

2026 ICEET

第十五屆 數位學習與教育科技國際研討會
International Conference on E-learning and Educational Technology

智慧教育×永續共榮

論文集 PROCEEDINGS



結合視線追蹤之網頁端學習注意力監測系統設計

Design of a Web-Based Learning Attention Monitoring System Integrated with Gaze Tracking

廖亞琳¹ 王岱伊^{2*}

Liao, Ya-Lin¹ Wang, Dai-Yi²

¹靜宜大學 人工智慧應用學系 研究生

¹Providence University Department of Applied Artificial Intelligence
Student

E-mail : cxghk581@gmail.com

²靜宜大學 人工智慧應用學系 副教授

²Providence University Department of Applied Artificial Intelligence
Associate Professor

E-mail : ophelia.wang@gmail.com

摘要

隨著線上學習日益普及，教師難以即時掌握學生在非同步影片學習過程中的注意力狀態。本研究設計並開發一套網頁端學習注意力監測系統，以普通網路攝影機結合 MediaPipe Face Mesh 與 TensorFlow.js，在瀏覽器端建立個人化視線預測模型，無需任何專用硬體即可實現即時視線追蹤。系統採雙端架構：教師端提供動態教材標註工具，可針對教學影片之特定時間段框選「關注區域（AOI）」；學生端即時偵測視線座標並與 AOI 進行比對，視線偏離時觸發視覺引導干預機制，並將完整行為軌跡同步至 Firebase 雲端資料庫。本文重點描述系統之整體架構設計、核心技術模組，以及使用流程，為線上學習注意力監測提供一套低成本、高部署性之解決方案。

關鍵字：視線追蹤；線上學習；注意力監測；MediaPipe

Abstract

As asynchronous online learning becomes increasingly prevalent, teachers face significant challenges in monitoring students' real-time attention states. This study designs and develops a web-based learning attention monitoring system that combines a standard webcam with MediaPipe Face Mesh and TensorFlow.js to construct a personalized gaze prediction model entirely within the browser, without requiring any specialized hardware. The system adopts a

dual-module architecture: the teacher module provides a dynamic content annotation tool allowing instructors to designate Areas of Interest (AOIs) for specific time segments within instructional videos; the student module performs real-time gaze coordinate detection and comparison against AOIs, triggering visual guidance interventions upon deviation and synchronizing full behavioral logs to a Firebase cloud database. This paper focuses on describing the system's overall architecture, core technical modules, and operational workflow, offering a low-cost, highly deployable solution for attention monitoring in online learning environments.

Keywords: gaze tracking; online learning; attention monitoring; MediaPipe

壹、緒論

在遠距教學快速普及的趨勢下，非同步線上影片成為主要教學媒介。然而，師生間缺乏即時面對面互動，使教師難以判斷學生是否維持適當的學習專注狀態，亦無法即時對教材吸引力進行科學化評估。此「教師無法感知、學生缺乏回饋」之雙盲困境，是制約線上學習品質提升的核心問題之一。

眼動追蹤（eye-tracking）技術長期被視為分析學習行為與視覺注意力之黃金工具，然而主流商業眼動儀售價高昂，且需嚴格的環境控制，使其應用場景幾乎侷限於設備完善之實驗室，難以普及至一般學習情境。近年來，網頁端機器學習框架（如 TensorFlow.js）與輕量化電腦視覺套件（如 MediaPipe Face Mesh）的快速發展，使純軟體視線追蹤成為可能(Vakunov & Lagun, 2020)，為解決上述困境提供了新路徑。

基於上述背景，本研究旨在設計並開發一套以普通網路攝影機為唯一硬體需求之網頁端學習注意力監測系統，整合即時視線偵測、教材動態標註與注意力干預機制，提供教育者低成本且具高部署性之學生注意力分析工具。本文聚焦於系統設計面向，包含整體架構、核心技術模組及使用流程。

貳、文獻探討

一、基於電腦視覺的視線追蹤技術

視線追蹤技術正經歷從紅外線感測向純軟體運算法轉型的過程。近年來，隨著網頁端運算能力的提升，研究者開始探索利用標準網頁攝影機擷取臉部特徵的可能性。

Chhimpa 等人 (2024) 提出以虹膜追蹤為基礎的低成本人機互動方案，展示在普通攝影機上以校正與回歸方式達成精確視線估計的可行性。Leblond-Menard 與 Achiche (2023) 則開發了非侵入式即時眼動追蹤系統，適用於輔助科技情境，證明視覺對齊方法在無專用硬體條件下的穩定性。

此類技術具備極高的部署性，特別是透過網頁端實作，能確保受測者的生理數據在本地端即時處理，符合隱私保護與低延遲的學術需求。Falch 與 Lohan (2024) 研究則進一步比較不同環境條件對網頁端視線追蹤精度的影響，為本研究在居家學習場景下的適用性提供重要參考依據。

二、眼部偵測與特徵擷取技術

眼部偵測的穩定性直接影響視線預測精確度。現代技術多採用臉部網格模型：Google 開發之 MediaPipe Face Mesh 能提供 468 個高維度臉部關鍵點，並透過 refine_landmarks 選項進一步細化眼部與虹膜區域，精確定位虹膜邊界(Vakunov & Lagun, 2020)。透過擷取虹膜中心相對於眼角與眼瞼的比例位移，並配合神經網路進行回歸運算，系統能建立個人化的眼-螢幕映射模型。

Liu、Chi 與 Yang (2024) 對視線追蹤個人化校正問題進行系統性回顧，指出九宮格校正程序 (nine-point calibration) 能有效修正個體生理差異並提升偵測穩定度。Khaleel、Abbas 與 Ibrahim (2024) 則比較多種低成本眼動偵測方法，發現 MediaPipe Face Mesh 在實時性與精確度上均優於基於 Haar 特徵的傳統方法，成為近年研究主流選擇。

此外，為了解決網頁端視線估計常見的高頻雜訊問題，本研究導入自適應平滑算法。此演算法會基於視線移動速度動態調整平滑係數，在保留快速眼動特徵的同時，有效過濾穩態下之微幅抖動，進一步確保座標預測之穩定性。

三、線上學習的注意力行為分析

在教育科技領域，視覺注意力是評估學習投入度的核心指標。Sharma、Giannakos 與 Dillenbourg (2020) 以 40 名學生為受試對象進行 MOOC 視線追蹤研究，透過定義「關注區域 (AOI)」來量化受測者在核心教材上的停留時間，發現 AOI 停留時長與學習表現呈顯著正相關。此概念構成本研究教師標註機制之理論基礎。

Trajkovska、Kljun 與 Pucihar (2024) 提出 Gaze2AOI，一套結合 YOLOv8 物件追

蹤與眼動數據的開源 AOI 分析工具，能自動計算首次注視時間（TTFF）、停留時長（dwell time）等指標。本研究採用類似概念，透過教師手動框選 AOI 並計算歸一化座標，以適應非同步影片教材之動態性需求。

在即時反饋干預方面，Schneider 與 Sung（2024）在線上影片教學研究中指出，將教師視線指引（Shared Gaze Visualization）與視覺提示（signalling）整合至學習介面，能顯著提升學生的注意力對齊程度並改善學習成效。本研究即基於此理論，在偵測到視線偏離 AOI 時，即時呈現紅色虛線框與引導文字，協助學生回歸學習焦點。

參、系統設計

一、整體架構設計

本系統採雙端架構設計，分為「教師端（教材製作端）」與「學生端（學生學習端）」兩大模組，透過 Firebase Realtime Database 作為橋樑實現即時資料同步，達成「教材標註 → 即時偵測 → 行為記錄」之完整閉環流程，如圖 1 所示。

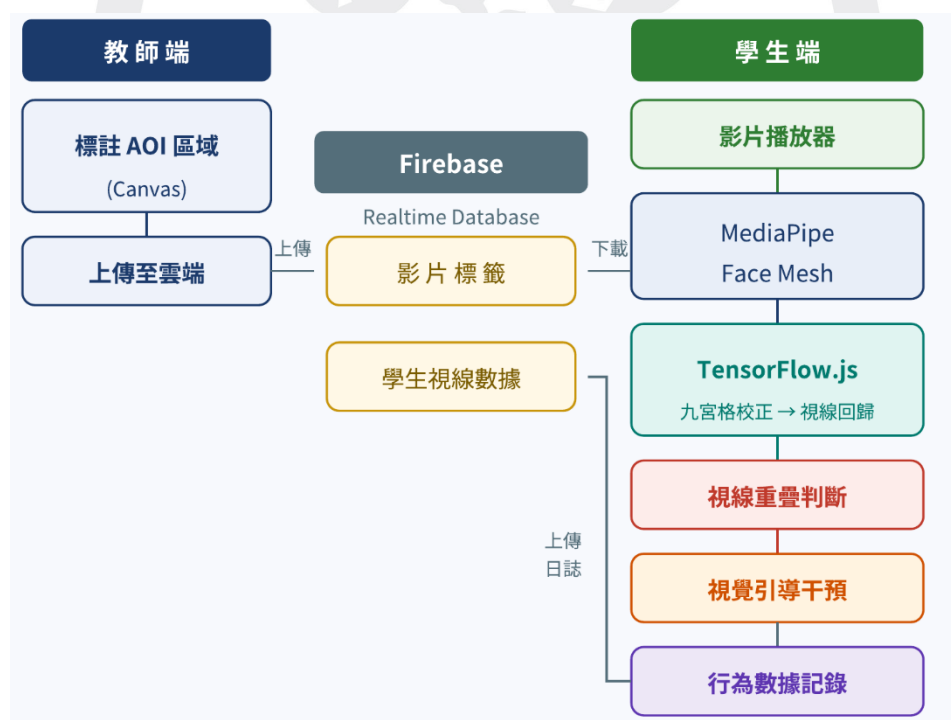


圖 1 系統架構圖

（一）**教師端**：教師匯入教學影片後，可針對特定時間段以拖曳方式框選「關注區域（AOI）」並設定引導文字（如圖 2 所示）。系統自動計算歸一化比例座標（ $xMin, xMax, yMin, yMax \in [0,1]$ ），確保跨裝置、跨解析度均能精準對位，並將參數

即時上傳至 Firebase。

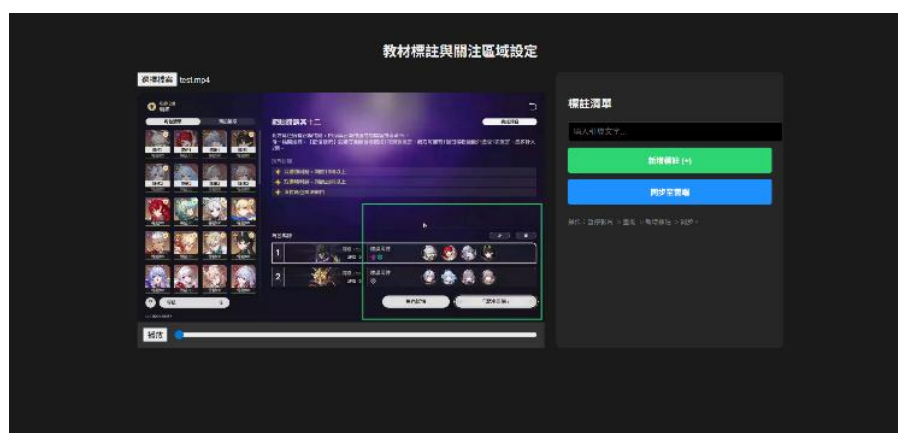


圖 2 教師端動態教材標註介面

(二) 學生端：學生完成個人化視線校正後，系統於背景持續推算視線位置並與當前活躍 AOI 即時比對；視線偏離時觸發視覺干預提示（如圖 3 所示），影片結束後多維度行為數據自動回傳至雲端。



圖 3 學生端視線偏移干預提示

底層採三層式架構：網路攝影機串流由 MediaPipe Face Mesh 即時解析輸出 468 個臉部關鍵點；眼部相對比例座標經 TensorFlow.js 神經網路回歸為螢幕座標（ $x, y \in [0,1]$ ）；預測結果同時驅動視覺游標顯示與 AOI 判斷邏輯；行為日誌經 Firebase SDK 同步至雲端資料庫。

二、核心技術模組

(一) **MediaPipe Face Mesh 臉部特徵擷取**：於瀏覽器端即時擷取 468 個臉部關鍵點，啟用 refine_landmarks 選項以精確定位虹膜邊界（landmarks 468–477），透過擷取虹膜中心相對於眼角與眼瞼的比例位移，形成個人化眼部特徵向量作為視線預測模型之輸入。

(二) **TensorFlow.js 個人化視線回歸模型**：採三層全連接神經網路（4 → 64 → 32 → 2），以 tanh 激活函數與 Adam 優化器（學習率 0.005）訓練 200 個 epoch，輸入為雙眼四維特徵向量，輸出為螢幕歸一化座標。另導入自適應平滑演算法，依視線移動速度動態調整平滑係數，有效過濾高頻雜訊。

(三) **Firebase Realtime Database 雲端資料同步**：同步儲存教師端 AOI 標註參數（含時間段與比例座標）與學生端多維行為日誌，包含每幀視線座標序列、時間戳記、所在 AOI 標籤及注意力統計摘要。

(四) **HTML5 Canvas API 視覺干預渲染**：即時繪製 AOI 框線、視線游標及干預提示文字，採用 requestAnimationFrame 確保繪製與影片幀率同步，提供流暢之視覺回饋體驗。

表 1 系統核心技術模組總覽

技術模組	工具／套件	功能說明
臉部特徵擷取	MediaPipe Face Mesh	即時取得 468 個臉部關鍵點，精確定位虹膜位置
視線預測模型	TensorFlow.js	瀏覽器端訓練輕量化神經回歸網路，映射虹膜特徵至螢幕座標
雲端資料同步	Firebase Realtime Database	同步教師 AOI 參數與學生學習軌跡至雲端
視覺干預渲染	HTML5 Canvas API	即時繪製 AOI 框線、視線游標與干預提示文字
前端環境	HTML5 / CSS3 / JavaScript (ES6+)	零依賴框架、支援所有現代瀏覽器

三、軟硬體環境

本系統採純前端技術實作，無需後端伺服器，可直接以瀏覽器開啟，具備極高跨平台相容性。開發語言採 HTML5、CSS3、JavaScript（ES6+），支援 Chrome / Firefox / Edge 等現代瀏覽器。核心套件包含 MediaPipe Face Mesh（v0.4）、TensorFlow.js（v4.x）及 Firebase SDK（v9.15.0）。硬體需求僅為具網路攝影機之一般筆電、桌機或前置鏡頭之平板裝置，無需紅外線攝影機或其他專用設備。

四、系統使用流程

(一) **教材標註與參數設定**：教師預先匯入教學影片，針對特定教學時間段框選

「注意力重點區域 (AOI)」並輸入引導文字，系統自動計算歸一化座標並即時上傳至 Firebase。

(二) 系統啟動與個人化視線校正：受測學生依序注視畫面上九個校正點 (如圖 4 所示)，每點採集四筆眼部特徵樣本，共 36 筆訓練資料。TensorFlow.js 即時訓練個人化回歸模型 (約 3 至 5 秒)，完成後進入學習模式。

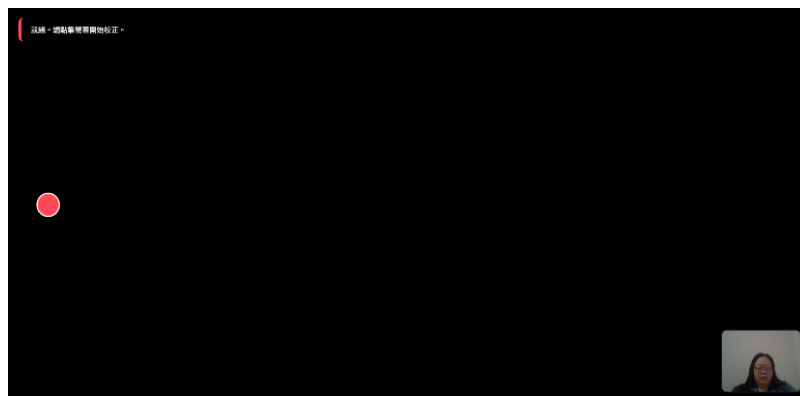


圖 4 學生端個人化視線映射校正程序

(三) 線上學習與即時視覺引導：學習過程中，系統每幀比對預測視線座標與當前活躍 AOI 範圍。視線落於重點區域內時，系統於背景靜默紀錄 (如圖 5(a)所示)；若視線連續偏離，則即時顯示紅色虛線邊框並暫停影片 (如圖 5(b)所示)，協助學生回歸學習焦點；偵測不到人臉時亦自動暫停影片播放。

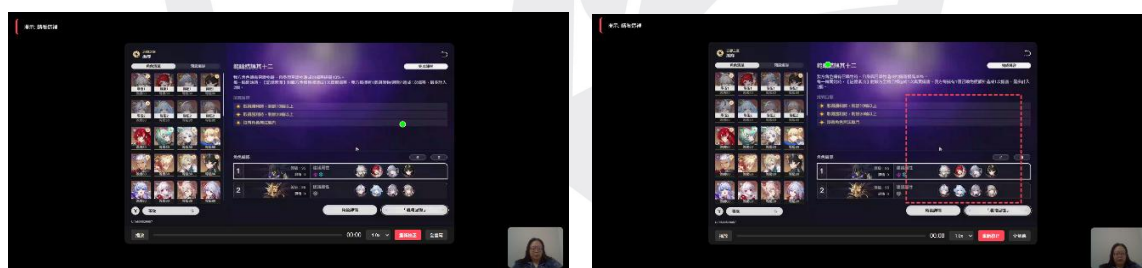


圖 5 系統即時偵測與視覺引導機制 (a) 視線專注於 AOI 之正常狀態 (b) 視線偏離觸發視覺引導

(四) 數據蒐集與行為日誌回傳：影片播放結束後，系統彙整詳細視線軌跡序列 (含每幀時間戳記、視線座標、所在 AOI)、視線移開次數、重新校正次數、分頁切換次數及手動暫停次數等資料，自動回傳至 Firebase 學生資料節點，供後續量化分析使用。

肆、討論與結論

本系統之設計充分體現「低成本、高部署性」之核心優勢。相較於傳統紅外線眼動儀動輒數十萬元之採購成本及嚴格的环境控制需求，本系統僅需一般網路攝影機即

可運行於任何現代瀏覽器，大幅降低了視線追蹤技術應用於教育現場的門檻。

在技術層面，本系統的核心挑戰在於如何在無專用硬體的情況下維持足夠的視線追蹤精度。透過個人化九宮格校正程序，系統能針對每位使用者的生理特徵建立專屬映射模型，有效修正個體差異。自適應平滑演算法的引入則進一步解決了高頻雜訊問題，在保留快速眼動特徵的同時，過濾穩態下之微幅抖動，確保座標預測的穩定性。

在應用層面，雙端架構設計使教師能靈活配置教材標註參數，而學生端的即時干預機制直接回應了線上學習中「缺乏即時回饋」的核心痛點。蒐集的行為日誌亦可作為教師優化教材設計之重要參考，實現教學與學習的雙向改善。

本研究設計並開發了一套結合視線追蹤技術之網頁端學習注意力監測系統，以普通網路攝影機取代高成本專用硬體，透過 MediaPipe Face Mesh 與 TensorFlow.js 實現瀏覽器端個人化視線預測，並整合動態教材標註、即時視覺引導干預與雲端行為數據蒐集等功能，構建完整之線上學習注意力監測解決方案。

本系統具備三項核心貢獻：其一，以純軟體方案複現傳統硬體眼動儀之核心功能，實現零門檻、零安裝之部署模式；其二，透過雙端閉環架構串聯教材標註、即時偵測與行為記錄之完整流程；其三，提供結構化之多維度學習行為數據，為教師優化教材設計與個人化學習分析提供科學依據。後續研究將進行量化驗證實驗，探討即時視覺引導干預對學習增益量之實際效果。

參考文獻

- Chhimpa, G. R., Kumar, A., Garhwal, S., Dhiraj, Khan, F., & Moon, Y.-K. (2024). Revolutionizing gaze-based human–computer interaction using iris tracking: A webcam-based low-cost approach with calibration, regression and real-time re-calibration. *IEEE Access*, 12, 168256–168269. <https://doi.org/10.1109/ACCESS.2024.3498441>
- Falch, L., & Lohan, K. S. (2024). Webcam-based gaze estimation for computer screen interaction. *Frontiers in Robotics and AI*, 11, Article 1369566. <https://doi.org/10.3389/frobt.2024.1369566>
- Khaleel, A. H., Abbas, T. H., & Ibrahim, A.-W. S. (2024). Best low-cost methods for real-time detection of the eye and gaze tracking. *i-com*, 23(1), 79–94. <https://doi.org/10.1515/icom-2023-0026>
- Leblond-Menard, C., & Achiche, S. (2023). Non-intrusive real time eye tracking using facial alignment for assistive technologies. *IEEE Transactions on Neural Systems and*

- Rehabilitation Engineering*, 31, 954–961. <https://doi.org/10.1109/TNSRE.2023.3236886>
- Liu, J., Chi, J., & Yang, Z. (2024). A review on personal calibration issues for video-oculographic-based gaze tracking. *Frontiers in Psychology*, 15, Article 1309047. <https://doi.org/10.3389/fpsyg.2024.1309047>
- Schneider, B., & Sung, G. (2024). Is seeing the instructor's face or gaze in online videos helpful for learning? *Journal of Learning Analytics*, 11(3), 210–223. <https://doi.org/10.18608/jla.2024.8235>
- Sharma, K., Giannakos, M., & Dillenbourg, P. (2020). Eye-tracking and artificial intelligence to enhance motivation and learning. *Smart Learning Environments*, 7(1), Article 13. <https://doi.org/10.1186/s40561-020-00122-x>
- Trajkovska, K., Kljun, M., & Pucihar, K. Č. (2024). *Gaze2AOI: Open source deep-learning based system for automatic area of interest annotation with eye tracking data*. arXiv. <https://arxiv.org/abs/2411.13346>
- Vakunov, A., & Lagun, D. (2020, August 6). *MediaPipe Iris: Real-time iris tracking & depth estimation*. Google Research. <https://research.google/blog/mediapipe-iris-real-time-iris-tracking-depth-estimation/>

