

一、相關研究

本章節中，將說明有關本研究之背景知識與文獻回顧，透過以下四章子節進行分述：嬰兒猝死症、嬰兒監測系統、殘差神經網路及人臉偵測演算法。

1.1 嬰兒猝死症

嬰兒猝死症 (The Sudden Infant Death Syndrome, 簡稱 SIDS) [kinney_sudden_2007.pdf](#) 其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。

目前醫界對嬰兒猝死症之直接致死原因尚未有統一的定義，但可統整出多項促使此症發生之風險因素，主要可分為兩類：其一為外在因素，包含嬰兒因俯臥及側睡姿勢、蓋住臉部的床單、嬰兒睡在沙發或其他容易陷入的柔軟家具上等，致使嬰兒呼吸困難而死亡；其二則為內在因素，包含成長因素（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，也造就了往後關於此症之探討多基於嬰兒呼吸或自主神經機制的缺陷。此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，造成反射性或阻塞性呼吸暫

停)，而將導致嬰兒窒息或腦部灌注不足，亦可能兩者皆發生。(2) 嬰兒無法自行轉頭，以應對窒息的情境，而導致無法從呼吸暫停中恢復。(3) 持續性窒息導致嬰兒失去意識或反射，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) 嬰兒的自主復甦能力受損，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，亦有研究人員使用 Triple-Risk Model 來解釋嬰兒猝死症 noauthor_what_nod 即嬰兒死於此症需同時包含以下三個因素：(1) 有風險的嬰兒：嬰兒含有一個未知的問題，可能是基因突變或腦部缺陷等，這使其面臨了嬰兒猝死症的風險；(2) 嬰兒發育的重要時期：嬰兒出生後的前六個月，將經歷許多快速成長的階段，此階段會改變身體控制和調節自身的能力，且嬰兒的身體也會於此時學習如何對環境做出應變；(3) 環境中的壓力源：即前述中之外在因素，包含嬰兒睡姿及接觸香菸煙霧等。若僅發生其中一項因素，將不足以導致嬰兒猝死症引發死亡。

因此，若能消除環境中的壓力源，將有利於嬰兒的生存。同時醫界亦發現俯臥睡姿勢將使嬰兒猝死症風險增加三倍以上，故在 1990 年代初期國際間即提倡嬰兒仰臥睡姿，此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

由於嬰兒尚無能力表達且無法避免危險，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

此種方式利用呼吸感測器、濕度感測器、溫度感測器、慣性感測器、三軸加速度計、一氧化碳感測器及二氧化碳感測器等多種不同設備，進行嬰兒呼吸頻率、出汗狀況、體溫、心率、身體位置或方向、睡眠姿勢、周圍一氧化碳濃度及呼出的二氧化碳濃度變化等訊號之偵測，且多會透過物聯網技術開發出可穿戴式裝置，以下介紹幾項基於感測器式偵測嬰兒之研究。

Linti 等人 **linti_sensory_2006** 將多個感官元件融入紡織品以開發嬰兒感測背心，此裝置可用來量測嬰兒之呼吸、心率、溫度及濕度，其示意圖見圖 1.1。



圖 1.1: Linti 嬰用感測背心之穿脫示意圖 **linti_sensory_2006**

Ferreira 等人 **ferreira_smart_2016** 將心律感測器、3D 加速度計、熱電堆感測器裝設於胸帶中（圖 1.2），以量測嬰兒之體溫、心率、呼吸頻率及身體位置，再透過 ZigBee 技術傳送數據至伺服器，並可由網頁查收訊息。

Lin 等人 **lin_wireless_2014** 於胸帶中嵌入三軸加速度計、溫度感測器及一氧化碳感測器，可用來量測嬰兒睡姿（面朝上、下、左或右）、體溫及其周圍一氧化碳濃度，且利用三軸加速度計 z 軸資訊可獲得嬰兒之

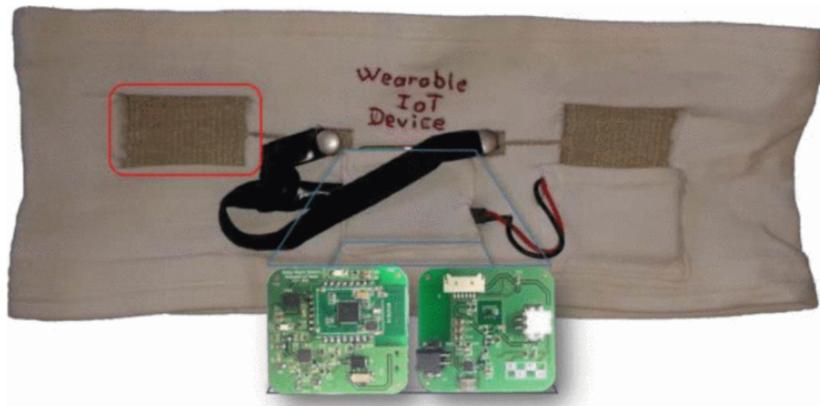


圖 1.2: Ferreira 嬰用感測胸帶 ferreira_smart_2016

呼吸頻率，最終驗證計算之呼吸頻率準確率達 100%。此系統架構圖見圖 1.3，

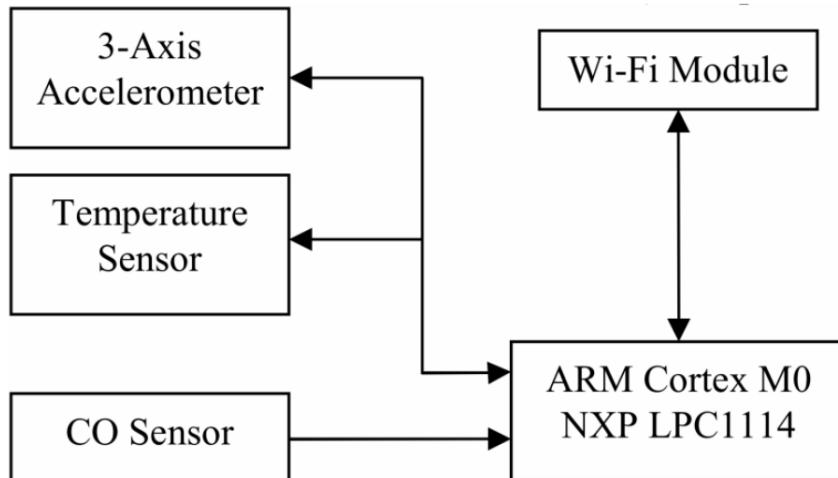


圖 1.3: Lin 系統架構圖 lin_wireless_2014

Ziganshin 等人 **ziganshin_uwb_2010** 基於超寬頻技術開發出可監測呼吸及心率之系統（如圖 1.4），以檢測嬰兒三種狀態：睡眠（無運動）、清醒（長期大幅度運動）及警示（偵測到呼吸頻率異常低或暫停）。

此種利用感測器監測嬰兒的方法，因為功能的單一性，為了擁有許多功能就須增加不同的感測器，而當這些裝置需穿戴在嬰兒身上時，將影響其活動或造成不適，更可能導致潛在的危險。

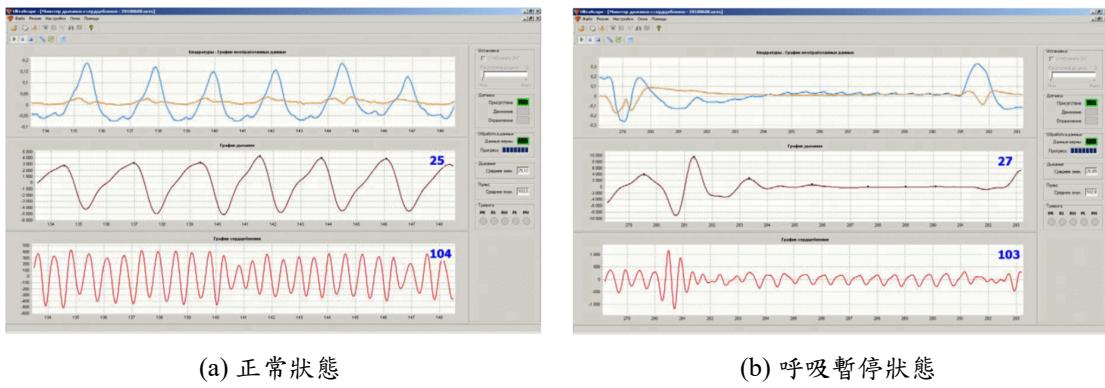


圖 1.4: Ziganshin 系統之嬰兒呼吸及心跳圖 ziganshin_uwb_2010

1.2.2 影像式偵測

現有研究中，基於電腦視覺技術之監測系統雖日漸廣泛，但多針對小孩、成人或老人之照護進行開發，而少數應用於嬰兒的偵測系統中，又多關注於呼吸頻率、面部特徵及趴睡姿勢等，以下介紹幾項基於影像式偵測嬰兒之研究。

有別於往的接觸式或非視覺技術之呼吸頻率偵測系統，Fang 等人 fang_vision-based_2015 基於視覺技術進行非接觸式之呼吸頻率偵測。此系統完整流程如圖 1.5a：先進行嬰兒運動偵測，其判斷流程見圖 1.5b，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；若系統未偵測到運動情形，則透過空間特徵擷取候選呼吸點，再利用模糊積分技術選擇呼吸點，即可計算呼吸頻率，進而判斷嬰兒呼吸是否異常。

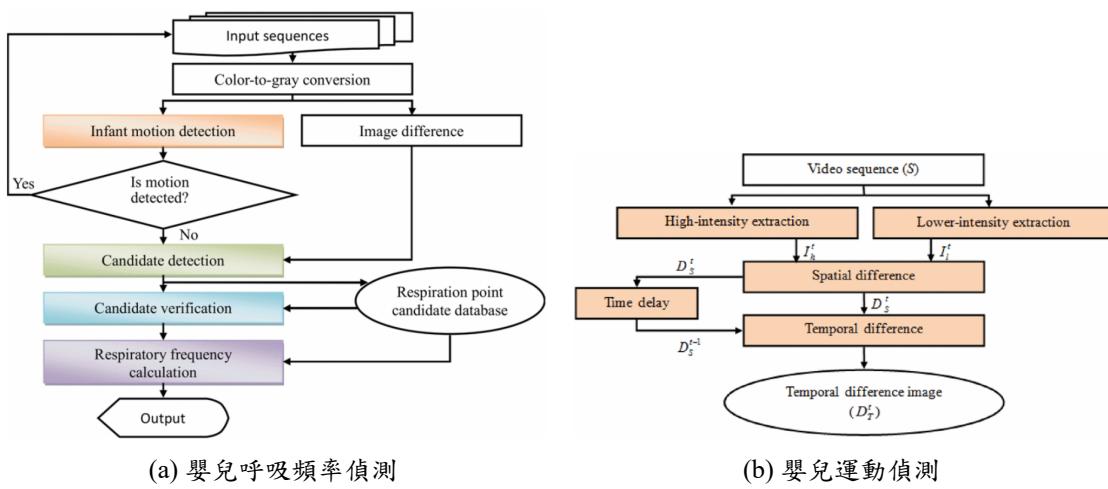


圖 1.5: Fang 系統流程圖 fang_vision-based_2015

Liu 等人 **liu_video-based_2017** 利用夜視攝影機拍攝在嬰兒床內的嬰兒，判斷其呼吸頻率異常與否。此系統主要包含三個部分：(1) 影片收集：為圖 1.6 中的前兩項，透過夜視攝影機連接到 Artik 板以收集嬰兒影片；(2) 呼吸偵測演算法：為圖 1.6 中的第三及四項，包含了使用 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，而若經正規化之像素差異值低於設定閾值，則判斷為呼吸頻率異常；(3) 警示：為圖 1.6 中的最後兩項，透過 Twilio 在演算法偵測到緊急狀況時，向父母手機發出警報。

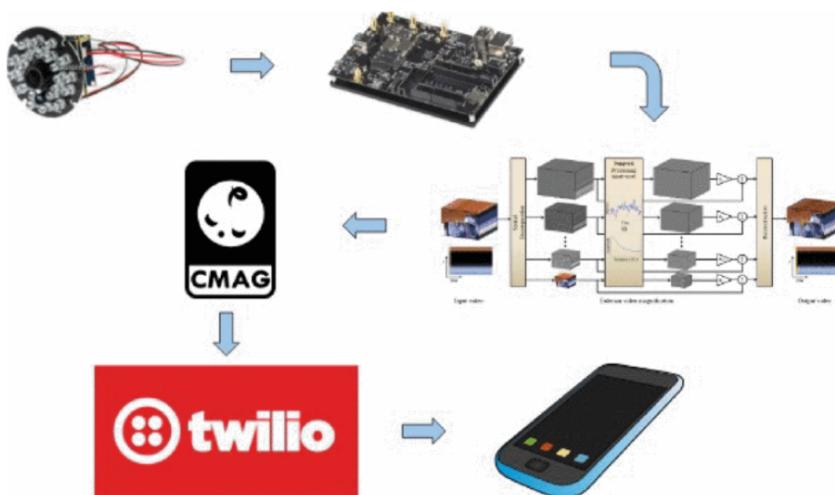


圖 1.6: Liu 系統設計圖 **liu_video-based_2017**

Gallo 等人 **gallo_marrsids_2019** 提出名為 MARRSIDS 的模型（架構如圖 1.7），其基於 Haar Cascade Classifier 進行嬰兒面部特徵偵測，若未偵測到其臉部，則認為嬰兒可能位於不良姿勢，需發出警報；而若偵測到嬰兒臉部且為睜眼狀態，則代表其處於清醒狀態，並非處於風險中。

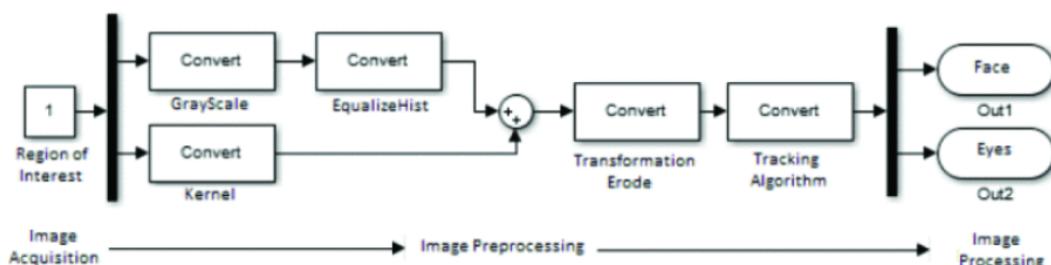


圖 1.7: MARRSIDS 模型架構圖 **gallo_marrsids_2019**

Wang 等人 **wang_multi-task_2019** 提出用以偵測嬰兒臉部遮擋的貝氏深度神經網路架構（圖 1.8），共包含四項子任務：(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。此框架使用了損失函數進行訓練，並考慮不同偵測任務間的不確定性，且有別於現有不專注於嬰兒影像的研究，此方法使用 MobileNetV2 針對自行收集之 YunInfants 資料集進行嬰兒頭部影像分析。

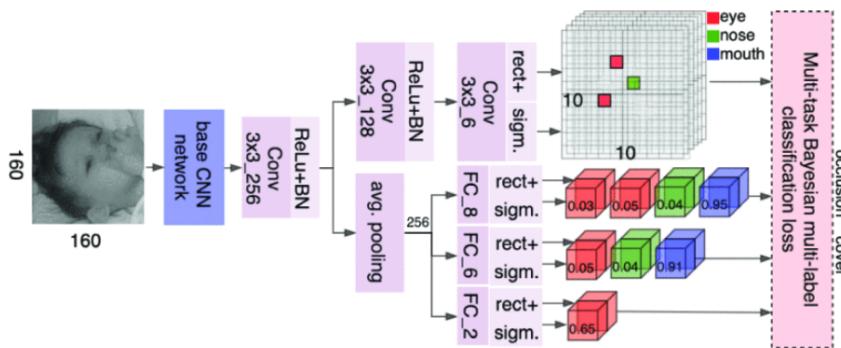


圖 1.8: Wang 系統之網路架構 **wang_multi-task_2019**

Bharati 等人 **bharati_efficient_2021** 開發基於卷積神經網絡之嬰兒睡眠姿勢電腦視覺系統，以評估仰臥（正常狀態）、從仰臥轉換到趴臥（警 示狀態）及趴臥（危險狀態）三種姿勢（如圖 1.9），並於嬰兒呈現趴臥時，透過手機提醒照護人員。此系統完整架構見圖 1.10a，其中卷積神經網路部分之架構如圖 1.10b：首先，輸入經轉換的 2D 嬰兒灰階影像；接著，經過多次的卷積層及最大池化層；再傳入全連接層後，最終輸出嬰兒的三種睡眠姿勢機率值。由於目前未有公開之嬰兒姿勢資料集，故此研究所使用的影像資料為和真實嬰兒相同比例之娃娃。



圖 1.9: Bharati 系統之嬰兒睡眠姿勢分類 **bharati_efficient_2021**

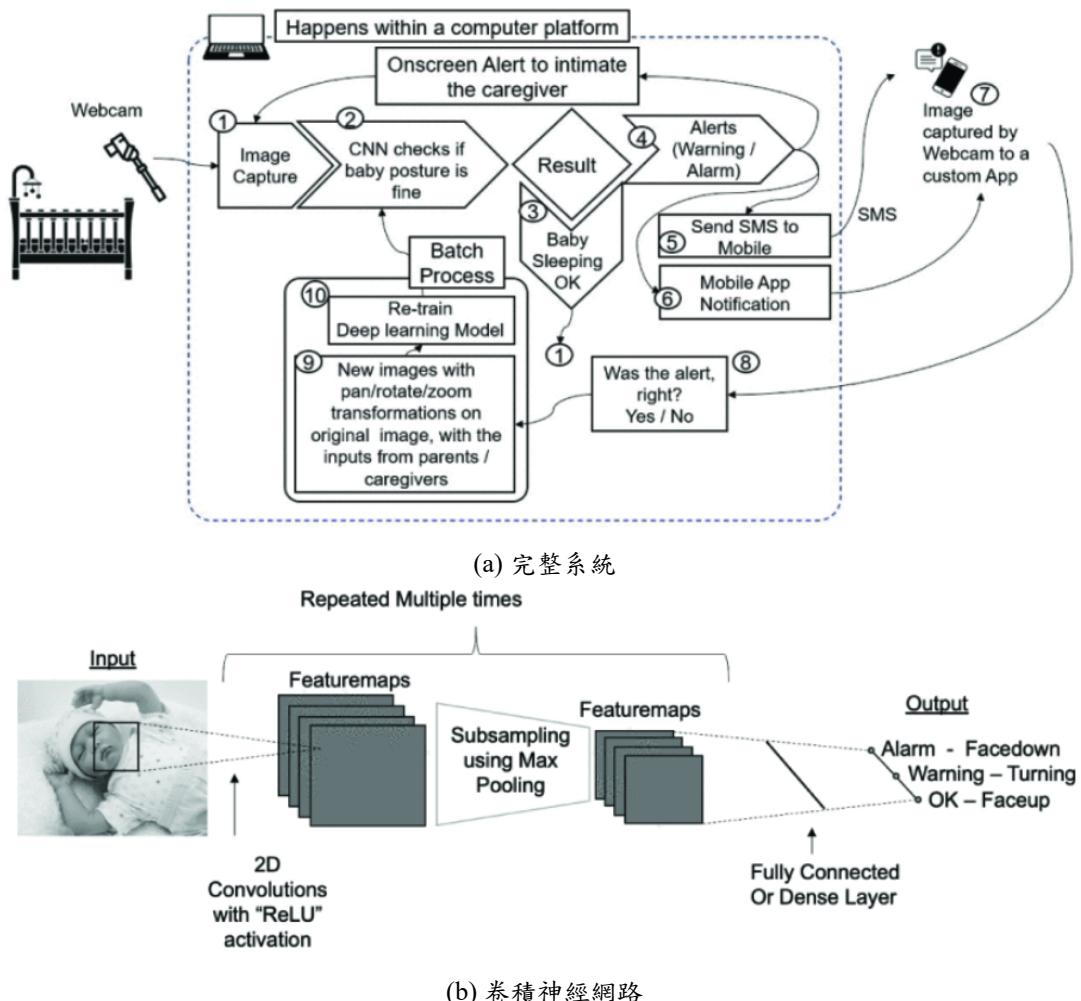


圖 1.10: Bharati 系統架構圖 bharati_efficient_2021

現有研究中，視覺化之嬰兒監測系統多關注其呼吸運動、面部特徵或單一姿勢，而未有辨識嬰兒常見動作之研究，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 殘差神經網路

既有研究中 simonyan_very_2015, szegedy_going_2015, ioffe_batch_2015，可知卷積神經網路其深度至關重要，在 ImageNet 資料集中結果表現良好的網路皆為十六至三十層之深層網路。

然而，當訓練更深層的神經網路時，卻會出現退化問題，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 1.11 可看到兩個不同層數的網路其訓練及測試誤差值。

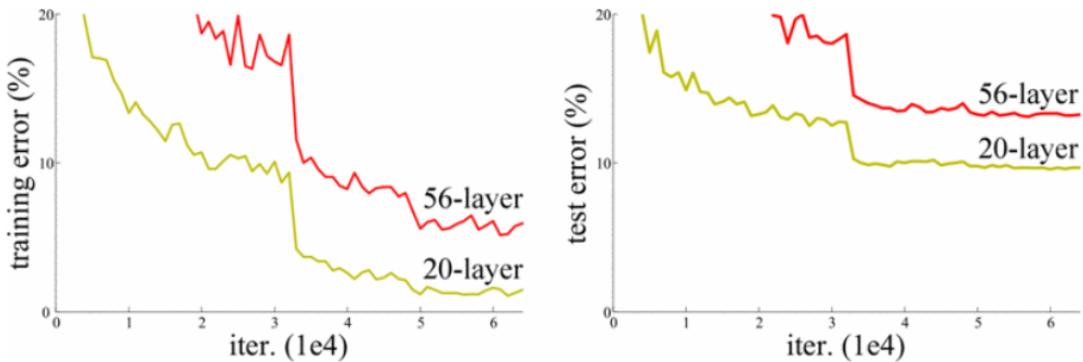


圖 1.11：網路深度與訓練誤差關係 he_deep_2016

因此，He 等人 **he_deep_2016** 提出使用深度殘差學習（圖 1.12）的網路架構，利用 shortcut connection 執行 identity mapping，且不需要增加額外的參數，亦即不增加計算複雜度。

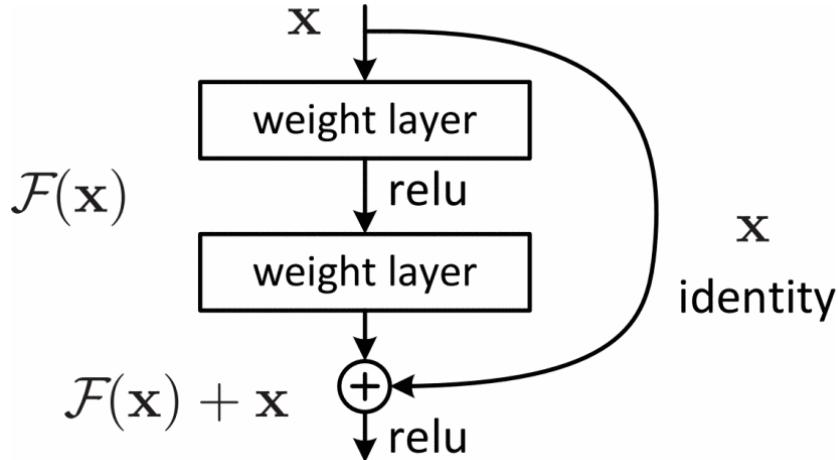


圖 1.12：殘差學習 he_deep_2016

此研究中，通過訓練 ImageNet 評估不同層數之普通網路與殘差網路，由圖 1.13（左）可觀察到普通網路中 34 層卻比 18 層有更高的驗證誤差，而圖 1.13（右）則可看出殘差網路中 34 層相對於 18 層有較低的訓練誤差，此結果說明退化問題獲得了解決。而文中亦透過訓練 CIFAK-10 進行更多的比較，圖 1.14（左）為普通網路不同層數之誤差曲

線圖，圖 1.14（中）為殘差網路不同層數其誤差曲線圖，圖 1.14（右）則為使用 110 層及 1202 層殘差網路之誤差曲線圖。

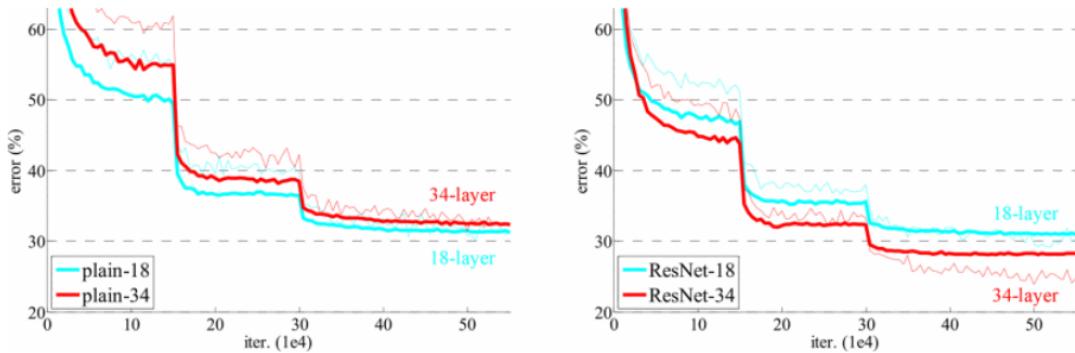


圖 1.13: 在 ImageNet 上訓練之誤差曲線圖 he_deep_2016

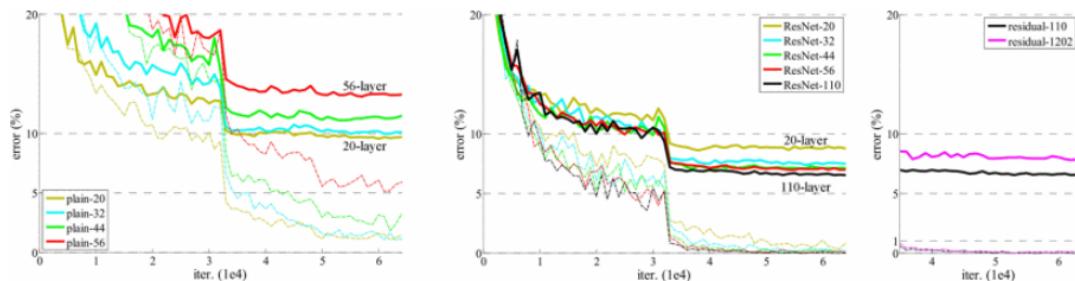


圖 1.14: 在 CIFAK-10 上訓練之誤差曲線圖 he_deep_2016

最終，本研究以 152 層之殘差網路在 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。

1.4 人臉偵測

現有研究中，有許多為偵測人臉之演算法，以下介紹幾篇論文。

Cuimei 等人提出使用 Haar cascade classifier 的人臉偵測演算法 cuimei_human_2017 並結合三個弱分類器：第一個為基於膚色直方圖匹配的決策節點，以拒絕大量僅使用 Haar cascade classifier 而偵測錯誤的非人臉；第二及第三個則為基於眼睛偵測及嘴部偵測的分類，進一步去除顏色與人類膚色相近卻沒有眼睛和嘴部的非人臉，以降低錯誤率。此演算法之偵測架構，如??所示。

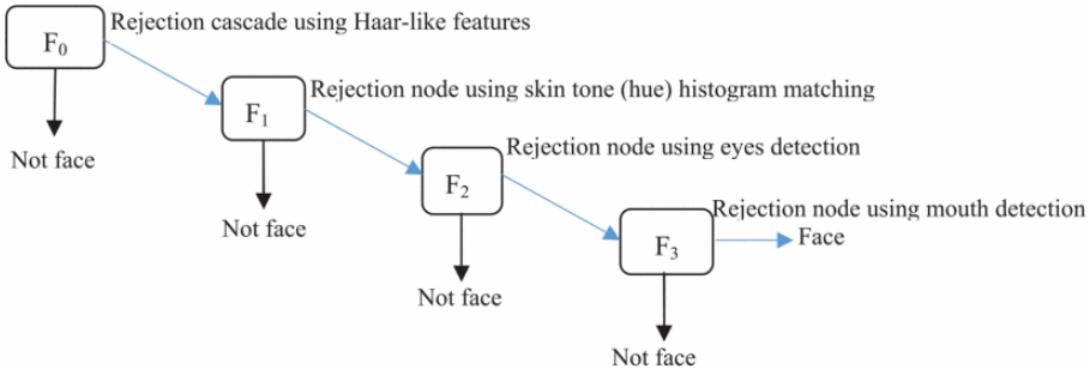


圖 1.15: Cuimei 人臉偵測演算法之偵測架構 cuimei_human_2017

Face SSD ye_face_2021是由 Ye 等人基於 single shot multibox detector (SSD) 提出的人臉偵測器，為了加快檢測速度及提高準確度，其作出了三項改進：其一為改進了 ShuffleNet V2 的架構，使用 Group Convolution 取代 Depthwise Convolution 以提高準確率，將此網路作為骨幹網路；其二提出由多個卷積層堆疊的 modified prediction module，並使用 anchor densification 策略以確保不同特徵圖的 anchor 密度相同，解決小人臉召回率較低之問題；最後引入了一個規模公平的人臉檢測框架，以更好的匹配人臉。此方法之偵測架構如??，並透過實驗得到在 CPU 及 GPU 上運行分別為每秒 40 帧以及每秒 110 帧的結果。

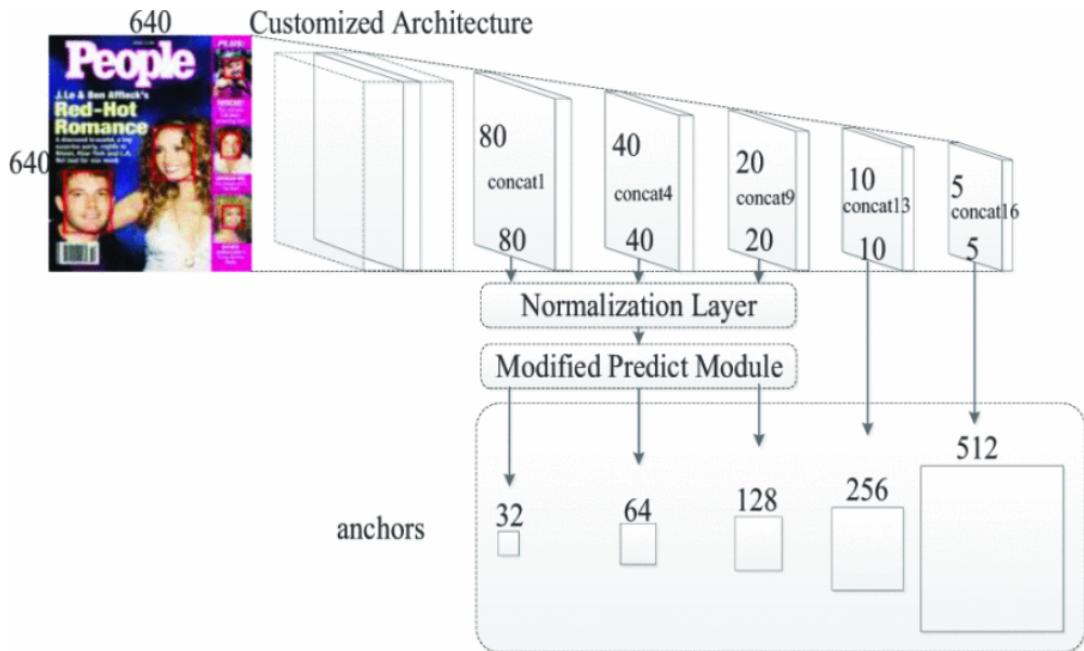


圖 1.16: Face SSD 之偵測架構 ye_face_2021

MTCNN ***zhang_joint_2016***是由 Zhang 等人提出的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。此演算法包含三階段架構（如圖 1.17）：第一階段為全卷積網路構成之 proposal network（簡稱 P-Net），用來獲得人臉區域的候選窗口及其邊界框回歸向量，根據此回歸向量校準候選者，再以非極大值抑制合併高度重疊的候選者；而第二階段，所有候選者皆饋送至另一個稱為 refine network（簡稱 R-Net）的卷積神經網路，以進一步拒絕大量錯誤候選者；第三階段中，則利用 output network（簡稱 O-Net）輸出五個臉部的座標位置，其目標為識別受更多監督的人臉區域。由圖 1.18，可清楚看出此網路是以粗到細的方式預測人臉及座標位置。

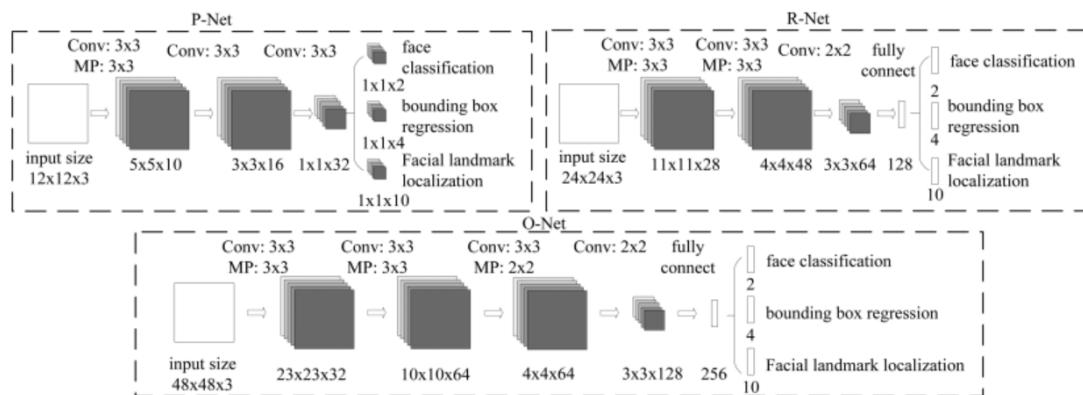


圖 1.17: MTCNN 架構圖 ***zhang_joint_2016***

RetinaFace ***deng_retinaface_2020***由 Deng 等人提出的人臉定位方法，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 1.19）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，由 feature pyramid network 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。

本研究中，將應用上述介紹的演算法進行嬰兒臉部的偵測，並於第四章節中的實驗進行各演算法之比較。

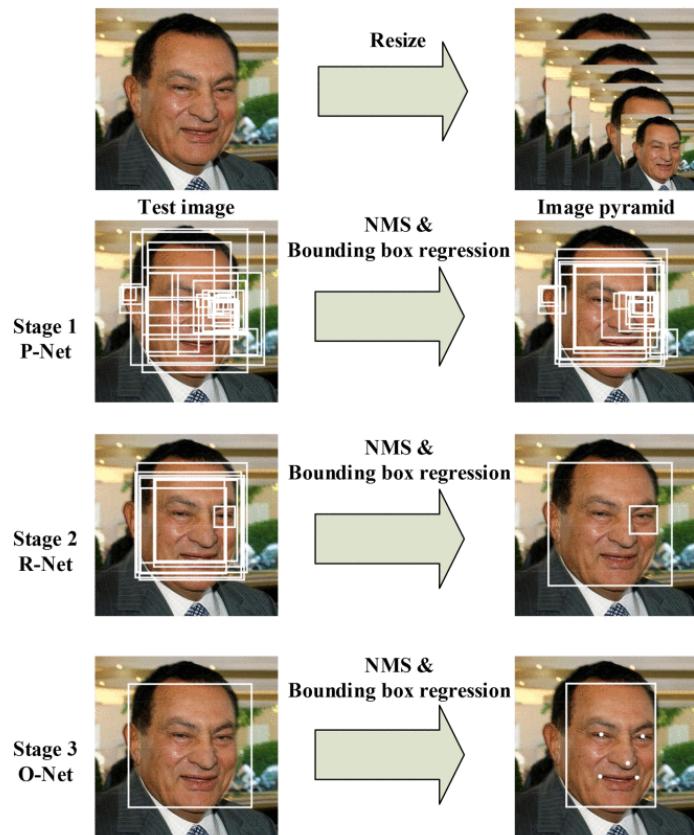


圖 1.18: MTCNN pipeline zhang_joint_2016

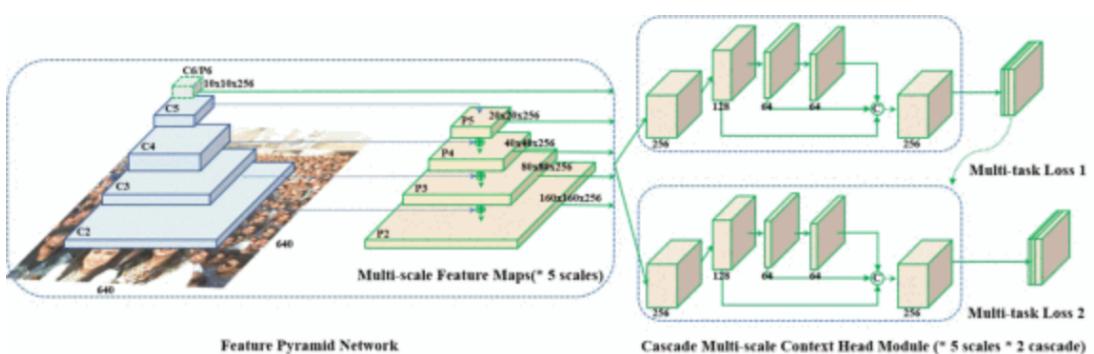


圖 1.19: RetinaFace 架構圖 deng_retinaface_2020