

一、 相關研究

1.1 嬰兒猝死症

嬰兒猝死症 (The Sudden Infant Death Syndrome, 簡稱 SIDS) kinney_sudden_2000 之特徵為一位看似健康的嬰兒在睡眠期間突然死亡，其真正致死之原因尚不明確且非單一。

目前醫界雖未有單一定義此症之直接致死原因，但可統整出多項促使嬰兒猝死症發生之風險因素，可分為兩類因素：其一為外在因素，包含嬰兒因俯臥、側睡或蓋住面部等致使呼吸困難；其二為內在因素，包含發展因素（如：早產）、遺傳因素（如：家族性之嬰兒猝死症）、性別（男性比例為女性的兩倍）或種族等。除此之外，嬰兒也可能因其他外在環境條件，如：產前或產後暴露於不良物質中（如：香菸煙霧、酒精或非法藥物等），而弱化嬰兒之內在條件。

在嬰兒猝死症研究中，有許多關於此症之死亡機制理論，其中心肺控制假說主導了多數研究，也造就了往後關於嬰兒猝死症之研究多基於嬰兒呼吸或自主神經機制的缺陷。這樣的論點主要包含了五個步驟：(1) 發生危及生命的事件（如：面部朝下或面部遭遮蔽，將造成反射性或阻塞性呼吸暫停），而將導致嬰兒窒息、腦部灌注不足或兩者皆發生。(2) 嬰兒無法自行轉頭，以應付窒息的情境，而導致無法從呼吸暫停中恢復。(3) 持續的窒息導致失去意識或反射，即低氧昏迷。(4) 發生心率過緩及缺氧喘氣，此現象在嬰兒因嬰兒猝死症逝世前將明顯發生。(5) 嬰兒之自主復甦能力受損，即因無效的喘氣而最終導致呼吸暫停及死亡。

因此，由嬰兒猝死症之紀錄中，可看出此症狀並非一種突發疾病，而是在嬰兒死亡前，即會出現心率不正常或呼吸暫停之惡性循環現象。

另外，醫界亦發現俯臥睡姿將會使嬰兒猝死症之風險增加三倍以上，故在 1990 年代初期國際間即提倡嬰兒仰臥睡姿，嬰兒猝死症之發病率也因此降低了 50% 以上，但仍為嬰兒主要死亡原因之一。

1.2 嬰兒監測系統

在照護嬰兒的過程中，由於嬰兒尚未發展出語言能力表達自己的不適，或尚無能力將自己避免於危險之外。因此，為了協助照顧者關注嬰兒狀態，現有許多為自動化監測嬰兒之研究，主要分為以感測器偵測生理訊號及以影像式偵測兩種方式。

1.2.1 感測器偵測

此種方式利用多種不同感測器進行生理訊號之偵測，包含利用呼吸感測器、濕度感測器、溫度感測器、非接觸式紅外溫度感測器、三軸加速度計、慣性感測器、一氧化碳感測器、二氧化碳感測器等，分別量測嬰兒之呼吸頻率、出汗狀況、體溫、心率、身體位置或方向、睡眠姿勢、嬰兒周圍的一氧化碳濃度、呼出的二氧化碳濃度的變化等，且多會透過物聯網技術開發出可穿戴式裝置之系統。

如：Linti 等人 **linti_sensory_2006** 所開發的嬰兒感測背心，其將多個感官元件融入紡織品中以用來量測嬰兒之呼吸、心率、溫度及濕度；Ferreira 等人 **ferreira_smart_2016** 開發了將感測器裝設於胸帶中，而得以量測嬰兒之體溫、心率、呼吸頻率及身體位置，並透過 ZigBee 技術將收集到的數據傳送至伺服器，用戶則可透過醫療網頁介面進行查看及收到緊急訊息；Ziganshin 等人 **ziganshin_uwb_2010** 基於超寬頻技術開發出可監測嬰兒呼吸及心率之系統，其可檢測嬰兒之睡眠、清醒及警示狀態；

Lin 等人 **lin_wireless_2014** 開發出一套在嬰兒胸帶上嵌入了三種不同感測器的系統，透過三軸加速度計可確定嬰兒睡姿（面朝上、下、左或右）與計算 z 軸資訊得出呼吸頻率、利用溫度感測器量測體溫以及使用一氧化碳感測器偵測嬰兒周圍之一氧化碳濃度，再藉由 WiFi 模組傳送收集之生理資料至伺服器，而其驗證所計算之呼吸頻率準確率達 100%。

此種利用感測器監測嬰兒的方法，雖然可直接量測嬰兒之生理訊號以判斷狀態正常與否，但仍可能因硬體設備之缺陷無法準確量測，進而有失判斷準確性，亦或者因嬰兒需額外穿戴裝置而造成不適，進而影響嬰兒活動或導致更多危險的發生。

1.2.2 影像式偵測

此種方式利用電腦視覺技術對於嬰兒影像畫面進行偵測，現有研究中包含了計算嬰兒之呼吸頻率、關注於嬰兒之面部特徵及嬰兒趴睡姿勢偵測。

Fang 等人 **fang_vision-based_2015** 開發了一基於視覺之非接觸式呼吸頻率偵測系統，先判斷嬰兒是否正在運動（包含頭部、四肢及身體運動，但不包含因呼吸引起的輕微運動），若未偵測到嬰兒運動，則系統開始進行呼吸頻率偵測：首先，透過空間特徵擷取呼吸之候選點；接著，利用模糊積分技術選擇呼吸點；最終，得以計算嬰兒的呼吸頻率，進而可判斷嬰兒是否發生呼吸異常之情形。

Liu 等人 **liu_video-based_2017** 利用夜視攝影機拍攝在嬰兒床內的嬰兒，並使用 MIT 所提出之 Eulerian Magnification 技術，放大影片中的細微運動以監測拍攝對象之胸部運動，若經正規化之像素差異值低於設定閾值，則判斷其呼吸頻率異常，進而透過手機裝置發出警報。

Gallo 等人 **gallo_marrsids_2019** 提出一名為 MARRSIDS 的模型，其利用 OpenCV 之 Haar-Like Features 偵測嬰兒之面部特徵。系統透過嬰兒臉部辨識與否及睜眼狀態，判斷其是否處於危險情境中，而需發出聲音

警示：若臉部未被偵測，則認為嬰兒可能位於不良姿勢需發出警示；而若嬰兒為睜眼狀態，則代表嬰兒處於清醒狀態，並非處於風險中。

Wang 等人 **wang_multi-task_2019** 提出了一個多任務貝氏深度神經架構，其使用 MobileNetV2 網路，針對自行收集之 YunInfants 資料集進行嬰兒頭部影像分析，包含了四項子任務以達成嬰兒面部遮擋之監測：(1) 眼睛、鼻子或嘴巴是否可見，(2) 不可見的原因是否為被外物（如：枕頭）遮擋，(3) 眼睛睜開與否，及 (4) 五個臉部座標之位置。

Bharati 等人 **bharati_efficient_2021** 提出一個基於卷積神經網絡的電腦視覺系統，可用來評估嬰兒三種睡眠姿勢：仰臥（正常狀態）、從仰臥轉換到趴臥（警示狀態）、趴臥（危險狀態），並於嬰兒呈現趴臥姿勢時，透過手機提醒照護人員。而此系統亦提供了驗證反饋機制，供照護人員對於系統警報是否誤報之回饋。另外，由於目前未有公開之嬰兒姿勢資料集，此文透過拍攝和真實嬰兒相同比例之娃娃進行資料收集。

現有研究中，多關注於嬰兒呼吸運動、面部特徵或單一姿勢偵測，而尚未有對於嬰兒常見動作之辨識模型，故我們提出一可偵測嬰兒基礎姿勢及面部遮擋之危險監測系統。

1.3 ResNet

過往神經網路訓練中，更深層的網路會有模型退化的問題，亦即隨著網路深度的增加，準確率達飽和後，反而迅速下降，而這樣的結果並非因過度擬合所致，如圖 1.1 可看到兩個不同層數的網路其訓練誤差值。

因此，He 等人 **he_deep_2016** 提出了一個深度殘差學習 (圖 1.2) 的架構，利用 shortcut connection 執行 identity mapping，如此並不需要增加額外的參數，亦即不增加計算複雜度。最終，本研究以 152 層的殘差網路在 ILSVRC 2015 中獲得第一名，此網路比 VGG 網路深八倍，卻仍擁有較低的複雜度。

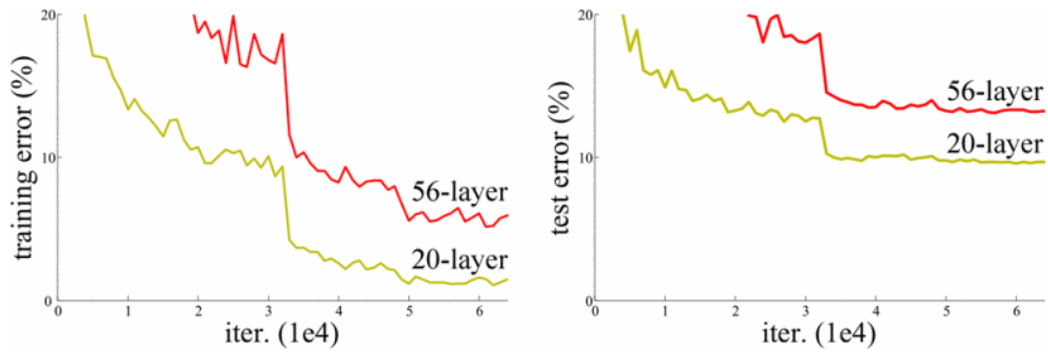


圖 1.1: 網路深度與訓練誤差關係

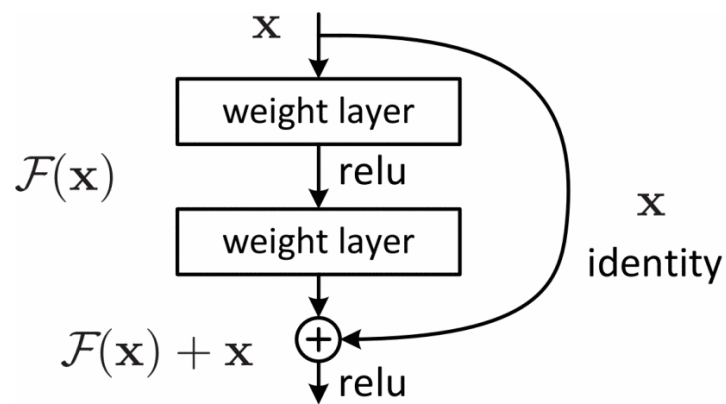


圖 1.2: 殘差學習

1.4 人臉偵測演算法

1.4.1 OpenCV

OpenCV `goyal_face_2017`

1.4.2 SSD

SSD `ye_face_2021`

1.4.3 MTCNN

MTCNN `zhang_joint_2016` 是由 Zhang 等人於 2016 年提出的一種多任務級聯卷積神經網路，可以同時處理人臉偵測及對齊任務；並提出可提升效能的 online hard sample mining 策略，其是否使用之效能差距如圖 1.3。

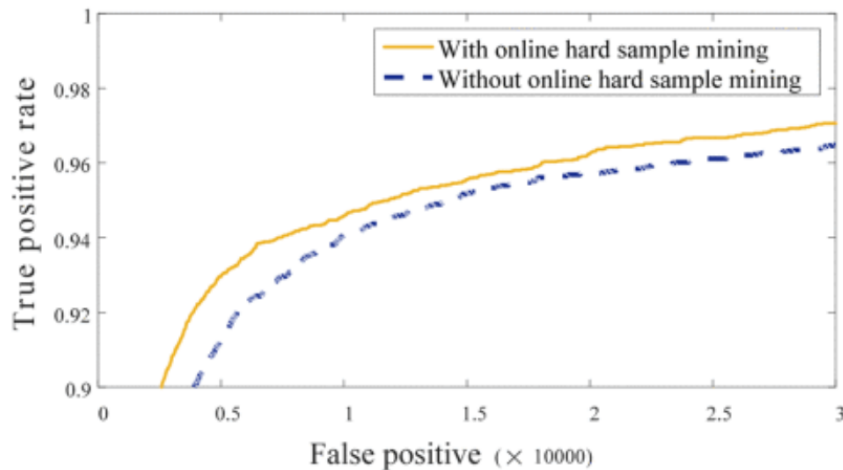


圖 1.3: 使用 online hard sample mining 策略效能比較

此網路包含三階段級聯架構的深度卷積網路，以粗到細的方式預測人臉及座標位置，其方法流程見圖 1.4：第一階段，由全卷積網路構成之 proposal network (P-Net) 獲得人臉區域的候選窗口及其邊界框回歸向量，並根據此估計回歸向量校準候選者，再以 nonmaximum suppression (NMS) 合併高度重疊的候選者；第二階段，所有候選者皆饋送至另一個稱為 refine network (R-Net) 的 CNN，其進一步拒絕大量錯誤候選者，並使用邊界框回歸進行校準及 NMS；第三階段，則利用 output network (O-Net) 輸出五個臉部的座標位置，其類似於第二階段，但不同處是為識別具有更多監督的人臉區域。此三階段網路的架構見圖 1.5，圖中”MP”為 max pooling、”Conv”為 convolution，而 pooling 及 convolution 的步長分別為 2 及 1。

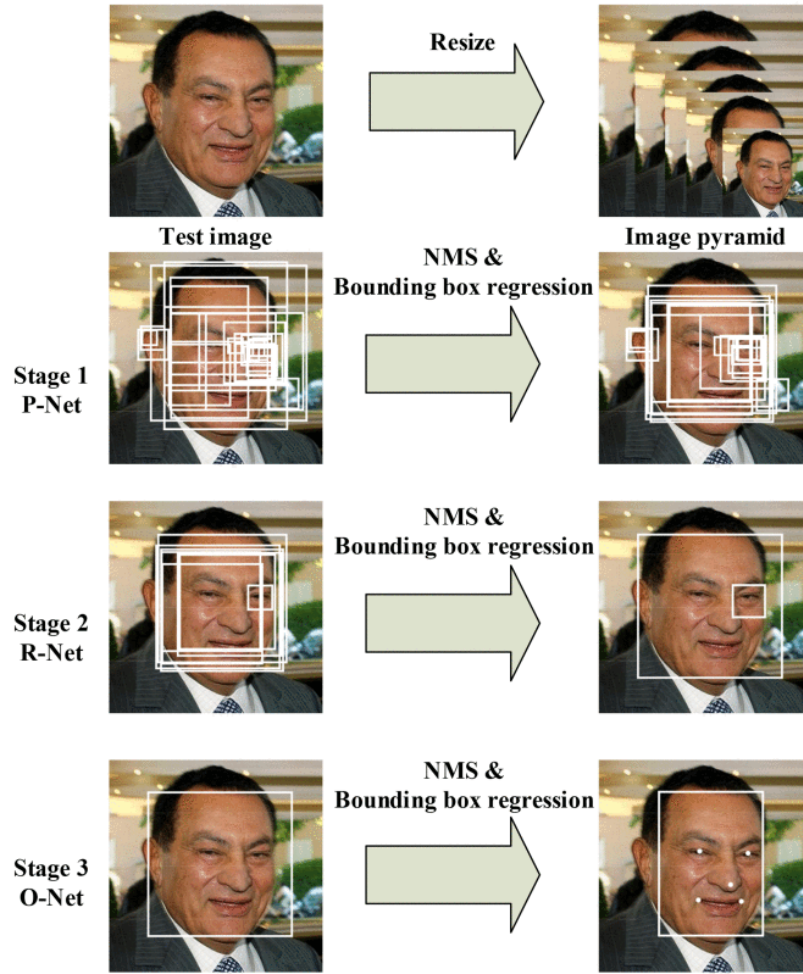


圖 1.4: MTCNN pipeline

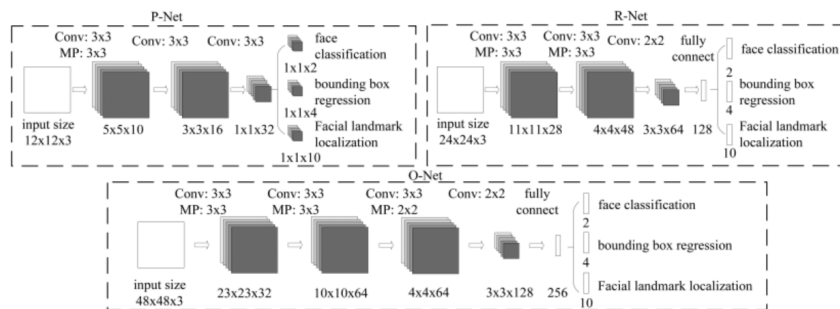


圖 1.5: MTCNN framework

1.4.4 RetinaFace

RetinaFace **deng_retinaface_2020** 是由 Deng 等人於 2020 年提出的 single-shot、multi-level 人臉定位方法，其基於影像平面之點回歸整合了人臉框預測、2D 人臉標示定位及 3D 頂點回歸。

此模型架構（見圖 1.6）中，主要由三個部分組成：（1）feature pyramid network、（2）context head module 及（3）cascade multi-task loss。首先，feature pyramid network 獲得輸入影像，並輸出五個不同比例的特徵圖；接著，context head module 獲得這些特徵圖以計算多任務的損失：亦即第一個模組會從一般的 anchor 預測範圍框，而後第二個模組利用第一個模組迴歸出的 anchor 以預測更精準的範圍框。

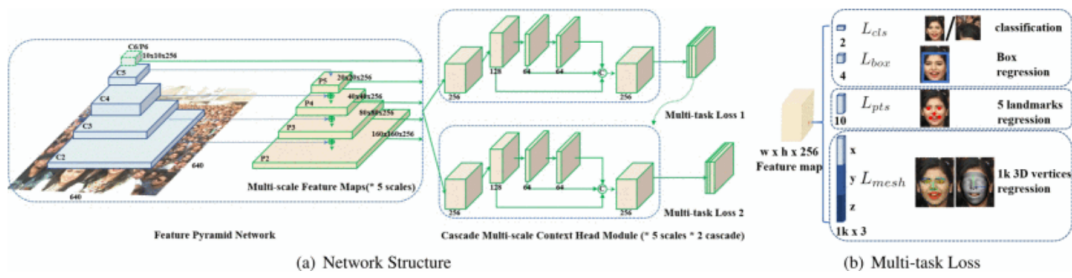


圖 1.6: RetinaFace 架構

本篇論文展示了 RetinaFace 和其他 29 種人臉偵測演算法之平均準確度（Average Precision）比較，如圖 1.7 所示，此演算法擁有 91.7% 的良好結果。

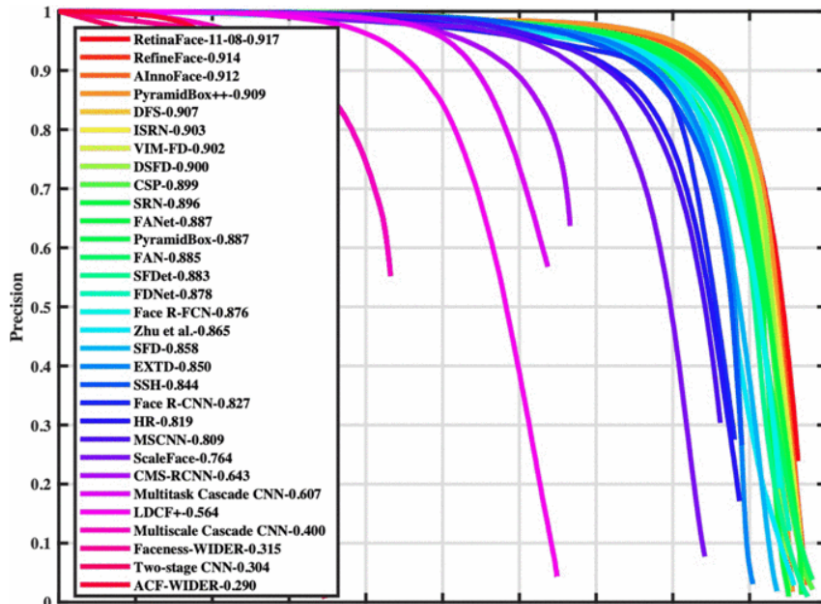


圖 1.7: RetinaFace(ResNet-152) 在 WIDER FACE 測試集之 Precision-Recall 曲線