

國立臺灣師範大學機電工程學系

碩士論文

指導教授：陳美勇博士

以類神經網路實現臉部影像疼痛水準即時估測

Implemented Rapid Pain Intensity Estimation from Facial Image  
using Artificial Neural Network



研究生：林高遠 撰

中 華 民 國 1 0 5 年 8 月

## 摘要

本論文提出一種以人類臉部影像為輸入資料，用類神經網路即時測得受測者疼痛水準的方法。情感運算在最近幾年來已經逐漸受到重視，而臉部影像疼痛水準自動估測，有助於醫療照顧、健康促進等領域自動化，有效減少第一線照顧者的負擔。但相關研究的數量與關注程度遠落後於其他表情辨識技術，使得相關應用發展受限。

本論文嘗試兩種輸入資料：一種是屬於低階外觀特徵的人類臉部眼、嘴區域 Uniform LBP 直方圖，取得 118-D 向量；另一種屬於使類神經網路自動尋找的高階抽象特徵，將臉部影像做最大池化(Max Pooling)處理後，從 32x32 灰階影像取得 1024-D 向量。將兩者正規化，再輸入類神經網路做迴歸訓練與測試。

實驗結果方面，將 The UNBC-McMaster Shoulder Pain Expression Archive Database 隨機分割為兩份，分別做為訓練與測試樣本。將本論文提出的方法與 Sebastian Kaltwang 等人與 Xiaopeng Hong 等人的研究比較，可以達到較小的均方誤差( $MSE=0.17$ )與較接近 1 的皮爾森相關係數( $r=0.94$ )。速率表現方面，本論文以 C#實作出的程式在 i5 雙核心的電腦上平均可以達 24FPS。

**關鍵詞：**監督式機器學習、特徵學習、痛苦偵測、電腦視覺

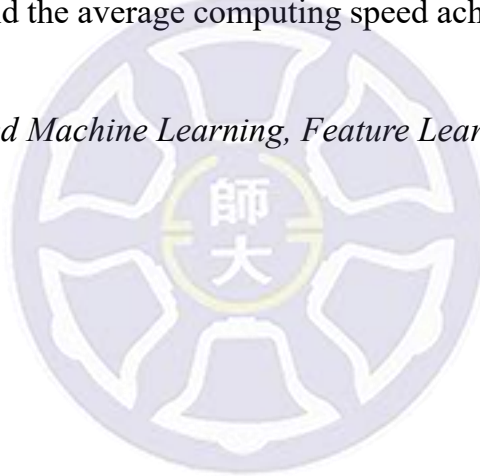
## Abstract

This thesis presents a method to estimate pain intensity which is revealed on human face image rapidly.

Two types of data are extracted from the human face image: one of which is the Uniform LBP, the belong low-level appearance features, which is extracted from the eyes and mouth area; the other is the 32x32 face image data which is extracted using Max Pooling. Both will be computed by the regression neural network, and the neural network is trained and the training result will be verified.

The data from the UNBC-McMaster Shoulder Pain Expression Archive Database is randomly assigned into two groups—one for training, another for testing. The result of this study achieves a MSE of 0.17 and a Pearson's correlation coefficient close to 1 ( $r = 0.94$ ), and the average computing speed achieves 24FPS on i5 dual-core computer.

**Keywords :** *Supervised Machine Learning, Feature Learning, Pain detection, Computer Vision.*



## 誌謝

首先要感謝你，就是正在看這篇論文的你，因為太多人認為碩士論文形同垃圾。為了感謝你，本論文準備了幾份特點如下：

- 不會只有自己研究室才看得懂的誌謝
- 因為參數詳細，所以容易重現的實驗過程
- 研究成果實用，且冷門到目前尚無公開 API 可達成相同功能
- 更加生動的口試投影片可以在我的個人網站找到

另外，為了讀者良好的使用者經驗(User Experience)，本論文在 Word 與 PDF 版皆可以使用目錄功能直接跳轉。

感謝指導教授陳美勇老師在我大學時的「機械視覺」課程引領我進入影像處理的領域、並且從大學專題到研究所都容忍我不斷犯錯，在我不圓滑的待人處事上給予很多建議與幫助。此外，在研究所階段老師總是鼓勵同學盡量到外系或外校修課、參與競賽等，讓我的研究所生涯總是能走出舒適圈並開拓眼界。

感謝口試委員練光祐老師、郭景明老師、黃正民老師的熱心參與，讓我在僅僅 30 分鐘的報告後，歷經一場長達 70 多分鐘的密集攻防，尤其郭老師與黃老師的意見可說是拳拳到肉。這樣的挑剔，對我來說是一種提攜後進的認同；即使看見的論文缺點變多了，但經過洗禮後卻覺得更有價值。

感謝吳順德老師在我大學時的「程式設計」和研究所的「物件導向程式設計」課程幫助我打下深厚的基礎，使本論文幾乎不曾因為 coding 問題停滯，讓我可以好好聚焦在各種演算法上；感謝資工系的葉梅珍老師在「多媒體系統設計」課程引領我認識情感運算，並在本論文最後階段給予資料擴增的建議，為本論文研究成果的實用性推了一把。

感謝本研究室博班的鍾秉剛同學，在口試預報給的諸多意見，讓我對口試準備更充分；感謝本研究室的校友楊智翔同學，即使畢業後仍利用下班時間多次幫我找出本論文中需要修改之處；感謝 MSDN 上的各方高手經常幫忙回答疑問，讓本論文實作上順利許多，使我能以較好的架構與重用性(Reusability)完成本論文所需的程式碼；感謝瑩瑩在論文定稿前與我一起重寫英文摘要，一邊改一邊告訴我更好的寫法，讓我知道自己的英文水準還遠遠不足。



# 目錄

摘要 .....	ii
Abstract.....	iii
誌謝 .....	iv
目錄 .....	vi
表目錄 .....	ix
圖目錄 .....	x
第一章 緒論 .....	1
1.1 研究動機.....	1
1.2 論文架構.....	2
1.3 論文貢獻.....	3
第二章 表情疼痛估測相關研究回顧與探討 .....	5
2.1 FACS 與 PSPI 兩大人臉疼痛水準量化系統 .....	5
2.2.1 Facial Action Coding System(FACS) : .....	5
2.2.2 Prkachin and Solomon pain intensity(PSPI) : .....	5
2.2 現有臉部表情資料庫.....	7
2.2.1 Japanese Female Facial Expressions (JAFPE) .....	7
2.2.2 Taiwanese Facial Expression Image Database .....	7
2.2.3 Infant COPE database .....	8
2.2.4 The UNBC-McMaster Shoulder Pain Expression Archive Database.....	8
2.3 疼痛表情辨識相關研究.....	9
2.3.1 疼痛表情特徵研究回顧.....	10
2.3.2 The UNBC-McMaster Shoulder Pain Expression Archive	

Database 的 Frame level Ground truth 品質探討.....	11
2.3.3 痛苦表情辨識速度研究回顧 .....	15
第三章 類神經網路介紹 .....	16
3.1 單層感知機與倒傳遞.....	16
3.2 多層感知機.....	18
3.3 反向傳播演算法.....	20
3.4 迴歸類神經網路.....	21
3.4.1 迴歸簡介.....	21
3.4.2 以類神經網路迴歸 $\sin(x)$ 範例 .....	22
3.5 學習模式.....	24
3.5.1 線上學習.....	24
3.5.2 批量學習.....	24
3.5.3 隨機學習.....	24
3.6 訓練數據調教.....	25
3.6.1 資料標準化.....	25
第四章 影像前處理 .....	27
4.1 眼嘴區域 Local Binary Pattern .....	27
4.2 池化運算.....	30
4.3 資料擴增.....	32
第五章 實驗結果與討論 .....	34
5.1 實驗設備.....	34
5.1.1 EmguCV 3.1.....	34
5.1.2 Microsoft Visual Studio Community 2015 .....	35
5.2 均方誤差.....	36
5.3 皮爾森相關係數.....	36

5.4	混淆矩陣.....	38
5.5	實驗方法.....	39
5.6	實驗結果.....	42
第六章	結論與未來展望 .....	48
6.1	結論.....	48
6.2	未來展望.....	49
參考文獻	.....	50





## 表目錄

表 2-1 各種情感與 AU 的對應關係.....	5
表 2-2 The UNBC-McMaster Shoulder Pain Expression Archive Database 的 ground truth 種類.....	9
表 2-3 The UNBC-McMaster Shoulder Pain Expression Archive Database 資 料缺陷分析.....	12
表 2-5 多篇疼痛表情辨識研究的硬體平台與執行速度.....	15
表 3-1 圖 3-1 的說明.....	17
表 3-2 類神經網路規格與訓練參數.....	22
表 3-3 資料標準化前後對照範例.....	26
表 3-4 建議的類神經網路資料標準化方式.....	26
表 4-1 推定的雙眼與嘴巴相對區域.....	27
表 4-2 以 TFEID 的 happiness 再擴增的實驗結果.....	33
表 5-1 本論文與 Sebastian Kaltwang 等人的研究結果比較.....	42
表 5-2 使用最大池化原始影像以不同參數練出的相關係數.....	43
表 5-3 使用最大池化原始影像以不同參數練出的均方誤差.....	43
表 5-4 實驗結果所用類神經網路的其他規格 .....	43

## 圖目錄

圖 1-1 微軟 Emotion API Demo 結果，右邊為輸出的 JSON 格式資料.....	1
圖 2-1 Facial Action Coding System 的部分編碼與對應臉部動作.....	6
圖 2-2 以 Action Unit 得到 PSPI 的影像範例.....	6
圖 2-3 JAFFE 部份樣本.....	7
圖 2-5 TFEID 部份樣本.....	8
圖 2-6 來自 UNBC-McMaster Shoulder Pain Expression Archive Database 的 部份樣本.....	9
圖 2-7 UNBC Pain Database 所有樣本的 PSPI 直方圖.....	12
圖 2-8 UNBC Pain Database 從原始影像計算出 PSPI 的流程圖.....	13
圖 2-9 AAM landmarks 與原始影像對照，檔案名稱為 hs107t2aaaff352。此樣 本 PSPI 為 15 分.....	13
圖 2-10 AAM landmarks 與原始影像對照，檔案名稱為 ll042t1aaaff001。此 樣本無法解析出 FACS，PSPI 被標示為 0 分。.....	14
圖 2-11 AAM landmarks 與原始影像對照，檔案名稱為 jk103t2aeaff232。此 樣本被解析出的 Action Unit 為 AU25 與 AU26，PSPI 被標示為 0 分。.....	14
圖 3-1 單層感知機.....	16
圖 3-2 口袋式單層感知機學習人體運動的分類結果.....	17
圖 3-2 一個二維 XOR 問題的範例.....	18
圖 3-3 不同層數感知機的活化函數，左邊屬於單層，右邊屬於多層.....	18
圖 3-4 具有三大層架構的多層感知機.....	20
圖 3-5 學習率、損失函數與迭代次數三者之間的關係示意圖.....	21
圖 3-6 線性迴歸、多項式迴歸.....	21
圖 3-7(a)(b) $\sin(x)$ 迴歸 a.訓練用資料 b.訓練用迴歸類神經網路.....	22

圖 3-8 訓練途中的回想結果.....	23
圖 4-1 以推定區域對樣本框選的範例.....	27
圖 4-2 LBP 的計算範例.....	28
圖 4-3 LBP 影像與原始影像對照.....	28
圖 4-4 LBP to Uniform LBP Lookup Table.....	29
圖 4-5 (a).雙眼區域高斯模糊影像 (b).對(a)取 Uniform LBP 影像 (c).高斯模糊前 Uniform LBP 直方圖 (d).高斯模糊後 Uniform LBP 直方圖.....	29
圖 4-6 最大池化運算範例與平移前後對照.....	31
圖 4-7 人臉影像最大池化前後對照.....	31
圖 4-8 UNBC Pain Database 所有人臉影像最大池化的平均影像.....	31
圖 4-9(a)(b)(c)(d) 資料擴增示意圖.....	33
圖 5-1 EmguCV 與 OpenCV 架構關聯示意圖.....	34
圖 5-2 本論文使用 Visual Studio Community 2015 的開發畫面.....	35
圖 5-3 迴歸誤差示意圖.....	37
圖 5-4 相關係數與散布圖範例.....	37
圖 5-5(a)(b) 混淆矩陣範例.....	38
圖 5-6 預處理資料.....	39
圖 5-7 無法擷取出人臉的部分樣本.....	39
圖 5-8 實驗流程圖.....	40
圖 5-9 輸入 LBP 直方圖的迴歸類神經網路.....	41
圖 5-10 輸入最大池化影像的迴歸類神經網路.....	41
圖 5-11 實驗結果混淆矩陣.....	44
圖 5-12 The UNBC-McMaster Shoulder Pain Expression Archive Database 中數個影片的連續估測與 Ground Truth 比較.....	45
圖 5-13 舉重與演員孕婦生產影片的估測過程截圖.....	47

# 第一章 緒論

## 1.1 研究動機

台灣已逐漸邁入高齡化社會，除了伴隨而來的各行各業人力短缺，對於醫療照顧的人力需求更是隨人口老化一同增加；在眾多產業力求自動化轉型的時代下，醫療照護仍然是自動化程度相對低落的勞力密集產業。不論自動化程度再高，若是不能讓機器感知人的情緒，機器將難以取代第一線的醫療產業勞工。

在情感運算(Affective Computing)的技術中，已經可以透過人臉影像辨識出多種情緒，諸如憤怒、輕蔑、厭惡、恐懼、快樂、中性、悲傷和驚奇等，且這些情緒的辨識技術已經成熟並商用化(如圖 1-1) [1]。

對於醫療照護產業而言，疼痛表情最具有檢出價值，但相關研究卻還在起步階段，不論數量與關注程度皆遠落後於其他表情辨識技術，使得相關應用發展受限。

比起應用在醫療上更加積極的方向，其中之一就是應用在運動產業上以促進國民健康，實踐”預防勝於治療”。當同一位運動器材使用者，長期保持相同的運動學表現時，可能由於體能水準進步，使運動表情日漸由痛苦轉為自然。若智慧運動器材藉由痛苦感知來評價使用者的主觀努力水準，可更精準了解使用者生理上的體適能變化。



圖 1-1 微軟 Emotion API Demo 結果，右邊為輸出的 JSON 格式資料

在這個行動裝置大行其道的時代，人機介面(Human Machine Interface)的重要性快速提升。繼攝影鏡頭後，觸控裝置與各種物理量的感測器在短短幾年之內，躍升為手機與筆記型電腦的主流周邊設備之一，無不就是為了讓人與機器的連結更加緊密。

大多數的人機互動演算法，是為了取得人類可自己查覺的行為訊息，例如：觸控板的背景電容估測演算法、以三軸加速規區分日常生活動作、睡眠狀態的辨識演算法等。但人類經常透過臉部表情不自覺的流露出各種情感，因此相較於其它的人機互動方法，情感運算可以提供更為細膩的人類活動資訊，使機器的服務決策更精準。

然而，相較於各種物理量的感測器，攝影鏡頭能帶來的資訊更加豐富，故有「眼睛是靈魂之窗」的諺語。然而，因為影像資訊太過豐富，需要耗費大量運算資源過濾才能截取出有用資訊，因此各種電腦視覺的研究蓬勃發展，近幾年甚至已經將基於電腦視覺的轎車自動道路駕駛實用化。在可預見的將來，會有更多人類勞動工作因電腦視覺的進步而被機器取代。

在心理學分類中「痛苦」是一種典型的情緒，但是在情感運算中卻較少被討論到。基於先前敘述的應用潛力，本論文認為以人臉影像估測「痛苦」水準有相當高的研究價值。

## 1.2 論文架構

一個典型的辨識系統 [2]可以分為「特徵擷取」與「辨識模型」兩大部分，擷取何種特徵往往由人工定義，只有辨識模型的訓練上較為自動化。但近年興起的「深度學習(Deep Learning)」已經不再強調由人工高度干涉的特徵選擇，而是對資料做適度的強化與降低維度後，交由機器自動學習特徵，讓電腦在人類所不能及的大量與快速運算中，發現資料裡難以察覺得抽象關聯。

相較特徵學習(Feature Learning)，由人工定義的特徵雖然被侷限在較具



體的關聯上，高度依賴對辨識目標的相關知識，但仍有容易理解、容易重現、資料維度低等優點。因此本研究以迴歸類神經網路(Neural Network Regression)為辨識模型，同時嘗試上述兩種特徵擷取法，比較不同方法對疼痛影像資料庫的逼近與泛化能力。

在人工定義特徵中，本論文首先介紹幾個知名的表情資料庫，針對有提供疼痛表情的資料庫詳細討論，再由疼痛表情估測的相關研究中，歸納何種外觀特徵與疼痛水準較具關連？

在特徵學習中，本論文會先介紹資料篩選與降低維度的方式。此處雖然減少人工對特徵選擇的干涉，但由於資料維度大幅增長，如何減少模型訓練時間？參數如何調教？將比過去更加重要，因此本論文也一併探討。

在類神經網路介紹中，本論文先從「單層感知機(Single-layer Perceptron)」帶讀者了解類神經網路的最小單元並介紹其應用實例，然後再進一步推廣到架構更複雜多變的「多層感知機(Multi-layer Perceptron, MLP)」與反向傳播機制，最後再用視覺化實例，生動呈現以類神經網路對  $\sin(x)$  進行回歸預測的過程，以由淺入深的過程帶領讀者認識類神經網路。

在實驗結果討論，本論文將用同一套疼痛影像資料庫與高引用數的國際期刊論文比較，證明本論文訓練出的類神經網路，對基於人臉影像的疼痛估測有較佳表現。對於光線干擾與資料庫變異性不足等問題，本研究也嘗試多種方法加以改善。最後在結論與未來展望的部分，將談到目前尚無法克服的客觀研究限制。

### 1.3 論文貢獻

本論文主要貢獻是證實特徵學習在臉部影像疼痛估測上的可行性；另一貢獻是實作出可使用的軟體，在 i5 雙核心的筆記型電腦上從 webcam 即時取像偵測可達 24fps 以上，若對影像資料庫在背景批次處理可達 60fps。

特徵學習過去已經應用在影像分類上 [3]，但多為各類間差異明顯的狀

況。本論文使用的疼痛患者影像資料庫為 The UNBC-McMaster Shoulder Pain Expression Archive Database [4]，該資料庫全部都是在同一種光線、背景、角度下所拍攝，差異僅有不同疼痛水準造成的細微表情變化，故類間差異極小，在此嚴苛條件下仍可給出誤差較低的估測。

為了讓研究成果為產業界所用，本論文的實驗幾乎都以免費工具軟體、Open Source 函式庫、或未經封裝的 C#程式碼完成，不須依賴龐大且昂貴的套裝軟體，就可重現本論文研究成果。



## 第二章 表情疼痛估測相關研究回顧與探討

### 2.1 FACS 與 PSPI 兩大人臉疼痛水準量化系統

#### 2.2.1 Facial Action Coding System(FACS)：

Facial Action Coding System [5]最早是由解剖學家提出的，以臉部肌肉動作與否作為臉部動作分類依據，臉部上的每條肌肉動作被簡化成只有”動”與”不動”，如同數位系統中的 0 或 1，每一個位元只存在其中一種狀態。由於這套編碼系統幾乎窮舉所有人類可能出現的臉部動作，可以用來當作組成人類表情的基本元素，因此把每一種臉部基本動作成分稱作 AU(Action Unit)(如圖 2-1)。這套系統已經被心理學實驗與動畫產業廣泛使用，對應到臉部表情的方式是透過組合，如表 2-1。

#### 2.2.2 Prkachin and Solomon pain intensity(PSPI)：

Prkachin and Solomon [6]在 2008 年的一篇醫學研究中，針對疼痛表情的解剖學結構進行分析，並利用 FACS 定義出疼痛水準的公式，稱為「Prkachin and Solomon pain intensity」，該公式已經被疼痛表情相關研究廣泛使用，計算出來的結果就是 PSPI，公式如(2-1)：

$$PSPI = AU4 + \max(AU6, AU7) + \max(AU9, AU10) + AU43 \quad (2-1)$$

PSPI 的範圍為 0~15 分，總共 16 個整數級別，除 AU43 為 0~1 分外，其餘三項都是 0~5 分。以圖 2-3 為例，PSPI 的算法如式(2-2)

$$PSPI = 4 + \max(3, 4) + \max(2, 3) + 1 = 12 \quad (2-2)$$

表 2-1 各種情感與 AU 的對應關係

情感	AU 組合	情感	AU 組合
幸福	6 + 12	恐懼	1 + 2 + 4 + 5 + 7 + 20 + 26
悲	1 + 4 + 15	憤怒	4 + 5 + 7 + 23
驚	1 + 2 + 5B + 26	厭惡	9 + 15 + 16



Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
					
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink
Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
					
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
					
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
					
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

圖 2-1 Facial Action Coding System 的部分編碼與對應臉部動作

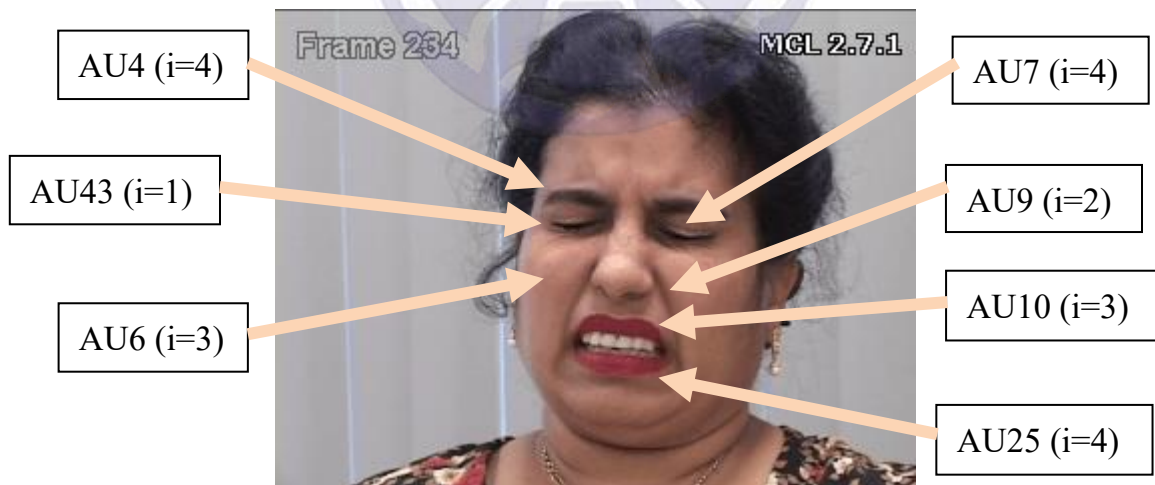


圖 2-2 以 Action Unit 得到 PSPI 的影像範例

## 2.2 現有臉部表情資料庫

為了更瞭解疼痛與自然的人臉有何不同？所以需要先取得大量的疼痛表情，從中找出疼痛表情特徵。以下將介紹較為知名的幾個表情資料庫。

### 2.2.1 Japanese Female Facial Expressions (JAFFE)

JAFFE [7]是日本九州大學心理學系(Psychology Department in Kyushu University)在 1997 年製作。由 10 位日本女性所拍攝的表情資料庫，每一位女性都會有 1+6 張不同表情的照片，分別是作為對照的自然表情，以及生氣、噁心、恐懼、快樂、難過、驚喜六種情緒。由於不包含痛苦類別，所以不適合本論文使用。

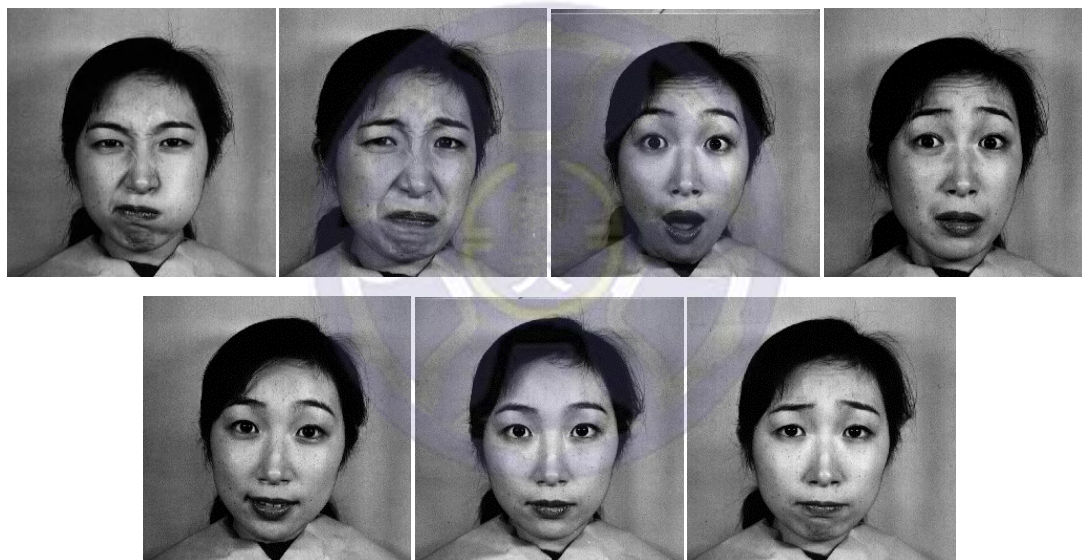


圖 2-3 JAFFE 部份樣本

### 2.2.2 Taiwanese Facial Expression Image Database

該資料庫 [8]是台灣陽明大學腦造影實驗室(the Brain Mapping Laboratory in National Yang-Ming University)在 2007 年製作。由 40 位台灣人所拍攝的表情資料庫，每一張照片大小都是 480\*600，每一位都會有 1+6 張不同表情的照片，分別是作為對照的 neutral(自然表情), anger(生氣), contempt(鄙視), disgust(噁心), fear(恐懼), happiness(快樂), sadness(難過) and

surprise(驚喜)，共七種表情。不包含疼痛類別，但本論文使用了 TFEID 的部分樣本作為資料擴增用。



圖 2-4 TFEID 部份樣本

### 2.2.3 Infant COPE database

該資料庫含有 206 張 24 位新生兒的照片，蒐集在他們出生後 18-36 小時內，於腳跟採血的疼痛表情照片。雖然這個資料庫在 Nanni 等人 [9]與 Gholami 等人 [10]的研究中被使用，但是並未發現任何可供下載或申請管道，可能不是供公開使用的資料庫。

### 2.2.4 The UNBC-McMaster Shoulder Pain Expression Archive Database

該資料庫 [4]是目前被使用最廣泛的疼痛表情的資料庫，由匹茲堡大學心理學系(Department of Psychology at the University of Pittsburgh)在 2011 年製作，每張影像都是來自肩痛患者在實施肩痛測試的臉部影片。隨著影格(frame)前進，約略可以看到疼痛程度逐漸起伏變化，絕大部分影格並不容易區分是疼痛或自然，這部分在稍後的章節有量化數據說明。

表 2-2 The UNBC-McMaster Shoulder Pain Expression Archive Database 的 ground truth 種類

來自 25 位受試者的 200 個影片，共 48398 張影格		
Frame_Labels	Sequence_Labels	
自動化計算	受試者自評	旁觀者評價
PSPI (0-15)	visual analogue scale (VAS) self-report code (0-10)	observer
	sensory (SEN) self-report code (0-10)	level code
	AFFECTIVE (AFF) self-report code (0-10)	OPR (0-5)

表 2-2 整理了該資料庫提供的各種 ground truth，其中 frame label 比 sequence label 更能提供細膩且富含變化的資料，更適合本論文動態即時辨識的需要，因此後面的章節將對 frame label 的 ground truth 分布做更詳細的探討。



圖 2-5 來自 UNBC-McMaster Shoulder Pain Expression Archive Database 的部份樣本

### 2.3 疼痛表情辨識相關研究

臉部情緒辨識已經相當發達，微軟也已經釋出付費使用的臉部情緒辨識 API(Application Interface)供開發者使用(如圖 1-1)，但目前已經商用化的研究成果僅限於憤怒、輕蔑、厭惡、恐懼、快樂、中性、悲傷和驚奇等八種情緒



[1]，其中並未包含疼痛類別。

截至目前為止，在 Google 學術上搜尋"face emotion recognition"，至少有 626 筆結果，但找遍國內外與疼痛表情辨識相關的研究，僅有 30 餘篇。

### 2.3.1 疼痛表情特徵研究回顧

在 Sebastian Kaltwang 等人 [11]的研究結論中提到：

*「Further-more, the current approach is static, and it cannot distinguish between eye blinks and eye closures, which are important cues for pain intensity estimation. These limitations of the proposed approach are the focus of our future research.」*

可見 Sebastian Kaltwang 等人的研究認為，「閉眼動作」與「眼睛周圍皮膚皺褶」和疼痛表情有高度相關。

而王佳琪 [12]透過臉部紋理分析提出更詳細的看法，認為疼痛患者會緊閉眼睛直到周圍皮膚出現皺褶，嘴巴則會緊閉或張開，並且只擷取眼睛和嘴巴的紋理特徵，以 Support Vector Machine 作為訓練與辨識用資料。

在 Hong Fu [13]等人的研究中提出以動態紋理(dynamic textures)擷取人類臉部的疼痛事件特徵，且認為沒有必要從整個臉部影像提取紋理，只需要眼、鼻與嘴周圍的紋理資訊。

Corneliu Florea [14]等學者在 2015 年的研究中，以地形特徵直方圖(Histograms of Topographical features)作為特徵空間進行疼痛表情辨識，並且對 PSPI 的定義提出兩點有待改進的觀察：

- AU9, AU10 這一組屬於疼痛表情的成分在 PSPI 中占了 0~5 分，但是在 disgust(噁心)表情中也可以發現 AU9 或 AU10。
- 不是所有疼痛患者都出現 AU43(closing eyes)，有些患者會以彎腰(bow)和抖動(jerk)來表現疼痛。

Corneliu Florea [14]最後認為，知名的 UNBC Pain Database 雖然開創了以臉

部表情進行非入侵式疼痛估計的研究，使這個領域奠定了相當的研究基礎，但該資料庫應該再提供更具變異性的疼痛表情案例。

### 2.3.2 The UNBC-McMaster Shoulder Pain Expression Archive Database 的 Frame level Ground truth 品質探討

UNBC Pain Database 提供的 frame level ground truth 為 PSPI，卻有高達 82.7% 的畫面 PSPI 都是 0 分。於是本論文針對此現象，進一步分析該資料庫所提供的各項數據。

由於 PSPI 為 FACS 依式 2-1 換算而來，本論文嘗試重新換算時，卻發現高達 69.1% 樣本的 FACS 都沒有資料；而這 69.1% 缺少 FACS 的樣本 PSPI 都剛好被標示為 0 分。由此可見，PSPI 為 0 分的樣本中，有高達 7 成是無資料可計算，並非真正 PSPI 為 0 分，是尚未經過 PSPI 計算疼痛程度的影像。(如表 2-3)(如圖 2-6)

而 FACS 是又怎麼來的？UNBC Pain Database 提供的資料包含每一張 frame 以 AAM(Active Appearance Model)取得的人臉特徵點座標資料(AAM landmarks)(如圖 2-7)。雖然該資料庫提供的 AAM landmarks 相當完整，每一個 frame 都有 66 點，並無短缺，但受限於捕捉當時角度與光線的干擾因素，AAM landmarks 不一定每一點都正確，即使都正確，也可能因為角度偏轉過大，不一定能換算出對應的 Action Unit(如圖 2-8, 圖 2-9, 圖 2-10)。

基於上述原因，FACS 的資料不完整是可推論的，也是難以避免的研究限制。但是對於缺乏 FACS 資料而無法換算出 PSPI 的樣本，就直接將 PSPI 標示為 0 分，這種作法是否妥當呢？

在抓不出任何 AU 的前提下，就將該人臉影像 PSPI 訂為 0 分，等於是找不出任何表情成分，卻又否定這張人臉影像存在任何疼痛成分。舉例來說，有一個內容物未知的盒子 A，由於各種技術限制而無法得知其內容，但不能

因為這樣就否定盒子 A 的內容物未含有 B 成分存在。

本研究認為，若某張人臉影像的 PSPI 被評為 0 分，其 FACS code 應該至少存在一個 AU；若是在擷取不出任何 AU 的前提下，該人臉影像的 PSPI 應該是未定義。

若在某張影像 FACS 資訊為空的前提下，就將 PSPI 定為 0 分的處理頗為草率；對於 FACS 資訊為空的影像，應該用其他標籤與「具有完整 FACS 資料並換算出 PSPI 為 0 分的樣本」作出區別。

表 2-3 The UNBC-McMaster Shoulder Pain Expression Archive Database 資料缺陷分析

	所有樣本	PSPI 被標示為 0 分的樣本	並無 FACS 資料可供換算出 PSPI 的樣本
數量	48398	40029	33484
百分比	100%	82.7%	69.1%
		100%	83.6%

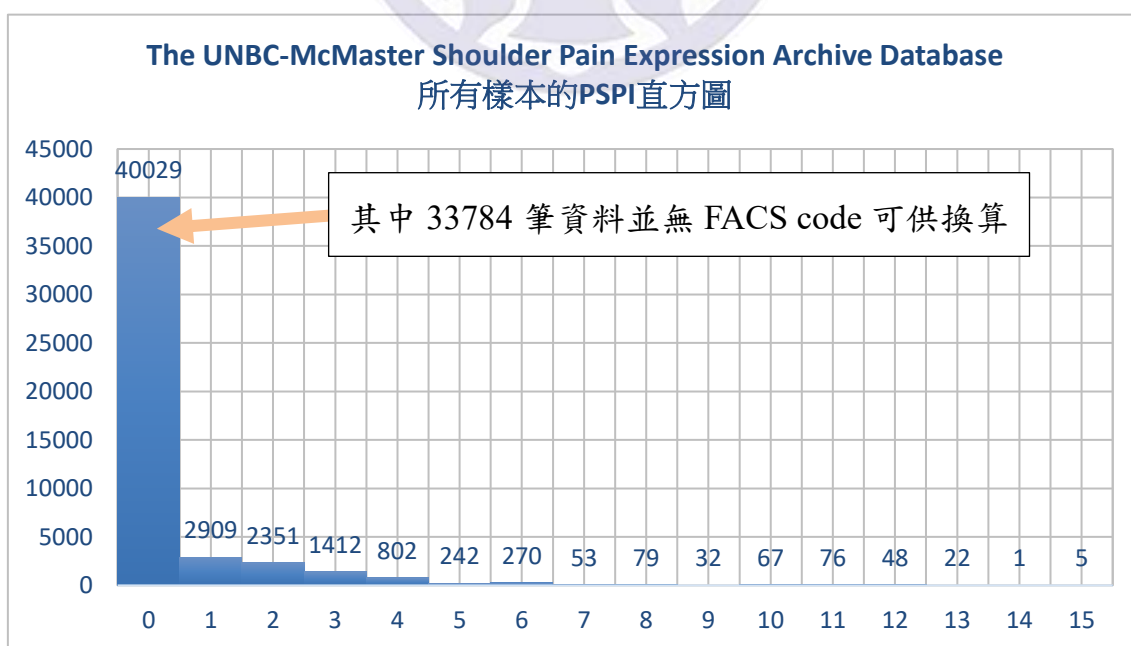


圖 2-6 UNBC Pain Database 所有樣本的 PSPI 直方圖

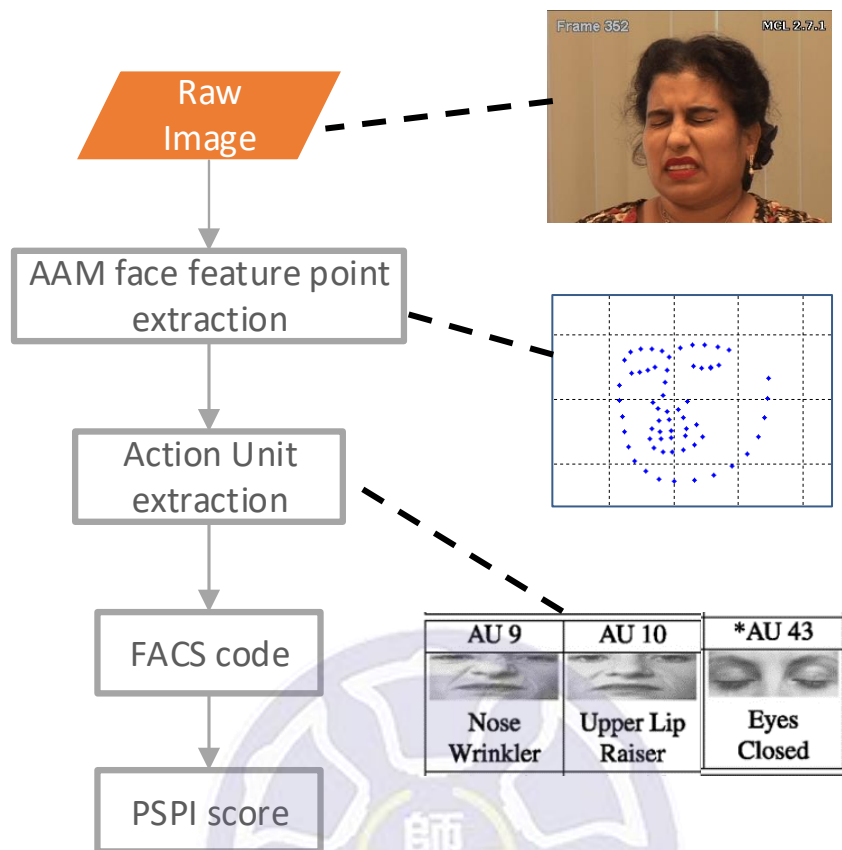


圖 2-7 UNBC Pain Database 從原始影像計算出 PSPI 的流程圖

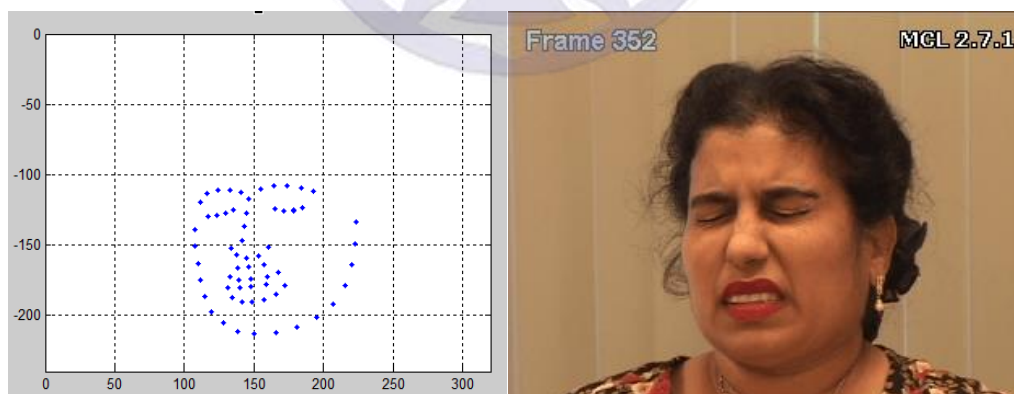


圖 2-8 AAM landmarks 與原始影像對照，檔案名稱為 hs107t2aaaff352。此樣本 PSPI 為 15 分。



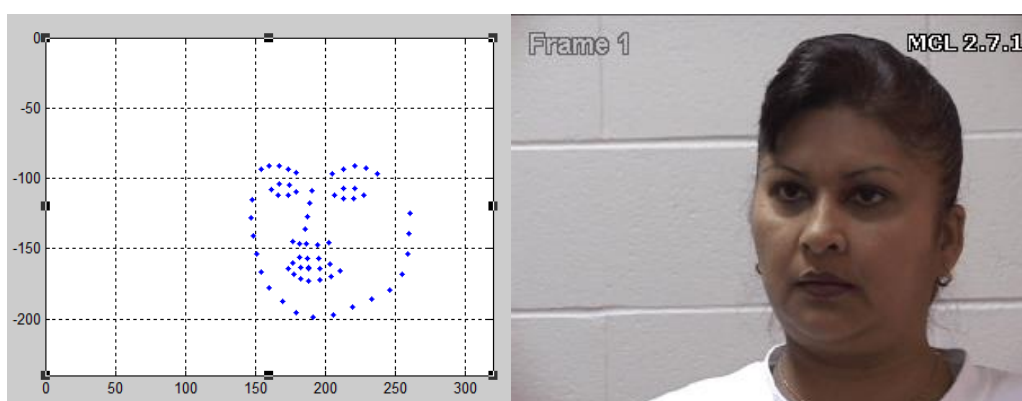


圖 2-9 AAM landmarks 與原始影像對照，檔案名稱為 ll042t1aaaff001。此樣本無法解析出 FACS，PSPI 被標示為 0 分。

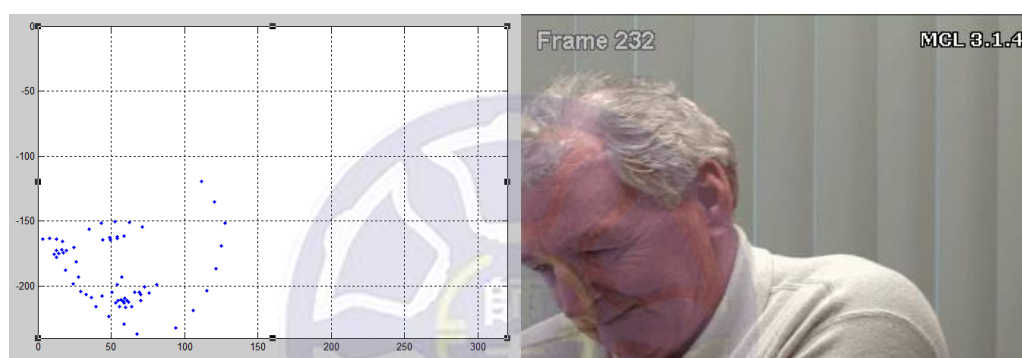


圖 2-10 AAM landmarks 與原始影像對照，檔案名稱為 jk103t2aeaff232。此樣本被解析出的 Action Unit 為 AU25 與 AU26，PSPI 被標示為 0 分。

### 2.3.3 痛苦表情辨識速度研究回顧

在為數不多的 20 餘篇疼痛表情相關研究中，且皆未提及演算速度。於是本論文逐一寫信向各疼痛表情辨識相關研究的通訊作者請教，詢問研究成果的演算速度、程式語言與硬體平台，最後僅有大約 1/3 作者回應，其中有提及演算速度得回應更少。得到的回應整理如表 2-5

表 2-5 多篇疼痛表情辨識研究的硬體平台與執行速度

通訊作者	文章標題	硬體	程式語言	速度
Liliana Lo Presti (2015)	Using Hankel Matrices for Dynamics-based Facial Emotion Recognition and Pain Detection	3.4GHz, 8core 12GB	C++	100fps
Corneliu Florea (2015)	Pain Intensity Estimation by a Self-Taught Selection of Histograms of Topographical Features	Intel Xeon 3.3 GHz	Matlab	0.15fps (on a single thread)
Kaltwang, Sebastian (2012)	Continuous Pain Intensity Estimation from Facial Expressions	3.3 GHz, 4core 16 GB	Matlab	1515fps
Muhammad Naufal Bin Mansor (2014)	Combination Filters for COPE Database	Intel® Core™ 2Duo CPU, 2.80 GHz	Matlab	12.5fps
Karan Sikka (2014)	Facial Expression Analysis for Estimating Pain in Clinical Settings	不記得	不記得	25fps
Rizwan Ahmed Khan (2013)	Pain detection through shape and appearance features	3GHz	Matlab	未回應
Maryam Moosaei (2014)	Naturalistic Pain Synthesis for Virtual Patients	i5, 8GB	C++	未回應

### 第三章 類神經網路介紹

在各種人工神經網路中，單層感知機(Single-layer Perceptron)是最簡單且最早實現的形式。因此，若要從頭了解類神經網路的原理，就要先從類神經網路的單細胞—單層感知機談起。

#### 3.1 單層感知機與倒傳遞

單層感知機(如圖 3.1)的結構依序為：

1. N 個輸入點  $a_1 \sim a_n$
2. N 個路徑權重  $w_1 \sim w_n$
3. 一個基底常數  $b$
4. 一個加總節點 sum
5. 一個活化函數  $f$
6. 一個純量輸出  $t$

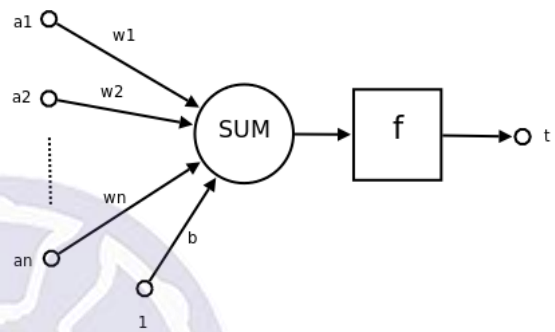


圖 3-1 單層感知機

輸出點  $t$  方程式如式 3-1、函數  $f$  的方程式如式 3-2。可以看出一個感知機本質上就是一組線性方程式， $f$  則是對方程式輸出做二元化處理。

$$t = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (3-1)$$

$$f(n) = \begin{cases} +1 & \text{if } n \geq 0 \\ -1 & \text{otherwise} \end{cases} \quad (3-2)$$

具學習能力的單層感知機，會將每一次輸出  $t$  與目標值之間的差異量，回傳到  $w_1 \sim w_n$  與  $b$ ，使得  $w(n+1) = w(n) - \text{err}$ ,  $b(n+1) = b(n) - \text{err}$ ，這個更新權重  $w_1 \sim w_n$  與基底常數  $b$  的過程也稱為倒傳遞；若輸出  $t$  與目標值吻合，權重就不會更新。經過不斷迭代的倒傳遞，使得單層感知機輸出越來越精準，

直到達成夠小的誤差、或一定的迭代量為止。此時，已訓練過的感知機就是一個良好的分類器，訓練所得成果就是路徑權重  $w_1 \sim w_n$  與基底常數  $b$ 。

筆者過去曾將身體活動運動學資料輸入單層感知機，訓練出效果良好的下樓梯動作分類器，並在 22MIPS 的 8 位元單晶片中實現即時回想，由此可見單層感知機的效果強大。

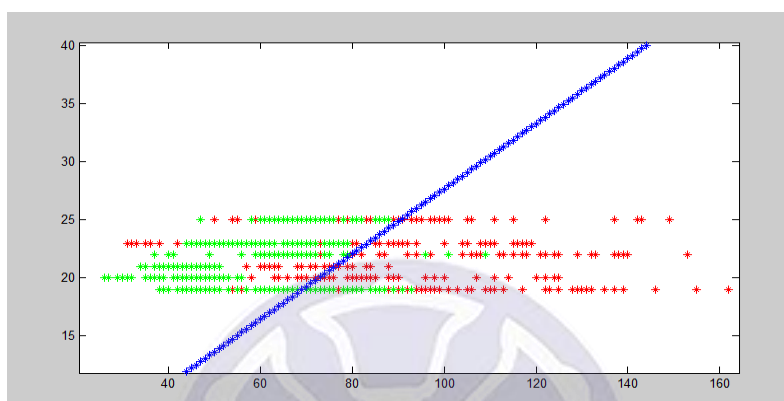


圖 3-2 口袋式單層感知機學習人體運動的分類結果

表 3-1 圖 3-1 的說明

X 軸	受試者下背垂直加速度 1 秒內的峰對峰 ADC 數值
Y 軸	受試者 BMI
紅色點	人工標記的下樓梯資料點
綠色點	人工標記的上樓梯或平地走資料點
藍色點	單層感知機經由倒傳遞習得的分類直線

從如圖 3-2 可以看出該兩類並非線性可分，有少數資料在特徵空間中交錯分布，造成訓練過程中無法保證誤差收斂，這也是感知機訓練實際資料幾乎都會遇到的問題，但使用口袋式感知機(Pocket Perceptron)仍可收斂到一組區域或全域最佳解。

### 3.2 多層感知機

多層感知機 (Multilayer Perceptron) 是單層感知機的推廣，由多個層組成，每一層可以有多個節點，每一個節點會連結到下一層的所有節點。此架構克服了單層感知機無法解決如 XOR 等線性不可分問題的弱點。

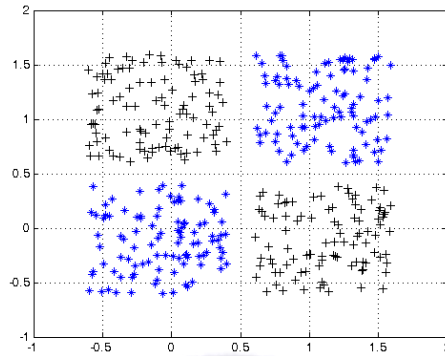


圖 3-2 一個二維 XOR 問題的範例。

如果只是簡單多層連結，那麼任意層數的多層感知機都可以被簡化成等效的單層感知機。因此，多層感知機的另一個重要差異，在於活化函數 (Activation Function) 必須可微且非線性，通常為 sigmoid function (如式 3-3)，本論文基於實作考量改以 Tanh(n) 代替。

$$f(n) = \frac{1}{1 + e^{-n}} \quad (3-3)$$

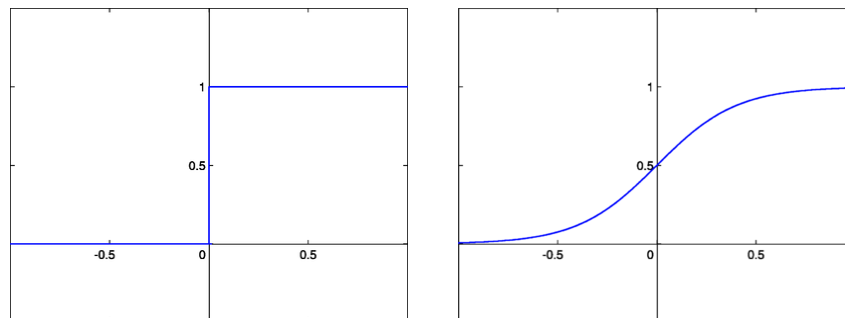


圖 3-3 不同層數感知機的活化函數，左邊屬於單層，右邊屬於多層。

一般的多層感知機可以粗分成三大層架構 (如圖 3-4)，細節分別敘述如下：

- 輸入層 (input layer)：輸入層只有一層，屬於不可訓練層，不具任何權重

參數，僅作為接收資料用。此層節點數目只與輸入資料維度有關，若輸入  $N$  維資料，輸入層就有  $N$  個節點。

- 隱藏層(hidden layer)：隱藏層可以有一層以上，寬度可任意決定，也可以每一個隱藏層都不同寬度，為調整彈性最大的一層。隱藏屬於可訓練層，絕大多數權重參數都集中在此層。但節點數或層數越多，訓練效果不一定越好，也有可能反而下降或過度擬合(Over fitting)，因此這一層的各项參數也是訓練時得調教重點，對訓練成效的好壞有直接影響。但如何設定參數？如何調教？並沒有公認標準，這部分相當依賴經驗與各領域的 domain knowledge。
- 輸出層(output layer)：輸出層只有一層，屬於可訓練層，但僅具少量權重參數，主要作為輸出資料用。這一層的節點數只與輸出資料有關，以下分別就兩大類監督式學習(Supervised Learning Network)探討輸出層：
  - 迴歸(Regression)：視被估測的純量數目而定。本研究就是迴歸類神經網路的實驗，因為被估測變數只有 PSPI 一項，所以輸出層只有一個節點(如圖 3-4)。由於迴歸計算的輸出沒有固定範圍，因此輸出層沒有活化函數。
  - 分類(Classification)：若有  $N$  類輸入資料，則輸出節點至少  $N$  個，也有可能多一個“以上皆非”的第  $N+1$  類輸出節點。由於分類器需要滿足  $P(C_i) \in [1,0]$  且  $1 = \sum_i P(C_i)$  的機率公設，因此分類用的類神經網路大多會以 Softmax function 作為輸出層的活化函數，使每一個輸出值都可以鬆散的解釋為當前輸入樣本屬於該類的機率。

至於第三種機器學習－聚類(Cluster)、分群等非監督式學習(Unsupervised Learning)的輸出層網路結構差異較大，非本研究討論的範圍。



上述多層感知機的架構，就是目前普遍被使用的類神經網路(如圖 3-4)。

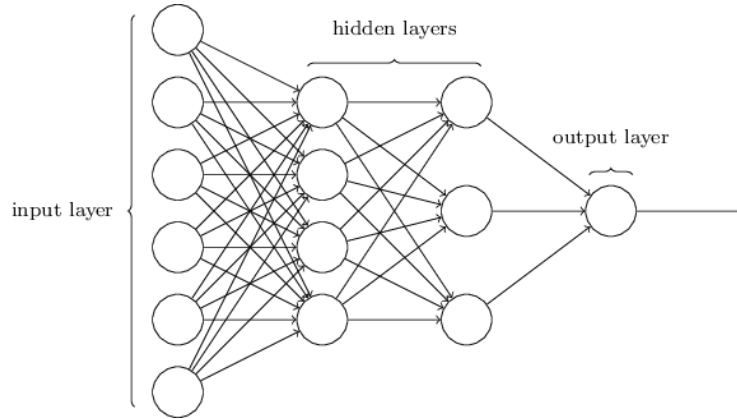


圖 3-4 具有三大層架構的多層感知機

### 3.3 反向傳播演算法

3.1 節談過單層感知機的倒傳遞過程，但在多層感知機有更複雜的學習過程，稱為反向傳播演算法。首先要計算順向傳遞的損失函數(loss function)，即平方差(square error)，取平方是為了避免正誤差與負誤差之間互相抵消。如式 3-4。其中E為平方誤差，t為樣本 Ground Truth，y為感知機實際輸出量。

$$E = \frac{1}{2}(t - y)^2 \quad (3-4)$$

有了損失函數後，便可以求得損失函數對類神經網路中各項參數的偏微分，並以此更新模型內的參數，在一次又一次的迭代訓練中降低損失函數的值，過程如下所示。其中 $\mathbf{W}$ 是包含權重與基底在內的向量， $\mathbf{W}(i)$ 是第i次迭代的結果， $\eta$ 是學習率(learning rate)。

$$\Delta \mathbf{W}(i) = -\eta \cdot \left. \frac{\partial E}{\partial \mathbf{W}} \right|_{\mathbf{W}=\mathbf{W}(i)} \quad (3-5a)$$

$$\mathbf{W}(i + 1) = \mathbf{W}(i) + \Delta \mathbf{W}(i) \quad (3-5b)$$

$$E(\mathbf{W}(i + 1)) \leq E(\mathbf{W}(i)) \quad (3-5c)$$

學習率設定是一個兩難問題，因為較大的學習率可以讓訓練初期收斂快，但後期可能會造成震盪誤差過大，不容易收斂到全域最佳解(global

minimum)；而較小的學習率除了收斂速度慢，也比較容易跳不出區域最佳解 (local minimum) [15]。

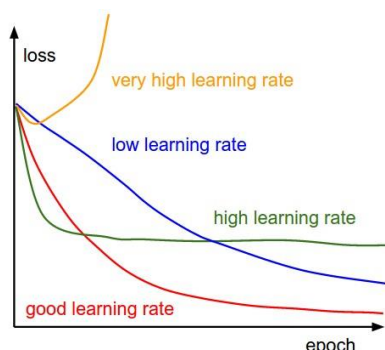


圖 3-5 學習率、損失函數與迭代次數三者之間的關係示意圖 [16]。

### 3.4 迴歸類神經網路

#### 3.4.1 迴歸簡介

迴歸是一種統計學上的預測方式，藉由歷史資料建立迴歸函數，再以此迴歸函數推論未知輸入值的輸出結果。常見的模型有線性迴歸、多項式迴歸等方法，本論文則是採用類神經網路迴歸。

以圖 3-6 為例，座標中的點就是歷史資料，座標中的線就是已建立的迴歸函數  $f(x)$ ，此迴歸函數可用  $x$  值預測任意未知資料的  $y$ 。星號代表一筆未知資料  $[x^*, y^*]$ ，模型預測所得數值為  $f(x^*)$ ，圖中的藍色雙箭頭長度為實際與預測的差距量，為  $y^* - f(x^*)$ 。

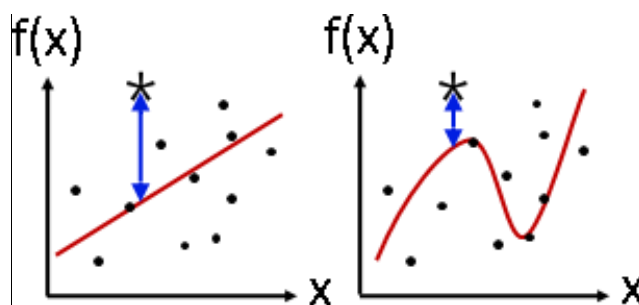


圖 3-6 線性迴歸、多項式迴歸 [17]

比起線性迴歸、多項式迴歸這兩種常見的模型，類神經網路迴歸對高維度複雜樣本有更佳描述力。



### 3.4.2 以類神經網路迴歸 $\sin(x)$ 範例

為了讓讀者具體了解類神經網路迴歸計算的迭代過程，本節將示範使用類神經網路對  $\sin(x)$  做迴歸訓練與估測。由於  $\sin(x)$  並不收斂，因此是非常難建模的函數。

首先取得訓練資料，在  $0 \sim 2\pi$  區間隨機挑選 80 個  $x$  值，並以  $\sin(x)$  產生  $y$  值作為 ground truth，如圖 3-7(a)。再以規格如表 3-2 與結構如圖 3-7(b) 的類神經網路進行訓練。最後，將訓練途中的回想結果視覺化(如圖 3-8)，可將類神經網路的機器學習過程生動呈現。圖 3-7(b) 所代表的多項式如式 3-6。

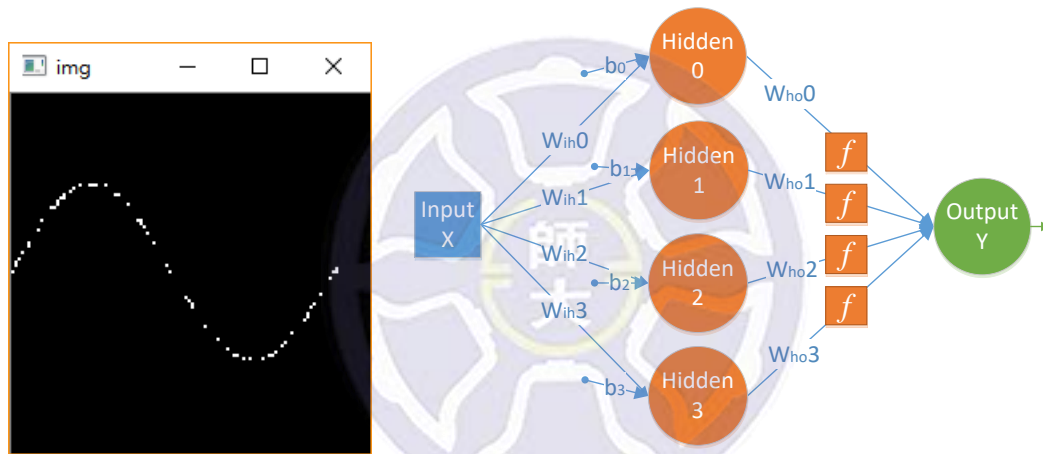


圖 3-7(a)(b)  $\sin(x)$  迴歸 a.訓練用資料 b.訓練用迴歸類神經網路

$$y = \sum_{j=0}^3 W_{hoj} \left( \text{Tanh}(W_{ihj}X + b_j) \right) \quad (3-6)$$

表 3-2 類神經網路規格與訓練參數

節點數與層數	1-4-1
活化函數	Tanh(x)
學習率	0.05
倒傳遞慣量	0.001
迭代次數	1000
初始權重	區間[-0.001, 0.001]內隨機數值

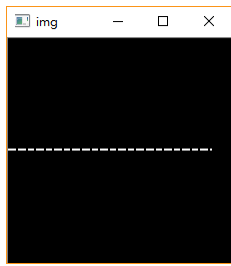
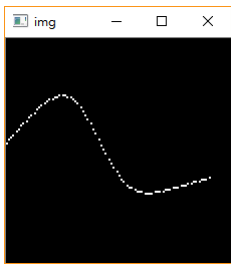
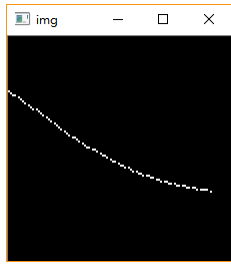
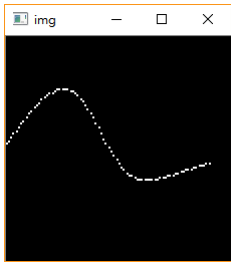
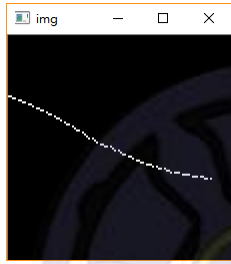
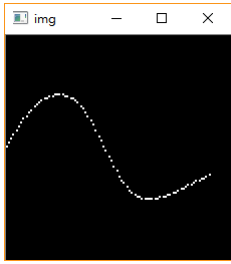
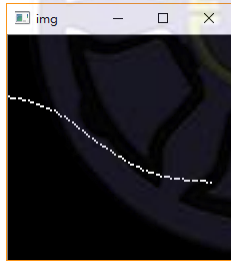
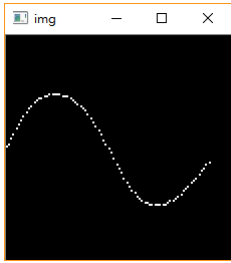
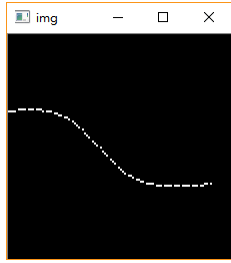
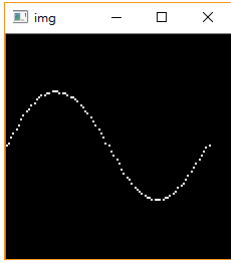
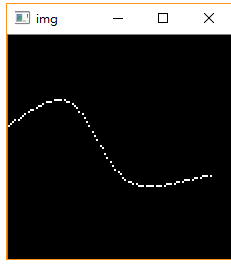
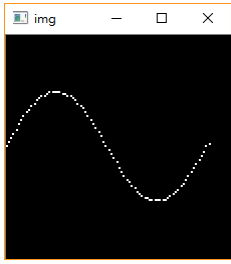
迭代第 0 次 MSE=0.4541		迭代第 60 次 MSE=0.0328	
迭代第 10 次 MSE=0.2121		迭代第 70 次 MSE=0.0480	
迭代第 20 次 MSE=0.1865		迭代第 100 次 MSE=0.0276	
迭代第 30 次 MSE=0.1496		迭代第 300 次 MSE=0.0125	
迭代第 40 次 MSE=0.1153		迭代第 600 次 MSE=0.0005	
迭代第 50 次 MSE=0.0548		迭代第 1000 次 MSE=0.0001	

圖 3-8 訓練途中的回想結果

### 3.5 學習模式

#### 3.5.1 線上學習

隨著網際網路蓬勃發展，數據越來越廉價，有很多即時、不間斷的數據隨時產生，線上學習(Online Learning)就是充分利用資料、隨時靈活更新權重的一種方式。

線上學習於每一次倒傳遞後會立即更新權重，用於沒有固定模式、或模式不停流動的狀況，所有資料用完一次即丟棄，例如運動控制所用的類神經網路。此模式下沒有明顯區別訓練與測試，即使是使用或驗證的過程中，權重依然會不斷被改寫，但會有一較佳的初始化權重。

#### 3.5.2 批量學習

批量學習(Batch Learning)於順傳遞後不會立即倒傳遞，會等待訓練集完成一輪訓練後，以所有樣本損失函數的平均值來倒傳遞更新權重。使用或驗證的過程中就不再倒傳遞，所以權重也固定不變。優點是當樣本普遍含有雜訊時，平均過的損失函數可以降低雜訊影響。

#### 3.5.3 隨機學習

隨機學習(Stochastic Learning)為本論文採用的方式。隨機學習每一個樣本訓練完後會立即倒傳遞更新權重。此輪訓練完後，將訓練集的順序隨機洗牌，然後再進行下一輪訓練。類似批量學習，使用或驗證的過程中就不再倒傳遞，所以權重也固定不變。

另有一種新型的隨機學習法稱為 Mini-batch [18]，為目前已知收斂相當快速的學習法之一。若一個訓練樣本  $M$  含有類別  $C = \{C_1, \dots, C_n\}$ ，將  $M$  隨機分割為多個子集  $m = \{m_1, \dots, m_k\}$ ，因此共有  $M/m$  個子集，但任意子集  $m_i$  中必須包含完整的類別  $C = \{C_1, \dots, C_n\}$ ，每次迭代只取其中一個子集  $m_i$ 。Mini-batch 學習法的原理是盡量避免連續訓練到同類的樣本，可有效提高平均學

習量。

## 3.6 訓練數據調教

### 3.6.1 資料標準化

類神經網路的資料標準化很少在公開文獻上探討。本論文為了讓研究成果有較高重用性，大多數實驗都以未經包裝的 C#程式碼進行，未將資料標準化使得研究初期效果不佳。資料科學家 James McCaffrey 就在自己的 Blog 上寫過這麼一段話 [19]:「Although data standardization is not a glamorous topic, understanding data encoding and normalization is an absolutely essential skill when working with neural networks」。

理論上，不做資料標準化還是有可能收斂到一組好的解，因為在倒傳遞過程中，所有可訓練權重(Weight)、基底(bias)都會相應的縮放。實例上，3.1 節所分享的身體活動分類實驗，就是一個未將資料標準化的類神經網路訓練過程。

那為何還要將資料標準化？基於三個原因：

- 較不容易陷在局部最優解(Local Optimum)
- 更快的誤差收斂：以個人電腦進行影像辨識的類神經網路訓練，通常動輒數小時到數十天 [20]，因此加快誤差收斂有莫大幫助。
- 修正因資料尺度造成的權重不平衡：以表 3-3 作為例子，可看出資料標準化前後的差異。若要以二維資料[台幣月收入,年齡]來預測是否已婚？則數值範圍可能在數千到數萬的台幣月收入，比起數值範圍在 0~100 之間的年齡，台幣月收入這一項數值可能對類神經網路產生較大影響，但台幣月收入與是否已婚的相關程度不見得較高。而經過標準化的資料只會呈現相對大小，因此可以消除上述尺度差異。

至於如何將資料標準化？表 3-4 是 James McCaffrey 建議的標準化方式 [19]。本論文的輸入、輸出資料都是數值，因此對輸入資料採用高斯正規化，輸出資料不需要處理可直接對應到 PSPI。

表 3-3 資料標準化前後對照範例

原始數據	標準化後 (數值採用高斯正規化)
[22000, 18, 未婚]	[-0.90, -1.15, -1]
[36000, 28, 已婚]	[-0.17, 0.67, +1]
[60000, 42, 已婚]	[1.08, 0.48, +1]

表 3-4 建議的類神經網路資料標準化方式

	資料類型	標準化方式
輸入	數值	高斯正規化、最大最小正規化
輸出	數值	不需要

## 第四章 影像前處理

本論文先以 Viola-Jones [21]方法取出人類正面臉部的 ROI(Region of Interest)，往後的所有步驟只處理該 ROI。為了排除誤偵測並加快速度，限制大小在 70\*70 以上。使用的分類器參數檔案是 OpenCV 2.4.6 所提供的 haarcascade\_frontalface\_alt2.xml。

### 4.1 眼嘴區域 Local Binary Pattern

從文獻回顧中得知，眼、嘴與其周圍紋路與疼痛水準有高度相關，因此本論文為了降低資料維度與計算量，針對雙眼與嘴巴與附近區域影像取出 LBP 特徵向量。由於本論文在實驗中發現，Viola-Jones 方法取出的人臉與五官的相對位置非常穩定(如圖 4-1)，因此以固定相對位置(如表 4-1)推定為雙眼或嘴巴區域。

表 4-1 推定的雙眼與嘴巴相對區域

	雙眼(綠色框)	嘴(紅色框)
寬範圍	5%~95%	25%~75%
高範圍	30%~50%	70%~95%

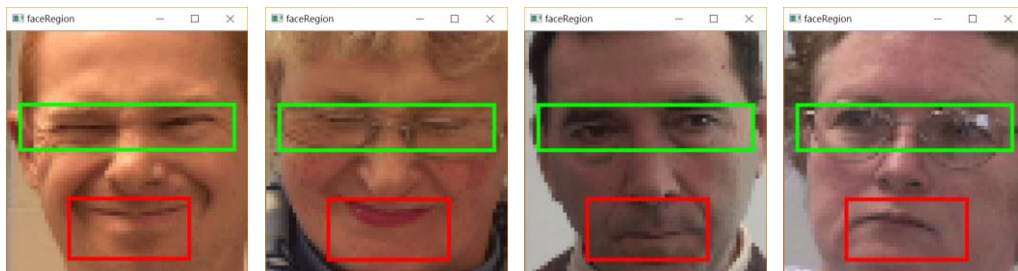


圖 4-1 以推定區域對樣本框選的範例

Local Binary Pattern 是一種對紋理有高度描述力的影像特徵。由於 LBP 數值是透過與周圍像素比較相對亮度所得，因此有非常強健的亮度不變性(Intensity Invariant)。



如圖 4-2，LBP 的計算方式，是將中心像素與周圍 8 個像素進行順時針或逆時針比較，比較結果依序存為一個 8bit 數值，若中心像素較大則為 1，反之為 0。

原始	比較結果	旋轉權重
1	0	1
0	0	128
9	1	64
0	3	32
6		16
8	0	
	1	
	1	

Binary = 0111 0100  
 Decimal = 64+32+16+4 =116

圖 4-2 LBP 的計算範例



圖 4-3 LBP 影像與原始影像對照

8 位元 LBP 共有 256 種，其中許多種經過旋轉後其實是同一種，例如 11100011, 11000111, 10001111, 00011111, 00111110, 01111100, 11111000 都可旋轉到最小值為 00001111。但是好的紋理特徵應該具備旋轉不變性(Rotation Invariant)，於是發展出 LBP-ROT，將所有 LBP 旋轉到最小值，進行多對一映射，於是不重複的 pattern 只剩下 36 種。LBP-ROT 不只增加了旋轉不變性，還可將原本 256 維的 LBP 降到 36 維。

但 LBP-ROT 只對材質有好的描述力，用在表情分析上效果不佳 [22]，所以又發展出將 8bit 視為環狀連結，只計算跳變次數、基於頻率分析的 Uniform LBP。Ojala [23]等學者發現在紋理影像中，跳變次數在兩次以下的

LBP 佔 90%，這些 LBP 共有 58 種，於是將這 58 種頻率較高的 LBP 各自成為一維，剩下在紋理影像中頻率較低的 198 種 LBP 湊成一維，總共 59 個維度，這也是 LBP 最常採用的降維方法。

圖 4-4 是本研究採用的 Lookup Table，以此將 LBP 快速轉換成 Uniform LBP。

```
public static int UniformLookupTable(int LBPvalue)
{
    int[] lookup = new int[256]
    { 0, 1, 2, 3, 4, 58, 5, 6, 7, 58, 58, 58, 8, 58, 9, 10, 11, 58, 58, 58, 58, 58, 58, 12, 58, 58, 58, 13, 58,
      14, 15, 16, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 17, 58, 58, 58, 58, 58, 58, 58, 18,
      58, 58, 58, 19, 58, 20, 21, 22, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58,
      58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 23, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58,
      58, 24, 58, 58, 58, 58, 58, 58, 58, 58, 25, 58, 58, 58, 26, 58, 27, 28, 29, 30, 58, 31, 58, 58, 58, 32, 58, 58,
      58, 58, 58, 58, 33, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 34, 58, 58, 58, 58, 58,
      58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 35,
      36, 37, 58, 38, 58, 58, 58, 39, 58, 58, 58, 58, 58, 58, 58, 58, 40, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58,
      58, 58, 58, 58, 41, 42, 43, 58, 44, 58, 58, 58, 45, 58, 58, 58, 58, 58, 58, 58, 46, 47, 48, 58, 49, 58, 58,
      58, 50, 51, 52, 58, 53, 54, 55, 56, 57 };
    return lookup[LBPvalue];
}
```

圖 4-4 LBP to Uniform LBP Lookup Table

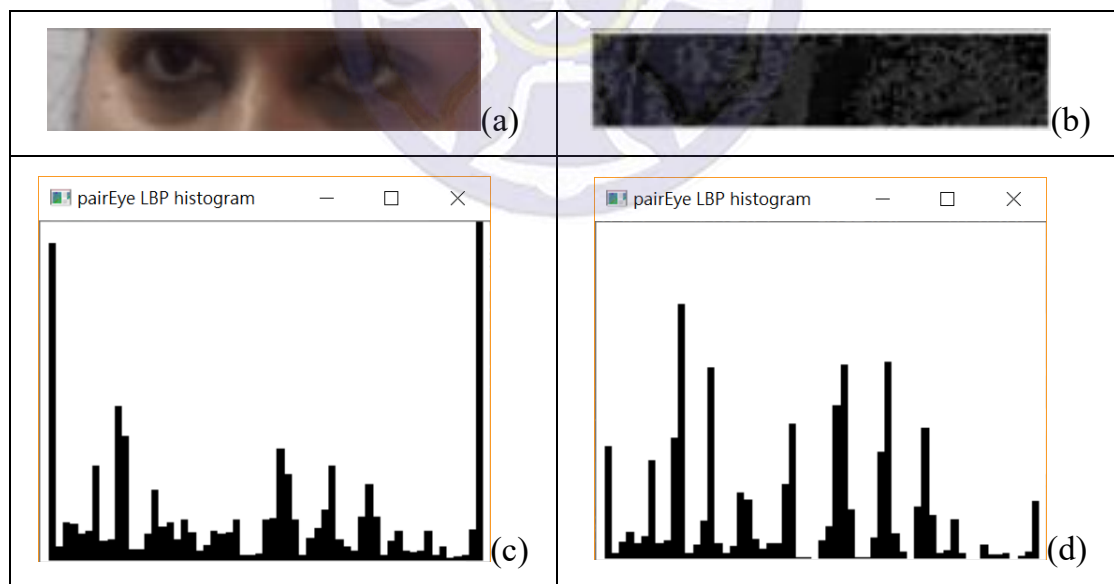


圖 4-5

- (a).雙眼區域高斯模糊影像
- (b).對(a)取 Uniform LBP 影像
- (c).高斯模糊前 Uniform LBP 直方圖
- (d).高斯模糊後 Uniform LBP 直方圖



本論文起初在取 Uniform LBP 直方圖時，發現 Uniform LBP 特徵與疼痛水準相關性極低。但仔細逐張檢視後，發現最大值與最小值(最左側與最右側)常態性偏高(如圖 4-5(c))，若中間數值與疼痛水準相關性高，權重可能會被左右兩側壓過。

回到 LBP 的定義來看，LBP 直方圖的最大值與最小值分別代表哪種外觀意義？LBP 二進位最大值為 11111111，最小值是 00000000，兩者都代表一種區域極端，外觀上會呈現較亮或較暗的獨立點，類似椒鹽雜訊(Salt-and-pepper noise)，但較微弱。

本研究對影像高斯模糊後，LBP 直方圖兩側的相對數值顯著下降(如圖 4-5(d))，表示區域極大或極小的像素點減少很多，實驗中也明顯提高眼嘴區域 Uniform LBP 特徵與疼痛水準的相關性。

## 4.2 池化運算

先前嘗試過眼嘴區域 LBP 特徵以後，並未取得突破性的效果，顯示尚有許多與疼痛水準相關的資訊紋理資訊未能描述，且不在眼嘴周圍的區域中。因此本論文借鏡深度學習 [3]的池化層(Pooling Layer)作為輸入資料。

若將整張影像都輸入類神經網路，資料維度將造成極大的訓練負擔，因此需要有一種能降低維度又能保持特徵的手法，池化就是常見的方法之一。

池化(Pooling)處理是將影像區分成 $m \times n$ 個不相交疊的區域，在每個區域中取其極大、極小或平均值，拋棄其他資訊，然後組成一個新的 $m \times n$ 大小影像(如式 4-1)，此新影像可保留較具代表性的資訊。

$$I_{pooling}(x, y) = \max(I(2x, 2y), I(2x + 1, 2y), I(2x, 2y + 1), I(2x + 1, 2y + 1)) \quad (4.1)$$

其中  $0 \leq x < m$  且  $0 \leq y < n$ , 2 為本論文選用的池化核大小

由於池化運算具高度平移不變性(Translation Invariant)，因此可以有效減

少過度擬合(Over fitting)；反過來說，若輸入資料對平移敏感，那資料只要稍微發生空間上(如影像)或時間上(如聲音)的平移，很容易被模型當作是不同目標看待。如圖 4-6，中間是一個 7x7 大小的隨機數值影像，分別取左上角與右下角，共兩個 6x6 的區域作為平移前後對照。在 6x6 的範圍以 2x2 為單位採用最大池化，輸出 3x3 的池化影像。這兩個 3x3 的池化影像數值相當接近，對照之下可看出池化處理的平移不變性實例。

一般來說，平均池化能夠保留較多背景資訊，最大池化能保留較多紋理資訊。因此本研究對 64x64 的灰階人臉影像，以 2x2 為單位採用最大池化，最後輸出 32x32 的池化影像。

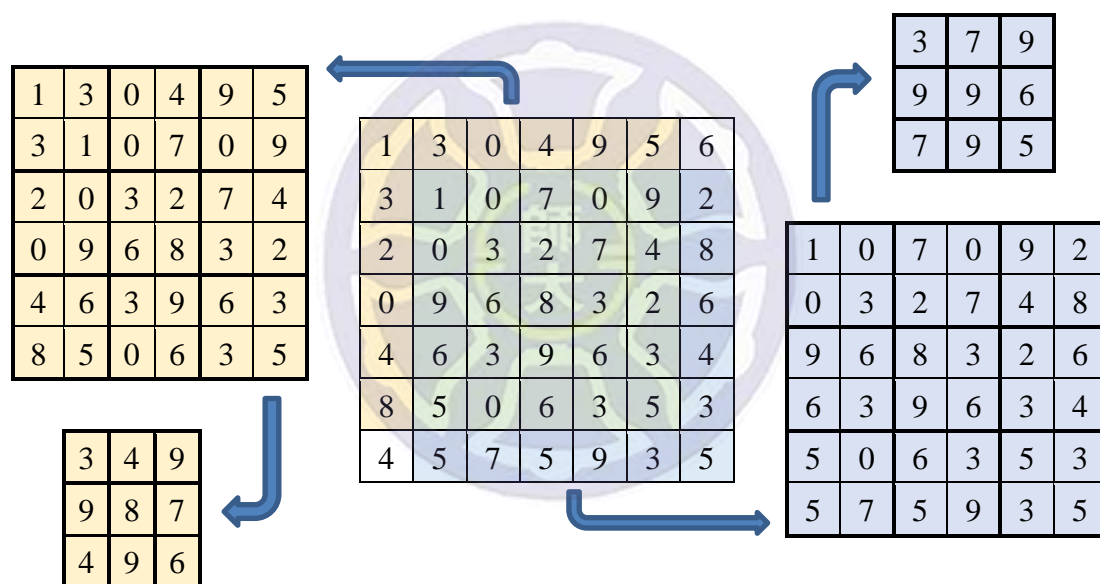


圖 4-6 最大池化運算範例與平移前後對照



### 4.3 資料擴增

在本論文前期的實驗中，以池化影像訓練出來的模型對測試集有不錯的迴歸表現。但實作上，發現估測準確度對於環境光源非常敏感，一般日常生活環境並不容易找到有效的估測位置。若環境光源不佳，類神經網路輸出的估測值通常會停在一個很低的數值附近保持不變，對人臉的各種動作只有微弱反應、甚至是無反應，必須經過多次試誤才能找到一個正確反應的位置。

雖然特徵學習法有找出高階抽象關聯的優點，但必須使用 Raw Pixel 作為類神經網路輸入值，因此對光源敏感的問題將不可避免。

由於池化影像仍屬於可為人類直接解讀的 Raw Pixel 層次資料，含有豐富亮度資訊，不像 LBP 特徵具有非常強健的亮度不變性，因此訓練出來的模型自然也對亮度敏感。本論文進一步推論是該影像資料庫中的光源照射的強度與角度不夠多元，這也呼應了 Corneliu Florea [14]對該資料庫提出變異性不足的看法。

本論文為了解決上述實作缺陷，因此採用資料擴增(Data Augmentation) [3]，用影像處理的方法，使原始資料衍生出許多能提高整體變異性的資料，以便進一步提升被訓練模型的泛化能力。資料擴增與線性插值法(linear interpolation)等數值方法技巧不同，線性插值法僅僅只以數學觀點用舊資料推論出新資料，忽略了應用場合的豐富可能性。資料擴增則是考慮實際情境的各種變化來延伸原始資料。本論文採用以下 3 種擴增方式，因此資料量也提升為原來的 4 倍(如圖 4-9)：

- 水平旋轉(Horizontal Flip)
- 直方圖等化(Histogram Equalization)
- 水平旋轉+直方圖等化

擴增方法	→直方圖等化	
↓ 水平旋轉	 (a)原始灰階影像	 (c)直方圖等化
	 (b)水平旋轉	 (d)水平旋轉+直方圖等化

圖 4-9(a)(b)(c)(d) 資料擴增範例。

另外在實測中發現，資料擴增對微笑略有反應。因此以 Taiwanese Facial Expression Image Database (TFEID) [8]中的快樂(happiness)表情做為新增的 0 分樣本再訓練。該資料庫的快樂表情共有 80 張，隨機分割為各 40 張的訓練集與測試集。實驗結果如表 4-2，顯示可降低對快樂表情的誤判情形。由於此測試集完全是 PSPI=0 的快樂表情，因此無法算出皮爾森相關係數，僅能以均方誤差(Mean Square Error)評估。

表 4-2 以 TFEID 的 happiness 再擴增的實驗結果

訓練集	40 張來自 TFEID 的 happiness Set-A+原資料擴增訓練集
測試集	40 張來自 TFEID 的 happiness Set-B
訓練前 MSE	5.257867
訓練後 MSE	4.307616

## 第五章 實驗結果與討論

### 5.1 實驗設備

#### 5.1.1 EmguCV 3.1

EmguCV 的官方介紹 [24]是「OpenCV in .NET」，是將著名的影像處理類別庫 OpenCV 重新包裝為 Dot Net 類別庫，版本號碼與 OpenCV 同步。特色是將不支援 Dot Net 系列語言(如 C#、VB.Net)的 OpenCV 重新封裝，使 Dot Net 系列語言也能使用 OpenCV。在本論文中主要用來進行 Viola-Jones 法擷取人臉，也用來作影像 IO 與簡單處理。至於主要的影像處理方法，如 LBP、池化等皆自行撰寫類別庫，並未採用 EmguCV 經過高度包裝的類別庫。

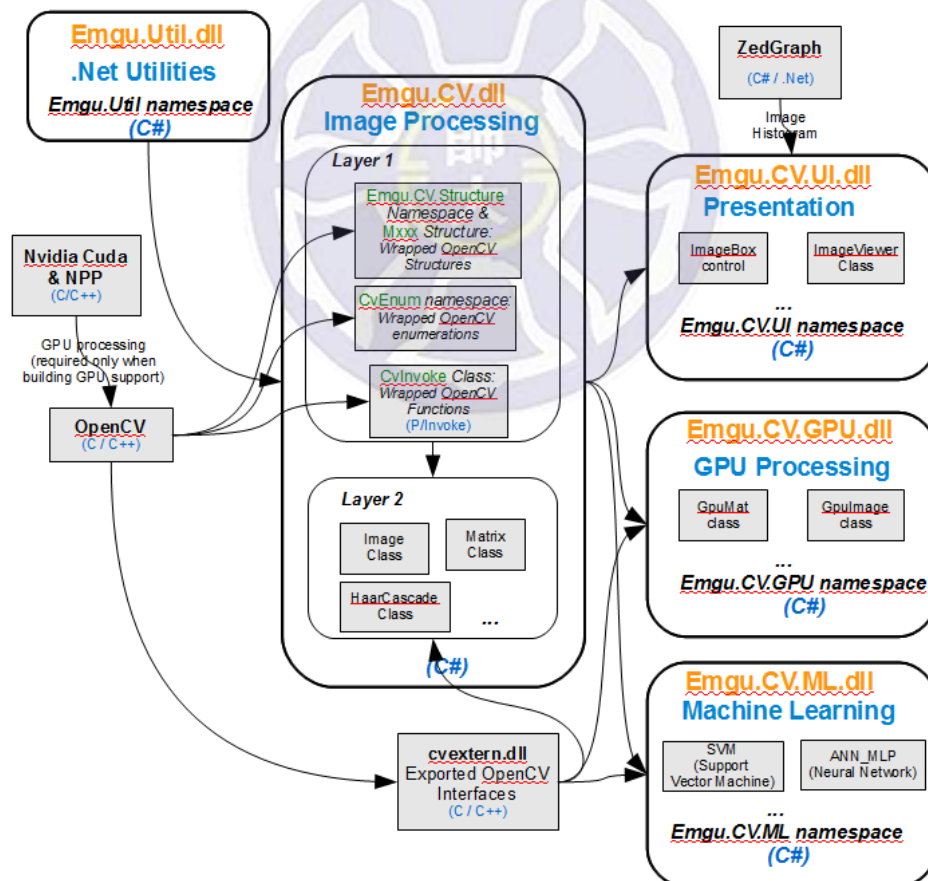


圖 5-1 EmguCV 與 OpenCV 架構關聯示意圖。



## 5.1.2 Microsoft Visual Studio Community 2015

Microsoft Visual Studio 是支援多種程式語言的整合開發環境(Integrated Development Environment，簡稱 IDE)，Community 代表免費版本。

本論文以 Visual Studio 搭配 EmguCV，撰寫 C#與 XAML(Extensible Application Markup Language)程式碼，進行多項實驗，完成最後可執行、可發布、並以 Windows Presentation Foundation (WPF)實作有使用者圖形介面(Graphical User Interface，簡稱 GUI)的即時疼痛水準估測程式。

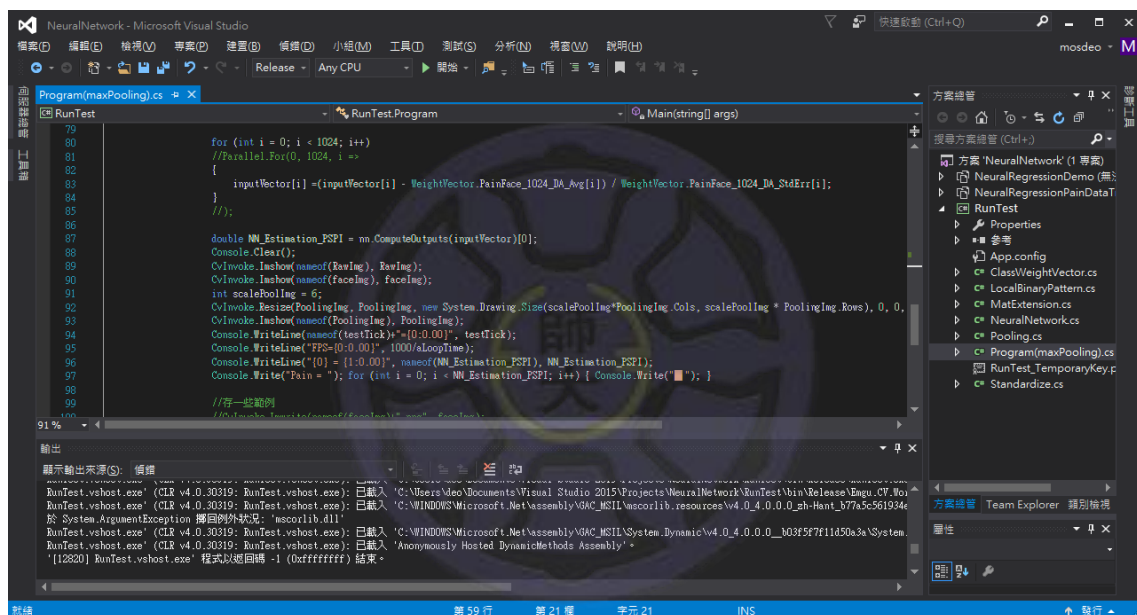


圖 5-2 本論文使用 Visual Studio Community 2015 的開發畫面。

## 5.2 均方誤差

迴歸是一種試圖以函數貼近多筆數據的計算。在機器學習中，希望能根據歷史資料訓練此函數貼近訓練樣本，再以此函數預測未來資料。預測力的指標之一，就是所有測試樣本的均方誤差(Mean Square Error)。

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (5.1)$$

取平方是為了避免正誤差與負誤差之間互相抵消。在本論文中， $X_i$ 為類神經網路輸出的估測值， $\bar{X}$ 為資料庫提供的 PSPI。

## 5.3 皮爾森相關係數

若觀察到  $X$  變化時，另一個變數  $Y$  也以某種方式變化，便可以稱  $X$  與  $Y$  之間彼此相關(correlation)，式 5.2 就是量化相關程度的皮爾森相關係數(Pearson's correlation coefficient)。

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (5.2)$$

式 5.2 在本論文實驗中對應的變數如下：

$\bar{X}$ ：測試集樣本 PSPI ground truth 的母體平均數

$X_i$ ：測試集第  $i$  個樣本 PSPI ground truth

$\bar{Y}$ ：測試集樣本 PSPI 估測值的母體平均數

$Y_i$ ：測試集第  $i$  個樣本 PSPI 估測值

式 5.2 分母可看作是向量  $X$  與  $Y$  的純量長度，分子可看作向量  $X$  與  $Y$

投影長度，所以  $r$  相當於向量  $X$  與  $Y$  的  $\cos(\theta)$ 。因此，從幾何觀點解釋皮爾森相關係數會更為具體 [25]：向量  $X$  與  $Y$  的變化方向越一致， $\cos(\theta)$  越接近 1 或 -1；若向量  $X$  與  $Y$  的變化方向越接近正交，則  $\cos(\theta)$  越接近 0。

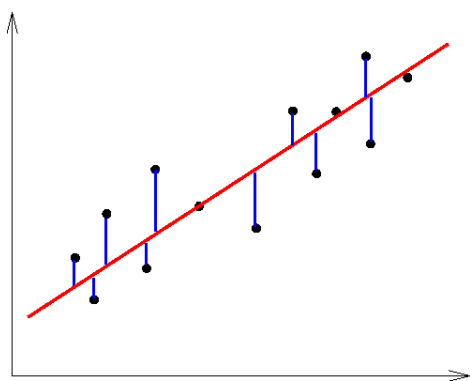


圖 5-3 迴歸誤差示意圖

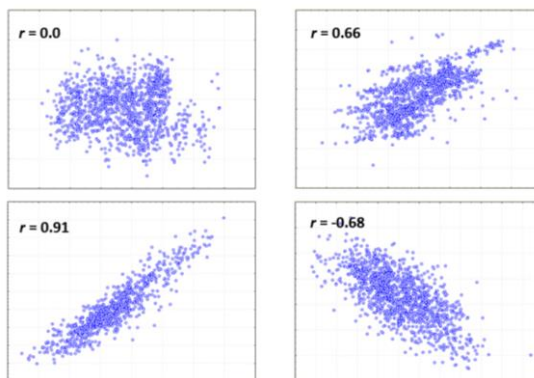


圖 5-4 相關係數與散布圖範例 [25]



## 5.4 混淆矩陣

在機器學習的多類辨識模型評估上，通常用混淆矩陣(Confusion matrix)給予視覺化參考，有時也稱為「分類矩陣」(Classification matrix)。除了能了解各類分類的正確性，也能了解分類錯誤的詳細情形。圖 5-5(a)的混淆矩陣所有數值都集中在對角線上，是一個完美分類的範例，圖 5-5(b)的混淆矩陣所有數值都隨機分散，完全看不出對角線輪廓，是最差的情況。

本論文所用的迴歸類神經網路，雖然不是多類辨識模型，仍可將迴歸結果四捨五入到整數，依 PSPI 的範圍分為 0~15 共 16 類。

為何要將評估多類辨識模型的工具，突兀的套用在迴歸模型上？依照均方誤差公式(式 5.1)，PSPI 高分樣本若有明顯誤差，將會被數百倍之多的 0 分樣本(如圖 2-7)稀釋掉。由於混淆矩陣給予各類相同權重，將可避免上述缺陷。

一般對混淆矩陣的解讀是：「不在對角線上，就是分類錯誤」；但本論文將迴歸估測的結果以混淆矩陣來表達細節，因此與對角線之水平距離同等於估測誤差量，可在混淆矩陣上看出測試樣本的整體分布情形。

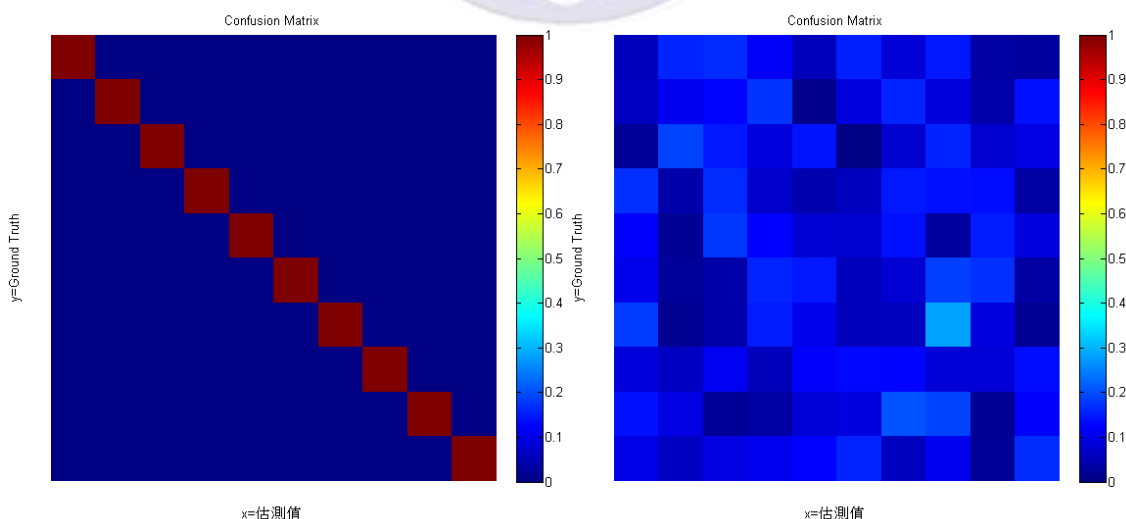


圖 5-5(a)(b) 混淆矩陣範例

## 5.5 實驗方法

如圖 5-6，先將 The UNBC-McMaster Shoulder Pain Expression Archive Database 擷取出人臉影像資料集，有 3.6%(1757 張)無法擷取出人臉，原因是臉部角度偏轉過大、頭髮蓋住額頭、或畫面全黑等嚴重缺陷(圖 5-7)。

對人臉影像資料擷取最大池化與眼嘴 Uniform LBP 直方圖以後，將預處理資料以\*.csv 存檔。預處理資料的內容包含如下：

- 影像檔案名稱，供追蹤估測誤差大的樣本用
- PSPI，作為計算誤差與相關係數用
- 代表人臉影像的 1024 維向量或 118 維向量

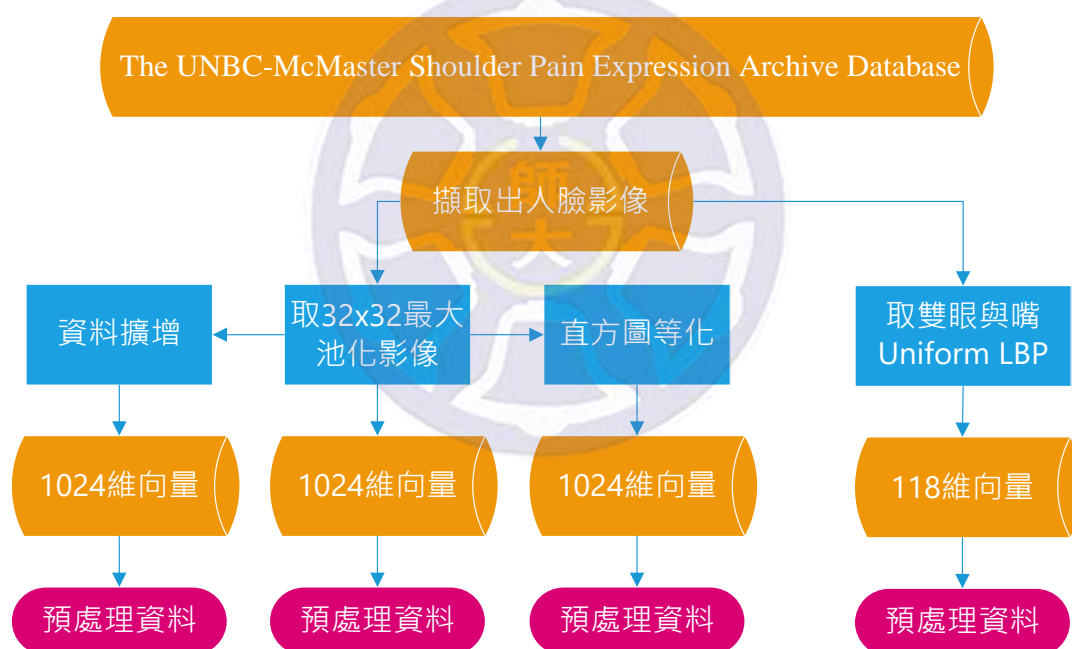


圖 5-6 預處理資料



圖 5-7 無法擷取出人臉的部分樣本



如圖 5-8，將預處理資料隨機分割成用於訓練的 Set-A 和用於測試評估的 Set-B，各占 50%。由於 PSPI 介於 14~15 分的樣本極稀少，尤其 14 分樣本僅一張(如圖 2-7 統計圖所示)，隨機分配容易發生不均勻情形，所以在隨機分配後用人工方式將 14~15 分的樣本調整至平均。

每一次訓練完後，自動讀取 Set-B，以訓練好的類神經網路對 Set-B 估測每一筆向量的 PSPI，並計算皮爾森相關係數與均方誤差，最後自動將計算結果、本次訓練參數、權重與基底等資料存檔。

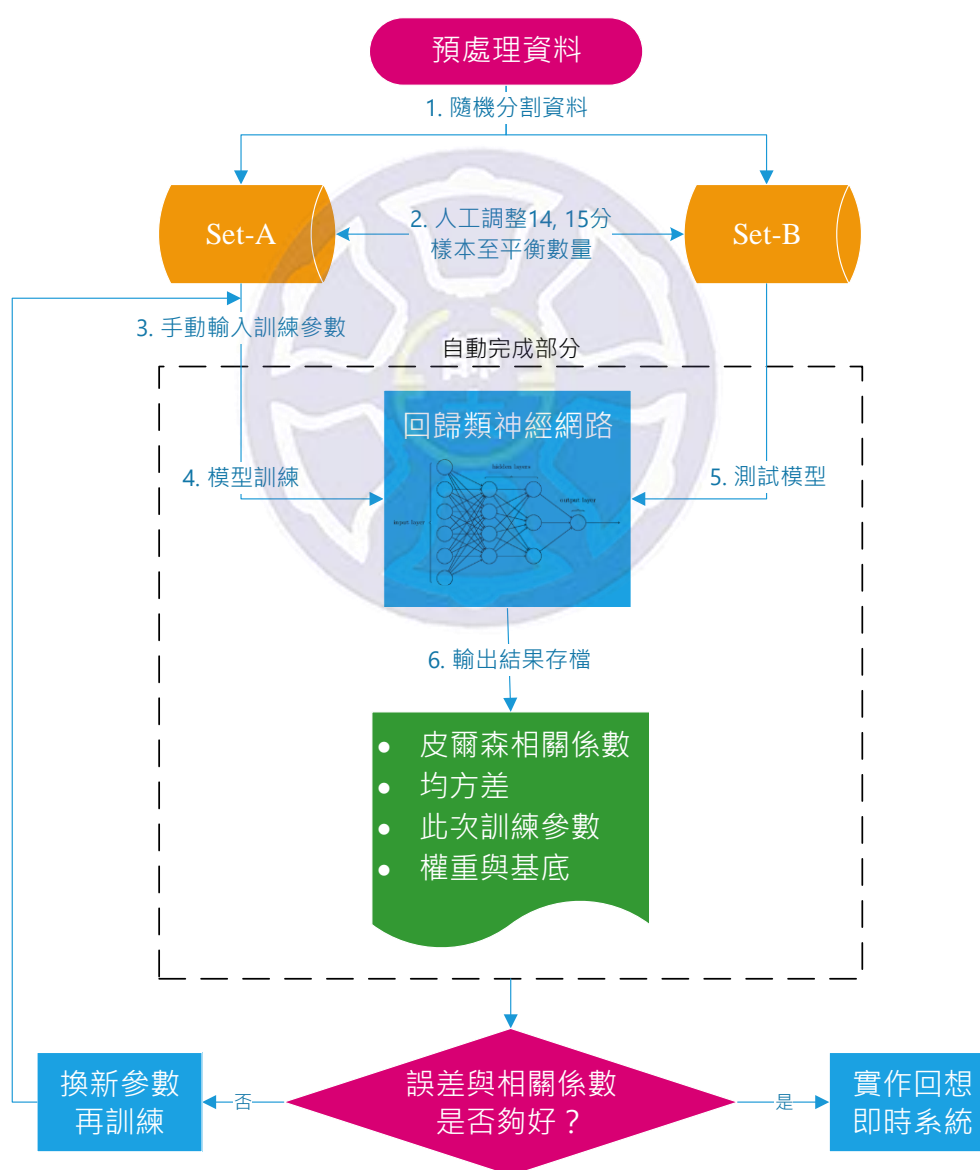


圖 5-8 實驗流程圖

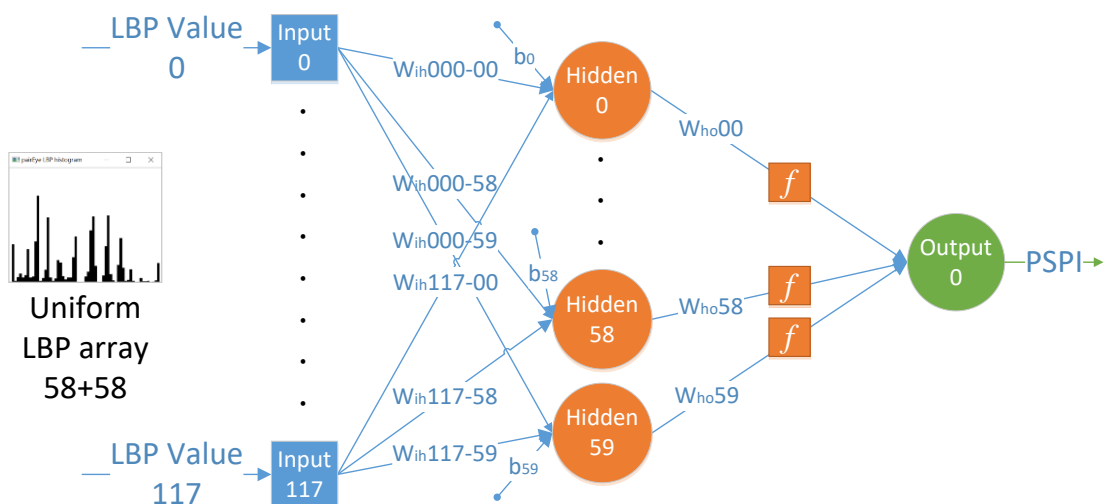


圖 5-9 輸入 LBP 直方圖的迴歸類神經網路

$$PSPI = \sum_{k=0}^{59} \tanh \left( W_{hok} \left\{ \left[ \sum_{p=0}^{117} (W_{ihj-k} X_j) \right] + b_k \right\} \right) \quad (5-3)$$

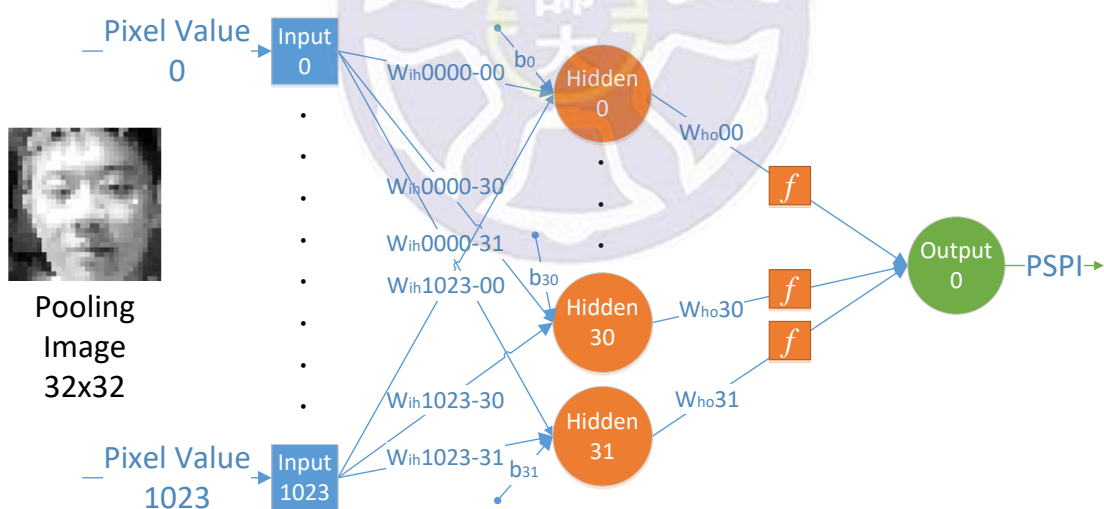


圖 5-10 輸入最大池化影像的迴歸類神經網路

$$PSPI = \sum_{k=0}^{31} \tanh \left( W_{hok} \left\{ \left[ \sum_{p=0}^{1023} (W_{ihj-k} X_j) \right] + b_k \right\} \right) \quad (5-4)$$

## 5.6 實驗結果

本研究提出的雙眼與嘴 Uniform LBP 搭配迴歸類神經網路，所使用的資料維度是三者中最少，但僅在均方誤差有稍佳表現，皮爾森相關係數為三者中表現最差。另外，由表 5-1 得知，本研究提出的最大池化臉部影像搭配迴歸類神經網路，不論在均方誤差或皮爾森相關係數都比 Sebastian Kaltwang [11]等人的研究有更好表現，且所需資料維度不到 1/4。

在最大池化臉部影像的方法中，直方圖等化前處理是一種亮度分布正規化，可以減少樣本間的亮度變異性，因此不論誤差或相關係數皆有改善；資料擴增則是同時提高樣本的亮度與方向變異性，故誤差或相關係數也如預期變差。

表 5-1 本論文與 Sebastian Kaltwang 等人的研究結果比較

作者	Sebastian Kaltwang [11]等人(引用數 62)	Xiaopeng Hong 等人 [26] (引用數 2)	本論文提出			
輸入資料	全臉 Uniform LBP	2Standmap (second-order standardized moment average pooling)	雙眼與嘴 Uniform LBP	Max Pooling		
				原始資料	直方圖等化	資料擴增
迴歸模型	相關向量迴歸(Relevance Vector Regression )	相關向量迴歸 (Relevance Vector Regression )	類神經網路迴歸 (Neural Network Regression)			
			118-60-1 (如圖 5-9)	1024-48-1	1024-32-1 (如圖 5-10)	
資料維度	4779-D (59*81)		118-D (59+59)	1024-D (32*32)		
均方誤差	1.484	1.42	0.38	0.14	0.16	0.23
皮爾森相關係數	0.518	0.55	0.422	0.95	0.95	0.93

表 5-2 使用最大池化原始影像以不同隱藏層節點與學習率訓練出的相關係數

Pearson's correlation coefficient				
Hidden Node Learning Rate	24	32	48	64
0.001	0.953	0.952	0.957	0.944
0.002	0.949	0.955	0.958	0.937
0.005	0.936	0.946	0.956	0.936
0.01	0.907	0.928	0.939	0.915

表 5-3 使用最大池化原始影像以不同隱藏層節點與學習率訓練出的均方誤差

MSE				
Hidden Node Learning Rate	24	32	48	64
0.001	0.1569	0.1588	0.1416	0.1876
0.002	0.1695	0.1495	0.1401	0.2095
0.005	0.2129	0.1801	0.1474	0.2127
0.01	0.3147	0.2444	0.2039	0.2386

表 5-4 實驗結果所用類神經網路的其他規格

隱藏層活化函數	Tanh(x)
輸出層活化函數	無
倒傳遞慣量	0.9
迭代次數	200

圖 5-11 的混淆矩陣為輸入最大池化臉部原始影像的測試樣本評比結果，可看出類神經網路對於 8~10 分的樣本有略為低估的情形，但絕大部分的估測值都準確集中在對角線附近，顯示以類神經網路迴歸能夠有效估計測試樣本的 PSPI。

右下角可看出對角線有一斷點，這是因為唯一一個 14 分樣本被估測為 15 分(向右移動一格)，因此形成圖形上的斷點。

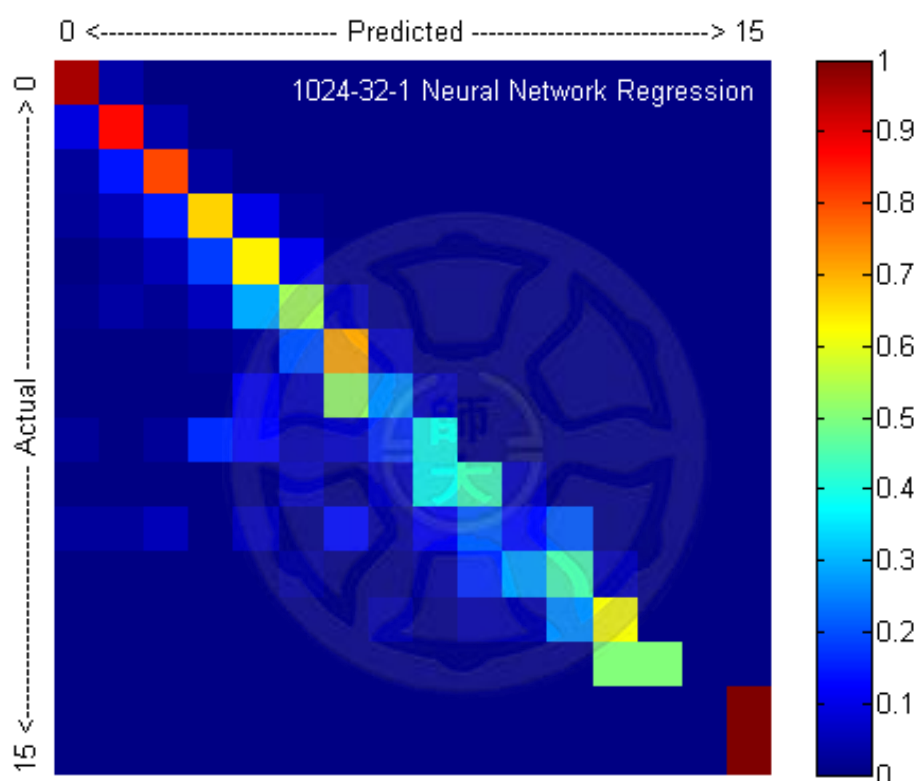
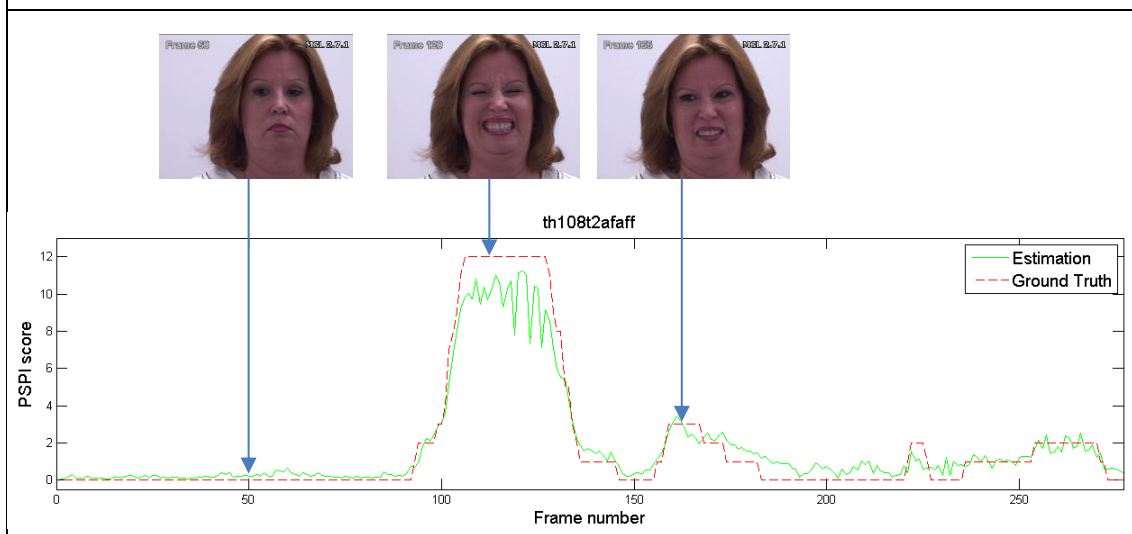
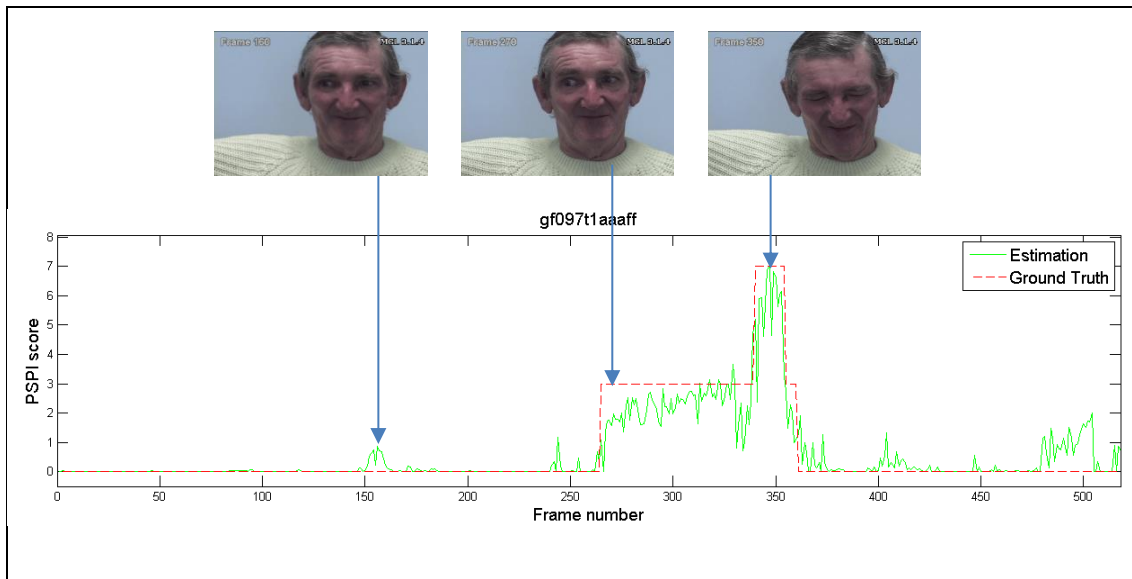


圖 5-11 實驗結果混淆矩陣

圖 5-12 為實驗資料庫其中 6 個影片的 ground truth 與估測值對照，大多數估測值都能跟隨 ground truth 起伏，偶有誤差較大情形。特別的是，在編號為 vw121t1aaaff 的影片中，約第 130~150 frame 之間患者有低頭動作，可能是痛苦造成的彎腰(bow) [14]，但是在此區間的 PSPI ground truth 皆為 0 分，而本論文提出的方法則給予平均約在 2 分左右的 PSPI 估測值。





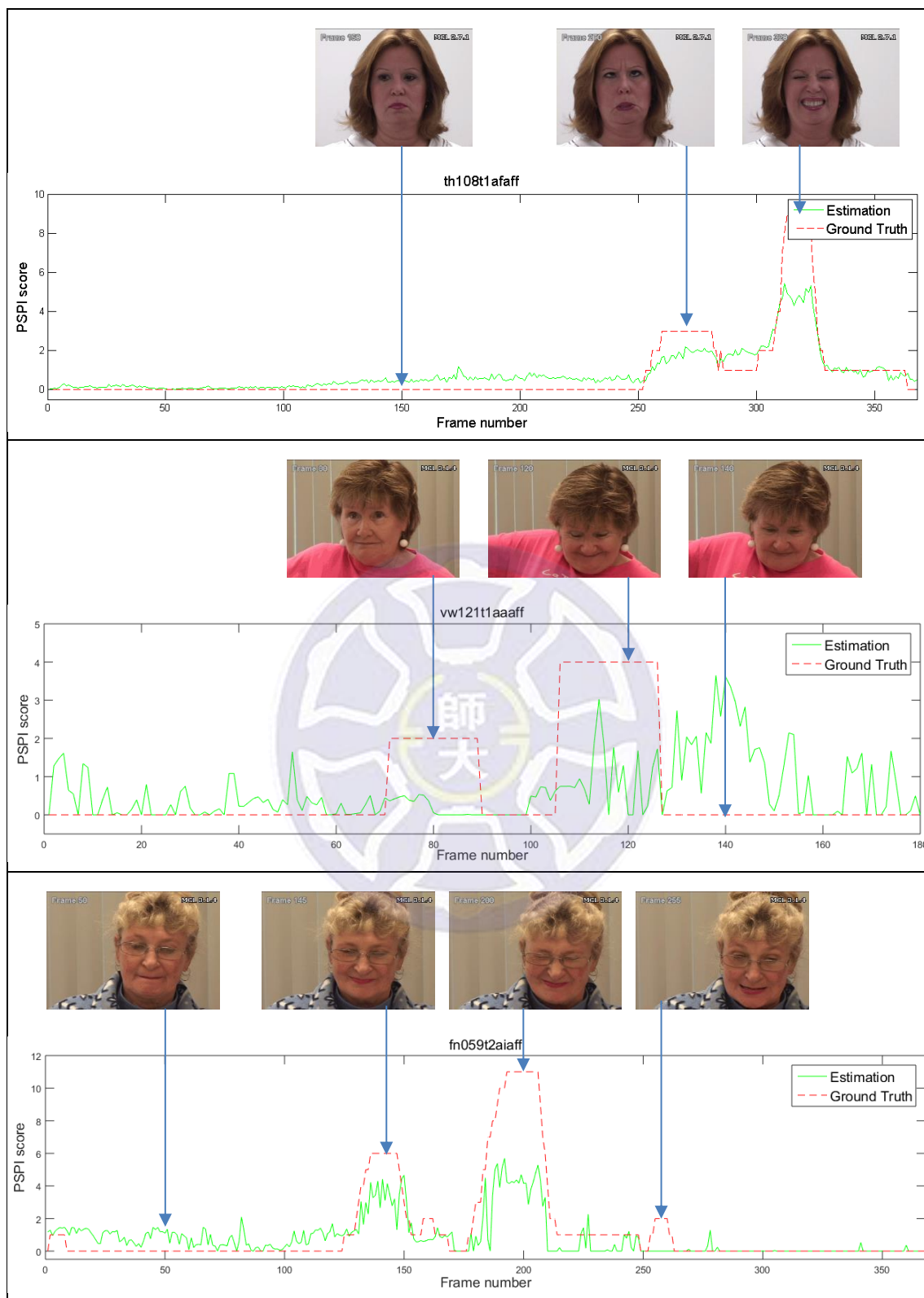


圖 5-12 The UNBC-McMaster Shoulder Pain Expression Archive Database 中數個影片的連續估測與 Ground Truth 比較

圖 5-13 為舉重與演員孕婦生產影片的估測過程截圖。由於畫面跳動相當快速，為了在影片中提供一個方便解讀的介面，畫面中以”\*”數目呈現 PSPI 估測值的四捨五入，將 PSPI 估測值視覺化為”\*”符號字串長度。左上角 FPS 為當下估測速度，維度越大的畫面估測越慢。



圖 5-13 舉重與演員孕婦生產影片的估測過程截圖

## 第六章 結論與未來展望

### 6.1 結論

本論文在文獻回顧中，得知眼嘴區域影像含有最豐富的疼痛水準資訊，因此對其取 Uniform LBP 並輸入類神經網路做迴歸訓練，但估測 PSPI 的效果卻不盡如人意，顯示臉部影像的中的疼痛資訊仍有許多人類難以察覺的抽象部分，僅僅依賴明顯的低階的外型特徵估測疼痛強度，仍有其不足之處。相較之下，Sebastian Kaltwang [11]等人雖然用全臉多區域 Uniform LBP 達到不錯的效果，但相應的代價是極高資料維度。

另一方面，本論文借鏡深度學習的部分知識，將人工對特徵定義的干預降到最低，試圖讓類神經網路自行學習特徵，找出臉部影像與疼痛水準的高階抽象關聯，因此對淺層的傳統類神經網路，輸入以 Max Pooling 簡單降取樣後的 Raw Pixel 進行人臉影像疼痛水準的迴歸估測，最後達到相當準確的估測效果，且實作上對於動態影像可以達到即時回應的 24FPS。

雖然後者的整體統計數據達到相當優異的表現，但仔細逐一檢視個案的 Ground Truth 與估測值(圖 5-12)，仍可發現相當多有待改善的空間。不論是從圖 5-11 或圖 5-12，皆可看出本論文提出方法，對於疼痛水準低估的情形大於高估許多，本論文推測此現象的成因在於 The UNBC-McMaster Shoulder Pain Expression Archive Database 高分樣本過於稀少(如圖 2-7)，造成迴歸訓練時高分樣本的權重修正量不夠，無法有效使超平面(Hyperplane)貼近高分樣本所在的局部特徵空間。高分樣本過於稀少也是本論文最大的研究限制。

最後在實作驗證上，發現特徵學習法會對環境亮度過於敏感，因此以直方圖等化預處理或資料擴增皆能改善此問題。其中，本研究的資料擴增以亮度、方向的組合變化，擴增多出 3 倍資料，不但改善模型的對亮度適應力，也同時提高對臉部方向的適應力。

## 6.2 未來展望

近數年來旋積類神經網路(Convolutional Neural Network)的在電腦視覺領域帶起一股旋風，使得此架構幾乎在影像辨識和偵測上幾乎無往不利。目前可供使用的旋積類神經網路函式庫中，較熱門的有 Caffe、TensorFlow 等。由於本論文未能自行實作出成功的旋積類神經網路，因此列出一些本研究可再以旋積類神經網路提升水準的方向：

- 以 CNN 取代 Viola-Jones 方法擷取出人臉 ROI

抓取臉部的準確程度會大幅影響本研究的量化表現與實用性，而 CNN 已經被證實 [26] 抓取臉部的表現超越 Viola-Jones 方法。

- 改用 CNN 與其他表情合併多類訓練

本論文在實測過程中，發現訓練出的良好模型對微笑表情偶爾會有假警報(false alarm)，雖然本論文已在  $PSPI=0$  的樣本中加入一些微笑表情樣本改善此誤差。但更加徹底的做法，是將疼痛表情估測，融入到目前已經成熟的各種表情辨識中。



## 參考文獻

- [1] "Microsoft Project Oxford Emotion API," Microsoft, [Online]. Available: <https://www.projectoxford.ai/emotion>. [Accessed 2016].
- [2] Theodoridis Sergios, Koutroumbas Konstantinos, in *Pattern Recognition (Fourth Edition)*, Academic Press, 2009, p. 7.
- [3] KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E., "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097-1105., 2012.
- [4] Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., Matthews, I., "Painful data: The UNBC-McMaster shoulder pain expression archive database," in *Automatic Face & Gesture Recognition and Workshops, 2011 IEEE International Conference on* (pp. 57-64), 2011.
- [5] "Facial Action Coding System," Wikimedia Foundation, Inc, [Online]. Available: [https://en.wikipedia.org/wiki/Facial\\_Action\\_Coding\\_System](https://en.wikipedia.org/wiki/Facial_Action_Coding_System). [Accessed 25 2 2016].
- [6] Prkachin, K. M., Solomon, P. E., "The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain," *Pain*, 139(2), pp. 267-274, 2008.
- [7] Lyons, M. J., Akamatsu, S., Kamachi, M., Gyoba, J., Budynek, J., "The Japanese female facial expression (JAFPE) database," 1998. [Online]. Available: <http://www.kasrl.org/jaffe.html>.
- [8] Chen, Li-Fen, and Yu-Shiuan Yen, "Taiwanese facial expression image

database, Brain Mapping Laboratory, Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan," 2007. [Online]. Available: <http://bml.ym.edu.tw/tfeid/>. [Accessed 16 6 2016].

- [9] Loris Nannia, Alessandra Luminia, Sheryl Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artificial intelligence in medicine* 49(2), pp. 117-125.
- [10] Gholami, Behnood, Wassim M. Haddad, and Allen R. Tannenbaum., "Agitation and pain assessment using digital imaging," in *IEEE Engineering in Medicine and Biology Society. Conference (Vol. 2009)*, 2009.
- [11] Kaltwang Sebastian, Ognjen Rudovic, and Maja Pantic, "Continuous pain intensity estimation from facial expressions," *Advances in Visual Computing*, pp. 368-377, 2012.
- [12] 王佳琪, Recognition of Painful Facial Expression using Multiple Kernel Learning, 臺灣新竹: 國立清華大學電機工程學系碩士論文, 2011.
- [13] MLA Chen, Junkai, Zheru Chi, and Hong Fu., "A new approach for pain event detection in video," in *Affective Computing and Intelligent Interaction (ACII), International Conference*, 2015.
- [14] Florea, C., Florea, L., Boia, R., Bandrabur, A., Vertan, C., "Pain Intensity Estimation by a Self-Taught Selection of Histograms of Topographical Features," 26 5 2016. [Online]. Available: <http://arxiv.org/abs/1503.07706>.
- [15] 周揚賀, 基於深度摺積神經網路之影像檢索技術, 臺灣桃園: 國立中央大學通訊工程學系碩士論文, 2015.

- [16] "CS231n Convolutional Neural Networks for Visual Recognition," [Online]. Available: <http://cs231n.github.io/neural-networks-3/>. [Accessed 15 6 2016].
- [17] 演算法筆記, "Regression," [Online]. Available: <http://www.csie.ntnu.edu.tw/~u91029/Regression.html>. [Accessed 26 6 2016].
- [18] Mu Li, Tong Zhang, Yuqiang Chen, Alexander J. Smola, "Efficient mini-batch training for stochastic optimization," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014.
- [19] McCaffrey, James, "How To Standardize Data for Neural Networks," 15 1 2014. [Online]. Available: <https://visualstudiomagazine.com/Articles/2014/01/01/How-To-Standardize-Data-for-Neural-Networks.aspx?Page=2>. [Accessed 31 5 2016].
- [20] Liu, P. H., Su, S. F., Chen, M. C., Hsiao, C. C., "Deep learning and its application to general image classification," in *Informative and Cybernetics for Computational Social Systems (ICCSS)*, 2015.
- [21] Paul Viola, Michael J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, pp. 137-154, 5 2004.
- [22] 廖文宏, "基於三元化樣式的通用型區域特徵描述方法," 國立政治大學資訊科學系行政院國家科學委員會專題研究計畫期末報告, 臺灣台北, 2011.
- [23] T. Ojala, M. Pietikainen and T. Maenpaa,, "Multi-resolution Gray-scale and

Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24(7), pp. 971-987, July 2002.

[24] "Emgu CV: OpenCV in .NET (C#, VB, C++ and more)," [Online].

Available: [http://www.emgu.com/wiki/index.php/Main\\_Page](http://www.emgu.com/wiki/index.php/Main_Page).

[25] 周志成, "相關係數," [Online]. Available:

<https://ccjou.wordpress.com/2011/03/09/相關係數/>. [Accessed 10 6 2016].

[26] Hong, X., Zhao, G., Zafeiriou, S., Pantic, M., Pietikäinen, M., "Capturing correlations of local features for image representation," *Neurocomputing*, pp. 99-106, 2016.

[27] H. e. a. Li, "A convolutional neural network cascade for face detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

