國立臺灣師範大學

資訊工程學系碩士論文

指導教授：方瓊瑤 博士

視覺式智慧型高爾夫揮桿動作姿勢分析系統

A Vision-Based Intelligent Golf Swing Posture Analysis System

研究生：石展兢 撰

中華民國 一百一十一 年 七 月

摘要

全球參與高爾夫這項運動的人口數量正在逐步上升，根據世界高爾夫管理機構皇家古老高爾夫俱樂部(The R&A)公布2021年的全世界高爾夫球人數為6,660萬人，超越了2012年的6,160萬人來到歷史高點，可見高爾夫球已經成為全世界普及的運動。近年來運動科技興起，將運動與科技兩者相互結合，利用智慧化訓練能夠有效幫助運動員提升訓練品質並降低運動傷害發生。本研究以高爾夫運動為基礎，為避免高爾夫揮桿姿勢錯誤導致運動傷害，因此開發出一套視覺式智慧型高爾夫揮桿動作姿勢分析系統，讓使用者能夠隨時隨地將自身和教練兩者的高爾夫揮桿姿勢相互比較，可達到自行修正高爾夫揮桿姿勢之目的。

視覺式智慧型高爾夫揮桿動作姿勢分析系統輸入使用者之高爾夫揮桿影片以及教練之高爾夫揮桿影片進行高爾夫揮桿姿勢比對分析。本系統主要分為兩大步驟：高爾夫揮桿分解動作擷取以及三維人體模型姿勢比對分析。在第一步驟中，本研究使用輕量級網路 ShuffleNetV2和循環神經網路 Bi-GRU 進行改良後擷取出使用者以及教練兩者的高爾夫揮桿八個分解動作。在第二步驟中，利用擷取出使用者以及教練兩者的高爾夫揮桿八個分解動作分別建構出可以表現出豐富人體資訊的三維人體模型，接著使用三維人體模型進行使用者以及教練的高爾夫揮桿姿勢比對分析。

本研究將高爾夫揮桿動作拆解成八個分解動作，依序是擊球準備(address)、起桿(toe-up)、上桿(mid-backswing)、上桿頂點(top)、下桿(mid-downswing)、擊球(impact)、送桿(mid-follow-through)以及收桿(finish)。本研究使用 GolfDB 資料集[Mcn19]所蒐集的高爾夫揮桿影片進行訓練及測試，實驗結果顯示高爾夫揮桿分解動作擷取之準確率為86.15%。另外，本研究採用之三維人體模型是由6,890個節點所組成的人體網格，該模型將人體分解成24個身體部位，實驗時利用該模型之擬真人體特性能夠更精準地判斷使用者及教練之高爾夫揮桿姿勢差異。如上所述，本研究所提出之視覺式智慧型高爾夫揮桿動作姿勢分析系統具有效性。

關鍵字: 高爾夫運動、高爾夫揮桿姿勢、運動科技、輕量級神經網路、循環神經網路、三維人體模型、深度學習

Abstract

Global participation in golf is gradually increasing. The world golf management organization, has already announced that the number of golfers in the world is 66.6 million in 2021, which exceeds the number of 61.6 million in 2012 to all-time high. In recent years, with the rise of sports technology, and the use of intelligent techniques can effectively help athletes improve training quality and reduce sports injuries. This study proposes a vision-based intelligent golf swing posture analysis system allowing the user to compare the golf swing of himself and the coach with each other anytime and anywhere. The proposed system can achieve the purpose of correcting the golf swing by the user and avoid sports injury caused by wrong swing posture.

The input of the proposed a vision-based intelligent golf swing posture analysis system is one user's golf swing video and one coach's golf swing video. The system is mainly provided with two stages: golf swing decomposition action extraction and golf swing action comparison analysis. In the first stage, this study modified a lightweight neural network (ShuffleNetV2) and a recurrent neural network (bidirectional GRU) to extract the eight decomposition actions of golf swings of both the user and the coach. In the second stage, the system uses the eight decomposition actions of the golf swing of both the user and the coach to construct 3D human models, respectively. The 3D human models are used to compare and analyze the golf swing of the user and the coach.

This study decomposes the golf swing into eight decomposed actions which in the order of address, toe-up, mid-backswing, top, mid-downswing, impact, mid-follow-through and finish. This study uses golf swing videos collected by GolfDB dataset for training and testing and the experimental results show that the accuracy of golf swing determination action capture is 86.15%. In addition, the 3D human model used in this study is composed of 6,890 vertices. The above model decomposes the human body into 24 body parts. In the experiment, the characteristics of the 3D model can be used to more accurately judge the difference in golf swing posture of users and coaches. In conclusion, the vision-based intelligent golf swing posture analysis system proposed in this study is effective and robust.

Keywords: Golf, Golf swing, Sports technology, Lightweight neural network, Recurrent neural network, 3D human pose and shape, Deep learning

誌謝

隨著論文的完成，研究所的生涯將告一段落，要感謝許多人一路上的指導、幫忙與鼓勵。首先要衷心感謝我的指導教授方瓊瑤教授，老師這兩年對我細心的指導、鼓勵及建議，更教導我做人處事的道理。並且在我在撰寫論文與研究實驗遇到瓶頸時，能夠再次指引我找到研究的方向，讓研究能順利的進行下去，也給予我相當多的啟發及收穫，使得本論文能夠順利完成。

同時也要感謝陳世旺教授，每週開會時給予我相當多的建議與提問，提醒在做研究時沒注意到的細節及研究盲點，還有教授對於學術研究的熱忱與嚴謹的態度更是令我非常欽佩。也要感謝口試委員，黃仲誼博士、羅安鈞博士以及許之凡博士能夠於繁忙中抽出時間替我審查論文，並且提供給我寶貴的指正與建議，使得本研究成果能夠更趨完善。

接著我要感謝實驗室博士班孟霖學長，碩士班皓中學長、日棠學長、后玲學姊、旭政學長、永權學長及秉琛學長，在課業及研究上都給予我相當大的幫助及建議。同時感謝同屆同學雅雯、妤涓以及柏恩，在一同修課時給予我相當多的幫助與教導，以及一起度過在實驗室裡奮鬥的日子，讓我在碩士生涯相當充實快樂。另外還要感謝實驗室哲緯學弟、信宏學弟以及育德學弟，幫忙協助處理實驗室事務。

最後我要感謝我摯愛的家人們，父親、母親以及哥哥，因為有你們一路上的支持以及鼓勵，也給予我相當多的溫暖與照顧，讓我能夠無後顧之憂地專心完成碩士學業。特別感謝葦庭陪伴我一起度過研究所的日子，不斷的給予我支持與鼓勵，讓我能夠堅持下去。

謹以此論文獻給每一位給予我幫助及鼓勵的人。

石展兢 謹致

國立臺灣師範大學 資訊工程學系研究所

中華民國 111 年 7 月

目錄

圖目錄

表目錄

第 1 章 緒論

近年來全世界高爾夫球運動人口數持續增長，根據外國媒體 Sport Show 報導高爾夫球運動以3.9億人口的關注度名列世界上最受歡迎的十大運動之一[1]。根據國際高爾夫總會統計[2]，國際高爾夫總會是由146個國家中的151個成員協會所組成，以及根據世界高爾夫管理機構皇家古老高爾夫俱樂部(The R&A)公布2021年全世界高爾夫球人數為6,660萬人[3]，超越了2012年6,160萬人來到歷史高點，可見高爾夫球已經成為全世界普及的運動。同時，高爾夫揮桿練習是一個可以在室內亦可以在戶外進行的運動，不受人數與場地的限制，非常適合在疫情期間發展與推廣。高爾夫揮桿是一項運用全身快速旋轉的運動，揮桿時需在短時間內使用全身力量將高爾夫球擊出，如果揮桿動作之姿勢長期錯誤，很容易造成下背部受傷或是手肘發炎。然而揮桿動作練習時的主要困擾是練習者看不到自己的揮桿姿勢，無法自我進行姿勢校正。若請專業教練一對一指導，則因時間與經濟因素，不容易符合練習者需求。因此，本研究提出視覺式智慧型高爾夫揮桿動作姿勢分析系統，將和專業高爾夫教練揮桿姿勢相比較，校正揮桿姿勢，避免使用錯誤的揮桿姿勢，減少運動傷害產生。

第一節 研究動機與目的

高爾夫球項目於2016年里約奧運重新被列為奧運比賽項目之一，在2020年東京奧運會上台灣高球好手潘政琮奪下台灣史上首面奧運高爾夫運動項目銅牌，成為一個新的里程碑，也高度提升國人投入高爾夫運動項目的熱情。以及在東京奧運中，中華隊一舉摘下2金、4銀以及6銅，獎牌數突破12面，寫下歷史新高紀錄。在每位選手的優秀表現之下，運動科技正是背後的隱形推手，更讓運動科技再次被重視。根據《科技戰已成常態，運動員背後的「神隊友們」》[4]報導指出，勝負通常都在分毫之間，科技與運動員的完美結合，主要透過運動科學、數據分析以及智慧科技來輔助運動員，不論是在個人化訓練或者模擬對手，都能讓運動員在訓練上達到事半功倍以及避免運動傷害發生的效果。

根據《運動科技導入，智能高爾夫增加擊球人口》[5]報導指出，高爾夫球導入運動科技可以讓新進學習打高爾夫球以及正在學習打高爾夫球的人，在更短的

時間內得到更大的學習效果。可以看出不只有高爾夫球選手需要運動科技的輔助，高爾夫初學者更可以透過運動科技來快速學習正確揮桿姿勢，避免揮桿姿勢錯誤造成運動傷害。

高爾夫球運動屬於低衝擊(low impact)的全身性鍛鍊運動，可以激活上肢和下肢肌肉以及提升身體靈活性和平衡感，是一項老少咸宜的運動。研究顯示[6]打高爾夫球可有效預防高血壓、降低膽固醇及腦中風發生的機率。打高爾夫球時，為了讓腦眼手合一，需要高度專注與身體協調能力，讓腦袋想的與身體做的一致，才能揮出好球。如果要揮出完美揮桿同時減少運動傷害，就必須要有正確的站姿以及正確的揮桿姿勢，身體任何一個部位不協調可能會使擊球失準甚至受傷。根據一三高爾夫教學中心創辦人黃玉潔所述[7]，養成錯誤的揮桿姿勢需要花三年的時間來進行姿勢修正，更有甚者，導致運動傷害。根據專業高爾夫球運動者以及業餘高爾夫球運動者的運動傷害原因[The98]調查結果如表1所示，專業高爾夫球運動者最常因練習過度而導致運動傷害，受傷比例達到79.90%。業餘高爾夫球運動者則是因為高爾夫揮桿技術錯誤與缺陷而導致運動傷害，受傷比例為62.70%。過度練習所導致受傷的比例只有28.90%，所以剛學習高爾夫球運動者因為高爾夫揮桿姿勢錯誤進而導致受傷比例非常高。休閒高爾夫球運動者的身體運動傷害部位統計[Mal95]則如表2所示，休閒高爾夫球運動者最常受傷的部位為下背部以及手肘，不正確的揮桿姿勢以及過度施力揮桿會導致腰椎承受太多力量，使得下背部受傷，受傷比例更高達52%。另外不正確的揮桿也會造成手肘受傷，被稱之為高爾夫球肘，當姿勢不良造成手肘內側受到大力拉扯，或者揮桿技術不佳導致球桿敲擊至地面，使肌肉在收縮時突然受到一個相反的震力而造成的創傷性損傷，都容易造成手肘內側發炎以及疼痛，引起前臂痠痛無法使力，受傷比例也達到24%。

McHardy 等人[Mch07]將高爾夫揮桿分為3個階段，分別為上桿(backswing)階段(如圖1(b)至(d))、下桿(downswing)階段(如圖1(d)至(f))以及送桿(follow through)階段(如圖1(f)至(h))。該研究調查73位高爾夫球運動者在揮桿導致運動傷害[Mch07](圖2)的階段，最常見為送桿(Follow through)階段。送桿階段在擊中高爾夫球後會藉由慣性使得身體持續的向左旋轉，直到身體正面向著目標。隨著手肘的彎曲而逐漸往上移，進入最後收桿的姿勢。不正確的揮桿姿勢或者不流暢的揮

桿過程都容易增加腰椎部位的負擔。圖2中的統計除了上述三個階段以外還有一個其他(Other)類，其他(Other)類的定義為超過1個揮桿階段或擊球時球桿撞擊地面使高爾夫球運動者造成運動傷害，其中撞擊造成運動傷害佔其他(Other)類的20%。由此可見，當高爾夫揮桿動作一旦不正確，就非常容易導致運動傷害。

表1：專業高爾夫球運動者與業餘高爾夫球運動者的運動傷害原因列表[The98]

受傷原因 / 級別 / 參考來源		過度使用	技術錯誤/缺陷	熱身不足	被球或球桿擊中	其他
專業高爾夫運動者	McCarroll [Mc96]	270 (79.90%)	58 (17.20%)	0 (0.00%)	3 (0.01%)	7 (0.02%)
業餘高爾夫運動者	Thériault et al. [The96]	24 (20.00%)	95 (77.00%)	0 (0.00%)	0 (0.00%)	2 (3.00%)
	McCarroll [Mc96]	204 (28.90%)	454 (62.70%)	60 (8.30%)	36 (5.00%)	50 (0.06%)

表2：休閒高爾夫球運動者的身體運動傷害部位列表[Mal95]

人數 / 部位	女(1160)	男(249)	總計 (1409)
下背部	617(53%)	111(45%)	728(52%)
左肘	273(24%)	68(27%)	341(24%)
左肩	101(9%)	10(4%)	111(8%)
左腕	71(6%)	35(14%)	106(8%)
左踝	34(3%)	7(3%)	41(3%)
視力問題	28(2%)	6(2%)	34(2%)
右臀	13(1%)	4(2%)	17(1%)
左臀	10(<1%)	4(2%)	14(1%)
右膝	7(<1%)	3(1%)	10(<1%)
左膝	6(<1%)	1(<1%)	7(<1%)

(a) (b) (c) (d) (e) (f) (g) (h)

圖1：高爾夫正面揮桿分解動作範例

運動傷害一般分為急性運動傷害以及慢性運動傷害二種。急性運動傷害主要是受到強烈外力所造成的傷害，慢性運動傷害主要是過度使用導致的傷害，而姿勢不正確也是慢性運動傷害所造成的原因，當運動傷害發生時，必須要停止訓練並適當休息才不會再次復發。根據高爾夫運動傷害造成部位受傷復原時間[The98]調查顯示(表3)，高爾夫運動傷害受傷復原時間小於一個月內為34.1%，受傷復原時間大於或等於一個月為65.9%，所以當受到運動傷害的影響而中斷高爾夫訓練，必須要修養好幾個禮拜讓身體復原，也會讓身體失去先前訓練的揮桿協調性。

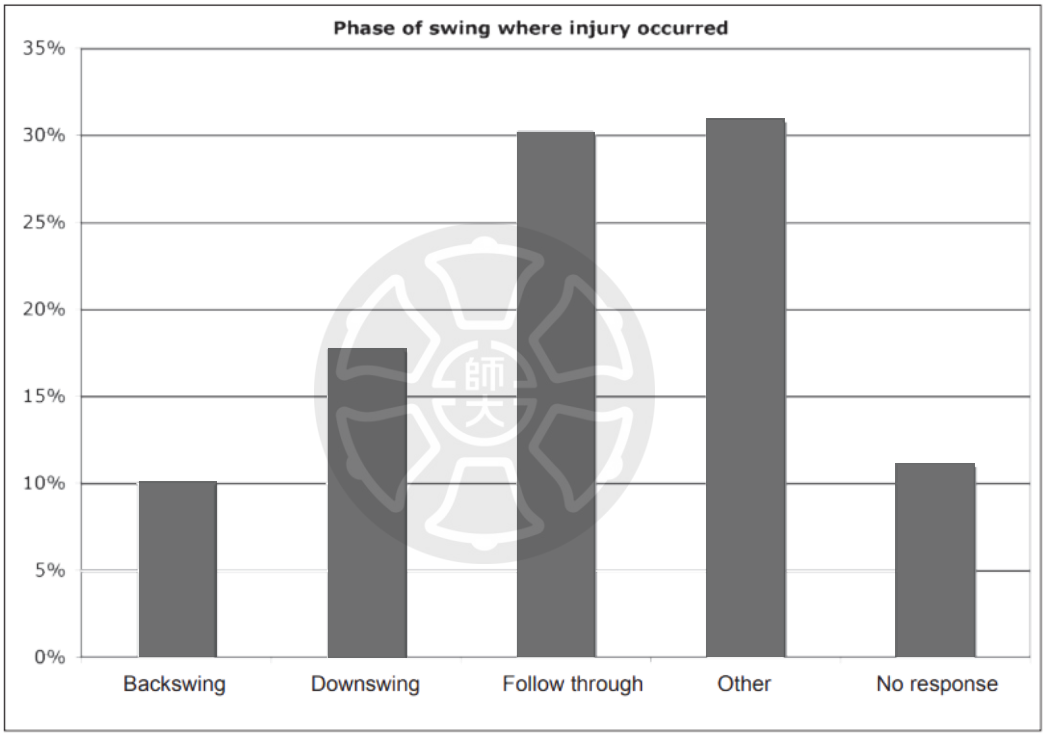


圖2：高爾夫揮桿階段造成運動傷害統計[Mch07]

高爾夫揮桿動作練習時主要的困擾是無法看到自己的揮桿姿勢，而且揮桿動作極為快速，因此自身無法找到需要改進的地方，致使無法進行姿勢校正。使用傳統高爾夫教練教學方式，每位教練教學風格有所不同，但是不一定符合學員的需求，導致效果有限進步緩慢，且必須配合教練的時間，無法隨時練習。因此，高爾夫揮桿動作分析系統和高爾夫揮桿追蹤器系統極具研發與商業的價值。現有的高爾夫揮桿動作分析系統有些需要安裝多台攝影機或感測器，在使用時有場地大小的限制且價格相對昂貴。另一方面，現有的高爾夫揮桿追蹤器系統，則將感

測器安裝在高爾夫球桿或者人體身上感測揮桿相關訊息，導致高爾夫球桿具重量差異或使用者感受到異物感造成揮桿不便或揮桿不準的問題。

綜整以上敘述，本研究將開發一基於深度學習技術對高爾夫揮桿動作姿勢分析之系統，利用類神經網路模型來對高爾夫揮桿動作進行比對分析，以利於校正高爾夫揮桿姿勢，避免產生運動傷害。

表3：高爾夫運動傷害造成部位的受傷復原時間[The98]

受傷復原時間(月)	脊椎	上肢	下肢	總計
<1	21 (41.2%)	15 (25.9%)	6 (42.9%)	42 (34.1%)
1-6	16 (31.4%)	28 (48.3%)	4 (28.6%)	48 (39.0%)
6-12	8 (15.7%)	11 (19.0%)	2 (14.3%)	21 (17.0%)
>12	6 (11.8%)	4 (6.9%)	2 (14.3%)	12 (9.8%)
總計	51	58	14	123

第二節 研究困難與限制

本研究以拍攝高爾夫揮桿畫面人數單人為前提，為提高系統的準確率，其研究限制如下：

1. 拍攝角度：拍攝高爾夫揮桿影片時，不是任意角度拍攝到的影像都能達到揮桿姿勢正確預測的目的，須拍攝到正確的角度才能預測出正確的姿勢並判別高爾夫揮桿動作是否正確。因此高爾夫揮桿動作姿勢分析系統主要以拍攝正面角度(如圖1所示)為前提。

2. 揮桿次數限制：部分高爾夫揮桿影片有可能會包含多次揮桿動作，然本研究並未處理影像片段分割部分，只針對單次高爾夫揮桿進行分析。

本研究利用攝影機所拍攝的高爾夫揮桿影片自動進行高爾夫揮桿分解動作擷取以及進行三維人體模型比對分析，但由於二維影像建構出三維人體模型資訊量不足，會遇到困難如下：

1. 自遮擋問題：在拍攝高爾夫揮桿影片時，會因為拍攝角度不同，導致揮桿者會因為自身身體遮擋了其他身體部位，例如圖3(g)所示，揮桿者左、右手被身體擋住導致無法取得雙手位置的資訊，使得建構三維人體模型時失真。

圖3：高爾夫側面揮桿分解動作範例

第三節 研究貢獻

本研究為視覺式智慧型高爾夫揮桿動作姿勢分析系統之開發，而本系統又可分為兩大步驟，分別為高爾夫揮桿分解動作擷取以及三維人體模型姿勢比對分析。以下說明本研究主要貢獻：

(1) 使用 GSNet(golf swing net)進行關鍵動作幀估計，以便未來可以嵌入行動裝置。GSNet(golf swing net)由輕量級網路 ShuffleNetV2和循環神經網路 Bi-GRU 組成，主要的改良為(1)採用幀差法(frame difference method)處理輸入之高爾夫揮桿影片，以便擷取人體運動輪廓並降低背景干擾，有效提升預測準確率。(2)利用群組正規化(group normalization)和 h-swish 激活函數來提升輕量級網路 ShuffleNetV2 模型預測的準確率。(3) 引入 ECA-Net(efficient channel attention for deep convolutional neural networks) 輕量級注意力模組改良 ShuffleNetV2模型，有效加速其收斂速度並提升預測準確率。

(2) 本研究研發出關鍵動作幀判定之校正方法。利用高爾夫揮桿時的時間連續性，代表高爾夫八個分解動作幀數會依序由小到大排序的特性，應用高斯分佈(gaussian distribution)技術將關鍵動作幀估計預測結果再次進行校正。

(3) 本研究研發出三維人體模型姿勢比對分析之方法。三維人體模型姿勢比對分析使用三維人體模型進行人體姿勢比對，由於使用者和教練的身型不同，二者的三維人體模型需對齊並調整後才能進行動作比對，最後輸出動作比對後差異最大的身體部位名稱以及兩者三維人體模型重疊之反饋圖片。

第四節 論文架構

本論文共分為五章，第一章闡述本研究的動機與目的、研究困難與限制；第二章為文獻探討，此章探討各種高爾夫揮桿分解動作系統以及三維人體模型應用；第三章為介紹視覺式高爾夫揮桿動作姿勢分析系統流程以及架構；第四章為實驗結果與討論；第五章為結論與未來工作。

第 2 章 文獻探討

在進行視覺式智慧型高爾夫揮桿動作姿勢分析系統開發之前，需要先瞭解目前視覺式高爾夫揮桿分解動作擷取的相關技術，包含高爾夫揮桿分解動作定義以及高爾夫揮桿分解動作擷取方法。另外，還需探究三維人體模型建構的相關技術，利用三維人體模型進行高爾夫姿勢的比對。本章第一節將探討高爾夫揮桿分解動作；第二節介紹高爾夫揮桿分解動作系統分析；第三節介紹輕量級神經網路；第四節探討人體結構表示法分析；而第五節則介紹三維人體模型建構及應用。

第一節 高爾夫揮桿分解動作

高爾夫揮桿動作一般可以仔細拆解成四至八個不等的分解動作，Chotimanus 等人[Cho12]將高爾夫揮桿動作分解為四個連續的分解動作如圖4所示，圖中示範者其動作由左至右分別為擊球位置(address position)、起桿(takeaway)、擊球(hit)以及揮桿結束(swing's finish)。

圖4：高爾夫揮桿動作拆解為四個分解動作之分解圖例[Cho12]

Ko 等人[Ko21]將高爾夫揮桿揮桿動作分解為七個連續的分解動作如圖5所示，圖中示範者之動作由左至右分別為擊球準備(address)、上桿(back swing)、上桿頂點(top of swing)、下桿(down swing)、擊球(impact)、送桿(follow through)以及收桿(finish)。

圖5：高爾夫揮桿動作拆解為七個分解動作之分解圖例[Ko21]

McNally 等人[Mcn19]將高爾夫揮桿動作分解為八個連續的分解動作，如圖6所示。圖中之示範者其揮桿動作由左至右分別是擊球準備(address)、起桿(toe-up)、上桿(mid-backswing)、上桿頂點(top)、下桿(mid-downswing)、擊球(impact)、送桿(mid-follow-through)以及收桿(finish)。由上述三個不同的分解動作範例可知揮桿動作的拆解其實具有一致性，揮桿動作中重要的細節步驟都被包含在分解動作中。

圖6：高爾夫揮桿動作拆解為八個分解動作之分解圖例[Mcn19]

一般而言，分解動作愈多代表其揮桿動作拆解的細緻度愈高。因此本研究以八個分解動作作為揮桿動作姿勢分析時的基礎。以 McNally 等人[Mcn19]所採用的八個分解動作為例，整理並簡述揮桿擊球各分解動作的動作重點[8]。

1. 擊球準備(address)：即包含瞄準、站距、球位、握桿及站姿而成的準備姿態[7]。首先挺直站立，雙腿放輕鬆，就瞄球姿勢時，上半身俯身，雙臂自然下垂，雙手在下巴的正下方握住球桿，臀部向後頂出，下背部伸直。身體重心落在腳弓前端，膝蓋略為彎曲。在擊球準備時常用的站姿有三種，分別為方正站姿(square stance)、開放式站姿(open stance)以及關閉式站姿(close stance)。其定義為啟動揮桿動作前一刻。

2. 起桿(toe-up)：即為上桿的開始動作。起桿動作亦為揮桿模式的開始，此時揮桿的模式與節奏已經開始運作。其定義為高爾夫球桿身與地面平行。

3. 上桿(mid-backswing)：即為揮桿向上的動作，此時身體上半身已經開始大幅轉動。在上桿時右膝應該保持擊球準備時略彎的角度，只要將肩膀和上半身旋轉90度以上即可。其定義為左手臂與地面呈現平行。

4. 上桿頂點(top)：即為上桿的終點，亦即下桿的起點。上半身轉動到達極限，下半身此時抗力最強，身體扭轉姿勢之展現。其定義為從上桿到下桿過程中改變方向的那一瞬間。

5. 下桿(mid-downswing)：即為揮桿向下的動作。下桿時將重心由右腳轉移至左腳，配合些微的腰部側向運動，雙臂加速向下揮桿。要注意的是這個動作的標準與否，教練必須利用攝影設備才能檢視[7]。其定義為左手臂與地面呈現平行。

6. 擊球(impact)：即擊到球的瞬間。擊球時雙臂及球桿同時到達目標球的位置，此時雙臂回到身體的正面以及球桿面亦朝向目標球回正，肩膀於擊球瞬間應平行於目標球。其定義為擊到球的瞬間。

7. 送桿(mid-follow-through)：即用球桿揮送目標球的動作。球桿握把末端離開身體，球桿及上半身在此階段和上桿有相對稱的動作。送桿時左腿向後伸直，重心後移到左腳根，右腳拉成腳尖頂地，左手肘向內自然彎曲，右手腕保持微彎的角度。其定義為高爾夫球桿身與地面平行。

8. 收桿(finish)：即揮桿結束的動作。揮桿結束時，身體的姿勢是先前所有動作造成的結果。此時幾乎全身的重量都落在左腳上，而右腳只有腳尖碰觸地面。收桿時手的位置決定揮桿路徑，若雙手與左耳呈水平，表示揮桿路徑較平坦(flat)；若雙手在頭部上方，則是揮桿路徑較為高直(upright)。其定義為揮桿結束動作身體肌肉放鬆前一刻。

這八個分解動作中較值得注意是第五個下桿動作，它是教練亦必須利用攝影設備才能檢視是否正確的動作，下桿動作的存在突顯了視覺式智慧型高爾夫揮桿動作姿勢分析系統的必要性。另一方面，這些分解動作的說明可以了解一個揮桿擊球動作要成功，有許多需注意的細節，很容易顧此失彼。高爾夫揮桿練習時常面臨的問題包含校準錯誤、上桿速度過快、右膝直立、重心逆轉或重心位置不正確、身體中軸(頭部)明顯橫移、雙手在揮桿過程中過分用力、雙手啟動下桿或轉動肩膀啟動下桿以及沒有送桿或送桿不充分等[4]，這些問題將導致揮桿無力或揮桿不穩定，更有可能導致身體受傷。因此，本研究所提出之視覺式智慧型高爾夫揮桿動作姿勢分析系統可以適時地比對分析練習者高爾夫揮桿和教練高爾夫

揮桿之間姿勢的差異性，提醒練習者注意正確揮桿動作的各種細節，才能夠揮出完美揮桿以及避免運動傷害的產生。

第二節　高爾夫揮桿分解動作系統分析

在高爾夫揮桿分解動作系統中，Chotimanus 等人[Cho12]使用傳統影像處理技術偵測高爾夫球桿以及高爾夫球來擷取高爾夫揮桿分解動作。該方法是利用(1)高爾夫球桿和高爾夫球之間的歐幾里得距離以及(2)高爾夫球是否還在同一個位置二項條件作為揮桿分解動作的判斷依據。由上可知，高爾夫球桿以及高爾夫球偵測的正確率高度影響高爾夫揮桿分解動作擷取時的準確率。而影像中的高爾夫球桿會因背景複雜以及高爾夫揮桿速度過快導致產生模糊，進而導致高爾夫球桿偵測時的誤判，使得系統無法進行高爾夫揮桿分解動作擷取。

Noiumka 等人[Noi13]則採用光學式動作捕捉方法捕捉高爾夫揮桿動作。該研究使用九台攝影機以及92個光學標識點(markers)進行高爾夫揮桿動作檢測和空間定位，再利用收集到的數據進行分析。值得一提的是這92個光學標識點不只分佈在使用者全身還設置在高爾夫球桿上。因此，該方法有較高的準確率，但需要使用到較多的硬體設備並限制收集訊息的特定場地，不僅設備成本高且會造成使用時的不便。

近年來深度學習技術蓬勃發展，Ko 等人[Ko21]使用卷積神經網路(CNN)進行高爾夫揮桿分解動作擷取。Ko 等人首先將高爾夫揮桿動作拆解成七個動作如本章第一節中的圖5所示。該研究使用的卷積神經網路是由3個卷積層(convolution layer)、3個池化層(pooling layer)以及2個全連接層(fully connected layer)所組成。使用卷積神經網路能夠保留影像中的位置資訊，但高爾夫揮桿動作具有時間序列關係，只使用該研究者提出的卷積神經網路無法保留高爾夫揮桿時間序列關係，會有應在排在前面的揮桿分解動作卻在後面的揮桿分解動作之後被辨識出來的不合理情況發生，例如將擊球準備動作排在上桿動作之後。

McNally 等人[Mcn19]於西元2019年提出一 SwingNet 架構進行高爾夫揮桿分解動作擷取的系統。McNally 等人在研究中定義了八個高爾夫揮桿分解動作，其高爾夫揮桿分解動作曾在本章的第一節說明(如圖6)。該研究首先使用輕量級神

經網路 MobileNetV2[San18]擷取影像中的空間結構的特徵，接著使用 Bi-directional Long Short-Term Memory (Bi-LSTM)得到影像序列間的時間特徵，利用這二種特徵進行高爾夫揮桿分解動作擷取。該高爾夫揮桿分解動作擷取系統並不限制高爾夫揮桿影片拍攝角度，使用 MobileNetV2和 Bi-LSTM 分別處理空間特徵與時間特徵具有互補的效果。且實驗結果顯示 MobileNetV2在降低參數量的同時仍能維持良好的準確率，因此能夠將該系統移植到行動裝置上。

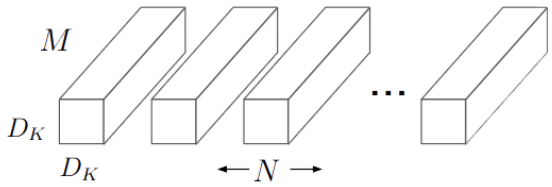
總而言之，本研究認為 McNally 等人[Mcn19]所提出的八個高爾夫揮桿分解動作其分解細緻度最高，且其技術不需限制高爾夫揮桿影片拍攝角度，系統可同時整合空間特徵和時間特徵，並採用輕量級神經網路降低參數量同時維持良好的準確率。故本研究決定以 McNally 等人研發的 SwingNet 作為基礎進行改良，開發高爾夫揮桿分解動作擷取系統。

第三節 輕量級神經網路

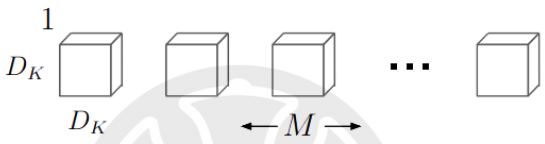
輕量級神經網路研發的核心思想就是在維持準確率的前提下，將卷積神經網路模型結構與速度兩個方面進行輕量化的改造。卷積神經網路在影像辨識、影像分割以及影像分類等相關領域擁有廣泛應用，隨著對準確率要求越來越高，網路層數不斷增加，使得結構越來越複雜。雖然卷積神經網路準確率提升，但是模型結構加深以及學習與執行速度變慢問題導致其在應用上產生許多限制。所以輕量級神經網路提出的主要概念為透過架構上的改良來降低參數量與浮點數運算次數(FLOPs)，在維持一定辨識準確率的情況下提升其學習與執行速度，使其能符合嵌入式裝置的應用需求。

Howard 等人[How17]於西元2017年提出 MobileNetV1的輕量級神經網路架構。MobileNetV1使用深度可分離卷積(depthwise separable convolution)來達到降低參數量的目的，其方式如圖7所示。圖7中，圖7(a)所示的3D 卷積運算有N個卷積核(kernel)，M個輸入通道(channel)，且每個3D 卷積核的參數量為$D_K \times D_K \times M$。因此在進行學習時所需調整的參數量為$N \times D_K \times D_K \times M$。而深度可分離卷積技術則是將3D 卷積運算以深度卷積運算與逐點卷積運算的連續二步驟來取代。圖7(b)所示為深度卷積(depthwise convolution)運算的卷積核，而圖7(c)則為逐點卷積(pointwise convolution)運算的卷積核。其中深度卷積的M個輸入通道採用不同的

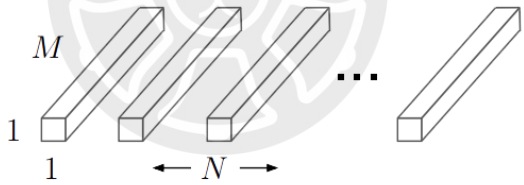
$D_K \times D_K \times 1$大小的卷積核進行卷積運算，換句話說就是一個卷積核只對應一個通道獨立各自做卷積運算。當每一個輸入通道都做完深度卷積運算後，再使用逐點卷積運算融合時間訊息，此時就是使用$1 \times 1 \times M$的卷積核做卷積運算。因為採用二步驟完成類似圖7(a)所示的3D 卷積運算的功能，其學習時所需調整的參數量僅為$D_K \times D_K \times M + N \times M$。由此可知深度可分離卷積與3D 卷積比起來深度可分離卷積可以減少許多的參數量以及計算量。



(a) Standard Convolution Filters



(b) Depthwise Convolutional Filters

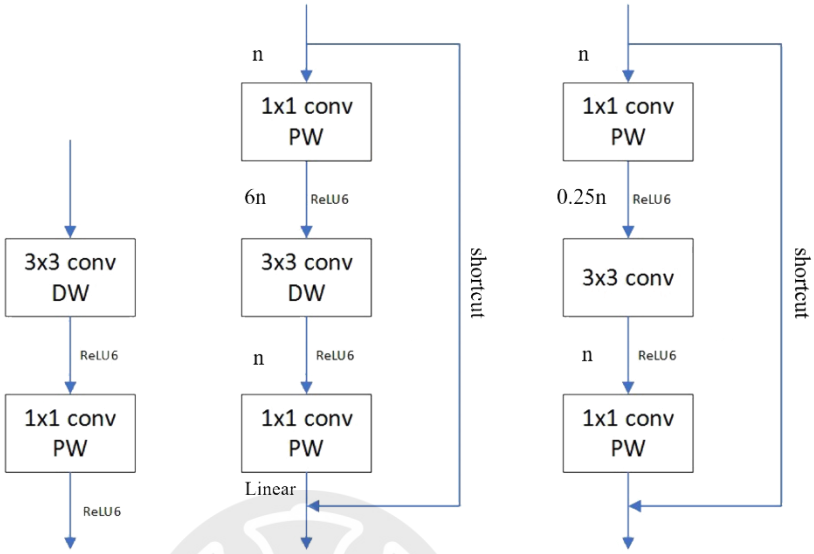


(c) 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

圖7：3D 卷積核和深度可分離卷積核比較圖[How17](a)3D 卷積核(b)深度卷積核(c)逐點卷積核

Sandler 等人[San18]改良 MobileNetV1[How17]提出一 MobileNetV2架構，改良的方式是在 MobileNetV1架構中嵌入線性瓶頸(linear bottleneck)以及反向殘差(inverted residual)兩種技術。MobileNetV1架構圖如圖8(a)所示，而改良後的MobileNetV2架構圖則如圖8(b)所示。線性瓶頸技術概念為將 MobileNetV1架構最後輸出的 ReLU 函數(activation function)替換成線性(linear)函數，而反向殘差技術和已知 ResNet 中的殘差(residual block)技術架構使用方式正好相反，是一種先升維再降維的神經網路架構。

Sandler 等人引入線性瓶頸技術是因為發現將特徵圖從高維空間映射至低維空間時，如果接著使用非線性激勵函數會容易丟失重要資訊。換句話說，ReLU 函數在低維空間中使用效果較差。因此該研究將最後一層逐點卷積從 ReLU 函數更改使用線性激勵函數，稱為線性瓶頸技術。



(a) MobileNetV1　　(b) MobileNetV2　　(c) ResNet

圖8：MobileNetV1、MobileNetV2與 ResNet 架構圖[San18](a)MobileNetV1架構圖(b)MobileNetV2架構圖(c)ResNet 架構圖

另一方面，反向殘差技術是因為使用深度卷積運算時無法改變通道維度，如果維持在低通道維度的情況下，做深度卷積運算容易導致訓練效果不佳，所以在每次做深度卷積之前，宜先使用逐點卷積將通道維度提升。該作法和已知 ResNet 中的殘差設計方法正好相反。ResNet 架構如圖8(c)所示，圖8(c)所示的 ResNet 架構有n個輸入通道，使用$1 \times 1 \times 0.25n$的卷積核做卷積運算進行降維至$0.25n$個通道之後才會進行下一梯次的卷積運算(卷積核大小為(3×3))，最後再使用$1 \times 1 \times n$的卷積核做卷積運算升維至原來的通道維度。ResNet 整個過程先降維、卷積運算再升維，此技術的目的是為了減少卷積運算次數。而反向殘差技術則是先使用$1 \times 1 \times 6n$的卷積核做卷積運算進行升維至$6n$個通道之後，接著進行下一梯次的卷積運算(卷積核大小為(3×3))，最後再使用$1 \times 1 \times n$的卷積核做卷積運算降維至原來的通道維度。反向殘差技術整個過程先升維、卷積運算再降維，此方式是為了在高通道維度上進行卷積運算，提升模型的訓練效果。總而言之，殘差技術

的整體概念為先降維、卷積運算再升維，而反向殘差技術則是先升維、卷積運算再降維。

Howard 等人[How19]接著改良 MobileNetV2[San18]提出一 MobileNetV3架構。MobileNetV3使用神經結構搜索(neural architecture search)技術，它可以根據使用者的需求自動設計出最合適該應用的神經網路架構，是自動化機器學習(automated machine learning)的一種技術。MobileNetV3架構如圖9所示，它在MobileNetV2架構基礎下引入 SENet[Hu18]架構中的 squeeze and excitation 技術，如圖9紅色框內所示。Squeeze and excitation 技術可以學習各通道裡特徵圖之間的資訊關係，並計算各個特徵圖的權重來提升較重要特徵圖的影響力，同時降低較不重要特徵圖的影響力。Squeeze and excitation 技術的作法先假設輸出特徵圖之寬度為w，高度為h，通道數量為 c，再使用全局平均池化(global average pooling)將每個通道的二維特徵$w \times h$壓縮為單一個特徵值，壓縮後的特徵圖大小即為$1 \times 1 \times c$。接著透過兩個全連接層以及兩個非線性激活函數來學習權重值，每個通道都會生成一個對應的權重值，最後再將權重值加權整合到通道的二維特徵中。

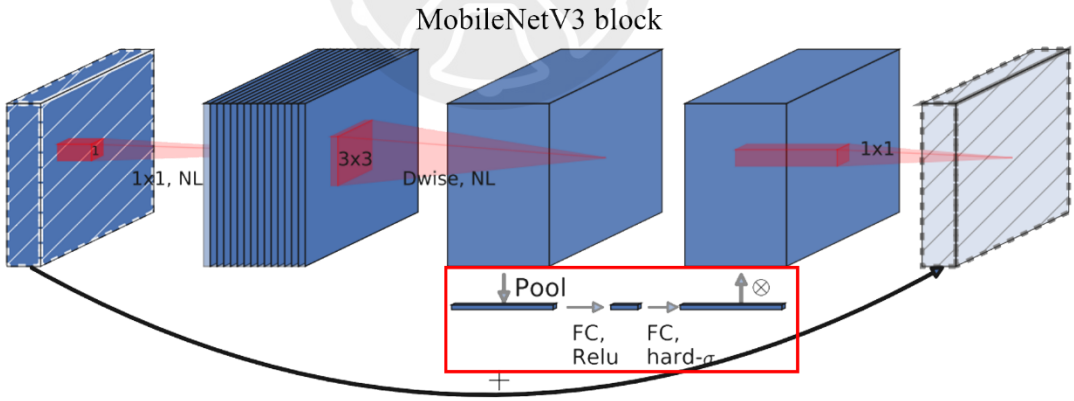


圖9：MobileNetV3架構示意圖[How19]

該研究者在實驗中發現將 ReLU 函數替換成 swish 函數可以提高MobileNetV3模型準確率。Swish 函數如公式(1)所示，其中$\sigma(x)$為 sigmoid 函數。由於 swish 函數中的 sigmoid 函數需要進行指數運算，計算成本較高，因此設計出 h-swish 函數，以符合輕量級神經網路中運算速度的需求。在 h-swish 函數中使用 $\frac{\mathrm{ReLU6}(x+3)}{6}$替代並近似 sigmoid 函數，如公式(2)所示。

$$\text{swish}(x) = x \cdot \sigma(x) \tag{1}$$

$$\text{h-swish}(x) = x\frac{\text{ReLU6}(x+3)}{6} \tag{2}$$

圖10則顯示 swish 函數和 h-swish 函數的比較結果。觀察圖10可以發現，swish 與 h-swish 函數兩者具有相似性，而 h-swish 運算速度更快同時可以維持一定的準確率。

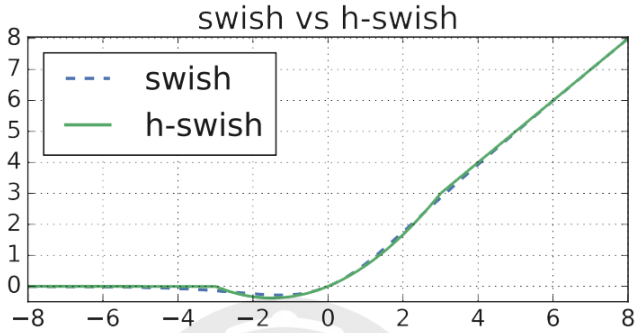


圖10：Swish 函數與 h-swish 函數比較圖[How19]

輕量級神經網路架構除了上述 MobileNet 系列外，還有 ShuffleNet 系列。Zhang 等人[Zha18]於西元2018年提出 ShuffleNetV1輕量級網路架構。該研究者在 ShuffleNetV1 架構中引入群組卷積(group convolution)以及通道洗牌(channel shuffle)兩種技術。群組卷積技術可用來降低計算量，而通道洗牌技術可使得特徵資訊能在不同特徵通道中進行流動。群組卷積技術是將特徵圖進行分群，以分群後的群組為單位分別進行卷積運算，最後再將其輸出按照原先的順序進行合併如圖11(a)所示。這樣分群的優點是可以減少卷積運算的計算量，但由於每個輸出通道只具該群的特徵資訊，因此會造成特徵資訊無法在通道之間進行流動。

圖11(b)及圖11(c)則是針對上述現象的改良方式。在圖11(b)中 GConv1中各群組輸出特徵圖將分割成子群組後進行所屬群組的交換，來達到各群組間訊息交換的目的。圖11(c)為使用通道洗牌技術示意圖。假設卷積層分為g組，每組有n個通道，總共有$g \times n$個輸出通道，通道洗牌技術首先將輸出特徵圖重塑(reshape)為(g, n)，再轉置(transpose)為(n, g)，最後再攤平(flatten)分為g組做為下一層的輸入。使用通道洗牌技術亦可以透過上述流程交換特徵資訊，達到各群組間訊息交換的目的。

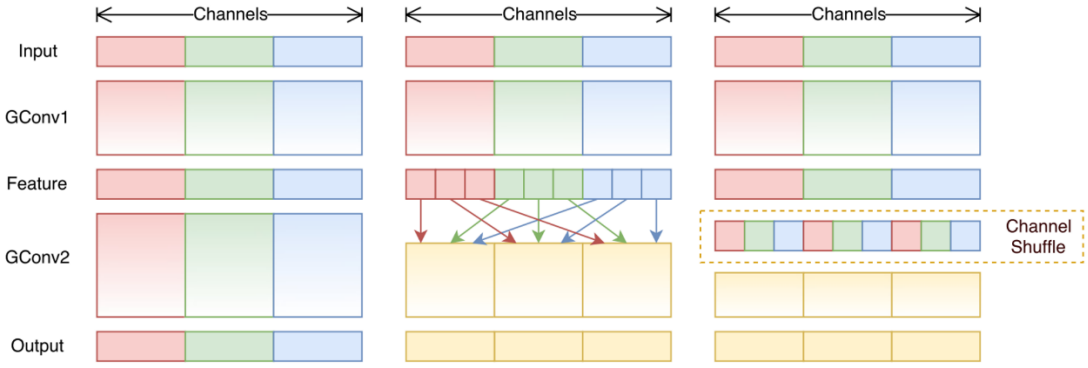


圖11：群組卷積與通道洗牌技術示意圖[Zha18](a)群組卷積架構(b)群組進行交換示意圖(c)通道洗牌技術示意圖

Ma 等人[Ma18]改良 ShuffleNetV1[Zha18]提出一 ShuffleNetV2架構。該研究利用圖12所示之實驗結果說明浮點數運算次數不能作為衡量系統運算速度的唯一指標。圖12中橫軸為每秒一百萬次浮點數運算次數(MFLOPs)，縱軸為運算速度。由圖12可以觀察到使用不同輕量級神經網路時，就算擁有大致相同的浮點數運算次數，運算速度仍會有所差異。以上實驗得出系統運算速度還需要考慮到其他各種因素的影響，例如記憶體存取成本(memory access cost)、GPU 平行度(degree of parallelism)等。

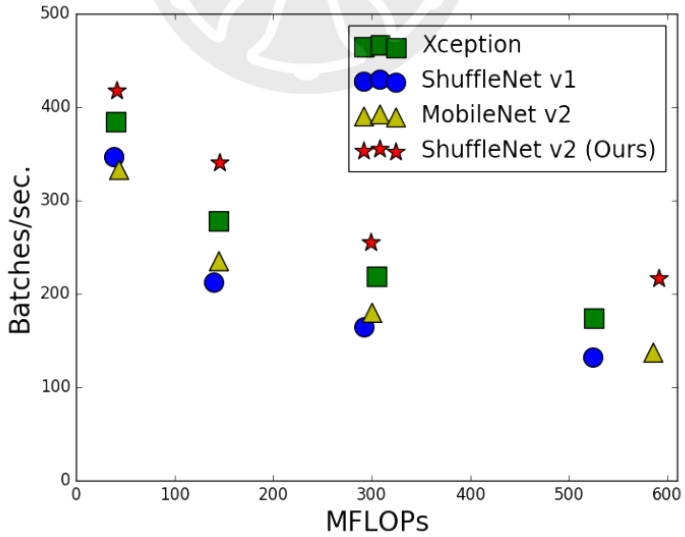


圖12：不同神經網路模型下運算速度與浮點數運算次數關係圖[Zha18]

Ma 等人針對上述觀察提出四種輕量級網路設計原則，第一種設計原則為輸入通道數量以及輸出通道數量兩者數量應相同。當輸入通道與輸出通道兩者數量相同時可最小化系統的記憶體存取成本。假設輸入通道個數為c_1，輸出通道個數為c_2，特徵圖長寬為$h \times w$，利用$1 \times 1 \times c_1$的卷積核做卷積運算時所產生浮點數

運算次數為B，B如公式(3)所示。此時記憶體存取成本 *MAC* 的公式則如(4)所示。根據均值不等式得出公式(5)。由公式(5)得知當輸入通道數量和輸出通道數量相等時，即$c_1 = c_2$時 *MAC* 為最小，使得記憶體存取成本最低。

$$B = h \times w \times c_1 \times c_2 \tag{3}$$

$$MAC = h \times w \times (c_1 + c_2) + c_1 \times c_2 \tag{4}$$

$$MAC \geq 2\sqrt{h \times w \times B} + \frac{B}{h \times w} \tag{5}$$

第二種設計原則為應將輸入資訊適度分組，不宜過度細分。雖然使用群組卷積技術可以降低計算量，但是輸入資訊分組數量過多時反而會增加記憶體存取成本。假設，輸入通道個數為c_1，輸出通道個數為c_2，特徵圖長寬為$h \times w$，分組數量為g，卷積核大小為$1 \times 1 \times c_1$，群組卷積的浮點數運算次數公式如(6)所示，記憶體存取成本 *MAC* 公式如(7)所示。由公式(7)可知，當h, w, c_1, B為固定的情況下，g和 *MAC* 會成正比，因此 *MAC* 會隨著分組數變多而增加。

$$B = \frac{h \times w \times c_1 \times c_2}{g} \tag{6}$$

$$MAC = h \times w(c_1 + c_2) + \frac{c_1 \times c_2}{g} = h \times w \times c_1 + \frac{B \times g}{c_1} + \frac{B}{h \times w} \tag{7}$$

第三種設計原則為應適度減少網路結構分支數量。網路結構分支數量過高會降低平行處理的能力，進而導致運算速度降低。圖13為多個分支結構上使用不同小型運算符組合。圖13顯示網路結構的碎片化(network fragmentation)示意圖，圖13(a)顯示一個碎片結構；圖13(b)顯示兩個碎片串列結構；圖13(c)顯示四個碎片串列結構，以上的網路結構中並沒有包含多分支結構。而圖13(d)與圖13(e)分別顯示兩個碎片與四個碎片的並列結構，是多分支的網路結構。雖然分支結構有助於提高系統的準確性，但對輕量級神經網路而言其在高度平行化運算下仍會降低運算速度，因此該研究建議適度減少網路結構分支數量。

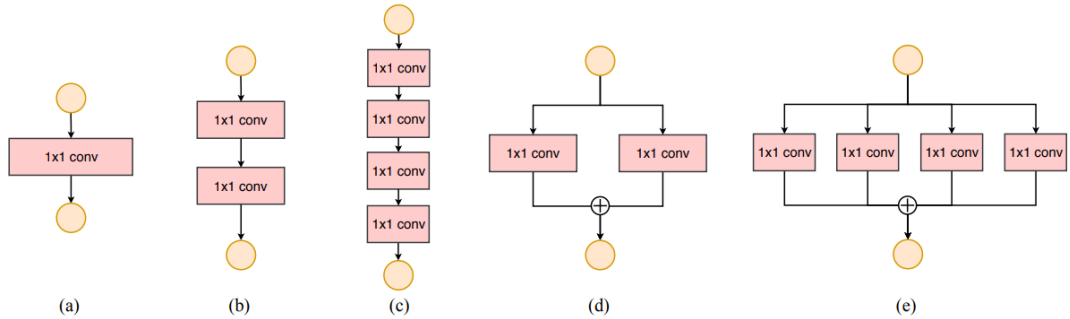


圖13：網路結構的碎片化示意圖[9](a)一個碎片結構(b)兩個碎片串列結構(c)四個碎片串列結構(d)兩個碎片並列結構(e)四個碎片並列結構

第四種設計原則為應減少逐元素(element-wise)操作，逐元素操作包含 ReLU 函數和加法運算。因為逐元素操作會增加記憶體存取成本，而運算速度大多只考慮到卷積運算時間並未考慮記憶體存取成本。圖14中顯示 ShuffleNetV1與 ShuffleNetV2模型之使用 GPU 時間分析，觀察圖14可以發現逐元素操作也占了相當多的 GPU 時間(圖14(a)顯示為15%；圖14(b)顯示為23%)，因此該研究建議要儘量避免使用逐元素操作。

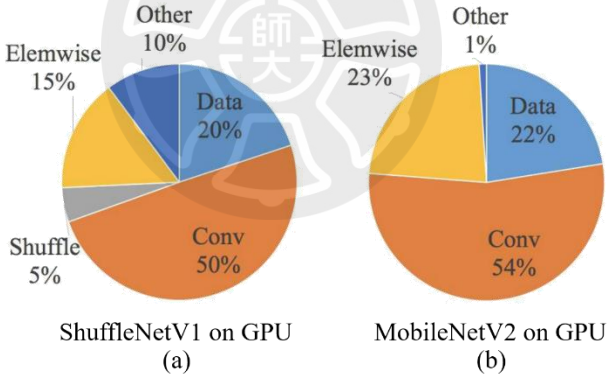
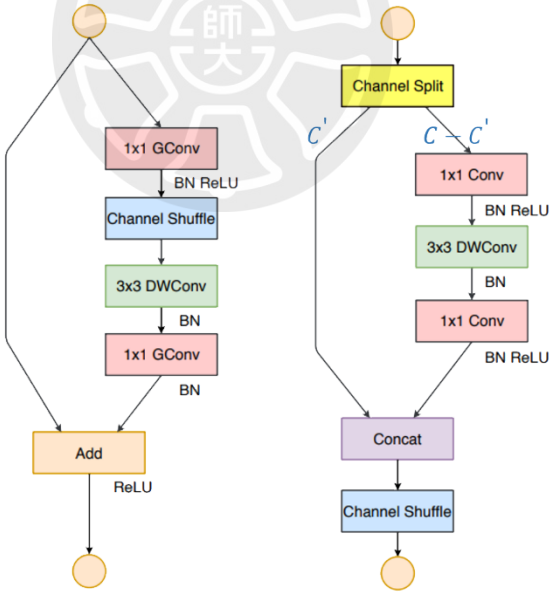


圖14：ShuffleNetV1與 ShuffleNetV2模型之使用 GPU 時間分析[Ma18]

在基於上述所提出的四種輕量級網路設計原則，Ma 等人針對 ShuffleNetV1[Zha18]架構(圖15(a))進行改良提出 ShuffleNetV2架構(圖15(b))。圖15(b)顯示在 ShuffleNetV2架構中假設輸入通道數量為C，它將輸入通道進行通道分割(channel split)後，將輸入通道拆解成二組分別為C'以及$C - C'$個通道，將其中一組C'個通道資訊使用 shortcut(採用第三種設計原則)直接向下傳送。將另一組$C - C'$個通道資訊送進右邊分支進行卷積運算(此時採用第二種設計原則，不再分群組)，而其卷積運算具有相同的輸入通道數量以及輸出通道數量，滿足第一種設計原則。之後將二個分支資訊合併從 ShuffleNetV1的 add 運算更改為

ShuffleNetV2的拼接(concat)方式，並將 ShuffleNetV1中 add 運算後的 ReLU 函數調整至 ShuffleNetV2右邊分支最後一層的卷積層之後。上述兩種作法都可以減少逐元素操作(滿足第四種設計原則)，並保持通道數量不變(滿足第一種設計原則)。ShuffleNetV2架構最後再進行通道洗牌操作(channel shuffle)讓兩個分支進行特徵資訊流動。實驗結果顯示使用上述四個設計原則，ShuffleNetV2會比其他神經網路在相同或者更少浮點數運算次數的情況下維持良好的準確率。

在實驗測試時，該研究將各式輕量級神經網路使用 CIFAR-10資料集進行測試[10]。其實驗測試分析結果分為準確率比較、浮點數運算次數比較、參數量比較、模型大小比較、記憶體存取成本比較、運算速度比較以及訓練模型時間比較等七項指標，其中運算速度與準確率這兩個指標被視為輕量級網路最重要的指標。ShuffleNetV2是在這兩個指標評比當中表現最好的輕量級神經網路。因此本研究決定採用 ShuffleNetV2作為高爾夫揮桿分解動作系統的基礎並且再進一步的改良使其更符合本應用的需求。



(a)ShuffleNetV1　　(b) ShuffleNetV2

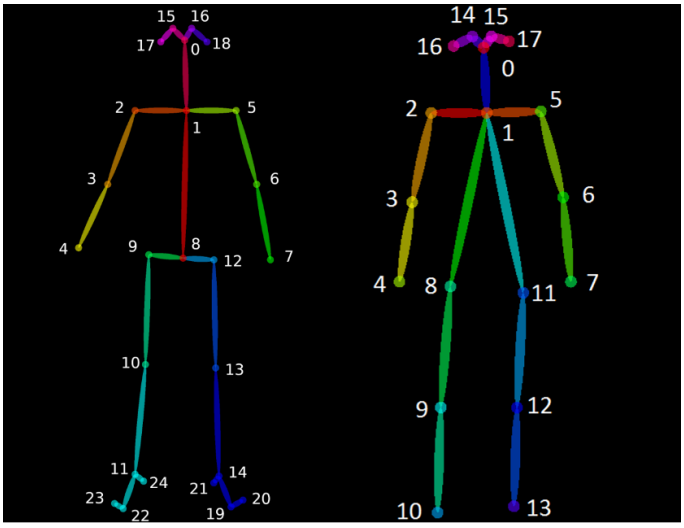
圖15：ShuffleNetV1與 ShuffleNetV2架構圖之比較[Ma18](a)ShuffleNetV1架構圖 (b)ShuffleNetV2架構圖

第四節 人體結構表示法分析

一般而言，影像中的人體結構資訊可以分為二維平面空間以及三維立體空間兩種。以二維平面空間來說，影像中的人體最常以二維人體骨架來描述；而在三維立體空間中則以三維人體骨架以及三維人體模型來描述。人體結構資訊可以用來進行人體監控、人機互動與動作辨識等應用。以下將分別介紹二維人體骨架結構、三維人體骨架結構及三維人體模型結構。

1. 二維人體骨架結構

二維人體骨架結構主要是以骨架中關節點的二維座標來表示。二維人體骨架結構的描述以人體關節點二維座標按照特定順序相連而成。Cao 等人[Cao17]於西元2017年提出 Openpose 系統來進行二維人體骨架估計。Openpose 系統內建兩種二維人體骨架結構描述方式，如圖16所示。圖16(a)顯示25個二維人體關節點的骨架描述方式，而圖16(b)則顯示18個二維人體關節點的骨架描述方式。上面所述並不是唯二的二維人體關節點的骨架描述方式，研究者因應用領域的不同需求，可以制訂各種不同數量以及不同連結順序的關節點來描述人體骨架。由於人體骨架中關節點的位置是以投影的方式投影至二維座標空間，所以關節點間的距離有可能忽長忽短，不像真實的人體骨架其關節點間的距離具固定的長度，因此在人體骨架估計時較容易失真，不易校正。

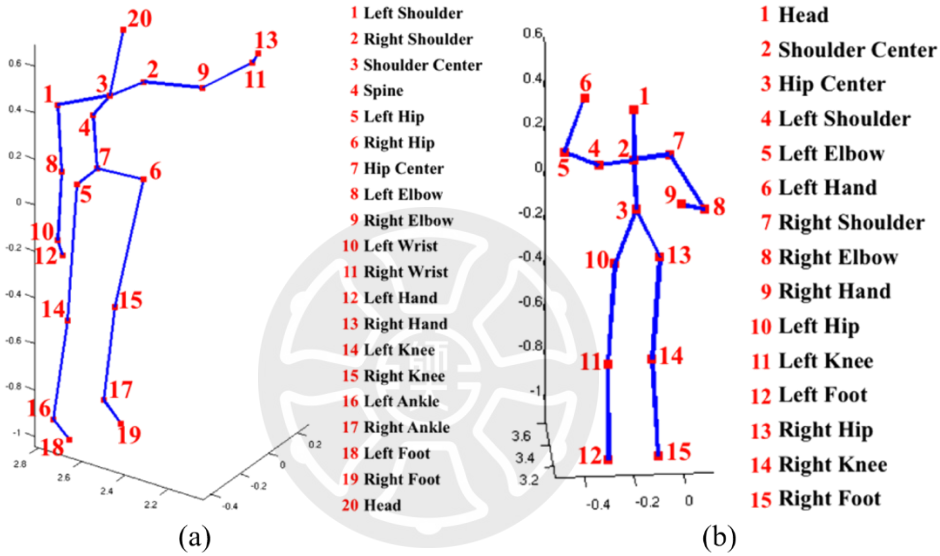


(a) (b)

圖16：Openpose 系統之二維人體骨架描述方式[Cao17](a)25個人體關節點描述示意圖(b)18個人體關節點描述示意圖

2. 三維人體骨架結構

三維人體骨架結構主要是以骨架中關節點的三維座標來表示。三維人體骨架結構的描述以人體關節點三維座標按照特定順序相連而成。圖17為兩種三維人體骨架結構描述方式，圖17(a)為 MSRA-3D dataset 顯示20個三維人體關節點的骨架描述方式，以及圖17(b)為 UCF kinect dataset 顯示15個三維人體關節點的骨架描述方式。將三維人體骨架與二維人體骨架比較可知，三維人體骨架中關節點之間的距離較具真實性與固定性，同時可以反應出三維空間中身體各部位所呈現的骨架角度，但在人體動作估計時亦有精確度不足的缺點。



(a) (b)

圖17：三維人體骨架及對應關節點示意圖[Pre16](a)MSRA-3D dataset 20個三維人體關節點描述示意圖(b) UCF kinect dataset 15個三維人體關節點描述示意圖

3. 三維人體模型結構

Loper 等人[Lop15]於西元2015年提出 SMPL 模型(a skinned multi-person linear model)如圖18所示，是一個三維人體模型結構。SMPL 模型使用6890個節點(vertices)連接組成人體網格(mesh)，24個關節點(joint)控制人體姿勢，同時使用10維向量描述人體體型(shape)。圖18(a)顯示一標準三維人體模型，其中$\bar{T}$為平均模板形狀(mean template shape)，W為各個關節點的權重，此標準三維人體模型採用平均模板形狀和權重建構，尚未添加姿勢參數和形狀參數。

圖18(b)是以圖18(a)為基礎經過高矮胖瘦調整後的三維人體模型，其中$B_S(\vec{\beta})$為 blend shape function，將形狀參數映射到6890個節點位置上；$J(\vec{\beta})$則會將形狀

參數映射到24個關節點的位置上。圖18(c)是以圖18 (b)為基礎添加不同人體姿勢對人體局部部位體態變化的三維人體模型，其中$B_P(\vec{\theta})$會將姿勢參數映射6890個節點位置。圖18(d)是以圖18(c)為基礎加上不同人體姿勢中皮膚所發生改變的三維人體模型。當人體姿勢改變導致關節點移動時，由節點所組成的皮膚(skin)將會隨著關節點移動而產生變化，此過程稱為蒙皮(skinning)。

三維人體模型與前兩種方法相比之下，三維人體模型不僅擁有人體關節點資訊，還可以依據人體結構，隨著變換身形以及姿勢動作，使得身體部位產生自然的變形，較具真實性與精確性。總體而言，三維人體模型結構相較人體骨架結構更為擬真人體。

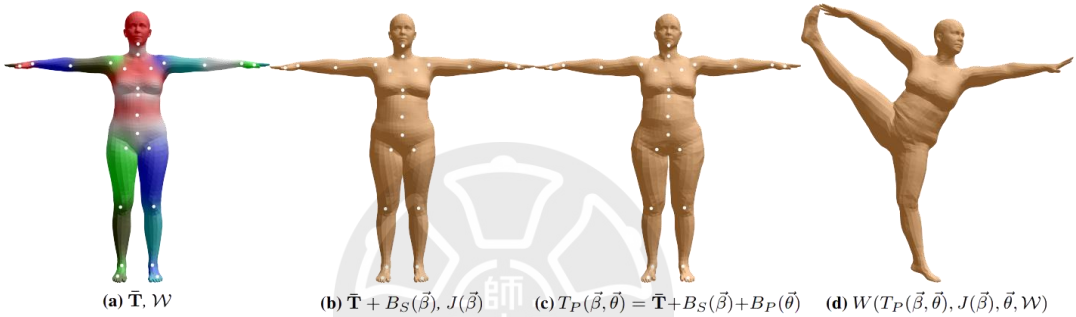


圖18：SMPL 模型示意圖[Lop15]

第五節 三維人體模型建構及應用

本節將討論三維人體模型建構及應用，首先從影像建構出三維人體模型有兩種方式，分別為直接估計法以及兩階段法。直接估計法並沒有使用到二維人體資訊，而是直接從影像預測出三維人體模型。兩階段法先將影像預測出二維人體資訊，接著再使用迴歸分析(regression analysis)或者模型擬合(model fitting)將二維人體資訊預測三維人體關節點，最後再使用三維人體關節點預測人體姿勢估計模型。

(1) 三維人體模型直接估計法

Kanazawa 等人[Kan18]於西元2018年提出 HMR(human mesh recovery)的三維人體模型建立之架構，其架構圖如圖19所示。首先將二維影像輸入後，經過 encoder 進行特徵擷取，接著使用迴歸器(regression)直接估計出三組不同的參數值。這三組參數分別為攝影機(camera)參數、形狀(shape)參數以及姿勢(pose)參數。

攝影機參數包括攝影機距離、角度及平移量；形狀參數包括代表人體高矮胖瘦等10個參數，而姿勢參數則包含人體骨架24個關節點的資訊。接著利用這些參數值建構三維人體模型後，重新投影到影像平面中則可以得到該模型二維的人體投影位置，經損失函數(loss function)計算輸入影像中人體形狀與三維人體模型投影形狀的差異後可再次調整參數值。另一方面，三維人體模型的參數可輸入到一個鑑別器(discriminator)中，由鑑別器來判斷模型是否來自於真實的人體參數。

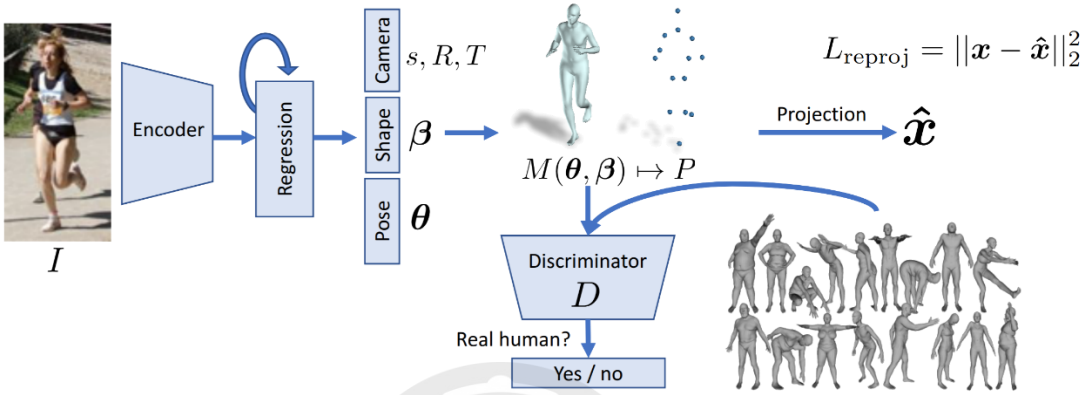


圖19：Kanazawa 等人所提出之 HMR 3D 人體姿態模型[Kan18]

Kocabas 等人[Koc20]於西元2020年提出名為 VIBE(video inference for human body pose and shape estimation)的三維人體模型建立之架構，VIBE 架構如圖20所示。先輸入二維影像經過 CNN 進行擷取影像中空間結構的特徵，並使用 GRU(gate recurrent unit)學習影像序列間的時間特徵，接著使用迴歸器估計出三維人體模型。並將三維人體模型參數輸入進鑑別器中，而在鑑別器中則加入自我注意力機制(self-attention)，以及採用 AMASS(archive of motion capture as surface shapes)資料庫[Mah19]所提供的人體動作資訊，利用鑑別器來判別出輸入的模型是否來自於真實的人體參數。此方法會使得生成的影像序列看起來更真實以及更為流暢。

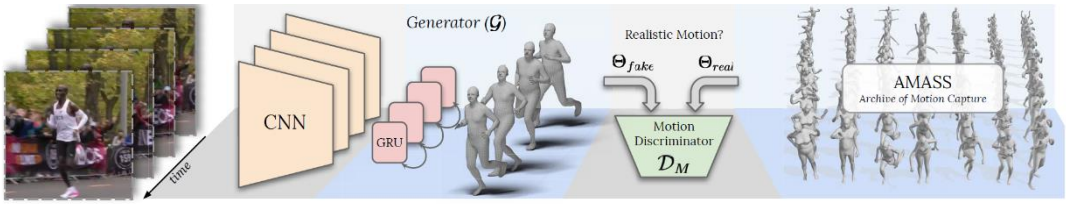


圖20：Kocabas 等人所提出之 VIBE 架構示意圖[Koc20]

Choi 等人[Cho21]改良 VIBE[Koc20]提出一 TCMR (temporally consistent mesh recovery system)架構，其架構圖如圖21所示。在該架構中影像輸入後皆使用 CNN 中的 ResNet 進行空間特徵擷取，之後使用三個 GRU 得到分別為整部影片的時

間特徵、當前幀以外過去的時間特徵以及當前幀以外未來的時間特徵。利用上述三種時間特徵分別預測出當前的三維人體模型，再將三個三維人體模型參數結合，即可產生當前幀的三維人體模型。此方法主要是考慮人體動作的連續性，採用過去與未來的人體動作資訊來校正當前幀的人體動作預測結果，會使得人體動作更具流暢性。

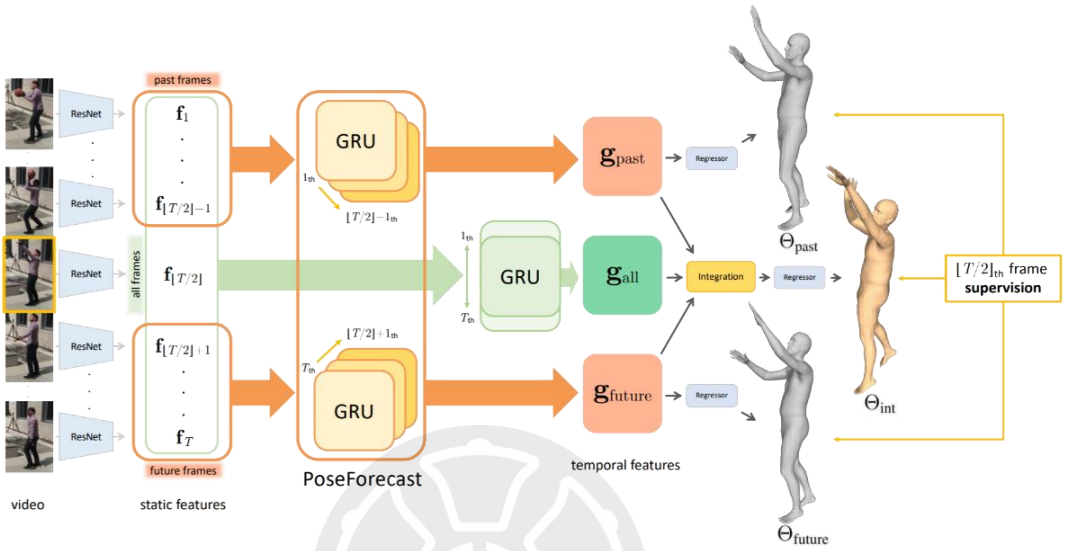


圖21：Choi 等人所提出之 TCMR 架構示意圖[Cho21]

Zhang 等人[Zha21]於西元2021年提出名為 PyMAF(3D human pose and shape regression with pyramidal mesh alignment feedback loop)的架構，其架構如圖22所示。由圖22可知，當二維影像輸入該系統後經過 encoder 可擷取第一階段人體動作空間特徵，接著利用特徵金字塔逐層擷取呈現在不同階段的動作空間特徵。這些不同大小的特徵圖將被進一步用來擷取 grid feature，grid feature 送入迴歸器後估計出三維人體模型。這一個三維人體模型會再度被投影(projection)至二維的特徵圖上進行差異比較，差異比較的結果將用來調整三維人體模型，直到系統穩定為止，此時即可輸出三維人體模型的預測結果。該預測三維人體模型整體而言分為兩大部分，即 auxiliary pixel-wise prediction 以及 mesh alignment feedback loop。其中 auxiliary pixel-wise prediction 部分是預測出人體動作的空間特徵，藉此利用預測出的人體資訊使得預測更為準確。而 mesh alignment feedback loop 則可不斷調整三維人體模型的參數值，讓三維人體模型逐漸導向正確的預測位置。

三維人體模型直接估計法可以直接從影像預測出三維人體模型，因此系統處理時間較為快速。但是直接估計法需要大量三維人體姿勢資料作為訓練，而這些

資料需要以架設多台攝影機並在拍攝者全身標註光學標識點的方式進行收集，因此大多數只能在特定的實驗室拍攝及蒐集。上述的資料收集方式具拍攝環境中擁有相同背景或是背景不複雜的缺點。因此根據實驗室拍攝的三維人體姿勢資料集訓練出來的模型很難有效利用在野生環境(wild)中，容易導致準確率下降。

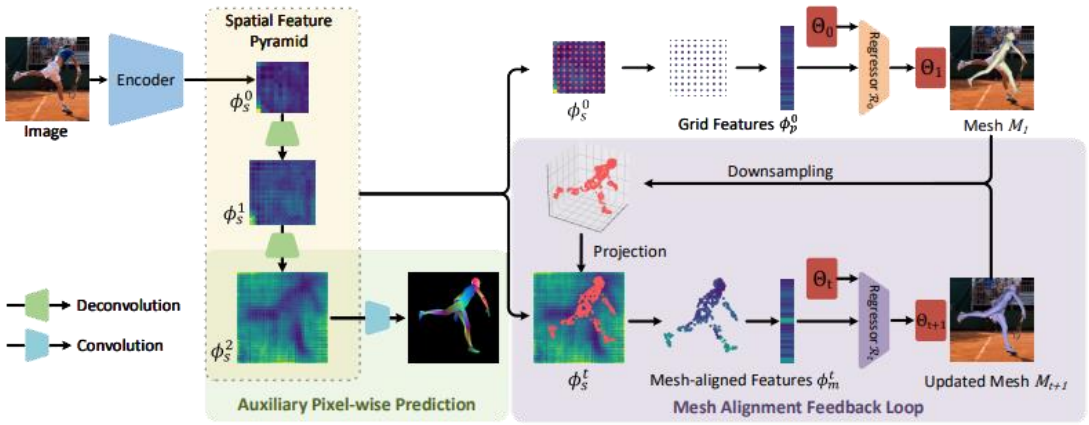


圖22：Zhang 等人所提出之 PyMAF 架構示意圖[Zha21]

(2) 三維人體模型二階段法

Bogo 等人[Bog16]於西元2016年提出名為 SMPLify(automatic estimation of 3D human pose and shape from a single image)的架構，其架構之執行範例如圖23所示。系統首先將影像圖23(a)輸入至 DeepCut[Pis16]預測影像中人體14個二維人體關節點的位置(圖23(b))，再使用三維人體模型的三維人體關節點投影到影像平面中擬合至預測的二維人體關節點位置(圖23(c))，找出誤差最小之三維人體模型。此方法是使用二維人體骨架來預測三維人體模型。圖23(d)是從不同角度顯示的最終三維人體模型。

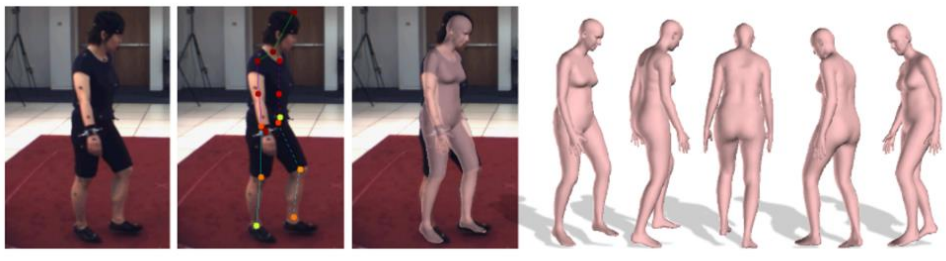


圖23：Bogo 等人所提出之 SMPLify 架構執行範例[Bog16](a)輸入影像(b)人體二維關節點預測(c)三維人體模型估計(d)不同角度三維人體模型展示

Omran 等人[Omr18]於西元2018年提出名為 Neural Body Fitting(unifying deep learning and model-based human pose and shape estimation)的深度學習架構來估計人體姿勢和形狀，其架構如圖24所示。系統首先使用 CNN 模型進行影像中的語意分割(semantic segmentation)，分別切割出人體和背景。語意分割時會使用不同顏色代表人體不同身體部位，此方法將二維影像中的人體劃分成12個人身體部位，如圖24所示。再將人體語意分割結果輸入另一個 CNN 預測出三維人體模型的參數，建構三維人體模型。將此三維人體模型重新投影至影像中可以得到該模型二維的人體投影位置，經過損失函數計算輸入影像中人體形狀與三維人體模型投影形狀的差異後可進行參數值調整。使用語義分割技術可以減少 RGB 影像中如背景複雜與光線等因素之干擾，比直接預估法更具有穩定性。

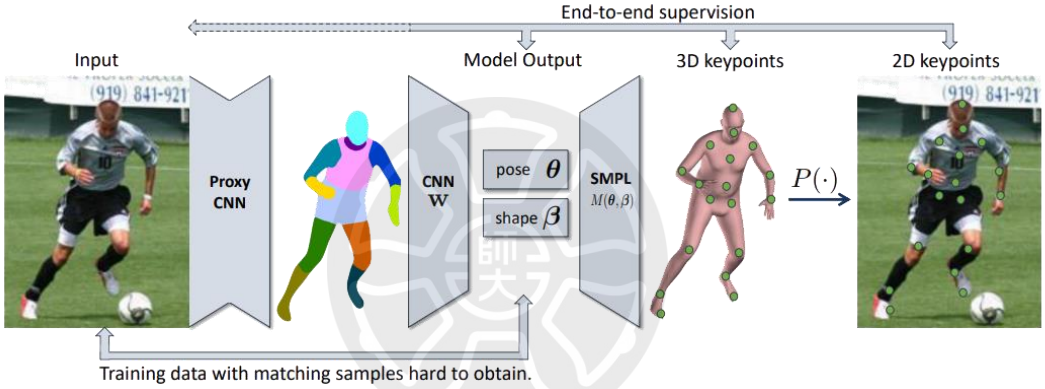


圖24：Omran 等人所提出之 Neural Body Fitting 架構圖[Omr18]

Kolotouros 等人[Kol21]於西元2021年提出名為 ProHMR (probabilistic modeling for human mesh recovery)的架構，其架構如圖25所示。圖25中第一行顯示的流程為單張影像輸入後，經過 encoder 進行特徵擷取，接著使用 normalizing flow[Rez15]技術預測三維人體模型參數的機率分佈(distribution)，再使用最大概似估計(maximum likelihood estimation)從機率分佈中找出投影後和二維影像最為相似的三維人體模型，達到三維人體模型建構的目的。圖25中第二行顯示另一種類似的流程，但增加二維人體骨架資訊協助預測三維人體模型。該流程顯示單張影像輸入後，一樣經由特徵擷取步驟並預測出三維人體模型機率分佈，接著將可能的三維人體模型投影至影像平面中，經過損失函數計算其三維人體模型投影後和二維人體骨架資訊的差異，依此選出損失函數差異最小的三維人體模型。

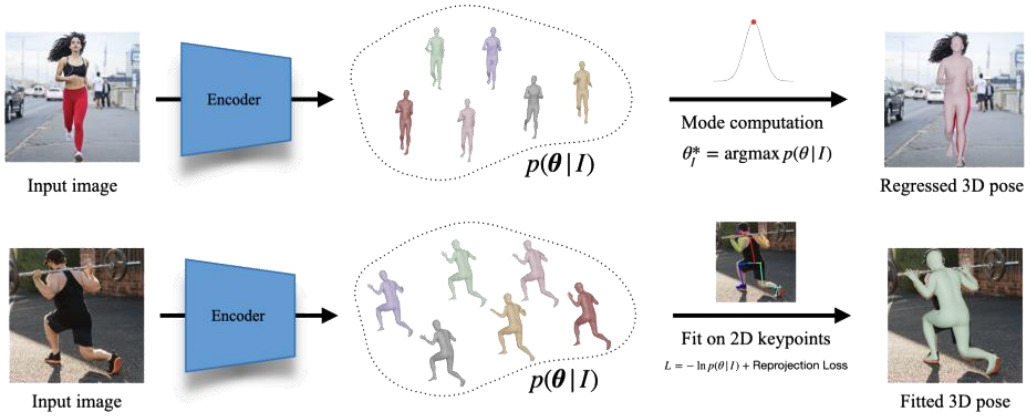


圖25：Kolotouros 等人所提出之 ProHMR 架構示意圖[Kol21]

如上所述，三維人體模型直接估計法可以直接從二維影像預估出三維人體模型，因此系統處理時間相對快速。而三維人體模型二階段法雖然需要花費些許時間處理二維人體資訊，但是利用二維人體資訊再預測出三維人體模型，可以獲得更多人體資訊進而提升準確率。三維人體模型二階段法與三維人體模型直接估計法兩者相比之下，三維人體模型二階段法準確率較高。由於本研究所開發之系統為高爾夫揮桿動作姿勢分析系統，對預測準確率具相當要求，因此選擇三維人體模型二階段法來預測三維人體模型。

而近年來利用三維人體模型模擬人體或預測人體動作的技術更加成熟準確後，開始被應用至如生物醫學、線上會議、互動遊戲、電影製作與運動訓練等各個領域。因此本研究亦利用三維人體模型較具真實性、精確性及擬真人體等特性，將三維人體模型預測人體動作的技術應用在高爾夫揮桿動作姿勢分析系統。

Fieraru 等人[Fie21]於2021年提出三維人體模型健身訓練自然語言反饋系統。該系統利用三維人體骨架將學員健身動作和教練健身動作相互比對，比對出二者動作上之差異，再利用自然語言呈現出學員需修正的健身動作資訊。Fieraru 等人所提出的健身訓練反饋示意圖如圖26所示。圖26中第一行為學員健身動作影像；第二行為學員三維人體模型；第三行為教練相對應的健身動作影像；而第四行則為學員和教練健身動作相比之下，其健身動作需修正的自然語言反饋內容。

Xie 等人[Xie19]於2019年提出三維人體模型核心訓練視覺反饋系統。該系統將學員和教練兩者三維人體模型以腰部位置為基準進行重疊對齊，並且將三維人體模型標記出十個身體部位，標記位置分別在雙手、手肘、肩膀、膝蓋以及腳踝。接著使用不同顏色來表示學員和教練兩者動作之間的差距，從動作差距由遠到近

的距離設定為紅色、橙色、黃色以及黃綠色。Xie 等人開發的三維人體核心訓練系統介面如圖27所示，此系統為使用圖片反饋方法來進行姿勢校正。

圖26：Fieraru 等人所提出之三維人體模型健身訓練反饋示意圖[Fie21]

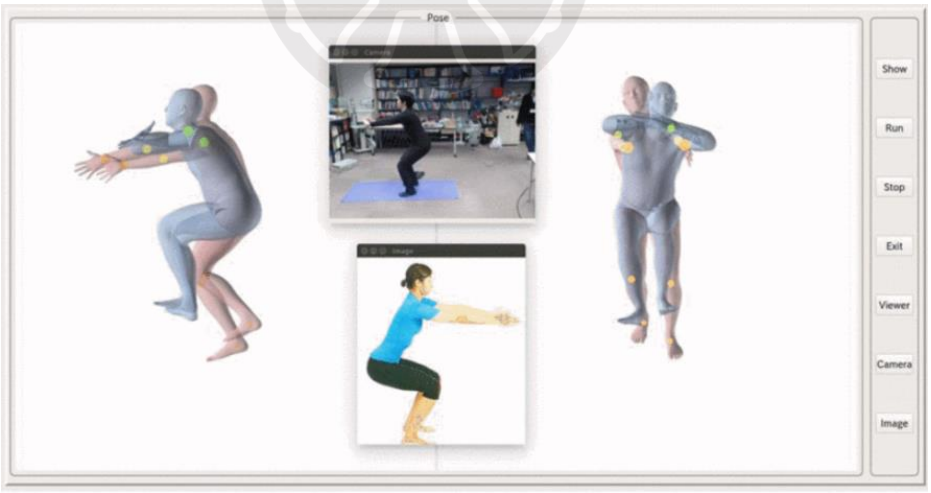


圖27：Xie 等人所提出之三維人體核心訓練系統介面[Xie19]

從以上文獻可以觀察到利用三維人體模型來模擬人體動作較具真實性、精確性及擬真性等特性。另一方面，使用文字反饋方法無法有效利用三維人體模型將人體資料視覺化的優勢，根據 Humans Process Visual Data Better[11]此篇內容，人類大腦處理影像的速度比起文字吸收快60000倍，並且有90%的資訊是由影像進

入到大腦中。圖片反饋和文字反饋相較之下，使用圖片反饋方法不但具精緻性，也更容易使人理解並更精準的傳遞訊息。

總而言之，在本章第一節討論之高爾夫揮桿分解動作中，本研究選擇使用將高爾夫揮桿動作分解為八個連續的分解動作。第二節討論之高爾夫揮桿分解動作系統分析中，McNally 等人[Mcn19]使用空間特徵和時間特徵兩者來達到互補的效果，因此本研究決定以 SwingNet [Mcn19]架構作為基礎開發高爾夫揮桿分解動作擷取。在本章第三節討論之輕量級網路中，ShuffleNetV2為運算速度和準確率兩者評量當中最優秀的輕量級網路，因此本研究決定採用 ShuffleNetV2作為基礎開發高爾夫揮桿分解動作擷取。在本章第四節討論之人體姿勢估計分析中，本研究決定使用可以表現出豐富人體資訊的三維人體模型。在本章第五節討論之三維人體模型建構及應用中，本研究決定使用兩階段法的 ProHMR[Kol21]預測三維人體模型以及使用圖片反饋方法進行姿勢校正。

本研究使用 GolfDB 資料集[Mcn19]進行訓練及測試。該資料集主要由來自 PGA、LPGA 以及 Champions Tours 等職業賽事中的職業高爾夫球選手組成，總共有1400部由不同角度拍攝之高爾夫揮桿影片，以及該資料集影片之拍攝速度區分為即時(real-time)及慢動作(slow-motion)，高爾夫揮桿影片解析度為720以及總長度超過390k 幀數。

本研究利用上述所提之技術改良並整合成視覺式智慧型高爾夫揮桿動作姿勢分析系統，以期能讓使用者隨時關注自身高爾夫揮桿動作的正確性，達到校正高爾夫揮桿動作之目的。

第 3 章 視覺式高爾夫揮桿動作姿勢分析系統

本研究提出之視覺式智慧型高爾夫揮桿動作姿勢分析系統主要是由輕量級神經網路和循環神經網路建構而成，並搭配三維人體模型進行高爾夫揮桿姿勢分析。本章將會介紹視覺式智慧型高爾夫揮桿動作姿勢分析系統的架構流程，以及說明本研究所改良之相關技術。

第一節 系統流程

視覺式智慧型高爾夫揮桿動作姿勢分析系統之流程如圖28所示。本系統主要分為兩大步驟：高爾夫揮桿分解動作擷取以及三維人體模型姿勢比對分析。第一大步驟高爾夫揮桿分解動作擷取主要分為三個步驟，分別為資料前處理(data preprocessing)、關鍵動作幀估計(key pose frame estimation)以及關鍵動作幀判定(key pose frame determination)。第二大步驟三維人體模型姿勢比對分析主要分為三個步驟，分別為二維人體骨架估計(2D pose estimation)、三維人體模型估計(3D human model estimation)以及人體模型對齊比對(alignment and comparison)。使用者影片或教練影片輸入時，上述第一大步驟可以平行處理(如圖28藍色區域所示)，而第二大步驟則將第一大步驟二者之處理結果整併處理(如圖28綠色區域所示)。

在高爾夫揮桿分解動作擷取時，系統輸入使用者之高爾夫揮桿影片(student golf swing video)與教練高爾夫揮桿影片(coach golf swing video)，分別對影片進行資料前處理，即使用目標檢測方法擷取出影像中的前景物件(即人體)，並調整輸入影像大小以符合後續步驟之要求。接著再分別將二個影片輸入至同一個ShuffleNetV2模型擷取動作特徵，根據擷取的特徵使用 Bi-GRU 進行各分解動作關鍵動作幀估計(key pose frame estimation)，為方便論文後續說明，本研究將ShuffleNetV2和 Bi-GRU 原型架構命名為 GSNet(golf swing net)。系統初步獲得使用者與教練的八個高爾夫揮桿分解動作關鍵動作幀之後，最後使用關鍵動作幀判定(key pose frame determination)將預測結果使用本系統研發出的校正方法調整八個高爾夫揮桿分解動作的關鍵動作幀。

在三維人體模型姿勢比對分析時，首先系統分別預估出使用者與教練高爾夫揮桿八個分解動作的關鍵動作幀之二維人體骨架(2D pose estimation)。接著利用八個高爾夫揮桿分解動作關鍵動作幀和它們對應的二維人體骨架估計出八個三維人體模型(3D human model estimation)，再把使用者以及教練揮桿分解動作所對應的三維人體模型一一進行對齊比對(alignment and comparison)分析出每一個揮桿分解動作之差異。最後系統輸出(system output)分解動作比對後差異最多的身體部位名稱以及兩者三維人體模型重疊之反饋圖片。

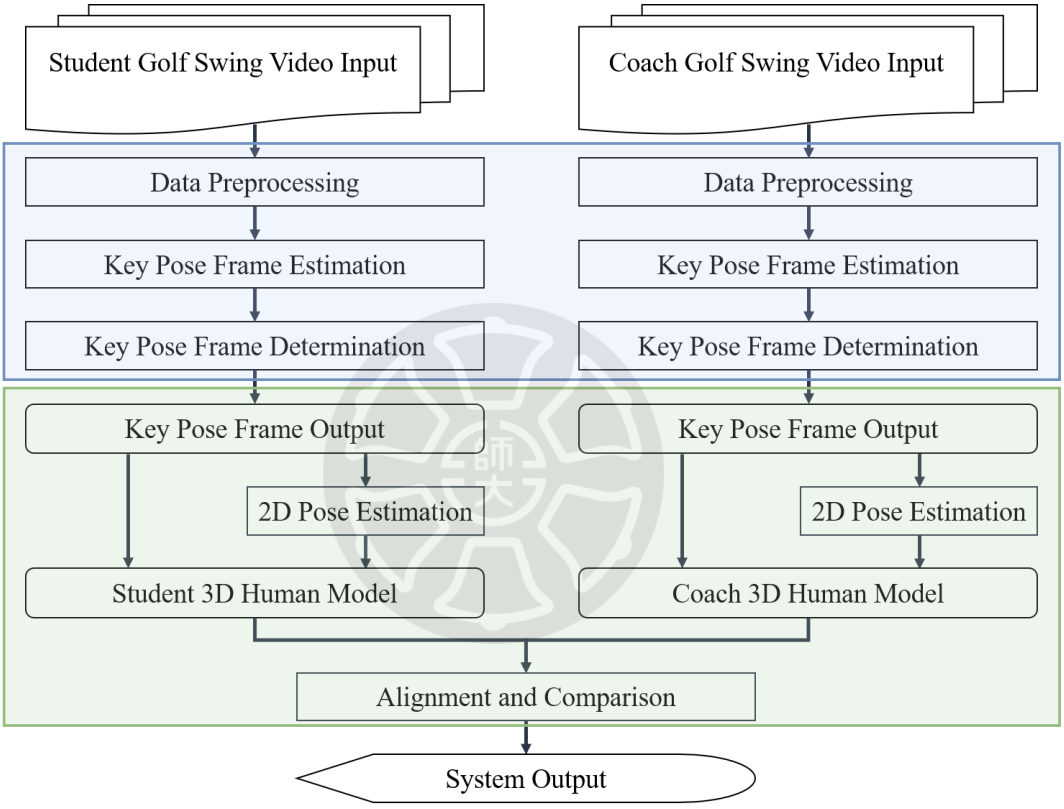


圖28：視覺式高爾夫揮桿動作姿勢分析系統流程圖

以下針對高爾夫揮桿分解動作擷取以及三維人體模型姿勢比對分析兩大步驟進行較為詳細的說明。如前所述，高爾夫揮桿分解動作擷取主要分為資料預處理、關鍵動作幀估計以及關鍵動作幀判定三個步驟；而三維人體模型姿勢比對分析主要可分為二維人體骨架估計、三維人體模型估計以及人體模型對齊比對三個步驟。

(1) 資料前處理(data preprocessing)

本系統的資料前處理方式，為調整影像大小並利用幀差法(frame difference

method)[Sin14]擷取二張影像差異。首先將輸入之彩色影片大小調整為224×224像素並轉成灰階，然後使用幀差法找出影片中連續二幀影像間的差異，並利用該差異偵測前景物是否移動。給定連續二幀影像I_k和I_{k+1}，針對座標位置為(x, y)之像素，其幀差之計算公式如下:

$$\mathrm{I}_{d(k,k+1)}(x, y) = |\mathrm{I}_{k+1}(x, y) - \mathrm{I}_k(x, y)| \tag{9}$$

其計算結果如圖29所示。圖29為一高爾夫揮桿影片影像序列由第k幀至第$k + N$幀，上排為 RGB 影像序列，而下排則是幀差影像序列。值得一提的是調整影像大小的原因是為輕量級網路 ShuffleNetV2模型所需，而採用幀差法是為突顯前景物輪廓，分離動態前景和背景，降低背景干擾，擷取人體動作特徵。若 GSNet 架構之輸入影片為經由幀差法計算後之影像，則本研究將此架構命名為 GSNetV1。

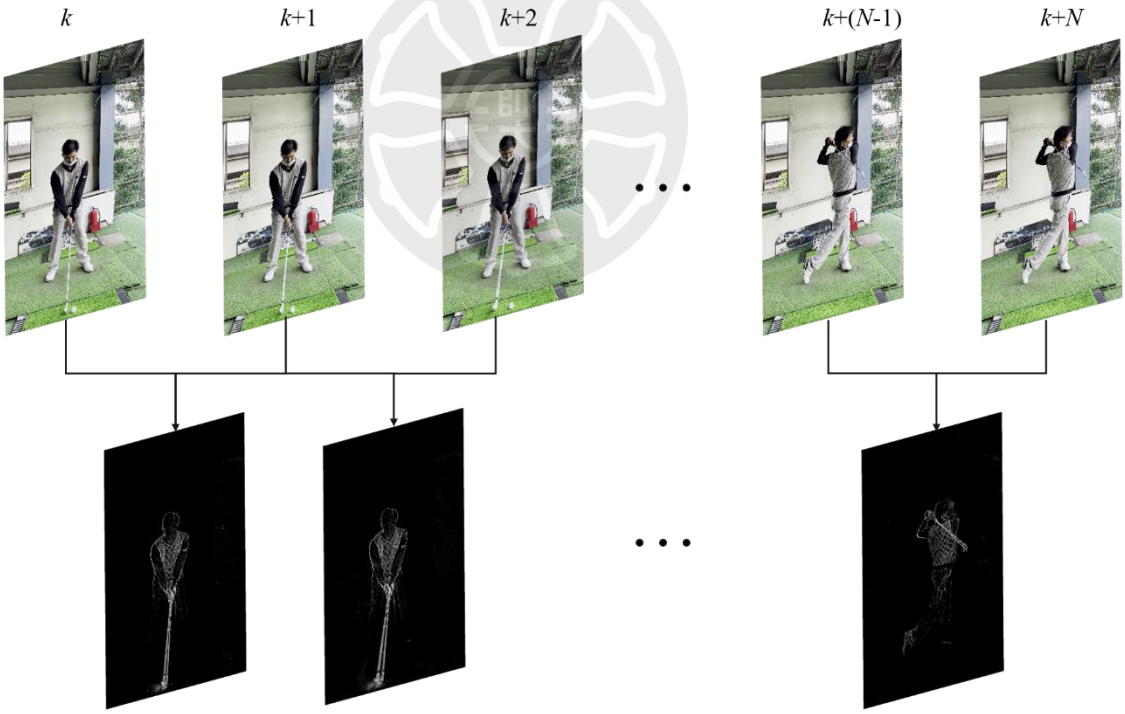


圖29：幀差法範例示意圖

(2) 關鍵動作幀估計(key pose frame estimation)

基於第2章第三節中所提輕量級神經網路各項指標評比結果，本研究決定採用 ShuffleNetV2模型進行改良。系統首先使用 ShuffleNetV2擷取影像中的空間特徵。另外，因高爾夫揮桿動作存在著時間前後順序，而選擇使用能夠考慮到上下

文關係的 Bi-GRU 得到影像序列的時間特徵，利用這二種特徵進行關鍵動作幀估計，而本研究將 ShuffleNetV2和 Bi-GRU 原型架構命名為 GSNet，架構圖如圖30所示。

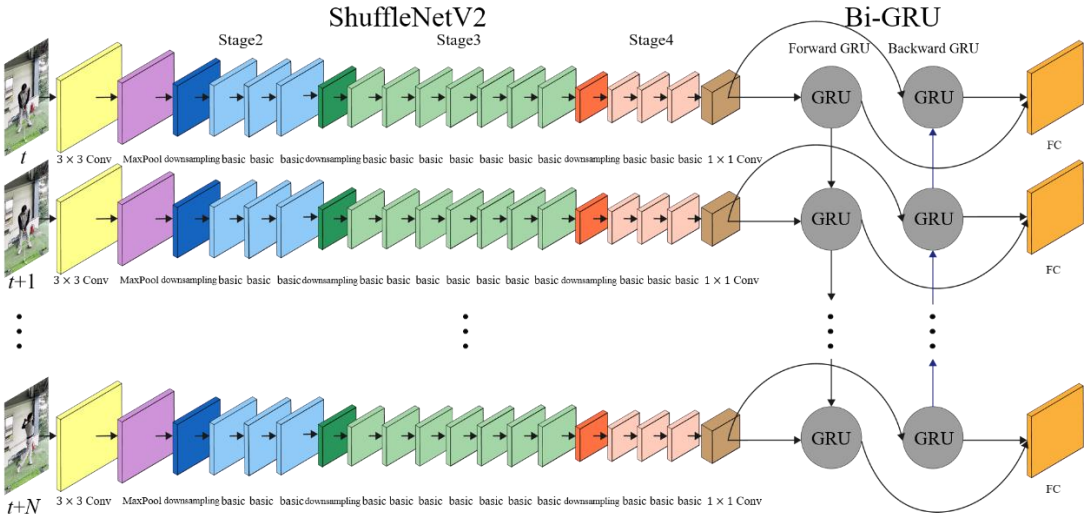


圖30：GSNet 架構圖

表4：GSNet 架構測試結果

編號	分解動作名稱	真實影片準確率	慢動作影片準確率	總體準確率
1	擊球準備	33.64%	25.08%	29.31%
2	起桿	89.13%	82.71%	85.88%
3	上桿	90.98%	81.69%	86.28%
4	上桿頂點	89.25%	69.60%	79.31%
5	下桿	97.68%	98.30%	98.00%
6	擊球	99.30%	98.53%	98.91%
7	送桿	97.11%	97.06%	97.08%
8	收桿	21.50%	20.22%	20.85%
	平均準確率	**77.32%**	**71.65%**	**74.45%**

本研究使用 GSNet 架構進行關鍵動作幀估計訓練。訓練完成後的測試結果如表4所示，表中列出真實影片準確率(real-time PCE)、慢動作準確率(slow-motion PCE)以及總體準確率(total PCE)。測試結果顯示除了擊球準備、上桿頂點以及收桿三個動作以外的其餘動作皆有達到80%以上的總體準確率，但擊球準備和收桿兩個動作的總體準確率分別只有29.31%以及20.85%。另外，真實影片準確率和慢

動作準確率相比之下，慢動作準確率的八個高爾夫揮桿分解動作大致上準確率都較低。而本研究將改良 GSNet 架構，將在後面的章節進行較詳細的介紹。

(3) 關鍵動作幀判定(key pose frame determination)

關鍵動作幀判定步驟是將關鍵動作幀估計後之八個高爾夫揮桿分解動作作為輸入，因為高爾夫揮桿動作具有時間前後順序，因此本研究應用高斯分佈(gaussian distribution)技術開發一校正方式用來調整關鍵動作幀估計之結果。該技術為本研究的改良技術，將在後面的章節進行較詳細的介紹。

(4) 二維人體骨架估計(2D pose estimation)

二維人體骨架估計步驟中系統輸入高爾夫揮桿分解動作擷取的八個高爾夫揮桿關鍵動作幀，透過使用 Openpose [Cao17]系統選用二維人體骨架25個關節點如圖16(a)所示，預測出其關節點的二維座標。本步驟每一關鍵動作幀皆需擷取出一個二維人體骨架，依前一步驟關鍵動作幀判定結果，每一個高爾夫揮桿動作將產出八組二維人體骨架的關節點座標。

(5) 三維人體模型估計(3D human model estimation)

三維人體模型估計步驟中輸入八個高爾夫揮桿分解動作關鍵動作幀和其對應的二維人體骨架一同輸入至 ProHMR[Kol21]系統中估計出三維人體模型。本研究主要目的在開發高爾夫揮桿動作姿勢分析系統，對三維人體預測結果準確率相當要求，因此選擇使用準確率較高的三維人體模型二階段法中的 ProHMR[Kol21]系統。

(6) 人體模型對齊比對(alignment and comparison)

本研究設定拍攝角度為高爾夫正面揮桿如圖1所示，因此將上個步驟估計出的使用者和教練兩者揮桿動作關鍵動作幀三維人體模型之三維人體關節點的0號關節點(圖33)位置對齊後，以及將教練的三維人體模型之體型調整至和使用者的三維人體模型為相同體型，接著嘗試將教練的三維人體模型進行多次小幅旋轉以便和使用者的三維人體模型角度對齊，儘可能達到兩者拍攝角度完全相同的情況。最後本系統使用三維人體模型節點座標做為比對依據，把使用者和教練兩者

的三維人體模型中每個對應的節點計算其歐幾里得距離(Euclidean distance)差，藉此分析使用者揮桿姿勢的正確性。本比對技術為本研究的改良技術，將在後面的章節進行較詳細的介紹。

綜整以上所述介紹之系統流程以及相關步驟，本章後續章節將詳細說明關鍵動作幀估計改良、關鍵動作幀判定以及人體模型對齊比對之作法。其中第二節說明關鍵動作幀估計改良；第三節說明關鍵動作幀判定；第四節說明三維人體模型姿勢比對。

第二節 關鍵動作幀估計改良

本系統處理的資訊為高爾夫揮桿影片，在使用深度學習技術時有 GPU 記憶體空間的需求，在批量正規化(batch normalization)時若批量大小(batch size)參數設定過大會導致系統訓練時 GPU 記憶體無法負荷。本研究引入群組正規化(group normalization)取代批量正規化(batch normalization)來解決這個問題。批量正規化是以單批輸入的資料進行正規化，而如果批量大小參數設定較小時反而會導致其訓練效果不佳。群組正規化能夠在批量參數設定較小時在系統訓練時依然維持良好的收斂速度及穩定性，因此本研究考慮以群組正規化技術來取代批量正規化技術。

另外，在第2章第三節曾提及 Howard 等人[How19]提出之 MobileNetV3架構將 ReLU 函數替換成 h-swish 函數來提升準確率同時維持輕量級神經網路運算速度，本研究亦引入 h-swish 函數取代 ReLU 函數來觀察其提升的效果。

最後本研究在 GSNet 架構中之 ShuffleNetV2模型加入 Wang 等人[Wan20]提出的 ECA-Net(efficient channel attention for deep convolutional neural networks)注意力模組。 ECA-Net 注意力模組的引入目的在有效提升模型的收斂速度。由於 ECA-Net 屬於輕量級注意力模組，因此可以保持輕量級神經網路運算速度並取得良好的準確率。以下將敘述本研究針對 GSNet 架構中之 ShuffleNetV2模型應用在視覺式智慧型高爾夫揮桿動作姿勢分析系統相關改良。

(1) 群組正規化(group normalization)

卷積神經網路模型在訓練時各個卷積層參數會不斷進行更新，而在淺層卷積層參數調整會對深層卷積層參數具有高度影響力。因此淺層卷積層產生的少許誤差會隨著卷積層深度增大導致誤差顯著增加，而誤差隨著層數越深而越擴大，使得模型正確收斂變得困難。為了解決此問題模型訓練時只能降低學習效率，進而導致收斂速度過慢，此現象稱為內部共變相平移(internal covariate shift)。Ioffe 等人[Iof15]於西元2015年提出批量正規化技術來解決內部共變相平移的問題。

批量正規化技術能夠有效提升卷積神經網路模型的收斂速度，但批量正規化時批量大小參數的設定值必須足夠高才能夠獲得良好的效能。因此 Wu 等人[Wu18]於西元2018年提出群組正規化技術。群組正規化技術不考慮批量大小參數值，是透過對同一個特徵圖的通道維度分組後進行正規化的技術。此方法能夠在批量大小參數設定較小情況下也能夠提升卷積神經網路模型收斂速度以及準確率。

群組正規化技術首先將特徵圖的所有通道分成多個群組，再針對以群組為單位的資料進行正規化，降低群組中資料之間的差異，此技術在實現時與批量大小無關。圖31為群組正規化示意圖，圖31中特徵圖高度為H，寬度為W，通道個數為C，批量大小個數為N。假設將群組分成G組，則每群組內的通道個數為C/G。以圖31為例，通道個數為6個，圖中將3個通道個數為一群組，因此總共有2個群組。系統將會以群組為單位分別對未來2個群組中的資料進行正規化。

假設群組中的資料輸入值為x_i，平均數為μ_i，變異數為σ_i^2，ϵ為防止除數為0所設計的極小常數，計算群組正規化公式如(10)所示。

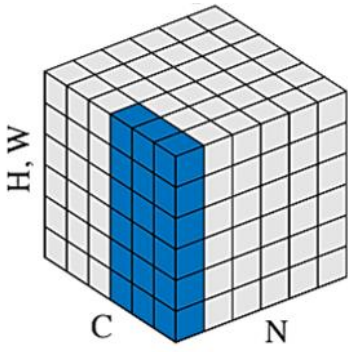


圖31：群組正規化示意圖[Wu18]

$$\hat{x}_i = \frac{x_i - u_i}{\sqrt{\sigma_i^2 + \epsilon}} \tag{10}$$

最後再引入兩個參數，分別是伸縮量(scale)參數為γ以及平移量(shift)參數為β，這兩個參數會和類神經網路模型的其他參數一樣利用自行學習的方式進行調整，因此加入可學習參數之群組正規化公式如(11)所示。

$$y_i = \gamma \hat{x}_i + \beta \tag{11}$$

(2) H-swish 函數

多數類神經網路模型是由一個輸入層、一個輸出層以及多個隱藏層所組合而成。類神經網路在不使用非線性激勵函數的情況下會以線性的方式組合運算僅能處理線性問題。因為隱藏層與輸出層皆是由上層之輸出結果作為輸入，並且是以線性組合做運算，使得上層輸出與下一層的輸入只存在線性關係，不管再深層的網路也只具線性性質，最後都只是輸出線性組合，因此訓練出來的神經網路也只能處理線性問題，極大地限制了這類神經網路的能力。解決的方式就是採用非線性激勵函數。因此在類神經網路中，非線性激勵函數對於類神經網路模型去學習、理解複雜和非線性函數來說具有十分重要的作用。總而言之，使用非線性激勵函數可以讓類神經網路從資料中學習到非線性關係以便處理非線性問題。

Ramachandran 等人[Ram17]於西元2017年提出 swish 函數做為類神經網路中的激勵函數，給定 sigmoid 函數$\sigma(x)$，swish 函數的計算方式如(12)所示。

$$\text{swish}(x) = x \cdot \sigma(\beta x) \tag{12}$$

其中β為可訓練參數。Swish 函數除了具有類似 ReLU 函數的性質外，swish 函數之特色為無上界有下界，其具有非單調性(non-monotonic)及平滑特性，可以提升梯度的表現度(expressivity)，因此 swish 函數在類神經網路模型上的訓練效果優於 ReLU 函數。

但由於 swish 函數計算複雜，會降低系統的執行速度，不適合使用於輕量級網路，因此 Howard 等人[How19]提出之 MobileNetV3採用 h-swish 函數來取代 swish 函數。h-swish 函數公式如(13)所示。

$$\text{h-swish}(x) = x\frac{\text{ReLU6}(x+3)}{6} \tag{13}$$

大致來說，h-swish 函數和 swish 函數具有其相似性，但是 h-swish 函數可讓類神經網路模型在避免使用計算量大的 sigmoid 函數的情形下依然維持一定的準確率。因此本研究將延續 GSNetV1架構將 ShuffleNetV2模型中所使用的批量正規化替換成群組正規化以及 ReLU 函數替換成 h-swish 函數，並稱為 GSNetV2架構。

(3) ECA-Net

前面曾提及 Wang 等人[Wan20]於西元2020年提出 ECA-Net 輕量級注意力模組，可提升卷積神經網路辨識準確率。注意力機制源於對人類視覺之研究，當外在環境訊息不斷經由人類視覺系統湧入腦神經架構中，人類腦部會過濾掉不重要的訊息，只專注在有意義的重點資訊上。注意力機制在自然語言領域中的應用已經獲得相當成功，因此近年來在電腦視覺領域中也引入上述概念。總而言之，使用注意力機制能夠有效提高模型性能。

ECA-Net 架構圖如圖32所示。圖32中，假設輸入特徵圖之寬度為W，高度為H，通道數量為C，先使用全局平均池化將每個通道的二維特徵$W \times H$壓縮成單一個特徵值，壓縮後的特徵圖大小即為$1 \times 1 \times C$，接著使用$1 \times 1 \times k$(圖中的$k = 5$)的卷積核做卷積運算，接著經過 sigmoid 函數得到權重值，每個通道都會生成一個對應的權重值，最後再將權重值加權整合到原輸入的特徵圖中。

使用$1 \times 1 \times k$的卷積核做卷積運算主要是用來擷取區域性的跨通道資訊(local cross-channel interaction)。其中k代表跨通道資訊所涵蓋的範圍，因此在不同卷積神經網路或者在不同通道個數中，參數k設定多寡會影響效果好壞。若要找出k參數的最佳設定值會非常耗費計算資源，因此 Wang 等人[Wan20]建議使用自適應性的方法找尋最適合的參數。而參數k和通道數量C兩者擁有映射關係如公式(14)所示。

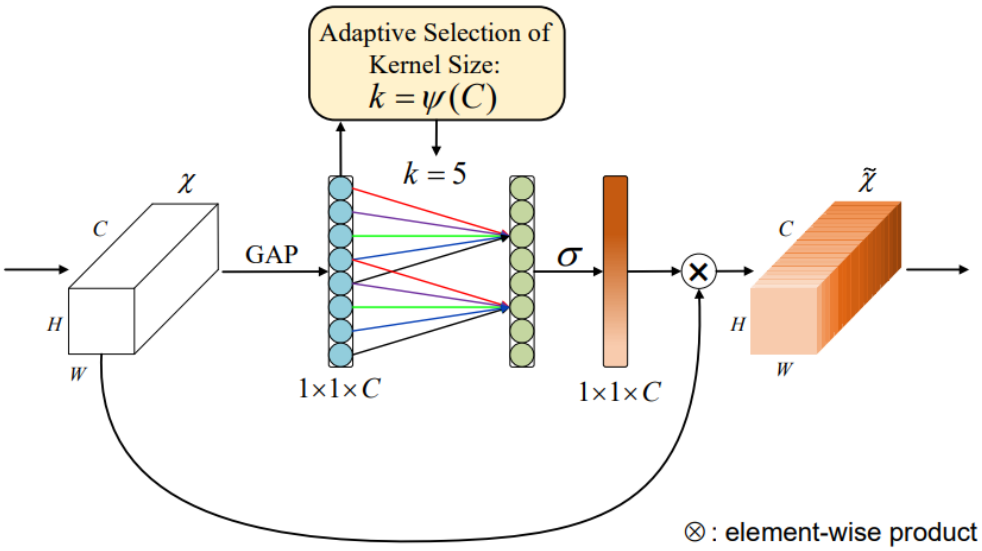


圖32：ECA-Net 架構示意圖[Wan20]

$$C = \emptyset(k) = \gamma * k - b \tag{14}$$

Wang 等人[Wan20]先以最簡單的映射關係設計線性方程式為$\emptyset(k) = (\gamma * k - b)$進行實驗，但在神經網路中通道個數通常會是以2為底的指數做為設計，所以後來將線性方程式延伸為非線性方程式如公式(15)所示。

$$\emptyset(k) = 2^{(\gamma * k - b)} \tag{15}$$

假設已經得到通道數量為C，那麼參數k的計算如公式(16)所示。

$$k = \psi(C) = \left|\frac{\log_2(C)}{\gamma} + \frac{b}{\gamma}\right|_{odd} \tag{16}$$

其中$|x|_{odd}$表示如果x為奇數，則$|x|_{odd} = x$；如果x為偶數，則$|x|_{odd}$的值設定為$x + 1$。另外，本研究實驗時將參數k設定為3、參數γ設定為2及參數b設定為1。本研究將延續 GSNetV2 架構加入 ECA-Net 輕量級注意力模組，並命名為GSNetV3架構。

第三節 關鍵動作幀判定

關鍵動作幀估計的方式是針對高爾夫揮桿影片裡的每一幀都分別預測出八個高爾夫揮桿分解動作的信心程度，同時選擇信心程度最高者做為該分解動作之關鍵動作幀。本研究在關鍵動作幀估計中雖然使用了能夠考慮到上下文關係的Bi-GRU 進行分類，但系統設定一次輸入為64幀，而一部即時(real-time)高爾夫揮桿影片拍攝時間為3至10秒鐘不等，約為90至300幀，因此系統無法一次同時處理所有輸入幀進行關鍵動作幀估計，容易造成關鍵動作幀估計結果預測錯誤。

因此本研究研發出關鍵動作幀判定技術，應用高斯分佈調整技術將前述關鍵動作幀估計預測結果再次進行校正。本研究是利用高爾夫揮桿時的時間連續性，即高爾夫八個分解動作幀數會依序由小到大排序的特性，來進行關鍵動作幀校正。

本研究統計 GolfDB 資料集[Mcn19]中的高爾夫揮桿影片其八個分解動作個別關鍵動作幀所在位置分佈，採用高斯分佈去模擬其相鄰倆倆關鍵動作幀間之幀差分佈，建立關鍵動作幀位置分佈模型。假設資料集中第 t 個高爾夫揮桿影片的八個關鍵動作幀以$x_i^t, i = 1,2,\dots,8$表示，而給定i、j兩關鍵動作幀間之幀數差其平均數為μ_{ij}，標準差為σ_{ij}，$i,j = 1,2,\dots,8$，則高斯分佈公式如下:

$$f_{ij}(x) = \frac{1}{\sigma_{ij}\sqrt{2\pi}} e^{-\frac{(x-\mu_{ij})^2}{2\sigma_{ij}^2}} \tag{17}$$

利用上述的方式針對某一特定分解動作，它和其它七個分解動作之間的幀數差可以建構七個高斯分佈模型，同時其自身也會有一個高斯分佈，因此一個分解動作會建構出八個高斯分佈模型。由於全部的分解動作有八個，所以系統中總共建構六十四個高斯分佈模型。

值得一提的是在估計高斯分佈模型時，拍攝影片時所設定的幀率可能有所不同，進而導致用不同幀率拍攝的影片，其關鍵動作幀間之幀差會有所極大的差別，因此真實影片和慢動作影片兩類需分別計算。

如上所述，每個分解動作會擁有八個高斯分佈模型，接著以八個高斯分佈模型作為基礎，利用上一步驟關鍵動作幀估計輸出的結果對影像序列中的每一幀分別預測出八個分解動作的信心程度。舉例而言，針對某一特定分解動作的信心程度分別和八個關鍵動作幀預估該特定分解動作之高斯分佈模型中該幀對應位置的值兩者一一相乘，其相乘後結果再做相加，最終可獲得每一幀加權後信心程度。由於全部的分解動作有八個，因此上述方式會重複進行八次，最後再使用加權後信心程度選擇信心程度最高者做為八個高爾夫揮桿分解動作之關鍵動作幀。

第四節 三維人體模型姿勢比對

前述曾提及本研究所設定之拍攝角度為高爾夫正面揮桿如圖1所示。系統首先將使用者和教練兩者揮桿動作關鍵動作幀三維人體模型利用兩者的三維人體骨架中的0號關節點進行位置對齊，以及將教練和使用者的三維人體模型之體型調整至相同體型，接著考慮到每位拍攝者的拍攝角度會有所差異，進而導致比對產生誤差，因此本研究在比對之前會先將使用者和教練兩者的三維人體模型修正為同一視角。

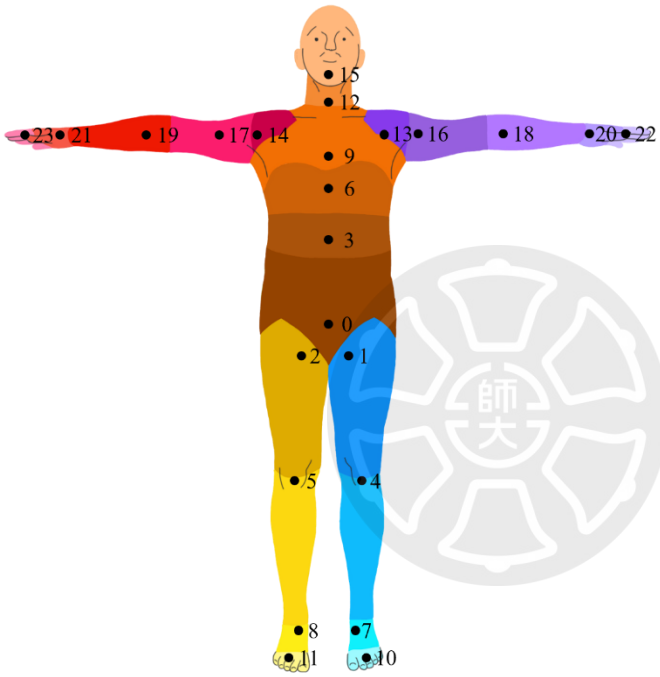
本研究所使用的三維人體模型是由6890個節點所組成的人體網格，其關節點位置與身體部位節點個數如圖33所示。本研究在比對時需計算出使用者和教練兩者全身6890個對應節點在三維空間中的歐式距離。給定三維空間中的二點$p_1(x_1, y_1, z_1)$和$p_2(x_2, y_2, z_2)$，其歐式距離公式如(18)所示。

$$d_{p_1p_2} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \tag{18}$$

系統將教練三維人體模型以 Y 軸為基準每間隔1°進行±10°範圍內的旋轉，利用上述公式計算出每一角度的教練與使用者兩者之間的三維人體模型6890個對應節點的歐式距離總合，最後挑選出歐式距離總合最短之角度所對應的三維人體模型，當作該教練比對時之三維人體模型，此方法可以有效解決使用者與教練拍攝角度不同之問題。

將使用者和教練兩者三維人體模型旋轉成相同視角解決拍攝角度差異問題之後，系統接著將進行使用者和教練兩者的姿勢比對。圖33中，三維人體模型在

建構時將人體分解成24個身體部位，其各身體部位節點個數分佈如圖33右邊所示。由圖33可以觀察到各個身體部位擁有的節點數量並不相等，因此在比對時需先將每個身體部位的節點數量進行正規化，正規化後將使用者和教練兩者的三維人體模型各個身體部位分別計算對應節點間之歐式距離總合，最後顯示出對應節點間之歐式距離總合最高的三個身體部位進行文字反饋，同時將兩者模型重疊進行可視化圖片反饋。



	No.	Part	No. of vertices
	0	Hips	486
	1	Left Up Leg	253
	2	Right Up Leg	253
	3	Spine	232
	4	Left Leg	216
	5	Right Leg	216
	6	Spine1	266
	7	Left Foot	142
	8	Right Foot	142
	9	Spine2	614
	10	Left Toe Base	121
	11	Right Toe Base	121
	12	Neck	155
	13	Left Shoulder	150
	14	Right Shoulder	150
	15	Head	1193
	16	Left Arm	283
	17	Right Arm	283
	18	Left Fore Arm	245
	19	Right Fore Arm	245
	20	Left Hand	323
	21	Right Hand	323
	22	Left Hand Index	477
	23	Right Hand Index	477

圖33：三維人體模型關節點位置與身體部位節點個數示意圖

總體而言，本研究在第一大步驟高爾夫揮桿分解動作擷取中，使用輕量級網路 ShuffleNetV2和 Bi-GRU 進行關鍵動作幀估計，將 ShuffleNetV2模型所使用的批量正規化技術以及 ReLU 函數分別替換成群組正規化技術和 h-swish 函數來提升準確率。同時加入 ECA-Net 輕量級注意力模組以期有效加速模型的收斂速度，最後本研究依照高爾夫揮桿的時間特性研發出關鍵動作幀判定之校正方法。接著在第二大步驟三維人體模型姿勢比對分析中，先將使用者和教練兩者之三維人體模型利用0號關節點進行位置對齊，以及將教練和使用者的三維人體模型之體型調整至相同體型，接著將教練三維人體模型旋轉成與使用者之三維人體模型具相

同視角來解決拍攝角度差異問題，再對使用者和教練兩者之三維人體模型進行差異比對，最終以文字和圖片兩種反饋方式呈現。

第 4 章 實驗結果與討論

本章將進行實驗結果與討論，總共將劃分成七節。本章第一節將介紹使用資料庫與研究設備；第二節探討採用幀差法之改良成果；第三節探討使用群組正規化及 h-swish 函數對於 GSNet 架構準確率的影響；第四節探討引入 ECA-Net 輕量級注意力模組對於 GSNet 架構收斂速度的影響；第五節探討使用關鍵動作幀判定對於將 GSNet 架構的預測結果再次進行修正的改良成果；第六節探討使用三維人體模型進行姿勢比對的成果；最後，第七節探討本研究在各項模型之改良分析與討論。

第一節 資料庫介紹與研究設備

本研究使用 McNally 等人[Mcn19]所提出的 GolfDB 資料集進行訓練及測試，GolfDB 資料集是針對高爾夫運動之高爾夫揮桿影片資料集，高爾夫揮桿影片主要來自 PGA、LPGA 以及 Champions Tours 等職業賽事中的職業高爾夫球選手所組成，該資料集總共含有1400部由不同角度拍攝之高爾夫揮桿影片，以及該資料集影片之拍攝速度區分為即時(real-time)及慢動作(slow-motion)，1400部高爾夫揮桿影片當中含有758部為即時速度之高爾夫揮桿影片以及642部慢動作速度之高爾夫揮桿影片。每部高爾夫揮桿影片解析度為720以及影片總長度超過390k幀數。

本研究中將 GolfDB 資料集其中的1050部影片用於類神經網路之訓練當作訓練集，相當於資料庫占比70%，另外將350部用於類神經網路之測試當作測試集，相當於資料庫占比30%。而本研究實驗使用的硬體規格為 i5-9400F 處理器與 RTX 2080Ti 顯示卡。其中本研究訓練 GSNet 模型架構中可調整的模型訓練參數為學習率(learning rate)、批量大小(batch size)以及 epoch。學習率為控制模型中梯度下降的速度，本研究預設學習率為0.001；批量大小是指將訓練集中一次輸入至類神經網路中的影片數量，本研究預設批量大小為一次輸入6部影片；epoch 是指在訓練過程中將訓練集中的所有影片都訓練過一次，本研究預設 epoch 為100次。而在本實驗中，每種改良之測試數據為五次測試的平均。

第二節 幀差法分析

本節主要分析在本系統中的第一大步驟高爾夫揮桿分解動作擷取將輸入影片採用幀差法之準確率。首先介紹本研究第一大步驟高爾夫揮桿分解動作擷取的準確率評估方式，本研究遵循 McNally 等人[Mcn19]所提出的 PCE(percentage of correct events) 準確率計算方式。一般而言，影片可以使用不同幀率(frame rate)進行拍攝，例如30fps、60fps 以及120fps 等拍攝幀率，30fps 為最常使用的拍攝幀率。而一個完整高爾夫揮桿平均時間為1至3秒鐘，當使用30fps 拍攝高爾夫揮桿影片時，影像序列可能無法完整定義八個高爾夫揮桿關鍵動作幀，高爾夫揮桿分解動作的關鍵動作幀可能落在兩幀影像之間，因此準確率的計算公式需要具備容錯能力。給定n為第一個高爾夫揮桿動作擊球準備幀數，f為採樣幀率，McNally 等人[Mcn19]所定義準確率計算式中的容錯率δ如公式(19)所示。

$$\delta = \max\left(\left\lfloor \frac{n}{f} + \frac{1}{2} \right\rfloor, 1\right) \tag{19}$$

其中$\lfloor \cdot \rfloor$表示向下取整數。本研究中採樣幀率f設定為30。

本節後續將分析輸入影片採用幀差法之改良成果。為方便文中說明，本研究將 ShuffleNetV2和 Bi-GRU 原型架構命名為 GSNet(golf swing net)。使用 GSNet 進行高爾夫揮桿分解動作擷取測試結果如表4所示，從表4中得知 GSNet 模型架構測試結果平均準確率為74.45%。表4中可以觀察到第一個分解動作擊球準備、第四個分解動作上桿頂點以及第八個分解動作收桿這三個分解動作其準確率低於80%，尤其在第一個分解動作擊球準備以及第八個分解動作收桿兩者準確率偏低，分別只有29.31%以及20.85%，因此將特別討論這三個分解動作。

第一個分解動作擊球準備其定義為啟動揮桿動作前一刻，因此還在揮桿前的準備動作時，其具有相當多幀為揮桿前準備動作，系統較難判斷出高爾夫球運動者要啟動揮桿動作前一刻。圖34為第一個分解動作擊球準備 RGB 影像序列與幀差影像序列之測試結果，圖34(a)至(d)為擊球準備 RGB 由第k幀至第$k+3$幀之影像序列，圖34(e)至(h)為擊球準備由第k幀至第$k+3$幀之幀差影像序列。由圖中可以觀察到圖34(a)至(d)很難判斷出高爾夫球運動者要啟動揮桿動作前一刻，而圖

34(e)至(h)為將四張影像採用幀差法計算的結果，可以看出連續二幀影像間的高爾夫球桿之運動輪廓，因此幀差影像序列較易判斷出高爾夫球運動者要啟動揮桿動作前一刻。

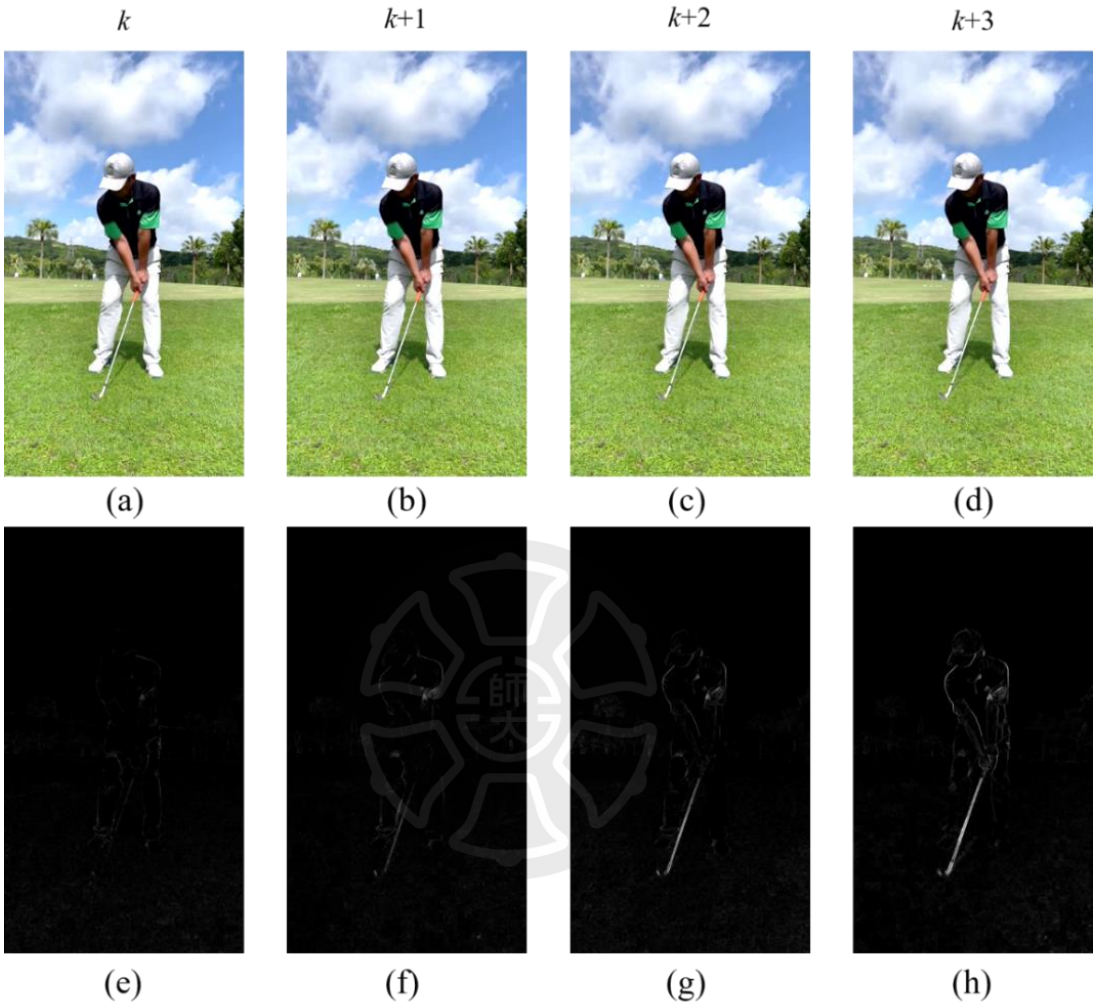


圖34：擊球準備動作 RGB 影像序列與幀差影像序列之測試結果示意圖
(a)(b)(c)(d)為 RGB 影像序列(e)(f)(g)(h)為幀差影像序列

第四個分解動作上桿頂點其定義為從上桿到下桿過程中改變方向的那一瞬間，主要是觀察高爾夫球桿揮擊方向改變的那一剎那，因此使用 RGB 影像序列會較難分辨出上桿頂點動作。圖35為第四個分解動作上桿頂點 RGB 影像序列與幀差影像序列之測試結果，圖35 (a)至(d)為 RGB 上桿頂點動作由第k幀至第$k+3$幀之影像序列，圖35(e)至(h)為上桿頂點動作由第k幀至第$k+3$幀之幀差影像序列。幀差影像序列可以看出連續二幀影像間的高爾夫球桿方向改變的那一瞬間，因此幀差影像序列較易判斷出上桿到下桿過程中高爾夫球桿改變方向的那一瞬間。

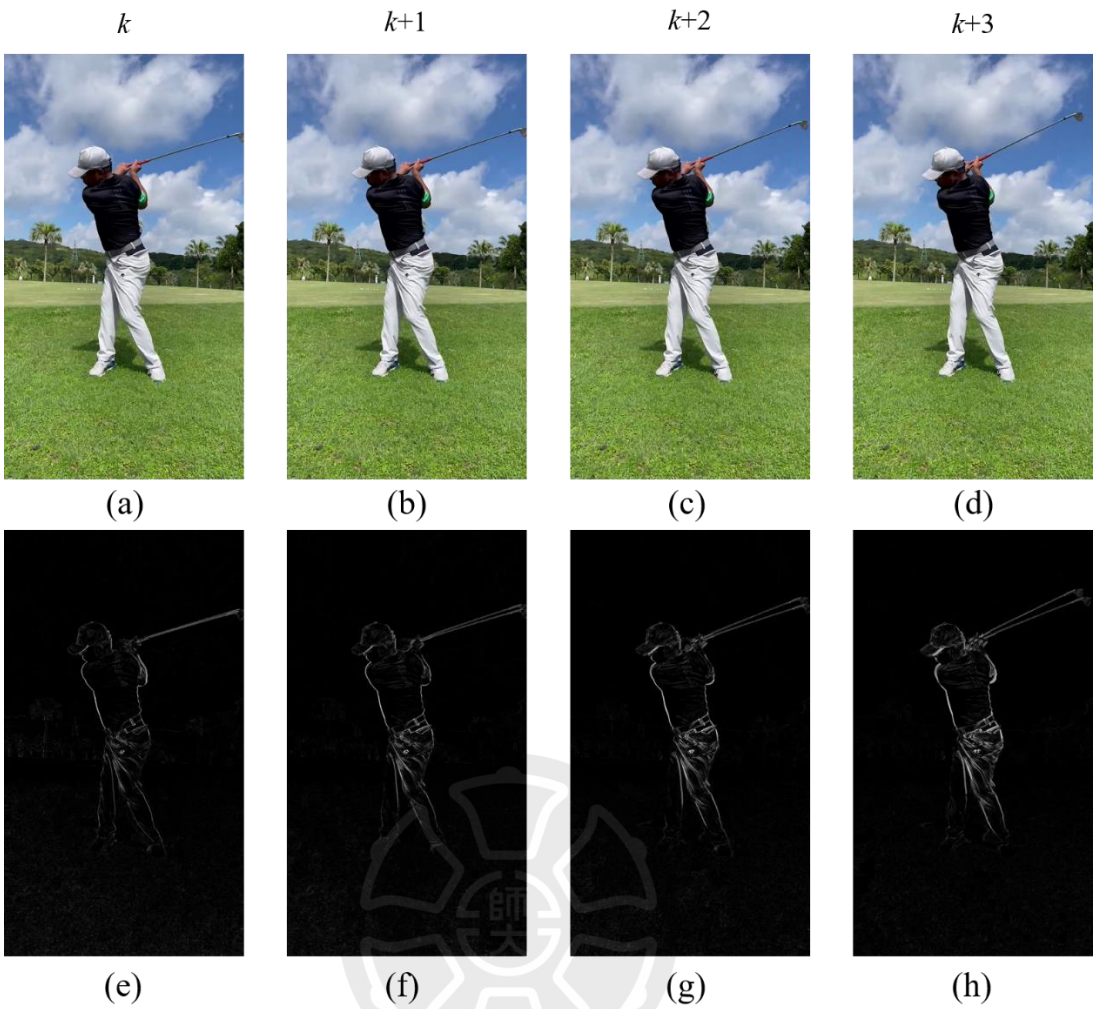


圖35：上桿頂點動作 RGB 影像序列與幀差影像序列之測試結果示意圖
(a)(b)(c)(d)為 RGB 影像序列(e)(f)(g)(h)為幀差影像序列

第八個分解動作收桿其定義為揮桿結束動作身體肌肉放鬆前一刻，主要每位高爾夫球運動者之收桿動作可能不盡相同，因此使用 RGB 影像序列會較難分辨出收桿動作。圖36為第八個分解動作收桿 RGB 影像序列與幀差影像序列之測試結果，圖36 (a)至(d)為 RGB 收桿動作由第k幀至第$k+3$幀之影像序列，圖36(e)至(h)為收桿動作由第k幀至第$k+3$幀之幀差影像序列。幀差影像序列可以看出連續二幀影像間的身體以及高爾夫球桿兩者之差異，可以觀察出身體以及高爾夫球桿是否有移動產生，因此幀差影像序列較易判斷出高爾夫球運動者之收桿動作。

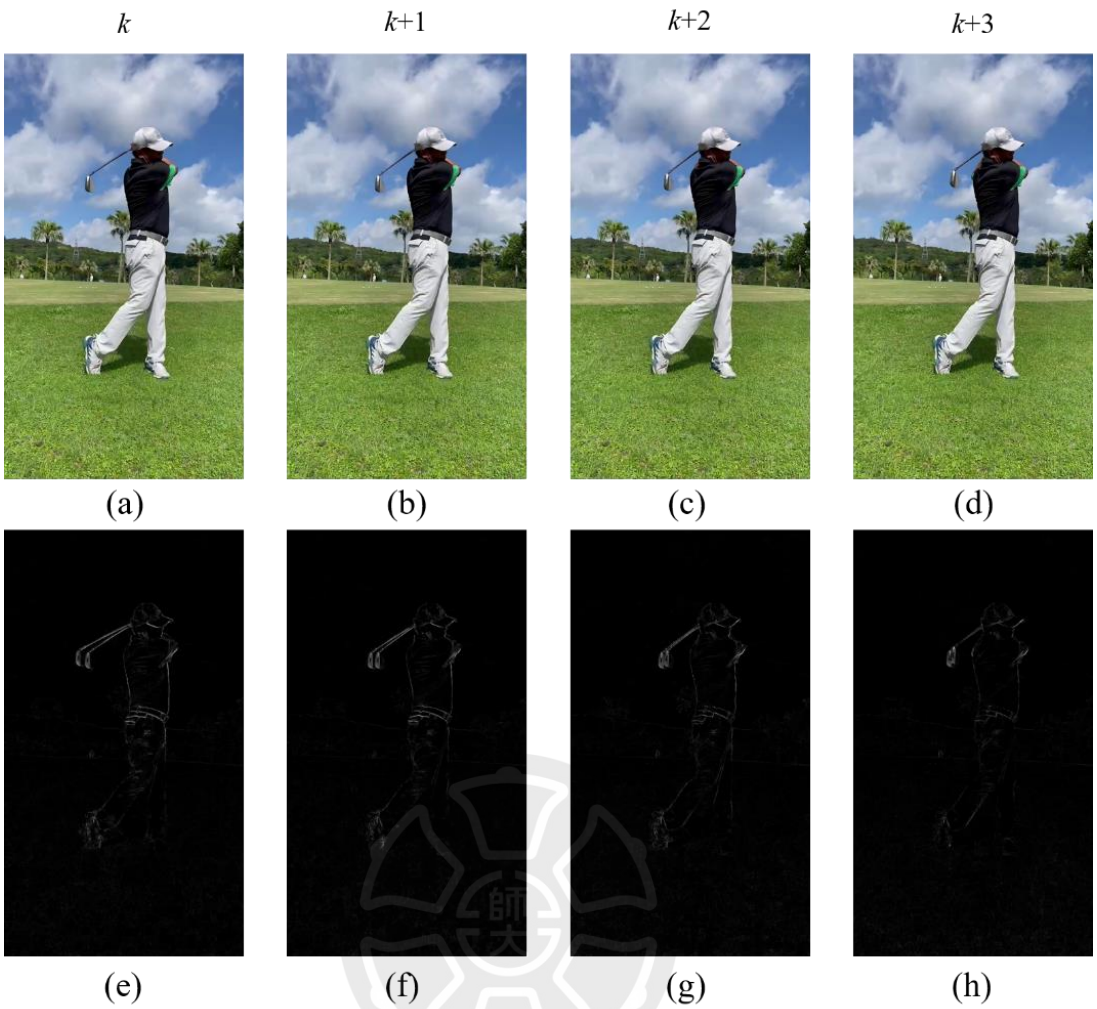


圖36：收桿動作 RGB 影像序列與幀差影像序列之測試結果示意圖(a)(b)(c)(d)為 RGB 影像序列(e)(f)(g)(h)為幀差影像序列

表5：GSNetV1架構測試結果

編號	分解動作名稱	真實影片 準確率	慢動作影片 準確率	總體準確率
1	擊球準備	58.15%	56.16%	57.14%
2	起桿	94.33%	92.09%	93.20%
3	上桿	91.90%	84.07%	87.94%
4	上桿頂點	93.52%	89.15%	91.31%
5	下桿	97.45%	98.53%	98.00%
6	擊球	98.84%	99.21%	99.03%
7	送桿	98.15%	98.53%	98.34%
8	收桿	42.89%	40.90%	41.88%
	平均準確率	**84.40%**	**82.33%**	**83.35%**

綜上所述，幀差影像序列更易判斷第一個分解動作擊球準備、第四個分解動作上桿頂點以及第八個分解動作收桿，而當 GSNet 架構之輸入影片為幀差影像序列(本研究將此架構命名為 GSNetV1)時，其總體平均準確率為83.35%，如表5所示。表5為 GSNetV1架構之測試結果，表中顯示八個分解動作中的起桿、上桿頂點、下桿、擊球以及送桿等五個動作的總體準確率已經都達到90%以上，而且第六個分解動作擊球的總體準確率更高達99%，具相當高的準確性。

表6：GSNet 架構與 GSNetV1架構各分解動作準確率比較表

編號	分解動作名稱	GSNet 準確率	GSNetV1 準確率	準確率提升 幅度
1	擊球準備	29.31%	57.14%	27.83%
2	起桿	85.88%	93.20%	7.32%
3	上桿	86.28%	87.94%	1.66%
4	上桿頂點	79.31%	91.31%	12.00%
5	下桿	98.00%	98.00%	0.00%
6	擊球	98.91%	99.03%	0.12%
7	送桿	97.08%	98.34%	1.26%
8	收桿	20.85%	41.88%	21.03%
	總體平均準確率	**74.45%**	**83.35%**	**8.90%**

表6為 GSNet 架構與 GSNetV1架構各分解動作準確率比較表。由表6中可以觀察出採用幀差法的情況下八個分解動作的準確率都有提升，特別是第一個分解動作擊球準備從29.31%提升至57.14%，以及第八個分解動作收桿從20.85%提升至41.88%，兩個分解動作之準確率皆有大幅度的提升。第四個分解動作上桿頂點的準確率也從79.31%提升至91.31%，獲得12.00%的成長。最後總體平均準確率從 GSNet 架構的74.45%提升至 GSNetV1架構的83.35%，總體平均準確率上升8.90%，有顯著的進步。

第三節 群組正規化及 h-swish 函數分析

本節將分析本研究 GSNet 架構中 ShuffleNetV2模型所改良的群組正規化技術和 h-swish 函數對準確率的影響。本研究延續 GSNetV1架構將 ShuffleNetV2模型中所使用的批量正規化替換成群組正規化以及 ReLU 函數替換成 h-swish 函數，

並且命名為 GSNetV2。圖37為改良模型 GSNetV2之架構圖，圖37(a)為改良後的 basic unit 以及圖37(b)為改良後的 downsampling unit。

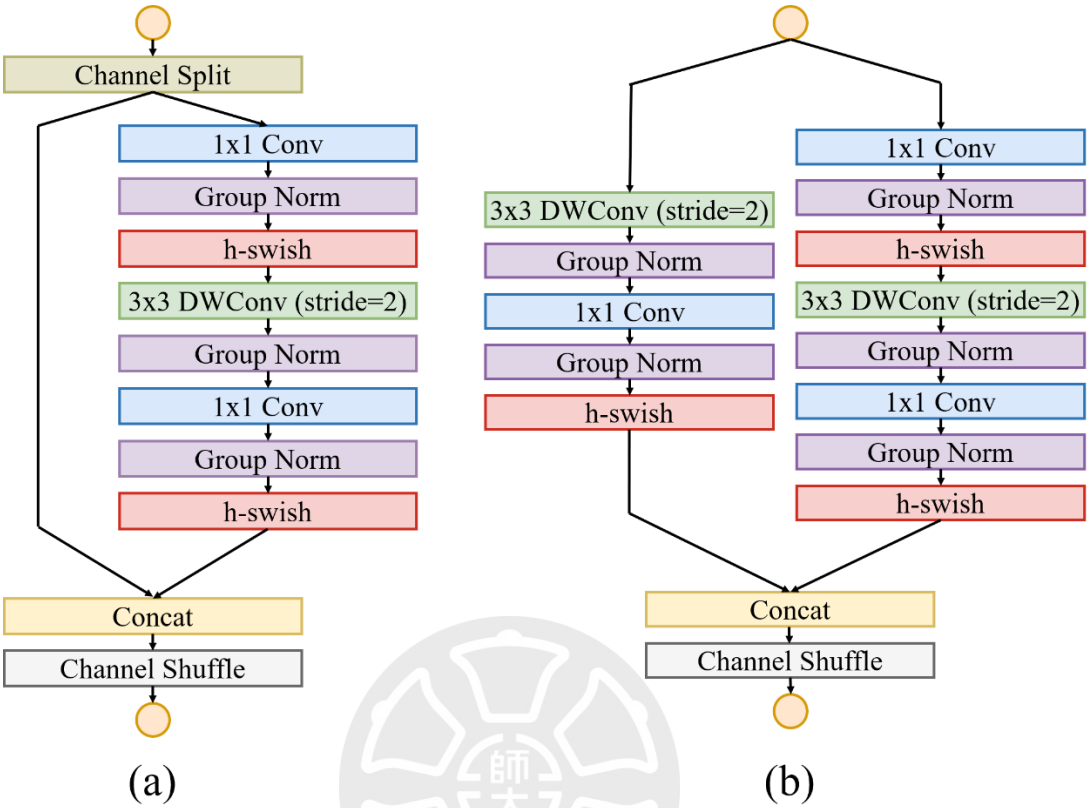


圖37：GSNetV2模型架構圖(a)basic unit(b)downsampling unit

表7：GSNetV2架構測試結果

編號	分解動作名稱	真實影片準確率	慢動作影片準確率	總體準確率
1	擊球準備	61.27%	60.56%	60.91%
2	起桿	95.72%	94.23%	94.97%
3	上桿	91.67%	85.65%	88.62%
4	上桿頂點	94.91%	90.73%	92.79%
5	下桿	97.34%	98.42%	97.88%
6	擊球	99.88%	99.44%	99.65%
7	送桿	99.42%	99.09%	99.25%
8	收桿	44.04%	49.60%	46.85%
	平均準確率	**85.53%**	**84.71%**	**85.12%**

表7所示為 GSNetV2架構之測試結果，其總體平均準確率為85.12%。表中顯示在八個分解動作中有五個分解動作總體準確率達到90%，而且其中有三個分解

動作的總體準確率超過95%，並且擊球和收桿兩個分解動作的總體準確率高達99%。

表8：GSNetV1架構與 GSNetV2架構各分解動作準確率比較表

編號	分解動作名稱	GSNetV1	GSNetV2	準確率提升幅度
1	擊球準備	57.14%	60.91%	3.77%
2	起桿	93.20%	94.97%	1.77%
3	上桿	87.94%	88.62%	0.68%
4	上桿頂點	91.31%	92.79%	1.48%
5	下桿	98.00%	97.88%	-0.12%
6	擊球	99.03%	99.65%	0.62%
7	送桿	98.34%	99.25%	0.91%
8	收桿	41.88%	46.85%	4.97%
	總體平均準確率	**83.35%**	**85.12%**	**1.77%**

表8為 GSNetV1架構與 GSNetV2架構各分解動作準確率比較表。由表8可以觀察出將 ShuffleNetV2模型改良後八個分解動作中大部分的準確率皆有提升，特別是第一個分解動作擊球準備從57.14%提升至60.91%，以及第八個分解動作收桿從41.88%提升至46.85%，兩個分解動作之準確率有明顯的提升。最後總體平均準確率從 GSNetV1架構的83.35%提升至 GSNetV2架構的85.12%，總體平均準確率上升1.77%，因此本研究群組正規化技術和 h-swish 函數之改良有效提升準確率。

第四節 ECA-Net 分析

本節將分析在本系統中將 GSNet 架構中之 ShuffleNetV2模型加入 ECA-Net(efficient channel attention for deep convolutional neural networks)輕量級注意力模組對模型收斂速度以及準確率的影響。本研究將延續 GSNetV2架構之 ShuffleNetV2改良模型加入 ECA-Net 輕量級注意力模組，並且命名為 GSNetV3。要注意的是本研究只有在 ShuffleNetV2模型中的 basic unit 加入 ECA-Net 輕量級注意力模組。圖38為 GSNetV2架構與 GSNetV3架構中 ShuffleNetV2改良模型之 basic unit 架構圖，其中圖38(a)為 GSNetV2架構中 ShuffleNetV2改良模型之 basic

unit 架構圖；圖38(b)為 GSNetV3架構中 ShuffleNetV2改良模型加入 ECA-Net 輕量級注意力模組(紅色粗框)之 basic unit 架構圖。

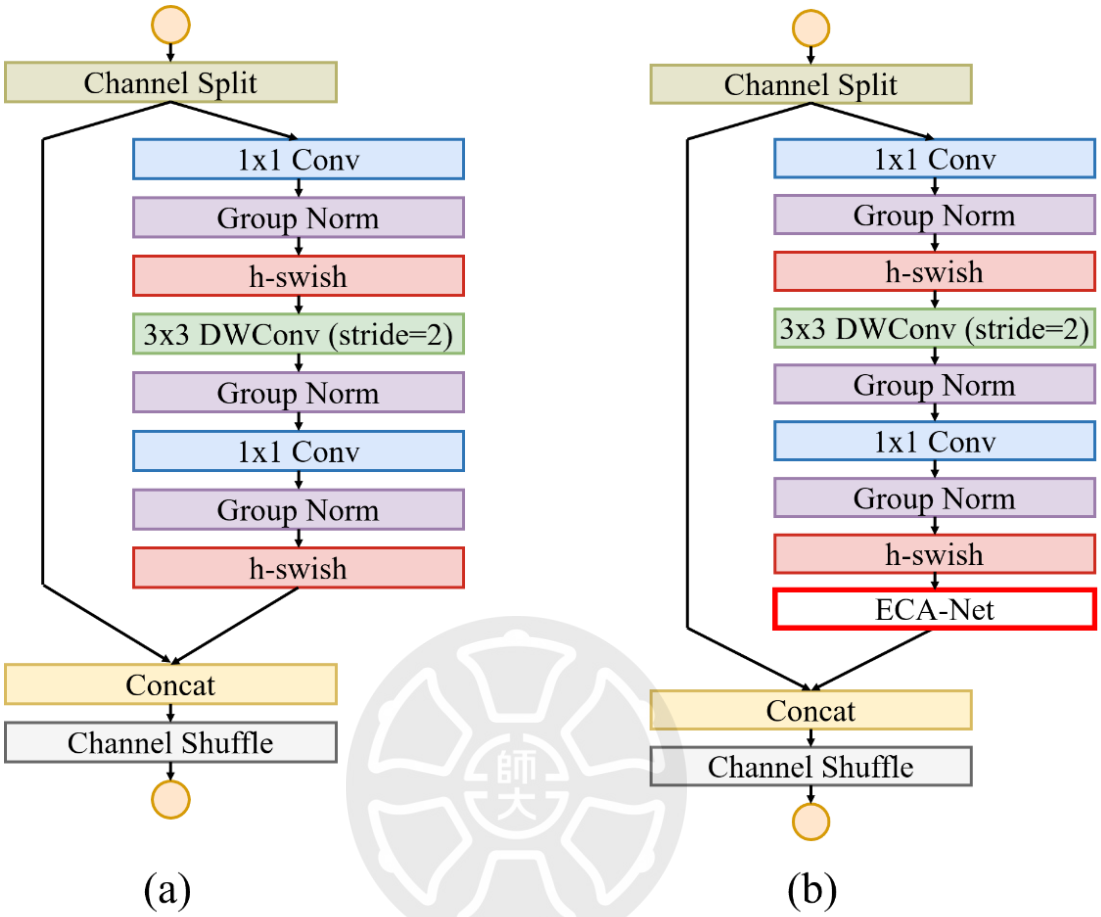


圖38：GSNetV2架構與 GSNetV3架構之 basic unit 架構圖(a)GSNetV2架構之 basic unit 架構圖(b)GSNetV3架構之 basic unit 架構圖

圖39為 GSNetV2架構與 GSNetV3架構之損失函數值折線圖。圖39中紅色折線為 GSNetV2架構損失函數值折線，藍色折線則為加入 ECA-Net 輕量級注意力模組之 GSNetV3架構損失函數值折線。GSNetV2架構和 GSNetV3架構兩者總共都訓練100個 epoch 並繪制其損失函數值折線。從圖中可以看出加入 ECA-Net 輕量級注意力模組之 GSNetV3架構的藍色折線比 GSNetV2架構的紅色折線之損失函數值下降更快。同時可以觀察到當 GSNetV3架構的藍色折線損失函數下降至0.2時在第20個 epoch，但 GSNetV2架構的紅色折線損失函數下降至0.2時是在第80個 epoch，因此 GSNetV3架構的藍色折線下降速率比 GSNetV2架構的紅色折線快上四倍。

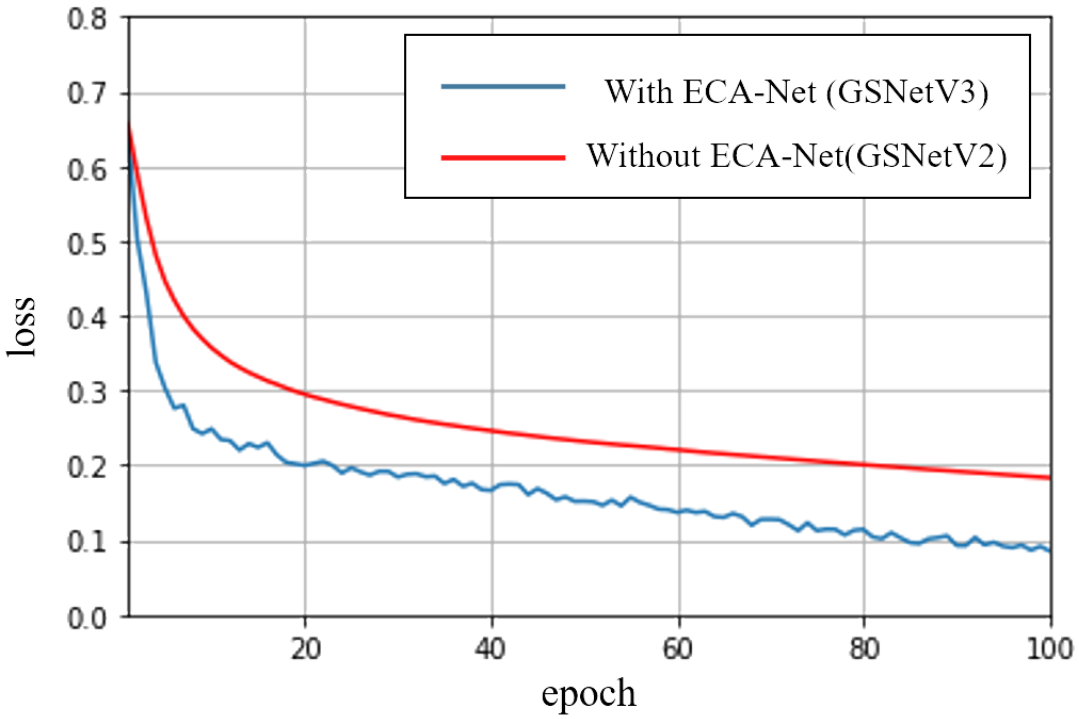


圖39：GSNetV2架構與 GSNetV3架構損失函數值折線圖

在輕量級網路中模型的參數量也是相當重要的衡量指標，因此在改良模型時亦必須要適時地限制參數量。表9顯示為沒有加入 ECA-Net 的 GSNetV2架構以及有加入 ECA-Net 的 GSNetV3架構總參數量比較表。從表中可以看出 GSNetV3只增加了48個參數量，因此 ECA-Net 在本系統的引入具合適性。

表9：GSNetV2架構與 GSNetV3架構總參數量比較表

	GSNetV2	GSNetV3	總參數量差異
總參數量	3,227,373	3,227,421	48

GSNetV3架構將在延續 GSNetV2架構改良後將 ShuffleNetV2模型之 basic unit 架構中加入 ECA-Net 輕量級注意力模組時，其總體平均準確率為85.39%，如表10所示。表10為 GSNetV3架構之測試結果，表中顯示在第八個分解動作收桿的總體準確率達到了51.02%，同時在八個分解動作中有五個分解動作總體準確率仍然達到90%，並且擊球和收桿兩個分解動作的總體準確率高達99%。

表10：GSNetV3架構測試結果

編號	分解動作名稱	真實影片 準確率	慢動作影片 準確率	總體準確率
1	擊球準備	61.73%	60.79%	61.25%
2	起桿	95.37%	91.97%	93.65%
3	上桿	91.44%	83.84%	87.60%
4	上桿頂點	94.79%	92.43%	93.60%
5	下桿	97.45%	97.97%	97.71%
6	擊球	99.42%	99.21%	99.31%
7	送桿	98.95%	99.09%	99.02%
8	收桿	49.94%	52.09%	51.02%
	平均準確率	**86.14%**	**84.67%**	**85.39%**

表11：GSNetV2架構與 GSNetV3架構各分解動作準確率比較表

編號	分解動作名稱	GSNetV2	GSNetV3	準確率提升 幅度
1	擊球準備	60.91%	61.25%	0.34%
2	起桿	94.97%	93.65%	-1.32%
3	上桿	88.62%	87.60%	-1.02%
4	上桿頂點	92.79%	93.60%	0.81%
5	下桿	97.88%	97.71%	-0.17%
6	擊球	99.65%	99.31%	-0.34%
7	送桿	99.25%	99.02%	-0.23%
8	收桿	46.85%	51.02%	4.17%
	總體平均準確率	**85.12%**	**85.39%**	**0.27%**

表11為 GSNetV2架構與 GSNetV3架構總體準確率比較表。表11中，GSNetV2架構與 GSNetV3架構分別訓練40個 epoch 所得到的總體準確率，可以觀察出在第八個分解動作收桿從46.85%提升至51.02%，獲得4.17%的成長，而其他分解動作準確率下降比例則可以計在誤差範圍之內。最後總體平均準確率從 GSNetV2架構的85.12%提升至 GSNetV3架構的85.39%。總體平均準確率上升0.27%，並且GSNetV3架構之收斂速度和 GSNetV2架構相比加快4倍，因此本研究加入 ECA-Net 輕量級注意力模組有效加速模型收斂並有效提升準確率。

第五節 關鍵動作幀判定分析

本節將探討使用本研究所研發之關鍵動作幀判定對於將 GSNet 架構的預測結果再次進行修正的改良成果。圖40為 GSNetV3架構之高爾夫揮桿分解動作擷取預測結果圖以及基準真相(ground truth)圖，其中圖40 (a)為八個分解動作的基準真相圖；圖40(b)為八個分解動作的預測結果圖。圖中的箭頭表示為時間軸，而每個分解動作上方的數字代表此動作的幀數。

本研究將八個分解動作使用不同顏色框所表示。第一個分解動作擊球準備為紅色表示；第二個分解動作起桿為黃色表示；第三個分解動作上桿為綠色表示；第四個分解動作上桿頂點為青色表示；第五個分解動作下桿為藍色表示；第六個分解動作擊球為紫色表示；第七個分解動作送桿為灰色表示；第八個分解動作收桿為黑色表示。

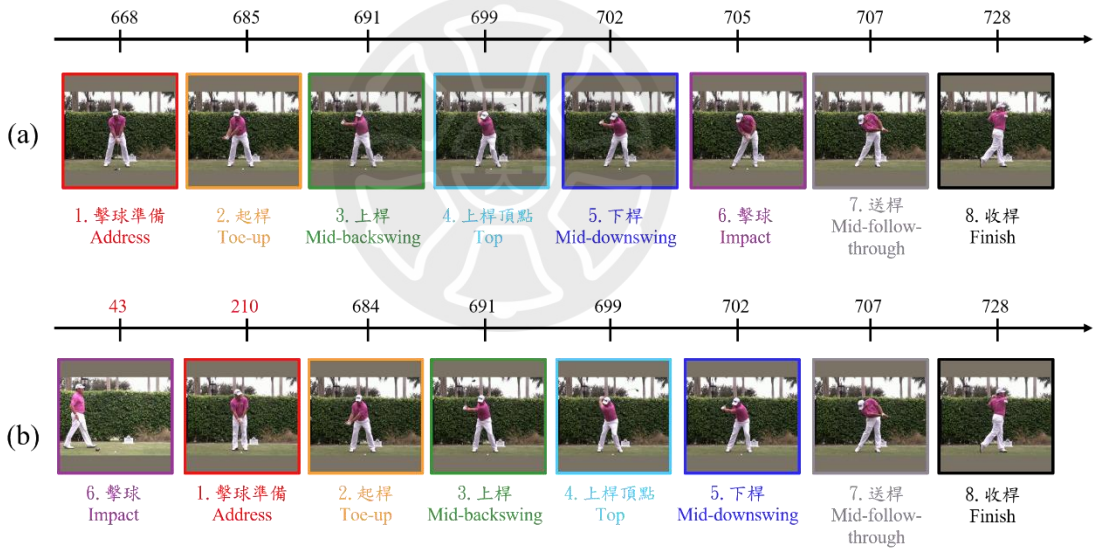


圖40：GSNetV3之高爾夫揮桿分解動作擷取之基準真相以及預測結果(a)八個分解動作的基準真相(b)GSNetV3擷取八個分解動作的預測結果

由高爾夫揮桿具有的時間特性可以看到圖40(a)由第一個分解動作擊球準備到第八個分解動作收桿的幀數為遞增排序，同時可觀察到圖40(b)從第一個分解動作擊球準備到第八個分解動作收桿並沒有遞增排序，因此可以推斷出圖40(b)有分解動作預測錯誤。圖40(b)幀數標為紅字(43和210)代表 GSNetV3架構預測八個分解動作幀數錯誤，它們分別是第一個分解動作擊球準備和第六個分解動作擊球。第一個分解動作擊球準備其定義為啟動揮桿動作前一刻，而第六個分解動作

擊球其定義為擊到球的瞬間，由圖40(b)中可見系統所預測出的第一個分解動作擊球準備和第六個分解動作擊球並不符合高爾夫揮桿的定義。

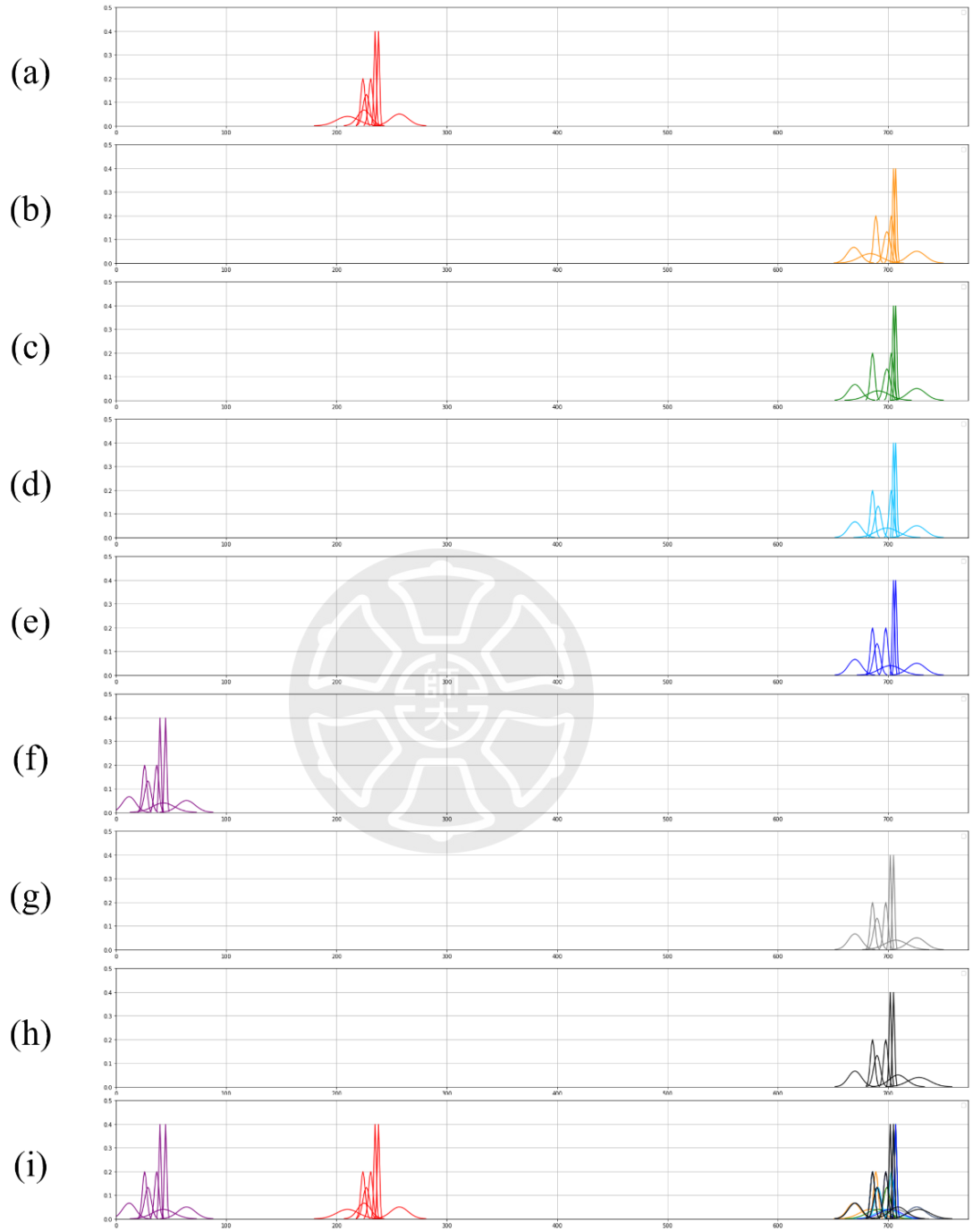


圖41：關鍵動作幀判定技術實驗結果(a)第一個動作擊球準備所對應的八個高斯分佈(b)第二個動作起桿所對應的八個高斯分佈(c)第三個動作上桿所對應的八個高斯分佈(d)第四個動作上桿頂點所對應的八個高斯分佈(e)第五個動作下桿所對應的八個高斯分佈(f)第六個動作擊球所對應的八個高斯分佈(g)第七個動作送桿所對應的八個高斯分佈(h)第八個動作收桿所對應的八個高斯分佈(i)八個分解動作所對應的高斯分佈合併示意圖

接著系統使用如圖40(b)GSNetV3架構之高爾夫揮桿分解動作擷取之關鍵動作幀來預測應對應的八個分解動作的幀數。這八個關鍵動作幀所分別對應的八個分解動作的幀數所建構出的八個高斯分佈模型如圖41所示。圖41(a)為第一個分解動作擊球準備所建構出對應八個分解動作的高斯分佈模型示意圖；同理，圖41(b)至(h)則為第二個分解動作至第八個分解動作所建構出對應八個分解動作的高斯分佈模型示意圖。而圖41(i)為八個關鍵動作幀所分別對應的八個分解動作的幀數其高斯分佈模型的合併示意圖。

由圖41(i)可以觀察到為紫色代表的第六個分解動作擊球與紅色代表的第一個分解動作擊球準備分別所建構出的八個高斯分佈模型的分佈位置與其他預測正確動作的分佈位置大不相同，而預測為正確的分解動作個別所建構出的八個高斯分佈模型的分佈為大致相同。因此利用此性質可以將預測結果再次進行校正，接著將 GSNetV3預測結果使用關鍵動作幀判定再次校正結果如圖42所示。首先觀察圖42之八個分解動作的幀數已校正為由小到大排序，且和圖40(a)八個分解動作的基準真相圖兩者相比的結果大致相同，因此關鍵動作幀判定具有將預測結果校正的效果。

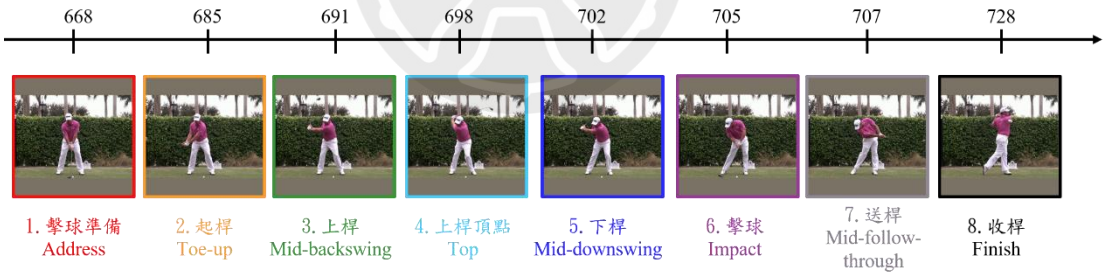


圖42：關鍵動作幀判定校正實驗結果

表12為 GSNet 架構、GSNetV3架構及 GSNetV3架構使用關鍵動作幀判定之測試影片由小到大正確排序之數量比較表。測試影片為350部高爾夫揮桿影片，包含173部高爾夫揮桿真實影片以及177部高爾夫揮桿慢動作影片。首先測試未改良之 GSNet 架構所預測出的八個分解動作幀數為由小到大排序的預測結果全部影片為339部影片，而測試 GSNetV3架構所預測出的八個分解動作幀數為由小到大排序的預測結果全部影片為344部影片。因為影片的預測結果幀數並未全部皆由小到大排序，也意味著至少6部影片有預測錯誤的情形發生。最後測試GSNetV3架構使用關鍵動作幀判定的八個分解動作幀數為由小到大排序之校正結果為350

部影片，代表所有預測結果幀數全部皆為由小到大正確排序，符合高爾夫揮桿動作具有的時間次序性。

表12：GSNet 架構、GSNetV3架構及 GSNetV3架構使用關鍵動作幀判定之測試影片由小到大排序比較表

	GSNet	GSNetV3	GSNetV3+KD
真實影片正確排序數(173)	168	170	173
慢動作影片正確排序數(177)	170	174	177
全部影片正確排序數(350)	339	344	350

關鍵動作幀判定是在 GSNetV3架構的預測結果之後所使用，其總體平均準確率為86.15%，如表13所示。表13為 GSNetV3架構使用關鍵動作幀判定之測試結果，表中顯示真實影片以及慢動作影片的兩者平均準確率都提升至85%以上，以及實驗結果顯示在八個分解動作中有五個分解動作總體準確率達到90%，而且其中有四個分解動作的總體準確率95%，並且擊球和收桿兩個分解動作的總體準確率高達99%。

表13：GSNetV3架構使用關鍵動作幀判定測試結果

編號	分解動作名稱	真實影片 準確率	慢動作影片 準確率	總體準確率
1	擊球準備	61.38%	62.26%	61.82%
2	起桿	95.95%	92.32%	94.11%
3	上桿	91.44%	86.32%	88.85%
4	上桿頂點	95.72%	94.46%	95.08%
5	下桿	97.80%	97.96%	97.88%
6	擊球	99.65%	99.32%	99.48%
7	送桿	99.07%	99.32%	99.20%
8	收桿	50.63%	54.91%	52.80%
	平均準確率	**86.45%**	**85.86%**	**86.15%**

表14：GSNetV3架構與 GSNetV3架構+KD 總體準確率比較表

編號	分解動作名稱	GSNetV3 準確率	GSNetV3+KD 準確率	準確率提升 幅度
1	擊球準備	61.25%	61.82%	0.57%
2	起桿	93.65%	94.11%	0.46%
3	上桿	87.60%	88.85%	1.25%
4	上桿頂點	93.60%	95.08%	1.48%
5	下桿	97.71%	97.88%	0.17%
6	擊球	99.31%	99.48%	0.17%
7	送桿	99.02%	99.20%	0.18%
8	收桿	51.02%	52.80%	1.78%
	總體平均準確率	**85.39%**	**86.15%**	**0.76%**

表14為 GSNetV3架構與 GSNetV3架構使用關鍵動作幀判定總體準確率比較表，表14中將關鍵動作幀判定縮寫為 KD。由表14中可以觀察出在第四個分解動作上桿頂點從93.60%提升至95.08%，獲得1.48%的成長。最後總體平均準確率從GSNetV3架構的85.39%提升至 GSNetV3架構使用關鍵動作幀判定的86.15%，總體平均準確率上升0.76%。同時實驗結果顯示利用高爾夫揮桿具有的時間特性，350部測試影片中的八個分解動作預測幀數都已校正成由小到大的正確排序。

第六節 三維人體模型姿勢比對分析

本節將探討使用三維人體模型進行姿勢比對的結果。首先，圖43為教練與使用者利用第一大步驟高爾夫揮桿分解動作擷取所獲得的高爾夫揮桿八個分解動作的實驗結果。圖43(a)為教練之高爾夫揮桿八個分解動作實驗結果；圖43(b)為使用者之高爾夫揮桿八個分解動作實驗結果。如前一節所述將八個分解動作使用不同顏色框表示。

而教練與使用者的高爾夫揮桿八個分解動作找出的對應二維人體骨架則如圖44所示。圖44(a)與圖44(b)分別為教練與使用者高爾夫揮桿八個分解動作之對應二維人體骨架。系統接著使用圖43的高爾夫揮桿八個分解動作影像以及圖44的八個分解動作對應二維人體骨架資訊來估計出每個分解動作的三維人體模型，其結果如圖45所示。圖45(a)為教練的高爾夫揮桿八個分解動作之三維人體模型建構結果，接下來的實驗中教練的三維人體模型將用藍色顯示。另外，圖45(b)為使用

者的高爾夫揮桿八個分解動作之三維人體模型建構結果。同理，接下來的實驗中使用者的三維人體模型則是用膚色顯示。

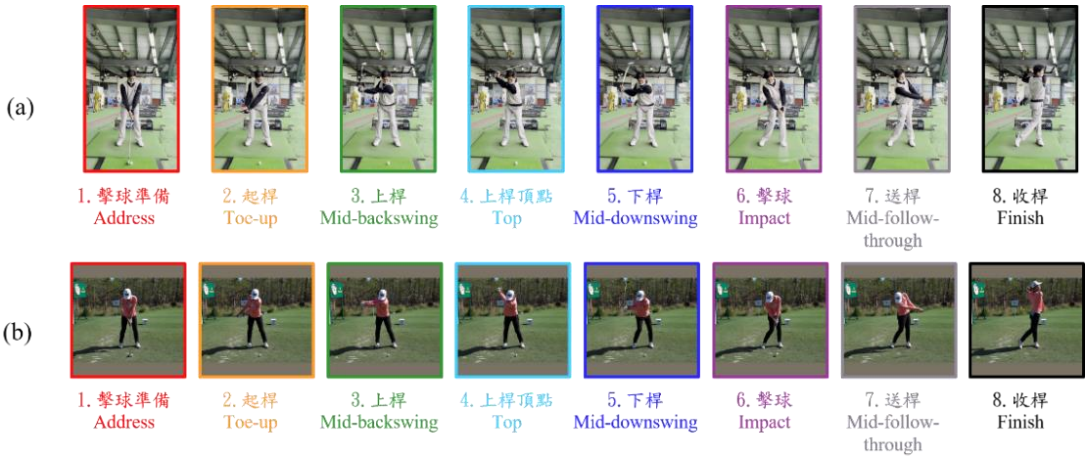


圖43：教練與使用者之高爾夫揮桿八個分解動作實驗結果(a)教練之高爾夫揮桿八個分解動作實驗結果(b)使用者之高爾夫揮桿八個分解動作實驗結果

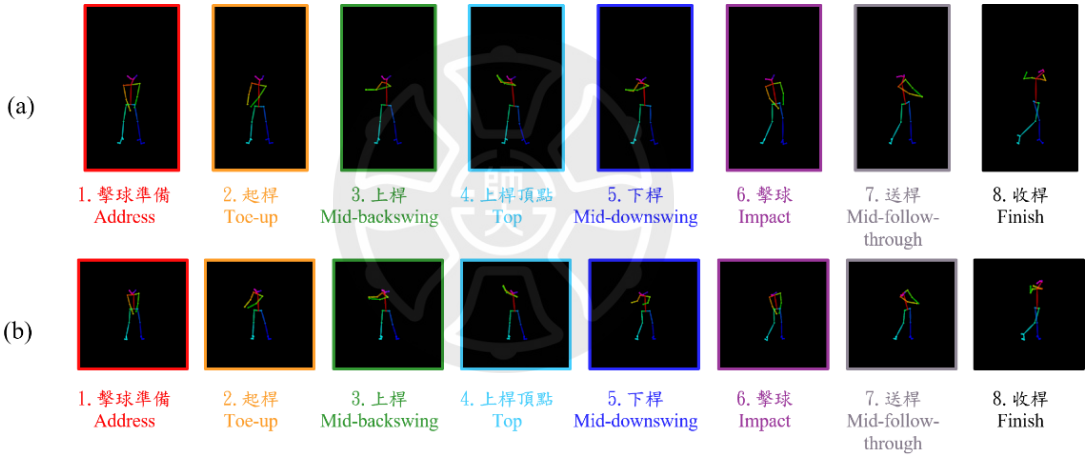


圖44：教練與使用者高爾夫揮桿八個分解動作之二維人體骨架實驗結果(a)教練高爾夫揮桿八個分解動作之二維人體骨架實驗結果(b)使用者高爾夫揮桿八個分解動作之二維人體骨架實驗結果

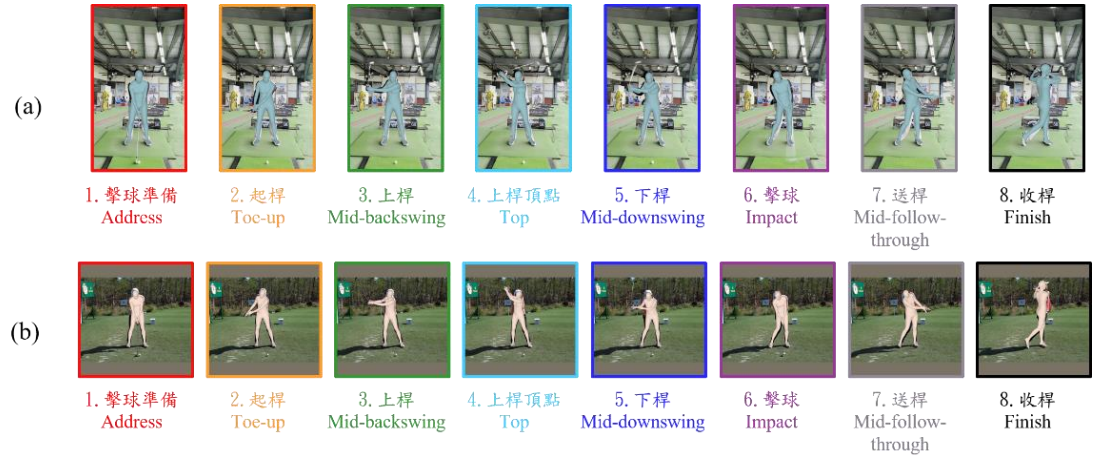


圖45：教練與使用者高爾夫揮桿八個分解動作之三維人體模型實驗結果(a)教練

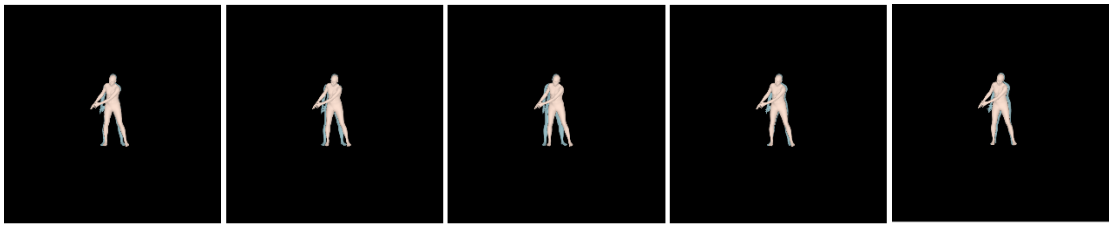
高爾夫揮桿八個分解動作之三維人體模型實驗結果(b)使用者高爾夫揮桿八個分解動作之三維人體模型實驗結果

在三維人體模型比對時，首先會將使用者和教練兩者的三維人體模型依第3章第四節所述之方式調整體型大小以及位置對齊如圖46所示，圖46(a)為使用者之三維人體模型原始體型大小實驗結果；圖46(b)為教練之三維人體模型原始體型大小實驗結果；圖46(c)為教練之三維人體模型調整至使用者體型大小實驗結果。調整體型大小完畢後再進行位置對齊，圖46(d)為使用者與教練之三維人體模型位置對齊實驗結果，兩者之三維人體模型對齊重疊後，將教練的三維人體模型擺放至後方，而使用者的三維人體模型會擺放至前方，可以讓使用者容易觀察到自身動作，以及重疊可以更容易觀察出使用者和教練兩者的揮桿姿勢差異進行可視化圖片反饋。

接著考慮到每位拍攝者的拍攝角度會有所差異，因此在比對之前會先將使用者和教練兩者的三維人體模型修正為同一視角如圖47所示。圖47(a)為使用者與教練之三維人體模型原始角度對齊實驗結果；圖47(b)為將教練之三維人體模型旋轉5度對齊實驗結果；圖47(c)為教練之三維人體模型旋轉10度對齊實驗結果；圖47(d)為教練之三維人體模型旋轉-5度對齊實驗結果；圖47(e)為教練之三維人體模型旋轉-10度對齊實驗結果。接著依第3章第四節所述之方式挑選出歐式距離總合最短之角度所對應的三維人體模型，當作該教練比對時之三維人體模型。

(a) (b) (c) (d)

圖46：使用者與教練上桿動作之三維人體模型體型大小及位置對齊實驗結果(a)使用者之三維人體模型原始體型大小實驗結果(b)教練之三維人體模型原始體型大小實驗結果(c)教練之三維人體模型調整至使用者體型大小實驗結果(d)使用者與教練之三維人體模型位置對齊實驗結果



(a) (b) (c) (d) (e)

圖47：使用者與教練之三維人體模型角度對齊實驗結果(a)使用者與教練之三維人體模型原始角度對齊實驗結果(b)教練之三維人體模型旋轉5度對齊實驗結果(c)教練之三維人體模型旋轉10度對齊實驗結果(d)教練之三維人體模型旋轉-5度對齊實驗結果(e)教練之三維人體模型旋轉-10度對齊實驗結果

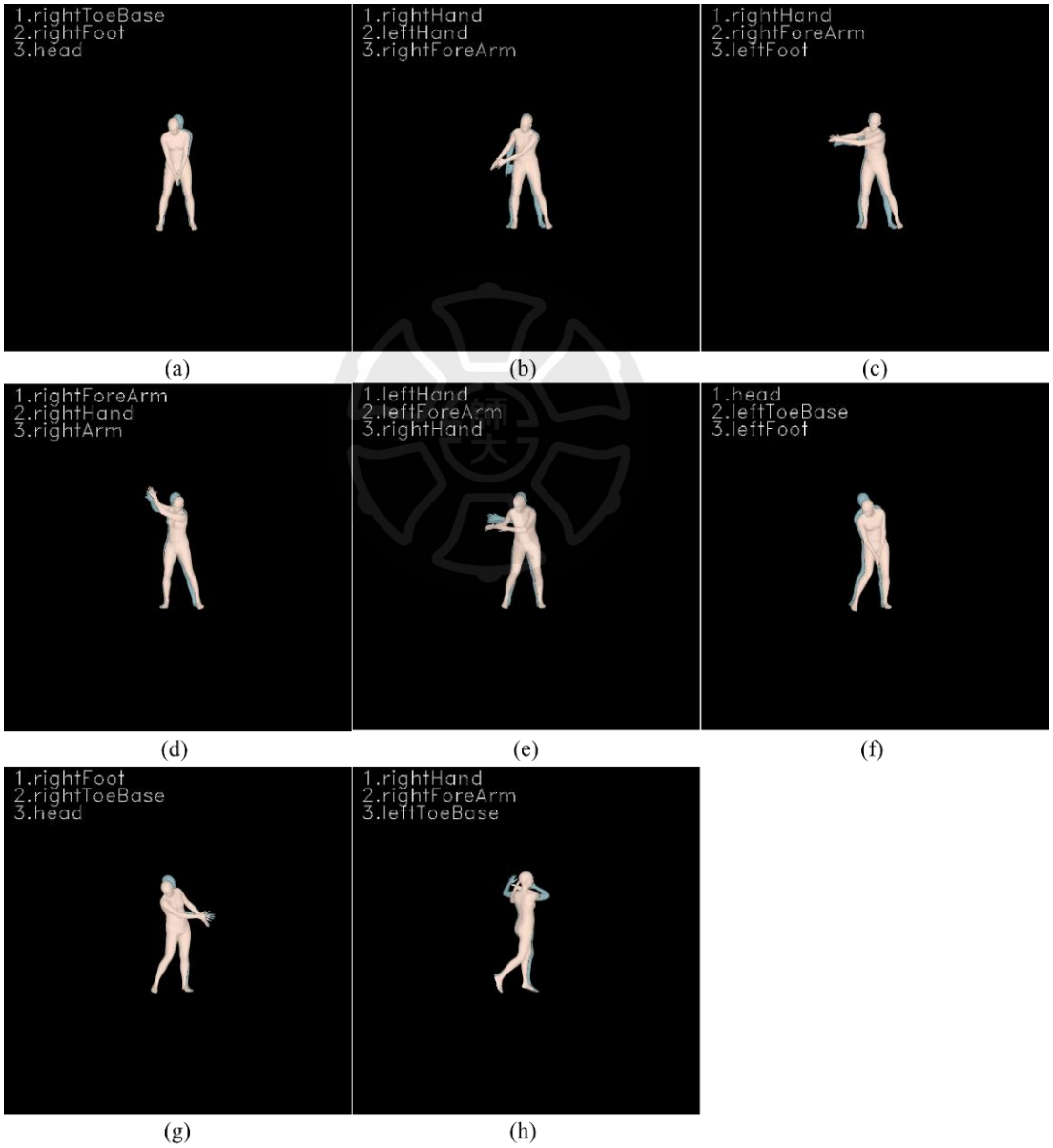


圖48：使用者和教練高爾夫揮桿八個分解動作之三維人體模型比對實驗結果(a)擊球準備比對結果(b)起桿比對結果(c)上桿比對結果(d)上桿頂點比對結果(e)下桿比對結果(f)擊球比對結果(g)送桿比對結果(h)收桿比對結果

最後再將使用者與教練之三維人體模型進行比對，輸出為可視化圖片，反饋中的圖片左上角會列出使用者和教練兩者姿勢差距最大的三個身體部位進行文字反饋，身體部位名稱的對應為圖33所示。

圖48為將使用者和教練兩者對應分解動作之三維人體模型比對結果。圖48(a)為第一個分解動作擊球準備比對實驗結果；圖48(b)為第二個分解動作起桿比對實驗結果；圖48(c)為第三個分解動作上桿比對實驗結果；圖48(d)為第四個分解動作上桿頂點比對實驗結果；圖48(e)為第五個分解動作下桿比對實驗結果；圖48(f)為第六個分解動作擊球比對實驗結果；圖48(g)為第七個分解動作送桿比對實驗結果；最後，圖48(h)為第八個分解動作收桿比對實驗結果。

三維人體模型不僅可以表現出豐富的人體資訊，並且可以將三維人體模型任意旋轉，以此來觀察了解使用者的揮桿姿勢。圖49所示第一個分解動作擊球準備四種旋轉角度實驗範例。其中圖49(a)為原始正面角度實驗結果；圖49(b)為三維人體模型旋轉90度實驗結果；圖49(c)為三維人體模型旋轉180度實驗結果；而圖49(d)則為旋轉270度之實驗結果。而且三維人體模型可以完整旋轉360度，因此三維人體模型的建構可以使用者或教練觀察出更多角度以及更細緻的揮桿姿勢。

圖49：第一個分解動作擊球準備四種旋轉角度實驗範例(a)原始正面角度實驗結果(b)旋轉90度實驗結果(c)旋轉180度實驗結果(d)旋轉270度實驗結果

第七節 各項改良分析與討論

本研究中第一大步驟高爾夫揮桿分解動作擷取中是由 ShuffleNetV2和 Bi-GRU 所組成的原型架構 GSNet。之後本研究陸續提出之改良系統分別為GSNetV1、GSNetV2、GSNetV3以及 GSNetV3加上使用關鍵動作幀判定技術(GSNetV3+KD)。因此本節將探討並比較上述各種改良模型架構的準確率、訓練時間以及執行時間。

表15顯示 GSNet、GSNetV1、GSNetV2、GSNetV3以及 GSNetV3+KD 各分解動作準確率比較表，表15中將關鍵動作幀判定技術縮寫為 KD。由表15可以觀察出每次進行改良後總體平均準確率都有所提升，特別是第一個分解動作擊球準備從 GSNet 架構的29.31%提升至 GSNetV3使用關鍵動作幀判定的61.82%，大幅成長32.51%；而第八個分解動作收桿從 GSNet 架構的20.85%提升至 GSNetV3使用關鍵動作幀判定的52.80%，也大幅成長31.95%。另外，第二個分解動作起桿從 GSNet 架構的85.88%提升至 GSNetV3使用關鍵動作幀判定的94.11%，提升8.23%；而第四個分解動作上桿頂點從 GSNet 架構的79.31%提升至 GSNetV3使用關鍵動作幀判定的95.08%，提升15.77%。如上所述，說明本研究將 GSNet 架構中輸入影片為幀差影像序列、引入群組正規化技術、h-swish 激活函數以及 ECA-Net 輕量級注意力模組等改良皆有效提升高爾夫揮桿分解動作擷取準確率。

表15：GSNet、GSNetV1、GSNetV2、GSNetV3及 GSNetV3+KD 各分解動作準確率比較表

編號	分解動作名稱	GSNet 準確率	GSNetV1 準確率	GSNetV2 準確率	GSNetV3 準確率	GSNetV3 +KD 準確率
1	擊球準備	29.31%	57.14%	60.91%	61.25%	61.82%
2	起桿	85.88%	93.20%	94.97%	93.65%	94.11%
3	上桿	86.28%	87.94%	88.62%	87.60%	88.85%
4	上桿頂點	79.31%	91.31%	92.79%	93.60%	95.08%
5	下桿	98.00%	98.00%	97.88%	97.71%	97.88%
6	擊球	98.91%	99.03%	99.65%	99.31%	99.48%
7	送桿	97.08%	98.34%	99.25%	99.02%	99.20%
8	收桿	20.85%	41.88%	46.85%	51.02%	52.80%
	總體平均準確率	**74.45%**	**83.35%**	**85.12%**	**85.39%**	**86.15%**

表16為 GSNet、GSNetV1、GSNetV2與 GSNetV3各個架構訓練模型所需時間以及執行架構所需時間的比較表。表中訓練時間計算為完整訓練100個 epoch，可以觀察到每種改良的訓練時間大致相同，顯示本研究所進行的改良並不會增加過多參數或增加浮點數運算次數導致訓練時間拉長。接著，由表中可觀察 GSNetV1 架構的執行時間和 GSNet 架構的執行時間相比之下增加約3秒鐘。導致此現象的原因是因為將影片採用幀差法計算需要時間處理，但後續進行的改良 GSNetV2 架構以及 GSNetV3架構與 GSNetV1架構的執行時間相比之下則大致相同。由於

本系統並不屬於即時系統，因此本研究改良後所增加的執行時間屬於可接受的範圍。

表16：GSNet、GSNetV1、GSNetV2與 GSNetV3訓練時間及執行時間比較表

	GSNet	GSNetV1	GSNetV2	GSNetV3
訓練時間	1.40小時	1.40小時	1.45小時	1.51小時
執行時間	1.13秒	4.28秒	4.31秒	4.31秒

接著討論在本研究中高爾夫揮桿分解動作擷取與三維人體模型姿勢比對分析的實驗結果，並且對實驗結果進行討論。圖50為高爾夫揮桿影片中第六個分解動作擊球連續影像範例以及該擊球動作定義示意圖。首先看到圖50(a)為使用30fps 所進行拍攝的高爾夫揮桿影片第六個分解動作擊球附近之連續幀數，而圖50(b)為高爾夫揮桿第六個分解動作擊球定義示意圖，依前述第六個分解動作擊球其定義為擊到球的瞬間。本範例顯示若選擇使用較低的 fps 進行拍攝，因高爾夫揮桿速度過快可能無法擷取到高爾夫揮桿分解動作的正確影像，即擊到球的瞬間之影像。而以圖50(a)所示的範例中，系統只能選擇第 *t*+2幀當作是第六個分解動作擊球，這情況會使得第 *t*+2幀與圖50(b)的擊球姿勢有所差距，進而導致比對誤差。

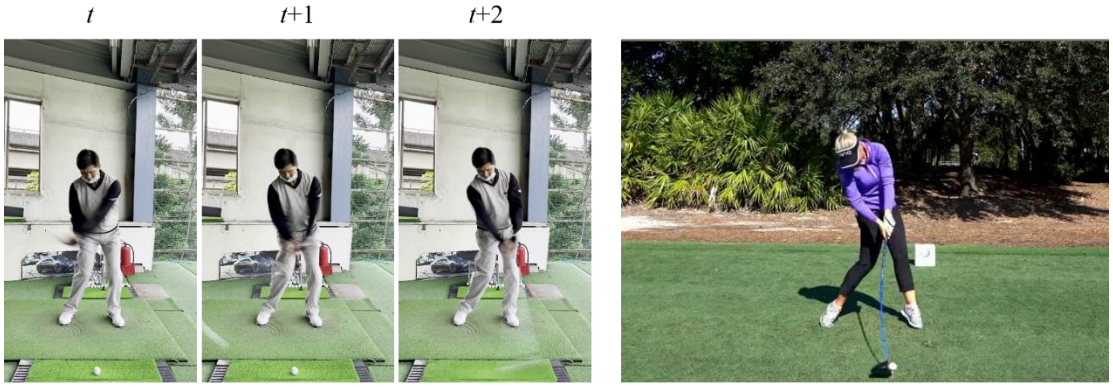


圖50：第六個分解動作擊球連續影像範例(a)第六個分解動作擊球連續影像範例(b)第六個分解動作擊球定義示意圖

圖51則是三維人體模型預測錯誤的實驗範例。圖51(a)為要預測三維人體模型的高爾夫揮桿原始影像；圖51(b)為高爾夫揮桿預測三維人體模型影像；圖51(c)為

三維人體模型預測結果。從圖51(b)與圖51(c)可以觀察出系統預測的三維人體模型之左手穿過自身身體的不正常現象，此情況發生的原因可能是身體四肢太靠近身體或者穿著長袖衣物使得四肢和身體顏色太過相近造成的預測錯誤，進而導致比對誤差。本研究期望未來能夠對上述問題加以改進。

(a) (b) (c)

圖51：三維人體模型預測錯誤實驗範例(a)高爾夫揮桿影像(b)高爾夫揮桿預測三維人體模型影像(c)三維人體模型預測結果

第五章 結論與未來工作

第一節 結論

近年來運動科技興起，將運動與科技兩者相互結合，利用智慧化訓練能夠有效幫助運動員提升訓練品質以及降低運動傷害發生。本研究以高爾夫運動為基礎，為避免高爾夫揮桿姿勢錯誤導致運動傷害，開發出視覺式智慧型高爾夫揮桿動作姿勢分析系統。本系統開發目的主要為使用者在自主練習時無法看到自身揮桿姿勢，若透過本系統讓使用者隨時隨地將自身和教練兩者的高爾夫揮桿姿勢相互比較，可達到自行修正高爾夫揮桿姿勢之目的。

本系統將使用者及教練兩者的高爾夫揮桿影片作為輸入，系統主要分為兩大步驟：高爾夫揮桿分解動作擷取以及三維人體模型姿勢比對分析。第一大步驟高爾夫揮桿分解動作擷取所使用的架構為輕量級網路 ShuffleNetV2以及 Bi-GRU 原型架構組成的 GSNet。將影片採用幀差法計算後作為系統輸入以突顯動態前景物輪廓(GSNetV1)，接著引入群組正規化技術、h-swish 激活函數(GSNetV2)以及 ECA-Net 輕量級注意力模組(GSNetV3)來改良 GSNet 架構。最後使用本研究研發之校正技術將預測之關鍵動作幀再次進行校正，最終擷取出使用者以及教練兩者的高爾夫揮桿八個分解動作。

第二大步驟為三維人體模型姿勢比對分析，首先系統分別預估出使用者以及教練高爾夫揮桿八個分解動作的關鍵動作幀之二維人體骨架。接著利用八個關鍵動作幀和它們對應的二維人體骨架估計出八個三維人體模型，再把使用者以及教練揮桿分解動作所對應的三維人體模型一一進行對齊比對，分析出每一個揮桿分解動作之差異。最後系統輸出分解動作比對後差異最多的三個身體部位名稱以及兩者三維人體模型重疊之反饋圖片。

本研究實驗結果顯示，GSNetV1架構採用幀差法計算後的影像做為輸入，準確率從原先 GSNet 架構的74.45%提升至83.35%。而 GSNetV2架構再採用群組正規化技術以及 h-swish 激活函數後，其準確率則上升至85.12%。接著 GSNetV3架構中再引入 ECA-Net 輕量級注意力模組後，其收斂速度是 GSNetV2架構的四倍同時準確率提高至85.39%。最後將 GSNetV3架構的預測結果使用本研究研發之

關鍵動作幀判定進行校正，使其準確率再提升至86.15%。總而言之，本系統在GolfDB 資料集上進行高爾夫揮桿分解動作擷取最終達到86.15%的準確率。另外，本研究後半部採用三維人體模型進行姿勢比對分析，該三維人體模型是由6,890個節點組成24個身體部位的人體網格，可以呈現豐富的人體資訊。實驗結果顯示利用該模型之特性能夠更精準地判斷使用者及教練之高爾夫揮桿姿勢差異。

第二節 未來工作

本研究所提之視覺式智慧型高爾夫揮桿動作姿勢分析系統依然有可優化之處。第一個可優化的項目為分析各分解動作的特性，持續提升高爾夫揮桿分解動作擷取之準確率。本研究之改良模型在第一個分解動作擊球準備以及第八個分解動作收桿這兩個分解動作和其他分解動作相比之下準確率還是偏低，未來也許可以針對這兩個分解動作之特性進行特殊處理，以提升準確率。

第二個可優化的項目為改善三維人體模型估計模型避免估計異常。本研究實驗時發現當三維人體模型的四肢和身體兩者位置相互距離太近時，會導致預測出的三維人體模型產生四肢穿過身體的不正常現象。為了避免出現此狀況，必須要將此三維人體模型估計方法再加以改良。

第三個可優化的項目為新增高爾夫球桿之三維模型。高爾夫球桿之三維模型若能與三維人體模型整合，就能夠讓使用者在觀察比對結果時，也能參考高爾夫球桿正確的位置資訊，更易了解使用者和教練兩者高爾夫球桿的位置差異，進一步校正高爾夫揮桿姿勢。

本研究在第一大步驟高爾夫揮桿分解動作擷取中已經選擇使用輕量級網路，但是在第二大步驟三維人體模型姿勢比對分析中所使用的網路模型目前還無法達到輕量級的要求。未來為了達成將本系統應用至嵌入式行動裝置中的目的，必須要研發輕量級三維人體模型估計法，且該模型必須維持良好的三維人體模型重建之準確率，才能達到將本研究的最終目標。期望未來視覺式智慧型高爾夫揮桿動作姿勢分析系統能夠被實際運用於日常生活中，得以幫助高爾夫球運動者能夠提升訓練品質以及降低運動傷害發生。最後期盼本系統功能能夠更加完善。

參考文獻

[Mal95] W. J. Mallon and A. J. Colosimo, “Acromioclavicular Joint Injury in Competitive Golfers,” *Journal of the Southern Orthopaedic Association*, pp. 227-82, 1995.

[The98] G. Thériault and P. Lachance, “Golf Injuries. An overview,” *Sports Med*, pp. 43-57, 1998.

[Mch07] A. McHardy, H. Pollard, and K. Lou, “The Epidemiology of Golfrelated Injuries in Australian Amateur Golfers - A Multivariate Analysis,” *South African Journal of Sports Medicine*, pp. 12-19, 2007.

[Cho12] P. Chotimanus, N. Cooharojananone, and S. Phimoltares, “Real Swing Extraction for Video Indexing in Golf Practice Video,” *Proceedings of Computing, Communications and Applications Conference*, Hong Kong, China, pp. 420-425, 2012.

[Noi13] S. Noiumkar and S. Tirakoat, “Use of Optical Motion Capture in Sports Science: A Case Study of Golf Swing,”*2013 International Conference on Informatics and Creative Multimedia*, Hong Kong, China, pp. 310-313, 2013.

[Mcn19] W. McNally, K. Vats, T. Pinto, C. Dulhanty, J. McPhee, and A. Wong, “GolfDB: A Video Database for Golf Swing Sequencing,” *Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), Long Beach, CA, pp. 2553-2562, 2019.

[How17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” *arXiv preprint arXiv: 1704.04861*, 2017.

[How19] A. Howard, M. Sandler, G. Chu, L. C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, “Searching for MobileNetV3,” *Proceedings of International Comference on Computer Vision* (*ICCV*), Seoul, Korea, pp. 1314-1324, 2019.

[Hu18] J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” *Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Salt Lake City, UT, pp. 7132-7141, 2018.

[Zha18] X. Zhang, X. Zhou, M. Lin, and J. Sun, “ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices,” *Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Salt Lake City, UT, pp. 6848-6856, 2018.

[Ma18] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, “ShuffleNetV2: Practical Guidelines for Efficient CNN Architecture Design,” *Proceedings of the European Conference on Computer Vision* (*ECCV*), pp. 116-131, 2018.

[San18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. -C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Salt Lake City, UT, pp. 4510-4520, 2018.

[Ko21] K. R. Ko and S. B. Pan, “CNN and bi-LSTM based 3D Golf Swing Analysis by Frontal Swing Sequence Images,” *Multimedia Tools and Applications*, pp. 8957-8972, 2021.

[Lop15] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, “SMPL: A Skinned Multi-Person Linear Model,” *Association for Computing Machinery* (*ACM*), pp. 248:1-248:16, 2015.

[Kan18] A. Kanazawa, M. J. Black, D. W. Jacobs, and J. Malik, “End-to-end Recovery of Human Shape and Pose,” *Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Salt Lake City, UT, pp. 7122-7131,2018.

[Koc20] M. Kocabas, N. Athanasiou, and M. J. Black, “VIBE: Video Inference for Human Body Pose and Shape Estimation,” *Proceedings of 2020 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, WA, pp. 5252-5262, 2020.

[Mah19] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. Black, “AMASS: Archive of Motion Capture as Surface Shapes,” *Proceedings of*

2019 IEEE International Conference on Computer Vision (*ICCV*), Seoul, Korea, pp. 5441-5450,2019.

[Cho21] H. Choi, G. Moon, J. Y. Chang, and K. M. Lee, "Beyond Static Features for Temporally Consistent 3D Human Pose and Shape from a Video," *Proceedings of 2021 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, TN, pp. 1964-1973, 2021.

[Zha21] H. Zhang, Y. Tian, X. Zhou, W. Ouyang, Y. Liu, L. Wang, and Z. Sun, "PyMAF: 3D Human Pose and Shape Regression with Pyramidal Mesh Alignment Feedback Loop," *Proceedings of 2021 IEEE International Conference on Computer Vision* (*ICCV*), Quebec, Canada, pp. 11446-11456, 2021.

[Bog16] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. J. Black, "Keep It SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image," *Proceedings of 2016 European Conference on Computer Vision* (*ECCV*), pp. 561-578, 2016.

[Pis16] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler, and B. Schiele, "Deepcut: Joint Subset Partition and Labeling for Multi Person Pose Estimation," *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, NV, pp. 4929-4937, 2016.

[Omr18] M. Omran, C. Lassner, G. Pons-Moll, P. V. Gehler, and B. Schiele, "Neural Body Fitting: Unifying Deep Learning and Model-Based Human Pose and Shape Estimation," *Proceedings of 2018 International Conference on 3D Vision* (*3DV*), pp. 484-494, 2018.

[Kol21] N. Kolotouros, G. Pavlakos, D. Jayaraman, and K. Daniilidis, "Probabilistic Modeling for Human Mesh Recovery," *Proceedings of 2021 IEEE International Conference on Computer Vision* (*ICCV*), Montreal, QC, pp. 11605-11614, 2021.

[Fie21] M. Fieraru, M. Zanfir, S. C. Pirlea, V. Olaru, and C. Sminchisescu, "AIFit: Automatic 3D Human-Interpretable Feedback Models for Fitness

Training," *Proceedings of 2021 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, TN, pp. 9914-9923, 2021.

[Xie19] H. Xie, A. Watatani, and K. Miyata, "Visual Feedback for Core Training with 3D Human Shape and Pose," *2019 Nicograph International* (*NicoInt*), pp. 49-56, 2019.

[Cao17] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, HI, pp. 7291-7299, 2017.

[Pre16] L. L. Presti and M. L. Cascia, "3D Skeleton-Based Human Action Classification: A Survey," *Pattern Recognition*, pp. 130–147, 2016.

[Rez15] D. J. Rezende and S. Mohamed, "Variational Inference with Normalizing Flows," *Proceedings of the 32nd International Conference on Machine Learning* (*ICML*), pp. 1530-1538, 2015.

[Sin14] N. Singla, "Motion Detection Based on Frame Difference Method," *International Journal of Information & Computation Technology*, pp. 1559-1565, 2014.

[Iof15] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *Proceedings of the 32nd International Conference on Machine Learning* (*ICML*), pp. 448-456, 2015.

[Wu18] Y. Wu and K. He, "Group Normalization," *Proceedings of the European Conference on Computer Vision* (*ECCV*), pp. 3-19, 2018.

[Ram17] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for Activation Functions," *arXiv preprint arXiv: 1710.05941,* 2018.

[Wan20] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," *Proceedings of 2020 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, WA, pp. 11531-11539, 2020.

[1] Top 10 Most Popular Sports In The Word: https://sportsshow.net/top-10-most-popular-sports-in-the-world/，2021年。

[2] International Golf Federation National Members: https://www.igfgolf.org/about-igf/nationalmembers/。

[3] The Royal & Ancient Golf Club of St Andrews：Record Numbers Now Playing Golf Worldwide: https://api.randa.org/en/news/2021/12/record-numbers-now-playing-golf-worldwide。

[4] 科技戰已成常態，運動員背後的「神隊友們」: https://www.inside.com.tw/feature/digi-plus/25004-digi-plus-science-sport，2021年。

[5] 運動科技導入，智能高爾夫增加擊球人口: http://www.taiwangca.org.tw/news/data.php?id=1145，2021年。

[6] 孫藝珍、玄彬、全智賢都愛打「高爾夫球」，輕輕一揮竿全身瘦：https://www.womenshealthmag.com/tw/fitness/work-outs/g35690208/golf-benefit/，2021年。

[7] 揮桿姿勢多重要？一三高爾夫創辦人：「1 年錯誤要花 3 年修正！」: https://news.ebc.net.tw/news/living/159443，2019年。

[8] 揮桿技術及原理(Swing Fundamentals & Skills):http://www.garoc.org/images/referee_coach/201910181834563702.pdf。

[9] Reading: ShuffleNet V2 — Practical Guidelines for Efficient CNN Architecture Design: https://sh-tsang.medium.com/reading-shufflenet-v2-practical-guidelines-for-e-fficient-cnn-architecture-design-image-287b05abc08a。

[10] 誰才是輕量級 CNN 的王者？7個維度全面評測 mobilenet/shufflenet/ghostnet: https://www.bilibili.com/read/cv8801259。

[11] Humans Process Visual Data Better:https://www.t-sciences.com/news/humans-process-visual-data-better。