

國立中興大學 資訊管理學系

碩士學位論文

生成對抗網路應用於多角度學習情緒辨識之研究

The Study on Recognizing Multi-angle Learning

Emotion Based on Generative Adversarial Network



指導教授：林冠成 Kuan-Cheng Lin

研究生：楊詒鈞 Yi-Chun Yang

中華民國一百一十年一月

國立中興大學資訊管理學系

碩士學位論文

題目：生成對抗網路用於多角度學習情緒辨識之研究

姓名：楊詒鈞 學號：7107029097

經 口 試 通 過 特 此 證 明

論文指導教授

林冠成

論文考試委員

林冠成

吳宗慶

洪啟祥

蔡毛山

吳惠珠

中華民國 109 年 06 月 30 日

## 誌謝詞

當初誤打誤撞進了中興研究所，短短兩年好像稍縱即逝，在撰寫本篇論文時，卻覺得時間怎麼過那麼快，做了無數個生成對抗網路，總是失敗，數度想換題目，好在最後終於有了新的發現。

在這個過程中要感謝的人有很多，其中最要感謝的是指導教授林冠成老師，從剛進研究所時，甚麼都不知道，從理解人工學習，到第一次接觸深度學習，都是經過老師的詳細解釋。另外，在第一次口試的時候幫助我很多，即便模型結果不是那麼順利，老師依然替我想了很多解決方案，並不厭其煩的鼓勵我，讓我更有心情去面對。

次之，感謝實驗室的學弟妹們在這半年的時間，犧牲一台 GPU 完全讓我使用，讓我在這學期能無止盡的找方法；另外，感謝實驗室的朋友們彥希在我壓力大的時候聽我碎念，且在模型有困難的時候的幫助，這份恩情我沒齒難忘。

最後我要感謝我的家人們，即便遇到挫折也不曾催促我，讓我在這半年沒有後顧之憂，以及我的妹妹挪出了房間的一半讓我繼續留在台中專心寫論文，即便焦躁、不耐煩也都能體諒我。終於現在要邁向新的人生階段了，希望在未來的職場中，我能保有寫此篇論文堅持不懈的敬神繼續下去！

楊詒鈞 謹致於

國立中興大學 資訊管理學系碩士班

中華民國一百一十年 一月

## 摘要

情感狀態，如困惑、沮喪、無聊、焦慮、好奇、投入和快樂，會在整個學習過程中不斷發生，然而這樣的情感狀態不僅是偶然，它們通過多種方式調節認知，間接影響學習成效。一個有效的學習媒介，無論是人還是電腦，都應培養對學習有益的情感狀態，例如，專注、有趣和好奇等情感狀態，且應盡量減少可能會產生於學習過程中干擾情感狀態的情況發生，例如，無聊和絕望等較劣於學習過程的情感狀態。為了讓講者能夠觀察聆聽者們的情緒，並利用這些信息決定如何調整學習材料的節奏及內容，因此本論文將學習情緒辨識作為研究主題，再對於本研究室目前擁有的資料集之漏洞加以調整。

目前將 Dense\_FaceLiveNet 之網路架構作為辨識學習情緒資料集的學習情緒分類，而該架構之學習情緒分類正確率達到 91.93%，在此架構所使用的學習情緒資料集中，所有受試者之學習情緒圖片皆為正臉。在學習情緒場域中，無論是被遮擋的受試者臉部圖片，例如，掩眉、托腮等，或極端(90 度)側臉的受試者臉部圖片，對學習場域來說皆可能做為有效的關鍵資訊。因此本研究使用域內生成對抗網路反演方法將受試者之非正臉學習情緒圖片生成出其正臉的學習情緒圖片，完成圖像編輯任務，例如，語義處理及圖像插值，再使用 Dense\_FaceLiveNet 之網路架構加以辨識該生成圖片之學習情緒。

透過域內生成對抗網路反演方法生成基本情緒之正臉，包含七種基本情緒，及學習情緒之正臉，包含六種學習情緒。在本次實驗中，發現 Dense\_FaceLiveNet 模型無法對非正臉圖像進行有效的情緒辨識，當正臉的基本情緒辨識準確率為

70%時，使用域內生成對抗網路進行臉部旋轉後，當旋轉角度到達正負 60 度時，辨識準確率為 42.67%、40.57%，因此判斷該模型對於非正臉圖片較不適用。

最後將學習情緒資集內的非正臉圖像用域內生成對抗網路進行旋轉成正臉後，輸入 Dense\_FaceLiveNet 模型做學習情緒判斷，學習情緒的辨識準確率從 42.67% 提升為 57.33%，因此判斷在 Dense\_FaceLiveNet 模型中，輸入的圖象若參雜非正臉時，會導致臉部情緒判斷準確率下降，須將臉部圖像轉至正臉，以達到相對準確的情緒判斷。



關鍵字：學習情緒、生成對抗網路、生成對抗網路反演、圖像編輯、語義處理

## ABSTRACT

Emotional states, such as confusion, depression, boredom, anxiety, curiosity, engagement, and happiness, occur throughout the learning process. And such emotional states are not only accidental, they regulate cognition in a variety of ways and indirectly affect learning outcomes. Whether it's a human or a computer, an effective learning medium should cultivate emotional states beneficial to learning, for example, emotional states such as focus, fun, and curiosity. In addition, the occurrence of situations that may interfere with the emotional state during the learning process should be minimized, such as boredom, despair and other emotional states inferior to the learning process. In order to allow speakers to observe the emotions of listeners and use this information to determine how to adjust the rhythm and content of learning materials, this paper takes learning emotional recognition as the research topic, and then adjusts the gaps in the dataset currently available in our laboratory.

At present, the Dense\_FaceLiveNet network is used as the classification of learning emotions in the learning emotion dataset, and the accuracy of learning emotion classification reached 91.93%. In the learning emotion dataset used in this architecture, all the subjects' learning emotion pictures are frontal. In the learning emotional field, both occluder images of subjects' faces, such as eyebrow masking, gill-holding, etc., and extreme (90 degree) side-facing images of subjects' faces may serve as key information for the learning field. In this study, I use In-Domain GAN network to generate the learning emotion of frontal images from the learning emotion of profile face images of the subjects to finish like Image Editing and Semantic Processing, and then use the Dense\_FaceLiveNet network to identify the learning emotion of the generated images.

In-domain GAN is used to generate the positive face of basic emotions, which

contains seven kinds of basic emotions, and the positive face of learning emotions, which contains six kinds of learning emotions. In this experiment, found Dense\_FaceLiveNet model cannot be emotional identification is a face image effectively, when the basic emotions, positive face recognition accuracy rate was 70%, using field generated against Internet face after rotation, when the rotation Angle arrived at plus or minus 60 degrees, the recognition accuracy is 42.67%, 40.57%, so the judgment for the model is a face image is not applicable.

Finally, the non-positive face image in the learning emotion information was rotated into a positive face by using the domain generated antagonism network, and then the Dense\_FaceLiveNet model was input to judge the learning emotion, and the recognition accuracy of learning emotion increased from 42.67% to 57.33%. Therefore, it is judged that in the Dense\_FaceLiveNet model, if the input image is mixed with non-positive faces, the accuracy of facial emotion judgment will decrease. Therefore, the face image must be transferred to the positive face to achieve relatively accurate emotion judgment.

Keywords: Learning Emotion, Generative Adversarial Network, In-Domain GAN, Image Editing, Semantic Processing

## 目 次

摘要.....	i
ABSTRACT.....	iii
目次.....	v
表目次.....	vi
圖目次.....	vii
<b>第一章 緒論.....</b>	<b>1</b>
1.1 研究背景.....	1
1.2 研究動機與目的.....	3
<b>第二章 文獻探討 .....</b>	<b>5</b>
2.1 情緒.....	5
2.1.1 基本情緒與學習情緒.....	5
2.1.2 情意計算及其應用於情緒識別.....	6
2.2 生成對抗網路(GANs).....	6
2.3 StyleGAN 模型及利用分層噪聲進行反演.....	7
2.4 域內生成對抗網路反演方法(in-domain GAN).....	12
2.5 Dense_FaceLiveNet 網路.....	16
2.6 遷移學習.....	21
2.7 資料集.....	23
<b>第三章 研究方法 .....</b>	<b>25</b>
3.1 實驗流程 .....	25
3.2 資料預處理 .....	26
3.2.1 模型超參數設定 .....	26
<b>第四章 研究結果 .....</b>	<b>28</b>
4.1 實驗環境.....	28
4.2 基本情緒旋轉.....	29
4.3 學習情緒旋轉結果.....	41
<b>第五章 結論與建議 .....</b>	<b>45</b>
5.1 研究結果與討論.....	45
5.2 未來研究方向.....	46
<b>參考文獻 .....</b>	<b>47</b>



## 表 目 次

表一、學習情緒操作型定義 .....	24
表二、In-domainGAN 超參數設定 .....	26
表三、Dense_FaceLiveNet 超參數設定 .....	26
表四、開發環境整體規格 .....	28
表五、各基本情緒總張數、彩色及黑白分別張數 .....	29
表六、基本情緒-害怕旋轉至各角度 .....	29
表七、基本情緒-生氣旋轉至各角度 .....	30
表八、基本情緒-厭惡旋轉至各角度 .....	31
表九、基本情緒-開心旋轉至各角度 .....	32
表十、基本情緒-中性表情旋轉至各角度 .....	33
表十一、基本情緒-傷心旋轉至各角度 .....	34
表十二、基本情緒-驚訝旋轉至各角度 .....	35
表十三、學習情緒資料庫各學習情緒張數 .....	41



## 圖目次

圖 2.1、傳統生成對抗網路生成器與基於樣式的生成器 .....	9
圖 2.2、映射網路圖示 .....	10
圖 2.3、樣式模塊(AdaIN)圖示 .....	11
圖 2.4、隨機變化圖示 .....	12
圖 2.5、(a)傳統編碼器與域導向編碼器比較(b)傳統優化與域正則化優化比較 .....	13
圖 2.6、卷積計算圖示 .....	18
圖 2.7、最大池化圖示 .....	18
圖 2.8、Dense Inception Block 架構示意圖 .....	20
圖 2.9、Dense_FaceLiveNet 架構示意圖 .....	21
圖 2.10、遷移學習示意圖 .....	22
圖 3.1、本論文實驗流程 .....	25
圖 4.1、原先正臉之基本情緒辨識混淆矩陣 .....	36
圖 4.2、基本情緒圖片旋轉正負 5 度混淆矩陣 .....	37
圖 4.3、基本情緒圖片旋轉正負 10 度混淆矩陣 .....	37
圖 4.4、基本情緒圖片旋轉正負 15 度混淆矩陣 .....	38
圖 4.5、基本情緒圖片旋轉正負 30 度混淆矩陣 .....	39
圖 4.6、基本情緒圖片旋轉正負 45 度混淆矩陣 .....	39
圖 4.7、基本情緒圖片旋轉正負 60 度混淆矩陣 .....	40
圖 4.8、基本情緒圖片多角度辨識準確率 .....	41
圖 4.9、彩色學習情緒圖片旋轉前後 .....	42
圖 4.10、黑白學習情緒圖像旋轉前後 .....	42
圖 4.11、旋轉前學習情緒混淆矩陣 .....	43
圖 4.12、旋轉過後的學習情緒混淆矩陣 .....	44

# 第一章 緒論

本篇論文主要將非正臉的臉部圖片生成出正臉，並加以辨識其臉部其學習情緒，以下將於研究背景、目的、動機中詳細介紹。

## 1.1 研究背景

情緒，是由感覺、思想以及行為所自然產生的生理及心理反應，也是經過外界的刺激所引發出綜合的心理與生理狀態[1]，因這些反應所導致的情緒進而體現於不同的外在行為，且可由此外在行為提前推測人們的情緒，預測出人們後續的行為模式。

而由情緒所造成的行為模式之意義可用情意計算來表達。情意計算是美國學者 Picard[2]於 1995 年提出，簡單來說，即為利用科技去辨識情感，而情感的表達方式有很多種，例如，體溫[3]、腦波[4]、臉部表情等。透過各種機器取得任何情感資訊，並利用這些資訊進行識別，理解人們的情感後做出適當的回應。

情感表達方式中的臉部表情出現在眾多有關情緒的文獻中[6][7]，透過臉部表情去分析的情緒有兩種，分別是「基本情緒」及「複雜情緒」。第一個提出基本情緒的是 1962 年心理學家 Silvan Tomkins[5]，他所提出的基本情緒包含驚訝、有趣、愉悅、憤怒、害怕、嫌惡、羞愧、痛苦等八種情緒，之後 Paul Ekman[6]基於 Tomkins 的研究成果，在 1994 年也提出了喜悅、生氣、悲傷、恐懼、難過與驚訝等六種基本情緒。基本情緒代表本能的反應，與原始人類之生存息息相關，不需後天額外的任何學習。

而需後天額外學習，並透過後天社會化過後的情緒稱為複雜情緒[7]，其相對於基本情緒來說更為複雜。人們可能在社會化過後，處於相同情況中會產生不同的狀態，不同的狀態進而形成不同的情緒。在各種不同的情境中，與學習相關的情緒更是受到重視，稱為學習情緒，而學習情緒[8]也是複雜情緒的其中一種。學習情緒為探討學生在學習的過程中所反應出來的學習狀況，讓一位老師對多位學生的講述式教學能夠掌握學生們的學習效果，而提供老師有幫助的教學回饋，成為課堂教學進度與內容之改善依據。Sidney D'Mello[7]在 2006 年提出了學習情緒 (Learning Emotion) 的概念，在學習情境中，基本情緒無法準確反應學習者的學習狀態，而學習情緒較能表現學習者在學習過程中所透露的情緒，並將情緒分為挫折(Frustration)、困惑(Confusion)、無聊(Boredom)、投入(Flow)、喜悅(Delight)及驚訝(Surprise)六種情緒。

由於機器學習與深度學習不斷進步而改良臉部辨識方法，傳統機器學習需用特徵點學習，如此不能取得未被標定的特徵資料，因此透過深度學習的圖像處理器、卷積神經網路來解決此問題，不用進行特徵標定，即能學習到原始資料的特徵[9]。

而近年學習情緒辨識應用中，賴念祥改良 Zuheng 等人[10]提出之 FaceLiveNet 卷積神經網路架構，提出 Dense\_FaceLiveNet 之架構，以卷積神經網路( Convolution Neural Network, CNN ) 加上遷移學習(Transfer Learning)辨識學習情緒[10]。首先，使用資料較為簡單的 JAFFE 與 KDEF 資料集建立的基本情緒辨識模型，再將模型遷移學習至 FER2013 基本情緒資料集，最後遷移學習至學習情緒資料集，形成完全正臉的學習情緒辨識模型，該模型平均辨識準確率到達 91.93%。

然而在賴念祥論文[10]中，發現兩處可以改善的地方。(1)Shan Li 及 Weihong Deng 提出[12]若缺乏足夠的訓練數據或含有和表達無關的變化會導致過度擬合的問題，所以必須讓每筆資料都有它的價值，而不是刪除缺失資料，例如，非正臉圖像，且目前實驗室的學習情緒資料集的臉部圖像含有部分圖像為托腮、掩眉，導致投入、無聊的錯分比例相比於其他學習情緒較多[10]。(2)在實際的學習場域中，攝像頭安裝於學生們電腦上，導致臉部圖像存在小部分遮擋，或是大範圍遮擋，甚至是極端側臉之圖像(90 度)都是可能發生的，而以上情況對於學習場域來說都是不可忽略的資訊。在賴念祥的實驗中，未對非正臉到極端側臉的多角度臉部圖像進行學習情緒識別，只採用學習者的正臉進行學習情緒識別。

因此本研究使用生成對抗網路來改善上述問題，在進行監督式的深度學習模型訓練時，往往需要大量人工標籤好的資料才能用於訓練神經網絡辨識模型。隨著時代資訊科技化，人工智慧的崛起、神經網絡的深化，另外還有「對抗」想法的產生，Ian J. Goodfellow 等人於 2014 年提出生成對抗網路(Generative Adversarial Networks, GANs)[11]之設計方法，其能使神經網絡之訓練大幅減少人力介入。GAN 設計理念為訓練兩個相互競爭的神經網絡，分別為生成器(Generator)及判別器(Discriminator)，生成器學習產生擬真資料，以欺騙判別器；而判別器學習增強其真實資料之辨識能力對抗生成器之欺騙，如此生成器得以產生出許多擬真資料，擬真資料可彌補真實資料集資料樣本數不足，減少收集大量資料的負擔，且判別器也能同時完成對等訓練，且亦能利用生成器增廣深度學習訓練。

## 1.2 研究動機與目的

由於在賴念祥的 Dense\_FaceLiveNet 之架構中[10]，對於學習情緒辨識所使用

的學習情緒資料集圖像未含有對受試者正臉以外的表情做學習情緒判斷，而本論文使用域內生成對抗網路進行受試者非正臉生成至正臉之圖像處理。

Jiapeng Zhu 等人[13]提出域內生成對抗網路反演方法(in-domain Generative Adversarial Network, in-domain GAN)，此方法從 Karra 等人[14] 利用分層噪聲對 StyleGAN 模型進行反演延伸而來。域內生成對抗網路反演方法學習域引導編碼器，將輸入圖像投影至 GAN 潛在空間，再把編碼器作為正則化器將域正則化優化，達到完整重建輸入圖像，並確保其語義。將非正臉圖像產生出其正臉圖像後，再將其圖像作為 Dense\_FaceLiveNet 網路[10]之輸入進行測試，對該位學生的學習情緒做判斷。

本論文引用域內生成對抗網路反演方法[13]生成學習場域中之非正臉圖像至正臉，並對生成正臉之表情做情緒判斷，而本論文目的為建立基於臉部表情偵測之生成正臉學習情緒預測模型。

## 第二章 文獻探討

本章將於 2.1 節介紹何謂情緒、基本情緒、複雜情緒裡的學習情緒以及情意計算，2.2 節介紹基本生成對抗網絡(GANs)，2.3 節介紹 StyleGAN 模型及利用分層噪聲對 StyleGAN 模型進行反演，2.4 節介紹域內生成對抗網路反演方法(in-domain GAN)，2.5 節介紹 Dense\_FaceLiveNet 網路，2.5 節介紹遷移學習，2.6 節為本研究所使用的資料集介紹。

### 2.1 情緒

情緒是被定義為主觀認知的統稱，藉著許多不同的感覺、想法和行為統整後形成的心理加生理反應，在日常生活中最頻繁發生的情緒為喜怒哀樂等，當然也存在更細部的情緒分類，本章節將探討基本情緒及複雜情緒中的學習情緒，2.1.1 節說明何謂基本情緒與複雜情緒，2.1.2 節說明何謂情意計算以及在識別情緒上的應用。

#### 2.1.1 基本情緒與學習情緒

基本情緒與複雜情緒之差別在於，基本情緒並未經過社會化，而複雜情緒為已社會化後每個人會有不同的情緒表現。此節分別介紹基本情緒及複雜情緒中的學習情緒。

##### A. 基本情緒

基本情緒不需經後天之社會化，並自然產生，Paul Ekman[6]基於 Tomkins[5]的研究成果提出所有人在面對相同情況時，會產生相同的情緒、生理狀態，六種基本情緒為憤怒 (Angry)、快樂 (Happiness)、恐懼 (Fear)、厭惡 (Disgust)、悲

傷 (Sadness)、驚訝 (Surprise)，而這六種基本情緒也被廣泛使用，定義為實驗中基本情緒的辨識模型基礎。

## B. 複雜情緒(學習情緒)

複雜情緒[7]是經過社會化而產生的情緒，在經過人們相處後學習到的情緒，然而每個人在社會化之後會對同一件事情有不同的情緒，稱為複雜情緒。複雜情緒通常因為道德標準或價值觀的差異不同而有所不同，複雜情緒包含很多種，像是在學習情景下的複雜情緒則稱為學習情緒。

### 2.1.2 情意計算及其應用於情緒識別

情意計算是在希望機器具有情感能力的情況下產生，最初由麻省理工學院多媒體實驗室(Media Lab)的 Picard 教授[2]所提出，情意計算所代表的是與情緒有相關的，由情緒所引發的，能調整情緒因素的運算。情意計算去蒐集、分析人們情緒的表現，像是臉部表情[15]、聲音起伏[16]、腦波[4]等，利用這些表現來探討人們情緒的變化過程。

## 2.2 生成對抗網路(GANs)

生成式對抗網路 GAN(Generative Adversarial Networks)是一種非監督式學習的方法，其基本架構由兩個網絡組成，一個為生成器(Generator)，另一個為判別器(Discriminator)，最初是 2014 年由 Ian Goodfellow 等人所提出[11]。生成器將從潛在空間(latent space，又稱隱空間，生成對抗網路利用噪聲  $z$  產生圖片，而噪聲  $z$  所在的空間即為潛在空間)隨機抽樣當作輸入，生成出與訓練集圖像相似的圖片。



判別器將真實圖像或生成圖像作為輸入，判斷出真實圖像與生成圖像。而兩個網絡需要不斷相互對抗，直到判別器無法判斷出是否為生成圖片。雖生成對抗網絡是為了非監督式學習而提出，但它也可以用於半監督式學習[17]、完全監督式學習[18]及強化學習[19]。

## 2.3 StyleGAN 模型及利用分層噪聲進行反演

2017 年，NVIDIA 提出[20]ProGAN 解決生成高分辨率圖像之問題，例如  $1024 \times 1024$ ，而其關鍵創新為漸進式訓練，開始於訓練低分辨率( $4 \times 4$ )圖像之生成器與判別器，每次皆增加更高的分辨率層。ProGAN 與多數 GAN 一樣存在控制特定圖像的特定特徵之能力有限，即便微調輸入，也因特徵相互糾纏而導致生成圖像的多個特徵受到影響。

2018 年，StyleGAN[14]為 NVIDIA 在 ProGAN 後提出新的網路架構，該模型達到在不影響模型中其他層級的情況下，主要透過各別修改每層之輸入來達到控制該層級所代表的特徵，而這些特徵可以是範圍的特徵，例如姿勢、臉型等，亦可是小範圍的特徵，例如瞳孔、髮色等。

StyleGAN 提出[14]以下四點貢獻：

### (1) 提出基於樣式的生成器

- a. 達成無監督分離高級特徵，例如臉部姿勢、身分等，以及隨機變化，例如雀斑、頭髮等。
- b. 達成對生成圖像特定尺度的特徵控制。
- c. 讓生成器由可學習的常量輸入開始，在每個卷積層用噪聲調整圖像樣式，

(4) 新高質量臉部資料集(FFHQ，七萬張 1024\*1024 的臉部圖片)。

Latent  $z \in \mathcal{Z}$

Normalize

Latent  $z \in \mathcal{Z}$

Normalize

① → Const 4\*4\*512

Synthesis is network  $g$

Noise

Diagram of the second stage of the U-Net architecture. It shows a skip connection from the encoder's output to the decoder's input, followed by a Pixel Norm layer, two FC (Fully Connected) layers, and an Upsample layer. A red circle with the number 4 is next to the Upsample layer.



## 圖2.1、傳統生成對抗網路生成器與基於樣式的生成器

圖 2.1、[14]為傳統生成器(a)與基於樣式的生成器(b)結構圖，基於樣式的生成

器共有 18 層，每個分辨率包含兩個卷積層，而基於樣式的生成器將較於傳統生成器有下列改良：

### (1)移除傳統輸入

傳統生成器的初始輸入為隨機輸入(Latent Code)，StyleGAN 則將可學習的常數作為生成器的初始輸入。假使如此減少特徵糾纏，對於模型來說，不依賴糾纏的數向量相較容易學習。

### (2)映射網路(Mapping Network)

映射網路如圖 2.2、[14]所示，由八個全連接層構成，其輸入與輸出大小相同，都是(512x1)，目標是將輸入向量編碼成中間向量，利用中間向量的不同元素來控制相異特徵，而利用輸入向量  $z$  來控制特徵是有限的，因為它需依訓練資料的機率密度，像是圖像中為黑髮的人在資料集中更常出現，那麼輸入值會映射更多到黑髮的特性，導致模型無法將輸入向量  $z$  的一部份映射到特徵中，而形成特徵糾纏。而利用映射網路可生成不須依訓練資料的分布向量  $w$ ，且減少特徵之間的相關性，達到解耦、特徵分離的目的。

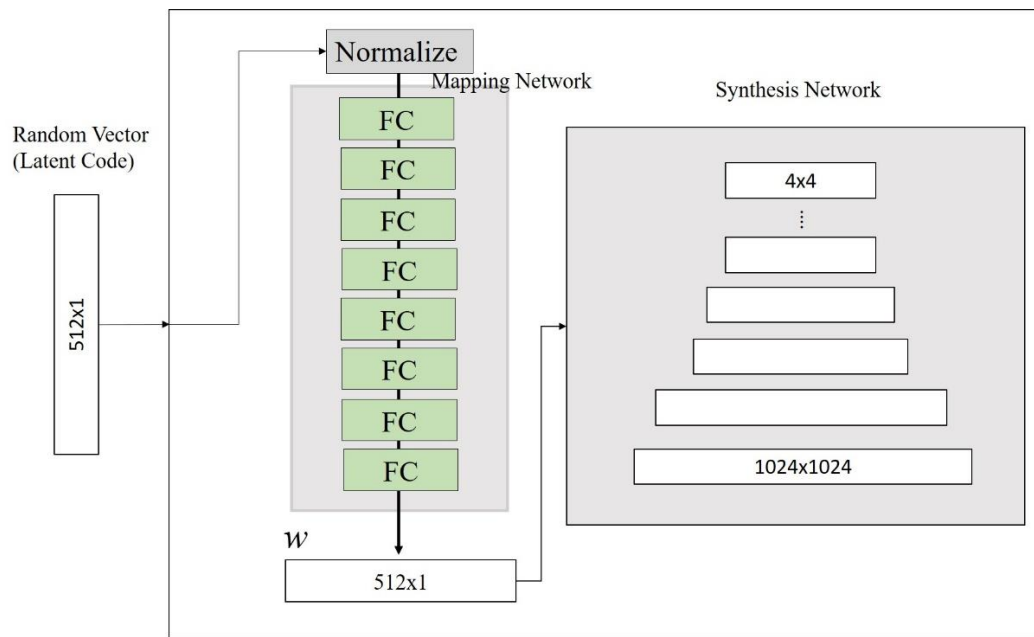


圖2.2、映射網路圖示

### (3)樣式模塊(Style Modules, AdaIN，自適應實例歸一化)

潛在空間通過卷積層中的 AdaIN 輸入到生成器中的每一層，圖 2.3、[14]中 A 代表可學習的仿射變換，其計算方法先將每個特徵圖  $X_i$  (feature map)各自進行歸一化  $(x_i - \mu(x_i)) / \sigma(x_i)$ ，特徵圖中的值每個皆減去該特徵圖的均質再除以方差。可學習的仿射變換 A 把潛在空間化成 style 裡 AdaIN 的縮放及平移因子  $y = (y_s, i, y_b, i)$ ，再將每個特徵圖分別使用 style 中學到的縮放及平移因子達成平移及尺度變換。

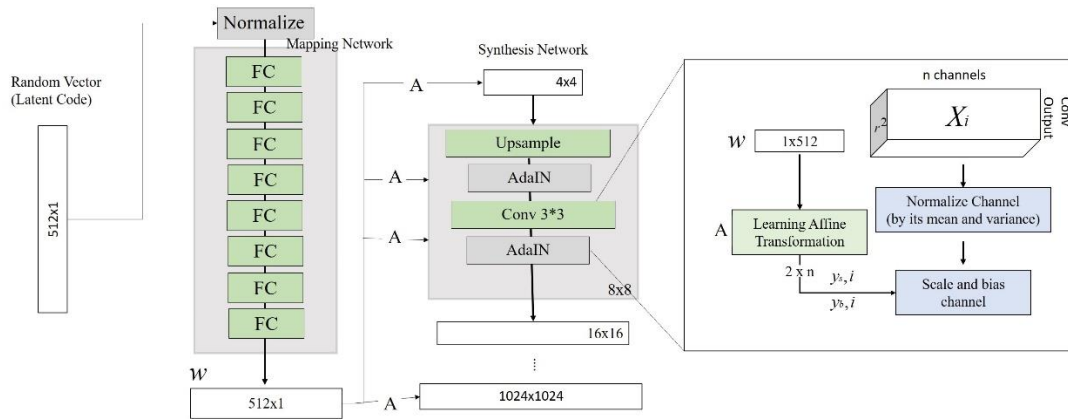


圖2.3、樣式模塊(AdaIN)圖示

#### (4)隨機變化(Stochastic Variation，利用加入噪聲來為生成器生成隨機細節)

臉部很多特徵可以視為隨機，例如頭髮精準的位置，而讓圖像更為逼真，而提高輸出多樣性，將該類隨機特徵的常用方法為將輸入向量增加隨機噪聲，再通過輸入層進到生成器，但如此可能會造成特徵糾纏的狀況發生，使得其他特徵也受到影響，因此圖 2.4、[14]架構是利用在合成網路的每個分辨率上加上尺度化的噪聲去避免特徵糾纏的問題，該噪聲是由高斯噪聲構成單一通道圖片，使噪聲圖像供給合成網路的特徵圖。而在卷積後、AdaIN 前，把高斯噪聲加進生成器網路之中，在圖 2.4、中，B 利用可學習的縮放參數將輸入進來的高斯噪聲進行變換，再把噪聲圖像各自加到每一特徵圖上，而每一特徵圖對應一可學習的尺度參數。



GAN Inversion)。在對生成對抗網路進行反演時，除了依照像素值來恢復輸入圖像之外，該模型亦關注反演代碼是否有語意意義，語意意義為生成對抗網路從觀察到的資料中學習到的新知識，因此先訓練域導向的編碼器，再利用該編碼器當作正則化器進行更進一步的域正則化優化，如圖 2.5、所示。

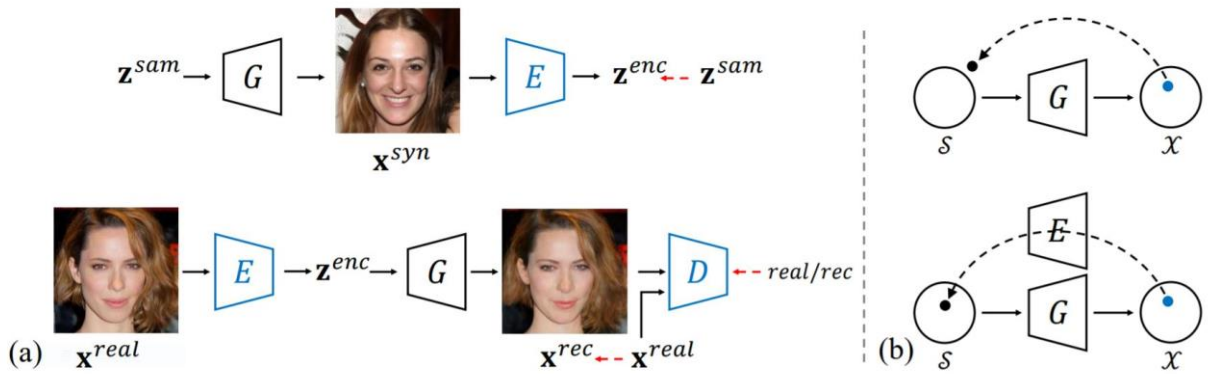


圖2.5、(a)傳統編碼器與域導向編碼器比較(b)傳統優化與域正則化優化比較

生成對抗網路模型通常由生成器  $G(\bullet): Z \rightarrow X$  及判別器組成  $D(\bullet)$ ，前者用於合成高品質的圖像，後者用於區分真實圖像及合成圖像，而生成對抗網路的反演達到  $G(\bullet)$  的反向映射，該映射找到最佳的潛在代碼  $z^{inv}$  來恢復給定的真實圖像  $x^{real}$ 。生成對抗網路大部分從預定義的分布式空間  $Z$ ，例如正態分布，中採樣潛藏代碼  $z$ ，StyleGAN 提出[14]利用多層感知將潛在空間  $Z$  映射到第二個潛在空間  $W$ ，再將  $w \in W$  輸入到生成器中進行圖像合成，如此達成附加映射，且已證明此方法可以學到更多的解糾纏語意[50][14]。Jiapeng Zhu 等人選擇  $W$  空間作為反演空間有下列原因，第一個原因為  $W$  空間更適合分析，因為需要注意的是反演代碼的語意屬性(域內)；第二個原因是反演到  $W$  空間比反演到  $Z$  空間的性能更好[51]；最後，將  $W$  空

間引入到其他任何的生成對抗模型是容易的，只需再生成器前學習一個額外的多層感知器(MLP)，因此 $W$ 空間不會對泛化能力有所影響。在圖 2.5、的 a 中為傳統編碼器與域編碼器之比較，藍色為可訓練的模型，而紅色虛線箭頭表示監督，域導向的編碼器不是使用合成資料來恢復潛在代碼，而是利用目標來訓練恢復，另外，固定生成器(非藍色框)是為了確保編碼器生成的代碼剛好在生成器本身的潛在空間裡，並保持語意意義。圖 2.5、的 b 為傳統優化與域正則化優化的比較，而 $S$ 表示生成對抗網路學習到的語意空間，在優化過程中引入訓練好的域導向編碼器當作正則化器，將潛在代碼引入語意。

#### 域導向編碼器(Domain-Guided Encoder)

訓練編碼器常用在生成對抗網路反演問題，主要因其推演速度快，但現有的方法只學習一個特定的模型[21][22]，不考慮編碼器生成的代碼是否與 $G(\bullet)$ 學習的語意知識相同，如圖 2.5、(a)上方所示，把一組潛在代碼 $z^{sam}$ 隨機採樣，送入 $G(\bullet)$ 得到相應的合成 $x^{syn}$ ，再將編碼器 $E(\bullet)$ 分別以 $x^{syn}$ 和 $z^{sam}$ 當作輸入及監督，以公式一進行訓練，

$$\min_{\theta_E} L_E = \| z^{sam} - E(G(z^{sam})) \|_2 \quad (\text{公式一})$$

而式中 $\|\bullet\|_2$ 表示 $l_2$ 距離， $\theta_E$ 表示編碼器 $E(\bullet)$ 的參數，而 Jiapeng Zhu 等人認為若僅通過 $z^{sam}$ 進行監督將不足以訓練出精確的編碼器，且無考慮 $G(\bullet)$ 的梯度，導致生成器被省略，進而影響提供其領域知識達成編碼器訓練之成效。

為了解除上述問題，因此 Jiapeng Zhu 等人提出域導向編碼器，如圖 2.5、(a)下圖所示，與傳統編碼器有以下三個主要不同的差異：



- (1) 以編碼器之輸出輸入進生成器以重建圖像，如此目標函數來自圖像空間而非潛在空間，而這會影響到生成器在訓練時的語義知識，並提供更精確的監督，以達到輸出代碼與生成器的語義一致。
- (2) 用真實圖像訓練域導向編碼器而非合成圖像，使得編碼器更適用在真實環境。
- (3) 使用判別器與編碼器競爭，以保證重構圖像的真實性。

以上三種差異可以讓生成對抗網路獲取更多訊息，對抗性的訓練方式使輸出代碼更好的學習生成器的語知識，另外，Jiapeng Zhu 等人使用 VGG[53]提取特徵，且引入感知損失[52]，如公式二及公式三所示，式中  $P_{data}$  表示真實的資料分布， $\gamma$  為梯度正則化之超參數， $\lambda_{vgg}$  以及  $\lambda_{adv}$  為感知器及判別器的損失權值， $F(\bullet)$  為 VGG 特徵提取模型。

$$\min_{\theta_E} L_E = \|x^{real} - G(E(x^{real}))\|_2 + \lambda_{vgg} \|F(G(E(x^{real})))\|_2 \quad (\text{公式二})$$

$$- \lambda_{adv} \sum_{x^{real} \sim P_{data}} [D(G(E(x^{real})))]$$

$$\begin{aligned} \min_{\theta_E} L_D = & \sum_{x^{real} \sim P_{data}} [D(G(E(x^{real})))] - \sum_{x^{real} \sim P_{data}} [D(x^{real})] \quad (\text{公式三}) \\ & + \frac{\gamma}{2} \sum_{x^{real} \sim P_{data}} \left[ \|\nabla_x D(x^{real})\|_2^2 \right] \end{aligned}$$

#### 域正則化優化(Domain-Regularized Optimizatoin)

生成對抗網路的生成過程即在分布層學習映射，例如從潛在分布到真實圖像

分布，而生成對抗網路反演就是實例任務，最完整的重建給定的圖像，但因為編碼器有限的表示能力，很難單獨去學習完整反向映射，因此有域導向編碼器可重建基於預訓練生成器的輸入圖像，確保代碼自己的語義上是有意義的之外，還需要其他代碼，讓反演更好的達到目標圖像的像素值。

先前使用梯度下降優化代碼[26][24]，圖 2.5、(b)頂部圖片說明，若只基於生成器自由的優化潛在代碼的優化過程可能會導致產生出域外反演的情況，因為這樣的做法對潛在代碼沒有任何約束，因此依域導向編碼器，Jiapeng Zhu 等人提出[13]域正則化優化，如圖 2.5、(b)下圖所示，其中包含兩點改善：

(1) 使用域導向編碼器的輸出作為起點，免去受到局部最小值代碼的侷限，並縮短優化過程。

(2) 將域導向編碼器當作正則化器，達到保留生成語義域內的潛在代碼。

上述之優化的目標函數如公式四，式中  $x$  為要反演的目標圖像，而  $\lambda_{vgg}$  及  $\lambda_{dom}$  分別是感知損失與編碼器正則化所對應的損失權值。

$$z^{inv} = \arg \min_z \|x - G(z)\|_2 + \lambda_{vgg} \|F(x) - F(G(z))\|_2 + \lambda_{dom} \|z - E(G(z))\|_2 \quad (\text{公式四})$$

## 2.5 Dense\_FaceLiveNet 網路

此模型由賴念祥提出[10]，而改良自 Zuheng[51]等人提出的 FaceLiveNet，圖 2.9、[10]為 Dense\_FaceLiveNet 架構圖，該架構相較 FaceLiveNet 改良下列三項：

### (1) 使用 GlobalAveragePooling 取代全連接層

全連接層將前面的卷積層與池化層彼此相互交疊之後學習到的特徵變成標記輸出，把卷積層所輸出的特徵圖轉成向量，並用將此向量做相乘降低維度，再使用 softmax 輸出到相對應的分類。參數量過多就是全連接層最大的問題，進而形成過度擬合，而卷積層與池化層為 CNN 裡特有的，用於保有原資料多維度的相對應特徵。

### 卷積層

在影像處理時，卷積是常見的操作，用來選取影像中的重要資訊，在傳統影像處理中常用到 Sobel、中值濾波等進行卷積，該運算即把原始輸入影像與卷積核做相乘，卷積核又稱為特徵擷取器或濾波器。CNN 中不須手動設計特徵擷取器的參數，在 CNN 訓練的過程中，每個特徵擷取器會學習不同的影像特徵，例如一張人臉圖像，一些特徵擷取器學習臉部輪廓，另一些學習鼻子等，CNN 會自動學習影像中需要學習的地方，卷積的計算過程如圖 2.6、[10]，原資料為 6\*6 的二為矩陣，而 3\*3 的二為矩陣為卷積核，把滑動窗口設為 1 後將原影像和卷積核做矩陣相乘，形成 4\*4 二維矩陣，經過此矩陣相乘得到的二維矩陣稱為特徵圖。

0	0	0	0	0	0	0
0	1	0	0	0	1	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	1	0	0	0	1	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0

原始影像

 $\otimes$ 

0	0	1
1	0	0
0	1	1

卷積核

 $=$ 

0	1	0	0	0
0	1	1	1	0
1	0	1	2	1
1	4	2	1	0
0	0	1	2	1

卷積後影像

圖2.6、 卷積計算圖示

### 池化層

池化在 CNN 中主要是將原始輸入的圖像維度降低，把區塊內像素壓成代表性的像素點，壓縮方法分為平均池化和最大持化，平均池化將固定範圍中的全部像素值做平均後保留，而最大池化即維固定範圍內將最大的像素值保留起來，兩種池化也分別有模糊及銳化的效果。池化層常放在卷積層後面，把從卷積後的圖像維度降低，讓過濾出來的特徵既不失去原特性，且減少參數量，如此更能讓模型降低計算量，並減少過度擬合的情況產生。圖 2.7、[10]為最大池化圖示，6\*6 的二維矩陣做為原始輸入的影像大小，3\*3 為池化層大小，透過滑動窗口設為 3 後的運算，形成 2\*2 的圖像，達成影像有效縮小之外，並保有原影像中的關鍵特徵。



圖2.7、 最大池化圖示

[38]提出全域平均池化(Global Average Pooling, GAP)，利用取代全連接層

達到降低網路參數量的目的。它主要透過輸入的特徵圖大小設為平均池化的窗口大小，而輸出則是與分類類別個數一樣的特徵圖通道支向量。Dense\_FaceLiveNet 將 Dense\_LiveNet 中的全連接層換成全域平均池化，達到減少網路參數量與避免過度擬合的目的。

## (2) 使用 Dense Block 取代 Residual Block

使用不同大小的卷積核去學習不同大小的特徵為 Inception[39]的主要特點。Dense Block[40]認為神經網路不是只能成為遞進階層網路，在  $i$  層學習到的特徵，不一定是仰賴  $i-1$  層傳遞來的特徵。Dense Block 將網路裡每層都做串聯，達到特徵重複利用，其公式如公式五所示[40]， $X_I$  當作輸出，而第  $I$  層的函數為  $H_I$ ，Dense\_FaceLiveNet 中的每一層的 Inception 當作  $X$ ，利用串聯而成  $H_I$ ，因為 Dense Block 達成將每一層 Inception 學到的特徵傳給在後面全部的 Inception 使用。

$$X_I = H_I([x_0, x_1, \dots, x_{I-1}]) \quad (\text{公式五})$$

Dense\_FaceLiveNet 將 Inception 彼此串聯形成該架構，稱做 Dense Inception Block，如圖 2.8、中在最後的  $X_I$  加入 Translate Layer，以此達成把所學習到的特徵做降維，而在 Dense\_FaceLiveNet 中共用兩層 Dense Block，以及一層 Translate Layer。

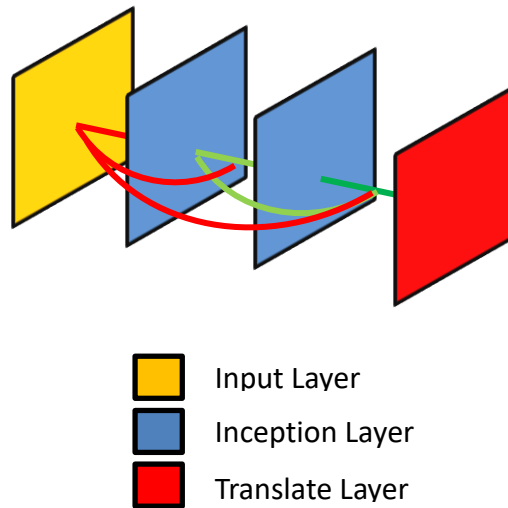


圖2.8、 Dense Inception Block 架構示意圖

(3) 把 ReLU 替換成 Swish 當作激活函數

以上第一及第二點皆是為了使網路架構最佳化，Dense\_FaceLiveNet 模型中，透過替換激活函數提升模型質量。2017 年，Google Brain[37]提出 Swish 激活函數，該激活函數之特性與 ReLU 相同，皆為平滑且非單調的激活函數，因此使用 Swish 直接替換 ReLU，且在 ImageNet 的測試數據中，Swish 激活函數相較 ReLU 激活函數高出 0.9%。



而已有知識又稱原有領域，未知知識則稱目標領域，遷移學習主要任務讓原有領域遷移到目標領域，而在機器學習中，遷移學習則是利用已建立的模型應用在位置但有關連性的領域如圖 2.10、所示。

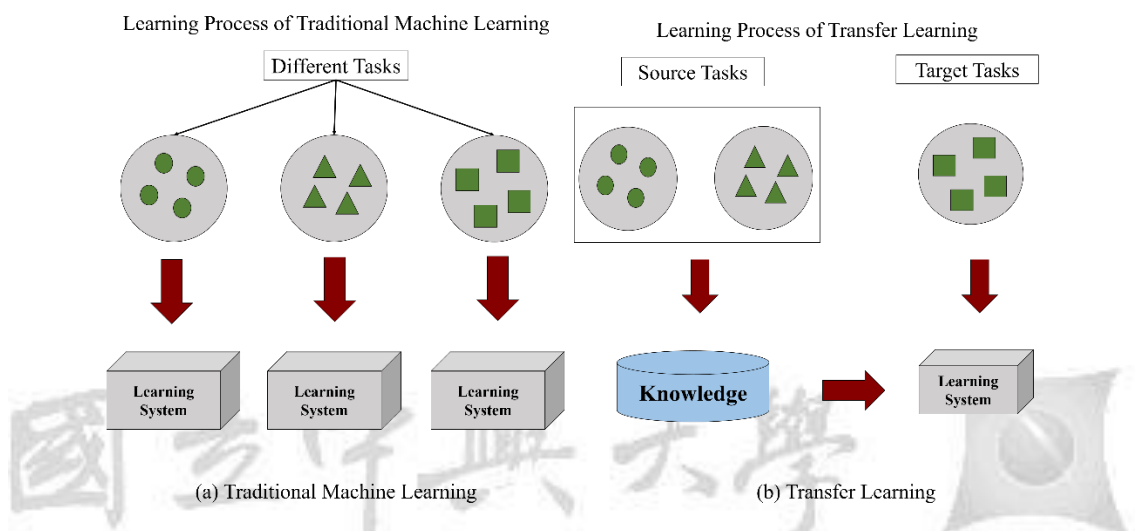


圖2.10、 遷移學習示意圖

遷移學習可依相異的學習方式分成四個類型，基於特徵、基於模型、基於關係，以及基於樣本，基於特徵的遷移學習是把原有領域及目標領域對應到相同地方；基於模型是把模型跟樣本結合後去調整模型參數；基於關係則是將把原有領域的學習概念對應到目標領域；最後，基於樣本的遷移學習則是對原有領域有標定樣本做加權使用，而賴念祥在實驗中使用的是基於模型的學習[10]。傳統在實作遷移模型時，先移除整個模型的最後一層，再把它層的參數固定，設成不可訓練。然後，加入一輸出層，把目標輸出數量設為結點，再將模型訓練到收斂。完成後，開放其他層的訓練，且流程一樣，重新訓練整個網路直到收斂。以上作法是為了達到在模型執行遷移學習時，初始參數值差異過大，新加入的輸出層示初



始化之參數，而其他層則為已訓練的參數。兩中參數的差異位導致整個模型有較大的梯度更新，進而造成原本訓練好的參數在訓練初期沒有效用。而賴念祥提出的 Dense\_FaceLiveNet 的最後一層沒有使用全連接層，所以在做遷移學習時，只需把最後的輸出層節點替換成新資料及類別數量就能直接訓練。

## 2.7 資料集

過往研究中使用公開基本情緒資料集例如 CK+[30]、KDEF[31]、JAFPE[32] 與 FER2013[33]等，來訓練基本情緒辨識模型，而在 In-domainGAN 實驗中所使用的資料集為 StyleGAN 自帶的臉部資料集 FFHQ[14]，該資料集包含 52,000 張 512\*512 分辨率之高品質圖像，且在種族、年齡及圖像的背景方面皆存在很大差異，此資料集也含有眼鏡，太陽鏡，帽子等附件。資料集的圖像是從 Flickr 抓取的，因此繼承了該網站的所有偏見，並使用 dlib 自動對齊和裁剪。

而在 Dense\_FaceLiveNet 的模型訓練中，使用的是 JAFPE 資料集以及 KDEF 資料集做為訓練資料，再遷移學習到 FER2013 資料集進行訓練。JAFPE 資料集是在 1998 年建立，此資料集屬於比較小的資料集，其中包含 10 位日本女性在控制的實驗環境中依照指令做出的表情，再用相機獲取該 10 位日本女性的臉部表情圖像，此資料庫共包含 213 張影像，七種表情，分別為傷心、快樂、生氣、厭惡、驚訝、害怕，以及中性表情，每組含有 20 張圖片。KDEF 資料集於 1998 年建立，最初用於心理及醫學方面，主要用在注意、情緒、知覺等方向，而在建立資料集時有特意用較為均勻且柔和的光照，且受試者的衣服顏色統一，其中包含 70 個人，男性女性各佔一半，年齡介於 20 到 30 之間，受試者位包含鬍鬚、眼鏡、耳環，且無鮮豔妝容，一樣含有七種不同表情，五種角度，共 4900 張 562\*762 的彩色圖

片。FER2013 資料集建立於 2013 年，此資料集含有 26190 張 48\*48 的灰階圖片，且圖片解析度較低，一樣共有種七表情。而 FER2013 資料集相較於 JAFFE 資料集以及 KDEF 資料集的圖片較為複雜的，其中含有不同人種、角度，且同一張圖片會出現在不同的分類中。

FER2013 的模型訓練完後，最後遷移學習到實驗室所建立的學習情緒資料集，此學習情緒資料集是以鍾沛儒[48]提出的操作型定義當作基準，並透過人工標記進行，學習情緒操作型定義如表一、[48]。

表一、 學習情緒操作型定義

學習情緒	操作型定義	
挫折	臉部表情	皺眉且嘴部動作緊閉
	敘述	好難喔，好難懂
困惑	臉部表情	皺眉，嘴部動作微張或閉上
	敘述	讓我來思考看看
無聊	臉部表情	上眼皮下降，看起來無精打采、分心
	敘述	可以趕快跳過這個概念嗎?還有其他有趣的嗎?
喜悅	臉部表情	微笑或是笑的很開心
	敘述	我懂了!好有趣!
驚訝	臉部表情	眼睛睜大或嘴巴微開且眉毛上升
	敘述	歐!不，這是甚麼
投入	臉部表情	眼睛專注在注視著螢幕
	敘述	無

### 第三章 研究方法

本章節說明建立生成對抗網路之多角度學習情緒辨識模型之方法。3.1 節介紹實驗流程；3.2 節為資料預處理。

#### 3.1 實驗流程

首先，將圖片進行資料預處理，其中包含圖片大小的裁切及對齊，最後作為 In-domainGAN 之輸入，旋轉臉部角度。對於模型適用性，以基本情緒資料集 JAFFE 以及 KDEF 來驗證模型適用性，再以學習情緒資料集生成正臉圖片，最後以 Dense Face Live Net[10]做基本情緒及學習情緒判斷，本論文之研究流程圖如圖 3.1、所示。

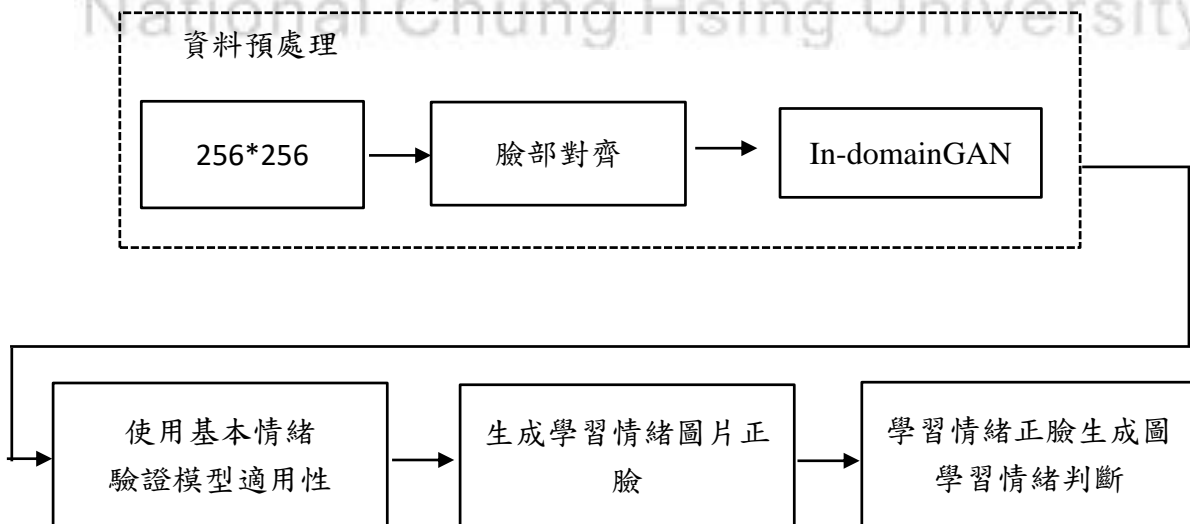


圖3.1、本論文實驗流程

## 3.2 資料預處理

本研究使用生成對抗網路進行資料預處理，首先，將選取到的基本情緒圖像以及學習情緒圖像，調整為相同尺寸，256\*256，再使用 Dlib 官方訓練好的 shape\_predictor\_68\_face\_landmarks.dat 模型使選取的圖片進行對齊的步驟，以方便 In-domainGAN 抓取特徵，接著將處理好的圖像作為 In-domainGAN 之輸入，進行臉部旋轉。

### 3.2.1 模型超參數設定

超參數(Hyper Parameter)和一般的參數不一樣，超參數是用來定義模型，在深度學習中的參數定義為模型可以自動的學習變量，例如偏差。下表為本實驗之超參數設定，表二、為 In-domainGAN 之超參數設定，而表三、為 Dense\_FaceLiveNet 之超參數設定。

表二、 In-domainGAN 超參數設定

超參數名稱	值
初始學習速率	0.01
最大迭代次數	1000(官方)
Batch Size	32
損失函數	G_logistic_nonsaturating、D_logistic_simplegp
最佳化算法	域正則化優化(Domain-Regularized Optimizaton)

表三、 Dense\_FaceLiveNet 超參數設定

超參數名稱	值
-------	---

初始學習速率	0.01
最大迭代次數	100
Batch Size	32
損失函數	Categorical Cross Entropy
最佳化算法	Adamax, SGD

學習速率對於模型訓練影響很大，當學習速率越大模型越快往最佳解前進，以增加學習效率，但當學習速率過大時，可能會讓模型無法達到最佳解，在某區域來回震盪的情況產生，以至於無法收斂；當學習速率越小時，雖訓練時間會拉長，但可以讓模型用更精確的方式往最佳解邁進，而當學習速率過小時，模型則會在區域最佳解收斂，因此學習速率過大或過小都對模型的訓練不好。

National Chung Hsing University

## 第四章 研究結果

本章共分成四部分，4.1 節說明本論文所建立之模型所用的軟硬體設備；4.2 節為基本情緒旋轉結果；4.3 節為學習情緒旋轉結果；4.4 節為兩者結果比較及學習情緒辨識分析。

### 4.1 實驗環境

本論文使用 Windows10 企業版 64 位元作業系統進行實驗，中央處理器使用 Intel(R) Core(TM) i7-7700K CPU @ 4.20GHz，顯示卡使用 NVIDIA GeForce GTX 1080 8GB，記憶體為 64GB。本論文使用 Python 程式語言進行開發，並以 Facebook 人工智慧研究團隊開發的 PyTorch[41]進行網路架構及演算法實驗，PyTorch 主要有兩個特徵[45]，它類似於 NumPy 的張量計算，可使用 GPU 加速，且是基於帶自動微分系統[43][44]的深度神經網路[45]。深度學習需極大的運算資源，因此使用 NVIDIA 的 GPU 加他們的 CUDA 進行模型的訓練來減少模型運算的計算時間。本論文所使用的軟硬體設備如表四、所示。

表四、 開發環境整體規格

名稱	規格
處理器	Intel(R) Core(TM) i7-7700K CPU @ 4.20GHz
顯示卡	NVIDIA GeForce GTX 2080 11GB
記憶體	64GB
作業系統	Windows10 企業版

開發語言	Python 3.6
深度學習運算軟體	1. PyTorch 0.4.1 2. Torchvision 0.2.1 3. Tensorflow 1.8 4. Keras 2.1.6 5. CUDA 10.0



## 4.2 基本情緒旋轉

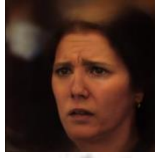
本實驗從 JAFFE 資料集以及 KDEF 資料集隨機選出共 150 張臉部圖片，七種基本情緒包含彩色及黑白圖片如表五、所示。而表六、到表十二、分別為基本情緒中害怕、生氣、厭惡、開心、中性、傷心與驚訝旋轉至各角度之圖示。

表五、 各基本情緒總張數、彩色及黑白分別張數

基本情緒	張數	彩色張數	黑白張數
害怕	24	15	9
生氣	20	16	4
厭惡	23	18	5
開心	24	19	5
中性	18	14	4
難過	21	20	1
驚訝	20	19	1

表六、 基本情緒-害怕旋轉至各角度

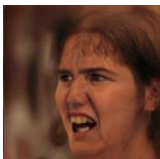












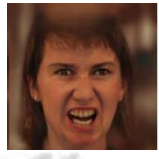



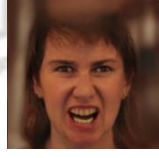


角度		
----	---	--

正負 60 度				
正負 45 度				
正負 30 度				
正負 15 度				
正負 10 度				
正負 5 度				

表七、 基本情緒-生氣旋轉至各角度


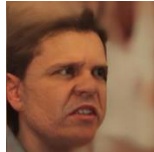

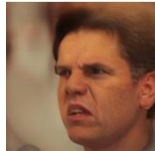
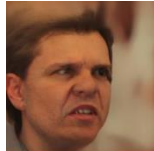


角度				
正負 60 度				

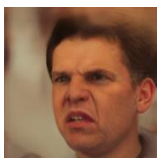

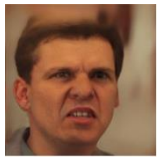

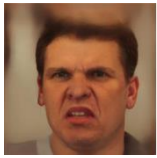


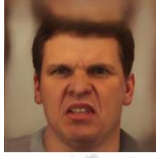
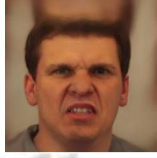



正負 45 度				
正負 30 度				
正負 15 度				
正負 10 度				
正負 5 度				















National Chung Hsing University

表八、 基本情緒-厭惡旋轉至各角度

角度				
正負 60 度				
正負 45 度				



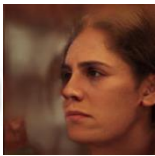



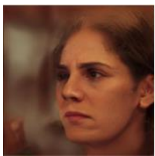
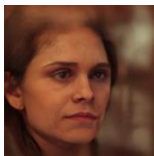


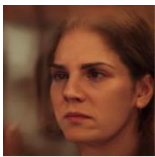
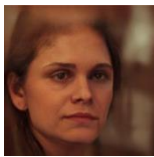


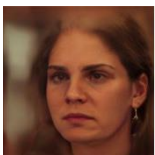
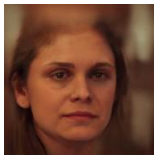
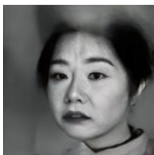
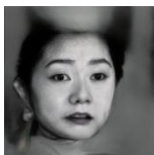
正負 30 度				
正負 15 度				
正負 10 度				
正負 5 度				

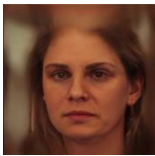
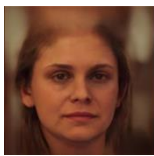

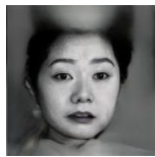
表九、 基本情緒-開心旋轉至各角度

角度				
正負 60 度				
正負 45 度				
正負 30 度				

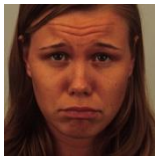


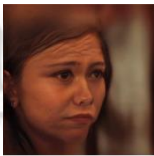


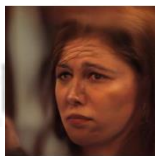
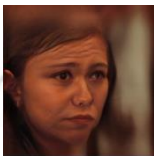


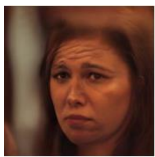
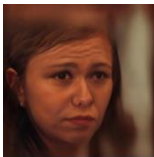


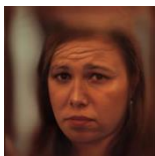
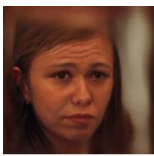


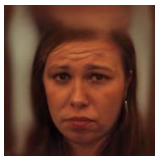
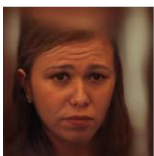


正負 15 度				
正負 10 度				
正負 5 度				

表十、 基本情緒-中性表情旋轉至各角度





角度				
正負 60 度				
正負 45 度				
正負 30 度				
正負 15 度				

正負 10 度				
正負 5 度				

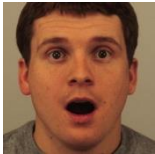


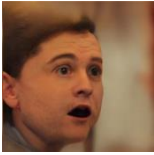



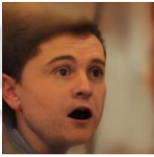








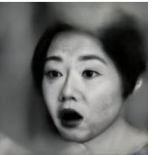

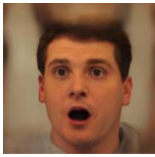



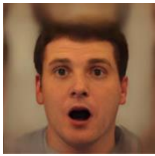
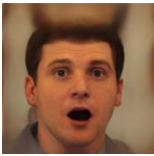


表十一、基本情緒-傷心旋轉至各角度

角度				
正負 60 度				
正負 45 度				
正負 30 度				
正負 15 度				
正負 10 度				



正負 5 度				
--------	---	---	--	---

表十二、基本情緒-驚訝旋轉至各角度

角度				
正負 60 度				
正負 45 度				
正負 30 度				
正負 15 度				
正負 10 度				
正負 5 度				

將七種基本情緒做多角度旋轉後，分別將不同角度的基本情緒做為 Dense\_FaceLiveNet 模型之輸入，進行測試後比較。首先，圖 4.1、為將原先 JAFFE 資料集以及 KDEF 資料集中隨機選取的共 150 張基本情緒輸入進 Dense\_FaceLiveNet 模型所產生的混淆矩陣，其中厭惡或錯分成生氣之外，害怕也錯分成傷心，中性表情錯分成害怕及傷心，在基本情緒資料集之完全正臉的基本情緒辨識率為 70%。

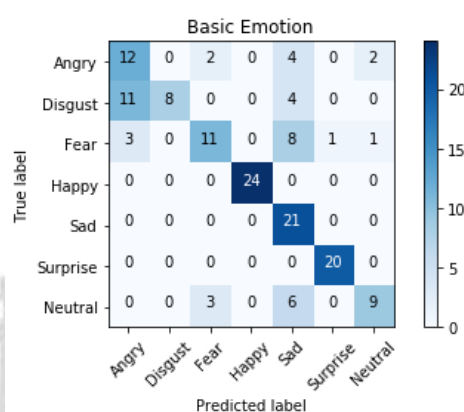


圖4.1、原先正臉之基本情緒辨識混淆矩陣

當基本情緒之完全正臉辨識完成後，將依序把旋轉過後的基本情緒圖片也輸入進 Dense\_FaceLiveNet 模型進行測試。首先，圖 4.2、為將基本情緒圖片做正負 5 度旋轉形成的混淆矩陣，圖中(a)為基本情緒圖片正五度的旋轉，圖中(b)為基本情緒圖片負五度的旋轉，分別為 53.33%以及 54%的基本情緒辨識準確率。經過旋轉正負五度的基本情緒圖片，厭惡錯分到生氣比原先更多，且害怕直接錯分成生氣以及傷心，而中性表情甚至錯分增加到驚訝。

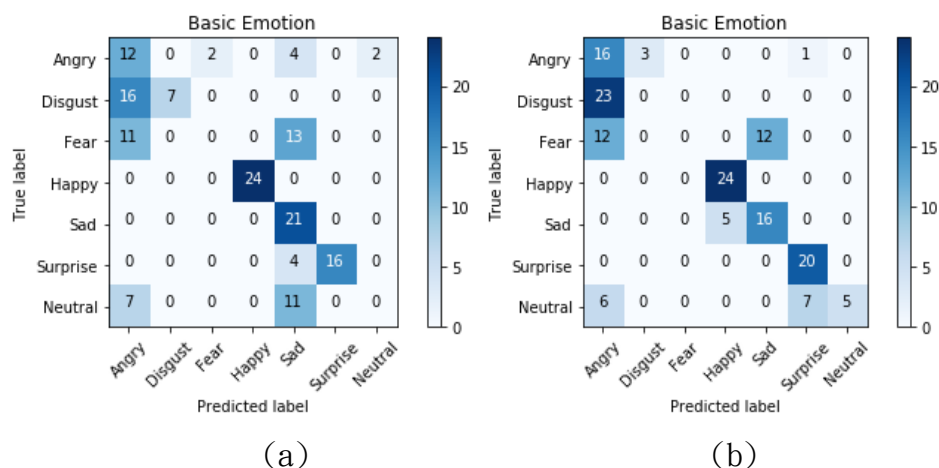


圖4.2、基本情緒圖片旋轉正負 5 度混淆矩陣

再來是將基本情緒圖片經過正負 10 度的旋轉後，輸入進 Dense\_FaceLiveNet 模型進行基本情緒辨識。圖 4.3、中(a)為將基本情緒圖片旋轉正 10 度的混淆矩陣，而(b)為旋轉負 10 度後的混淆矩陣，而基本情緒辨識準確率分別為 53.33%及 64%。旋轉正負 10 度後的基本情緒圖片，在厭惡的基本情緒圖片中依舊錯分到生氣，害怕也依舊錯分到生氣及傷心，反而中性表情在旋轉負 10 度後準確分類到中性表情的圖提高了。

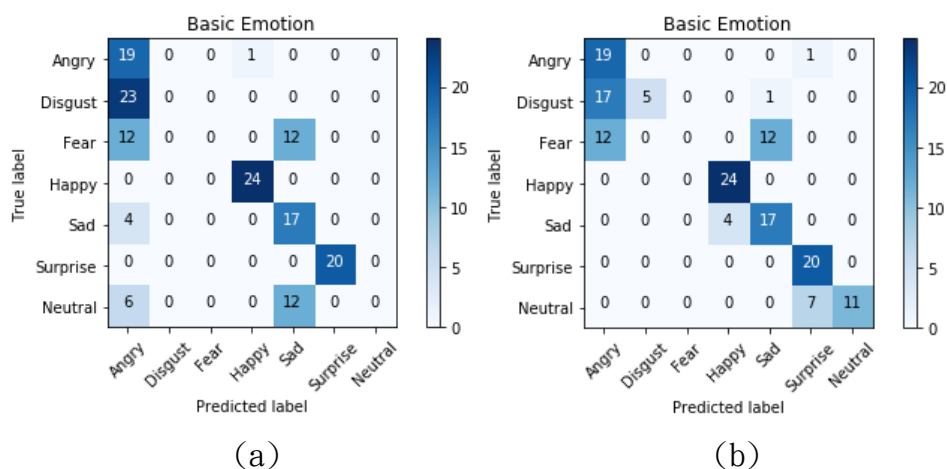


圖4.3、基本情緒圖片旋轉正負 10 度混淆矩陣

然後是將基本情緒圖片旋轉正負 15 度的基本情緒辨識準確率測試，圖 4.3、

(a)

為將基本情緒圖片旋轉正 15 度的混淆矩陣，圖 4.3、(b)則是將基本情緒圖片旋轉為負 15 度的混淆矩陣，經過兩種方向的旋轉，兩者的基本情緒辨識準確率皆為 55.33%。其中與上面兩個旋轉不同的是，害怕錯分的表情涉及到高興。

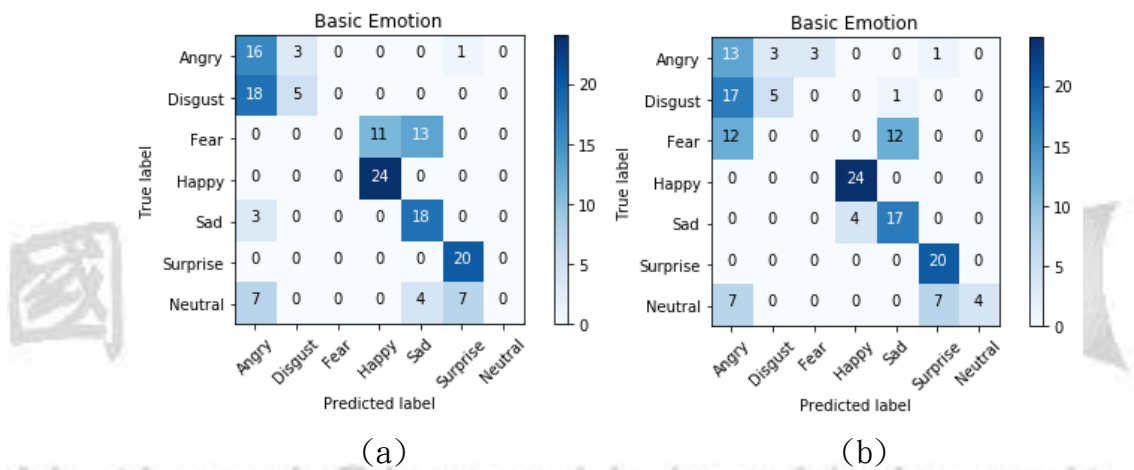


圖4.4、基本情緒圖片旋轉正負 15 度混淆矩陣

再來是將基本情緒圖片旋轉成正負 30 度的臉部圖片輸入進 Dense\_FaceLiveNet 模型進行基本情緒辨識，圖 4.5、(a)為將基本情緒圖片旋轉正 30 度的混淆矩陣，而圖 4.5、(b)則是將基本情緒圖片旋轉成負 30 度的混淆矩陣，兩者分別的基本情緒辨識準確率為 46.67%及 56.57%。旋轉到 30 度後的基本情緒辨識準確率相較原先未旋轉過的基本情緒辨識準確率兩者已相差 23.33%，且旋轉正 30 度過後基本情緒圖片中的害怕已經完全辨識為難過，中性表情也完全錯分成其他類別。



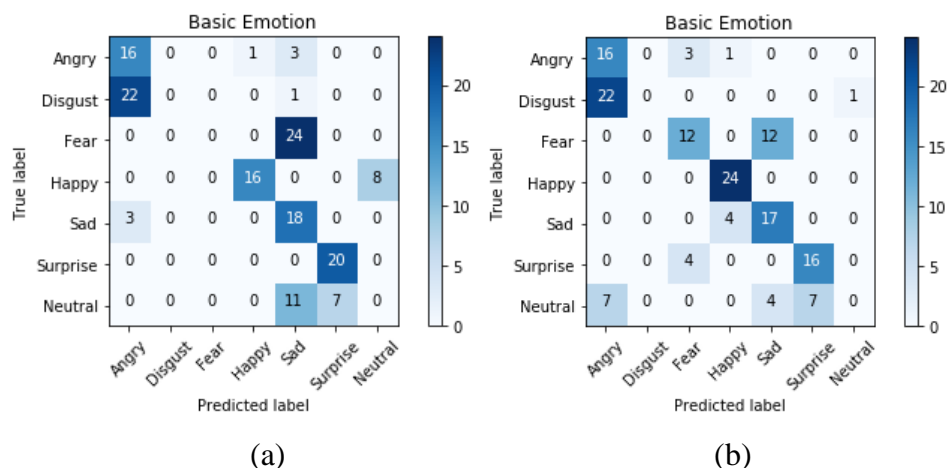


圖4.5、基本情緒圖片旋轉正負 30 度混淆矩陣

再來是將基本情緒辨識模型做正負 45 度的旋轉，輸入進 Dense\_FaceLiveNet 模型進行基本情緒辨識準確率測試。圖 4.6、(a)為基本情緒圖片旋轉正 45 度後的混淆矩陣，而圖 4.6、(b)為基本情緒圖片旋轉負 45 度的混淆矩陣，兩者的基本情緒辨識準確率皆為 46.67%。相較於旋轉 30 度後，將基本情緒圖片旋轉 45 度時，兩者皆與未旋轉過的基本情緒圖片之辨識準確率相差 23.33%，且厭惡幾乎完全被錯分到生氣，而害怕則是完全被錯分到傷心。

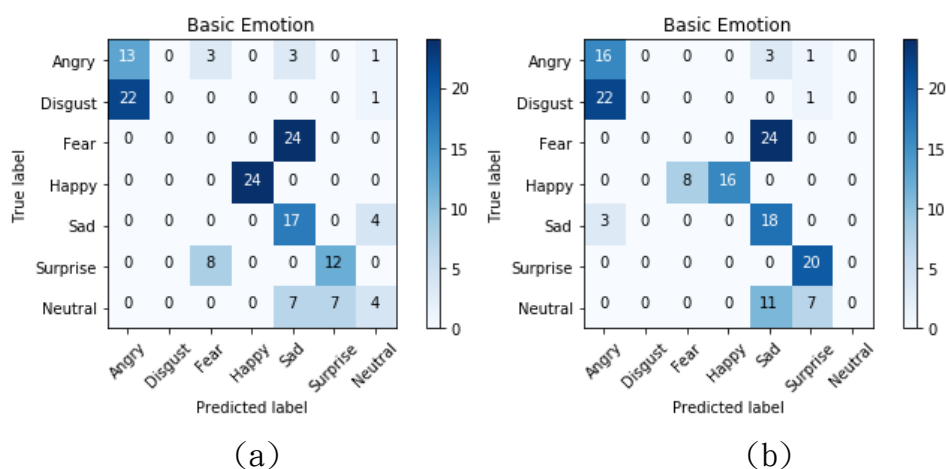


圖4.6、基本情緒圖片旋轉正負 45 度混淆矩陣

最後是將基本情緒旋轉正負 60 度的混淆矩陣，圖 4.7、(a)為將基本情緒圖片旋轉至正 60 度的混淆矩陣，而圖 4.7、(b)為將基本情緒圖片旋轉成負 60 度的混淆矩陣，該基本情緒辨識準確率分別為 42.67% 以及 40.57%，兩方向的旋轉皆讓基本情緒的辨識準確率下降。

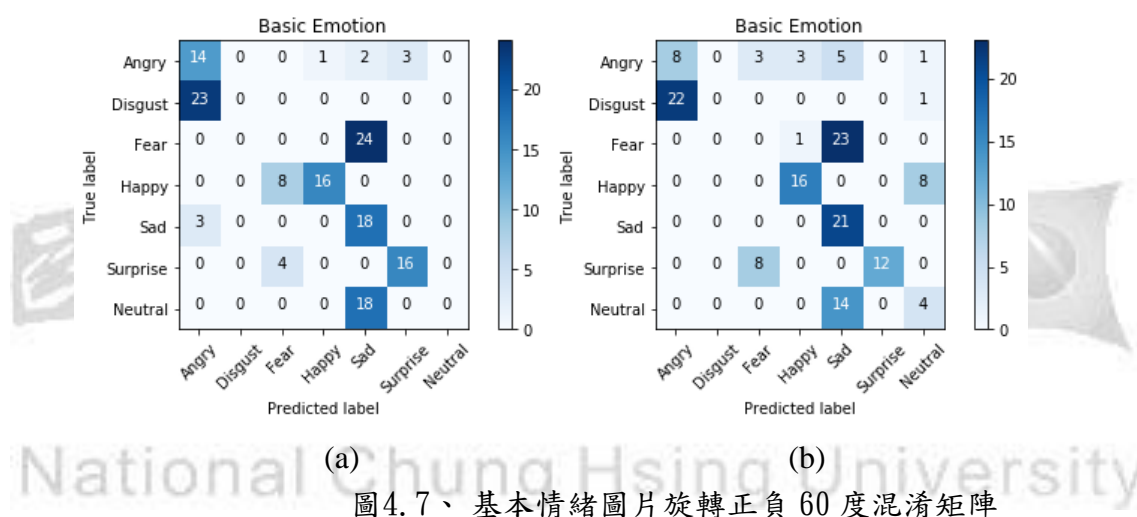


圖4.7、基本情緒圖片旋轉正負 60 度混淆矩陣

從上述實驗中，可以知道使用 Dense\_FaceLiveNet 模型進行情緒辨識時，若其中含有非正臉圖像的情況，將會導致該測試的情緒辨識準確率下滑如圖 4.8、所示。

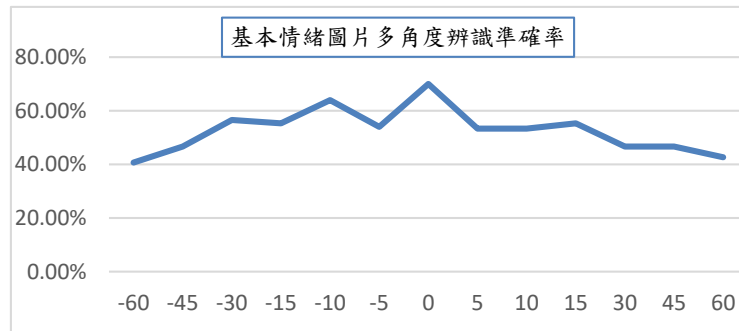


圖4.8、基本情緒圖片多角度辨識準確率

### 4.3 學習情緒旋轉結果

此實驗所用的學習情緒圖片來自實驗室所建立的學習情緒資料庫，學習資料庫包含五種學習情緒，分別為挫折、困惑、無聊、喜悅、投入、以及驚訝，共 150 張，一樣包含彩色及黑白的學習情緒圖片。

表十三、學習情緒資料庫各學習情緒張數

學習情緒標籤	圖片張數
挫折	15
困惑	21
無聊	28
喜悅	22
投入	45
驚訝	19
圖片總張數	150

當將學習情緒資料庫非正面的圖片經過裁切成 256\*256，再經過使用 Dlib 官方訓練好的 shape\_predictor\_68\_face\_landmarks.dat 模型將圖片進行對齊後，輸入 In-domainGAN 做該學習情緒圖片的正臉生成。圖 4.9、以及圖 4.10、分別為彩色學習情緒圖片在經過旋轉後的生成圖與原圖比較，以及黑白學習情緒圖像在經過旋轉後的生成圖與原圖比較。



圖4.9、彩色學習情緒圖片旋轉前後



圖4.10、黑白學習情緒圖像旋轉前後

學習情緒資料庫在旋轉之前，該學習情緒辨識準確率為 42.67%，圖 4.11、為旋轉前的學習情緒混淆矩陣，在矩陣中，挫折完全錯分第一到無聊，第二到投入；而無聊會錯分到投入；投入有些會錯分到困惑；而驚訝有些會錯分到投入。

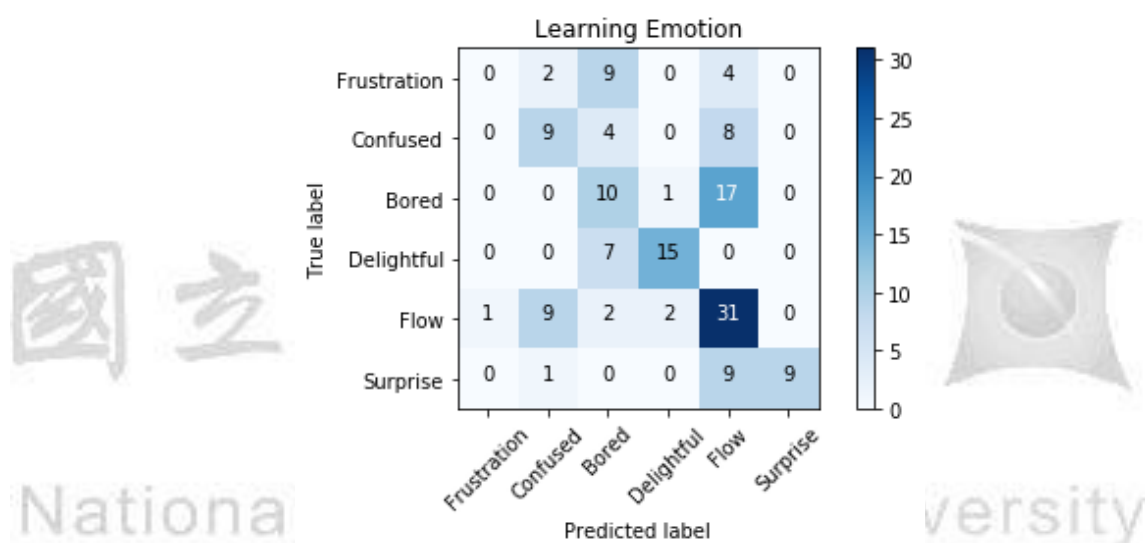


圖4.11、旋轉前學習情緒混淆矩陣

而在經過旋轉的學習情緒辨識準確率從 42.67% 提升到 57.33%，圖 4.12、為經過旋轉後的學習情緒混淆矩陣，挫折在困惑分類中增加兩張，在無聊分類中減少張數，因為挫折及困惑的差異僅在於嘴巴是否緊閉，而在輸入進 In-domainGAN 生成正臉圖片時會有些為模糊，尤其是黑白圖像；而投入在經過旋轉後分類到投入分類中的張數增加；驚訝在經過旋轉後將錯分到困惑的圖片成功分類回驚訝。

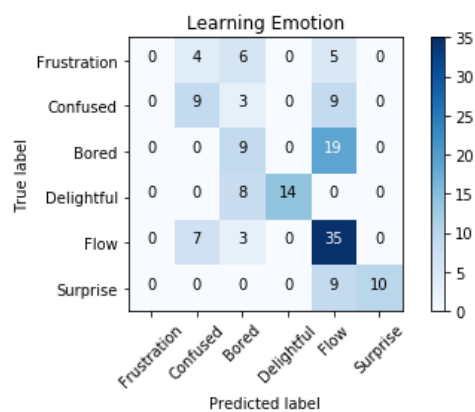


圖4.12、旋轉過後的學習情緒混淆矩陣

國立中興大學



National Chung Hsing University

## 第五章 結論與建議

### 5.1 研究結果與討論

本論文引用一個生成對抗網路，以及一個經過遷移後的卷積神經網路。此研究的目的是在於在學習場域下，台下的聆聽者不會完全專心，進而發現在本實驗室的學習情緒資料庫理沒有包含非正臉的學習情緒圖像，進而提出這個實驗。本實驗先證明，Dense\_FaceLiveNet[10]網路是否對非正臉圖像一樣適用，而實驗證明，從正臉基本情緒中得到的基本情緒辨識準確率為 70%，但是當到臉部轉為正負 60 度時，基本情緒的辨識準確率下降成 42.67%、40.57%。因此，判斷 Dense\_FaceLiveNet 網路對於非正臉圖像並不受用。

因此本實驗將學習情緒圖像中非完全正臉的圖象進行了旋轉，為旋轉前的學習情緒辨識準確率為 42.67%，而經過 In-domainGAN 旋轉後，學習情緒辨識準確率提升到 57.33%。

不過，使用 In-domainGAN[13]這個生成對抗網路時發現，此模型對於黑白圖片的旋轉表現與彩色圖片的旋轉相較之下較無法完整呈現，經過旋轉後，一些小特徵在黑白圖片時會模糊，但在彩色圖片卻不會有這樣的問題。且在使用 In-domainGAN[13]網路之前需先將要輸入的臉部圖像做一個精確地對齊，倘若圖片沒有對齊 In-domainGAN 網路[13]所要的特徵點位置，該圖片所生成出來的臉部圖像將會是一片模糊，這個問題不是只有黑白圖片會發生，在彩色圖片沒有對準的情況下也會發生如此情況。

另外，在學習情緒資料庫，因為是透過人工標定，且自行定義操作型定義，所

以有些分類會較為模糊，例如挫折及困惑，兩者的差別只在於嘴巴特徵是否緊閉或微張，而學習情緒圖像中，很多挫折中的學習情緒圖像輕閉，在分類時卻被判斷為困惑；而投入及無聊也是分類不夠明確，分類在投入的學習情緒圖片因為該受試者原先的眼睛本身較為下垂，而被判斷為無聊。

## 5.2 未來研究方向

從結果得知，In-domainGAN 此網路對於黑白圖片及或是沒有對齊的圖片較不友善，因此，未來可將做以下改良：

### (1) 去模糊方法

在研究結果中，有些錯分是因為一些小特徵經過旋轉後模糊化，因此，未來希望使用去模糊的方式，來對此問題做改善，達到辨識準確率更有所提升。

### (2) 對比增強

因為本次實驗在黑白圖片中，有些圖片的呈現較不清楚，因此無法精準對齊，倘若使用對比增強，或許能改善這類問題，



## 參考文獻

- [1] E. Fox, "Emotion Science: An Integration of Cognitive and Neuroscientific Approaches", New York: Palgrave MacMillan, 2008.
- [2] R. Picard, Affective Computing, MIT, Media Laboratory, 1995.
- [3] Mittelmann Bela, Wolff, Harold G., "Emotions and Skin temperature: Observations on Patients During Psychotherapeutic (Psychoanalytic) Interviews", 1943.
- [4] Paul Salvador Inventado, Roberto Legaspi, The Duy Bui, Merlin Suare, "Predicting student's appraisal of feedback in an ITS using previous affective states and continuous affect labels from EEG data", 2010.
- [5] Tomkins, Silvan S. (Silvan Solomon),. "Affect, imagery, consciousness.. Affect, imagery, consciousness"New York,: Springer Pub. Co. pp.1962-63.
- [6] Ekman, Paul, Davidson, Richard J, "The nature of emotion: Fundamental questions.", 1994.
- [7] Graesser, A., Mcdaniel, B., Chipman, P., Witherspoon, A., D'Mello, S & Gholson, B., "Detection of emotions during learning with AutoTutor," in Cognitive Science Society, 2006.
- [8] Rafael A. Calvo, Sidney D'Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications," IEEE Transactions on Affective Computing, pp. 18-37, 2010.
- [9] R. Breuer, "A Deep Learning Perspective on the Origin of Facial Expressions", 2017.
- [10] Nian-Xiang Lai, "The Study on Recognizing Learning Emotion Based on Convolutional Neural Networks and Transfer Learning", 2018.
- [11] Ian Goodfellow, Jean PougetAbadie, Mehdi Mirza, Bing Xu, David Warde-Farley,

- Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets." In NIPS, pp. 2672–2680, 2014.
- [12] Shan Li, Weihong Deng, "Deep Facial Expression Recognition: A Survey", 2018.
- [13] Jiapeng Zhu, Yujun Shen, Deli Zhao, Bolei Zhou, "In-Domain GAN Inversion for Real Image Editing", 2020.
- [14] Karras, T., Laine, S., Aila, T., "A style-based generator architecture for generative adversarial networks." In: CVPR, 2019.
- [15] Ira Cohen, Nicu Sebe, Ashutosh Garg, Lawrence S. Chen, Thomas S. Huang, "Facial expression recognition from video sequences: temporal and static modeling," Computer Vision and Image Understanding, vol. 91, pp. 160-187, 2003.
- [16] Tong Zhang, Mark Hasegawa-Johnson, Stephen Levinson, "Children's emotion recognition in an intelligent tutoring scenario," in Proceedings of 8th European Conference on Spoken Language Processing, Korea, 2014.
- [17] Salimans Tim, Goodfellow Ian, Zaremba, Wojciech, Cheung Vicki, Radford Alec, Chen Xi. "Improved Techniques for Training GANs," 2016.
- [18] Isola Phillip, Zhu Jun-Yan, Zhou Tinghui, Efros Alexei, "Image-to-Image Translation with Conditional Adversarial Nets. Computer Vision and Pattern Recognition," 2017.
- [19] Ho Jonathon, Ermon Stefano, "Generative Adversarial Imitation Learning. Advances in Neural Information Processing Systems," 2016.
- [20] Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," 2018
- [21] Zhu, J.Y., Krähenbühl, P., Shechtman, E., Efros, A.A., "Generative visual manipulation on the natural image manifold," In: ECCV, 2016
- [22] Perarnau, G., Van De Weijer, J., Raducanu, B., Alvarez, J.M., "Invertible

- conditional gans for image editing," In: NeurIPS Workshop, 2016
- [23] Bau, D., Strobel, H., Peebles, W., Wulff, J., Zhou, B., Zhu, J.Y., Torralba, A., "Semantic photo manipulation with a generative image prior," SIGGRAPH, 2019
- [24] Creswell, A., Bharath, A.A., "Inverting the generator of a generative adversarial network," TNNLS, 2018.
- [25] Lipton, Z.C., Tripathi, S., "Precise recovery of latent vectors from generative adversarial networks," In: ICLR Workshop , 2017
- [26] Rameen, A., Yipeng, Q., Peter, W., "Image2stylegan: How to embed images into the stylegan latent space?," In: ICCV, 2019.
- [27] Bau, D., Strobel, H., Peebles, W., Wulff, J., Zhou, B., Zhu, J.Y., Torralba, A., "Semantic photo manipulation with a generative image prior," SIGGRAPH, 2019.
- [28] Alqahtani, H., Kavakli-Thorne, M., "Adversarial disentanglement using latent classifier for pose-independent representation," ICIAP, 2019.
- [29] Alqahtani, H., Kavakli-Thorne, M., Kumar, G., "An analysis of evaluation metrics of gans. In: International Conference on Information Technology and Applications," ICITA, 2019.
- [30] Patrick Lucey , Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, Iain Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, pp. 13-18, 2010.
- [31] Lundqvist, D., Flykt, A., Ö hman, A, "The Karolinska Directed Emotional Faces - KDEF," 1998.
- [32] M. J. Lyons, M. Kamachi and J. Gyoba, "Japanese Female Facial Expressions (JAFFE)," Database of digital images, 1997.

- [33] Ian J. Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Ruifan Li, Xiaojie Wang, "Challenges in Representation Learning: A Report on Three Machine Learning Contests," ICONIP 2013: Neural Information Processing, pp. 117-124.
- [34] P. Ekman and W. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement," in Consulting Psychologists Press, Palo Alto, 1978.
- [35] Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S., "Multi-pie. Image Vis. Comput," pp. 807–813, 2010.
- [36] Zhu, X., Lei, Z., Liu, X., Shi, H., Li, S.Z., "Face alignment across large poses: a 3d solution," In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 146–155, 2016.
- [37] Prajit Ramachandran, Barret Zoph, Quoc V. Le, "Searching for Activation Functions," 2017.
- [38] Min Lin, Qiang Chen, Shuicheng Yan, "Network In Network," 2013.
- [39] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, "Rethinking the Inception Architecture for Computer Vision," 2015.
- [40] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, "Densely Connected Convolutional Networks," 2016.
- [41] Robert Guthrie, "Natural Language Processing (NLP) with PyTorch — NLP with PyTorch documentation," 2017.
- [42] Pan S J, Yang Q, " A survey on transfer learning," IEEE Transactions on knowledge and data engineering, pp. 1345-1359, 2010.
- [43] R.E. Wengert, "A simple automatic derivative evaluation program. Comm," ACM,

1964.

- [44] Bartholomew-Biggs, Michael; Brown, Steven; Christianson, Bruce; Dixon, Laurence, "Automatic differentiation of algorithms," Journal of Computational and Applied Mathematics, pp. 171-190, 2000.
- [45] 機械工業出版社, "神經網絡與 PyTorch 實戰 Application of Neural Network and PyTorch," 2018.
- [46] Brandon M. Smith Li Zhang, Jonathan Brandt Zhe Lin, Jianchao Yang , "Exemplar-Based Face Parsing," University of Wisconsin–Madison, 2013.
- [47] John Duchi, Elad Hazan, Yoram Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," Journal of Machine Learning Research, 2011.
- [48] Daniel Gutierrez, "RMSprop Optimization Algorithm for Gradient Descent with Neural Networks," 2017
- [49] 鍾沛儒, "利用臉部動作單元辨識學習情緒之研究", 台中市: 國立中興大學, 2018.
- [50] Shen, Y., Gu, J., Tang, X., Zhou, B., "Interpreting the latent space of gans for semantic face editing," In: CVPR , 2020.
- [51] Zhu, J., Zhao, D., Zhang, B, "Latently invertible autoencoder with adversarial learning," 2019.
- [52] Johnson, J., Alahi, A., Fei-Fei, L., "Perceptual losses for real-time style transfer and super-resolution," In: ECCV, 2016.
- [53] Simonyan, K., Zisserman, A., "Very deep convolutional networks for large-scale image recognition.," ICLR, 2015.