AI在超音波影像之應用(物件偵測篇)

連琬菁醫師[1]、周信宏副教授[2]

/台大醫院 急診醫學部[1]、國立暨南國際大學 資訊工程學系[2]

超音波因為具有非侵入性、無游離輻射及容易取得的特性，讓超音波成為臨床上常使用的臨床診斷工具之一，但是由於超音波診斷常受限於操作者經驗不足，使其診斷率變異極大。因此，近年來醫界一直關注是否有更能提升超音波診斷率之方法，除了增加超音波訓練、累積操作個案外，人工智慧(artificial intelligence, AI)亦是可能解決方案之一。

AI 超音波輔助診斷系統是目前之新興研究領域，投入此研究主題之論文發表數量，在近五年內快速成長，而超音波影像 AI 相關研究主題可歸納為四類，第一類為分類(classification)，從影像判斷是否有病灶，如區分腫瘤組織為良性或惡性腫瘤；第二類為偵測(detection)，由影像中偵測出異常的位置，例如偵測超音波影像中腫瘤存在的區域邊界(bounding box)；第三類為語義切割(semantic segmentation)，於影像中找出異常區域並標出其輪廓，例如手術過程於超音波影像標記出切除區域的輪廓，輔助醫師確定解剖結構之距離與位置；第四類為重建(reconstruction)，主要應用於強化及改善超音波設備所拍攝的影像品質。

由於超音波影像中可能會包含其他不需要處理的器官或背景雜訊，為了加速 AI 模型的訓練成效，前處理時通常會先運用物件偵測的深度學習模型（或稱為物件偵測器）定位目標器官的所在區域，以去除背景雜訊。現代物件偵測器通常由兩部分組成，包含使用 ImageNet 資料集預訓練的骨幹(Backbone)和用於預測物件的類別和邊界框的頭部(Head)。常見的骨幹深度學習架構包含 VGG[1]、ResNet[2]和 DenseNet[3]等。至於頭部則分為兩類，分別是單一階段(one-stage)物件偵測器和兩階段(two-stages)物件偵測器。

著名的單一階段物件偵測器包含 YOLO (You Only Look Once)系列 YOLOv1 ~ YOLOv7[4-10]、SSD[11]和 RetinaNet[12]等，而著名的兩階段物件偵測器則包含 R-CNN 系列 R-CNN[13]、Fast R-CNN[14]和 Faster R-CNN[15]等。在訓練和預測的速度上，單一階段物件偵測器會比兩階段物件偵測器還要快。近年來開發的物件偵測器經常在骨幹和頭部之間插入一些神經層，這些神經層通常用於收集不同階段的特徵圖，稱為物件偵測器的頸部(Neck)。常用的頸部深度學習架構包含 FPN (Feature Pyramid Network)[16]和 PANet(Path Aggregation Network)[17]等。

由於超音波影像的判讀不僅需要準確率還需要快速，這邊我們來介紹兼具快速與準確特性的 YOLOv4。YOLOv4 於 2020 年四月公開發表於 arXiv 網站[7]，迄今被論文引用的次數超過一萬次，第一位作者是來自俄羅斯曾經參與 YOLO github 項目維護的 Alexey Bochkovskiy，另外兩位則是來自臺灣中央研究院資訊科學研究所的王建堯博士和廖弘源所長。YOLOv4 修改 YOLO v3 的骨幹網路，以 CSPDarknet53 替換 Darknet53，其優點是減少參數量進而減少運算量，使訓練及辨識速度更快且能保持準確性。YOLOv4 的頸部深度學習架構採用 SPP 和 PANet 的結合，主要目的在於提升局部特徵和全局特徵的融合，增強最終特徵圖的表達能力。至於頭部深度學習架構則是採用和 YOLOv3 一樣的 Dense Prediction 單一階段預測。YOLOv4 模型的架構詳見圖 1。關於 YOLOv4 的程式碼可

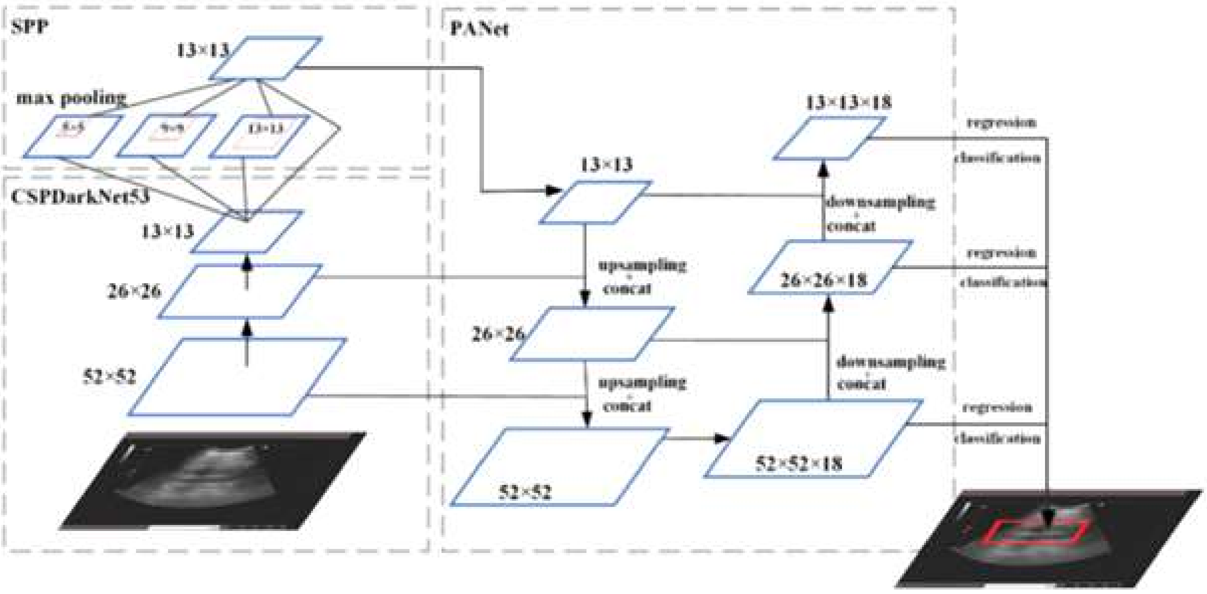

圖 1. YOLOv4 模型架構

以參考下列網址：https://github.com/AlexeyAB/darknet。

台大急診團隊亦利用 YOLOv4 進行前處理，建置水腎 AI 超音波診斷模型，研究成果已發表在 Ultrasound in Medicine and Biology [18]。

參考文獻

[1] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770–778, 2016.

[3] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 4700–4708, 2017.

[4] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 779– 788, 2016.

[5] Joseph Redmon and Ali Farhadi. YOLO9000: better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 7263– 7271, 2017.

[6] Joseph Redmon and Ali Farhadi. YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.

[7] Alexey Bochkovskiy and Chien-Yao Wang and Hong-Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934, 2020.

[8] Github: YOLOv5. https://github.com/ultralytics/yolov5.

[9] Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, Yiduo Li, Bo Zhang, Yufei Liang, Linyuan Zhou, Xiaoming Xu, Xiangxiang Chu, Xiaoming Wei, Xiaolin Wei. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. arXiv preprint arXiv:2209.02976, 2022.

[10] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696, 2022.

[11] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision (ECCV), pages 21–37, 2016.

[12] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Doll´ar. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 2980–2988, 2017.

[13] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 580–587, 2014.

[14] Ross Girshick. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 1440–1448, 2015.

[15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems (NIPS), pages 91–99, 2015.

[16] Tsung-Yi Lin, Piotr Doll´ar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2117–2125, 2017.

[17] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 8759–8768, 2018.

[18] Wan-Ching Lien, Yi-Chung Chang, Hsin-Hung Chou, Lung-Chun Lin, Yueh-Ping Liu, Li Liu, Yen-Ting Chan, Feng-Sen Kuan. Detecting Hydronephrosis Through Ultrasound Images Using State-of-the-Art Deep Learning Models. Ultrasound in Medicine and Biology. 2023 Mar; 49(3):723-733.