

元智大學資管系

學術類畢業專題頂石課程(二)期末報告

研究主題：**高爾夫揮桿姿勢即時檢測系統**

公司代號：ZE6

實習/專題單位：邱昭彰

指導老師：邱昭彰

姓 名：楊宸瑋

學 號：1113502

Abstract

本研究旨在開發一套結合電腦視覺與深度學習技術的高爾夫揮桿自動化分析系統，以提供球員在非專業場域中亦能獲得客觀、量化且具一致性的動作評估與訓練回饋。傳統高爾夫揮桿學習模式多依賴教練經驗與學員主觀體感，缺乏可量化的姿勢監測機制，且市面上具量測功能之感測裝置往往價格高昂、使用限制多，導致一般球友或初學者難以取得完整姿勢分析資訊。為改善上述問題，本研究採用 SwingNet 模型進行揮桿影片時序事件偵測，能自動辨識揮桿影片中八個標準揮桿階段，建立具一致性的動作節奏分段基準。接著利用 MediaPipe Pose 進行 33 個人體骨架關鍵點定位，並以幾何向量運算推算各階段之肘、膝、體幹、骨盆與肩線相關角度，以形成具生物力學依據的姿勢分析資料。

本系統不需額外感測器或穿戴裝置，即可在一般攝影環境下取得量化姿勢資訊，具備可擴充性、可重複性與使用便利性。透過階段性姿勢量測與角度資料，本研究希望協助不同經驗層級之使用者檢視自身揮桿姿勢，並提供後續整合策略以支援個人化動作改善建議的生成。最終，本研究期望能降低專業訓練門檻，並提升揮桿姿勢分析之普及性、客觀性與效率，成為非侵入式、低成本之運動科技輔助工具原型，作為未來體育智慧化訓練系統之研究基礎。

關鍵字：SwingNet、MediaPipe Pose、自動辨識、姿勢分析

Contents

Abstract.....	i
Contents	ii
Chapter 1 緒論.....	1
Chapter 2 相關技術與研究.....	2
2.1 SwingNet 揮桿動作分段模型	2
2.2 MediaPipe Pose — 用來找出身體骨架位置與動作角度.....	3
Chapter 3 研究方法.....	5
Chapter 4 實驗結果或系統展示.....	7
4.1 系統展示	7
Chapter 5 結論.....	10
Chapter6 專題工作內容和心得	12
Reference	14

Chapter 1 緒論

近年來人工智慧（Artificial Intelligence, AI）與電腦視覺（Computer Vision）技術快速發展，影像辨識已廣泛應用於醫療診斷、運動科學、智慧教練輔助等領域，使動作分析逐漸從傳統依賴經驗判斷轉向客觀化與量化的數據分析。高爾夫揮桿是一個複合性與連續性的精細動作，涉及軀幹旋轉、上下肢協調與穩定控制，其姿勢差異即可能影響擊球距離、準確度與球路軌跡。然而一般學習者多仰賴教練口頭指導或自我體感評估，缺乏實際量化的生物力學參數作為依據，因此容易造成長期錯誤動作習慣與訓練效率不佳的問題。

現有揮桿分析系統多依賴穿戴式感測器、動作捕捉儀或高階量測儀器，雖具高精度，但成本昂貴、使用限制高且不便於一般練習場域普及，因此發展以純影像為基礎的低成本揮桿分析工具具有實用價值。本研究以揮桿影像自動化分析技術為核心，採用 SwingNet 進行揮桿影片的八階段動作分段辨識，建立標準化時序基準，並使用 MediaPipe Pose 偵測 33 個骨架關節點，以計算肘關節、膝關節、體幹傾角、骨盆旋轉與肩線方向等量化運動參數。本研究期望透過此流程達成無需額外裝備即可取得姿勢評估資訊的目標，提高訓練便利性、可及性與一致性。

本系統具有低成本、可跨場域操作、可重複量測等優點，能協助學習者以更客觀的方式理解動作差異，並作為後續導入智能分析與訓練建議的基礎。然而本研究亦可能受影像角度、光線、背景複雜度及人物遮蔽程度影響，導致姿勢估測存在誤差。本研究期望成果能作為運動科技與個人化訓練輔助系統的初步驗證，並提供未來延伸至更高維度生物力學分析或即時教練輔助系統的發展方向。

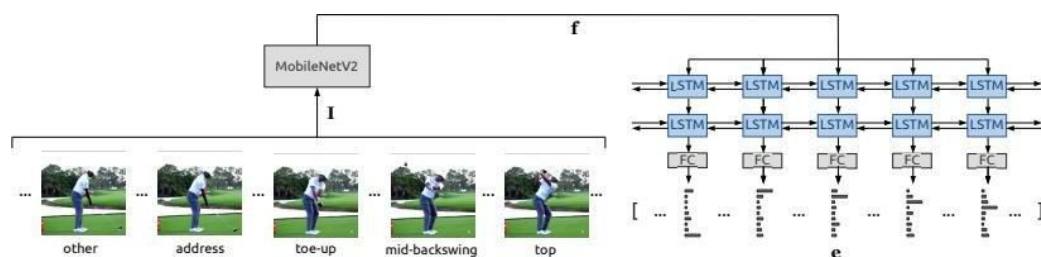
Chapter 2 相關技術與研究

本研究旨在透過影像辨識與深度學習技術進行高爾夫揮桿分析，因此本章將介紹研究中所使用之核心技術與相關背景，包括揮桿階段事件偵測模型 SwingNet、人體骨架與動作偵測模型 MediaPipe Pose，以及基於骨架資訊之姿勢角度計算方法，並簡述其在運動科學中之應用價值。

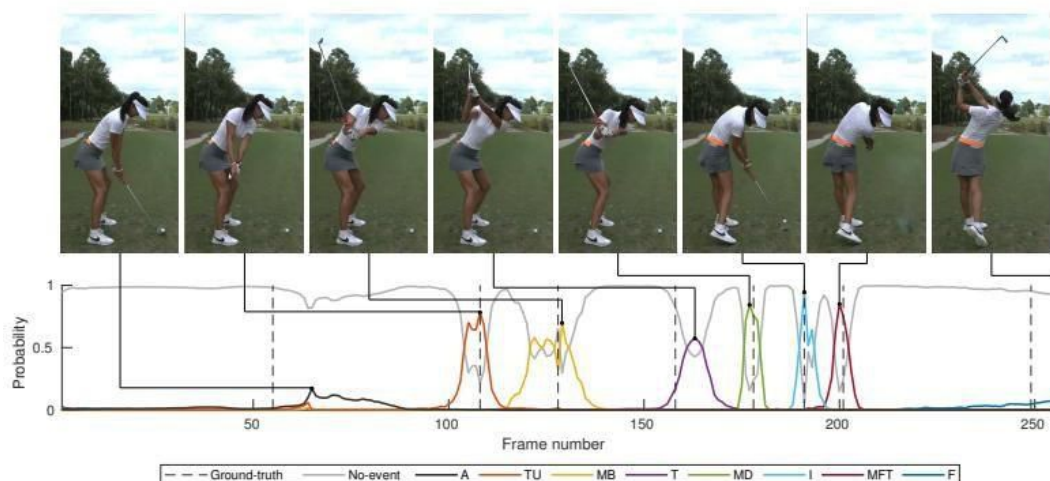
2.1 SwingNet 揮桿動作分段模型

SwingNet 是一個用來分析「影片中動作進行到哪一個階段」的模型。高爾夫揮桿並不是單一動作，而是依時間順序分成多個重要階段，例如準備動作、上桿、下桿、擊球與收尾等。SwingNet 會觀看整段揮桿影片，並自動判斷每一幀影像所對應的階段，讓系統知道整個揮桿的流程與節奏，而不需要靠人工逐格判斷。

簡單來說：SwingNet 的工作就像「幫影片做動作標記的人」。



▲SwingNet 是一個結合卷積神經網路(CNN)與雙向長短期記憶網路 (Bi-LSTM) 的混合式深度模型，用於從高爾夫揮桿影片中逐影格預測八個揮桿事件的發生位置。



▲該圖片是 LPGA 女子高爾夫巡迴賽選手 Michelle Wie 的慢動作揮桿，研究者將此影片輸入 SwingNet 模型，並比較 SwingNet 所預測的揮桿事件（8 個動作階段）是否與人工標註相符。

整體結果為：

1. 8 個階段中有 6 個成功預測，且誤差在 ± 5 個影格（frames）之內。

→ 表示模型具有一定程度的準確度與實用性。

2. 兩個階段預測誤差較大：

- Address (起始準備)：比正確位置慢 10 個影格
- Finish (收尾動作)：比正確位置慢 7 個影格

→ 表示 SwingNet 對「開始」與「結束」的定位相對較不精準，可能與動作模糊界線或個人姿勢風格有關。

2.2 MediaPipe Pose — 用來找出身體骨架位置與動作角度

MediaPipe Pose 是一套可以從影片畫面中偵測人體骨架的技術，它可以觀察出使用者的頭、肩膀、手肘、膝蓋、手腕、腳踝等 33 個身體重要點。透過這些點

的座標，就可以計算出手腳彎曲角度、身體前傾程度或旋轉方向。

簡單來說：**MediaPipe** 就像一個會從影像中畫出「人體火柴人」的工具，讓我們能根據關節位置做分析。



▲該圖片是 **MediaPipe** 應用在高爾夫球揮桿上的示意圖

Chapter 3 研究方法

3.1 系統開發流程

本研究之整體實作流程如圖所示，可分為四個主要階段：

1. 揮桿影片輸入與前處理
2. SwingNet 分段預測
3. MediaPipe Pose 骨架偵測與角度計算
4. 階段姿勢分析與匯出結果

此流程使原始影像資料依序轉換為具有訓練價值之量化姿勢參數，為後續比對與建議產生建立基礎。

3.2 影片前處理

使用者可提供一般攝影設備拍攝之揮桿影片。為提高模型判斷穩定性，本研究在前處理階段執行下列操作：

- 影片解析度統一調整
- 影格抽取 (Frame Extraction)
- 時間序列輸出以符合 SwingNet 需求

此步驟確保資料輸入模型時格式一致並具適當時間序列密度。

3.3 SwingNet 揮桿階段辨識

本研究採 SwingNet 模型對揮桿影片進行八階段動作辨識。透過 MobileNetV2 萃取影格視覺特徵，再由 Bidirectional LSTM 分析時間關聯性，最終以 Softmax 輸

出各影格所屬之揮桿階段機率。本步驟之目的為：

- 將整段揮桿影片**標準化分段**
 - 避免依賴人工判讀造成**不一致與主觀誤差**
 - 提供後續角度分析之**時間定位基準**
-

3.4 MediaPipe Pose 骨架關鍵點偵測與角度計算

於 SwingNet 分段結果中選取各階段代表影格後，使用 MediaPipe Pose 偵測該影格中之 **33 個人體關鍵點位置**，接著透過向量運算與三角函數計算以下生物力學參數：

- 上肢關節：肘角、腕部線性方向
- 下肢關節：膝角、腳踝角度支撐穩定性
- 軀幹核心：骨盆旋轉角度、體幹前傾角度、肩線方向

此步驟之目標為將揮桿姿勢具體轉化為**可量化數值資料**。

3.5 姿勢分析與結果呈現

本研究最後將各階段關節角度整合後進行初步判讀與可視化輸出，包含：

- **每階段關節角度參數表**
- **文字化姿勢描述**
- **常見揮桿動作偏誤指標對照**

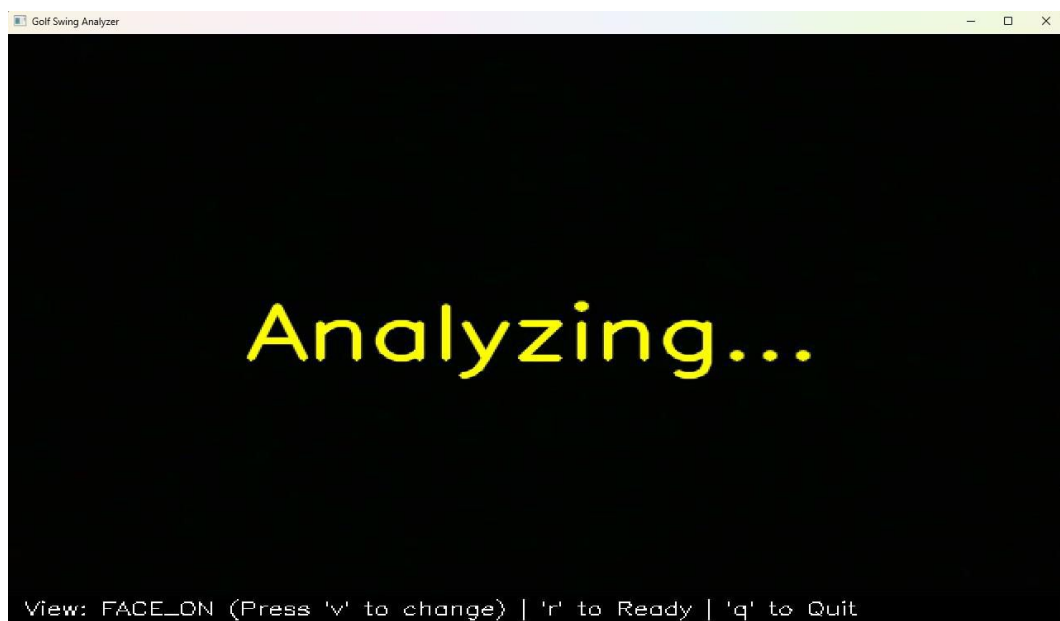
此階段可支援後續深度研究或延伸為訓練建議生成模組。

Chapter 4 實驗結果或系統展示

4.1 系統展示



▲上圖為在家中模擬高爾夫球揮桿姿勢



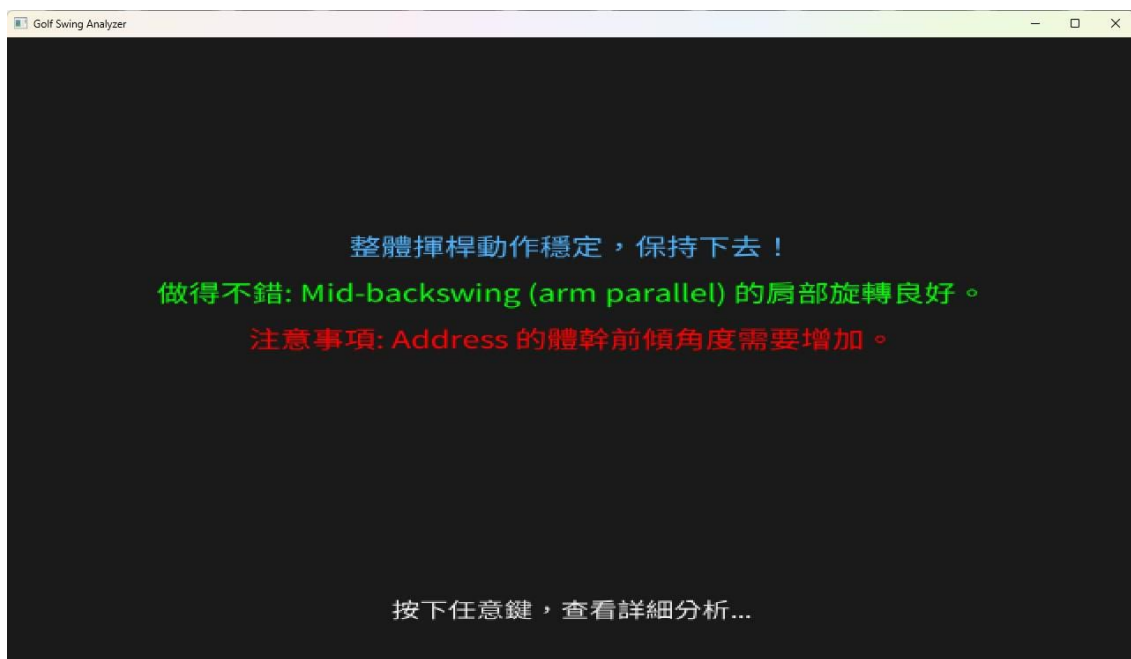
▲在完成揮桿動作後系統開始分析

```

ID:12 state changed to -> ARMING (請在 3 秒內就定位...)
ID:12 state changed to -> READY
ID:12 state changed to -> SWINGING
ID:12 揮桿開始! (Frame: 17929)
ID:12 state changed to -> COOLDOWN
ID:12 state changed to -> IDLE
OK ID:12 揮桿影片已儲存: swing_videos\swing_p12_20251118_030243_512455.mp4
[OpenAI] 開始分析 ID:12 的影片 swing_p12_20251118_030243_512455.mp4 (View: face_on)
[SwingNet] 正在使用設備: cpu
OK [SwingNet] 已載入模型權重: models/swingnet_1800.pth.tar
OK [SwingNet] 分析完成: swing_videos\swing_p12_20251118_030243_512455.mp4
--- [ID:12] 本地 8 階段關節角度分析 (計算完成) ---
Address: R.Elbow: N/A, Trunk: 20.4°
Toe-up: R.Elbow: 177.4°, Trunk: 9.6°
Mid-backswing (arm parallel): R.Elbow: N/A, Trunk: 15.1°
Top: R.Elbow: 120.5°, Trunk: 7.6°
Mid-downswing (arm parallel): R.Elbow: 160.8°, Trunk: 7.8°
Impact: R.Elbow: 113.4°, Trunk: 4.1°
Mid-follow-through (shaft parallel): R.Elbow: 179.7°, Trunk: 10.1°
Finish: R.Elbow: 120.1°, Trunk: 8.3°
-----
[OpenAI] ID:12 正在呼叫 gpt-4o API (View: face_on)...
[LandmarkSmoother] 已啟用 - 平滑強度 alpha=0.2
OK [OpenAI] ID:12 8 階段分析完成

```

▲分析期間程式載入 SwingNet 權重和開始抓取圖片中角度並呼叫 GPT 進行分析



▲分析完會先會給出一段簡評，按下任意按鍵即可查看詳細分析結果



▲8 階段詳細分析結果

Chapter 5 結論

綜合本專題的設計與實作歷程，可以將成果概括為：建立了一套從「揮桿影片」一路走到「文字化教練建議」的完整分析流程。系統首先透過 SwingNet 將連續揮桿影片切分為八個動作階段，解決了「不知道這一幀到底算不上桿還是下桿」的時間定位問題；接著利用 MediaPipe Pose 擷取人體 33 個關鍵點，計算出肘、膝、體幹、骨盆、肩線等角度指標，將原本主觀的姿勢印象，轉換為可以量化比對的數值。最後，再由 GPT 依據每一階段的角度與標準範圍落差，生成結合專業用語與一般人易懂描述的自然語言建議，讓使用者不只看到「角度差了幾度」，也能理解「該怎麼調整動作」。整體而言，本專題已完成一個由「影像 → 階段 → 骨架 → 角度 → 文字回饋」的端到端揮桿分析雛型。

從實作角度來看，這樣的做法有幾項明確的優點。首先，在硬體成本上，只需要一般攝影設備即可取得分析所需的資料，不必安裝感測器或進入實驗室級的量測環境，對一般球友或練習場的導入門檻相對較低。其次，利用 SwingNet 自動判定揮桿階段，避免人工逐格標記，讓之後要分析更多影片時，具備可擴充性與重複使用性。再者，結合 MediaPipe Pose 的骨架偵測與自訂角度計算公式，使得每一次揮桿都能留下可比較的數據紀錄，未來無論是同一位使用者的進步追蹤，或是不同使用者之間的差異分析，都有了共同的量化基準。最後，引入 GPT 產生文字回饋，讓「冷冰冰的數字」轉換成具體的動作提醒與說明，對沒有生物力學背景的使用者來說，更容易理解問題出在哪裡。

然而，在肯定成果之餘，也可以清楚看到本專題仍有不少限制。SwingNet 的分段結果，仍會受到影片拍攝角度、揮桿節奏差異與資料集分佈的影響，有些階段

（尤其是起始準備與結束姿勢）較容易出現偏移，導致後續角度對應到的階段不一定完全精準。MediaPipe Pose 雖然提供了方便的骨架資訊，但本質上仍以單一視角的 2D/2.5D 推估為主，當身體有較大旋轉或部分關節被遮蔽時，角度計算難免存在誤差。此外，目前標準角度範圍與 GPT 的建議內容，仍以整理文獻與教學觀念為主，尚未大量結合實際職業選手以及不同族群（年齡、身材、柔軟度）的資料來微調，嚴格來說仍屬於一個「輔助參考系統」，而非取代專業教練的完整診斷工具。針對這些限制，未來可以有幾個具體的改進方向。技術層面上，可嘗試蒐集更多不同球員與拍攝條件的揮桿影片，對 SwingNet 進行再訓練或微調，提升在非理想場景下的分段穩定度；也可以嘗試加入多視角影像或深度攝影機，改善骨架與角度推估的空間精度。在分析層面上，則可以發展「個人化標準」的概念，讓系統不只與一般標準做比較，也能參考使用者自己的歷史最佳表現，提供更貼身的進步建議。在回饋設計上，可進一步結合圖表與動作示意，例如用顏色標示各階段偏差程度，或提供「前後兩次揮桿的差異對照」，讓使用者一眼就能看出調整是否朝正確方向前進。長遠來看，若能整合更多實際訓練數據與教練回饋，建立一個持續學習與修正的系統，便有機會從現在的分析工具，進一步演化成一套具備教學策略的智慧教練平台。

總結而言，本專題並不是要否定傳統教練的價值，而是希望透過程式與模型，替高爾夫揮桿多準備一套「會記錄、會比較、會說明」的第二雙眼睛。即使目前仍有精度與泛化上的限制，但已初步證實：只要善用既有的深度學習模型與姿勢偵測工具，就能在一般環境中，為運動訓練帶來更科學、可量化的輔助方式，這也是本專題最重要的意義所在。

Chapter6 專題工作內容和心得

● 楊宸瑋

在本專案中，我負責系統開發流程中的核心技術研究、程式實作與分析模組設計，包含從資料取得、模型運作到姿勢分析與結果呈現等關鍵任務。首先，我建立揮桿影片的前處理流程，包含影格擷取、格式統一與模型輸入格式調整，確保系統能正確讀取與分析使用者影像。接著，我實作 SwingNet 動作分段流程，將揮桿影片自動切分為八個標準階段，並建立階段標記與結果對應邏輯。於姿勢分析部分，我整合 MediaPipe Pose 進行 33 個骨架關節點偵測，並以向量與三角函數設計角度計算函式，用以量化肘、膝、體幹、骨盆與肩線等重要動作參數。最後，我將各階段分析結果視覺化與模組化呈現，並進行系統整合測試，以確保整體流程能完整執行與正確輸出。

在整個執行過程中，我不僅投入程式開發，也負責技術查證、參數調整、流程邏輯修正與分析準則的設計。這段期間遇到許多挑戰，如模型參數調整困難、骨架數據不穩定與結果分析方式需一再修改，但也因此累積了更多實作與問題解決的經驗。經過一個學期的努力，最終成功完成以影像為主的揮桿分析系統原型，並使整體架構具備可擴充性與後續優化基礎，這對我個人在 AI 與電腦視覺領域的能力成長具有很大助益。

透過本次專案，我深刻體會到實作系統並非僅是程式能執行，而是必須兼顧可用性、流程連貫性與分析可信度，同時也需要持續調整、驗證與思考。此外在這段學習歷程中，我十分感謝指導教授的耐心指導與方向建議，讓我在遇到瓶頸時能重新釐清目標與思考方向；也感謝實驗室的同學與周遭夥伴於討論與測試上的協助，

使專案能順利推進。透過這次專案，我不僅增強了技術能力，也建立了持續改進的觀念，這將成為我日後面對更多專案與研究時的重要基礎。

Reference

- [1] W. McNally, K. Vats, T. Pinto, C. Dulhanty, J. McPhee, A. Wong, "GolfDB: A Video Database for Golf Swing Sequencing," in *Proc. IEEE CVPR Workshops*, June 2019.
- [2] Google AI, "On-device, Real-time Body Pose Tracking with MediaPipe BlazePose," Google Research Blog, 13 Aug. 2020..
- [3] J.-W. Kim, J. Y. Choi, E.-J. Ha, J.-H. Choi, "Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model," *Applied Sciences*, vol. 13, no. 4, article 2700, Feb. 2023.
- [4] Google AI Edge, "Pose Landmark Detection Guide | MediaPipe Pose," Jan. 13 2025.
- [5] W. McNally, K. Vats, T. Pinto, J. McPhee, and A. Wong, "SwingNet: A Deep Learning Approach for Golf Swing Sequencing," *arXiv preprint*, arXiv:1903.06528, 2019.
- [6] S. Nesbit and S. Serrano, "Work and Power Analysis of the Golf Swing," *Journal of Sports Science & Medicine*, vol. 4, no. 4, pp. 520–533, 2005.
- [7] T. Inoue et al., "Automatic Golf Swing Evaluation Using 3D Pose Estimation and Deep Learning," *IEEE Access*, vol. 9, pp. 100653–100663, 2021.
- [8] F. Zhang et al., "Simple Baselines for Human Pose Estimation and Tracking," *European Conference on Computer Vision (ECCV)*, 2018.
- [9] G. Pavllo et al., "3D Human Pose Estimation in Video with Temporal Convolutions and Semi-Supervised Training," *CVPR*, 2019.
- [10] K. Simonyan and A. Zisserman, "Two-Stream Convolutional Networks for Action Recognition in Videos," *NIPS*, 2014.