Face Mask Detection using Transfer Learning of YOLOv3

Cheng-Zhe Wu

ABSTRACT

近期，新冠病毒蔓延全世界，造成了很嚴重的危機，為了防止人與人之間的接觸造成傳染擴大，只要是人群聚集地、室內空間都需要強制戴上口罩，減少病毒傳染的機率，防止疫情的增長，因此許多地方需要檢測人們進出是否佩戴口罩，為了讓檢測更方便快速且準確，越來越多人投入這方面的研究主題。在本實驗中，使用的訓練資料集為 kaggle 中的 Face Mask Detection 資料集，利用 deep learning 來偵測人們有無配戴口罩，使用最先進的檢測模型之一"YOLOv3"，且以 transfer learning 的方式進行 finetune training [1]，最終得到 mAP 為 77.86%。

Keywords: Face Mask, Detection, Deep Learning, YOLO

1. INTRODUCTION

在新冠病毒盛行之前，鮮少人們外出會配戴口罩，不過隨著新冠病毒大肆傳播，世界衛生組織(WHO)開始呼籲民眾外出應配戴口罩，口罩在保護個人免受呼吸道疾病的健康中起著至關重要的作用，這是在沒有對新冠病毒免疫的情況下可用的少數預防措施之一，台灣衛服部也呼籲民眾應盡量配戴口罩，甚至在某些場所是需要強制配戴口罩，像是醫院、公車、賣場等人群眾多或密閉的地方，

人工智慧的技術在近幾年蓬勃發展，機器學習與深度學習被應用在各種領域上，也包含醫療領域，透過這樣的技術，可以用來防止新冠病毒的傳播，建立監測系統，像是人臉口罩偵測，減緩病毒的傳播。

本文使用深度學習的方式來檢測人們有無配戴口罩，使用的模型架構為 YOLOv3，透過攝影機拍攝在公共場所的民眾使否有正確配戴口罩。

2. DATASET

本實驗所使用的資料集來自於 kaggle 的 Face Mask Detection 資料集 [2]，總共有 853 張影像，總共分為 3 類，分別為 with mask、without mask 和 mask weared incorrectly，每張影像都有一個 xml 檔案，裡面描述目標的邊界框，為 PASCAL VOC 格式。

3. RESEARCH METHOD

3.1 Model Architecture

YOLOv3[3] 為最先進也最被廣泛應用的模型之一。在邊界框預測中， 延續 YOLO9000 使用 dimention cluster 作為 anchor box 來預測 bounding box，Bounding box 為 tx、ty、tw、th 四個座標(Fig 1)。訓練中，使用的 loss function 為 mean square error，。YOLOv3 使用 Logistic Regression 來預測每個 bounding box，如果先驗 bounding box 與 ground truth 的重疊大於其他 bounding box，則分數為 1，如果先驗 bounding box 不是最好的，且若無法超過某個設定的閾值，則忽略該 bounding box。

每個 box 使用多標籤分類來預測 bounding box 可能包含的類別，在訓練中，取代 softmax，YOLOv3 只使用獨立的 logistic classifier，loss 則是使用 binary cross entrpy。

YOLOv3 使用了三種不同尺寸的 box 做預測，並且使用類似於 pyramid networks 的概念去提取特徵，在特徵擷取器中加入了許多的捲積層，而最後預測出一個 3d tensor encoding bounding box、objectness 跟類別。接下來使用來自前 2 層的特徵圖，並且上採樣 2 倍，然後還使用了來自更早的特徵圖跟上採樣的特徵作融合，這樣的方式能夠從上採樣的特徵中獲得更有意義的語義信息，並從較早的特徵圖中獲得更細的信息，最後再添加一些捲基層來處理這些組合的特徵來預測相似的 tensor，而現在的尺寸為一開始的 2 倍。最後我們再執行一次一樣的設計來預測 box 跟最後的尺寸，因此，第 3 次的預測受益於先前的計算以及來自網路早期的細粒度特徵。 一樣使用 K means clustering 來定義先驗 bounding box。

YOLOv3 使用一種新的網路來做特徵擷取，這個新網路為 YOLOv2、DarkNet-19 以及 newfangled residual nerwork 的混和，本架構使用了連續的 3x3 跟 1x1 的捲基層和 shortcut conection，而且規模更巨大，他有 53 的捲基層，因此稱為 Darknet-53(Fig 2)。

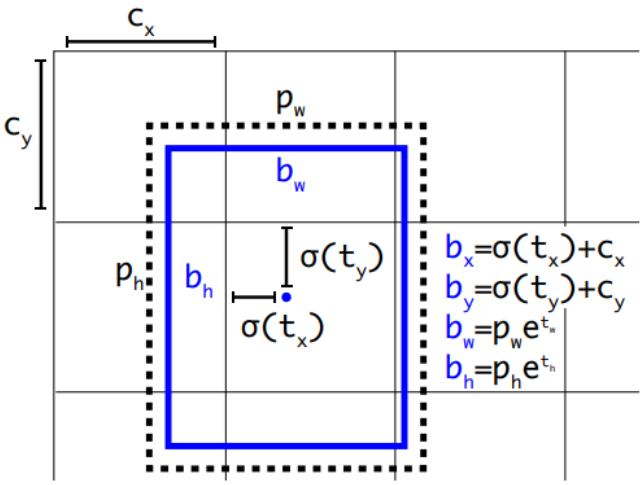

Fig 1: Bounding box 的計算公式與位置預測[3]

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig 2: Darknet-53[3]

3.2 Train

先將資料分為 training set 跟 validation set 以 testing set，比例為 8:1:1，採用 transfer learning 的方法，本實驗使用 darknet53 的 pretrained weights 做 finetune training，我們將訓練拆分兩個階段，第一階段先 freeze YOLOv3 除了輸出層以外網路的來訓練，Batch size 為 32，Epochs 為 50，Optimizer 使用 Adam，Learning rate 為 0.001，Loss function 為 YOLO loss，在這一階段會訓練出一個還不錯的 model，接著第二階段訓練，將整個網路設定成可學習的模式，訓練整個 model，將 initial learning rate 改為 0.0001，batch size 改為 4，epoch2 仍為 50，來訓練出一個更好的 model，第二階段的訓練中，訓練中會監測 validation loss 的數值來對 learning rate 做調整，設定為連續 3 epochs 都沒進步就將 learning rate 乘上 0.1，並且使用 early stopping 機制，設定為連續 10 epochs 都沒再進步就停止訓練，最後停止在 90 epochs，此時 training loss=22.1873、validation loss=25.4915。

4. RESULTS

在 Fig 3 中可以清楚看到前 50 epochs 利用較大的 learning rate，且只訓練輸出層，使 model 較容易訓練，因此很快的下降，收斂在一個不錯的低點，然後在第二階段的訓練中，

為了讓 mode 更接近 global minimum 而調降了 learning rate，且讓 learning rate 在學習受阻的時候下調，使 loss 達到更低。

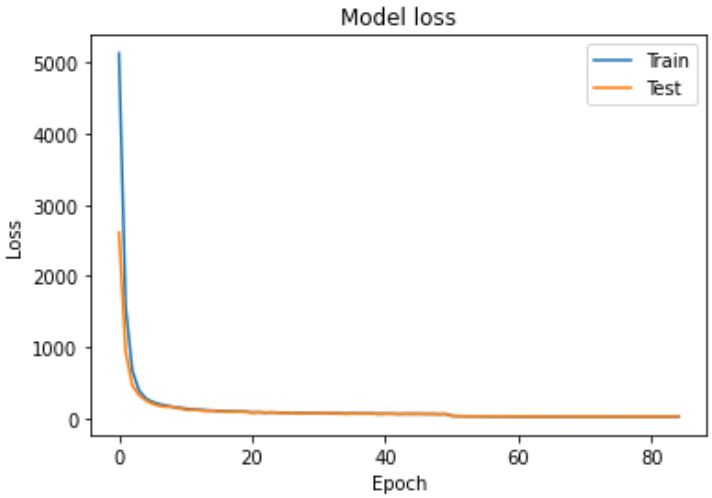

Fig 3: Loss with training and validation

Tabel 1 為評估的結果，mAP 達到 77.89。Fig 4 為預測的結果，透過這些偵測後的影像，可以看到 YOLOv3 準確地分辨人們有沒有戴上口罩，以及判斷戴口罩的方式正不正確。接下來，可以將此研究應用在攝影裝置上，設置在公共場合或室內場所的出入口，使機器成為人們防疫的好幫手。

label	**AP (%)**
With mask	81.43
Without mask	90.44
Mask weared incorrect	71.72
mAP (%)	**77.89**

Table 1: Performance

Input Prediction

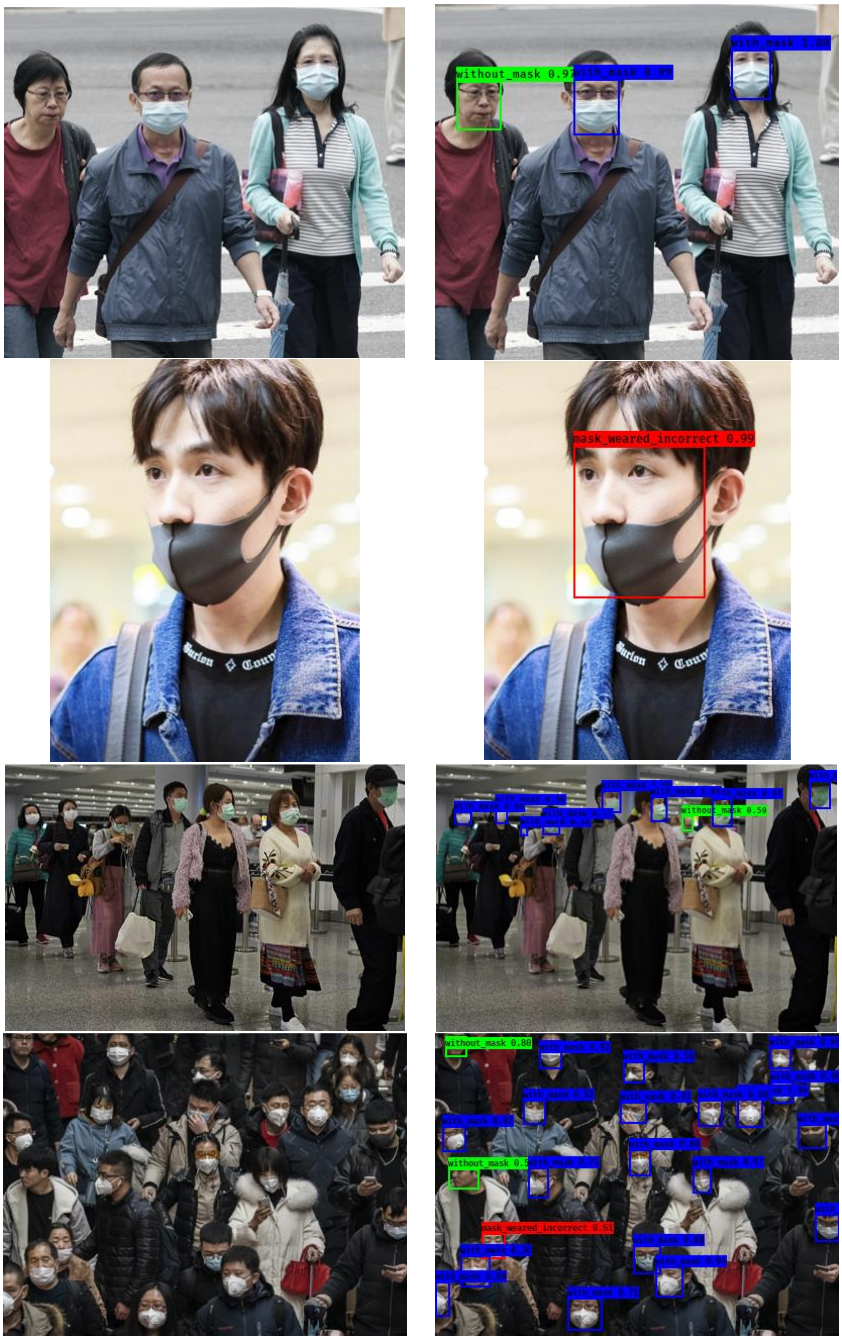

Fig 4: Examples of the prediction

REFERENCES

[1] G. Jignesh Chowdary, Narinder Singh Punn, Sanjay Kumar Sonbhadra, and Sonali Agarwal: Face Mask Detection using Transfer Learning of InceptionV3, 2020

[2] Dataset, https://www.kaggle.com/andrewmvd/face-mask-detection, online accessed Dec 30, 2020

[3] Joseph Redmon, Ali Farhadi: YOLOv3: An Incremental Improvement