

# Face Mask Detection using Deep Learning Method

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，部分特地區域需要檢測人們進出是否佩戴口罩。在本文中，將使用深度學習的方式來偵測人們有無配戴口罩，所使用的模型為最先進的模型之一 YOLOv3，然後進行微調，以適合此偵測任務訓練，使用的訓練資料集為 kaggle 中的 Face Mask Detection 資料集，實驗尚未做完，因此準確度尚未算出來。

**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，總共有 853 張影像，總共分為 3 類，分別為 with mask、without mask、mask worn 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 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 dimension 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 entropy。

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 network 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)。

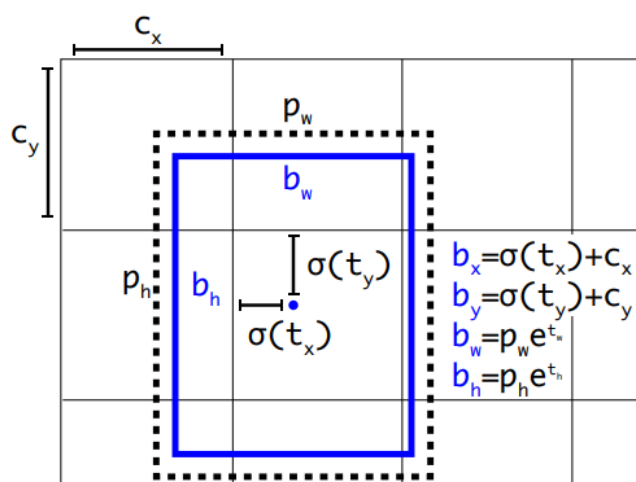


Fig 1: Bounding box 的計算公式與位置預測[3]

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

Fig 2: Darknet-53[3]

### 3.2 Train

先將資料分為訓練集跟驗證集，比例為 9:1，本實驗使用[4]的預訓練 tiny model，我們將訓練拆分兩次，第一次先 freeze YOLOv3 除了輸出以外的 Darknet-53 backbone 來訓練網路，優化器使用 Adam，學習率為 0.001，損失函數為 YOLO loss，batch size 為 3，epoch 為 50，在這會得到一個還不錯的網路，接著第二次訓練，使用所有的層並微調，將學習率改為 0.0001，batch size 改為 4，epoch 改為 100，然後訓練出最終的網路。

## 4. RESULTS

在 loss 的圖中，可以清楚看到前 50 層利用較大的學習率，因此很快的下降，收斂在一個不錯的位置，然後在第二次的訓練中，因為調降了學習率，因此在 epoch 等於 50 左右的地方又往下降了一點，並穩定的收斂。下圖 Fig4 為預測的結果，由上角及左下角的人沒有戴口罩，左上角及右下角的人有戴口罩，YOLOv3 準確地預測出正確的結果。此網路不僅可以用在影像上的偵測，還能夠預測影片中的人有無配戴口罩，做到即時偵測的效果，目前也被應用在許多公共場合，使機器成了人們防疫的好幫手。

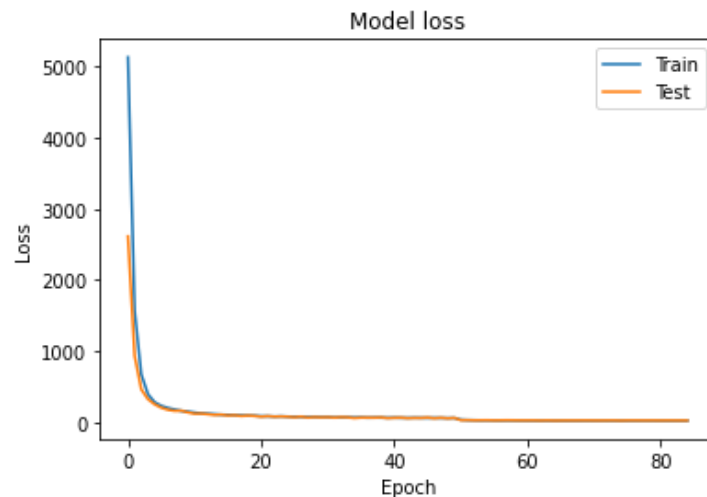


Fig 3: Loss

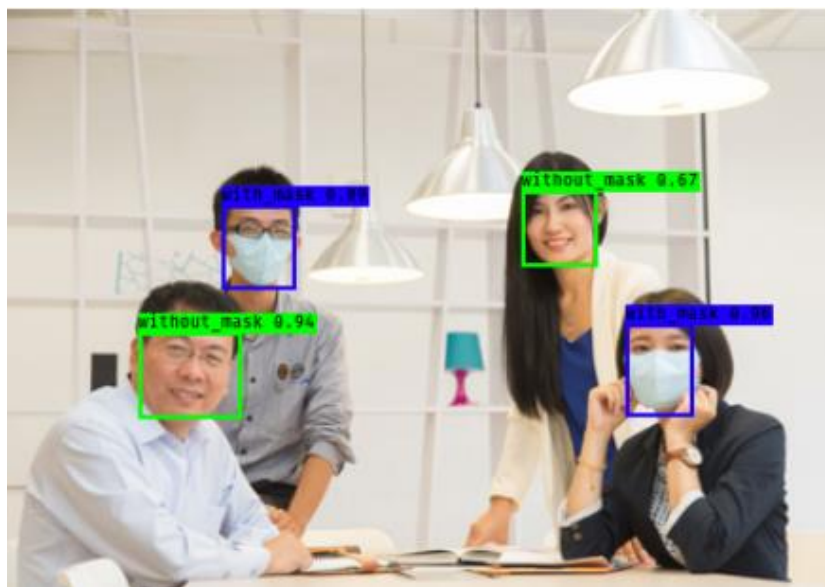


Fig4: prediction

## REFERENCES

- [1] Datasets, <https://www.kaggle.com/andrewmvd/face-mask-detection>, online accessed Dec 30, 2020
- [2] G. Jignesh Chowdary, Narinder Singh Pun, Sanjay Kumar Sonbhadra, and Sonali Agarwal: Face Mask Detection using Transfer Learning of InceptionV3, 2020
- [3] Joseph Redmon, Ali Farhadi: YOLOv3: An Incremental Improvement
- [4] GitHub <<https://github.com/qpwweee/keras-yolo3>>