第 19 卷第 1 期，09 - 14 頁 (2022 年 03 月)

基於深度學習進行模擬健身房單一器材與整體環境之動作辨識

[1]陳正鑫*、[2]陳五洲、[1]何金山、[3]李仁軍

[1]國立體育大學、[2]台北海洋科技大學、[3]國立高雄科技大學

投稿日期：2021 年 05 月；通過日期：2021 年 06 月

摘要

科技日新月異人們生活品質逐漸在改變，人們從以前書信往來到現在的視訊電話。人類所聽到的聲音、所感受到的溫度及看到的畫面，都是受器接收後，經由神經傳導到大腦，再藉由大腦的綜整處裡所得到的資訊，對於電腦而言，這些聲音、訊息及畫面就是大量的類比及數位訊號組合而成的資訊。如何透過電腦進行影像監控，並將得來的影像進行解析，獲取我們所想要的結果，始終是一項熱門的課題。以目前的運動監測系統來說，現今許多人為了身體健康，會上健身房運動，而戴上運動手環進行監測目前生理機能如何，再配合巡場的人員觀察與指導，以達到運動健身的效果。本研究主要探討如何針對鏡頭攝影的每個人狀態進行標示。若有突發狀況則會提供警示，不再只是單純的監視畫面，而是能真正掌握每個人的狀態監測系統。以研究面向來說，R-CNN、Fast R-CNN、Faster R-CNN 及 Mask R-CNN 等許多以卷積類神經網路 Convolutional Neural Networks (CNN) 為主軸核心擷取圖像特徵，以達成物件辨識與追蹤技術，其成果已應用在各個領域中。所以本研究藉由羅技 C310 攝影機，並透過 YOLO v4 深度學習模型結合即時影像進行人體狀態辨識，配合使用者讓使用者能依據自身需求建立屬於自己的人體狀態資料庫，以達到人體狀態監測的效果。

關鍵詞：機械學習、YOLO、狀態辨識、影像處理

壹、 緒論

動作型態辨識在運動科學領域中是最基礎但不可缺少的技術，隨著深度學習、數據科學與物聯網技術越趨成熟，使原本無法完成的項目，慢慢有解決的方法，其中動作識別和姿態分析就是最明顯的例子，從 2007 年 Mizuno 等學者透過陀螺儀與加速度計針對人體一天活動狀態進行資料收集與計算藉此來辨識人體動作 (Mizuno et al., 2007)，到 2017 年 Zhe Cao 等學者，透過即時二維影像針對人體姿態，進行人體姿態骨架建立，藉此來達到人體骨架辨識 (Cao, Simon, Wei, & Sheikh, 2017)，從動作辨識相關研究得知，研究者會透會根據不同的應用環境需求，選擇適合的方式來進行人體姿態的捕捉，目前常見的方式可分為非視覺類型 (Non-Vision Based) 與視覺類型 (Vision Based) 兩種: 非視覺型 (Non-Vision Based) 的方式透過常見的感測模組，如：肌電模組、加速計、陀螺儀及測力板等透過不同電子元件特性組成能收集關節角度的變化、肌肉活化特性、運動表現與能量消耗等相關人體活動量與運動學參數 (Zhou & Hu, 2008)，而視覺型 (Vision Based) 的方式則是透過光學感測器，如相機或是紅外線高速攝影機等相關能將影像記錄下來的設備，但針對視覺型設備又分為有標記與無標記這兩項，有標記的會針對欲收集之身體相關部位貼上標示點後，影像收集完成再針對標記點進行標示，如：Vicon 和 motion。無標記點則是沒有標記點，直接將影像透過影片剪輯軟體或是運動影片分析系統進行相關人體關節參數收集，如：OpenPose (Cao, Hidalgo, Simon, Wei, & Sheikh, 2019)和 3D Pose Estimation (Vosoughi & Amer, 2018)。但隨著圖形處理器 Graphics Processing Unit (GPU) 效能提升、圖形識別的演算法類神經網路概念的普及與演算法的修正及精進，影像辨識的處理方法越來越多樣，已漸漸由傳統的影像處理，開始往人工智慧與機械學習方向發展，其中卷積神經網路 (CNN,Convolutional Neural Networks) 在近年舉辦的視覺辨識競賽總是名列前茅 (Girshick, 2015)，其主要原因在於此神經網路不像其他神經網路只有單純提取圖像資料進行運算比對，而是透過壓縮圖片畫素在進行特徵 (feature) 圖比較來識別出圖片，所以在針對人體動作辨識與骨架相關所使用到的深度學習技術大多與 CNN 有關，如 Openpose、Fast R-CNN、Faster R-CNN 和 YOLO 等等。

*通訊作者：陳正鑫 國立體育大學
地址：桃園市龜山區文化一路250號
E-mail：1080209@ntsu.edu.tw

DOI：10. 53106/207332672022031901002

但目前國內針對人體動作分析研究或檢測上都是透過精密儀器進行精密的動作分析及量化，其缺點價格昂貴且架設麻煩。並不適合使用在一般場域中，例如：一般民眾在健身房做肌力訓練或體能訓練等動作時。由於一般大眾可能沒有受過專業運動訓練，姿勢上可能會造成身體損傷的情況，亦可能在運動過程中會有突發意外的情況發生，所以這種場域需要簡單且即時的動作分析或是物件追蹤這類型的系統，來進行狀態檢測及監控。

有鑑於此，本研究希望藉由類神經網路及深度學習模型來取代傳統影像分析，做到人體狀態辨識工作。原因在於傳統影像只能監測該畫面的狀況，而無法針對該環境中每一位成員的狀態進行檢測。故本研究將會先針對目前眾多的物件辨識技術透過 COCO 數據測試集 (Bochkovskiy, Wang, & Liao, 2020) 進行比較，再挑選適合於本研究的深度學習之模型，表 1 為各模型之數據集。

表 1 COCO 數據集

Method	Size	AP	FPS
YOLO v1	448	21.6%	45
SSD	300	25.1%	43
	512	28.8%	22
YOLO v2	416	21.6%	40
YOLO v3	320	28.2%	45
	416	31%	35
YOLO v4	416	41.2%	38
	512	43%	31
	608	43.5%	23

經由 COCO 數據集發現 YOLO v4 不論是 AP (average precision,精度) 還是 FPS (frame per second, 影格率) 都具有一定的水準，反觀其他模型，雖有較高精準度，但是在影像處理速度上較慢，故在危險事故發生時無法在第一時間將危險資訊辨識出來。反觀辨識較快的模型，其精度較差，有較大機率辨識錯誤，故本研究選擇辨識速度與精度都具有一定水準的 YOLO v4 (Bochkovskiy, Wang, & Liao, 2020)。環境測試如圖 1 所示，本研究將透過 YOLO v4 研究一款可以根據環境需求，建立對應的人體狀態的資料庫與模型，並且根據當前每一位成員狀態反饋給監控者，而這套人體狀態監測系統的最大特色，在於只要建立起該系統，即能透過即時影像，讓監測者明白監測區域內每一位成員目前行為狀況或是單一器材上每個人員的狀態。

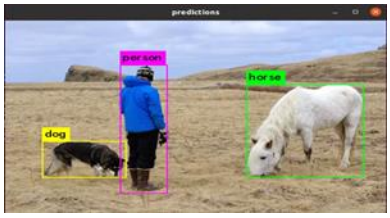

圖 1. Darknet YOLO v4 (Bochkovskiy et al., 2020)測試畫面。

貳、方法

本次研究主要會針對人們進健身房後使用比例較高的器材與容易發生危險的動作，如：硬舉、握推、伸展、深蹲、二頭彎舉、走、跑、站立等等。其中硬舉、握推和深蹲是健身房中常見發生意外的運動項目，其意外情況如：昏倒、重物壓傷與跌倒等對人體造成傷害的情況，所以本研究在國立體育大學運動科技實驗室模擬成健身房環境 (如圖 2 所示)，讓實驗室人員每周一次每次一小時共三周，在模擬的環境中進行運動並透過監視器進行錄製完後將監視器畫面提取出來後，透過 LabelImg (Singh & Bhushan, 2019) 進行動作名稱標示 (如圖 3 所示)，標記完後的文字檔放入 Darknet YOLO v4 (Bochkovskiy et al., 2020) 底下進行訓練，本研究針對整體環境所收集的資料中包含 19 個在健身房常見的動作以及 1 個異常動作，異常動作中包含跌倒、重物壓身或是暈倒等可能造成人受傷的情況都統一定義為異常動作。整體環境使用影像訓練資料有 5000 張訓練圖像，3000 張為測試圖像，共計 8000 張，另外針對單一器材所收集到的資料包含站、走、慢跑、跑以及非使用器材人員。而單一器材所使用影像訓練資料使用有 3000 張訓練圖像，1000 張為測試圖像，共計 4000 張。進行訓練過程中，每次迭代結果都會輸出在命令提示字元視窗中 (如圖 4 所示)，每 1000 次迭代後輸出的權重檔會存放在在 Darknet YOLO 資料夾 ./backup/目錄中。

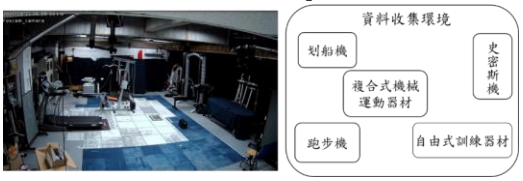

圖 2. 健身房模擬環境

圖 3. LabelImg 操作畫面

DOI：10. 53106/207332672022031901002

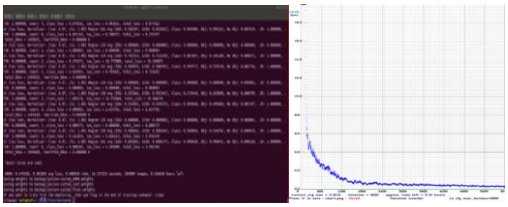

圖 4. 模型訓練

本研究中所使用到的電腦軟硬體相關設備，如表 2 所示。

表 2 研究設備與環境

中央處理器：Intel(R) Core(TM) i7-9700 @ 3.0-4.7GHz
記憶體：DDR4 16GB
顯示卡：NVIDIA RTX-2070
作業系統：Ubuntu18.04　64 位元
影像鏡頭：Logitech C310
運行環境：Python3、yolo v4、Darknet、LabelImg

參、結果

本研究經由訓練後分成圖片與影片進行測試 (如圖 5～圖 8 所示)。

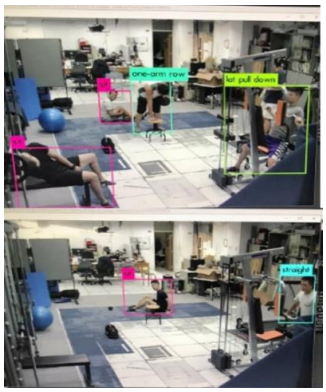

圖 5. 模擬健身房環境圖片辨識結果

圖 6. 模擬健身房環境影片辨識結果

圖 7. 模擬單一器材環境圖片辨識結果

圖 8. 模擬單一器材環境影片辨識結果

表 3 與表 4 針對模擬健身房環境與單向項目器材_跑步機訓練完後進行測試，並將各個動作項目的平均精度 (mAP, mean average precision) 列出。我們可以透過各項目的 mAP，來觀察各個項目訓練後的狀態是否良好。

表 3 模擬健身房環境各動作 mAP

動作	mAP
walk (走路)	.98
deadlift (硬舉)	.94
cable seat close row(划船運動)	.94
stand (站立)	.98
carry (提起物品)	.96
bench press (臥推)	.92
sit (坐)	.90
stretch (伸展)	.90
straight arm pull down (直臂下拉)	.85
lat pull down (滑輪下拉)	.84
squat (深蹲)	.95
change weight (更換重量)	.82
run (跑步)	.91
biceps curl (二頭灣舉)	.94
pecdeck (夾胸)	.91
one-arm row (單手啞鈴划船)	.82
chest push (坐式胸推)	.92
core training (核心訓練)	.84
hip push (臀推)	.85
dangerous (危險)- 異常動作	.97

DOI：10. 53106/20733267202203190100 2

表 4、模擬單一器材環境各動作的 mAP

動作	mAP
stand(站)	.99
walk(走)	.98
joe(慢跑)	.97
run(跑)	.97
other(其他_不再器材上人員)	.92

肆、討論

本次研究目的為能否針對畫面中每個人的狀態進行辨識，但更深層的含意在於假設發生危險情況時能是否能夠精確標示狀態。在本次研究中危險狀態大多都是模擬現實可能會發生的情況如：跌倒、暈倒或是休克等一些危險情形，但經由訓練與驗證，由結果發現其辨識精確程度仍就高達 9 成 7。所以透過 YOLO v4 (Bochkovskiy et al., 2020)進行人體動作行為辨識這方向是可行的，且本研究結果的各項運動狀態的 mAP 顯示本次訓練效果良好，若未來以實際的健身房的場景下，是有機會完成健身房內每個成員的運動狀況與警示系統。

依據圖 5 至圖 8 觀察得知本研究結果是可以針對一個模擬健身房環境或是單一器材中，將人體狀態能夠準確的標示並將狀態顯示出來。另外我們也將本研究針對模擬健身房環境中無論是整體環境或是單一器材項目人體動作辨識的精確性進行計算後發現其辨識平均精度 mAP 至少都有 8 成以上。其平均精度這項數據的好壞，對於本研究整體模型來說，除了該畫面這個人的動作狀態辨識精確程度之外，也可以透過該項數值來驗證該模型訓練對於同一個場景下不同情況下的人體動作狀態或物件辨識情況是否完善。

YOLO v4 (Bochkovskiy et al., 2020) 是否能夠針對模擬健身場域環境的人體狀態進行辨識所以將許多健身房常見的動作都納入模型訓練中，在未來可能對於動作辨識上，只會剩下正常與異常動作情況進行辨識，例如：目前應用於老人跌倒狀況。而在健身房的情況亦是如此，正常動作情況下，會將該人標記並標示為正常，如有發生意外，例如：跌倒、槓子壓在身上或是昏厥等等意外狀態，則會表示危險或是異常。在將其系統延伸至當有意外狀況發生時，可透過手機 APP、信箱以及聲音警報器進行警示通報，將意外降至最低，讓憾事不再發生。

針對於單一器材使用未來希望能透過這個環境再進行延伸，可結合目前熱門人體骨架辨識 OpenPose (Cao et al., 2019) 並結合手機 APP，來進行動作校正、計算關節活動角度及動作評估等，可透過這些資料，來協助日益蓬勃的健身產業上，增進該產業的相關產值。

而本研究的情景大多都是模擬健身房會發生的情況，雖然我們結果有圖像指出不論是危險的辨識精確性或是危險狀態的標示，精確性相當高，對於畫面中成員狀態也能夠確實標示其狀態出來。但終究還是沒有實際狀況中發生得來的那麼真實，所以無法預知實際危險狀況發生時，能否如同模擬場景一樣，那麼確實且有效。接下來，會透過與實際健身房的場域進行資料收集與訓練，屆時整體系統將會更完善，另外未來為了提高更好的辨識率，將 YOLO v4 (Bochkovskiy et al., 2020) 改成 YOLO v4-tiny (Jiang, Zhao, Li, & Jia, 2020) 雖說辨識精確會稍微降低，但辨識速度提升且 CPU 或 GPU 使用量同時也會降低 (Jiang et al., 2020)。對於系統上來說可以是更快確認每個人的狀態且架構能減省許多外，同時可以不用再用帶有 NVIDIA 顯示卡的桌機執行系統，而可改用 NVIDIA JETSON NANO (Jetson NANO,NVIDIA, State California,USA)，以節省整體系統空間與費用。

五、結論

本研究結論是可以透過YOLO v4 (Bochkovskiy et al., 2020) 環境進行模型訓練後再透過即時影像進行辨識影像中，每個人當前狀況。在未來我們可以藉由實際的健身房環境，收集大量的影像資料與該環境中容易發生意外狀況的影像，透過模型訓練後，放在該場域進行動作狀態檢測。

陸、參考文獻

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. IEEE transactions on pattern analysis and machine intelligence, 43(1), 172-186.

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Girshick, R. (2015). Fast r-cnn. Paper presented at the Proceedings of the IEEE international conference on computer vision.

Jiang, Z., Zhao, L., Li, S., & Jia, Y. (2020). Real-time object detection method based on improved YOLOv4-tiny. arXiv preprint arXiv:2011.04244.

Mizuno, H., Nagai, H., Sasaki, K., Hosaka, H., Sugimoto, C., Khalil, K., & Tatsuta, S. (2007). Wearable sensor system for human behavior recognition (First report: Basic architecture and behavior prediction method). Paper presented at the TRANSDUCERS 2007-2007 International Solid-State Sensors, Actuators and Microsystems Conference.

DOI：10. 53106/207332672022031901002

Singh, J., & Bhushan, B. (2019). Real Time Indian License Plate Detection using Deep Neural Networks and Optical Character Recognition using LSTM Tesseract. Paper presented at the 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS).

Vosoughi, S., & Amer, M. A. (2018). Deep 3d human pose estimation under partial body presence. Paper presented at the 2018 25th IEEE International Conference on Image Processing (ICIP).

Zhou, H., & Hu, H. (2008). Human motion tracking for rehabilitation—A survey. Biomedical signal processing and control, 3(1), 1-18. iley & Sons.

Volume 19 Issue 1, March 2022

Motion recognition based on deep learning to simulate single equipment and overall environment of gym

[1]Zheng-Xin Chen[*], [2]Wu-Chou Chen, [1]Chin-Shan Ho,[3]Jen-Chun Lee

[1]National Taiwan Sport University, Taoyuan, Taiwan
[2]Taipei University of Marine Technology, Taipei, Taiwan
[3] National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

Recevied: 2021/05; Accepted：2021/06

ABSTRACT

The quality of people's life is gradually changing with the rapid advances in technology, from the old days of correspondence to the current video phone. The sound we hear, the temperature we feel, and the images we see are all received by the receptors, transmitted to the brain through the nerves, and then integrated by the brain to obtain the information. It is always a hot topic to monitor the image through computer and analyze the image to get the result we want. In terms of the current exercise monitoring system, many people nowadays will go to the gym to exercise for their health, and wear exercise bracelets to monitor the current physiological function, and then with the observation and guidance of the patrolling staff, in order to achieve the effect of exercise and fitness. This study focuses on how to mark the status of each person for camera photography. If there is a sudden situation will provide a warning, no longer just monitor the screen, but can really grasp the status of each person monitoring system. In terms of research, R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, and many other Convolutional Neural Networks (CNN) are used as the main axis to capture image features to achieve object recognition and tracking technology, and their results have been applied in various fields. Therefore, this study uses Logitech C310 camera and YOLO v4 deep learning model combined with real-time images for human body status recognition, and allows users to build their own human body status database according to their needs to achieve the effect of human body status monitoring.

Keywords: Mechanical learning, YOLO, Status recognition, Image processing

*Corresponding author: Zheng-Xin Chen, National Taiwan Sport University, Taoyuan, Taiwan
Address: No 250, Wenhua 1st Rd, Guishan, Taoyuan 333, Taiwan.
E-mail: 1080209@ntsu.edu.tw