

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

# 基於深度學習之嬰兒危險監測系統

## 摘要

照護者在照顧嬰兒時，可能發生無法隨時關注其狀態的情形，使嬰兒因溢奶、翻身、趴睡等情形，致使呼吸不順而發生憾事。又因現有產品多為感測器式嬰兒偵測系統，功能單一且易干擾孩童；而既有的視覺式嬰兒偵測研究中，又多僅關注於呼吸頻率、面部特徵及單一動作，尚有許多值得探討之處。

因此，本論文提出基於深度學習技術，專注於嬰兒影像畫面之危險監測系統，包含兩大功能之偵測：(1) 臉部遮擋辨識：判斷嬰兒臉部是否遭非奶嘴之異物遮蔽，進而可能發生窒息危險、及(2) 姿勢辨識：分析嬰兒正躺、爬躺、坐姿及站立四種基礎姿勢，若為趴躺或站立之姿，則有可能發生呼吸不順或跌落床面等危險。綜上功能，當本系統讀取一段嬰兒影片後，可藉模型判斷嬰兒是否處於警示狀態，以提醒照護者。

本研究中，嬰兒臉部偵測部分，使用速度較快的 SSD 演算法，以及準確率較高的 RetinaFace 演算法，使整體系統在執行速度及準確度間達到平衡。而由於目前未有公開之嬰兒資料集，故本文收集網路真實嬰兒之不同視角圖片及影片，自製嬰兒臉部與姿勢資料集各 3475 張及 15416 張影像，再以 ResNet50 進行臉部遮擋辨識及姿勢辨識兩模型之訓練，其訓練及測試準確度皆達 99%。由此證明，本研究對於嬰兒危險監測系統具有良好的可用性及獨特性。

**關鍵字：**嬰兒危險監測，嬰兒臉部遮擋，嬰兒姿勢，深度學習，嬰兒猝死症

# A Deep-learning-based Danger Monitoring System for Infants

## Abstract

The babysitter may not focus on the status of the infant at any time. When unpredictable things happen to the baby, such as spitting up, rolling over, or sleeping on his stomach, the babysitter won't notice immediately. Most of the existing products are sensor-based infant detection systems, which are single-function and may disturb the movement of the baby. However, the existing vision-based infant detection studies only focus on breathing rate, facial features, and individual movements.

Therefore, this paper proposes a danger monitoring system based on deep learning technology. The system focuses on baby images and includes two major functions: (1) Facial Occlusion Recognition: Determine whether the infant's face is occluded by foreign objects, which may cause suffocation. (2) Posture Recognition: The four basic postures of infants are analyzed: lying on the back, lying on the stomach, sitting and standing. If the baby is lying on his stomach or standing, he may be at risk of breathing difficulties or falling off the bed. In summary, while monitoring the baby's video, the system can alert the babysitter when the infant is in an alarm state.

In this study, infant face detection uses the faster execution time SSD algorithm and the higher performance RetinaFace algorithm. With these algorithms, the system strikes a balance between execution speed and accuracy. There is

currently no open source infant dataset. Therefore, this paper collects real baby images and videos from different perspectives from the Internet to create an infant face dataset with 3475 images and an infant posture dataset with 15416 images. Then, two models of face occlusion recognition and posture recognition are trained using ResNet50, and the training and testing accuracy are 99%. This proves that this study has good utility and uniqueness for infant danger monitoring system.

**Keywords:** infant danger monitoring, infant face occlusion, infant posture, deep learning, sudden infant death syndrome

# 目錄

	頁次
<b>摘要</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>目錄</b>	<b>iv</b>
<b>一、 緒論</b>	<b>1</b>
1.1 研究動機 .....	1
1.2 研究目的 .....	2
1.3 論文架構 .....	3
<b>二、 相關研究</b>	<b>4</b>
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 .....	6
2.2.2 影像式偵測 .....	8
2.3 殘差神經網路 .....	11
2.4 人臉偵測 .....	13
<b>三、 研究方法</b>	<b>17</b>
3.1 嬰兒危險監測系統 .....	17
3.1.1 系統流程 .....	17
3.1.2 使用場域 .....	18

3.2 踝部遮擋辨識 .....	18
3.2.1 踝部偵測 .....	19
3.2.2 嬰兒臉部資料集 .....	20
3.2.3 模型訓練 .....	21
3.3 姿勢辨識 .....	21
3.3.1 嬰兒姿勢資料集 .....	23
3.3.2 模型訓練 .....	25
3.4 危險情境判斷 .....	25
<b>四、 實驗設計與結果</b>	<b>27</b>
4.1 臉部偵測準確度實驗 .....	27
4.1.1 實驗目的與設計 .....	28
4.1.2 實驗評估方式 .....	28
4.1.3 實驗結果與分析 .....	28
4.2 臉部偵測執行時間實驗 .....	30
4.2.1 實驗目的與設計 .....	30
4.2.2 實驗評估方式 .....	30
4.2.3 實驗結果與分析 .....	30
4.3 臍部遮擋辨識實驗 .....	31
4.3.1 實驗目的與設計 .....	31
4.3.2 實驗結果分析 .....	32
4.4 姿勢辨識實驗 .....	33
4.4.1 實驗目的與設計 .....	33
4.4.2 實驗結果分析 .....	33
4.5 影片危險偵測實驗 .....	34
4.5.1 實驗目的與設計 .....	34
4.5.2 實驗評估方式 .....	35
4.5.3 實驗結果分析 .....	35

<b>五、 結論與未來展望</b>	<b>38</b>
5.1 結論 .....	38
5.2 未來展望 .....	39
<b>參考文獻</b>	<b>40</b>

# 圖 目 錄

## 頁 次

2.1 Linti 嬰用感測背心之穿脫示意圖 [4] . . . . .	6
2.2 Ferreira 嬰用感測胸帶 [5] . . . . .	7
2.3 Lin 系統架構圖 [6] . . . . .	7
2.4 Ziganshin 系統之嬰兒呼吸及心跳圖 [7] . . . . .	7
2.5 Fang 系統流程圖 [8] . . . . .	8
2.6 Liu 系統設計圖 [9] . . . . .	9
2.7 MARRSIDS 模型架構圖 [10] . . . . .	9
2.8 Wang 系統之網路架構 [11] . . . . .	10
2.9 Bharati 系統之嬰兒睡眠姿勢分類 [12] . . . . .	10
2.10 Bharati 系統架構圖 [12] . . . . .	11
2.11 網路深度與訓練誤差關係 [16] . . . . .	12
2.12 殘差學習 [16] . . . . .	12
2.13 在 ImageNet 上訓練之誤差曲線圖 [16] . . . . .	13
2.14 在 CIFAK-10 上訓練之誤差曲線圖 [16] . . . . .	13
2.15 Cuimei 人臉偵測演算法之偵測架構 [17] . . . . .	14
2.16 Face SSD 之偵測架構 [18] . . . . .	14
2.17 MTCNN 架構圖 [19] . . . . .	15
2.18 MTCNN pipeline [19] . . . . .	15
2.19 RetinaFace 架構圖 [20] . . . . .	16

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	19
3.4	嬰兒臉部偵測結果	20
3.5	嬰兒臉部資料集	21
3.6	嬰兒平躺之骨架偵測結果	22
3.7	嬰兒趴躺之骨架偵測結果	22
3.8	不同視角之嬰兒骨架偵測結果	23
3.9	姿勢辨識流程圖	23
3.10	嬰兒腹面朝下之姿	24
3.11	嬰兒姿勢資料集	25
3.12	危險情境判斷流程圖	26
4.1	臉部辨識訓練及測試結果	32
4.2	姿勢辨識訓練及測試結果	33
4.3	姿勢辨識錯誤之影像：坐姿誤判為趴躺	34
4.4	姿勢辨識錯誤之影像：正躺誤判為趴躺	35
4.5	未偵測嬰兒臉部之影像	36
4.6	臉部遮擋誤判之為警示狀態	36

# 表目錄

## 頁次

4.1	RetinaFace [20] 偵測嬰兒臉部結果 . . . . .	28
4.2	MTCNN [19] 偵測嬰兒臉部結果 . . . . .	28
4.3	SSD [18] 偵測嬰兒臉部結果 . . . . .	29
4.4	Haar cascade [27] 偵測嬰兒臉部結果 . . . . .	29
4.5	人臉偵測準確度結果 . . . . .	30
4.6	人臉偵測執行時間結果 . . . . .	31
4.7	使用本文方法之人臉偵測執行時間結果 . . . . .	31
4.8	臉部遮擋辨識模型之混淆矩陣（單位：張 / 百分比） . . . .	32
4.9	姿勢辨識模型之混淆矩陣（單位：張 / 百分比） . . . .	34
4.10	實驗影片姿勢辨識之混淆矩陣（單位：張 / 百分比） . . . .	37
4.11	實驗影片臉部遮擋辨識之混淆矩陣（單位：張 / 百分比） . .	37

# 一、緒論

本章節中，將探討嬰兒因故逝世之相關現況及既有研究，並說明本論文應用於嬰兒危險監測之情境及目標，將透過以下三章子節進行分述：研究動機、研究目的及論文架構。

## 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。

當照護者視線離開嬰兒，如泡奶、如廁等情境，無法即時關注嬰兒狀態。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，將使嬰兒處於危險情境中，進而導致憾事發生。

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，則可透過影像進行更廣泛的嬰兒危險情境之監測。

因此，本論文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。此方法不僅可監測多種危險情境，也避免感測器影響嬰孩的問題發生，且對於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 進行嬰兒動作及臉部遮擋之模型訓練，且以 RetinaFace 等多種臉部偵測演算法進行嬰兒影像前處理以擷取其臉部畫面，對目標嬰兒進行危險監測。

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，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，包含完整系統之流程說明以及資料集之分類定義及前處理等。

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

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。

### 2.1 嬰兒猝死症

嬰兒猝死症（The Sudden Infant Death Syndrome, 簡稱 SIDS）[2] 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，但可統整出多項誘發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 [3]，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。

## 2.2 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。

## 2.2.1 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 [4] 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 2.1。



圖 2.1: Linti 嬰用感測背心之穿脫示意圖 [4]

Ferreira 等人 [5] 將心律感測器、3D 加速度計、熱電堆感測器裝設於胸帶中（圖 2.2），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 [6] 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 2.3，

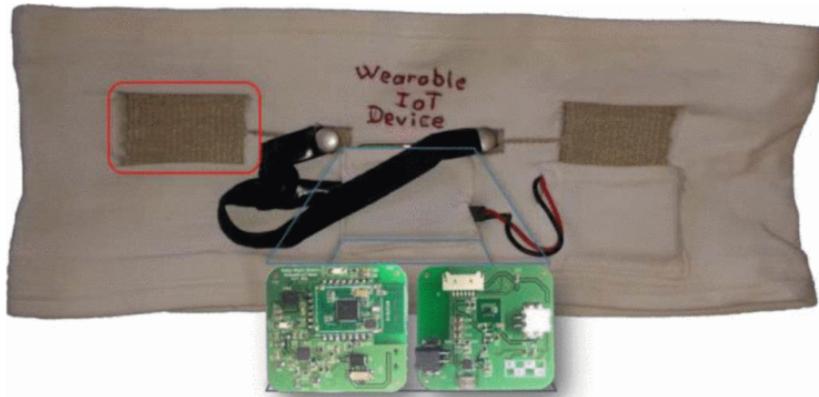


圖 2.2: Ferreira 嬰用感測胸帶 [5]

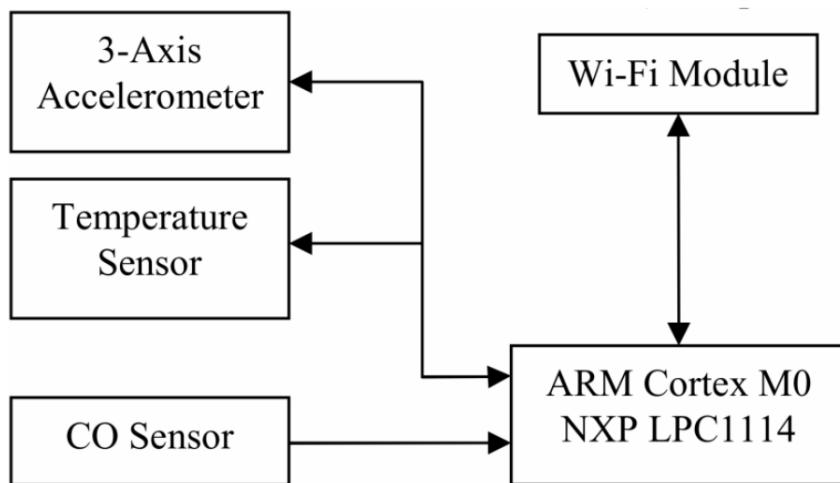


圖 2.3: Lin 系統架構圖 [6]

Ziganshin 等人 [7] 基於超寬頻技術開發出可監測呼吸及心率之系統（如圖 2.4），以檢測嬰兒三種狀態：睡眠（無運動）、清醒（長期大幅度運動）及警報（偵測到呼吸頻率異常低或暫停）。

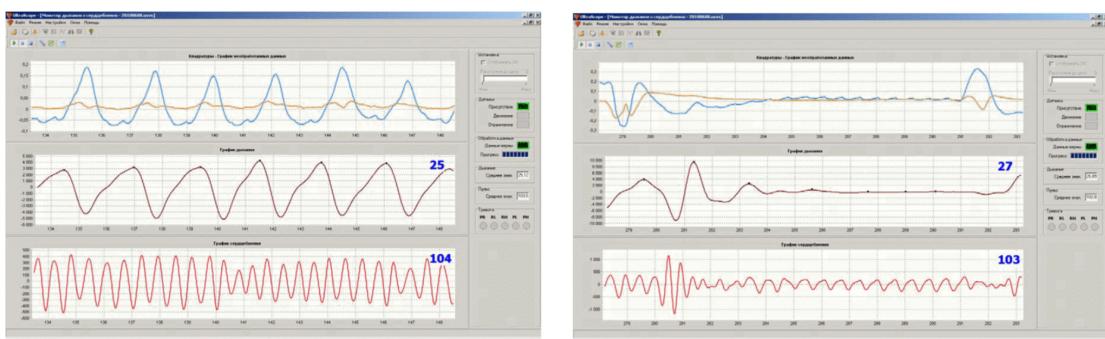


圖 2.4: Ziganshin 系統之嬰兒呼吸及心跳圖 [7]

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。

## 2.2.2 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 [8] 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 2.5a：先進行嬰兒運動偵測，其判斷流程見圖 2.5b，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。

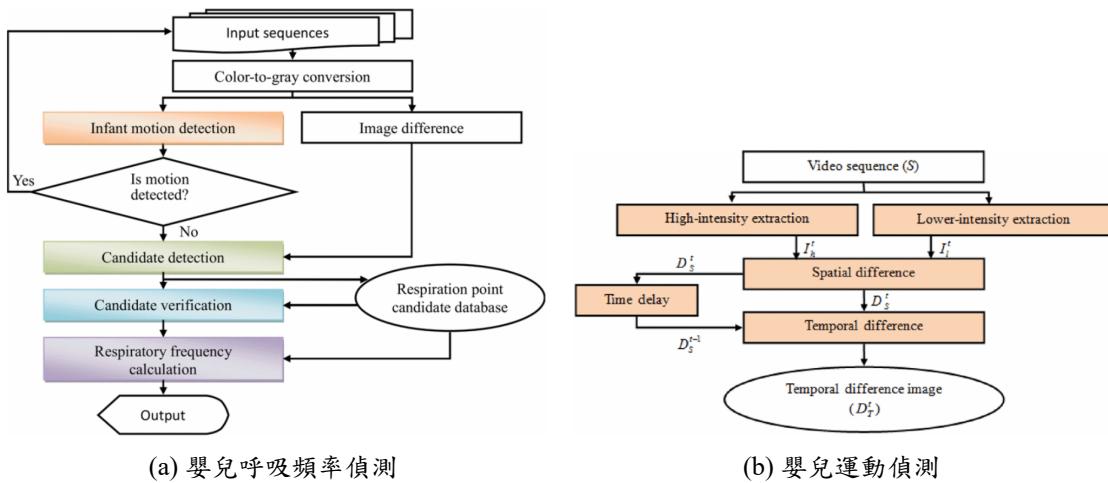


圖 2.5: Fang 系統流程圖 [8]

Liu 等人 [9] 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 2.6 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 2.6 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 2.6 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。

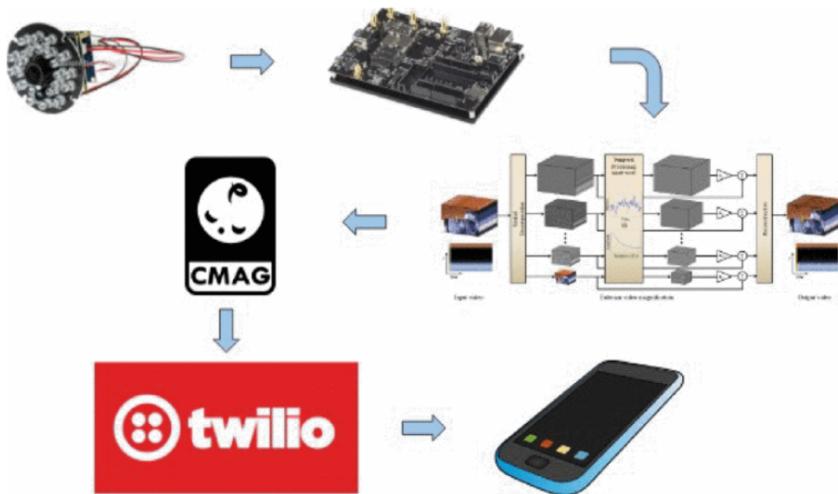


圖 2.6: Liu 系統設計圖 [9]

Gallo 等人 [10] 提出名為 MARRSIDS 的模型（架構如圖 2.7），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，需發出警示；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。

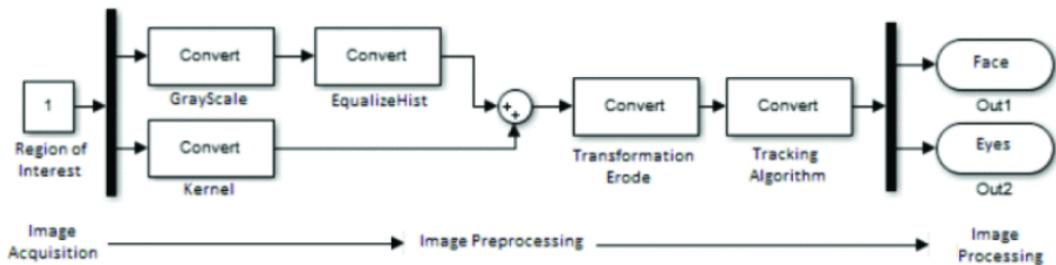


圖 2.7: MARRSIDS 模型架構圖 [10]

Wang 等人 [11] 提出用以偵測嬰兒臉部遮擋的貝氏深度神經網路架構（圖 2.8），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。

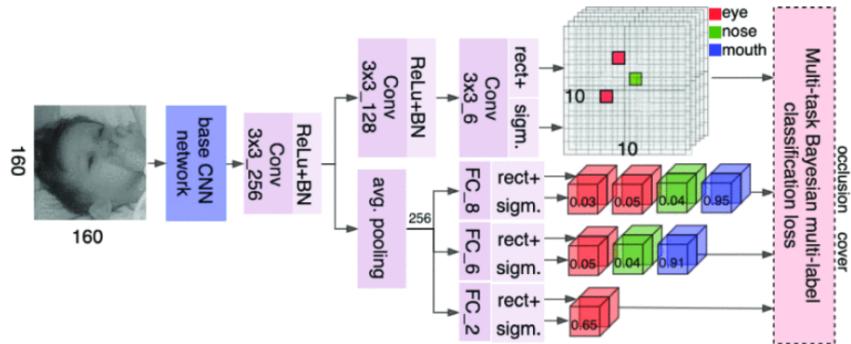


圖 2.8: Wang 系統之網路架構 [11]

Bharati 等人 [12] 開發基於卷積神經網路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 2.9），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 2.10a，其中卷積神經網路部分之架構如圖 2.10b：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 2.9: Bharati 系統之嬰兒睡眠姿勢分類 [12]

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。

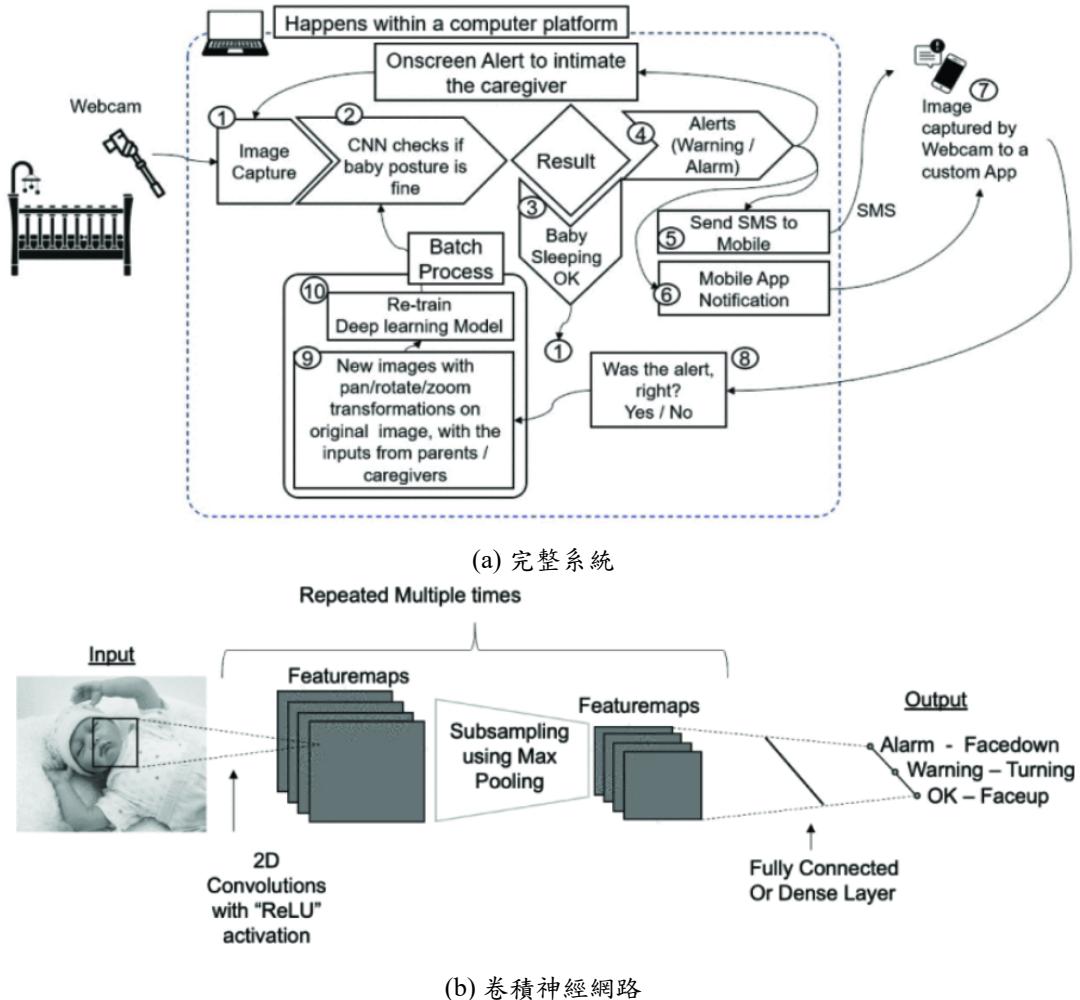


圖 2.10: Bharati 系統架構圖 [12]

## 2.3 殘差神經網路

既有研究中 [13]–[15]，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 2.11 可看到兩個不同層數的網路其訓練及測試誤差值。

因此，He 等人 [16] 提出使用深度殘差學習 (圖 2.12) 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。

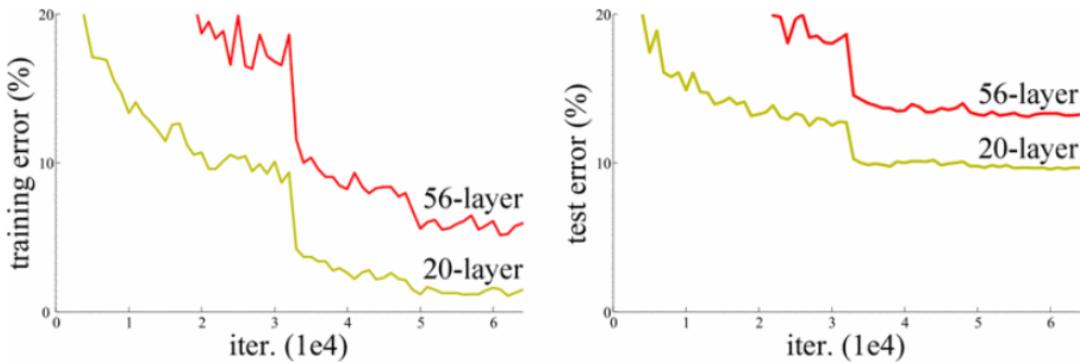


圖 2.11: 網路深度與訓練誤差關係 [16]

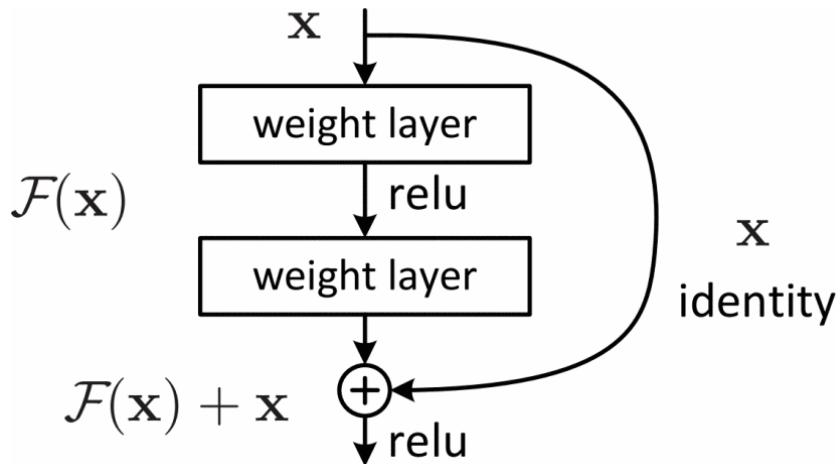


圖 2.12: 殘差學習 [16]

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 2.13（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 2.13（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 2.14（左）為普通網路不同層數之誤差曲線圖，圖 2.14（中）為殘差網路不同層數其誤差曲線圖，圖 2.14（右）則為使用 110 層及 1202 層殘差網路之誤差曲線圖。

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。

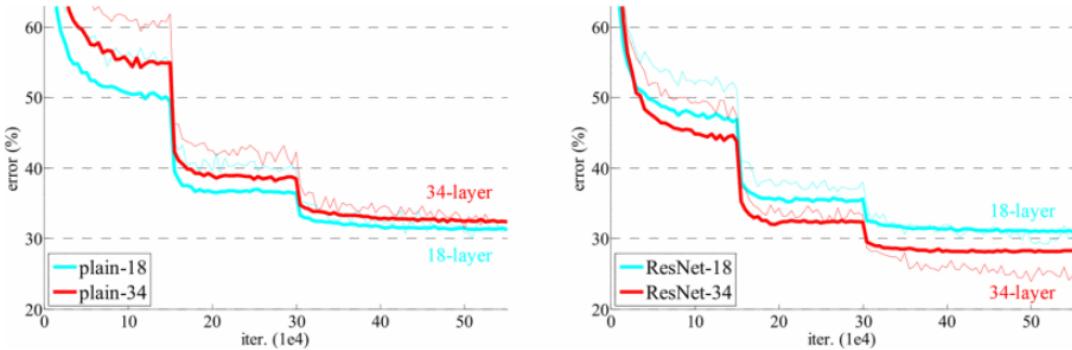


圖 2.13: 在 ImageNet 上訓練之誤差曲線圖 [16]

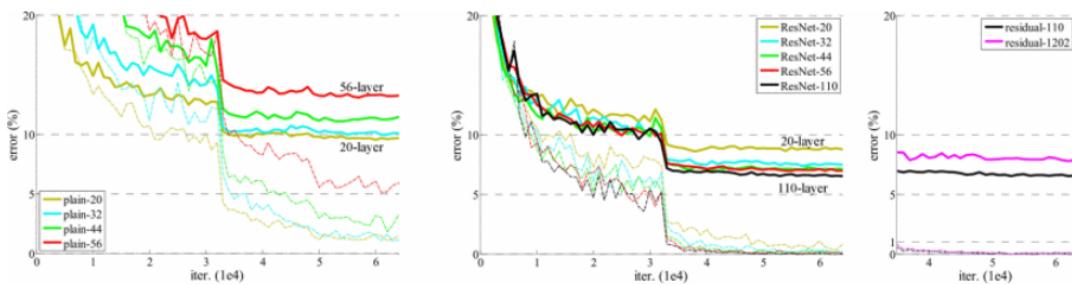


圖 2.14: 在 CIFAR-10 上訓練之誤差曲線圖 [16]

## 2.4 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 [17]，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，如圖 2.15 所示。

Face SSD [18] 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

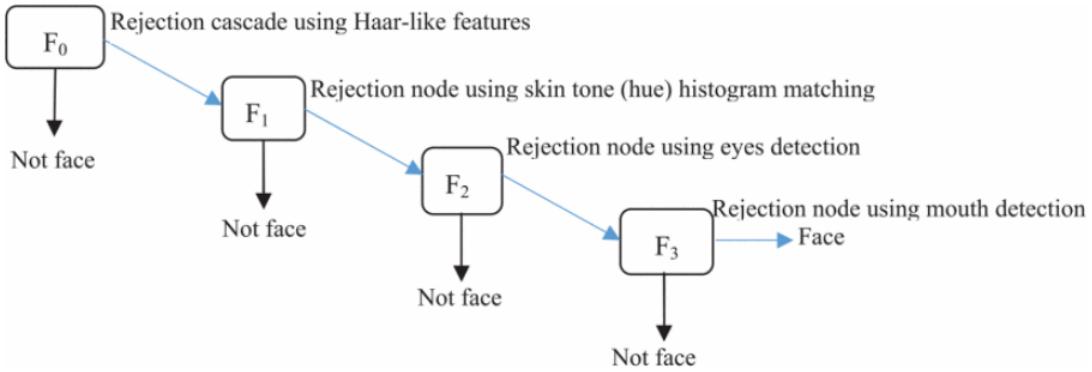


圖 2.15: Cuimei 人臉偵測演算法之偵測架構 [17]

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圖 2.16，並透過實驗得到在 CPU 及 GPU 上運行分別為每秒 40 帀以及每秒 110 帀的結果。

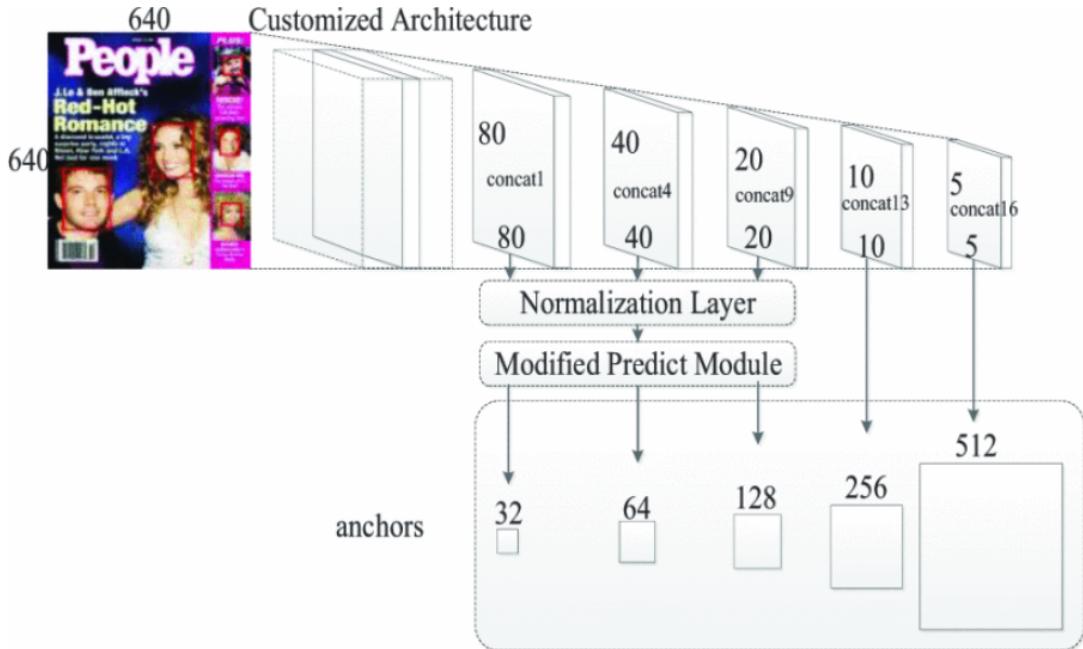


圖 2.16: Face SSD 之偵測架構 [18]

MTCNN [19]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 2.17）：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 2.18，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。

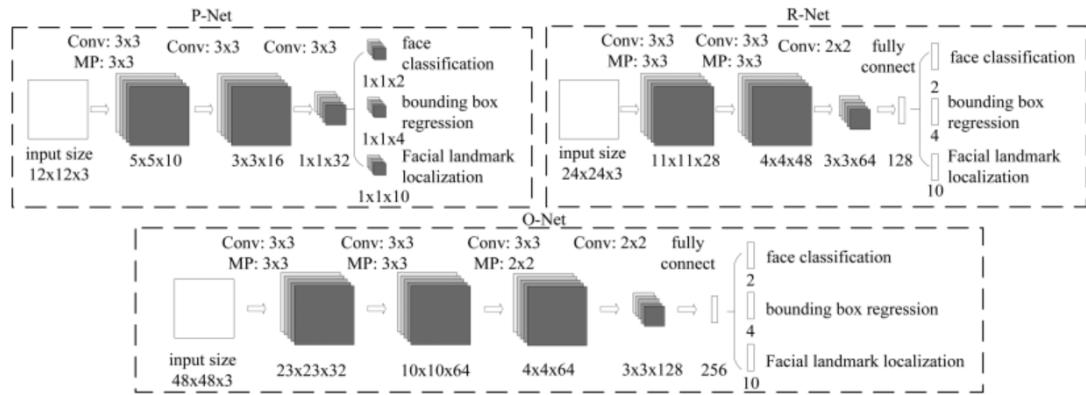


圖 2.17: MTCNN 架構圖 [19]

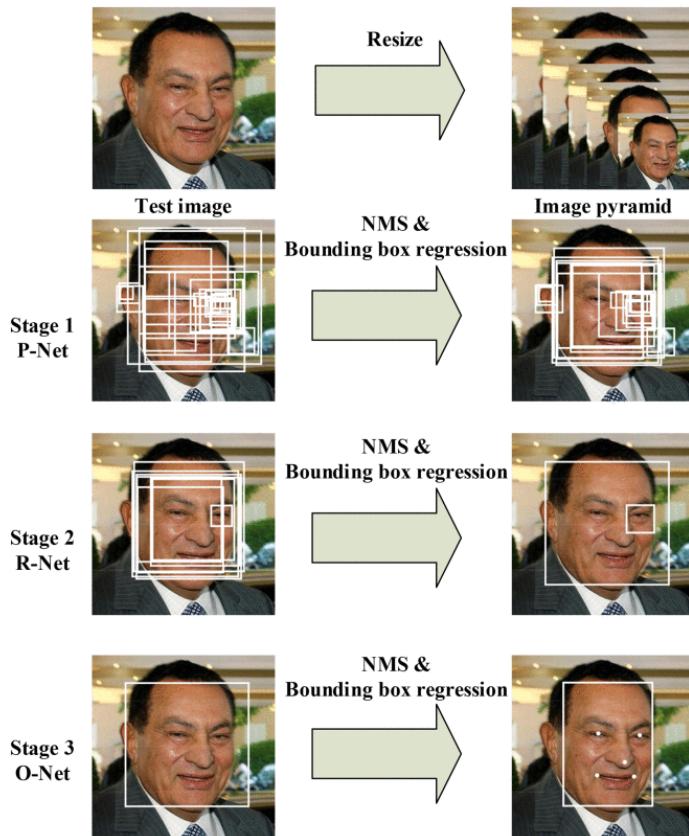


圖 2.18: MTCNN pipeline [19]

RetinaFace [20] 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 2.19）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。

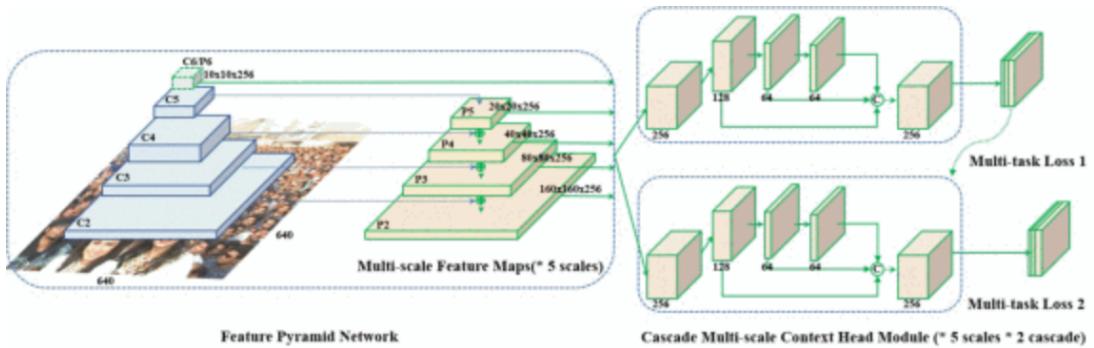


圖 2.19: RetinaFace 架構圖 [20]

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。

## 三、研究方法

本章節中，將闡述本文所開發之嬰兒危險監測系統，與其兩項核心辨識功能，透過以下四章子節進行說明：嬰兒危險監測系統、臉部遮擋辨識、姿勢辨識及危險情境判斷方法。

### 3.1 嬰兒危險監測系統

本研究針對嬰兒影像畫面進行偵測，判斷其臉部遮擋與否或姿勢是否不當，而處於危險情境中。

#### 3.1.1 系統流程

本系統之完整流程如圖 3.1所示：首先，讀取一段待觀測之嬰兒影片，將影片切成影像以進行後續危險偵測。針對每幀嬰兒畫面，系統對其臉部遮擋及姿勢進行辨識，分別輸出模型判斷之結果。接著，若分析嬰兒為警示狀態，則進行危險情境判斷，以決定系統是否發出警報。最後，檢查所觀測之影片結束與否，若尚未結束，則同前述步驟接續進行偵測。

而系統包含之兩項辨識功能及危險情境判斷，將分別於3.2節、3.3節及3.4節進行詳細的介紹。

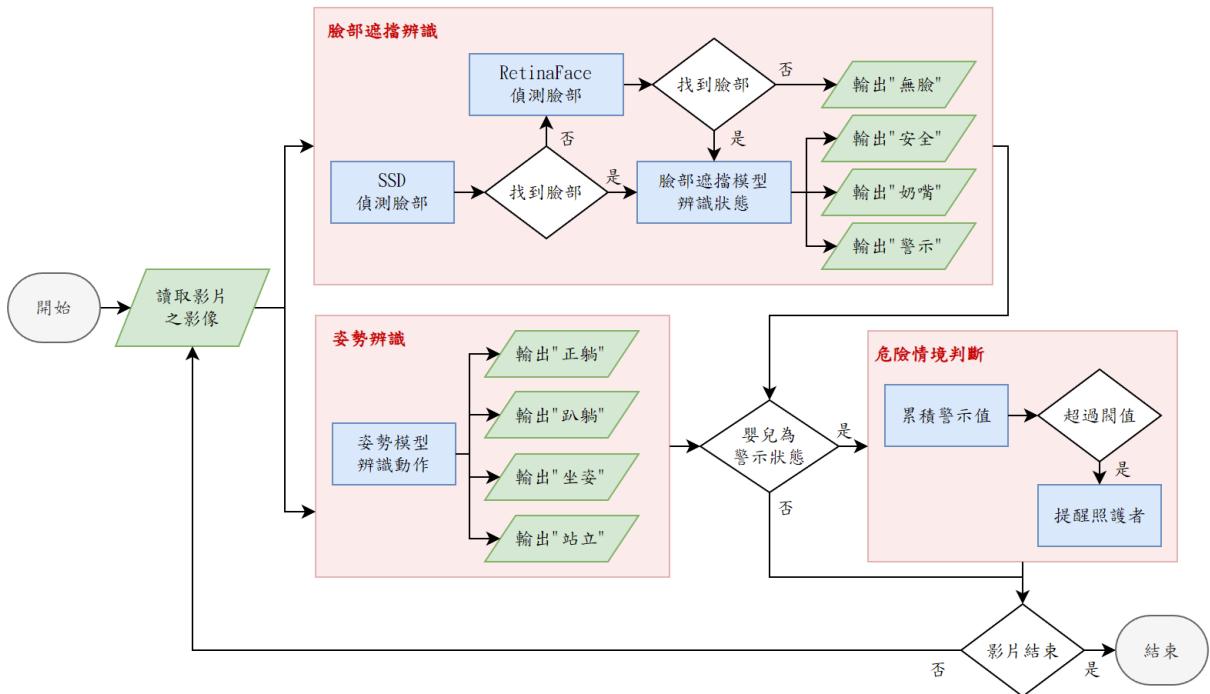


圖 3.1: 系統流程圖

### 3.1.2 使用場域

本研究中，目前僅針對單一嬰兒之情境進行辨識。而系統所讀取的影片可為俯視、平視等不同視角之畫面，不限定需架設於嬰兒床上方或房間某處，故當嬰兒在較大空間之環境活動時，可同時透過不同視角進行危險監測。另外，建議嬰兒佔據畫面比例一半以上，且穿著之服飾及背景環境顏色與膚色相異度較大，則能有較佳的辨識結果。

## 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指出，注意嬰兒臉部是否遭遮蔽，將有助於降低此症的發生；另亦有研究發現嬰兒使用奶嘴，對於預防嬰兒猝死症有幫助。因此，嬰兒使用奶嘴之情境，將不列入本文對於臉部遮擋的定義。

起初，基於電腦視覺及影像處理技術，例如：利用  $C_b, C_r$  色彩空間及 ellipse clustering [21]–[24] 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。

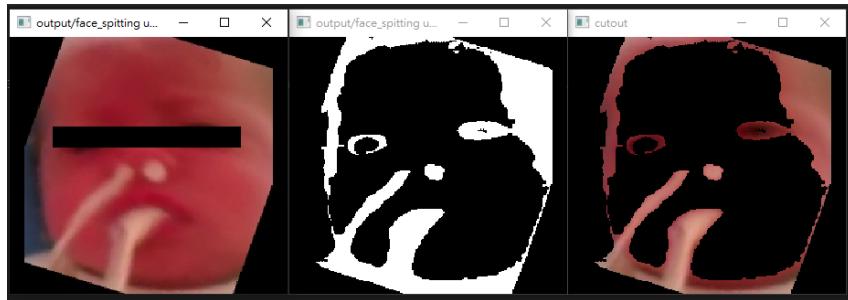


圖 3.2: 臉部膚色偵測

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進行臉部遮擋辨識，針對嬰兒面部影像自製資料集，以訓練可辨識三種嬰兒臉部狀態之模型。而資料集之詳細內容將於3.2.2節進行介紹。

本部分之流程如圖 3.3：首先，讀取嬰兒影像，並透過 SSD 演算法 [18] 偵測臉部，若未找到面部，則再以 RetinaFace 演算法 [20] 偵測之，若仍未能找到嬰兒臉部，則輸出「無臉」之結果；接著，將透過人臉偵測演算法擷取出之臉部畫面，使用本文訓練的臉部遮擋模型辨識面部遮蔽情形，並輸出判斷結果；若輸出為「安全」或「奶嘴」，此部分判斷嬰兒為安全狀態，而若結果為「警示」，則分析嬰兒為警示狀態。

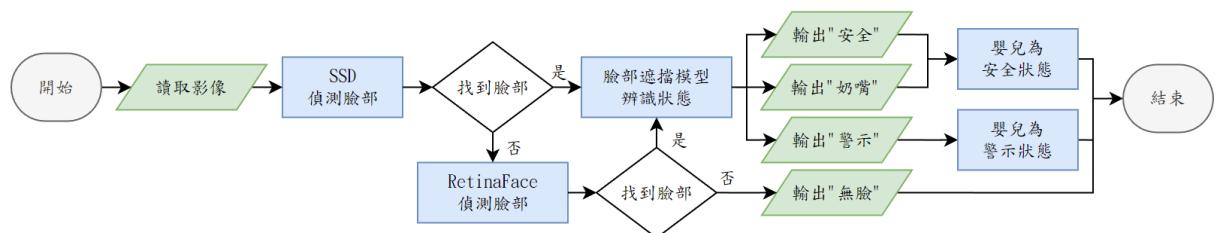


圖 3.3: 臉部遮擋辨識流程圖

### 3.2.1 臉部偵測

由於本研究中之臉部遮擋辨識僅需關注嬰兒臉部畫面，故本文會先透過人臉偵測演算法進行前處理，以獲得只涵蓋嬰兒面部範圍之影像。經多方實驗現有人臉偵測演算法後，同時考量嬰兒臉部偵測之正確率及

執行時間，最終本研究選用 SSD 及 RetinaFace 等演算法進行人臉偵測，其效果如圖 3.4。



圖 3.4: 嬰兒臉部偵測結果

### 3.2.2 嬰兒臉部資料集

由於目前未有公開之嬰兒資料集，故本論文使用之嬰兒影像，皆收集自網路上真實嬰兒的彩色照片或影片，再經前處理及分類標示而成。

本部分資料集將嬰兒臉部狀態分為三類，分別為臉部無遮蔽、臉部遮蔽物為奶嘴及臉部遮蔽物非奶嘴，各類別含 1197 張、1146 張及 1132 張，總共 3475 張影像。而三類詳細定義如下：

1. 臉部無遮蔽：嬰兒五官皆未被遮擋，為安全狀態，如圖 3.5a。
2. 臘部遮蔽物為奶嘴：嬰兒正在使用奶嘴，為安全狀態，如圖 3.5b。
3. 臍部遮蔽物非奶嘴：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，為警示狀態，如圖 3.5c。

並將完整資料集分成訓練、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。

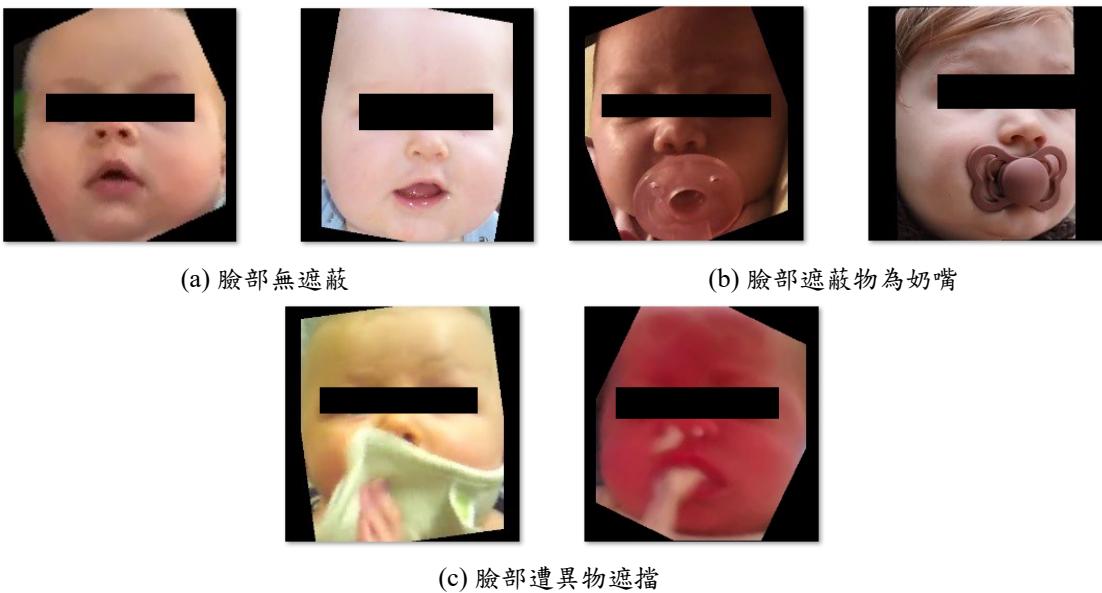


圖 3.5: 嬰兒臉部資料集

### 3.2.3 模型訓練

本論文使用3.2.2節之嬰兒臉部資料集，以 ResNet50 [16] 進行臉部遮擋辨識模型之訓練，最終達成辨識三種嬰兒臉部狀態：安全、使用奶嘴及警示。

## 3.3 姿勢辨識

承前言所述，除了臉部遮蔽可能造成嬰兒猝死症外，嬰兒做出不適當的動作也常為意外發生原因，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。

在現有成人動作辨識的研究中，多會以骨架偵測開始著手。本文使用 OpenPose [25] 及 MediaPipe Pose [26] 等演算法進行嬰兒骨架之偵測，發現嬰兒需在特定情境下才能有較佳的偵測效果，結果如圖 3.6及圖 3.7，可看到嬰兒平躺之偵測效果尚可，但趴躺等其他姿勢之效果則不如預期。又因本研究目標為能從非限定視角辨識嬰兒動作，然而嬰兒骨架圖在俯

視角與平視角中多有相似之處（如圖 3.8），而無法達到良好的應用結果，因此，最終本文並不使用骨架圖進行姿勢辨識。

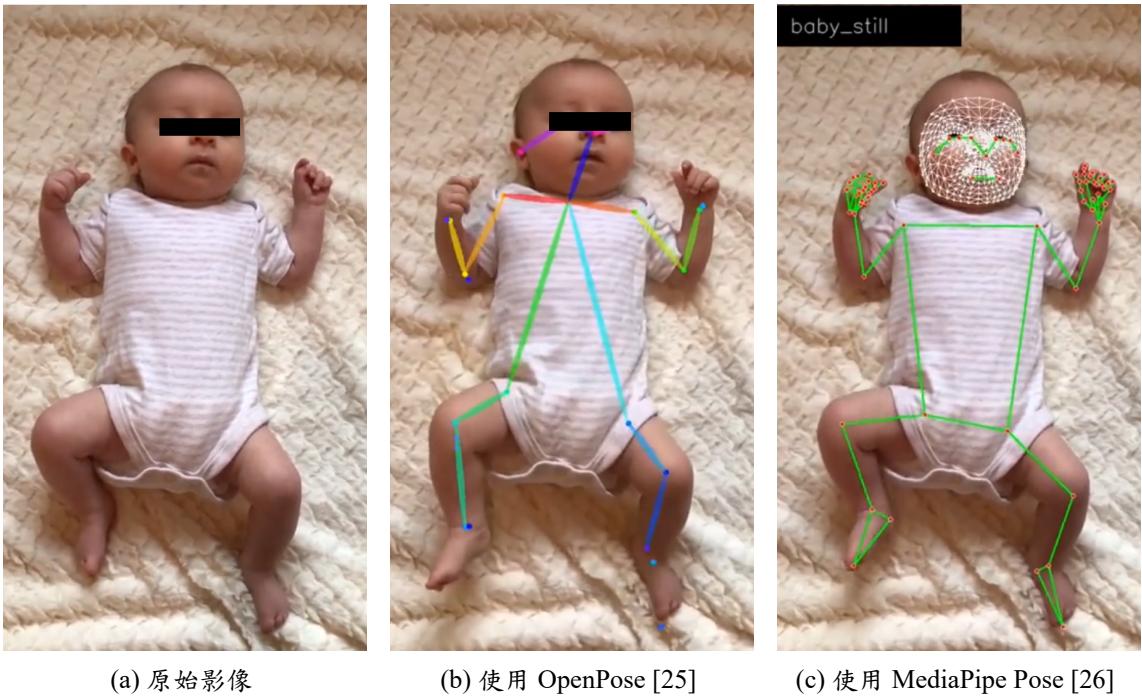


圖 3.6: 嬰兒平躺之骨架偵測結果



圖 3.7: 嬰兒趴躺之骨架偵測結果

而改為基於深度學習技術進行嬰兒動作辨識，使用自行收集之嬰兒影像資料集，訓練可辨識四種嬰兒基礎姿勢之模型。而資料集之詳細內容將於3.3.1節進行介紹。

本部分流程如圖 3.9：讀取嬰兒影像後，使用本文訓練的姿勢模型辨識嬰兒動作，並輸出判斷結果；若輸出為「正躺」或「坐姿」，此部分判斷嬰兒為安全狀態，而若結果為「趴躺」或「站立」，則分析嬰兒為警示狀態。

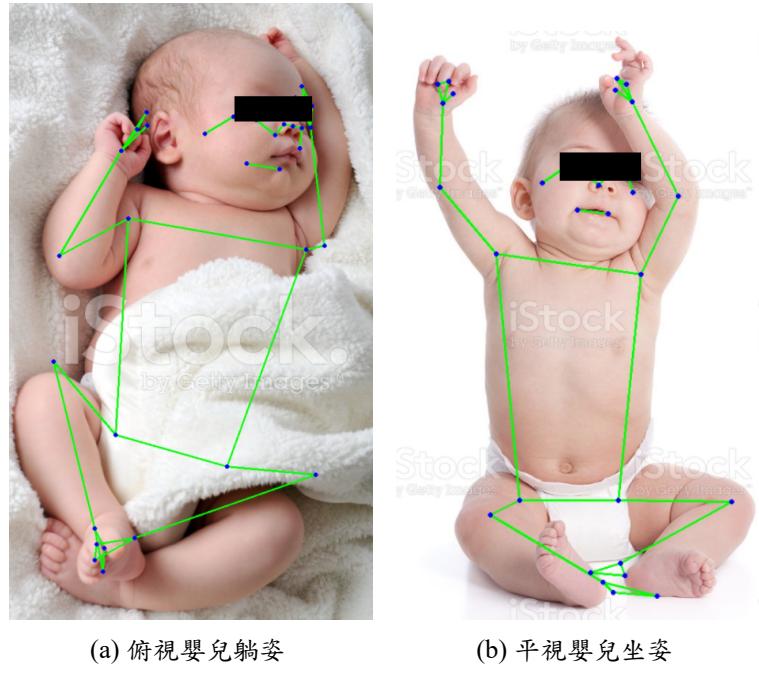


圖 3.8: 不同視角之嬰兒骨架偵測結果

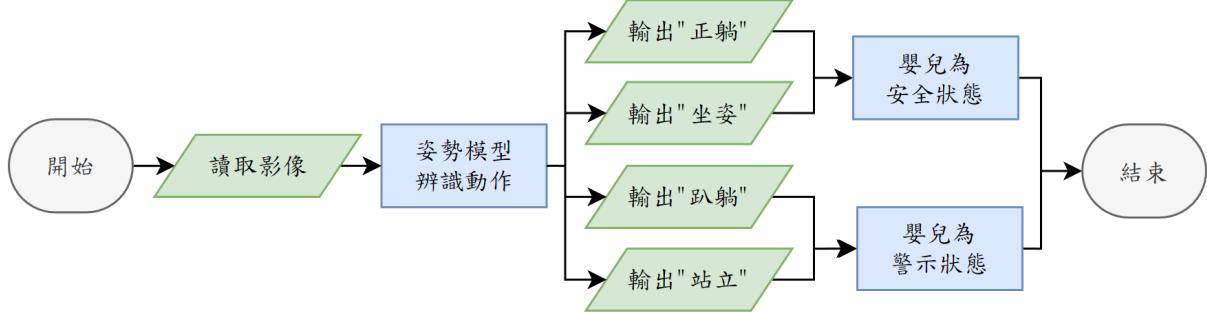


圖 3.9: 姿勢辨識流程圖

### 3.3.1 嬰兒姿勢資料集

由於目前未有公開之嬰兒資料集，故本論文使用之嬰兒影像，皆收集自網路上真實嬰兒的彩色照片或影片，再經前處理及分類標示而成。

起初，本研究將嬰兒姿勢分為五類：正躺、趴睡、爬行、坐姿及站立，而趴睡及爬行二類時常發生互相誤判，致使辨識錯誤率高。推測原因為此二類嬰兒皆呈現腹面朝下之姿，不同處在於四肢及軀體是否貼地，如圖 3.10 所示，但若接續細分姿勢，將使分類過細。

因此，最終本部分資料集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，各類別含 3774 張、3921 張、



圖 3.10: 嬰兒腹面朝下之姿

3900 張及 3821 張，總共 15416 張照片。且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，包含俯視及平視等。而對於此四類姿勢之詳細定義如下：

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 3.11a。
2. 趴躺：嬰兒腹部面朝下，包含趴睡及爬行等多動作，而頭部及四肢位置不限，如圖 3.11b。
3. 坐姿：嬰兒臀部貼於水平面，且背部未貼於同一平面，而頭部及四肢位置不限，如圖 3.11c。
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 3.11d。

並將完整資料集分成訓練、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。

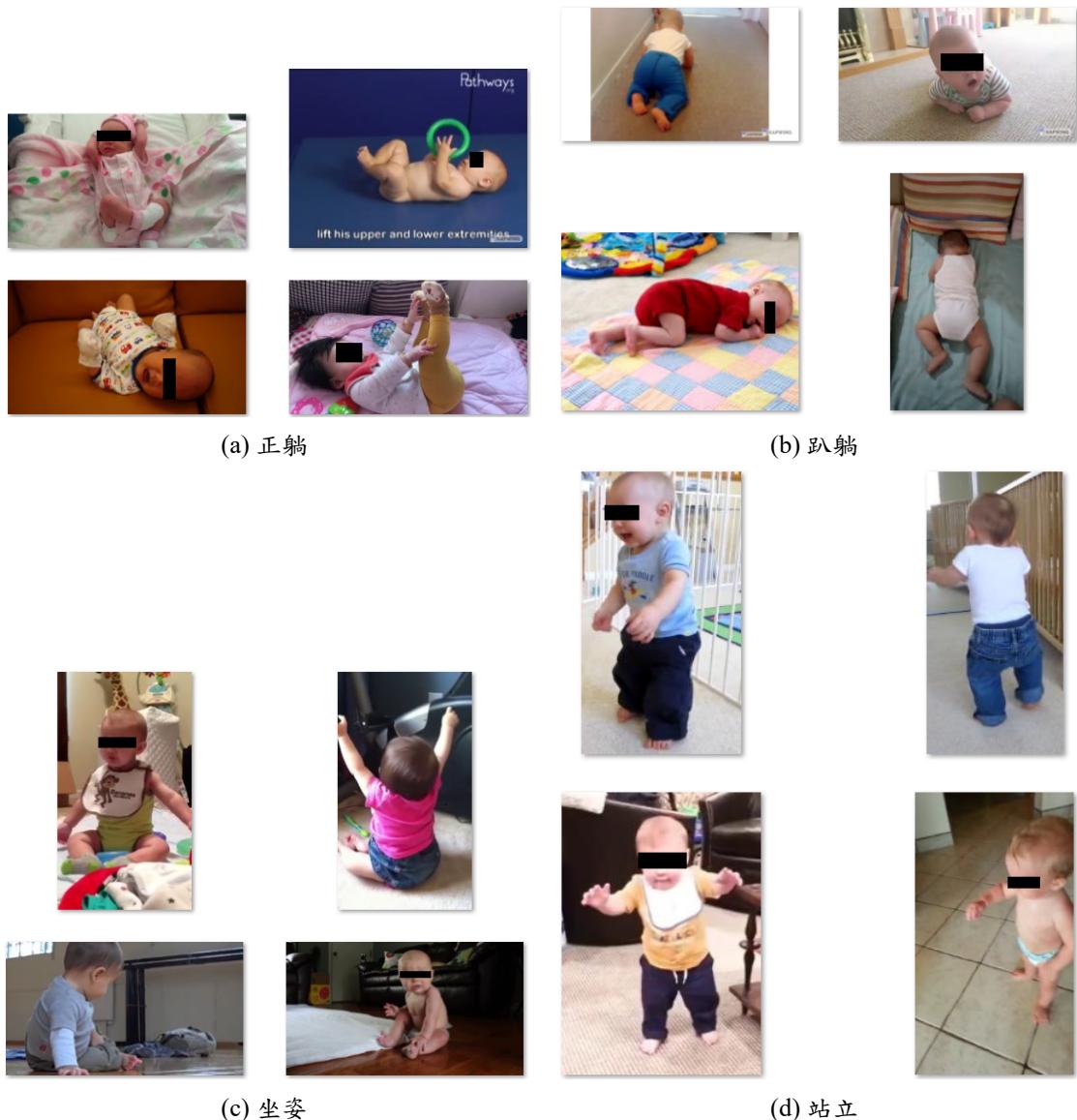


圖 3.11: 嬰兒姿勢資料集

### 3.3.2 模型訓練

本論文使用3.3.1節之嬰兒姿勢資料集，以 ResNet50 [16] 進行姿勢辨識模型之訓練，最終達成辨識四種嬰兒姿勢：正躺、趴躺、坐姿及站立。

## 3.4 危險情境判斷

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發生，並不須判斷一幀畫面為警 示狀態，就立即通知照護者。

因此，本系統使用一變數累積模型判斷嬰兒狀態為警示之幀數，當此變數超過設定閥值時，系統才會發出警示提醒照護者。本部分之流程圖，請見圖 3.12。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。

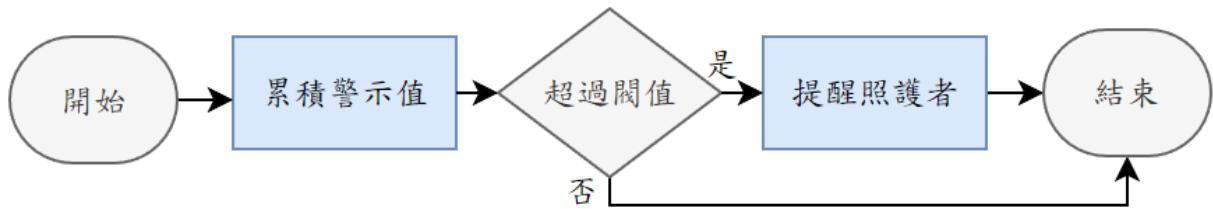


圖 3.12: 危險情境判斷流程圖

## 四、 實驗設計與結果

本章節中，根據第三章之研究方法以各項實驗驗證系統之設計，並介紹各實驗之目的與設計、評估方式以及結果與分析，透過以下五章子節進行說明：臉部偵測準確度實驗、臉部偵測執行時間實驗、臉部遮擋辨識實驗、姿勢辨識實驗及影片危險偵測實驗。

另外，本章各項實驗皆於相同硬體環境操作，詳細資訊如下：

- 作業系統：Windows 10
- CPU：Intel(R) Core(TM) i7-10700KF CPU @ 3.80GHz
- 記憶體：128GB
- GPU：NVIDIA GeForce GTX 1660

### 4.1 臉部偵測準確度實驗

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。

為了使本系統擁有較佳的臉部偵測準確性且兼具執行效能，本文透過4.1節及4.2節之實驗，分別進行臉部偵測演算法準確度與執行時間之比較，進而驗證以下設計：先使用 SSD 演算法偵測嬰兒臉部，此方法之召回率雖低，但其準確度很高，故能利用此算法之時間優勢；而若 SSD 演算法找不到嬰兒面部時，則接續使用 RetinaFace 演算法，利用其正確率及準確率皆高之優點進行嬰兒臉部偵測。

### 4.1.1 實驗目的與設計

本實驗為計算人臉偵測演算法之嬰兒面部擷取準確度，使用3.3.1節的嬰兒姿勢資料集共15416張影像，分析RetinaFace[20]、MTCNN[19]、SSD[18]及Haar cascade[27]等演算法之偵測結果。

### 4.1.2 實驗評估方式

透過分類標註四項演算法偵測嬰兒臉部之結果影像，計算出各演算法的accuracy、precision及recall。

### 4.1.3 實驗結果與分析

由表4.1、表4.2、表4.3及表4.4分別為RetinaFace、MTCNN、SSD及Haar cascade之詳細實驗結果。

表4.1: RetinaFace[20] 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	<b>12958</b>	<b>2458</b>

表4.2: MTCNN[19] 偵測嬰兒臉部結果

	True (預測有臉)	False (預測無臉)
True (實際有臉)	9361	3399
False (實際無臉)	517	2140
Total	<b>9877</b>	<b>5539</b>

而再經計算後，四項演算法之accuracy、precision及recall值如表4.5。

表 4.3: SSD [18] 偵測嬰兒臉部結果

	True (預測有臉)	False (預測無臉)
True (實際有臉)	4830	8141
False (實際無臉)	5	2440
Total	<b>4835</b>	<b>10581</b>

表 4.4: Haar cascade [27] 偵測嬰兒臉部結果

	True (預測有臉)	False (預測無臉)
True (實際有臉)	2882	9546
False (實際無臉)	725	2263
Total	<b>3607</b>	<b>11809</b>

故透過本實驗結果可得出選用 RetinaFace 演算法行嬰兒臉部偵測，可擁有較佳的偵測準確度。另外，值得注意的是，雖然 SSD 的 accuracy 及 recall 很低，但其 precision 達 99.90%，也就是說其將多數影像誤判為無臉，但判斷為有臉的結果幾乎正確。

因此，本研究接續進行4.2節的實驗，希望利用 SSD 這樣的特質。

表 4.5: 人臉偵測準確度結果

演算法	Accuracy	Precision	Recall
RetinaFace	99.71%	99.75%	99.91%
MTCNN	74.60%	94.78%	73.36%
SSD	47.16%	99.90%	37.24%
Haar cascade	33.37%	79.90%	23.19%

## 4.2 臉部偵測執行時間實驗

本研究進行嬰兒臉部偵測除了考量準確度外，亦希望提升整體系統之執行效率。

### 4.2.1 實驗目的與設計

本實驗為計算人臉偵測演算法之執行時間，使用3.3.1節的嬰兒姿勢資料集共 15416 張影像，分析 RetinaFace [20]、MTCNN [19]、SSD [18] 及 Haar cascade [27] 等演算法，其臉部偵測時間之結果；再以嬰兒影片進行實驗，比較本文所設計的兩步驟流程及僅使用 RetinaFace 演算法之偵測時間。

### 4.2.2 實驗評估方式

透過計算四項演算法偵測完整資料集所花費之時間，分析各演算法平均偵測一張影像之執行時間；並比較有無使用本文方法之臉部偵測時間。

### 4.2.3 實驗結果與分析

RetinaFace、MTCNN、SSD 及 Haar cascade 四項演算法之詳細實驗結果，請見表 4.6。

表 4.6: 人臉偵測執行時間結果

演算法	總花費時間 (15416 張影像)	每張影像平均時間
RetinaFace	5 小時 42 分 2.10 秒	1.33 秒
MTCNN	2 小時 8 分 22.05 秒	0.50 秒
SSD	9 分 17.26 秒	0.04 秒
Haar cascade	18 分 01.78 秒	0.07 秒

故透過本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 4.1 節實驗中，準確度最高的 RetinaFace 其平均偵測一張影像需 1.33 秒，為 SSD 的 33.25 倍。

而使用本文設計之兩步驟偵測流程及僅使用 RetinaFace 演算法之詳細實驗結果，請見表 4.7。

表 4.7: 使用本文方法之人臉偵測執行時間結果

方法	總花費時間 (1594 張影像)	每張影像平均時間
本文方法	17 分 51.21 秒	0.67 秒
僅使用 RetinaFace	32 分 2.69 秒	1.21 秒

因此，總結 4.1 節與 4.2 節之實驗結果，驗證本系統先使用 SSD 演算法偵測嬰兒臉部，未如期找到目標時，則改以 RetinaFace 演算法偵測，達成兼具準確性及執行效率之系統目標。

## 4.3 臉部遮擋辨識實驗

本研究中，利用深度學習技術辨識嬰兒臉部是否遭非奶嘴之異物遮蔽，進而判斷嬰兒是否處於危險情境中。

### 4.3.1 實驗目的與設計

本實驗為訓練針對嬰兒臉部遮擋辨識之模型，以 ResNet50 [16] 訓練 3.2.2 節的嬰兒臉部資料集，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、使用奶嘴及面部遭異物遮蔽之警狀狀態），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 optimizer 且將學習率設為 0.000001 以進行微調。

### 4.3.2 實驗結果分析

模型最終訓練準確率達 98.06%，而測試準確率達 99.43%，詳細訓練結果請見圖 4.1。

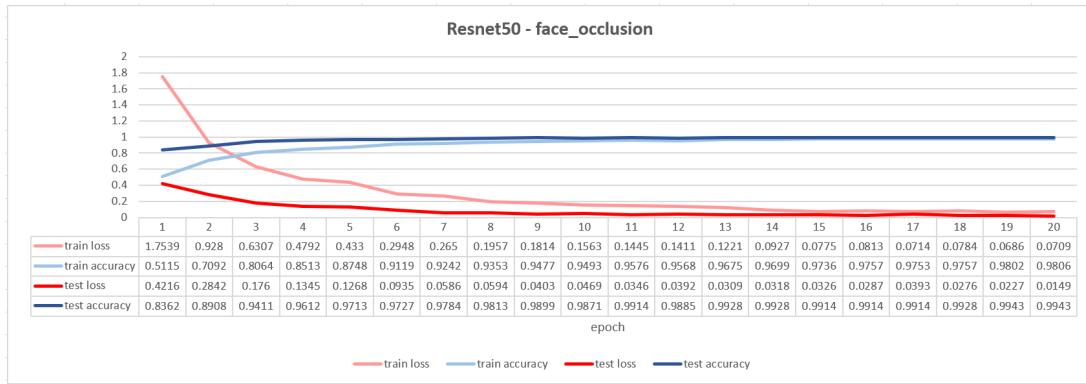


圖 4.1: 臘部辨識訓練及測試結果

我們使用 342 張之驗證集影像進行模型驗證，所有影像皆辨識正確。此模型之混淆矩陣如表 4.8，表中數字為各類影像之張數及比例，並於最右行展示各類別召回率。

表 4.8: 臥部遮擋辨識模型之混淆矩陣（單位：張 / 百分比）

		預測類別			召回率
		安全	奶嘴	警狀	
實際 類別	安全	120 (100.00)	0 (0.00)	0 (0.00)	100.00%
	奶嘴	0 (0.00)	115 (100.00)	0 (0.00)	100.00%
	警狀	0 (0.00)	0 (0.00)	107 (100.00)	100.00%

## 4.4 姿勢辨識實驗

本研究中，利用深度學習技術辨識嬰兒基礎姿勢，進而判斷嬰兒是否處於危險情境中。

### 4.4.1 實驗目的與設計

本實驗為訓練針對嬰兒姿勢辨識之模型，以 ResNet50 [16] 訓練3.3.1節的嬰兒姿勢資料集，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 optimizer 且將學習率設為 0.000001 以進行微調。

### 4.4.2 實驗結果分析

模型最終訓練準確率達 99.45%，而測試準確率達 99.71%，詳細訓練結果請見圖 4.2。



圖 4.2: 姿勢辨識訓練及測試結果

我們使用 744 張之驗證集進行模型驗證，包含了五張類別辨識錯誤的影像，其中三張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 4.3），而導致誤判。此模型之混淆矩陣如

表 4.9，表中數字為各類影像之張數及比例，並於最右行展示各類別召回率。



圖 4.3: 姿勢辨識錯誤之影像：坐姿誤判為趴躺

表 4.9: 姿勢辨識模型之混淆矩陣（單位：張 / 百分比）

		預測類別				召回率
		正躺	趴躺	坐姿	站立	
實際 類別	正躺	164 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)	100.00%
	趴躺	1 (0.52)	191 (99.48)	0 (0.00)	0 (0.00)	99.48%
	坐姿	0 (0.00)	3 (1.50)	196 (98.00)	1 (0.50)	98.00%
	站立	0 (0.00)	0 (0.00)	0 (0.00)	192 (100.00)	100.00%

## 4.5 影片危險偵測實驗

本研究基於嬰兒影像進行臉部遮擋及姿勢辨識，透過讀取嬰兒影片達成危險監測之目標。

### 4.5.1 實驗目的與設計

本實驗為驗證此系統能基於嬰兒影像進行危險監測，利用網路之真實嬰兒影片，包含不同之拍攝視角、嬰兒樣貌及狀態等，實驗臉部遮擋辨識模型與姿勢辨識模型之準確性。

## 4.5.2 實驗評估方式

透過輸出每幀影像之臉部遮擋及姿勢辨識結果，計算其 accuracy、precision 及 recall，以驗證此二模型得以應用在監測嬰兒危險情境。

## 4.5.3 實驗結果分析

本實驗使用之影片，包含嬰兒許多的不同情境，如：清醒與否、不同衣著、穿戴帽子與否及相異背景環境等，進行其臉部遮擋（無遮蔽、使用奶嘴及遭異物遮擋）與姿勢（正躺、趴躺、坐姿及站立）之危險辨識。

首先，姿勢辨識的部分，包含了多張誤判為趴躺姿勢的影像，推測原因為嬰兒身體遭棉被遮擋（如圖 4.4），而只拍攝到露出的嬰兒臉部，故造成姿勢辨識錯誤。



圖 4.4: 姿勢辨識錯誤之影像：正躺誤判為趴躺

其次，臉部遮擋辨識的部分，會先刪去嬰兒臉部未被偵測之影像（如圖 4.5），而後判斷有多張影像類別應為嬰兒正在使用奶嘴或安全狀態，但誤判為遭異物遮蔽之警示狀態，推測原因為影像中之奶嘴或嬰兒臉部遭手部等遮擋（如圖 4.6），而誤判類別。



圖 4.5: 未偵測嬰兒臉部之影像



圖 4.6: 臉部遮擋誤判之為警狀狀態

兩部分之混淆矩陣如表 4.10 及表 4.11，表中數字為各類影像之張數及比例，並於最右行展示各類別召回率。

表 4.10: 實驗影片姿勢辨識之混淆矩陣（單位：張 / 百分比）

		預測類別				召回率
		正躺	趴躺	坐姿	站立	
實際 類別	正躺	3223 (90.87)	324 (9.13)	0 (0.00)	0 (0.00)	90.87%
	趴躺	0 (0.00)	161 (100.00)	0 (0.00)	0 (0.00)	100.00%
	坐姿	0 (0.00)	0 (0.00)	214 (100.00)	0 (0.00)	100.00%
	站立	4 (1.87)	1 (0.47)	0 (0.00)	209 (97.66)	97.66%

表 4.11: 實驗影片臉部遮擋辨識之混淆矩陣（單位：張 / 百分比）

		預測類別			召回率
		安全	奶嘴	警示	
實際 類別	安全	1429 (85.83)	52 (3.12)	184 (11.05)	85.83%
	奶嘴	17 (0.84)	1615 (80.23)	381 (18.93)	80.23%
	警示	40 (11.90)	0 (0.00)	296 (88.10)	88.10%

# 五、結論與未來展望

本章節中，對本論文進行總結，並闡述本研究未來可發展與應用之處，將透過以下兩章子節進行分述：結論、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，而提供了關注於嬰兒臉部及動作之危險監測系統，將有助於協助照護者，並降低嬰兒猝死症發生風險。

## 5.2 未來展望

本研究中，目前僅針對單一嬰兒之情境進行辨識，未來可提供多嬰兒情境之危險偵測，則使用場景將可更廣泛。

本文中之姿勢辨識部分，目前僅辨識四種姿勢，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；而在臉部遮擋辨識部分，若加入偵測面部表情等其他資訊，亦可更詳盡的監測嬰兒狀態。

另外，系統方面，未來可提供設定觀測之年齡區間，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。

# 參考文獻

- [1] 統計處. “歷年死因統計.” 中文. (Mar. 2021), [Online]. Available: <https://dep.mohw.gov.tw/DOS/lp-5069-113.html>.
- [2] H. C. Kinney and B. T. Thach, “The sudden infant death syndrome,” *New England Journal of Medicine*, vol. 361, no. 8, pp. 795–805, 2009.
- [3] “What causes SIDS?” en, [Online]. Available: <https://www.nichd.nih.gov/health/topics/sids/conditioninfo/causes>.
- [4] 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)*, IEEE, 2006, 3–pp.
- [5] 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)*, IEEE, 2016, pp. 1920–1925.
- [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)*, IEEE, 2014, pp. 1–4.
- [7] E. Ziganshin, M. Numerov, and S. Vygolov, “Uwb baby monitor,” in *2010 5th International Conference on Ultrawideband and Ultrashort Impulse Signals*, IEEE, 2010, pp. 159–161.
- [8] 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)*, IEEE, 2015, pp. 1–8.
- [9] 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)*, IEEE, 2017, pp. 1–7.
- [10] 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)*, IEEE, 2019, pp. 1–4.

- [11] 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)*, IEEE, 2019, pp. 3006–3010.
- [12] 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)*, IEEE, 2021, pp. 286–291.
- [13] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [14] C. Szegedy, W. Liu, Y. Jia, *et al.*, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [15] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International conference on machine learning*, PMLR, 2015, pp. 448–456.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [17] L. Cuimei, Q. Zhiliang, J. Nan, and W. Jianhua, “Human face detection algorithm via haar cascade classifier combined with three additional classifiers,” in *2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, IEEE, 2017, pp. 483–487.
- [18] 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)*, IEEE, 2021, pp. 8445–8450.
- [19] 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, 2016.
- [20] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, “Retinaface: Single-shot multi-level face localisation in the wild,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5203–5212.
- [21] 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*, IEEE, vol. 2, 2008, pp. 758–761.
- [22] 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*, IEEE, 2011, pp. 335–339.
- [23] “Python opencv 膚色檢測的實現示例 \_ 程式設計 \_ 程式人生,” [Online]. Available: <https://www.796t.com/article.php?id=196625>.

- [24] WalkonNet. “Python opencv 膚色檢測的實現示例 - WalkonNet.” zh-TW, [Online]. Available: <https://walkonnet.com/archives/7903>.
- [25] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7291–7299.
- [26] “Pose.” en-US, [Online]. Available: <https://google.github.io/mediapipe/solutions/pose.html>.
- [27] K. Goyal, K. Agarwal, and R. Kumar, “Face detection and tracking: Using opencv,” in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, vol. 1, 2017, pp. 474–478.