

應用物件辨識於氣管插管關鍵解剖結構之研究

黃琨義

南臺科技大學資訊工程系
iamkyh77@stust.edu.tw

摘要

近年來，臨床主治醫師的插管手術正逐漸從傳統的直接喉鏡 (Dedo Laryngoscope, DL) 轉變為影像喉鏡 (Video Laryngoscope, VLS)，而兩者所使用的氣管插管難度評估系統或其他相關系統是否能互相兼容仍有待驗證。因此本論文希望透過 VLS 錄製的影像進行模型訓練，並使模型能夠辨識出氣管插管過程中所出現的氣管插管的關鍵目標物件。本論文資料集搜集 7 位臨床主治醫師實際對插管模型安妮進行氣管插管的影像，並且共標記 7 種氣管插管的關鍵目標，後續採用 YOLOv7 進行模型訓練，實驗以留一交叉驗證 (Leave-One-Out Cross-Validation) 的方式進行模型評估，模型 Precision、Recall 及 F1-score 的平均與標準差已達 $92.19\% \pm 3.88\%$ 、 $92.16\% \pm 2.31\%$ 、 $92.17\% \pm 3.01\%$ 。

關鍵詞：氣管插管、關鍵目標、物體偵測、影像喉鏡

Object Detection for Automated Recognition of Key Anatomical Structures in Tracheal Intubation: A Feasibility Study

Kun-Yi Huang

Department of Computer Science and Information Engineering, Southern Taiwan University of Science and Technology

Abstract

In recent years, clinical tracheal intubation has increasingly shifted from the use of traditional direct laryngoscopes to video laryngoscopes. However, it remains unclear whether existing difficulty assessment systems and related evaluation frameworks are equally applicable to both approaches. To address this gap, this study develops an object detection model trained on images captured using a video laryngoscope system (VLS) to identify critical anatomical structures encountered during tracheal intubation. The dataset comprises images collected from seven clinical attending physicians performing intubation procedures on a training mannequin, with seven key anatomical targets manually annotated. The YOLOv7 model was employed for object detection training, and performance was evaluated using a leave-one-out cross-validation strategy. The model achieved an average precision of 92.19% ($\pm 3.88\%$), recall of 92.16% ($\pm 2.31\%$), and F1-score of 92.17% ($\pm 3.01\%$), demonstrating robust and consistent recognition performance across evaluators. These results suggest that object detection techniques combined with video laryngoscope imagery can effectively support the automated recognition of essential anatomical structures during tracheal intubation, with potential applications in training, assessment, and clinical support systems.

Keywords: Tracheal intubation, Object detection, Video laryngoscope, Anatomical structure recognition

Received: Oct. 13, 2025; first revised: Dec. 10, 2025; accepted: Jan. 2026.

Corresponding author: K.-Y. Huang, Department of Computer Science and Information Engineering, Southern Taiwan University of Science and Technology, Tainan 710301, Taiwan.

壹、研究動機與目的

近年來，臨床主治醫師的插管手術正逐漸從傳統的直接喉鏡 (Dedo Laryngoscope, DL) 轉變為影像喉鏡 (Video Laryngoscope, VLS)，DL 在插管時能夠使臨床主治醫師的視線與口腔及喉腔呈現一直線，但此方法有可能使患者在插管過程中受傷甚至引發併發症，而 VLS 則透過鏡頭將清晰的影像傳回外部螢幕，使臨床主治醫師能夠了解目前插管的位置及進度。現有以 DL 影像建立的深度學習評估系統，多依賴固定視角、穩定光源與清晰邊界等影像特性；然而 VLS 因拍攝角度動態變化、亮度與反光不穩定、組織與器械遮擋頻繁，以及個體解剖差異大，使傳統評估指標在此環境下難以準確反映模型效能，形成 DL 評估方法在 VLS 中的主要局限。但 DL 與 VLS 在手術進行的方式具有一定的差異性，因此適用於 DL 的氣管插管難度評估系統或其他相關系統是否能在 VLS 達到良好的表現仍然有待驗證。因此希望透過 VLS 所拍攝的影像進行模型訓練，並使模型能對氣管插管過程中出現的氣管插管的關鍵目標進行辨識，也將透過模型表現進行訓練集的調整使模型能具有更好的表現。期望最終結果能協助未來在進行相關研究及開發插管難度評估系統或其他相關系統時提供更多協助。

貳、文獻探討

臨床主治醫師的氣管插管常常發生在患者有生命危險時的緊急醫療服務，且氣管插管屬於一項備受矚目的手術，在研究紀錄中有 22% 會發生錯誤，而在特定的救護車上時，發生錯誤的頻率甚至高達 40% [1]，因此對臨床主治醫師的插管訓練是一項重要的議題。臨床主治醫師在訓練氣管插管 (Endotracheal Intubation, ETI) 時，需要透過經驗的累積才能提高完成率 [2]，但能提供臨床主治醫師進行插管訓練的機會卻非常少，在一項研究中蒐集了 161 份來自專職健康教育計劃認證委員會認可的培訓計劃負責人的答覆，其中 156 份有進行手術室的插管培訓，但有一半的計畫在手術室的培訓時長低於 16 小時，且培訓過程中實際執行插管的次數也有限 [3]。而在另一份研究中針對 178 名醫學生進行插管技能熟練所需要的插管次數進行評估，每位學生進行 1 至 23 次的實際插管，所有學生實際的插管次數中位數為 9 次，平均的插管成功率為 75%，結果表明成功率要達到 90% 以上至少要有 17 次的實際插管經驗 [4]。

插管培訓機會需要專家講師指導，因此一套能夠培訓新手的插管評估系統能使新手有更多的實際插管經驗，從而提升整體插管成功率及降低因失誤導致的傷害或併發症，目前已有案例是透過插管時手部特徵進行新手與有經驗的插管者分類，該系統被認為具有提升插管培訓成效的功能 [5]，其原理是分析新手與有經驗者間特徵差異並改善差異的特徵，使新手能更快了解自己應該修改的動作，且該類系統在分類新手與有經驗者時已有案例透過相關演算法與人工神經網路達到 91.17% 的準確率 [6]，除僅透過手部特徵之外也有結合其他部位特徵的案例透過手部、舌頭、門牙及插管模型頭部角度四種特徵進行新手與有經驗者的分類，且最終也取得了 96.4% 的準確率 [7]。

在過往研究中已有許多將 Machine Learning (ML) 及 Computer Vision (CV) 應用於呼吸道的案例，例如曾有研究透過 Convolutional Neural Networks (CNN) 訓練模型，進行鼻腔插管的路徑標示，並獲得 97.8% 的準確率。另外也有案例透過多種模型架構針對聲帶的影像進行分類，其架構包括 CNN 的 ResNet、Inception 及 MobileNet，並且最終三種架構訓練的模型在進行分類時的準確率分別能達到 84%、78% 及 64%，而 ResNet 及 Inception 在經過遷移學習後的準確率提高至 98.5% 及 97.1% [8]。而 CNN 也曾被運用在聲門間隙的定位和量化，該研究透過 CNN 所訓練的模型取得良好的辨識結果 [9]。而 Artificial Intelligence (AI) 應用在手術室喉鏡相關的研究中最常見的應用為識別聲帶結節 [10]，曾有研究透過 CNN 的 CNN5、VGG19、MobileNetV2 和 Inception-ResNetV2 四種架構在聲帶結節辨識獲得平均 90.63% 的準確率，其中 CNN5 所訓練的模型表現最好，辨識能達到 97.75% 的準確率 [11]。另外也有其他的案例

針對正常、聲帶結節、息肉、白斑及惡性腫瘤五個項目進行分類，透過 CNN 架構所訓練的模型最終獲得 96.24% 的準確率 [12]。而 CNN 也曾被用於協助臨床主治醫師進行手術，如過往曾有研究透過 CNN 架構訓練模型，並且記錄醫師在是否有模型輔助時判斷支氣管影像的準確率，研究結果表明在有模型輔助時醫師的準確率會有明顯的提高 [13]。除此之外也有案例是針對良性和惡性病變的分類，透過 CNN 架構訓練的模型在分類時的準確率為 78.05%，針對陰性有 95.63% 的準確率，但針對陽性則僅有 32.51% 的準確率，但該系統仍然能夠協助醫師在臨床診斷喉癌時有更正確的判斷 [14]。

除呼吸道的應用外，ML 與 CV 也常被應用於其他醫學領域相關的研究如腦電圖 (EEG) 分類，該案例透過 ML 演算法進行聆聽音樂時的 EEG 情緒辨識，在 26 個受試者獲得了約 82.29% 的平均分類準確度 [15]。此外也有應用在心電圖 (ECG) 的研究，該研究透過 CNN 為特定患者訓練專用模型，後續便可快速為較長的 ECG 訊號進行分類，且該方法適用任何 ECG 數據集 [16]。另外也有研究透過 Self-organized operational neural networks (Self-ONN) 進行 ECG 的判讀，對於正常訊號的辨識模型在 Precision、Recall 及 F1-score 達到 99.21%、99.10% 及 99.15% 的表現，並且在 Supra-Ventricular Ectopic Beat (SVEBs) 也有 82.19%、82.50% 及 82.34% 的表現，該模型在辨識 Ventricular-Ectopic Beats (VEBs) 時也能獲得 94.41%、96.10% 及 95.20% 的表現，整體辨識有著極好的表現 [17]。

YOLO (You-Only-Look-Once) 是影像處理中常見的一種模型訓練架構，該架構所訓練的模型曾經被運用在氣管插管難度的評估上，並且能在插管時提供臨床主治醫師建議，該建議是透過專業醫師對過往相似狀況所提出的建議，該系統也成功提高插管的成功率，並降低併發症發生的機率 [18]。在過去的案例中 YOLO 訓練出許多不同目的模型，其中包含用於檢測醫療設備的 ME-YOLO [19] 及即時檢測肺結節的 STBi-YOLO [20]，用於進行植牙前檢測牙槽骨和下頷管的 Dental-YOLO [21]，且這些模型與其他機器學習的架構進行結合時也能獲得不錯的效果，例如深度遷移學習 (Deep Transfer Learning, DTL) 與 Efficient-YOLOv3 結合 [22]，該案例用於執行口罩配戴檢測，這些模型在各自負責的領域也都達到了至少 90% 的準確率。且在近兩年全球受 Covid-19 疫情所影響，YOLO 便開始被大量運用在口罩配戴狀況的辨識 [23–25]，其中最好的模型表現更是能達到 99.98% 的準確率。另外也曾有透過 YOLOv5 用於檢測喉鱗狀細胞癌的案例，該案例最終獲得了 66% 的精確率及 62% 的召回率 [26]。

雖然現有研究已廣泛運用機器學習與電腦視覺技術於呼吸道評估、醫療影像分析及臨床訓練輔助領域，但針對氣管插管之即時難度判別與操作評估的研究仍十分有限。現有插管相關研究多著重於預測困難插管的術前指標或利用離線影像進行分類，而缺乏能直接應用於 VLS 影像的即時目標辨識與動態評估框架。此外，現行的插管教學與訓練系統多依賴人工評分或模擬器回饋，無法提供客觀、量化且基於實際臨床影像的操作指標。文獻中亦尚未建立一套專門針對 VLS 影像特性所設計的深度學習評估機制，致使模型效能難以準確反映真實操作情境。因此，發展一套能於 VLS 影像中自動辨識關鍵解剖結構並提供量化評估的系統，對於提升氣管插管的臨床安全性、縮短學習曲線以及強化客觀訓練回饋具有重要研究必要性與實際價值。

參、研究架構

本論文研究架構圖如圖 1 所示，設計蒐集一氣管插管影像資料庫，並由臨床主治醫師對資料庫進行氣管插管的關鍵目標標記；並將資料庫所有資料用於模型訓練，經過訓練後產生氣管插管的關鍵目標模型，後續將透過氣管插管的關鍵目標模型對測試集影像進行氣管插管的關鍵目標辨識。

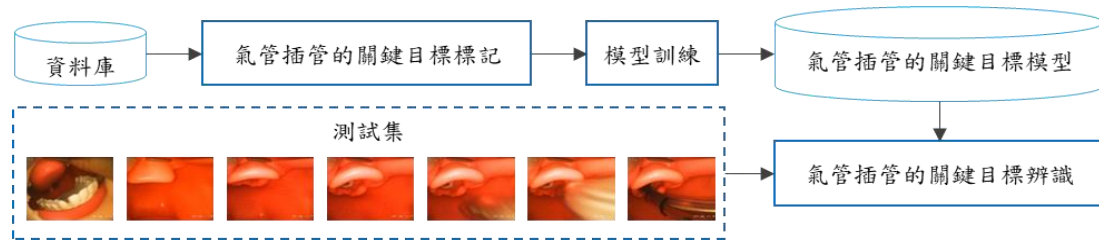


圖 1 研究架構流程圖

一、資料庫

氣管插管影像資料庫蒐集流程如圖 2 所示，首先將由 7 位臨床主治醫師實際進行氣管插管，每位醫師針對急救訓練模型安妮進行連續 3 次插管，其中插管難易度根據插管難度量表(Intubation Difficulty Scale, IDS)為第 0 級，臨床醫師插管時將透過錄影設備 Trachway 進行全程影像的錄製，所錄製之影像會直接存放至資料庫。錄製期間每秒擷取影像數量為 30 幀，且每張影像的解析度為 640×480 個像素。根據臨床醫師於氣管插管過程的連續影像定義牙齒、會厭、聲門、喉頭、管頭、管身及管身黑線共 7 種關鍵目標，並由臨床主治醫師進行標記。本論文定義的 7 種關鍵目標標記準則如圖 3 所示。會厭在最初進入氣管時的角度為垂直，後續接近物件時又會趨於平行，因此在影像上會有兩種不同的模樣；圖 3(i) 中會有多種物件同時出現，其原因為會厭、聲門、喉頭及管頭有可能會同時在影像中出現，而當四種物件重疊時，則被標記為管身如圖 3(g) 所示。部分影像因過曝或模糊等原因未被標記，因過曝而未被標記的影像如圖 3(j) 所示。

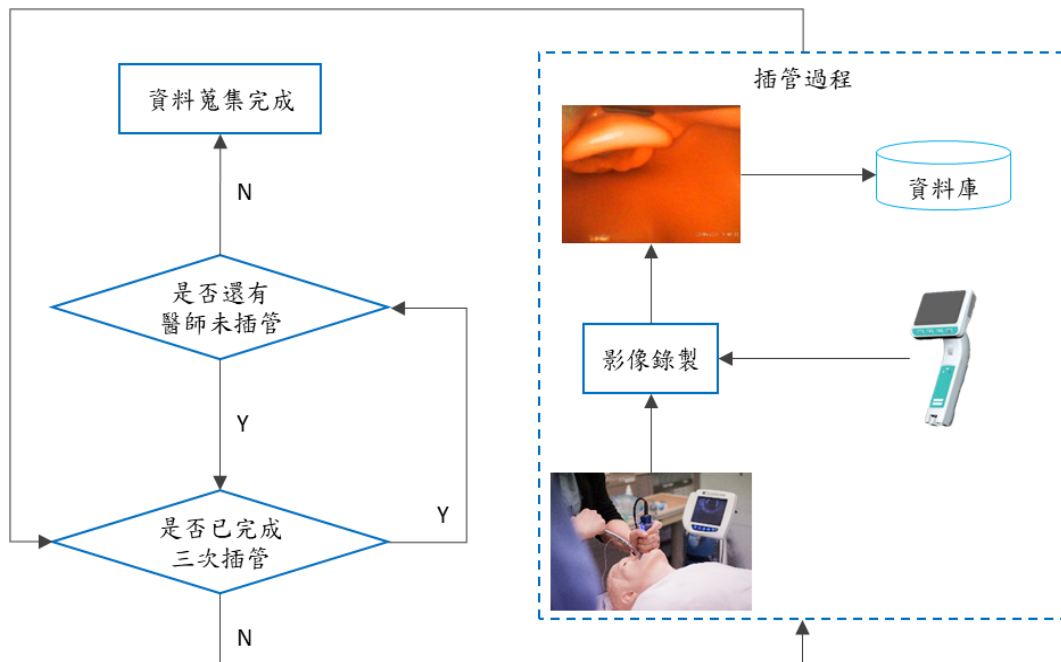


圖 2 資料蒐集流程圖

氣管插管影像共蒐集 21 個氣管插管案例，其中 1 個氣管插管影像因 Trachway 攝影機在插管前被關閉，因此刪除該份非全程氣管插管之影像，並將完整的 20 個氣管插管影像依序進行編號，各個插管影像數量以及臨床醫師標記的關鍵目標數量如表 1 所示，每次插管皆獨立為 1 個 case，並以 $D_{i,j}$ 進行編號，其中 i 代表醫師，而 j 代表該醫師第 j 次插管。共蒐集 15,216 張影像，標記 10,536 個關鍵目標。

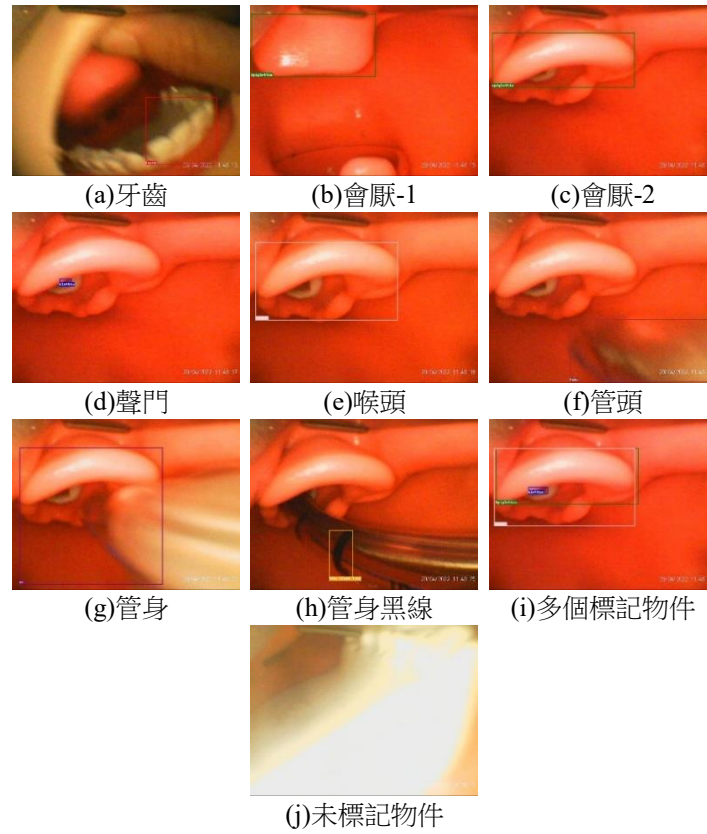


圖 3 關鍵目標標記準則示意圖

表 1 各個氣管插管影像統計數據

醫師編號	影像數量	關鍵目標標記數量
D _{1,1}	1,531	675
D _{1,2}	851	514
D _{1,3}	922	454
D _{2,1}	910	410
D _{2,2}	662	371
D _{2,3}	627	335
D _{3,1}	776	604
D _{3,2}	583	399
D _{3,3}	551	495
D _{4,1}	795	586
D _{4,2}	644	394
D _{4,3}	605	368
D _{5,1}	583	486
D _{5,2}	466	353
D _{5,3}	533	367
D _{6,1}	522	352
D _{6,2}	536	272
D _{7,1}	1,451	1,570
D _{7,2}	873	831
D _{7,3}	795	700
總和	15,216	10,536

表 2 為各個醫師在 3 次插管過程中所標記的關鍵目標數量，因每個醫師在插管時的速度不同，因此會有關鍵目標數量的差異，其中以 D₇ 的 3,101 個關鍵目標數量為最多，而缺少一個 Case 的 D₆ 為最少。

表 2 關鍵目標標記數量的統計數據

醫師編號	牙齒	會厭	聲門	喉頭	管頭	管身	管身黑線	總和
D ₁	69	511	309	272	12	83	387	1,643
D ₂	62	379	199	186	7	82	201	1,116
D ₃	57	329	192	170	12	324	414	1,498
D ₄	76	419	246	190	69	185	163	1,348
D ₅	185	350	212	210	19	75	155	1,206
D ₆	68	204	74	59	13	80	126	624
D ₇	140	1,090	708	668	42	400	53	3,101
總和	657	3,282	1,940	1,755	174	1,229	1,499	10,536

二、模型網路架構

圖 4 為 YOLOv7 網路架構，該架構相較 YOLOv5 最大的不同是將 Neck 層與 Head 層合併，而 YOLOv7 的輸入必須為解析度長寬相等的影像，其影像縮放的方式如圖 5 所示，影像解析度為 640×480，而訓練及辨識的解析度為 640×640，因此影像的寬度並不會被縮放，而高度缺失的 160 像素會在原始影像的上下方各補上 80 個值為 128 的像素，因此關鍵目標的比例不會被改變。

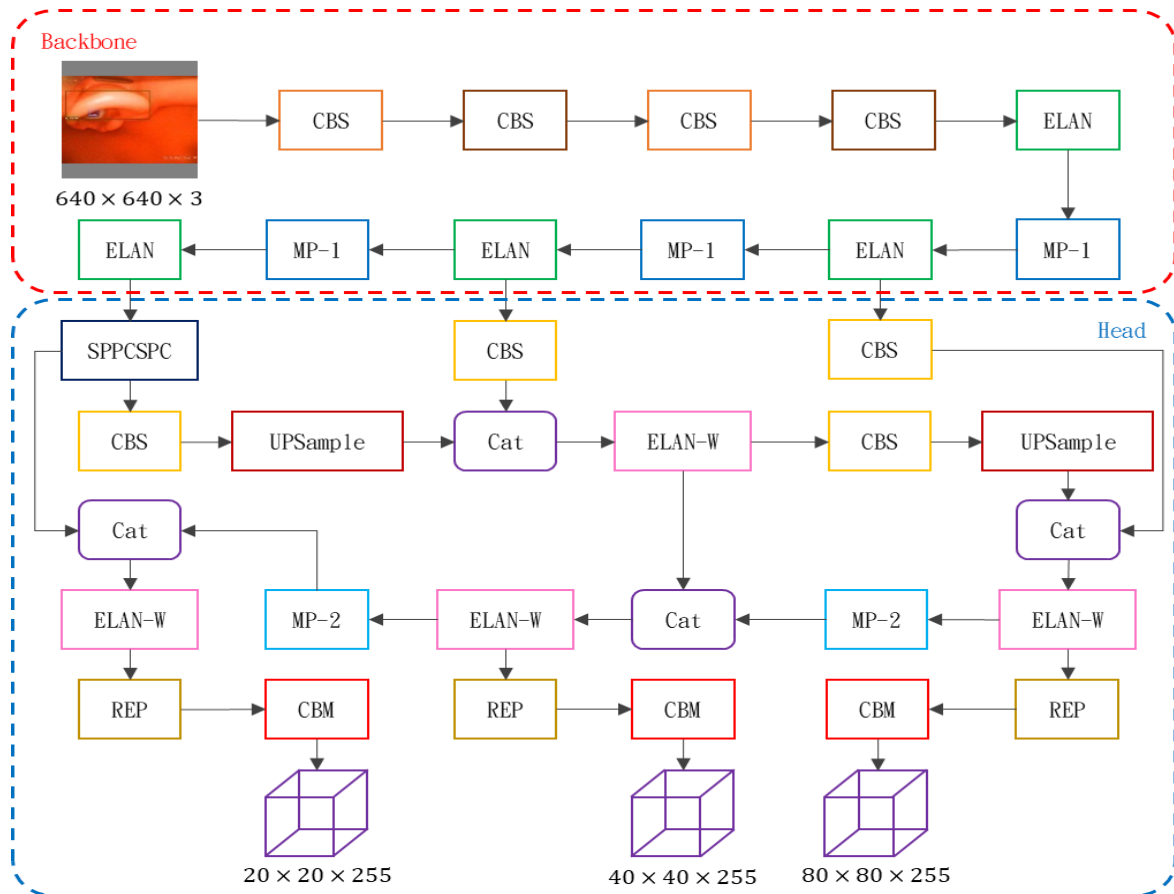


圖 4 模型網路架構圖

YOLOv7 [27] 中 Concatenation (CAT) 用於串接不同層的特徵，在 FPN/PAN 結構中促進多尺度訊息融合。Convolution 層、BatchNorm 層及 Mish 層 (CBM) 為 YOLOv7 backbone 的主要卷積結構，採用 Mish 作為激活函數以提升特徵表達能力。Convolution 層、BatchNorm 層及 SiLU 層 (CBS) 用於更改通道數、特徵提取與 stride=2 的影像縮小，其中 SiLU 為 YOLOv7 head 的主要激活函數。YOLOv7 中的 MaxPooling (MP) 具有兩種形式包含 MP-1 結合 MaxPool 與卷積；MP-2 使用 stride=2 的降採樣卷積。兩者因使用的卷積參數不同，使得影像縮小比例略有差異。Re-parameterization (REP) 模組在訓練與推論階段使用不同結構。訓練時包含特徵提取、平滑特徵與參數增強分支；推論時則將多分支合併為單一卷積以提升速度。Efficient Layer Aggregation Network (ELAN) 與 ELAN-W 都透過控制梯度路徑來獲取最多特徵。兩者差異在於卷積分支數不同，ELAN-W 採用更寬的特徵路徑。Upsample 用於特徵圖放大以利高低層特徵融合；SPPCSPC 則透過空間金字塔池化與跨階段部分連接，使不同尺度的特徵皆能整合進模型。相較於 YOLOv5 與 YOLOv8，YOLOv7 於網路設計上引入 ELAN 結構與 re-parameterization 技術，以提升梯度流穩定性及特徵表達能力。此外，其高效率的多尺度特徵聚合機制使模型在相同運算成本下維持更佳的準確率與推論速度。多項基準測試結果亦顯示，YOLOv7 在偵測效能與效能-運算量平衡方面均優於 YOLOv5 與 YOLOv8，展現其技術優勢。因此基於 YOLOv7 之優點，採用 640×640 的影像大小進行氣管插管的關鍵目標模型訓練。

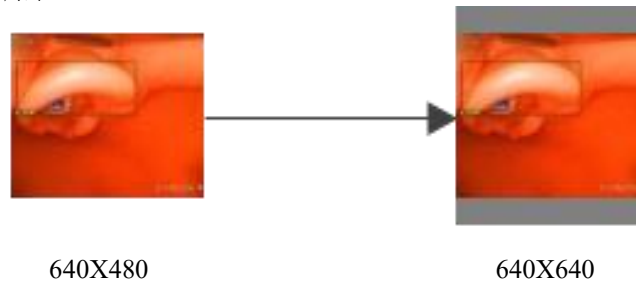


圖 5 YOLOv7 影像縮放示意圖

肆、實驗結果

本研究之實驗流程如圖 6 所示。資料庫首先依比例切分為訓練集與測試集，YOLOv7 於訓練集進行模型學習以建立氣管插管關鍵目標偵測模型，並將其應用於測試集進行推論與辨識。訓練設定方面，epoch 為 300、batch size 設定為 16，優化器採用 SGD，初始學習率為 0.01，推論階段之非極大值抑制 IoU 閾值設定為 0.45。實驗環境採用 NVIDIA GeForce RTX 2060 GPU，並以 YOLOv7 作為目標檢測模型，搭配 CUDA 11.1 與 PyTorch 1.8.1 進行模型訓練與推論。為評估模型之泛化能力，本研究採用留一交叉驗證策略。效能指標則包含 Precision、Recall 與 F1-score，以量化模型於氣管插管關鍵目標偵測任務中的準確性與穩定性。

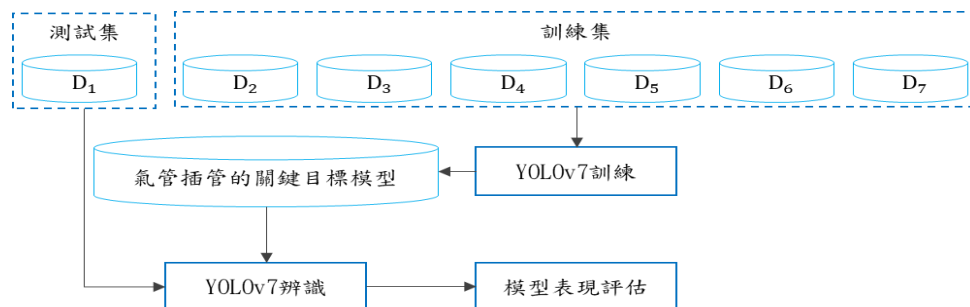


圖 6 模型評估流程

一、氣管插管的關鍵目標模型實驗結果

表 3 為氣管插管的關鍵目標辨識結果，表為各個醫師進行留一交叉驗證後的 Precision、Recall 及 F1-score，其中 D₁、D₂、D₃ 與 D₅ 辨識結果，3 個評估指標皆高達 85% 以上，D₄ 與 D₆ 表現較一般，D₇ 表現較差，整體 Precision、Recall、F1-score 平均與標準差為 82.98%±11.49%、72.94%±18.53%、77.64%±15.32%。表中數據顯示 D₇ 做為測試集時的關鍵目標辨識率相較其他醫師做為測試集時有較大差異，推測因為 D₇ 在氣管插管過程未開啟閃光燈，導致如圖 7 中影像的亮度較為不足的情況，因此在影像中的關鍵目標與其他醫師的影像差異甚大，其影像變異度太大導致的結果。

表 3 氣管插管的關鍵目標辨識實驗結果(7 位醫生)

醫師編號	Precision	Recall	F1-score
D ₁	93.30%	88.19%	90.67%
D ₂	88.99%	86.92%	87.94%
D ₃	86.20%	95.93%	90.81%
D ₄	85.98%	76.97%	81.23%
D ₅	96.80%	92.70%	94.71%
D ₆	71.92%	84.13%	77.55%
D ₇	61.75%	37.02%	46.29%
平均	82.98%	72.94%	77.64%
標準差	11.49%	18.53%	15.32%

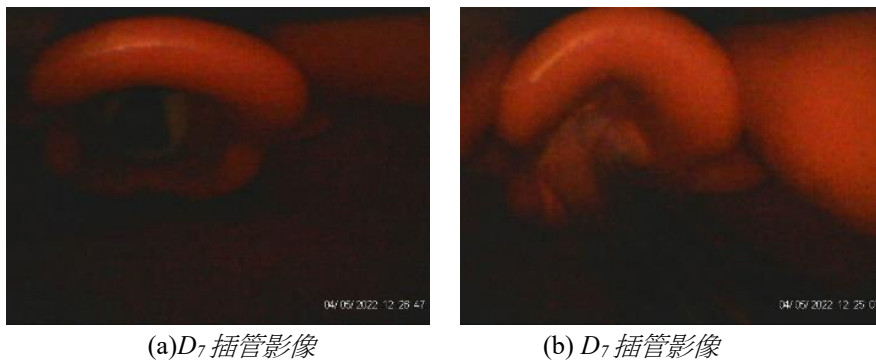


圖 7 D₇ 插管影像

因 D₇ 影像與其他醫師影像差異過大，因此屏除 D₇ 的影像並重新訓練氣管插管的關鍵目標辨識模型，在不考慮 D₇ 影像後的模型表現如表 4 所示，發現屏除 D₇ 資料後訓練的模型 Precision、Recall、F1-score 平均與標準差為 86.92%±3.76%、89.04%±5.95%、87.97%±4.34%。其中 D₆ 的 Precision 明顯提升，而 D₁、D₂、D₄ 及 D₅ 的 Precision 有些微降低的狀況，推測是因 D₇ 的關鍵目標數量在資料集中佔比高達 29.44%，導致剔除 D₇ 資料後提供給模型的訓練樣本不足，使得模型的 Precision 表現出現下降的狀況。而 D₆ 的 Precision、Recall 及 F1-score 提升達 18.39%、4.91%及 10.08%，推測 D₆ 的表現大幅提升是因為 D₆ 的影像亮度較其他醫師有較多的影像亮度偏高如圖 8(a) 與圖 8(b) 的差異，而 D₇ 影像亮度過低導致所訓練的模型無法正確辨識亮度較高的影像，也使得剔除 D₇ 影像後所訓練的模型有更好的表現。

表 4 氣管插管的關鍵目標辨識實驗結果(6 位醫生)

醫師編號	Precision	Recall	F1-score
D ₁	88.37%	94.34%	91.26%
D ₂	80.86%	86.29%	83.49%
D ₃	87.16%	91.99%	89.51%
D ₄	83.81%	78.08%	80.84%
D ₅	91.89%	94.94%	93.39%
D ₆	90.31%	85.10%	87.63%
平均	86.92%	89.04%	87.97%
標準差	3.76%	5.95%	4.34%



(a) D₆ 插管影像

(b) D₃ 插管影像

圖 8 插管影像

本論文初步完成關鍵目標辨識後以時序圖呈現如圖 9，由時序圖能發現初始影像容易發生如圖 10(a) 的誤辨情況，另外在插管完成並退出時會出現管身的誤辨如圖 10(b)。因此將所有未被標記關鍵目標的影像檔皆標上 nonTarget 並重新進行訓練，資料集將屏除 D₇ 的影像資料。表 5 為重新訓練後的模型表現，能發現整體 Precision、Recall 及 F1-score 為 92.19%±3.88%、92.16%±2.31%、92.17%±3.01%，相較未標記 nonTarget 的模型表現更好。

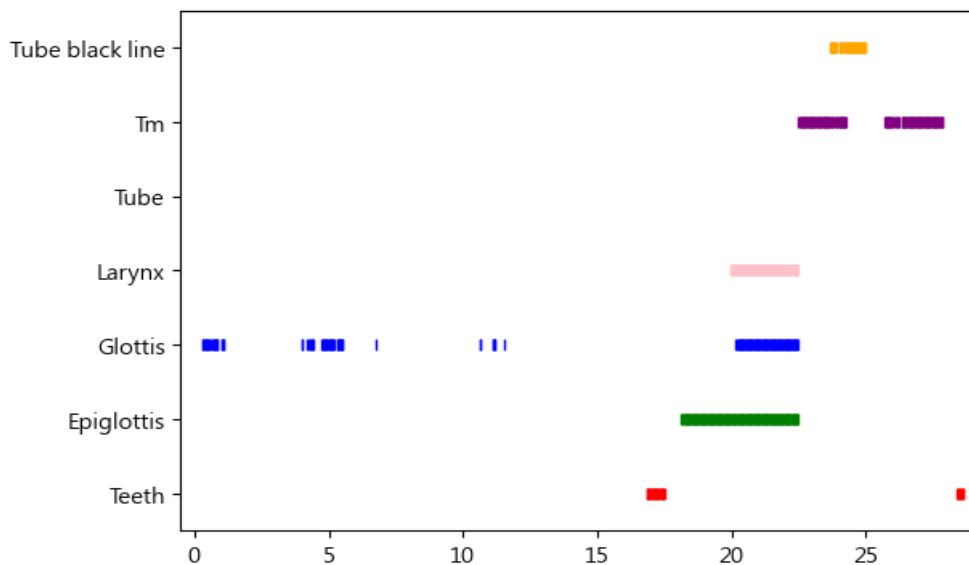


圖 9 氣管插管的關鍵目標辨識時序圖



(a) 誤辨影像 1

(b) 誤辨影像 2

圖 10 誤辨影像

表 5 包含 nonTarget 的模型表現

醫師編號	Precision	Recall	F1-score
D ₁	94.48%	94.06%	94.27%
D ₂	96.71%	93.03%	94.83%
D ₃	88.16%	90.11%	89.12%
D ₄	86.86%	88.14%	87.50%
D ₅	95.57%	94.87%	95.22%
D ₆	89.12%	91.89%	90.48%
平均	92.19%	92.16%	92.17%
標準差	3.88%	2.31%	3.01%

但表 5 中包含 nonTarget 的辨識表現，因此將 nonTarget 數據剔除後重新評估模型表現，其模型表現如表 6 所示，由表中數據能發現相較於沒有標記 nonTarget 的模型，Precision 及 F1-score 平均有成長，但 Recall 出現些微的下降，其主要原因是因為所有非氣管插管的關鍵目標皆屬於 nonTarget，而部分影像如插管完成時退出所拍攝的管身也屬於 nonTarget，因此導致管身與 nonTarget 出現混淆，使得 Recall 相較沒有標記 nonTarget 的模型更低。

表 6 扣除 nonTarget 的模型表現

醫師編號	Precision	Recall	F1-score
D ₁	88.93%	90.93%	89.92%
D ₂	93.01%	91.76%	92.38%
D ₃	89.99%	86.45%	88.18%
D ₄	84.88%	80.91%	82.85%
D ₅	97.32%	93.37%	95.30%
D ₆	94.48%	85.10%	89.55%
平均	90.82%	88.24%	89.51%
標準差	4.04%	4.33%	3.83%

伍、結論

本論文在實驗中將原始訓練集分為三種進行模型訓練，在最開始使用完整資料集進行訓練產生的模型能發現，當插管過程未開啟閃光燈使影像過暗時，其在被有開啟閃光燈的影像所訓練之模型辨識時，

可能會出現較差的表現，如 D₇ 影像即為插管過程中未開啟閃光燈的案例。因此後續將 D₇ 影像屏除後再次訓練，便能使模型的表現獲得提升。後續也發現模型會在插管過程中出現許多誤辨，因此將未被標記關鍵目標的影像標記 nonTarget 並重新進行模型訓練，並且透過模型表現的數據能發現標記 nonTarget 所訓練的模型能對插管過程出現的關鍵目標有更好的辨識，但因為在插管過程中所出現的管身與插管完成後退出攝影機時所出現的管身在影像上的差異較小，因此當退出時所出現的管身被標記上 nonTarget 後，會使模型出現混淆的狀況，並使模型的 Recall 表現出現下降，本論文認為若是後續研究能採用連續影像的方式進行分類，便能透過影像出現的順序辨別管身是否在退出攝影機時出現。本研究所使用之資料量有限，且影像來源受限，使模型於多變臨床環境中的泛化能力可能受到影響。未來將實際收集更多真人臨床影像，以涵蓋不同患者特徵、拍攝角度與亮度條件，增加資料多樣性。此外，將進一步導入資料擴充技術，如亮度調整、旋轉、尺度變化與對比度增強，以模擬多種環境情境。透過擴大資料集與強化資料多樣性，可提升模型在氣管插管影像辨識上的準確率與穩健性，增進其臨床應用價值。

參考文獻

- [1] H. E. Wang, J. R. Lave, C. A. Sirio, and D. M. Yealy, "Paramedic intubation errors: Isolated events or symptoms of larger problems?," *Health Affairs*, vol. 25, no. 2, pp. 501–509, Mar. 2006.
- [2] H. E. Wang, S. R. Seitz, D. Hostler, and D. M. Yealy, "Defining the learning curve for paramedic student endotracheal intubation," *Prehospital Emergency Care*, vol. 9, no. 2, pp. 156–162, Dec. 2005.
- [3] B. D. Johnston, S. R. Seitz, and H. E. Wang, "Limited opportunities for paramedic student endotracheal intubation training in the operating room," *Academic Emergency Medicine*, vol. 13, no. 10, pp. 1051–1055, Oct. 2006.
- [4] P. G. Tarasi, M. P. Mangione, S. S. Singhal, and H. E. Wang, "Endotracheal intubation skill acquisition by medical students," *Medical Education Online*, vol. 16, Aug. 2011.
- [5] J. N. Carlson et al., "Assessment of movement patterns during intubation between novice and experienced providers using mobile sensors: A preliminary proof-of-concept study," *BioMed Research International*, vol. 2015, Jun. 2015.
- [6] C. Lim et al., "Development of a hand motion-based assessment system for endotracheal intubation training," *Journal of Medical Systems*, vol. 45, no. 8, Aug. 2021.
- [7] C. Lim et al., "Multi-sensor feature integration for assessment of endotracheal intubation," *Journal of Medical and Biological Engineering*, vol. 40, no. 5, pp. 1–?, Jun. 2020.
- [8] C. Matava, "A convolutional neural network for real-time classification, identification, and labeling of vocal cord and trachea using laryngoscopy and bronchoscopy video," *Journal of Medical Systems*, vol. 44, no. 2, Dec. 2019.
- [9] M. Pedersen et al., "Localization and quantification of glottal gaps on deep learning segmentation of vocal folds," *Scientific Reports*, vol. 13, no. 1, Jan. 2023.
- [10] P. Yao et al., "Applications of artificial intelligence to office laryngoscopy: A scoping review," *The Laryngoscope*, vol. 132, no. 10, Sep. 2021.
- [11] C. F. Larsen and M. Pedersen, "Comparison of convolutional neural networks for classification of vocal fold nodules from high-speed video images," *European Archives of Oto-Rhino-Laryngology*, Nov. 2022.
- [12] T. Lu and J. Ren, "Automatic recognition of laryngoscopic images using a deep-learning technique," *The Laryngoscope*, vol. 130, no. 11, Jan. 2020.

- [13] Y. Li et al., “Development and validation of the artificial intelligence (AI)-based diagnostic model for bronchial lumen identification,” *Translational Lung Cancer Research*, vol. 11, no. 11, Jan. 2021.
- [14] P. Yan et al., “Automated detection of glottic laryngeal carcinoma in laryngoscopic images from a multicenter database using a convolutional neural network,” *Clinical Otolaryngology*, Jan. 2023.
- [15] Y.-P. Lin et al., “EEG-based emotion recognition in music listening,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.
- [16] S. Kiranyaz, T. Ince, and M. Gabbouj, “Real-time patient-specific ECG classification by 1-D convolutional neural networks,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, Mar. 2016.
- [17] M. U. Zahid, S. Kiranyaz, and M. Gabbouj, “Global ECG classification by self-operational neural networks with feature injection,” *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 1, pp. 205–215, Jan. 2023.
- [18] G. Hoyer et al., “Advanced video laryngoscope and automatic data collection system,” *Biomedical Engineering and Bioengineering Commons*, May 2020.
- [19] B. Wu, C. Pang, X. Zeng, and X. Hu, “ME-YOLO: Improved YOLOv5 for detecting medical personal protective equipment,” *Applied Sciences*, vol. 12, no. 23, Nov. 2022.
- [20] K. Liu, “STBi-YOLO: A real-time object detection method for lung nodule recognition,” *IEEE Access*, vol. 10, pp. 75385–75394, Jan. 2022.
- [21] M. Widiyasri et al., “Dental-YOLO: Alveolar bone and mandibular canal detection on cone beam computed tomography images for dental implant planning,” *IEEE Access*, vol. 10, pp. 101483–101494, 2022.
- [22] X. Su et al., “Face mask detection and classification via deep transfer learning,” *Multimedia Tools and Applications*, vol. 81, no. 36, pp. 1–20, 2022.
- [23] S. Degadwala et al., “YOLO-v4 deep learning model for medical face mask detection,” in *Proc. Int. Conf. Artificial Intelligence and Smart Systems*, Mar. 2021.
- [24] Z. Zhao et al., “PIS-YOLO: Real-time detection for medical mask specification in an edge device,” *Computational Intelligence and Neuroscience*, vol. 2022, Nov. 2022.
- [25] I. Javed et al., “Face mask detection and social distance monitoring system for COVID-19 pandemic,” *Multimedia Tools and Applications*, Sep. 2022.
- [26] M. A. Azam et al., “Deep learning applied to white light and narrow band imaging videolaryngoscopy: Toward real-time laryngeal cancer detection,” *The Laryngoscope*, vol. 132, no. 9, Nov. 2021.
- [27] 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. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 7464–7475.