

運用 YOLOv7 辨識口罩配戴狀況

郭智榮

長榮大學 資訊工程學系 3B

Email: 109B30612@mailst.cjcu.edu.tw

摘要

嚴重特殊傳染性肺炎 (Coronavirus disease 2019, COVID-19) 爆發以來已經過 2 年多的時間，各國開始鬆綁及開放各項防疫規定的同時，仍有許多小範圍的集體染疫事件正在發生。因此本報告將透過訓練口罩配戴辨識模型搭配現有體溫量測系統，期許透過結合口罩配戴辨識模型後，能增強現有系統的功能。

1. 研究動機與目的

COVID-19 疫情出現至今已長達 2 年，如今由於症狀減輕、中重症比例降低和經濟發展考量等原因，各國皆已陸續地放寬確診者、密切接觸者相關的防疫規定和入境隔離天數等。當社會正在逐步走向疫情前的步調時，仍然有許多地區正在爆發小範圍的集體染疫，因此防疫的強度仍然需要透過其他方式補足，避免一個疏漏導致原本可控的情況逐漸延變為失控。

1.1 研究動機

長榮大學在 2021 年時已開發出可協助教職員及師生進行體溫量測的系統[1]，透過該系統對全校師生進行體溫控管，也可幫助已發燒而自身沒有發覺的同學即早進行就醫，但現階段的體溫量測並不會對未配戴口罩或不正確配戴口罩的同學進行提醒。

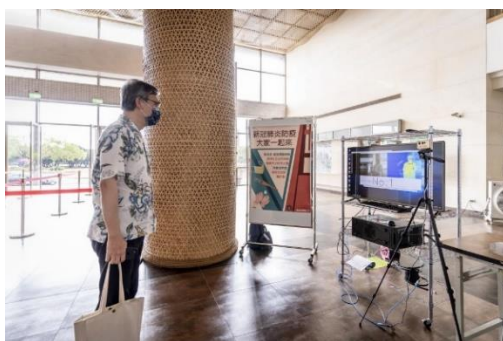


圖 1. 長榮大學校長進行體溫量測

1.2 研究目的

本報告將透過 YOLOv7[2] 進行模型訓練，並觀察在不同的訓練參數下進行訓練對模型辨識表現差異，透過調整參數的訓練及 outside 的測試，藉此找到能夠在辨識率保有一定水準且不會產生過擬合(Overfitting)的模型。

本報告希望能透過訓練口罩配戴狀況辨識模型與原有系統結合，進而在量測體溫的同時能協助使用者檢測口罩配戴狀況是否正常；因此本報告將針對口罩配戴狀況進行初步針對是否配戴口罩的模型進行訓練及測試，以利日後若有需求則可以銜接本報告之結論進行更深入的研究。

2. 文獻探討

本章節將針對 YOLOv7 進行相關文獻的探討，並延伸探討目前 YOLO 模型在針對配戴口罩進行辨識的相關文獻。

2.1 YOLOv7

YOLOv7 是由台灣中研院的王建堯博士和廖弘源所長以及來自俄羅斯的 Alexey Bochkovskiy 共同推出地最新版 YOLO，三位作者同時也是 YOLOv4[3] 的發布者。

YOLOv7 不僅在準確度上面有所提升，速度方面也相較 YOLOv5 提高了約 120%，並且 YOLOv7 透過優化模型架構及訓練過程，降低了約 4 成的參數量及 5 成的運算量。

YOLOv7 由 VoVNet[4] 衍生出 CSPVoVNet 架構，如圖 2 所示，除了 VoVNet 本身會考量參數量、運算量等，CSPVoVNet 會進一步分析梯度路徑讓不同層的權重能透過更多樣化的特徵提升準確率。

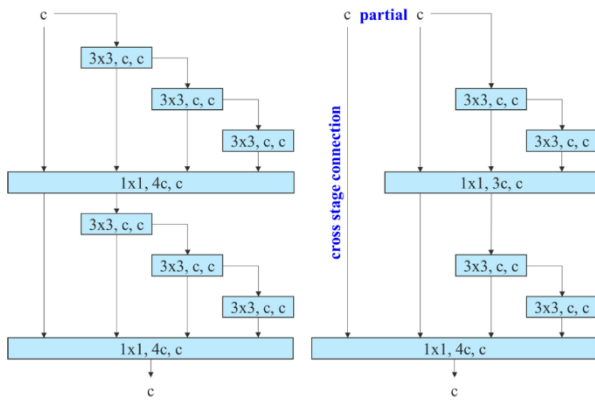


圖 2. VoVNet(左)與 CSPVoVNet(右)架構圖

ELAN 架構藉由控制最短與最長梯度路徑達到讓更深層的網路能有效地學習和收斂，而目前的 ELAN 架構中的計算模塊(Computational Blocks)堆疊數量已達穩定，強行疊加模塊可能產生反效果。因此 YOLOv7 提出了 E-ELAN(Extended-ELAN)架構，如圖 3 所示，透過 expand, shuffle, merge cardinality 等方式避免梯度路徑被破壞且能夠持續加強模型的學習能力。

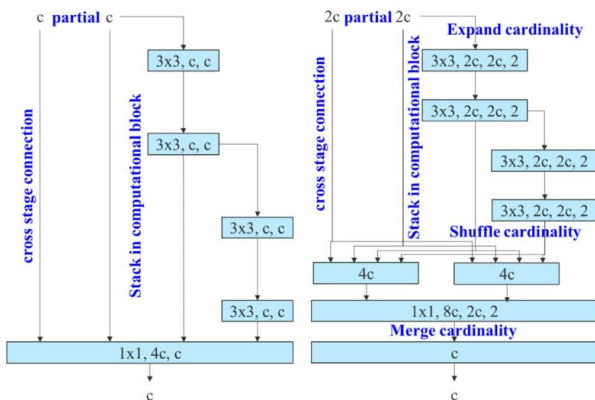


圖 3. ELAN(左)與 E-ELAN(右)架構圖

2.2 口罩配戴狀況辨識

由 Peishu Wu, Han Li, Nianyin Zeng 和 Fengping Li 所提出的 FMD-YOLO[5] 在一份含有「未配戴口罩」及「配戴口罩」的資料集中與 YOLOv3, YOLOv4 及其他 6 個物件辨識的演算法進行比較，FMD-YOLO 在 mAP, AP50, AP75, mAR, AR50 五項數據中都獲得極好的表現，五項數據依序獲得 4.6%、1.8%、5.5%、10% 及 2.3% 的提升。從數據中能發現 FMD-YOLO 在 AP75 的提升相較 AP50 更多，其原因是 FMD-YOLO 的訓練採用了 IoU-aware，使得 FMD-YOLO 在辨識上能夠更精準。

3. 研究架構

3.1 資料集說明

本報告資料來源為 Kaggle 的「[Face Mask Dataset \(YOLO Format\)](#)」，其原始資料集(dataset)共 920 張影像分為訓練集(training set)700 張、測試集(testset)120 張、驗證集(validation set)100 張，分類類別為配戴口罩的「mask」及未配戴口罩的「no_mask」。

3.2 實驗流程

本報告將透過 YOLOv7 訓練模型並將原始資料及合併為一個資料集，後續將資料集再另外分成 5 等份進行五折交叉驗證(Five-Fold Cross Validation)，藉此以達到較精確的模型評估，資料集重新分配後其影像及物件數量如表 1 所示。

表 1. 每份資料的影像及物件數量

k_n	影像數量	未配戴口罩	已配戴口罩
k_1	184	240	829
k_2	184	177	777
k_3	184	219	689
k_4	184	240	707
k_5	184	197	826
總和	920	1073	3828

實驗流程如圖 4 所示，且每次訓練將分為 50epoch、100epoch、150epoch 的方式訓練。

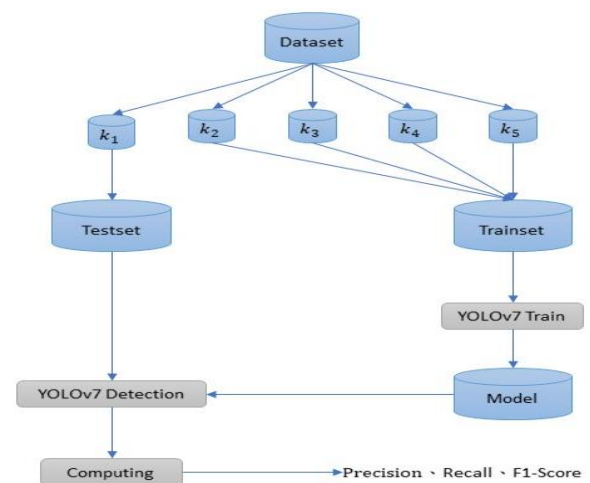


圖 4. 一輪訓練的實驗流程示意圖

4. 實驗結果

本報告採用精確率(Precision)、召回率(Recall)、F1-score 三個常見的評估指標進行效能的評估，其可透過辨識後得出的結果與原始資料進行比對得出的真陽性(True Positive, TP)、偽陽性(False Positive, FP)、偽陰性(False Negative, FN)進而計算後得出。

TP 代表在原始資料被標註，且模型也同時有標註之資料；FP 代表原始資料中並未標註，但模型有標註之資料；FN 代表原始資料中有標註，但模型並未標註之資料。Precision 代表模型辨識物件中正確的比例；Recall 代表原始資料標註的物件被模型辨識的比例。

4.1 預設的信心門檻進行辨識的表現

本報告首先採用 YOLOv7 的信心門檻(Confidence Thres)預設值 0.25 進行檢測，五份資料在 50、100 及 150epochs 下訓練後辨識表現如表 2 至表 6 所示。

由統計結果可以發現，每份資料集間的結果差異大多落在 10%之內，考量本報告所使用資料的物件數較少，因此浮動值的確有可能較大。而較值得關注的是每份資料的 Precision 表現皆達到 70%以上，但 Recall 表現卻皆落在 14~21%左右，本報告認為此狀況可能是因為模型所使用的訓練集樣本數不足導致模型無法考量到許多情況。

表 2. 第一份資料為測試集時的結果(Confidence 為 0.25)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	79.12%	20.13%	32.09%	83.66%	20.04%	32.33%	78.75%	21.06%	33.23%
Mask	92.23%	21.08%	34.32%	93.91%	20.95%	34.26%	92.00%	21.06%	34.27%

表 3. 第二份資料為測試集時的結果(Confidence 為 0.25)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	77.66%	14.17%	23.97%	76.10%	14.54%	24.42%	81.82%	14.26%	24.29%
Mask	86.33%	19.46%	31.76%	87.90%	19.36%	31.73%	88.96%	19.15%	31.52%

表 4. 第三份資料為測試集時的結果(Confidence 為 0.25)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	80.65%	18.64%	30.28%	77.27%	19.01%	30.51%	79.84%	19.11%	30.84%
Mask	91.85%	17.35%	29.19%	93.10%	17.27%	29.14%	93.87%	17.19%	29.06%

表 5. 第四份資料為測試集時的結果(Confidence 為 0.25)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	81.65%	20.32%	32.54%	82.56%	19.85%	32.00%	85.83%	19.66%	31.99%
Mask	86.32%	17.97%	29.75%	87.29%	17.95%	29.78%	90.48%	17.63%	29.51%

表 6. 第五份資料為測試集時的結果(Confidence 為 0.25)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	82.06%	17.05%	28.23%	88.24%	16.87%	28.32%	85.99%	16.68%	27.94%
Mask	87.39%	20.64%	33.39%	89.81%	20.72%	33.67%	90.83%	20.69%	33.70%

4.2 較低的信心門檻進行辨識的表現

本報告依據 4.1 節所得之結果進行延伸，測試信心門檻降低的情況下是否能使 Recall 上升，為避免信心門檻下降過多導致數據變化過大或下降過少導致數據變化不明顯之情況，本小節採信心門檻 0.1 進行辨識，並得出模型表現如表 7 至表 11 所示。

4.2.1 較低的信心門檻與預設之比較

表 12 所示之數據為第一份資料在 Confidence 為 0.1 時相較於 0.25 的表現變化，可觀察出 Precision 的表現因為信心門檻的降低導致下降許多，而 Recall 的表現提升相較 Precision 的變化而言相對較少許多，因此信心門檻的降低對本報告之模型的表現並不理想。

表 7. 第一份資料為測試集時的結果(Confidence 為 0.1)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	65.49%	20.69%	31.45%	77.27%	20.60%	32.53%	73.25%	21.44%	33.17%
Mask	84.98%	21.32%	34.09%	90.32%	21.21%	34.35%	85.56%	21.32%	34.13%

表 8. 第二份資料為測試集時的結果(Confidence 為 0.1)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	66.80%	15.10%	24.63%	67.36%	15.19%	24.79%	74.30%	14.82%	24.71%
Mask	81.18%	19.75%	31.77%	83.67%	19.67%	31.85%	86.32%	19.28%	31.52%

表 9. 第三份資料為測試集時的結果(Confidence 為 0.1)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	71.38%	19.48%	30.61%	71.04%	19.66%	30.80%	75.70%	19.76%	31.34%
Mask	85.92%	17.50%	29.08%	89.13%	17.35%	29.05%	90.14%	17.42%	29.20%

表 10. 第四份資料為測試集時的結果(Confidence 為 0.1)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	68.88%	21.06%	32.26%	75.60%	20.50%	32.25%	78.72%	20.60%	32.65%
Mask	78.59%	18.18%	29.53%	82.68%	18.08%	29.67%	86.11%	17.79%	29.49%

表 11. 第五份資料為測試集時的結果(Confidence 為 0.1)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	70.57%	17.43%	27.96%	78.30%	17.15%	28.14%	80.53%	17.05%	28.14%
Mask	80.16%	21.06%	33.36%	85.52%	20.98%	33.69%	88.34%	20.95%	33.87%

表 12. 第一份資料為測試集時在 Confidence 為 0.1 時的表現相較於 0.25 的表現

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	-13.63%	0.56%	-0.64%	-6.39%	0.56%	0.20%	-5.50%	0.38%	-0.06%
Mask	-7.25%	0.24%	-0.23%	-3.59%	0.26%	0.09%	-6.44%	0.26%	-0.14%

4.3 較高的信心門檻進行辨識的表現

本報告同樣讓模型在較高的信心門檻下進行物件的辨識，並從中觀察信心門檻提高對於辨識結果是否有更好的提升，模型在信心門檻 0.5 時的表現如表 13 至表 17 所示；而信心門檻提高至 0.75 時的表現則如表 19 至表 23 所示。

4.3.1 較高的信心門檻與預設之比較

由表 18 可得知信心門檻為 0.5 時辨識表現的 Precision 在未配戴口罩(No mask)的辨識有明顯提升，且 Recall 下降的幅度並不大。如表 24 及表 25 所示，若信心門檻提升至 0.75 時，與 0.5 的 Precision 和 Recall 變化幅度較為接近且與預設相比 Recall 有明顯下降。

表 13. 第一份資料為測試集時的結果(Confidence 為 0.5)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	89.81%	17.24%	28.93%	92.34%	19.11%	31.67%	89.32%	19.48%	31.98%
Mask	96.03%	20.25%	33.45%	96.91%	20.45%	33.77%	96.62%	20.19%	33.40%

表 14. 第二份資料為測試集時的結果(Confidence 為 0.5)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	89.68%	12.95%	22.63%	87.20%	13.33%	23.12%	85.14%	13.89%	23.88%
Mask	91.62%	18.86%	31.28%	92.71%	18.94%	31.45%	92.55%	18.83%	31.29%

表 15. 第三份資料為測試集時的結果(Confidence 為 0.5)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	93.78%	16.87%	28.60%	85.09%	18.08%	29.82%	84.87%	18.73%	30.69%
Mask	95.41%	16.82%	28.60%	96.14%	16.93%	28.79%	95.44%	16.95%	28.79%

表 16. 第四份資料為測試集時的結果(Confidence 為 0.5)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	91.54%	17.15%	28.89%	90.31%	19.11%	31.54%	89.91%	19.11%	31.52%
Mask	93.44%	17.48%	29.45%	91.28%	17.50%	29.37%	93.65%	17.35%	29.28%

表 17. 第五份資料為測試集時的結果(Confidence 為 0.5)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	94.41%	15.84%	27.13%	91.44%	16.03%	27.28%	91.71%	16.59%	28.10%
Mask	93.77%	20.04%	33.02%	93.86%	20.38%	33.49%	94.08%	20.35%	33.46%

表 18. 第一份資料為測試集時在 Confidence 為 0.5 時的表現相較於 0.25 的表現

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	10.69%	-2.89%	-3.16%	8.68%	-0.93%	-0.66%	10.57%	-1.58%	-1.25%
Mask	3.80%	-0.83%	-0.87%	3.00%	-0.50%	-0.49%	4.62%	-0.87%	-0.87%

表 19. 第一份資料為測試集時的結果(Confidence 為 0.75)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	99.19%	11.46%	20.55%	97.65%	15.47%	26.71%	96.59%	15.84%	27.22%
Mask	99.08%	16.80%	28.73%	99.04%	18.78%	31.57%	99.15%	18.23%	30.80%

表 20. 第二份資料為測試集時的結果(Confidence 為 0.75)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	100.00%	5.96%	11.25%	98.89%	8.29%	15.30%	92.31%	12.30%	21.71%
Mask	97.02%	15.31%	26.45%	96.64%	16.51%	28.20%	95.20%	18.13%	30.46%

表 21. 第三份資料為測試集時的結果(Confidence 為 0.75)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	96.30%	7.27%	13.52%	94.41%	15.75%	27.00%	91.37%	16.78%	28.35%
Mask	98.17%	14.03%	24.55%	98.69%	15.75%	27.16%	96.59%	16.27%	27.85%

表 22. 第四份資料為測試集時的結果(Confidence 為 0.75)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	98.96%	8.85%	16.25%	95.08%	16.22%	27.71%	94.03%	17.61%	29.66%
Mask	97.39%	14.63%	25.44%	95.65%	16.67%	28.39%	96.84%	16.82%	28.66%

表 23. 第五份資料為測試集時的結果(Confidence 為 0.75)

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	97.75%	8.11%	14.98%	96.05%	13.70%	23.98%	94.12%	15.00%	25.88%
Mask	97.48%	16.14%	27.69%	96.97%	19.25%	32.12%	96.66%	19.64%	32.65%

表 24. 第一份資料為測試集時在 Confidence 為 0.75 時的表現相較於 0.25 的表現

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	20.07%	-8.67%	-11.54%	13.99%	-4.57%	-5.62%	17.84%	-5.22%	-6.01%
Mask	6.85%	-4.28%	-5.59%	5.13%	-2.17%	-2.69%	7.15%	-2.83%	-3.47%

表 25. 第一份資料為測試集時在 Confidence 為 0.75 時的表現相較於 0.5 的表現

k_n	50epoch			100epoch			150epoch		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
No mask	9.38%	-5.78%	-8.38%	5.31%	-3.64%	-4.96%	7.27%	-3.64%	-4.76%
Mask	3.05%	-3.45%	-4.72%	2.13%	-1.67%	-2.20%	2.53%	-1.96%	-2.60%

5. 結論

5.1 在不同 Epochs 下訓練的模型表現差異

由第 4 節之數據可觀察出 50、100 及 150epochs 下訓練的模型表現差異並不大，且部分情況下 100 及 150epochs 所訓練的模型表現反而較差。本報告認為導致此狀況的原因是訓練集的樣本數較少導致模型提早收斂所致，後續持續訓練也造成出現過擬合的狀況使表現有所下降。

5.2 在 Recall 無法提升的改善方式

由第 4.2 與 4.3 節可觀察到不論信心門檻高低，都會出現 Precision 表現在水準之上甚至極好，但 Recall 表現則非常差的情況。而 Precision 高 Recall 低的情況也可以解釋為模型辨識到是否配戴口罩的正確率較高，但對於偵測影像中需要辨識的對象較不靈敏。因此若單純透過模型進行口罩配戴狀況辨識時，模型的表現可能較差，但若搭配其他專門用於人臉辨識的模型先判斷人臉位置，再由模型判定其是否配戴口罩，便能夠有效發揮出 Precision 高的優勢；且當信心門檻調高後所獲得的表現看來，若單純進行是否配戴口罩的辨識而不需要偵測，則模型或許能達到極好的表現。目前的

體溫量測系統的流程會先進行人臉辨識後再進行體溫判定，也剛好能夠符合模型的理想流程。

參考文獻

- [1] 人工智慧研究中心, “智慧科技防疫-AI 中心研發 AI 熱像儀體溫量測系統”, 長榮大學, 2021-05-27.
- [2] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors”, ArXiv, 2022-07-06.
- [3] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection”, ArXiv, 2020-04-23.
- [4] Youngwan Lee, Joong-won Hwang, Sangrok Lee, Yuseok Bae, and Jongyoul Park, “An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection”, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019-04-22.
- [5] Peishu Wu, Han Li, Nianyin Zeng, and Fengping Li, “FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public”, Image and Vision Computing 117(9):104341, 2021-11-19

本報告撰寫過程所設計之程式碼存放於 <https://reurl.cc/kq1xZG>

若該連結失效，請轉往 <https://github.com/Guo0911> 並找尋「CJCU_Course/111-1/人工智慧實務/Midterm_Report」