

AI 自動評分系統之研究

黃照恩，吳俊霖*

國立中興大學 資訊工程學系

g112056111@mail.nchu.edu.tw

jlwu@cs.nchu.edu.tw*

論文摘要

本研究提出一種基於 YOLOv7 的視覺追蹤系統，應用於桌球的球體軌跡分析與落點分類。系統首先利用 YOLOv7 偵測並追蹤桌球位置，透過比較相鄰影格間球體的距離、長寬比、交集比 (IOU) 及移動方向來建立或更新球體軌跡。當球體脫離軌跡的拋物線運動模式時，系統判定其落點，並透過標記球桌的四個角進行透視變形，以獲取球桌面上的實際落點位置。根據落點位置，系統將擊球結果分類為有效擊球、錯誤落點、擊球出界及未擊中。

本研究以 YOLOv7 為核心，具備高速、高準確性等特性，有效提高了桌球軌跡分析的準確性和即時性。透過拋物線運動模式與透視變形技術的結合，系統能夠更準確地定位球體落點，進一步提升擊球分類的可靠性。

關鍵詞：桌球、軌跡落點追蹤、YOLOv7

前言

隨著電腦視覺與人工智慧技術的迅速發展，影像辨識與物件追蹤技術在各領域中獲得了廣泛應用，尤其在運動分析領域中，這些技術的應用潛力逐漸受到重視。運動分析不僅有助於提升運動員的表現，對於賽事策略的制定、訓練效果的評估以及比賽後的數據分析等方面，也發揮著關鍵作用。然而，對於高速運動場景，如何在保持追蹤準確度的同時，確保即時性，仍然是現有技術面臨的重大挑戰。

桌球運動作為一項高速且高度動態的運動，對影像辨識與物件追蹤系統提出了更高的要求。桌球的球體運動速度極快，且場地背景複雜，這使得傳統物件追蹤演算法在此類場景下難以同時達到高準確度與即時性的平衡。尤其在桌球比賽中，如何精確追蹤小尺寸的球體，並能夠準確分析其運動軌跡，對現有技術提出了挑戰。

在此背景下，YOLOv7 (You Only Look Once Version 7) [1] 作為 YOLO 系列演算法，憑藉其卓越的實時物件偵測性能及高效的計算設計，已成為許多物件追蹤任務中的理想解決方案。YOLOv7 在速度與準確度之間達到了優良的平衡，並具有較強的適應能力，能夠有效應對高速運動場景中的挑戰。因此，YOLOv7 的引入為本研究提供了可行且高效的技術基礎。

除了物件偵測技術，本研究進一步透過軌跡分析以及透視變形等方法，來提升桌球運動數據的準確性與實用性。這樣的系統不僅能夠協助發球機

自動判定選手是否成功擊球與擊球落點是否符合規則，還能提供精確的運動數據分析，以優化選手訓練與技術評估，進一步提升發球機的輔助功能。

研究方法

本研究提出一套基於 YOLOv7 的桌球偵測與追蹤方法，旨在解決桌球運動中快速移動的球體追蹤問題。該方法結合了先進的物件偵測技術及精確的運動追蹤演算法，能夠實現即時且高準確性的球體追蹤。具體而言，YOLOv7 作為物件偵測的基礎模型，透過其強大的特徵提取與分類能力，有效識別桌球的存在及其位置。相比傳統的物件偵測方法，YOLOv7 具有更高的處理速度和準確性，能夠在複雜的桌球比賽環境中，實現高效且穩定的球體辨識。

此外，本研究引入了一套軌跡管理機制，該機制能夠在多目標追蹤情境中處理球體的運動問題。通過對每個球體軌跡的精確估算與更新，該方法不僅能保持對每一個球路的追蹤，還能在多個球體同時運動的情況下，避免追蹤混淆或丟失目標。

本研究的方法流程圖如圖 1 所示，首先會介紹 YOLOv7 的原理及架構，再來介紹如何透過相鄰影格間球體本身的特性來建立或更新球體軌跡，再來介紹利用脫離拋物線尋找桌球關鍵落點影格，最後介紹透過透視變形取得落點位於桌面上的哪一個點，並根據落點區域進行分類。

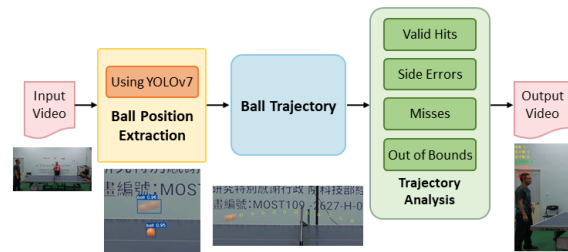


圖 1. 本研究方法流程圖。

3.1 物件偵測模型 (Object Detection Model)

近年來，深度學習技術已廣泛應用於影像分類 (Image Classification)、物件偵測 (Object Detection) 及語意分割 (Semantic Segmentation) 等領域。其中，卷積神經網路 (Convolutional Neural Network, CNN) 作為核心架構，結合特徵擷取 (Feature Extraction) 與分類網路，以實現高效的視覺任務處理。特徵擷取部分主要由卷積層 (Convolutional Layer) 與池化層 (Pooling Layer) 構成，其中卷積層透過帶有可學習參數的卷積核 (Kernel) 進行運算，並在設定的步長 (Stride) 下對輸入影像執行卷積操作，以擷取具高層次語意 (High-Level Semantic) 的特徵。這些特徵經過多層卷積與池化操作後，形成特徵圖 (Feature Map)，

進一步傳遞至分類網路進行最終的分類決策。卷積運算的流程可參考圖 2 所示。

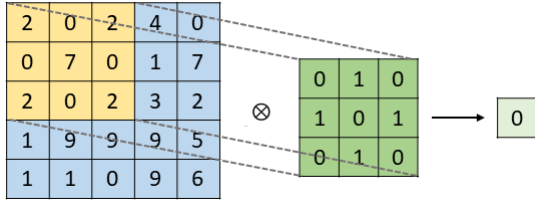


圖 2. 卷積網路運算示意圖。

為了提升模型的準確性，深度學習模型透過損失函數 (Loss Function) 計算預測結果與真實標籤之間的誤差，並利用反向傳播演算法 (Backpropagation) 調整神經網路中的權重 (Weight)，使誤差逐步下降，進而提升模型的學習能力與泛化效果。

在物件偵測 (Object Detection) 任務中，CNN 技術被廣泛應用於物件的定位與分類。其中，一階段 (One-Stage) 架構因其高效性與端對端 (End-to-End) 學習能力而受到廣泛關注。該方法將物件定位與分類整合至單一網路架構內，同時進行特徵擷取與目標識別，從而提升處理效率。代表性的一階段物件偵測模型包括 Single Shot MultiBox Detector (SSD) [2]、RetinaNet[3]、You Only Look Once (YOLO) [2] 及其變體。這些方法通常具有較為精簡的網路結構與較快的推理速度，使其適用於即時應用場景。

考慮到桌球運動中球體高速移動與頻繁變化的特性，YOLOv7 在速度與準確度之間達到了優良的平衡，並具有較強的適應能力，能夠有效應對高速運動場景中的挑戰。YOLOv7 架構如圖 3 所示。

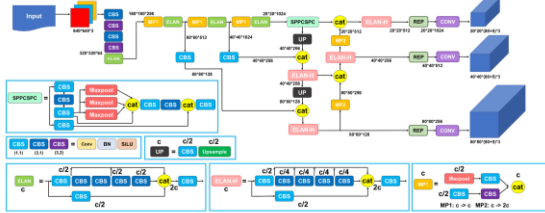


圖 3. YOLOv7 架構圖。

3.1.1 激勵函數 (Activation Function)

在影像特徵提取過程中，當前網路層的神經元透過對前一層輸出的線性組合來生成輸出，並將其傳遞至下一層。然而，此種線性映射方式使得神經網路的輸入與輸出之間僅具有線性關係，限制了其對高階特徵的表達能力及對複雜模式的學習能力。為了引入非線性變換並提升模型的特徵擷取能力，神經網路通常採用激活函數 (Activation Function)，使網路能夠學習更具表現力的映射關係。常見的激活函數包括修正線性單元 (Rectified Linear Unit, ReLU) [4] 與洩漏修正線性單元 (Leaky ReLU) [5]。

ReLU 的運作機制如圖 4 及式(1)所示，當輸入為正數時，輸出與輸入值相同，而當輸入為負數時，輸出則為零。由於在負值區域梯度消失，導致部分神經元無法更新，形成「神經元死亡問題」 (Dead ReLU Problem)。

為了解決此問題，Leaky ReLU 如圖 5 及式(2)所示，透過在負值區域引入一個小斜率 α 來維持非零梯度，使神經元仍然能夠學習。然而，Leaky ReLU 的非線性變換相對簡單，在複雜特徵的擷取上仍存在一定限制。

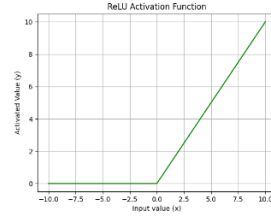


圖 4. ReLU 運作示意圖。

$$ReLU(x) = \max(0, x) \quad (1)$$

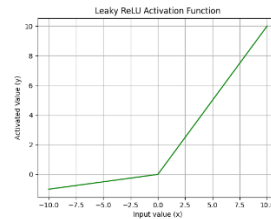


圖 5. Leaky ReLU 運作示意圖。

$$Leaky ReLU(x) = \max(\alpha x, x) \quad (2)$$

為了進一步提升神經網路的學習能力，YOLOv7 採用了 SiLU (Sigmoid Linear Unit) 激活函數[6]，其運作機制如圖 6 及式(3)所示。SiLU 函數的計算方式為將輸入值與其對應的 Sigmoid 函數值相乘，使其在輸入較大時趨近於線性變換，而在輸入較小時則逐漸逼近零。相較於傳統的 ReLU 或 Leaky ReLU，SiLU 具備更平滑的曲線特性，使其輸出隨輸入值連續變化，有助於穩定梯度更新並提升模型在深度學習過程中的收斂效果，進而增強對影像中複雜特徵的學習能力。

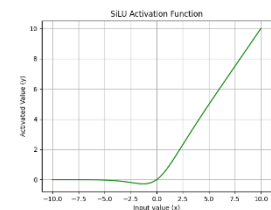


圖 6. SiLU 運作示意圖。

$$SiLU(x) = x \left(\frac{1}{1 + e^{-x}} \right) \quad (3)$$

3.1.2 邊界框損失函數 (Bounding Box Loss Function)

在深度卷積神經網路的訓練過程中，模型透過最小化損失函數來逐步逼近真實標籤，以降低預測誤差並提升準確性。在物件偵測任務中，傳統的 IoU (Intersection over Union) 用於衡量預測邊界框與真實邊界框之間的重疊程度，其計算方式為兩者的交集區域除以聯集區域，公式如式(4)所示。然而，IoU 僅關注邊界框的重疊比例，未能考慮物體內部可能存在的空洞或細節，導致其在某些情境下無法

準確評估偵測結果的精確性。物體內部的空洞指的是位於物體內部但屬於背景的区域，而非物體本身的一部分，這可能影響 IoU 對檢測準確度的衡量能力。

$$IoU = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad (4)$$

為了克服 IoU 的限制，Complete IoU (CIoU) [7] 在評估物件偵測準確性時額外納入對物體完整性的考量，將物體內部結構及細節納入計算範圍，以提供更全面的衡量標準。CIoU 的計算公式如式 (5) 所示。

$$CIoU = 1 - IoU + \frac{\rho^2(b^{pred}, b^{gt})}{c^2} + \alpha v \quad (5)$$

其中， b^{gt} 及 b^{pred} 分別為真實邊界框與預測邊界框的中心點， ρ 表示兩個中心點之間的歐式距離，而 c 則為同時包含真實與預測邊界框的最小區域的對角線長度。此外，CIoU 透過加入權重函數 α 以及長寬比相似度量 v (公式如式 (6) 所示)，進一步提升對目標形狀及比例的考量。

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w^{pred}}{h^{pred}} \right)^2 \quad (6)$$

其中， h^{gt} 和 w^{gt} 分別代表真實的長和寬的數值， h^{pred} 和 w^{pred} 分別代表預測的長和寬的數值。

本研究採用 CIoU 作為邊界框監督機制，以提升物件偵測模型的準確性。透過此方法，模型能夠更精確地學習邊界框的最佳位置，並有效減少誤差，使其逐步逼近真實答案，進而提升物件偵測的整體性能。

3.1.3 多尺度特徵融合框架 (Multi-Scale Feature Fusion Framework)

特徵金字塔網絡 (Feature Pyramid Network, FPN) 與路徑聚合網絡 (Path Aggregation Network, PAN) 共同構成了高效的多尺度特徵提取與融合模組，進一步提升了 YOLOv7 在不同尺度與形態的物件檢測能力。Lin 等人 [8] 首次提出 FPN 架構 (如圖 7 所示)，其核心思想在於透過自上而下 (top-down) 的特徵傳遞機制，將高階語義特徵逐層傳遞至低層特徵圖，以彌補低層特徵在語義資訊上的不足，從而確保模型在檢測小型物件時仍具備優異的分類能力。此外，FPN 透過層級間的資訊流動，使不同尺度的特徵圖均能獲得更豐富的表徵，進一步提升多尺度物件檢測的精準度。

相較之下，PAN 則負責自下而上 (bottom-up) 的特徵聚合，透過特徵金字塔內不同層級의 資訊交互，使低層特徵能夠接收來自高層的豐富語義資訊，而高層特徵則可融合來自底層的細節與空間資訊，從而提升物件定位能力。該結構確保了不同層級的特徵圖同時具備細節與語義信息，使模型在處理不同尺度與形態的物件時均能維持高準確度。此外，PAN 進一步透過額外的橫向連接 (lateral connections) 強化跨層級的信息流動，使特徵融合更加高效，進一步提升檢測效能。

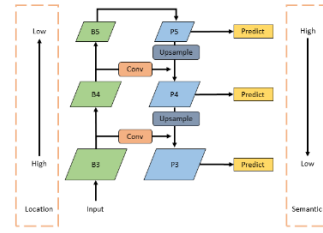


圖 7. 特徵金字塔架構圖。

3.1.4 跨階段局部池化網路 (SPPCSPC Network)

跨階段局部池化網路 (SPPCSPC) 結合了空間金字塔池化 (Spatial Pyramid Pooling, SPP) [9] 與跨階段局部網路 (Cross Stage Partial Network, CSPNet) [10] 的特性，以強化特徵提取能力並擴展感受野。在 SPP 模組中，透過不同大小的最大池化 (Maxpool) 操作對輸入張量進行處理。每次池化的步長 (Stride) 設為 1，以確保張量解析度不變。透過在多個尺度上執行池化操作，該模組能夠有效捕捉不同層級的特徵，並最終將原始輸入與池化結果串接，以提升模型對不同尺寸目標的偵測能力。

另一方面，CSPNet 透過跨階段的局部連接機制來提升特徵提取的效率，同時降低計算成本。其核心設計為將特徵圖通道分為兩部分，其中一部分經過 SPP 模組進行特徵提取，以增強模型的感受野，而另一部分則僅經過一層卷積操作以保留關鍵特徵資訊。最後，這兩部分特徵圖會進行融合，作為後續特徵提取的輸入。此架構設計不僅能夠充分利用不同層級的特徵資訊，還能有效減少參數數量，進而提升模型的推論速度與計算效率。

SPPCSPC 的結構如圖 8 所示，其中不同分支對應不同尺度的最大池化運算，此設計進一步擴展了模型的感受野，使其能夠更適應不同大小的目標物件，有助於提升目標檢測的準確性與穩健性。

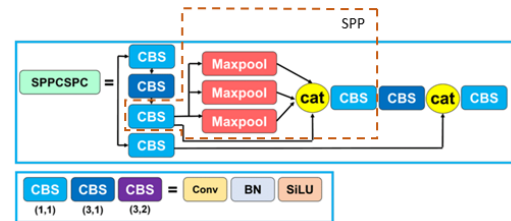


圖 8. SPPCSPC 架構圖。

3.1.1 高效層聚合網路 (Efficient Layer Aggregation Network)

Wang 等人提出的高效層聚合網路 (Efficient Layer Aggregation Network, ELAN) [11] 作為物件偵測網路的骨幹 (Backbone)，用於高效特徵提取。

ELAN 的架構設計基於梯度路徑規劃策略 (Gradient Path Design Strategies)，主要目標在於解決模型規模擴展後收斂速度下降的問題。透過分析整個網路中各層的最短與最長梯度路徑，ELAN 設計出一種能夠優化梯度傳遞效率的層級聚合機制，以確保深層網路的有效訓練與收斂。

該網路融合了 VoVNet [12] 與跨階段局部網路 (Cross Stage Partial Network, CSPNet) [10] 的優勢，

系統在高速運動場景下的適應性，並盡可能做到快速處理。

4.1 執行時間測量

為評估本系統之即時性，我們針對實驗影片進行處理時間分析，結果如表 1 所示。本研究提出之系統能夠在錄製影片長度約 1/3 的時間內完成所有球體軌跡追蹤與落點分類，顯示其優異的處理能力。

在處理 1 分 31 秒長度之輸入影片時，YOLOv7 物件偵測階段達到 58.1FPS，確保球體位置能夠即時擷取，進一步輸入至軌跡追蹤模組。透過計算相鄰影格間之長寬比變化、移動方向、距離與交集比 (IoU)，系統能夠準確建立或更新軌跡，並於球體脫離拋物線運動模式時判定落點。整體過程共計 39 秒，成功計算所有球體運動軌跡並完成落點分類，顯示本研究方法能有效應對高速運動場景，確保追蹤精準度與落點判定的準確性。

表 1. 系統處理時間

	Video Length	YOLO	Ball Trajectory Tracking	Total Processing Time
Time	1 min 31 sec	58.1 FPS	39 sec	39 sec

4.2 結果展示

本研究透過軌跡管理機制與落點座標偵測技術，精確判定球體落點位置，進而區分有效擊球與錯誤落點，對應結果如圖 12、圖 13 所示。當球體運動軌跡在相鄰 6 個影格內均無新增偵測點時，系統判定該軌跡已結束，並依據軌跡終點之運動方向來區分是擊球出界和未擊中，如圖 14、圖 15、圖 16、圖 17 所示。

此外，為評估選手擊球表現並進一步模擬比賽情境，本研究基於發球機設定，每 4 球為一回合，並依據擊球結果計算得分。若該回合內 4 球皆為有效擊球，則選手獲得 1 分；否則，發球機獲得 1 分。比賽共進行 7 個回合，最終比分將顯示於影片畫面上方，而球體軌跡分類結果則同步標示於畫面左上方，確保即時資訊呈現，提升系統於智慧桌球訓練與比賽分析之應用價值。



圖 12. 有效擊球。



圖 13. 錯誤落點。



圖 14. 擊球出界當下



圖 15. 判定擊球出界



圖 16. 未擊中當下



圖 17. 判定未擊中

結論與未來展望

本研究提出一套基於 YOLOv7 的桌球軌跡追蹤與落點分析系統，透過物件偵測與軌跡管理機制，實現球體追蹤與落點分類。系統結合透視變形技術提升落點計算準確度，並能在錄製影片長度的 1/3 時間內完成所有軌跡計算，展現良好的即時性與穩定性。實驗結果顯示，本系統可有效區分有效擊球、錯誤落點、擊球出界及未擊中，並具備比分顯示功能，適用於智慧桌球訓練與比賽分析。

本研究雖已成功實現桌球軌跡追蹤與落點分析，但仍有進一步優化空間。未來可透過多視角攝影與 3D 重建技術，提升球體軌跡與落點計算的準確性，減少視角變形誤差，並增強系統在不同場景與環境條件下的適應能力。

參考文獻

- [1] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2023, pp. 7464-7475.
- [2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, pp. 21-37, Springer International Publishing.
- [3] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. of the IEEE Int. Conf. on Computer Vision, 2017, pp. 2980-2988.
- [4] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," arXiv preprint arXiv:1803.08375, 2018.
- [5] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," in ICML Deep Learn., pp. 1-5, 2015.
- [6] S. Elfving, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement

- learning," *Neural Networks*, vol. 107, pp. 3-11, 2018.
- [7] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IOU loss: Faster and better learning for bounding box regression," in *Proc. of the AAAI Conf. on Artificial Intelligence*, vol. 34, no. 07, Apr. 2020, pp. 12993-13000.
 - [8] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2017, pp. 2117-2125.
 - [9] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904-1916, 2015.
 - [10] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 390-391.
 - [11] C. Y. Wang, H. Y. M. Liao, and I. H. Yeh, "Designing network design strategies through gradient path analysis," *arXiv preprint arXiv:2211.04800*, 2022.
 - [12] Y. Lee, J. W. Hwang, S. Lee, Y. Bae, and J. Park, "An energy and GPU-computation efficient backbone network for real-time object detection," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0-0.
 - [13] C. J. SOLER, "Table Tennis Ball Tracking and Bounce Calculation using OpenCV", in *Universitat de Barcelona*, June 22, 2017.