實現行為判斷，演算法需要多個人體關鍵點的座標資訊，以得知膝關節與髖關節角度。本研究主要使用以下五個關節點，分別為臀部 (H)、膝蓋 (K)、腳踝 (A) 和左右肩膀 (LS 和 RS)。

膝關節角度是由臀部 (H)、膝蓋 (K) 和腳踝 (A) 三個關鍵點所構成。假設這三個關鍵點的座標分別為 $H = (x_H, y_H)$ 、 $K = (x_K, y_K)$ 、 $A = (x_A, y_A)$ ，膝關節的角度 θ_K 可以使用以下的向量計算公式得出，膝關節角度 θ_K 可以反映膝蓋的彎曲程度。

$$\theta_K = \cos^{-1} \left(\frac{H - K}{\|H - K\|} \cdot \frac{A - K}{\|A - K\|} \right)$$

髖關節角度 θ_{hip} 則是由肩膀中點 (SM) 與臀部中點 (HM) 和兩膝蓋中點 (KM) 之連線所構成，描述軀幹與大腿之間的夾角。

$$SM = \left(\frac{x_{LS} + x_{RS}}{2}, \frac{y_{LS} + y_{RS}}{2} \right)$$

$$HM = \left(\frac{x_{LH} + x_{RH}}{2}, \frac{y_{LH} + y_{RH}}{2} \right)$$

$$KM = \left(\frac{x_{LK} + x_{RK}}{2}, \frac{y_{LK} + y_{RK}}{2} \right)$$

$$\theta_{hip} = \cos^{-1} \left(\frac{SM - HM}{\|SM - HM\|} \cdot \frac{KM - HM}{\|KM - HM\|} \right)$$

1. 站立行為

站立行為指膝關節與髖關節幾乎完全伸直，如圖 6。當髖關節接近 180° 時，表示身體接近直立。因此若角度滿足以下條件則可以判斷為站立，分別為膝關節角度 $\theta_K \geq 150^\circ$ ，與髖關節角度 $\theta_{hip} \geq 160^\circ$ 。

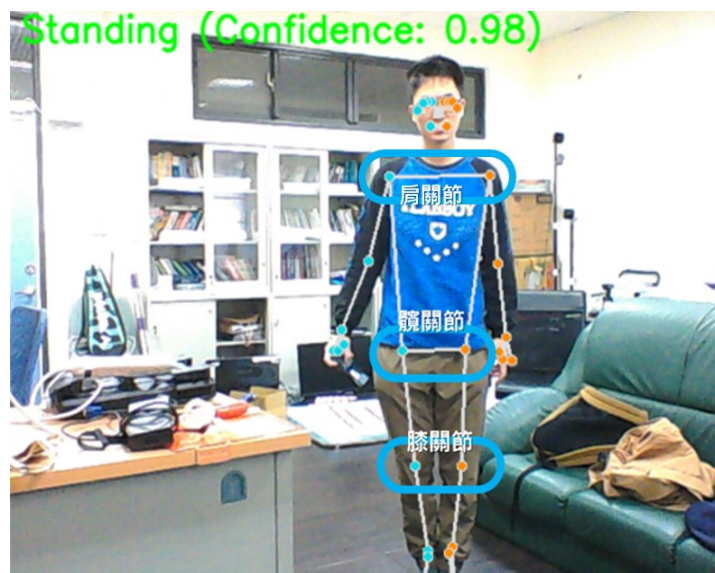


圖 6 站立行為示意圖

2. 坐下行為

坐下行為是膝關節明顯彎曲且髖關節形成一定夾角的狀態，如圖 8。膝關節 θ_K 與髖關節 θ_{hip} 呈明顯彎曲狀態。因此若角度滿足以下條件則可判斷為坐下，分別為膝關節角度 $\theta_K \leq 120^\circ$ ，與髖關節角度在範圍 $80^\circ \leq \theta_{hip} \leq 120^\circ$ 之間。

3. 行走行為

行走行為採用多個條件判定，每一幀的判定會依照以下幾個主要特徵來計算行走分數，這個分數用來量化每一幀是否符合行走的特徵，當某幀的行走分數大於或等於 6 分時，表示該幀符合行走的特徵，並記錄在移動歷史中。並根據更新過的移動歷史，演算法會判斷人物是否正在行走。具體的判定方式為計算最近 3 幀的行走判定結果，並判斷若其中有超過 2 幀符合行走的特徵，則視為人物正在行走。

行走判定條件分別為(1)人物明顯移動(+3 分)，檢測人物整體關節的移動幅度；(2)腿部交替(+1 分)，檢測雙腳踝的垂直位置差異顯示出行走時的擺腿動作；(3)站立(+2 分)，軀幹保持接近垂直，即站立時的直立狀態；(4)膝蓋有輕微彎曲四個條件(+2 分)，以下將詳細說明每個條件的判定方式：

(1) 人物明顯移動

將透過計算各關節的平均移動幅度來判斷人物是否在運動。對於某關節 P_i ，當前幀與上一幀的位置變化為 $d(P_i)$ 。

$$\begin{aligned} d(P_i) &= \|P_i(t) - P_i(t-1)\| \\ &= \sqrt{(x_i(t) - x_i(t-1))^2 + (y_i(t) - x_i(t-1))^2} \end{aligned}$$

對所有關節取平均移動量，記為 \bar{d} ，其中 n 為關節數量。關節移動判定需滿足條件 $\bar{d}_{movement} > \epsilon_{movement}$ ， $\epsilon_{movement}$ 為移動檢測的閾值，設定為 0.005。

$$\bar{d}_{movement} = \frac{1}{n} \sum_{i=1}^n d(P_i)$$

(2) 腿部交替運動

通過雙腳踝的垂直位置差異來檢測。假設 d_{left_leg} 和 d_{right_leg} 分別是左腳踝與右腳踝的垂直座標，其差異 Δd_{leg} 如下所示，腿部交替運動判定需滿足條件 $\Delta d_{leg}(t) > \epsilon_{leg}$ ，其中 ϵ_{leg} 為腿部交替閾值，設定為 0.05。

$$\Delta d_{leg}(t) = |d_{left_leg} - d_{right_leg}|$$

其中，(3) 站立與 (4) 膝蓋有輕微彎曲部分，演算法於前面站立與坐下行為中已提及，不再贅述，其中 (4) 膝蓋彎曲角度範圍為 $120^\circ \sim 170^\circ$ ，滿足此條件時可確定人物並非為坐下狀態。

四、資料管理

本研究開發一套整合性的視覺化資料分析系統，專注於處理和分析人體行為數據。使用者可以從前端介面下拉式選單中選擇特定日期時間的記錄檔案進行分析。這個選擇介面會自動掃描 behaviorLog 目錄，列出所有符合命名規則的 JSON 檔案，並將檔案名稱轉換為更易讀的日期時間格式顯示。當使用者選擇了特定的檔案後，系統會根據選擇的分析類型，將檔案名稱作為參數傳遞給相應的分析腳本，進而進行後續的數據分析和視覺化處理。

整個過程形成了一個完整的數據流：從原始記錄的存儲、處理、到最終的分析展示。此設計使得系統能夠有效管理和利用影片記錄數據，同時也為擴展更多分析功能提供了便利的基礎架構。透過這種模組化的設計，系統可以輕易地添加新的分析類型，或是修改現有的分析方法，而不需要改變基礎的數據存取架構。

(一) 視覺化資料分析

系統採用 PHP 作為後端開發語言，結合 MySQL 資料庫進行數據管理，並整合 Chart.js 框架實現視覺化功能。

1.分析系統類型

主要提供了六種不同的分析類型：30 分鐘行為變化分析(觀察行為 30 分鐘內的變化趨勢)；行為持續時間分析(分析各行為的持續時間分布)；主體行為分布分析(分析不同主體的行為模式)；行為轉換頻率分析(研究行為轉換的模式和頻率；各時間點行為人數分析)統計各時段的行為人數；連續行為模式分析(分析行為的持續性特徵)；其中連續行為模式分析如圖 7 所示。

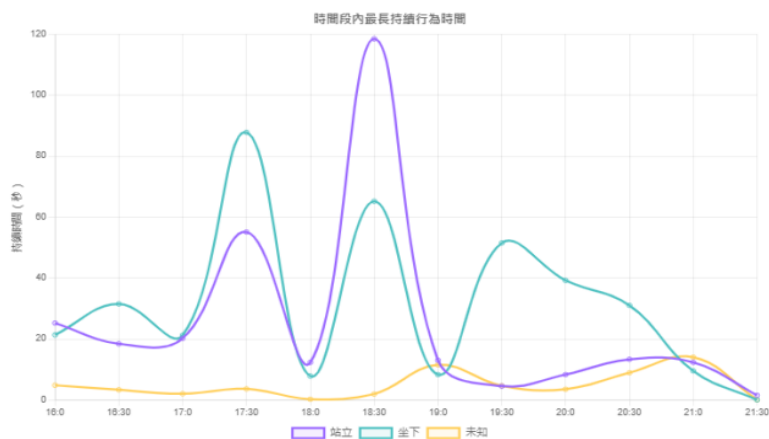


圖 7 連續行為模式分析圖

2.資料處理流程

系統採用分層處理架構設計：第一層進行數據預處理，負責數據格式驗證和清洗工作；第二層執行數據轉換，將 JSON 格式轉換為關聯式數據庫結構；第三層進行數據分析，產生統計結果和視覺化所需的數據格式。

(二)影片紀錄存取

影片紀錄存取的實現主要通過以下流程進行：首先，系統會將行為記錄以 JSON 格式保存於 behaviorLog 目錄下，檔案命名採用 "behavior_log_YYYYMMDD_HHMMSS.json" 的格式，這種命名方式確保了檔案的時序性和可追溯性。接著，透過 process_behavior_log.php 進行檔案處理，這個檔案會執行以下步驟：建立與 MySQL 資料庫的連線、讀取指定的 JSON 檔案內容、解析 JSON 數據、清空舊有的資料表內容，最後將新的數據逐條插入資料表中。

在數據存儲方面，每條記錄都包含了完整的行為信息：包括主體 ID (subject_id)、行為者名稱 (name)、行為類型 (behavior)、開始幀 (start_frame)、結束幀 (end_frame)、時間戳記 (timestamp) 以及來源檔案名稱 (log_file)。

肆、結果與討論

一、使用者介面

使用者介面主要分為(1)影片選擇區、(2)影片顯示區、(3)影片搜索區、(4)資料顯示區、(5)資料分析入口五個部分，如圖 8 所示。



圖 8 使用者介面

影片選擇區主要是提供使用者選擇影片的地方，在這區塊會將放入影片儲存資料夾的所有影片以排列顯示出來，影片顯示區則在使用者選擇影片後提供使用者觀看影片進行開始、暫停、快轉與影片進度調整的功能，而在影片顯示區下方的影片搜索區使用者可以透過影片產生的區間去搜索在使用者選擇區間的影片，右方則為資料顯示區並提供網頁伺服器的系統資訊與人臉與行為辨識資料，此區域呈列影片中行為人臉辨識的所有資料，使用者可以藉由搜索區間與點擊影片的方式來顯示資料，並藉由點擊資料來查看人臉與行為事件發生的時間，最後放置資料入口將使用者導向後端資料分析的頁面。

二、人臉辨識

本研究結合 YOLOv8、AWS Rekognition、supervision ByteTrack，完成一個人臉辨識追蹤、進出實驗室的辨識系統。根據實驗結果(如圖 9)，在人臉辨識方面，即使畫面中的人臉偏小，但是系統採用 AWS Rekognition 的辨識準確率平均可達 85%，明顯優於單一使用 YOLOv8 模型的效果。在人物追蹤方面，因為攝影機角度或是畫面中多人重疊等問題，可能導致誤判造成被追蹤者與其 ID 不匹配。在進出實驗室的判斷方面，由於防誤觸機制的加入，使得準確率大幅提升。

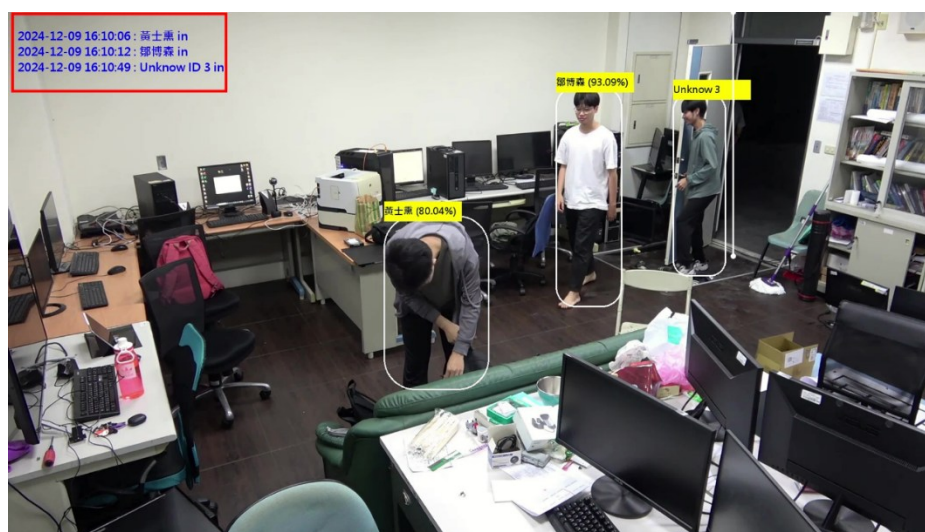


圖 9 人臉辨識展示圖

三、行為辨識

本研究針對站立、坐下及行走三種基本行為進行辨識，並且實作了基於 MediaPipe 與 YOLO 的多人骨架提取與行為辨識，實際辨識成果如圖 10 所示。

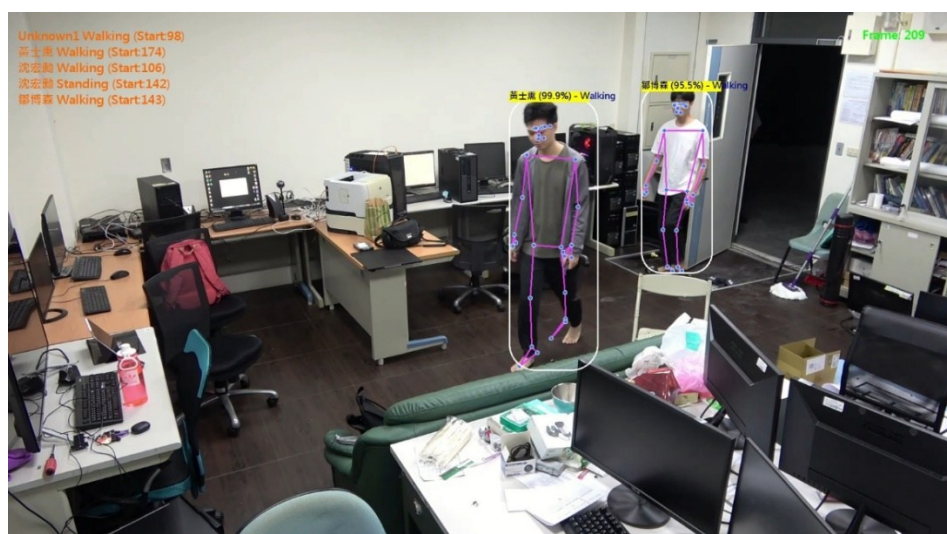


圖 10 行為辨識成果圖

實驗結果顯示，在站立行為的辨識方面，如表 1，系統於內容全為站立影片中的站立行為辨識率可達 92%。而在內容全為坐下影片中坐下行為的辨識率，如表 2，在全坐下影片中為 63%，顯著低於站立行為。這主要是由於採用了 2D 骨架繪製技術，該技術在拍攝者正面面向攝影機時，對坐下行為的關鍵動作捕捉效果較差。然而，當攝影機從側面拍攝時，系統的辨識效果明顯提升，顯示了角度對辨識結果的重要性。

表 1

內容全為站立影片中站立行為辨識率

行為	無法辨識	坐下	站立	總計	辨識率
辨識次數	163	190	4124	4477	92%

表 2

內容全為坐下影片中坐下行為辨識率

行為	無法辨識	坐下	站立	總計	辨識率
辨識次數	202	5194	2818	8214	63%

針對行走行為，即使人物在不同步態下移動，辨識結果依然具有相當的準確性。然而，當人物停止移動時，系統可能無法迅速做出反應，這在行為轉換的瞬間容易出現短暫的辨識失準。此外，在存在遮擋物的場景中，例如部分肢體被其他物件遮擋或人物處於擁擠環境，辨識的準確性大幅下降。這是因為遮擋會導致骨架追蹤的斷裂，進一步導致關鍵點錯位或跳動現象，進而影響辨識效果。

在簡單且無遮擋的環境中，本系統能夠穩定地辨識出站立、坐下及行走三種行為。然而，為了應對更加複雜的實際場景，未來需要進一步研究以提高遮擋情況下的辨識準確性，並改善辨識的即時性。此外，也可考慮結合 3D 骨架數據或其他深度學習模型，以提升系統在多角度、遮擋或動態背景下的辨識能力，使其更適用於真實世界中的應用場景。

四、資料庫管理

系統採用 LAMP (Linux/Apache/MySQL/PHP) 技術，實現了完整的開發環境整合。Apache 作為 Web 服務器，配合 PHP 的高性能運行時，為系統提供了穩定的執行環境。MySQL 的加入則為數據存儲和管理提供了強大支持，特別是在處理高併發訪問和複雜查詢場景時表現非常出色。

其次，在數據管理方面，系統透過 phpMyAdmin 實現了直觀的數據庫操作界面，大幅簡化了數據庫管理流程。透過精心設計的數據模型，包括行為記錄表 (behavior_records) 和分析快取表 (analysis_cache)，實現了高效的數據存取和分析處理。系統還採用 JSON 數據類型存儲複雜的行為詳情，既保持了數據結構的靈活性，又確保了查詢效率。

在性能優化層面，系統實施了多層次的優化策略。透過配置 PHP OpCache 和調整 MySQL 緩衝池大小，顯著提升了系統響應速度。實施的查詢優化機制，包括合理的索引設計和查詢快取策略，有效提高了數據檢索效率。此外，透過實現數據分區和定期維護機制，確保了系統在大規模數據處理場景下的穩定運行。

伍、結論

本研究的目的是建立一個適用於實驗室環境的智慧型人臉與行為辨識系統，該系統對實驗室中的人員進出與行為進行有效的監控。系統運用了人工智慧技術，如 YOLO、ByteTrack、AWS Rekognition 與 MediaPipe，成功實現了多項功能性目標。系統能透過人臉辨識準確記錄每位進出實驗室的人員，並檢測陌生人的出現，以提升實驗室的安全性。此外，系統利用骨架提取並進行行為分析，能夠監控人員的活動情況。

YOLO 被應用於人物檢測，作為一個快速且準確的物件偵測模型，能即時檢測出實驗室內的人員位置，接著由 ByteTrack 則負責追蹤人物的移動軌跡，透過其高效的追蹤演算法，系統能夠準確捕捉並持續分析人員的行為。AWS Rekognition 則提供人臉辨識功能，能夠辨別進出實驗室的人員身份，並識別是否有陌生人出入。最後 MediaPipe 提取人體的骨架，通過計算骨架角度與運動特徵來進行行為辨識，從而分析人員在實驗室內的活動情況。並將所有辨識數據儲存於系統資料庫中。根據行為辨識實驗結果，在全站立的影片中，站立行為辨識率達到 92%，而在全坐下影片中，因採用 2D 骨架提取，坐下的影片辨識率為 63%。

系統也提供了直觀且易於操作的網頁介面，使用者可以查看原影片與辨識後的影片，也可查詢進出紀錄與行為分析結果。系統提供了六種不同的圖表分析方式，使得使用者可以清晰了解不同時間段內人員的行為分佈情況。

參考文獻

- Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). BlazePose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204*. <https://arxiv.org/abs/2006.10204>
- Huang, A., Ju, X., Lyons, J., Murnane, D., Pettee, M., & Reed, L. (2023). YOLOv8: A high-performance object detection model. *arXiv preprint arXiv:2301.00501*. <https://arxiv.org/abs/2301.00501>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You only look once: Unified, real-time object detection. *arXiv preprint arXiv:1506.02640*. <https://arxiv.org/abs/1506.02640>
- Schlegel, K., Jiang, L., & Ni, H. (2024). Using joint angles based on the international biomechanical standards for human action recognition and related tasks. *arXiv preprint arXiv:2406.17443*. <https://arxiv.org/abs/2406.17443>
- Sengar, S., Kumar, A., & Ghosh, R. R. (2024). Efficient human pose estimation: Leveraging advanced techniques with MediaPipe. *arXiv preprint arXiv:2406.15649*. <https://arxiv.org/abs/2406.15649>
- Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X., (2021). ByteTrack: Multi-object tracking by associating every detection box. *arXiv preprint arXiv:2110.06864*. <https://arxiv.org/abs/2110.06864>