

國 立 中 央 大 學

資訊工程學系
碩士論文

基於深度學習之嬰兒危險監測系統

A Deep-learning-based Danger Monitoring System
for Infants

研 究 生：王佳君
指 導 教 授：蘇木春 博 士

中 華 民 國 一 百 一 十 一 年 六 月

國立中央大學圖書館學位論文授權書

填單日期： 111 / 06 / 30

2019.9 版

授權人姓名	王佳君	學 號	109522116
系所名稱	資訊工程學系碩士班	學位類別	<input checked="" type="checkbox"/> 碩士 <input type="checkbox"/> 博士
論文名稱	基於深度學習之嬰兒危險監測系統	指導教授	蘇木春

學位論文網路公開授權

授權本人撰寫之學位論文全文電子檔：

- 在「國立中央大學圖書館博碩士論文系統」.

() 同意立即網路公開

() 同意 於西元 _____ 年 _____ 月 _____ 日網路公開

() 不同意網路公開，原因是： _____

- 在國家圖書館「臺灣博碩士論文知識加值系統」

() 同意立即網路公開

() 同意 於西元 _____ 年 _____ 月 _____ 日網路公開

() 不同意網路公開，原因是： _____

依著作權法規定，非專屬、無償授權國立中央大學、台灣聯合大學系統與國家圖書館，不限地域、時間與次數，以文件、錄影帶、錄音帶、光碟、微縮、數位化或其他方式將上列授權標的基於非營利目的進行重製。

學位論文紙本延後公開申請 (紙本學位論文立即公開者此欄免填)

本人撰寫之學位論文紙本因以下原因將延後公開

- 延後原因

() 已申請專利並檢附證明，專利申請案號：

() 準備以上列論文投稿期刊

() 涉國家機密

() 依法不得提供，請說明： _____

- 公開日期：西元 _____ 年 _____ 月 _____ 日

※繳交教務處註冊組之紙本論文(送繳國家圖書館)若不立即公開，請加填「國家圖書館學位論文延後公開申請書」

研究生簽名：_____

指導教授簽名：_____

國立中央大學碩士班研究生
論文指導教授推薦書

資訊工程 學系/研究所 王佳君 研究生所
提之論文 基於深度學習之嬰兒危險監測系統 係
由本人指導撰述，同意提付審查。

指導教授_____ (簽章)

111 年 06 月 30 日

國立中央大學碩士班研究生
論文口試委員審定書

資訊工程 學系/研究所 王佳君 研究生所
提之論文 基於深度學習之嬰兒危險監測系統 經
本委員會審議，認定符合碩士資格標準。

學位考試委員會召集人 _____

委 員 _____

中 華 民 國 111 年 06 月 30 日

基於深度學習之嬰兒危險監測系統

摘要

嬰兒照護者在照顧嬰兒時，可能發生無法隨時關注嬰兒狀態的情形，使得嬰兒因溢奶、翻身、趴睡等情形，致使呼吸不順而發生憾事。

又因現有產品多用感測器偵測嬰兒狀態，功能單一且多有使用限制，便利性不佳。

因此，本論文提出基於深度學習技術，專注於嬰兒影像畫面進行臉部遮擋及姿勢辨識之危險偵測系統：利用 ResNet50 網路訓練模型，以分析嬰兒臉部是否遭異物遮蔽及辨識四種基礎姿勢：正躺、趴躺、坐姿及站立。

兩部分之辨識精確度皆達 98%。

故當系統輸入嬰兒影片時，可透過模型辨識嬰兒的姿勢可能處於危險狀態或臉部遭異物遮擋，則可即時警示照護者。

由於目前未有公開之嬰兒資料集，故本文中所使用的嬰兒照片皆為網路圖片及影片進行擷取並前處理而成。

關鍵字：嬰兒危險偵測，嬰兒臉部辨識，深度學習，嬰兒猝死症

A Deep-learning-based Danger Monitoring System for Infants

Abstract

When taking care of the infant, the baby caregiver may not be able to pay attention to the status of the baby at any time, which may cause the infant to suffer from unpleasant breathing due to overflowing milk, turning over, sleeping on the stomach, etc.

In addition, the existing products use multiple sensors to detect the state of the infant, which has a single function and many restrictions on use, resulting in poor convenience.

Therefore, this paper proposes a danger detection system based on deep learning technology, focusing on face occlusion and gesture recognition of baby images: The ResNet50 network is used to train the model to analyze whether the infant's face is obscured by foreign objects and recognize four basic postures: lying, lying on the stomach, sitting and standing.

The recognition accuracy of both parts is 98%.

Therefore, when the system inputs a infant video, the model can recognize that the infant's posture may be in a dangerous state or the face is covered by a foreign object, and the caregiver can be immediately alerted.

Since there is currently no public infant data set, the infant photos used in this article are all captured and pre-processed from online pictures and videos.

Keywords: Infant danger detection, Infant face detection, Deep learning, Sudden infant death syndrome

誌謝

就讀碩士班期間，接受了很多人的幫助與鼓勵，非常感謝這兩年的所有時光。

首先，感謝蘇木春老師的指導，在研究上給予了我很多的方向與教學，讓我在碩士期間獲益良多；也感謝蘇老師很信任學生，讓我在這兩年得以很好的妥善規劃時間。

接著，感謝實驗室的每位成員，讓我在 CILAB 擁有這麼珍貴的回憶。感謝佳菁、小花、熙琪、小烏龜，處理實驗室繁忙的事務，讓我們可以有這麼舒適的研究環境；感謝威任學長、偉倫學長，在研究中提供了很多想法，讓我在研究中能有更明確的方向；感謝子謙學長、政育學長、映如學姊、書仔學姊，在我有疑惑的時候，解答了許多的問題，帶領我們熟悉實驗室的生活；感謝鈞翔、昌翰、逸星、奕蘋、詩勻，交流彼此的意見與想法，讓我能激發出更多不同的思維；感謝智穎、景豐、季劫、譽鈞、姿瑩，一起參與了許多實驗室的事務，讓所有活動及計畫得以完成。

最後，也感謝我的家人、朋友、愛人以及我自己，在不同的人生階段中一起度過，給予我非常多的支撐與幫助，而得以逐漸成長為一位更成熟的個體。

目錄

	頁次
摘要	iv
Abstract	v
誌謝	vii
目錄	viii
一、 緒論	1
1.1 研究動機	1
1.2 研究目的	2
1.3 論文架構	3
二、 相關研究	4
2.1 嬰兒猝死症	4
2.2 嬰兒監測系統	5
2.2.1 感測器偵測	5
2.2.2 影像式偵測	8
2.3 ResNet.....	12
2.4 人臉偵測演算法	14
2.4.1 MTCNN.....	14
2.4.2 RetinaFace.....	16

三、 研究方法	18
3.1 系統流程介紹	18
3.2 臉部遮擋辨識	19
3.2.1 嬰兒臉部偵測	20
3.2.2 嬰兒臉部資料集	20
3.3 危險動作辨識	21
3.3.1 嬰兒姿勢資料集	22
3.3.2 危險動作判斷方法	26
四、 實驗設計與結果	27
4.1 嬰兒臉部偵測實驗	27
4.1.1 實驗目的與設計	27
4.1.2 實驗評估方式	27
4.1.3 實驗結果與分析	27
4.2 臉部遮擋辨識實驗	30
4.2.1 實驗目的與設計	30
4.2.2 實驗結果分析	30
4.3 姿勢辨識實驗	31
4.3.1 實驗目的與設計	31
4.3.2 實驗結果分析	31
4.4 影片危險偵測實驗	33
4.4.1 實驗設計	33
4.4.2 實驗評估方式	33
4.4.3 實驗結果分析	33
五、 結論與未來展望	34
5.1 結論	34
5.2 未來展望	35

參考文獻

36

圖 目 錄

頁次

2.1	嬰兒多感測器背心之穿脫示意圖	6
2.2	嬰兒穿戴式感測器裝置圖	6
2.3	嬰兒正常狀態之呼吸及心跳圖	7
2.4	嬰兒呼吸暫停狀態之呼吸及心跳圖	7
2.5	嬰兒監測系統架構圖	8
2.6	z 軸顯示吸氣及呼氣脈衝	8
2.7	嬰兒呼吸頻率偵測系統流程圖	9
2.8	嬰兒運動偵測流程圖	10
2.9	系統設計圖	10
2.10	硬體架構圖	10
2.11	人工視覺模型架構圖	11
2.12	網路架構	11
2.13	睡姿分類	12
2.14	CNN 架構	12
2.15	系統架構	12
2.16	網路深度與訓練誤差關係	13
2.17	殘差學習	13
2.18	使用 online hard sample mining 策略效能比較	14
2.19	MTCNN pipline	15
2.20	MTCNN pipline	15

2.21 RetinaFace 架構	16
2.22 RetinaFace(ResNet-152) 在 WIDER FACE 測試集之 Precision-Recall 曲線	17
3.1 系統流程圖	18
3.2 臉部膚色偵測	19
3.3 臉部遮擋辨識流程圖	20
3.4 嬰兒臉部無遮蔽	20
3.5 嬰兒臉部遮蔽物為奶嘴	21
3.6 嬰兒臉部遭異物遮擋	21
3.7 危險動作辨識流程圖	22
3.8 嬰兒正躺姿勢	23
3.9 嬰兒趴躺姿勢	23
3.10 嬰兒坐姿姿勢	24
3.11 嬰兒站立姿勢	25
4.1 臉部辨識訓練結果	30
4.2 臉部遮擋辨識模型之混淆矩陣	31
4.3 姿勢辨識訓練結果	32
4.4 姿勢辨識模型之混淆矩陣	32
4.5 姿勢辨識錯誤之影像：坐姿誤判為趴躺	33

表 目 錄

頁 次

4.1	OpenCV 演算法偵測嬰兒臉部結果	29
4.2	SSD 演算法偵測嬰兒臉部結果	29
4.3	MTCNN 演算法偵測嬰兒臉部結果	29
4.4	RetinaFace 演算法偵測嬰兒臉部結果	29

一、緒論

1.1 研究動機

根據衛生福利部統計處所發布的嬰兒主要死因統計 [1] 中，101 年至 105 年間每年至少 30 位嬰兒死於嬰兒猝死症候群（Sudden infant death syndrome，簡稱 SIDS），106 年至 109 年每年亦仍有超過 20 位嬰兒因此症狀逝世，是為嬰兒十大死亡原因之一。

三軍總醫院對於嬰兒猝死症的說明為：一個原本無異狀的嬰兒，突然且無法預期的死亡，常發生在嬰兒睡眠時，並在事後的屍體解剖檢查中找不到其真正致死原因。凡未滿一歲的嬰幼兒皆可能發生，其中二至四個月時期尤為常見，亦可能發生在嬰兒出生一至兩周內。醫界雖持續探討嬰兒猝死症的發生原因，但目前對於真正的成因仍不清楚，綜合醫界當前相關因素的研究中，包含了嬰兒因溢奶或嘔吐產生呼吸道緊縮反射及憋氣，或因翻身、趴睡致使呼吸困難，而窒息死亡等原因。

當照護者在嬰兒照護時，可能有許多不可避免的情形，而難免發生視線離開嬰兒的情形，如：泡奶、做飯、上廁所等，進而無法百分之百關注嬰兒的各種行為。而若此時嬰兒發生溢奶、物品遮蓋口鼻、自行翻身或站立等情形，將造成嬰兒處於危險情境中，而可能導致憾事發生。

國內外有許多為自動化監測嬰兒狀態之研究，主要包含兩種偵測方式：其一為使用感測器量測嬰兒之特定生理訊號，如：心率、呼吸頻率、體溫、身體位置或方向及嬰兒周圍之氣體濃度等，透過收集到的數值以判定所監測之嬰兒處於正常狀態與否；然而，使用此種監測方式具功能

單一性，若欲偵測其他生理訊號，則需增設更多不同種類的感測器，不僅可能影響嬰兒之活動，亦可能產生更多潛在的危險性，如：裝置纏繞嬰兒、孩童誤食裝置等。其二為透過電腦視覺偵測嬰兒影像，判定嬰兒是否處於危險狀態，而現有研究中多僅針對嬰兒之面部特徵或單一狀態進行偵測；然而，我們認為一張嬰兒影像包含了許多資訊得以應用，如：同時偵測嬰兒面部及姿勢等，則可透過影像偵測更廣泛的監測嬰兒不同危險情境。

因此，本論文開發出一透過嬰兒影像辨識其基礎姿勢與面部狀態，以監測嬰兒是否因姿勢不適當或面部遭異物遮擋，處於危險情境中而需警示照護者。此方法不僅擁有可監測多種不同危險情境之優點，亦可減少感測器式偵測將干擾嬰兒之缺點，且對於未來欲增加其他監測功能有良好的擴充性。

1.2 研究目的

本論文基於深度學習技術，利用 ResNet50 網路進行嬰兒動作及臉部遮擋之模型訓練，且以 DeepFace 演算法前處理影像擷取出嬰兒臉部畫面，而得以對嬰兒進行危險監測。

本研究預計達成以下目標：

- 針對嬰兒姿勢部分，辨識嬰兒之正躺、趴睡、坐姿及站立之四項基礎姿勢，進而判斷嬰兒是否做出具危險之動作。
- 針對嬰兒臉部部分，判斷嬰兒是否因嘔吐物、毛巾等非奶嘴之外物遮蓋其面部，而可能使嬰兒發生窒息危險。

綜上目標，本論文將建構出一可對嬰兒姿勢及臉部遮擋進行危險監測之系統。

1.3 論文架構

本論文分為五個章節，其架構如下：

第一章、緒論，敘述本論文之研究動機、研究目的及論文架構。

第二章、相關研究，敘述嬰兒猝死症之定義，並探討近年與嬰兒監測相關之研究以及深度學習模型架構與面部辨識網路。

第三章、研究方法，說明本研究之詳細內容，如：資料集之分類定義及前處理、以及完整系統之流程說明。

第四章、實驗設計與結果，說明各項實驗設計內容以及評估方法，並對於實驗結果進行探討。

第五章、結論與未來展望，對於研究結果進行總結，並討論研究的未來展望。

二、相關研究

2.1 嬰兒猝死症

嬰兒猝死症（The Sudden Infant Death Syndrome, 簡稱 SIDS）[2] 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。

2.2 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。

2.2.1 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 [3] 所開發的嬰兒感測背心（圖 2.1），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 [4] 開發了將感測器裝設於胸帶中（圖 2.2），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 [5] 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狀態（圖 2.3）、清醒狀態及警示狀態（圖 2.4）；



圖 2.1: 嬰兒多感測器背心之穿脫示意圖 [3]

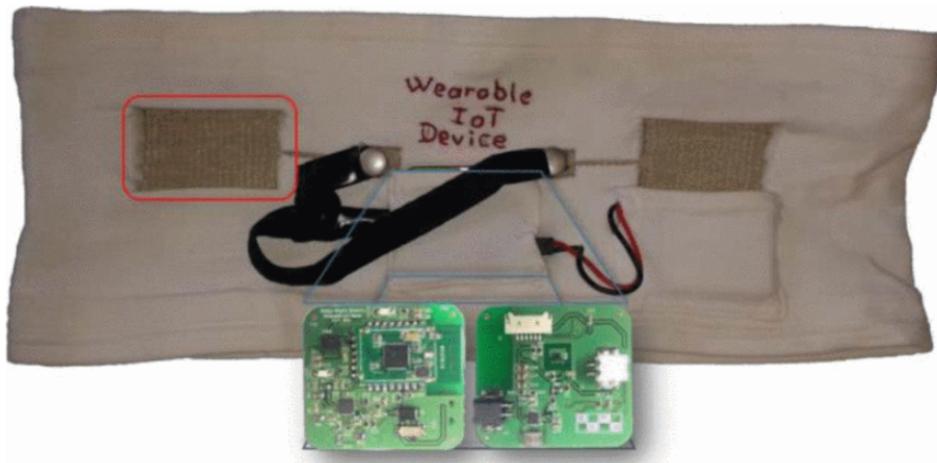


圖 2.2: 嬰兒穿戴式感測器裝置圖 [4]

Lin 等人 [6] 開發出一套在嬰兒胸帶上嵌入了三種不同感測器的系統（圖 2.5），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（圖 2.6）、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

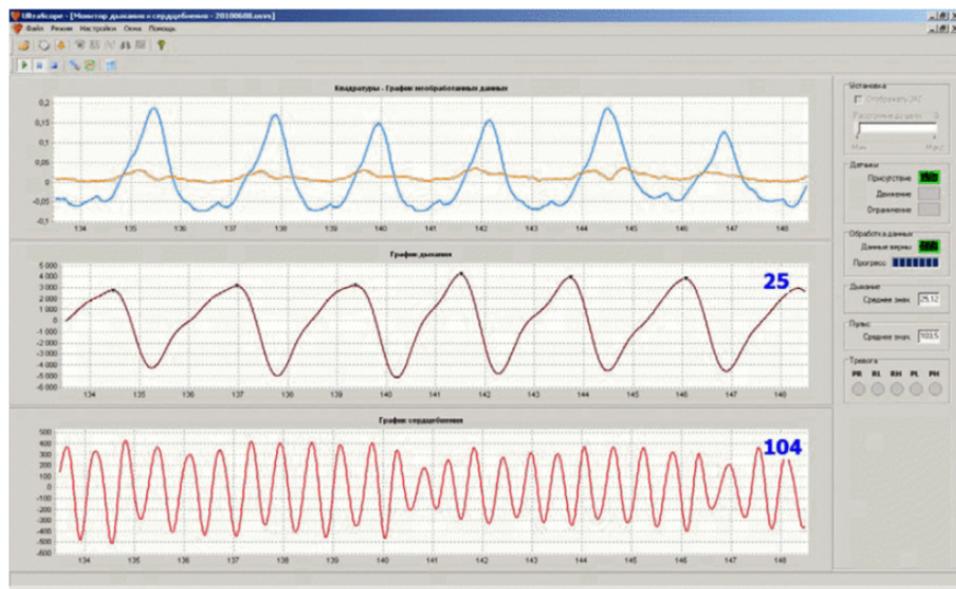


圖 2.3: 嬰兒正常狀態之呼吸及心跳圖 [5]

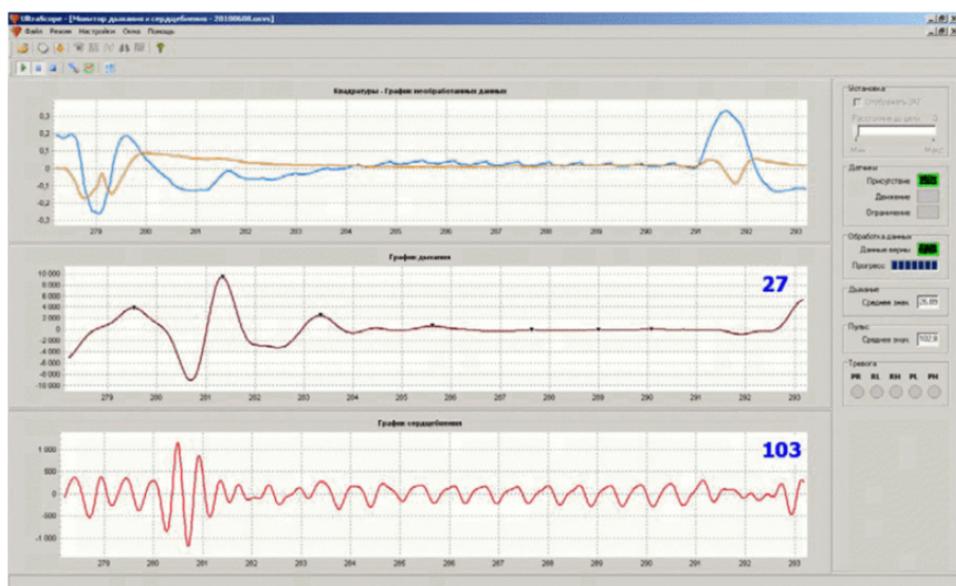


圖 2.4: 嬰兒呼吸暫停狀態之呼吸及心跳圖 [5]

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。

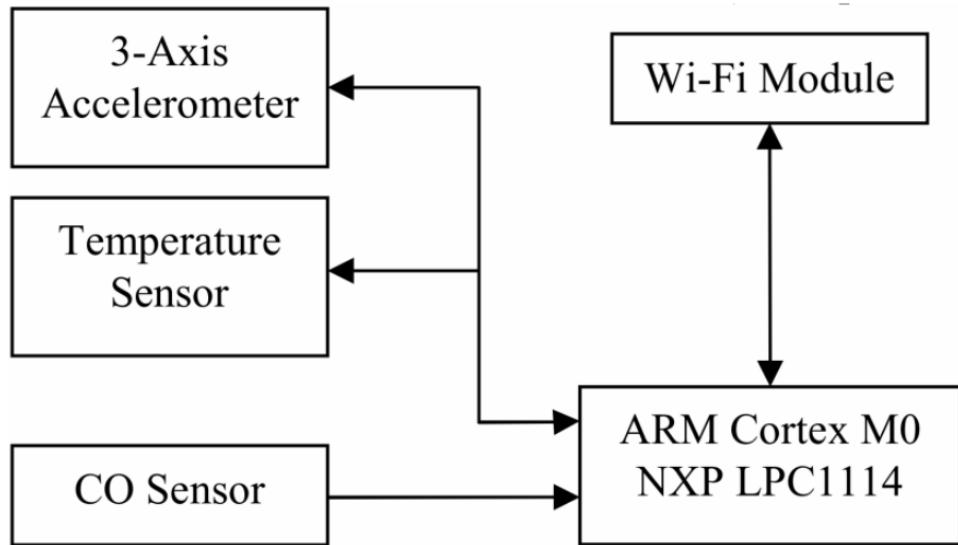


圖 2.5: 嬰兒監測系統架構圖 [6]

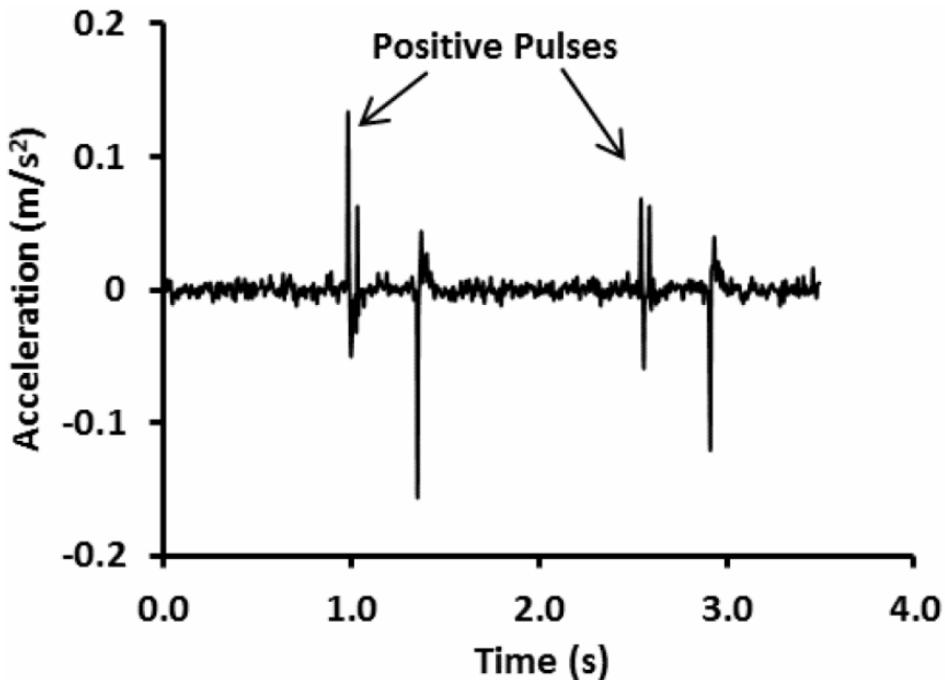


圖 2.6: z 軸顯示吸氣及呼氣脈衝 [6]

2.2.2 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 [7] 開發了一基於視覺之非接觸式呼吸頻率偵測系統，其

系統流程為（見圖 2.7）：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之輕微運動，其判斷流程圖見圖 2.8），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。

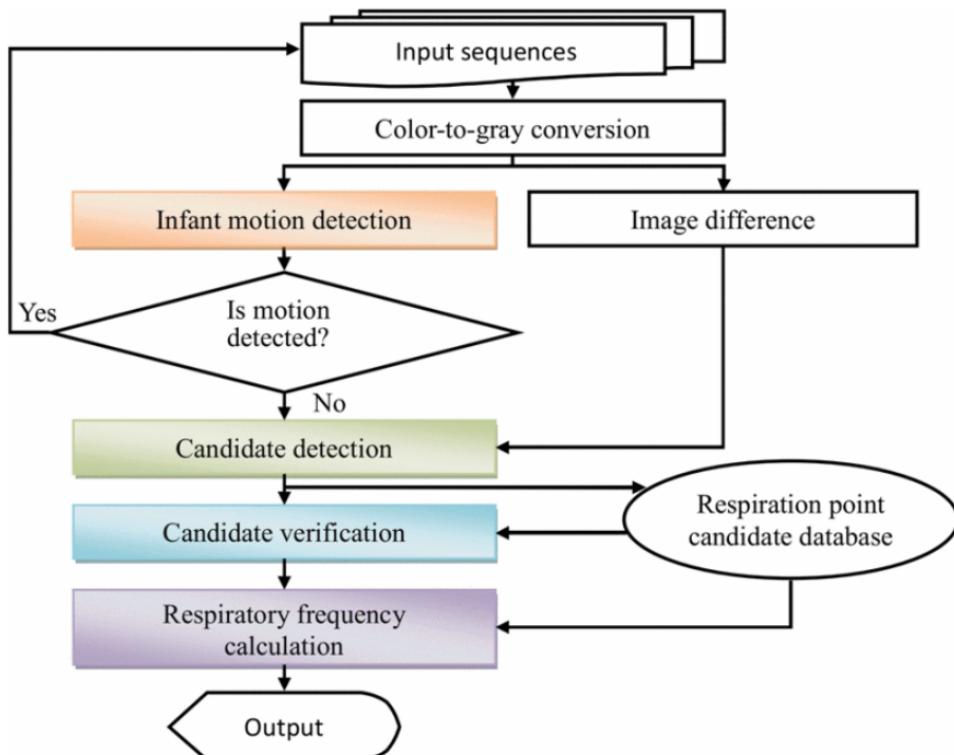


圖 2.7: 嬰兒呼吸頻率偵測系統流程圖 [7]

Liu 等人 [8] 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。系統整體設計圖見圖 2.9。

Gallo 等人 [9] 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。系統之硬體架構圖見

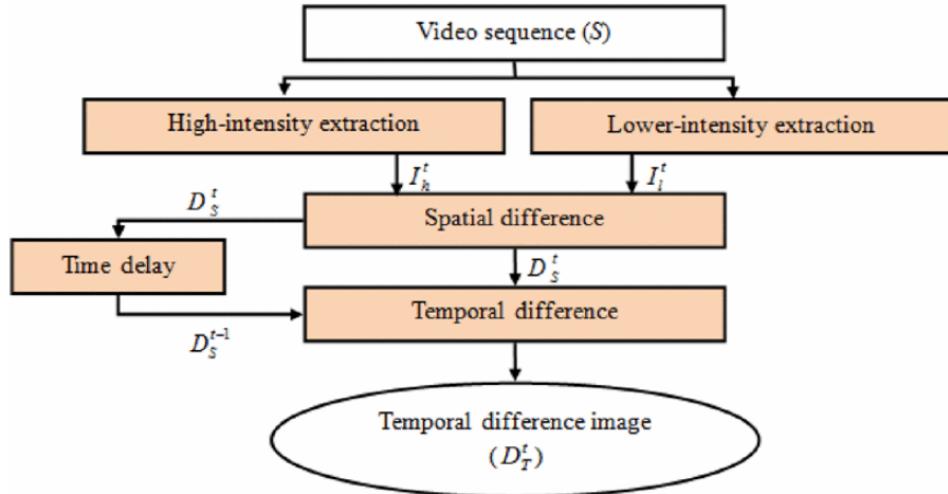


圖 2.8: 嬰兒運動偵測流程圖 [7]

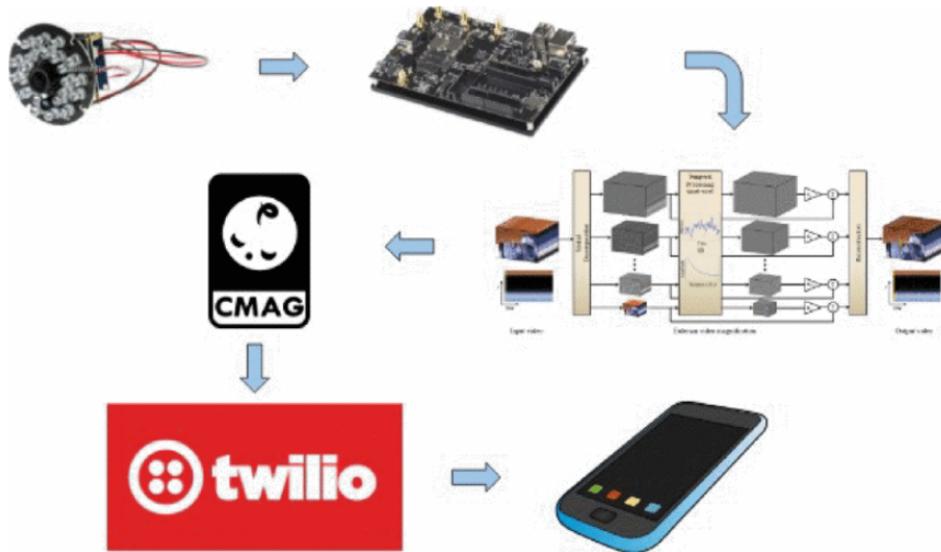


圖 2.9: 系統設計圖 [8]

圖 2.10，人工視覺模型架構見圖 2.11。

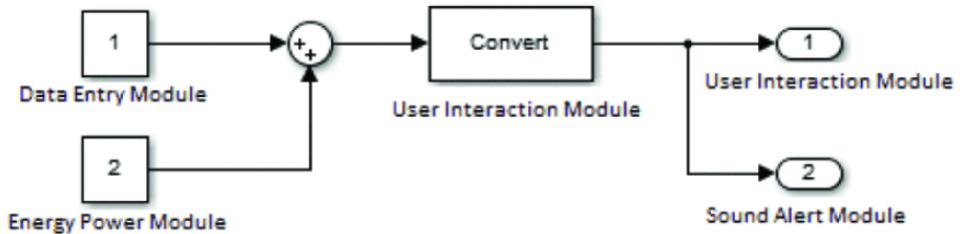


圖 2.10: 硬體架構圖 [9]

Wang 等人 [10] 提出了一個多任務貝氏深度神經架構（圖 2.12），其

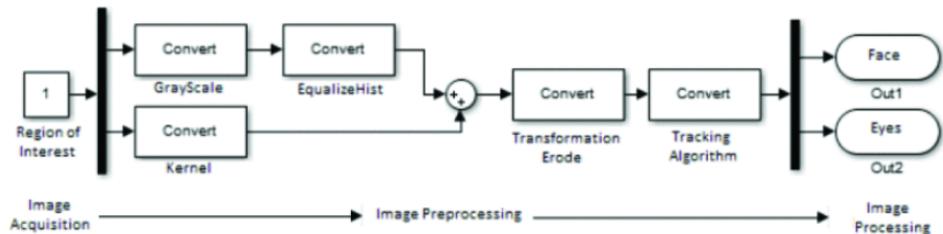


圖 2.11: 人工視覺模型架構圖 [9]

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。

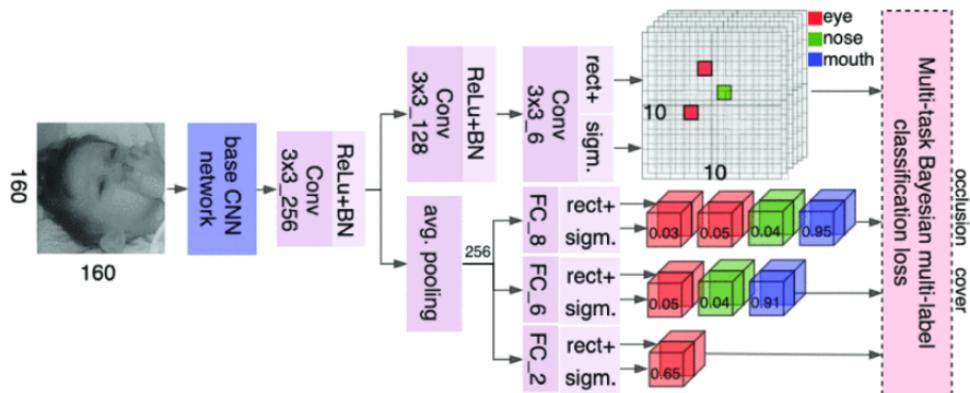


圖 2.12: 網路架構 [10]

Bharati 等人 [11] 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（圖 2.13）：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。此研究之 CNN 架構見圖 2.14，而系統完整架構見圖 2.15。

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。



圖 2.13: 睡姿分類 [11]

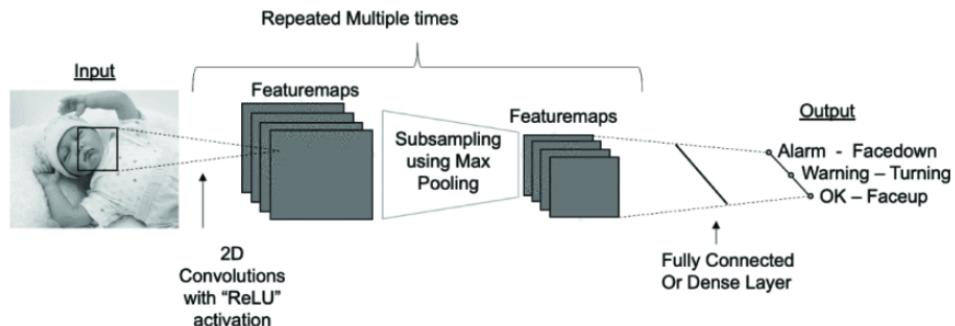


圖 2.14: CNN 架構 [11]

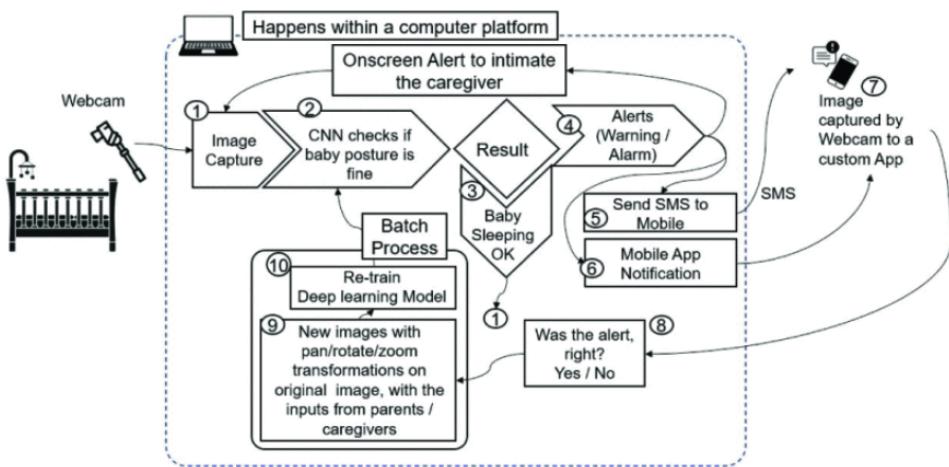


圖 2.15: 系統架構 [11]

2.3 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 2.16 可看到兩個不同層數的網路其訓練誤差值。

因此，He 等人 [12] 提出了一個深度殘差學習（圖 2.17）的架構，利用

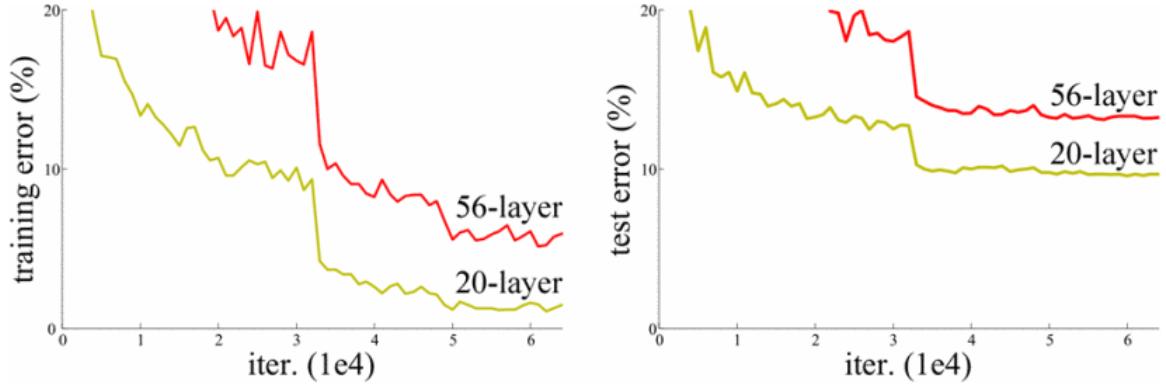


圖 2.16: 網路深度與訓練誤差關係

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。

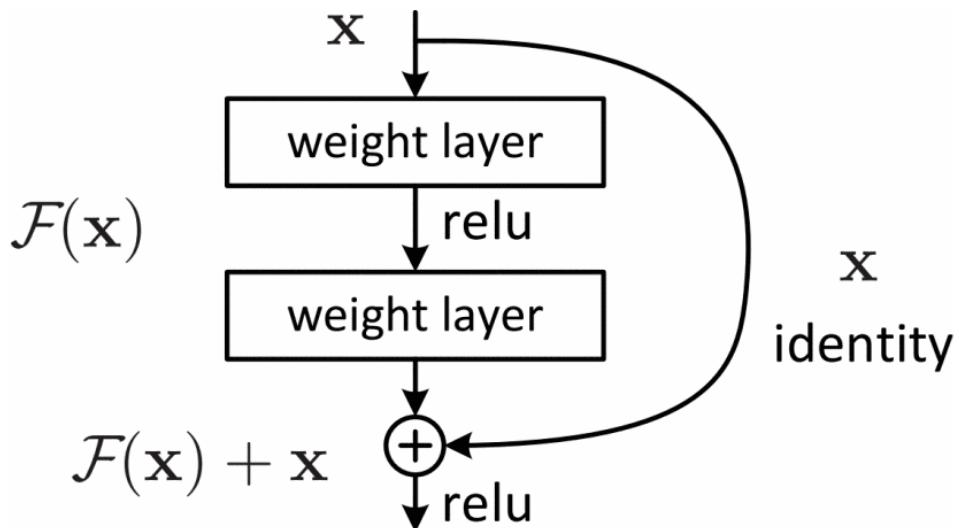


圖 2.17: 殘差學習 [12]

2.4 人臉偵測演算法

2.4.1 MTCNN

MTCNN [13] 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 2.18。

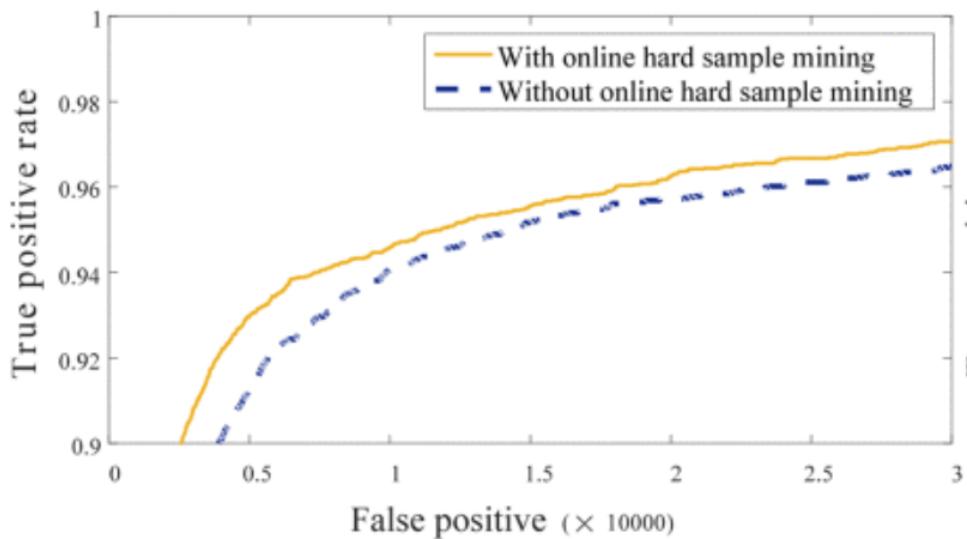


圖 2.18: 使用 online hard sample mining 策略效能比較 [13]

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 2.19：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 2.20，圖中”MP”為 max pooling、”Conv”為 convolution，而 pooling 及 convolution 的步長分別為 2 及 1。

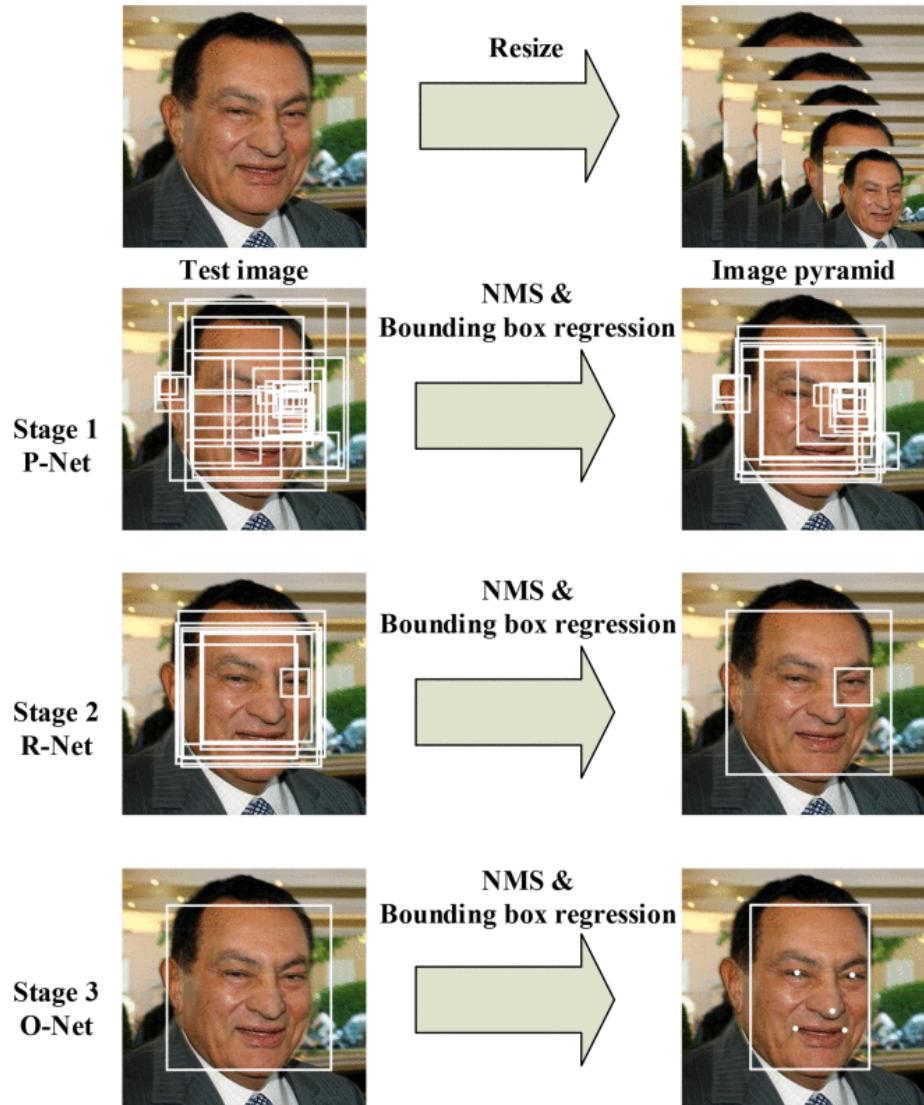


圖 2.19: MTCNN pipeline [13]

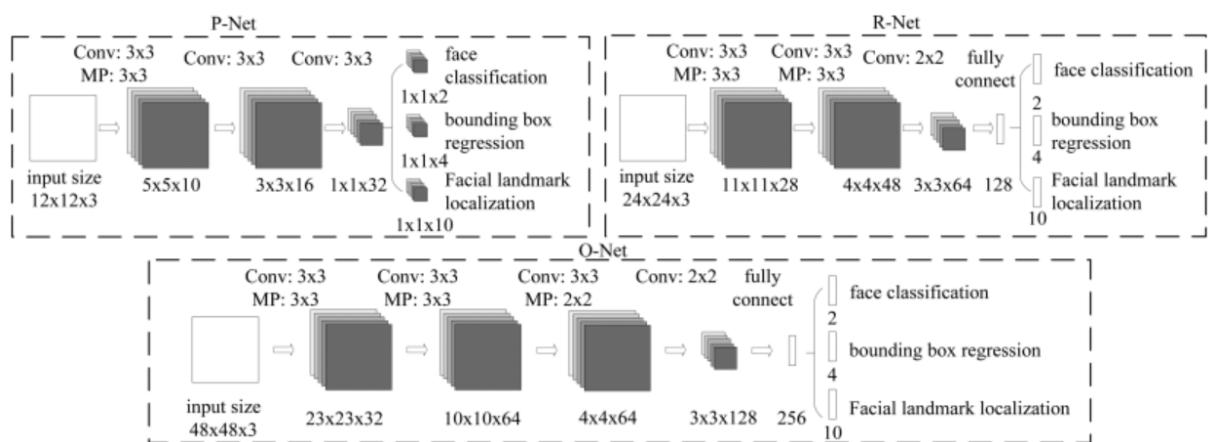


圖 2.20: MTCNN framework [13]

2.4.2 RetinaFace

RetinaFace [14] 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 2.21）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。

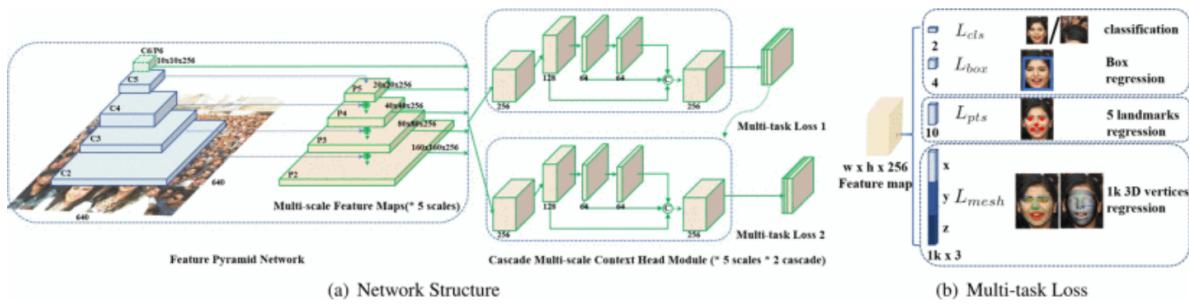


圖 2.21: RetinaFace 架構 [14]

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 2.22 所示，此演算法擁有 91.7% 的良好結果。

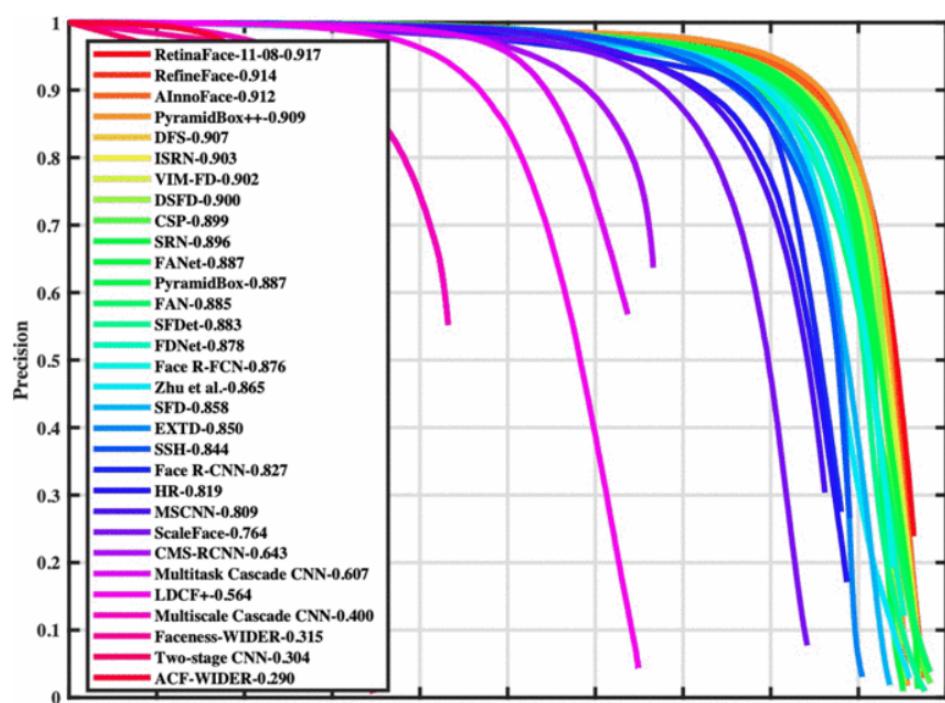


圖 2.22: RetinaFace (ResNet-152) 在 WIDER FACE 測試集之 Precision-Recall 曲線 [14]

三、研究方法

3.1 系統流程介紹

本論文所開發之嬰兒危險監測系統，其針對嬰兒影像畫面進行識別，以判斷嬰兒是否處於危險狀態，而須提醒照護者。

系統之完整流程為：首先，輸入一段待觀測之嬰兒影片，將影片切成數幀影像，並判斷影像存在與否，若不存在系統發出異常警告，反之則開始對該影像畫面進行危險偵測判斷。針對每幀嬰兒影像，系統針對其臉部遮擋及姿勢進行辨識，若透過模型分析為警示狀態，則再經後續步驟判斷是否提醒照護者；而若分析為安全狀態，則可接續下一幀之影像進行偵測。詳細之系統流程圖，請見圖 3.1。

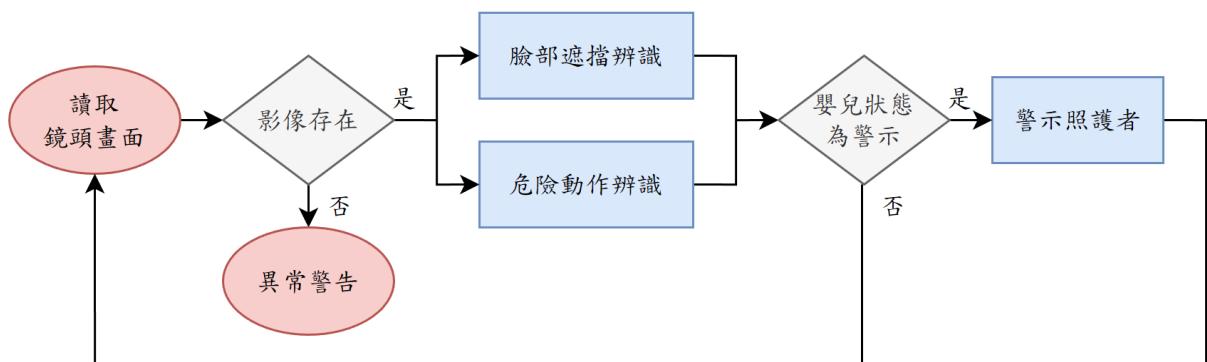


圖 3.1: 系統流程圖

而本系統中，包含兩項危險辨識模型：(1) 嬰兒臉部遮擋辨識：先將嬰兒畫面擷取出僅含臉部範圍之影像，再透過臉部遮擋模型判斷嬰兒臉部是否遭非奶嘴之異物遮蔽，若是，則嬰兒為警示狀態；(2) 嬰兒危險動作辨識：將嬰兒影像透過嬰兒姿勢模型進行辨識，判斷嬰兒為安全狀態：

正躺或坐姿，或為具危險性的狀態：趴躺或站立。而若兩模型結果皆為安全，則系統會判斷嬰兒狀態為安全，否則，嬰兒狀態則為警示。此二部分辨識之詳細方法，將於 3.2 節及 3.3 節進行介紹。

3.2 臉部遮擋辨識

如前言所述，目前醫界對於嬰兒猝死症之相關因素研究中，注意嬰兒臉部是否遭遮蔽，將有助於降低此症的發生。此外，亦有研究發現嬰兒使用奶嘴，對於預防嬰兒猝死症有幫助。

起初，基於電腦視覺及影像處理技術，例如：利用 Cb, Cr 色彩空間及 ellipse clustering [15] [16] [17] [18] 等偵測膚色，判斷嬰兒臉部是否出現非膚色之區塊，以進行臉部遮擋辨識，其效果如圖 3.2。

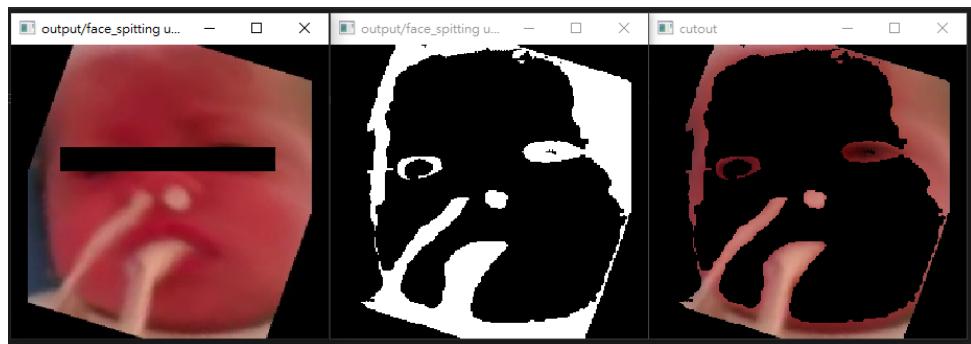


圖 3.2: 臉部膚色偵測 [18]

而後考量能有較佳的推廣性，因此，本研究改為使用深度學習技術，針對嬰兒臉部影像收集資料，進行臉部遮擋辨識模型之訓練，並且將嬰兒使用奶嘴之情形排除，亦即嬰兒臉部狀態共分為三種類別：(1) 面部無遮擋，為安全狀態、(2) 嬰兒正在使用奶嘴，亦為安全狀態、及 (3) 嬰兒面部遭嘔吐物或毛巾等外物遮蔽，為危險狀態，而需警示照護者。

本部分之流程圖，請見圖 3.3。

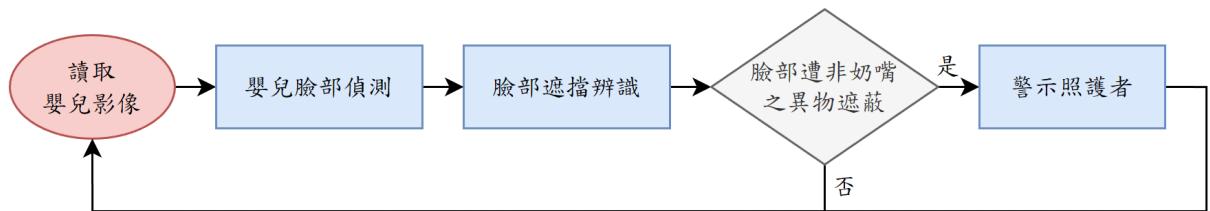


圖 3.3: 臉部遮擋辨識流程圖

3.2.1 嬰兒臉部偵測

嬰兒臉部遮擋辨識僅需關注嬰兒臉部畫面，故我們會先透過人臉偵測演算法進行前處理，以獲得只涵蓋嬰兒臉部範圍之影像。由於考量嬰兒臉部偵測之正確率及模型執行時間，因此，本研究選用 RetinaFace [14] 及 SSD [19] 等演算法進行此部分之處理。

3.2.2 嬰兒臉部資料集

本論文將嬰兒臉部狀態影像分為三類：面部無遮擋、有遮擋但遮蔽物為奶嘴及有遮擋且遮蔽物非奶嘴，前兩類判斷為安全狀態，最後一類則為危險狀態。對於臉部遮擋資料集之三類範例如下：

- (1) 臉部無遮蔽：嬰兒五官皆未被遮擋為安全狀態，如圖 3.4。
- (2) 臉部遮蔽物為奶嘴：嬰兒正在使用奶嘴為安全狀態，如圖 3.5。
- (3) 臉部遮蔽物非奶嘴：嬰兒臉部因溢奶遭嘔吐物遮蔽，或被毛巾等其他外物遮蓋，而可能造成窒息危險，如圖 3.6。

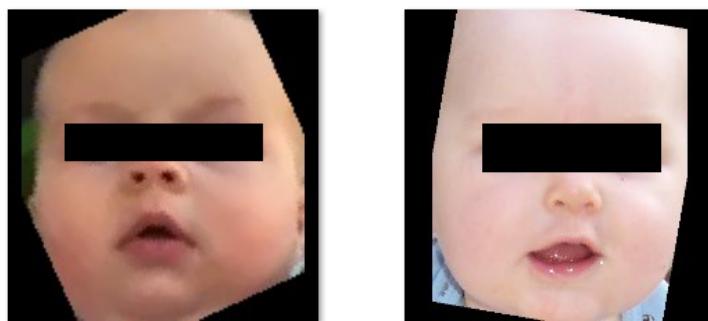


圖 3.4: 嬰兒臉部無遮蔽

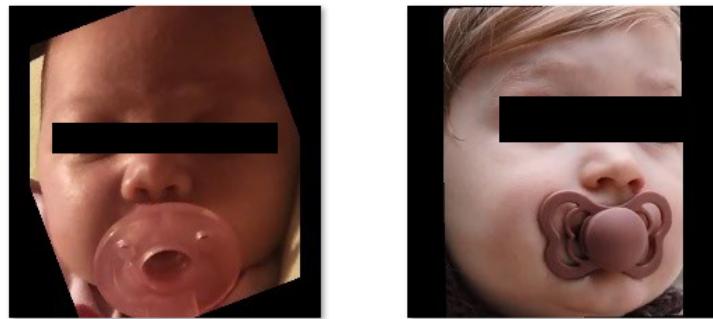


圖 3.5: 嬰兒臉部遮蔽物為奶嘴

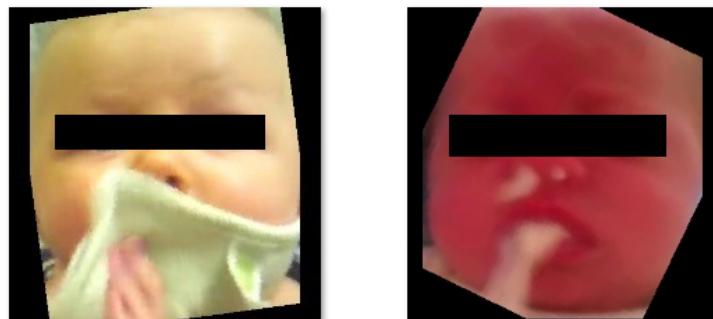


圖 3.6: 嬰兒臉部遭異物遮擋

此嬰兒臉部資料集包含嬰兒之正臉及側臉，共 3475 張照片。我們將所有影像分為訓練、測試及驗證集，各部分占比為 70%、20% 及 10%，即各有 2436 張、697 張及 342 張影像。

3.3 危險動作辨識

承前言所述，除了臉部遮蔽可能造成嬰兒猝死症外，嬰兒做出不適當的不適當也常為嬰兒逝世之原因。例如：嬰兒側躺或趴睡時，因頸部肌肉較弱等原因，無力自行將臉移開，造成呼吸困難而窒息死亡；或者當嬰兒自行站立，而有可能爬落嬰兒床等，亦可能使嬰兒處於危險情境中。

本部分之流程圖，請見圖 3.7。

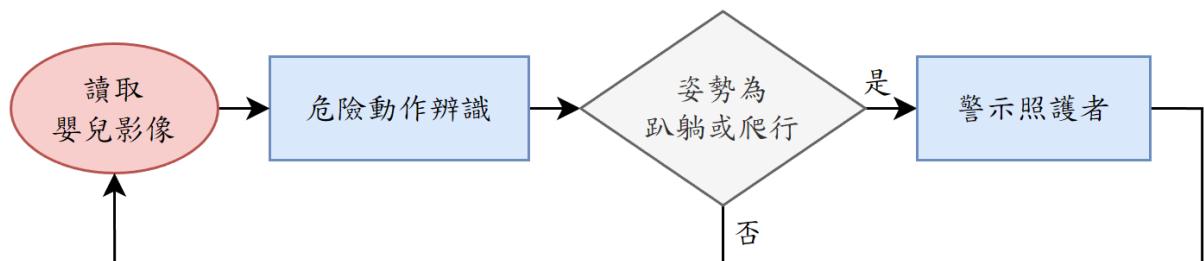


圖 3.7: 危險動作辨識流程圖

3.3.1 嬰兒姿勢資料集

起初，將嬰兒姿勢分為五類：正躺、趴著、爬行、坐姿及站立，而趴躺及爬行二類時常發生互相誤判，致使辨識錯誤率高。我們推測因為此二類嬰兒皆呈現腹面朝下之姿，而手腳位置亦有相同或相異之處，若接續細分姿勢，將導致動作分類過細。

因此，最終本論文將嬰兒姿勢分成基礎四類，包含正躺（腹面朝上）、趴躺（腹面朝下）、坐姿及站立，以供辨識嬰兒大部分之姿。對於此四類姿勢之詳細分類定義為：

(1) 正躺：嬰兒腹部面朝上，背部貼於水平面，而頭部及四肢位置不限，如 圖 3.8。

(2) 趴躺：嬰兒腹部面朝下，包含趴著或爬行等多動作，而頭部及四肢位置不限，如 圖 3.9。

(3) 坐姿：嬰兒臀部貼於水平面，而背部未貼於同一平面，頭部及四肢位置不限，如 圖 3.10。

(4) 站立：嬰兒腳掌貼於水平面，且腹部和背部皆未平行於此水平面，而頭部及上肢位置不限，如 圖 3.11。

而為了能有較廣泛的使用情境，所收集之嬰兒影像不限定拍攝視角，包含俯視、平視等，共 15416 張照片。我們將所有影像分為訓練、測試及驗證集，各部分占比為 70%、25% 及 5%，即各有 10815 張、3857 張及 744 張影像。

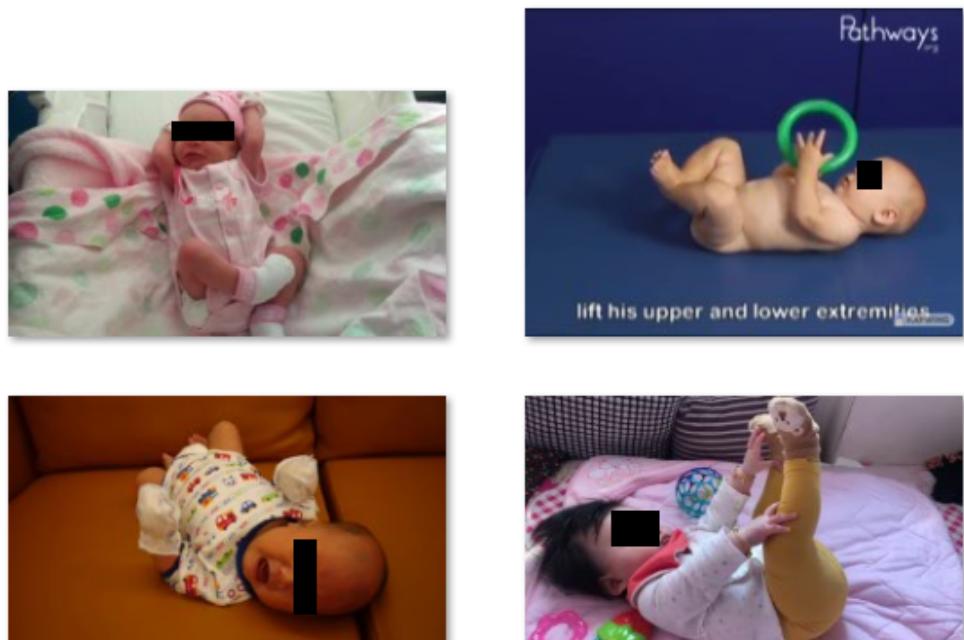


圖 3.8: 嬰兒正躺姿勢



圖 3.9: 嬰兒趴躺姿勢



圖 3.10: 嬰兒坐姿姿勢

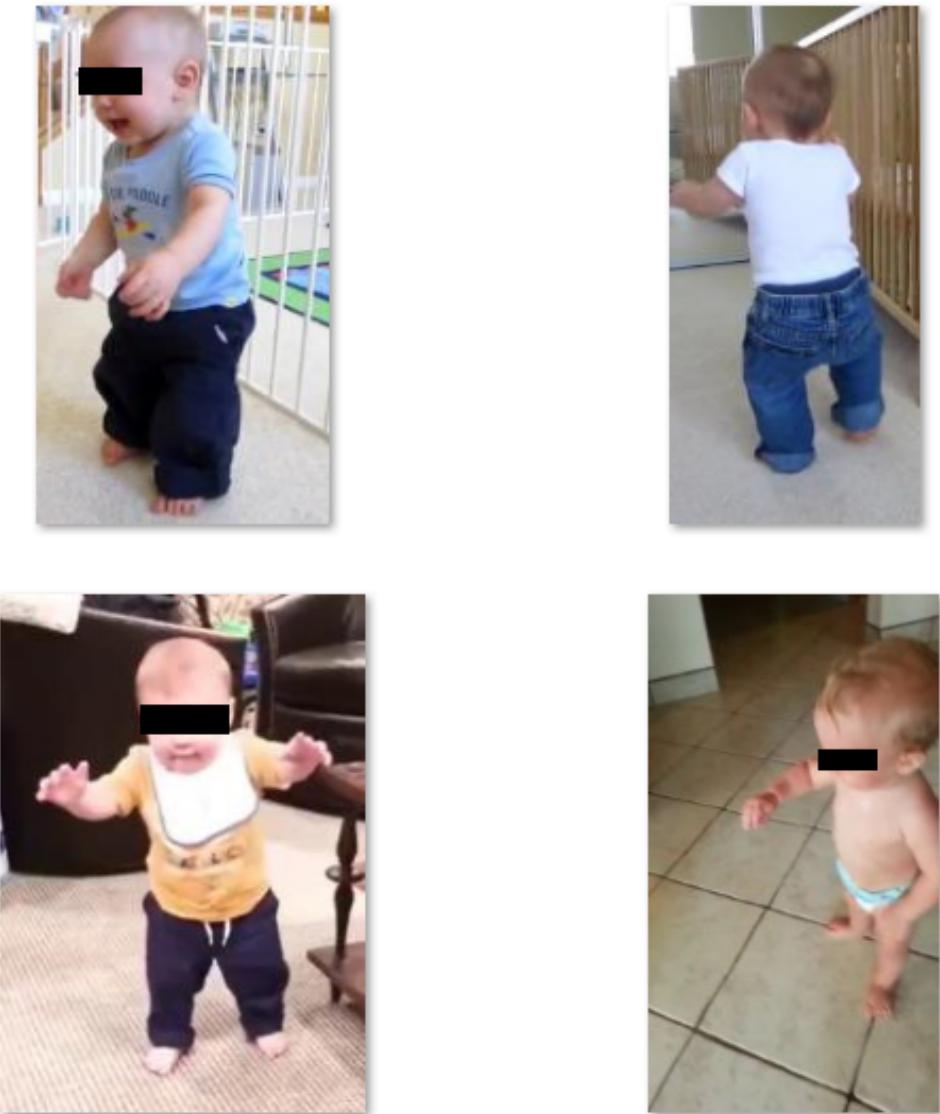


圖 3.11: 嬰兒站立姿勢

3.3.2 危險動作判斷方法

由於嬰兒做出趴躺及站立時，較容易發生危險，故當系統藉由模型辨識出嬰兒為上述兩種姿勢時，將警示照護者須關注嬰兒狀態。

然而，在實際情境中，當嬰兒做出具危險性之行為時，需持續一段時間才會導致危險的發生，故我們不須判斷一張影像畫面為警示狀態，就立即通知照護者。因此，本系統使用一變數累積結果為警示之幀數，當此變數超過一定值時，系統才會真正發出警示提醒照護者。此步驟不但更符合實際使用情境，同時亦可減少因模型辨識錯誤而誤判及誤發警報的情形。

四、 實驗設計與結果

4.1 嬰兒臉部偵測實驗

4.1.1 實驗目的與設計

在收集嬰兒臉部資料集時，需針對嬰兒影像擷取出臉部範圍，進而後續之臉部遮擋辨識階段。

因此，本實驗使用 3.3 節之嬰兒姿勢資料集，就 OpenCV [20] 、 SSD [19] 、MTCNN [13] 及 RetinaFace [14] 等臉部偵測演算法，分析其執行時間及臉部擷取準確度進行比較，以驗證適合本系統之演算法。

4.1.2 實驗評估方式

本實驗為驗證嬰兒臉部偵測演算法之實際可行性，將針對臉部偵測執行時間及偵測結果之準確度分別進行比較：透過計算演算法偵測所有資料集共 15416 張影像所花費之時間，得以算出各演算法平均每張需花費之時間；而準確度則將嬰兒臉部偵測之影像結果進行分類標註，分別計算出各演算法之 accuracy 、precision 及 recall 。

4.1.3 實驗結果與分析

首先，針對演算法之執行時間進行比較，透過實驗結果可得出使用 SSD 演算法進行嬰兒臉部偵測，將可擁有較佳的偵測速度。而四項演算

法偵測 15416 張影像之詳細實驗結果如下：

- (1) OpenCV 演算法：共花費 18 分 01.78 秒，平均每張影像需花 0.07 秒；
- (2) SSD 演算法：共花費 9 分 17.26 秒，平均每張影像需花 0.04 秒；
- (3) MTCNN 演算法：共花費 2 小時 8 分 22.05 秒，平均每張影像需花 0.50 秒；
- (4) RetinaFace 演算法：共花費 5 小時 42 分 2.10 秒，平均每張影像需花 1.33 秒。

接著，就偵測之精確度進行比較，透過實驗結果可得出選用 RetinaFace 演算法進行嬰兒臉部偵測，可擁有較佳的偵測準確度。而四項演算法進行嬰兒臉部偵測之詳細實驗結果如下：

- (1) 使用 OpenCV 演算法偵測結果如表 4.1，由於偵測效果不佳，將多數影像皆誤判為 False (無臉)，故僅計算其 precision 為 79.90%；
- (2) 使用 SSD 演算法偵測結果如表 4.2，由於偵測效果不佳，將多數影像皆誤判為 False (無臉)，故僅計算其 precision 為 99.90%；
- (3) 使用 MTCNN 演算法偵測結果如表 4.3，其 accuracy 為 90.20%、precision 為 94.76% 以及 recall 為 90.93%；
- (4) 使用 RetinaFace 演算法偵測結果如表 4.4，其 accuracy 為 99.78%、precision 為 99.75% 以及 recall 為 99.91%。

綜觀上述兩部分實驗結果，若系統欲擁有較迅速的執行速度又兼具偵測準確度，可得出以下結論：先使用 SSD 演算法找尋嬰兒臉部範圍，雖然此方法在許多狀況未能如期找到嬰兒臉部範圍，但其準確度很高，故能利用此算法之時間優勢；而若 SSD 演算法找不到嬰兒臉部時，則接續使用 RetinaFace 演算法，利用其很高之正確率及準確率之特質進行嬰兒臉部偵測。

表 4.1: OpenCV 演算法偵測嬰兒臉部結果

	True (預測有臉)	False (預測無臉)
True (實際有臉)	2882	11809
False (實際無臉)	725	

表 4.2: SSD 演算法偵測嬰兒臉部結果

	True (預測有臉)	False (預測無臉)
True (實際有臉)	4830	10581
False (實際無臉)	5	

表 4.3: MTCNN 演算法偵測嬰兒臉部結果

	True (預測有臉)	False (預測無臉)
True (實際有臉)	9361	994
False (實際無臉)	517	4544
Total	9878	5538

表 4.4: RetinaFace 演算法偵測嬰兒臉部結果

	True (預測有臉)	False (預測無臉)
True (實際有臉)	12925	11
False (實際無臉)	33	2447
Total	12958	2458

4.2 臉部遮擋辨識實驗

4.2.1 實驗目的與設計

本系統為偵測嬰兒臉部是否遭非奶嘴之異物遮擋，使用 3.2 節之資料集以 ResNet50 [12] 訓練模型，並透過驗證集進行模型驗證。

程式實作中，網路訓練回合數為 20，設定影像資料大小為 224x224，包含三個類別（臉部無遮擋之安全狀態、臉部遭奶嘴遮擋及臉部遭異物遮擋之危險狀態），且透過 data augmentation 技術生成訓練及測試資料，輸出層使用 softmax 作為激發函數，並使用 Adam 作為優化器且將學習率設為 0.000001 以進行微調。

4.2.2 實驗結果分析

本實驗訓練之模型其最終訓練準確率達 98.06%，而測試準確率達 99.43%，詳細訓練結果請見圖 4.1。

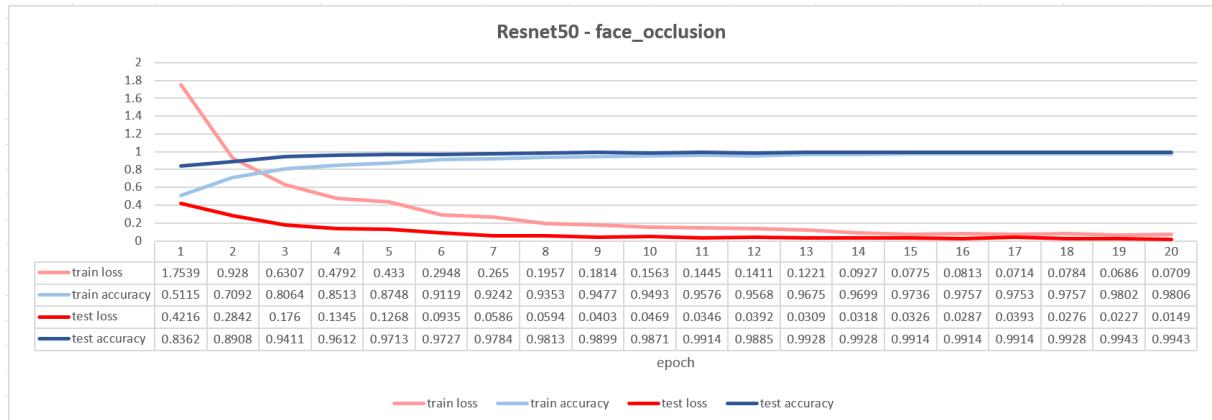


圖 4.1: 臉部辨識訓練結果

接著，再使用 342 張之驗證集影像進行模型驗證，所有影像皆辨識正確，其混淆矩陣如圖 4.2。

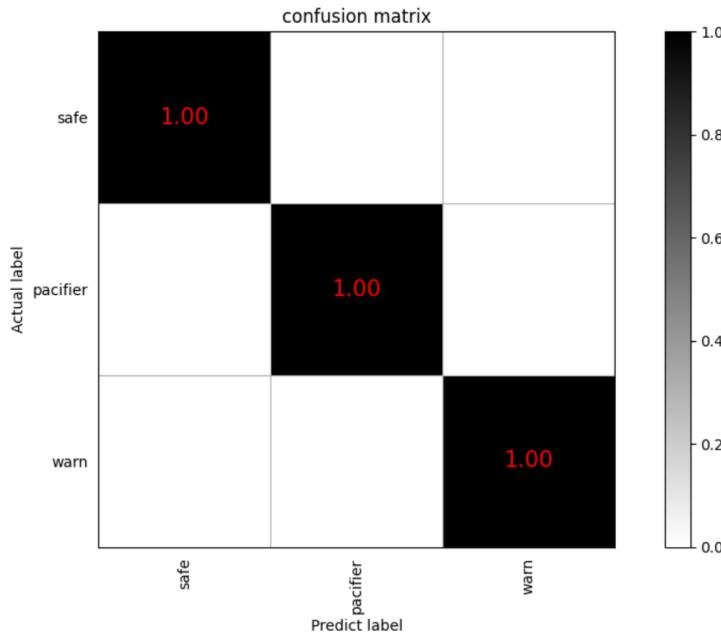


圖 4.2: 臉部遮擋辨識模型之混淆矩陣

4.3 姿勢辨識實驗

4.3.1 實驗目的與設計

本系統為辨識嬰兒姿勢是否處於危險狀態，使用 3.3 節之資料集以 ResNet50 [12] 訓練模型，並透過驗證集進行模型驗證。

程式實作中，網路訓練回合數為 20，設定影像資料大小為 224x224，包含四個類別（正躺、趴躺、坐姿及站立），且透過 data augmentation 技術生成訓練及測試資料，輸出層使用 softmax 作為激發函數，並使用 Adam 作為優化器且將學習率設為 0.000001 以進行微調。

4.3.2 實驗結果分析

本實驗訓練之模型其最終訓練準確率達 99.45%，而測試準確率達 99.71%，詳細訓練結果請見圖 4.3。

接著，再使用 744 張之驗證集影像進行模型驗證，包含了五張類別辨識錯誤的影像，其混淆矩陣如圖 4.4。辨識錯誤之五張影像中，有三張

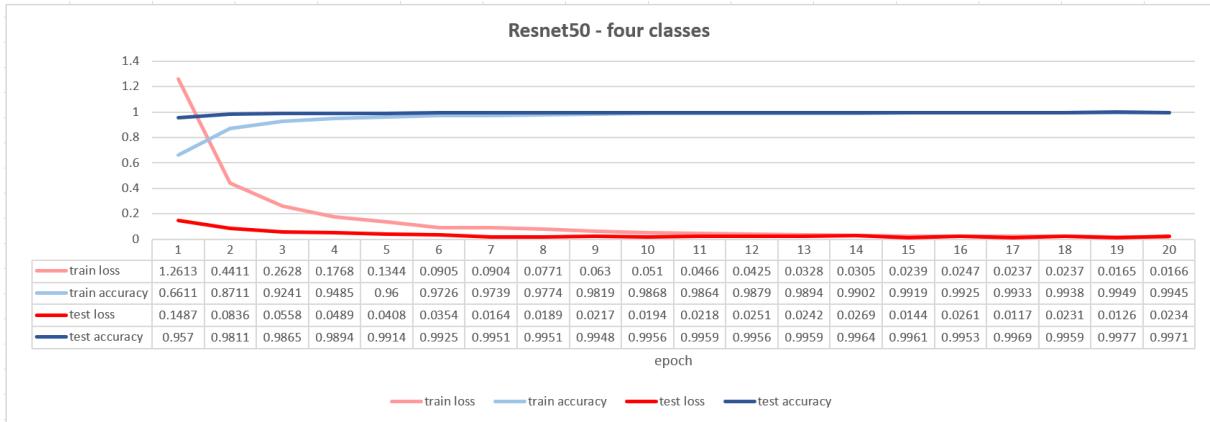


圖 4.3: 姿勢辨識訓練結果

將坐姿誤判為趴躺姿勢，推測原因為嬰兒雖呈現坐姿，但上半身貼近其腿部（如圖 4.5），而導致誤判。

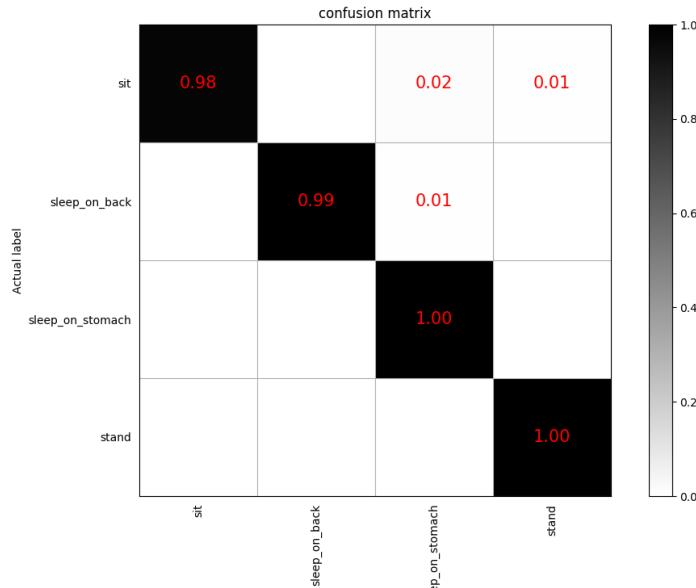


圖 4.4: 姿勢辨識模型之混淆矩陣



圖 4.5: 姿勢辨識錯誤之影像：坐姿誤判為趴躺

4.4 影片危險偵測實驗

4.4.1 實驗設計

影片危險偵測實驗
影片危險偵測實驗

4.4.2 實驗評估方式

影片危險偵測實驗
影片危險偵測實驗

4.4.3 實驗結果分析

影片危險偵測實驗
影片危險偵測實驗

五、結論與未來展望

5.1 結論

本論文基於深度學習技術，透過嬰兒影像畫面進行危險偵測，目前可進行兩大部分之偵測：(1) 嬰兒臉部遮擋辨識、及 (2) 嬰兒姿勢辨識，其訓練及測試準確率皆達 98% 以上。

本系統優於過往感測器式偵測之功能單一性及不便性，也不同於既有之影像式偵測僅關注嬰兒呼吸或單一動作之研究，提供了關注於嬰兒臉部及動作之危險監測系統，將有助於協助嬰兒照護者並降低嬰兒猝死症發生風險。

另外，由於目前未有公開之嬰兒資料集，故本論文使用之所有嬰兒影像，皆收集自網路上實際嬰兒之照片或影片擷取，再經前處理及分類標示而成。

5.2 未來展望

本論文中，目前僅辨識四項嬰兒姿勢，若在偵測姿勢時加入時間資訊，預期得以判斷更多嬰兒行為，如：翻身及爬行等動作，即可監測更多嬰兒危險情境；而除辨識嬰兒臉部遭異物遮蔽外，若加入偵測嬰兒之面部表情等其他資訊，亦可更詳盡監測嬰兒狀態，以提醒照護者；此外，亦可提供多嬰兒情境之危險偵測，則使用場景將可更廣泛。

而系統實作方面，預期未來可提供設定觀測之年齡區間，即可針對不同之特定年齡嬰幼兒警示其具危險性之動作，以達到更符合實際使用情境的危險偵測；亦可結合通訊社群軟體等，如：Line 或 Telegram 等，進行即時之推播訊息以通知嬰兒照顧者。

參考文獻

- [1] 統計處, 歷年統計, 中文, 文字, Mar. 2021.
- [2] H. C. Kinney and B. T. Thach, “The Sudden Infant Death Syndrome,” *The New England Journal of Medicine*, vol. 361, no. 8, pp. 795–805, Aug. 2009.
- [3] C. Linti, H. Horter, P. Osterreicher, and H. Planck, “Sensory baby vest for the monitoring of infants,” in *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*, Apr. 2006, 3 pp.–137.
- [4] A. G. Ferreira, D. Fernandes, S. Branco, *et al.*, “A smart wearable system for sudden infant death syndrome monitoring,” in *2016 IEEE International Conference on Industrial Technology (ICIT)*, Mar. 2016, pp. 1920–1925.
- [5] E. G. Ziganshin, M. A. Numerov, and S. A. Vygolov, “UWB Baby Monitor,” in *2010 5th International Conference on Ultrawideband and Ultrashort Impulse Signals*, Sep. 2010, pp. 159–161.
- [6] W. Lin, R. Zhang, J. Brittelli, and C. Lehmann, “Wireless Infant Monitoring Device for the prevention of sudden infant death syndrome,” in *2014 11th International Conference Expo on Emerging Technologies for a Smarter World (CEWIT)*, Oct. 2014, pp. 1–4.
- [7] C.-Y. Fang, H.-H. Hsieh, and S.-W. Chen, “A Vision-Based Infant Respiratory Frequency Detection System,” in *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Nov. 2015, pp. 1–8.
- [8] X. Liu, K. Takeuchi, T. Ogunfunmi, and S. Mathapathi, “Video-based IoT baby monitor for SIDS prevention,” in *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, Oct. 2017, pp. 1–7.
- [9] X. L. Gallo, S. Lechón, S. Mora, and D. Vallejo-Huang, “MARRSIDS: Monitoring Assistant to Reduce the Risk of Sudden Infant Death Syndrome,” in *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, Apr. 2019, pp. 1–4.
- [10] T.-J. Wang, J. Laaksonen, Y.-P. Liao, B.-Z. Wu, and S.-Y. Shen, “A Multi-Task Bayesian Deep Neural Net for Detecting Life-Threatening Infant Incidents From Head Images,” in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 3006–3010.

- [11] V. Bharati, “An Efficient Edge Deep Learning Computer Vision System to Prevent Sudden Infant Death Syndrome,” in *2021 IEEE International Conference on Smart Computing (SMARTCOMP)*, Aug. 2021, pp. 286–291.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778.
- [13] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks,” *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.
- [14] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, “RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 5202–5211.
- [15] H.-k. Tang and Z.-q. Feng, “Hand’s Skin Detection Based on Ellipse Clustering,” in *2008 International Symposium on Computer Science and Computational Technology*, vol. 2, Dec. 2008, pp. 758–761.
- [16] W. Li, Q. Yang, and X. He, “Face detection algorithm based on double ellipse skin model,” in *2011 IEEE 2nd International Conference on Software Engineering and Service Science*, Jul. 2011, pp. 335–339.
- [17] *Python opencv 肌色檢測的實現示例_ 程式設計_ 程式人生*.
- [18] WalkonNet, *Python opencv 肌色檢測的實現示例-WalkonNet*, zh-TW.
- [19] B. Ye, Y. Shi, H. Li, L. Li, and S. Tong, “Face SSD: A Real-time Face Detector based on SSD,” in *2021 40th Chinese Control Conference (CCC)*, Jul. 2021, pp. 8445–8450.
- [20] K. Goyal, K. Agarwal, and R. Kumar, “Face detection and tracking: Using OpenCV,” in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, vol. 1, Apr. 2017, pp. 474–478.