

二、研究計畫內容（以 10 頁為限）：

（一）摘要

深度學習（Deep Learning, DL）已發展成為這波人工智慧（Artificial Intelligence, AI）熱潮的主流技術。現今社會中，食衣住行都少不了人臉辨識的 AI 應用，從門禁管控、快速通關、智慧型手機中的人臉驗證（Face ID）到刷臉支付...等等，其中基於深度學習的神經網路作為辨識人臉是十分有用的工具。深層神經網路（Neural Network）的眾多模型其中之一的架構--孿生網路（Siamese Network），在人臉識別任務中扮演重要的角色，著名的 DeepFace 深度學習人臉辨識系統，就是基於孿生網路所建構的，實務上確實也得到了相當高的辨識準確率。然而對於深度學習領域上，往往需要大量的資料對模型進行訓練，以達到更好的學習結果，但在某些實際情境中卻無法得到足夠多的儲備資料，例如：嚴重特殊傳染性肺炎（COVID-19）感染後胸部 X 光表現徵狀[1]，自然災害後建築物的損壞程度[2]。這時需要可以使用少量資料就使神經網路訓練出良好結果的學習方法，元學習（Meta Learning）就是為此目的提出的一個有效且辨識結果不錯的學習模式。

在本研究計畫中，將研究透過設計一個孿生網路架構，並基於幾個常用的損失函數之下嘗試提議可用於孿生網路上更有效的損失函數，且採用 Python 程式語言及 TensorFlow 開發框架來實作所有演算法，期望提高在元學習模式下人臉識別的準確率，並深入探討孿生網路中各項因素對辨識率的影響。

（二）研究動機與研究問題

2.1 研究動機

談到深度學習，首先聯想到的就是大數據（Big Data），透過海量的資料訓練網路以達到預測、規劃的目的。然而對於只有少量資料，甚至是一個或者零個資料的資料集，例如在人臉識別應用中，神經網路見到的人臉，有很大的比例是沒有海量的資料庫儲備的，因此，如果能用更少的訓練樣本進行模型訓練，將是方便有效的。元學習就是為此問題建立的學習架構，它可以使用很少甚至零個資料就達到訓練目的，但是這在已知的深度學習概念是相互矛盾的，因為神經網路的主要概念就是依靠電腦強大的計算能力，在短時間內且不間斷的學習大量資料並不斷試錯，最終得到準確的結果，使用過少的資料會讓網路過度學習造成過擬合（overfitting），又或者是學習不足造成欠擬合（underfitting），都會造成網路表現不佳，本計畫想學習究竟元學習的架構有何特殊之處，使它可以僅使用幾個資料訓練又達成不錯的準確率。

2.2 研究問題

在人臉辨識並分類的任務中，元學習固然是一個有效的學習模式，但是在辨識準確率中，依然很難達到如同擁有海量學習資料集的網路訓練出來的準確率，甚至可能會低於人類辨識的準確率。由理論推斷，基於小數據集訓練出的模型常見問題有：(1) 泛化性不足、(2) 過擬合、(3) 樣本分佈不均以及(4) 受離群值（outlier）影響太大。因此，本研究計畫先由 Taigman 等人於 2014 年發表的 DeepFace [3] 入手，此為人臉識別並分類的研究，如果將其在元學習的模式下使用孿生網路架構訓練，對於準確率有何影響，分析與資料集大小的關係，再實作調整後的孿生網路架構與損失函數，觀察調整後的網路架構與損失函數能否解決少量訓練資料帶來的問題，以及提高模型泛化性。在少量訓練資料的元學習的概念中，輸入的訓練任務與真正想執行的辨識任務並不是相同的，所以訓練完成的模型能否在其他辨識任務中依然達到不錯的分類準確率很重要，最後比較調整前後的孿生網路在準確率的差異及應用在不同資料集中的表現。

(三) 基礎知識

3.1 元學習 (Meta Learning)

3.1.1 基本概念

一個應用在小資料集中的有效學習模式，也被稱為“學習如何學習”(Learn to Learn)，與常見的深度學習模式不同，不是學習可以達成良好結果的權重分佈，而是學習良好權重分佈的學習模式。這麼說有些抽象，舉例來說：人類的學習模式，在知識建構的學習過程中，人類沒辦法學習所有未知的知識，而是通過方法推論，例如學習過貓、豹、老虎的長相，猓猓是至今為止未曾見過的動物，我們不知道它的名字，但可以藉由它與貓、豹、老虎這類動物的相似的輪廓、外型，推論它可能也屬於貓科的動物。將學習過的模式套用在其他問題中，這種融會貫通的能力，更接近人類的學習方式。

而元學習具體的概念是讓網路訓練多種任務，若資料集內有三角形、圓形、長方形，藉由辨識三角形與圓形、三角形與長方形、圓形與長方形，找到一個演算法在各種任務間表現良好，也就是學會了如何辨識不同形狀的物體，並將其應用在辨識蘋果與香蕉上得到一個不錯的辨識結果。使用元學習學習模式的網路架構根據使用資料的數量(N)與資料類別數(K)來命名，稱為 N -shot K -way Learning。例如：One-shot 5-way Learning 的訓練資料集，擁有五個類別，每個類別只有一個資料。

因為元學習訓練資料的單位是一個任務，而不是單個資料，而每個訓練、驗證與測試子任務裡都有訓練資料與測試資料，為了不造成稱呼的混淆與更利於閱讀理解，以下將子任務裡的訓練資料命名為支持集(support set)，測試資料命名為訪問集(query set)。

3.1.2 元學習的種類

要找到一個表現良好的深度學習網路，可以通過學習權重如何設計或學習如何初始化權重等，根據學習參數的種類有不同的網路架構，具體可以分為三類：

3.1.2.1 學習特徵空間：

輸入兩個圖片進入網路，分別提取並建立特徵空間後，再比較兩者間差異來判斷輸入的圖片對是否相似。例如：1994 年由 J. Bromley 等人提出的孿生網路 (Siamese networks)[4]，以兩個結構相同、權重相同的網路分別提取兩圖片的特徵後，使用 energy function 計算兩者間的相似度，對圖片對的相似程度敏感，適合用來進行辨識人臉的任務。2017 年 Snell 等人提出的原型網路 (Prototypical networks) [5]，是先創建每個類 (support set) 的原型表示，相當於一個已分類好的原型特徵空間，如圖 1 所示，再根據訪問集與每個類的原型表示的差異度 (距離) 對訪問集分類，與聚類的概念相似，適合多元分類的任務，也可以進行半監督式學習對未標籤的物件做分類。而 2017 年由 Santoro 等人提出了另一種與原型網路概念相似的特徵空間學習網路，稱為關係網路 (Relation networks) [6]，將支持集與訪問集的特徵向量串聯 (Feature concatenation)，再用關係函數計算兩者間的關係分數 (Relation score)，它代表了這組特徵向量的關聯程度，根據類別的多寡決定有多少個串連的特徵向量，再依據關係分數的高低決定該將輸入圖片分至哪一類。由於關係網路是使用支持集與訪問集的特徵向量串聯的原理，可以用在 zero-shot 學習中，將各類別使用語意向量表示與訪問集串聯，計算兩者間的關係分數以達到分類的目的。2016 年 Vinyals 等人提出的配對網路 (Matching Network) [7]，就已經使用了前兩個網路的概念，不過是用雙向長短期記憶型網路 (LSTM) 對支持集與訪問集做特徵向量嵌入 (embedding)，再用餘弦距離度量兩者間的距離，藉此來決定輸入物件屬於何種類別，可以用作 one-shot 學習的網路模型。

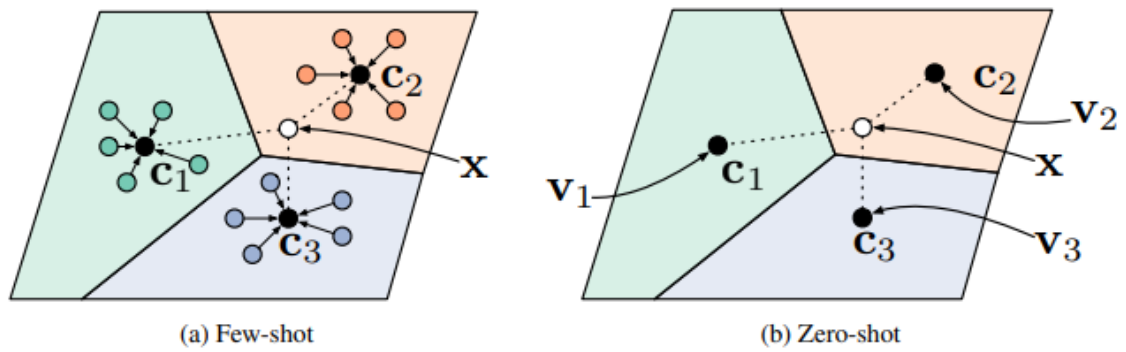


圖 1：支持集的原型特徵空間[5]

3.1.2.2 學習初始權重設定：

固定學習網路，在不同學習任務中，網路的初始參數不同，使用計算出的結果對初始參數做調整，再用調整後的初始參數進行下一輪學習任務。2017 年由 Finn 等人發表的無關模型的元學習模型 (Model-Agnostic Meta-Learning, MAML) [8]，為不斷輸入的支持集進行權重更新後，再使用訪問集計算總損失函數，然後進行權重更新。這種學習如何設定初始化權重以更快更好的達到高準確率的網路，而不是隨機初始化權重，將學習重點放在參數的初始化，因此稱為與模型無關的元學習，在元學習中做為非常指標的一種學習模式。更進一步的，2018 年 Nichol 等人提出的 Reptile [9]簡化了 MAML 的學習模型，原本 MAML 中更新參數需要加入訪問集計算總損失函數，Reptile 學習模型則不需要將資料分為支持集和訪問集，只需要不斷輸入學習任務，進行快速的權重更新，因為沒有損失函數的加入，因此更新權重的方向就由梯度下降的方向決定，使結構比起 MAML 簡單很多。而在 2017 年 Li 等人提出的 Meta-SGD [10]，則是另一種 MAML 改進的方向，除了更新初始權重外，更進一步更新學習率與更新優化方向，前兩項網路都需要通過手動設定這兩個參數，經過不斷的測試以得到合適數值，而 Meta-SGD 添加了學習率的矩陣，使學習率成為權重更新方向的參數之一，使訓練網路更加便利。

3.1.2.3 學習優化器：

假設有一個要用來訓練的基本網路，但因為學習資料量太少，導致無法使用常用的梯度下降法進行權重更新，因此需要引入一個新的優化器 (Optimizer)。這種網路稱為長短期記憶型 (Long short-term memory, LSTM) [11]，可使用遞歸神經網路 (Recurrent Neural Networks, RNN) 作為此基本網路的優化器，又稱為元學習器 (Meta Learner)。概念是初始化遞歸神經網路的權重，輸入圖片得到一個網路權重後，傳遞給基本網路，基本網路計算自己使用遞歸神經網路的參數得到的損失，回傳給遞歸神經網路，遞歸神經網路使用該損失更新自己的權重，算出新的更新權重後再傳遞給基本網路，不斷的迭代讓 RNN 作為梯度下降的角色，決定基本網路的權重更新方向。

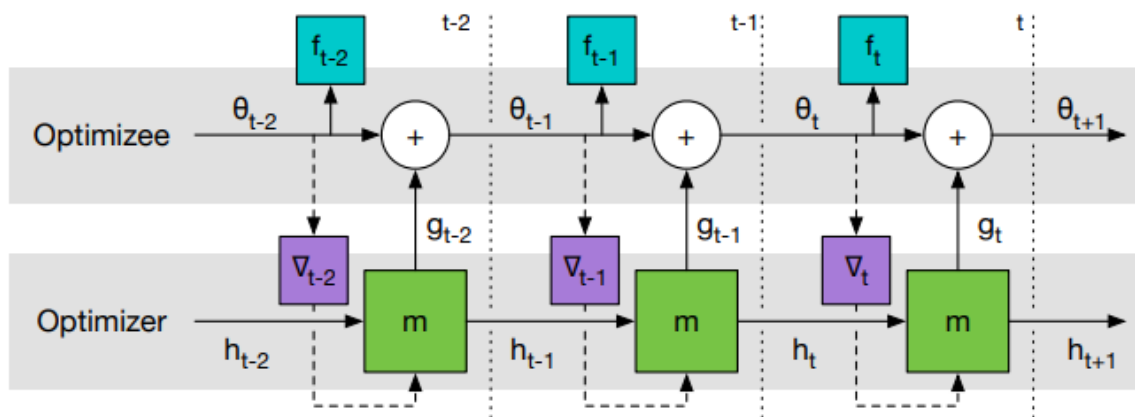


Figure 2: Computational graph used for computing the gradient of the optimizer.

圖 2：LSTM 的學習架構[11]

3.2 孿生網路 (Siamese Networks)

3.2.1 基本概念

網路架構由兩個一模一樣的神經網路，且彼此權重共享，並同時輸入兩個來自不同類別的訓練資料，生成特徵向量後通過比較兩者的差異。孿生網路的目標為讓兩個相似的圖片不會映射到特徵空間中相差太遠的位置，或拉近兩個類似的圖片的距離，以在辨識任務中達到較高的準確率。孿生網路的基本架構包含：兩個一模一樣的神經網路、energy function、loss function 及 Sigmoid 正規化或 Softmax 分類層或全連接層。

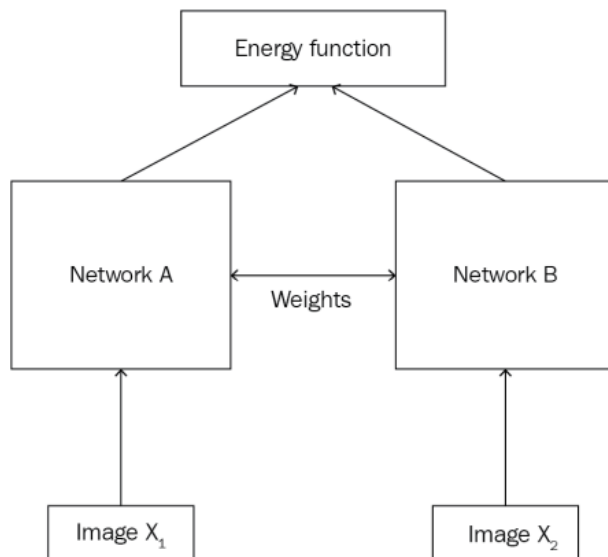


圖 3：基本孿生網路架構[12]

3.2.2 模型架構

3.2.2.1 基本網路

基本網路的選擇，往往會根據任務類型而定，常見的神經網路包含：LeNet、VGG16、Inception及ResNet。

LeNet [13]作為CNN最早的結構被美國銀行用於識別鈔票上的數字，特色為其中一層卷積層沒有使用上一層產生的所有特徵，可以降低網路的計算要求並讓卷積核學習不同的模式。VGG16 [14]是牛津大學 Visual Geometry Group 團隊的 Simonyan 與 Zisserman 承襲 AlexNet 的思路，設計在模型中建立更多層的網路，達到 16 及 19 個隱藏層，並使用小尺寸卷積核（例如：3×3）疊加起來取代大尺寸卷積核（例如：5×5），可

以增加非線性變換並減少參數，使網路對特徵的學習力更強，將準確率提高至90%，並在ILSVRC上取得很好的結果。Inception [15]使用多層 1×1 的卷積核，並用平均池化取代傳統CNN最後一層的全連接層，可以減少參數並強制Feature Map學習不同類別的特徵，提高網路的深度與寬度。在此之前根據常理，訓練層次越深訓練誤差應該更小，但事實上是當層數過多時反而效能會變差，還容易遇到梯度消失的問題，導致無法繼續訓練下去。2016年K. He等人引進殘差單元（residual unit）的概念設計ResNet [16]，成功地訓練出超深層神經網路，對深度學習的發展帶來很大的影響。

3.2.2.2 距離函數

度量兩輸入影像間的特徵向量相似度的工具稱為距離函數（distance function），如果輸出數值接近零，代表輸入對之間的相似度越高。常見的距離函數：曼哈頓距離、歐基里德距離與餘弦距離。

曼哈頓距離（Manhattan distance, L_1 ）定義兩點之間的距離由歐幾里得空間裡的兩點座標相減計算。歐基里德距離（Euclidian distance, L_2 ）的距離度量由兩點之間的直線距離計算，也就是兩點座標相減的平方。餘弦距離（cosine distance）則透過測量兩輸入向量夾角的餘弦值作為相似度度量，兩個向量方向相同時，餘弦相似度的值為1；兩個向量夾角為 90° 時，餘弦相似度的值為0；兩個向量方向相反時，餘弦相似度的值為-1。

3.2.2.3 損失函數

用來做為權重更新方向與算法性能度量的工具稱為損失函數（loss function），通常會根據任務類型改變，常見的包含：對比損失、交叉熵損失與三元損失。

對比損失（Contrastive Loss）[17]對兩輸入圖像之間的相似度非常敏感，應用於人臉驗證任務中有很好的表現。

交叉熵損失（Cross entropy Loss）[18]最常用於二元分類問題，概念是量化資料機率分佈的差異。更詳細的說，若資料的熵值低，代表資料的變動性不大很容易預測，反之，資料的熵值越高則越難預測，也就越需要學習預測此種資料的分布。而在計算資料的預測機率分佈與實際機率分佈的差異時，若差異越大則交叉熵（Cross-entropy）越高，所以交叉熵損失優化的目標與其他損失函數一樣越小越好。

三元損失（Triplet Loss）[19]是特徵空間學習中最常用到的損失函數之一，可以拉近同一類物件的距離，並分開不同類物件的特徵空間距離，因此常用於人臉辨識任務，還可以用來壓縮特徵，好處是不需要每一個類別都有大量的訓練資料，所以於元學習架構中也經常被拿來使用。除上述提到的損失函數，也可以自己設計新的損失函數或混合使用不同種類的損失函數。

3.2.2.4 應用實例

1993年，Bromley等人首先將孿生網路的概念用於支票上的手寫簽名辨識[4]，基本網路架構為時延神經網路（Time delay neural network, TDNN），並使用餘弦距離作為相似度度量函數，目標是識別簽名的真實性。而將孿生網路應用於人臉辨識相關的研究有很多，Sumit等人的研究(2005)[17]為其中之一，該方法使用的網路架構由LeNet-5與 L_1 距離和對比損失組成。在自然語言處理領域中，Ren與Xue [20]使用三元損失（triplet loss）作為損失函數的孿生網路，在口語對話理解中達到99.28%的準確率。除了使用兩個一模一樣的網路，也可以擴展成三個一模一樣的網路並行，稱為三元網路（triplet network）[21]，由Hoffer與Ailon於2015年提出。Facebook也採用孿生網路來建立DeepFace作為人臉識別系統。

3.3 人臉辨識（Face Recognition）

人臉辨識又稱人臉識別，利用提取人臉特徵信息對臉部進行生物辨識的技術，也可以提取掌紋、指紋、簽名、視網膜的生物特徵作為識別身份的依據。詳細的步驟由四個階段組成：人臉檢測=>人臉對齊=>特徵表示=>人臉分類偵測，其中涉及到的技術

都有相關的研究，人臉偵測[22]、人臉對齊[23]和人臉驗證[17][24]。

人臉特徵信息提取的技術，會使用深度神經網路作為提取工具，透過不斷的卷積計算，可以快速且有效地找出人臉的特徵信息，但人臉的視覺表現有結構相似的特色，只有五官有些微不同的定位，因此面部特徵在不同個體間的差異不是很大，所以如果是可以提取特徵並比較特徵間差異性的神經網路，可以更好的執行人臉識別任務，孿生網路就為此類神經網路架構。Facebook 的研究小組創建的 DeepFace 人臉識別系統，採用的即是孿生網路，基本網路架構為卷積層-最大池化-卷積層-3 個局部連接層-全連接層，使用 L_1 做為距離度量函數，並在野外標記人臉（LFW）數據集上達到 97.35% 的準確率，證明了孿生網路很適合用於人臉識別中，此外在 2014 年的論文中[3]，在網路末尾添加了額外的全連接層，將輸入未知人臉圖像分類在網路在訓練期間見過的 4,030 個可能的人之一。隨著人臉識別的準確率增高，與使用神經網路進行學習的便利性，隨之而來的是對於隱私與安全的爭議以及人臉特徵提取用於不法犯罪手段中，因此還需要發展相關技術進行防範。

(四)相關文獻

文獻[26]為基於元學習架構下的孿生網路應用於圖像辨識十分代表性的研究，展現了孿生網路在少資料學習下依然可以達到良好的準確率，改變了一直以來對機器學習的固有想法。該篇論文運用到的資料集為 Omniglot，由 50 個來自不同地區的字母表組成，如拉丁語和韓語到鮮為人知的地方方言。在實作時僅在各類別中抽取一個數據，稱為 one-shot 學習。提取特徵使用的基本網路如圖 4 所示，由七個卷積層，三個最大池化層，最後連接全連接層所組成，其中卷積層所使用的卷積核大小由 10×10 、 7×7 到 4×4 ，每次步長為 1，激活函數為 ReLU，池化層卷積核大小和步長為 2。而損失函數則選擇交叉熵損失（Cross Entropy Loss），但為了增加資料的多樣性，作者還有加入小的仿射轉換（Affine distortions）增強訓練集，最後在 Omniglot 的辨識結果準確度為 92%，比表現最好的 Hierarchical Bayesian Program Learning 網路略差一些，接近於人類辨識的準確度(95.5%)。將訓練後的網路架構用於 MNIST 資料集時，準確度也還有 70.3%。

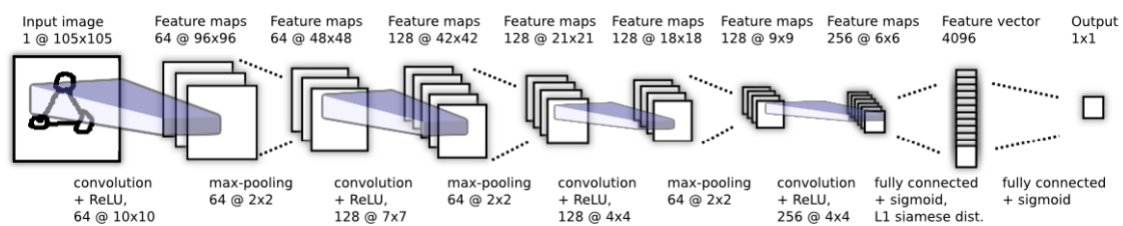


圖 4：論文[26]中使用到的孿生網路

在人臉辨識任務中，文獻[17]為最早將孿生網路應用於小量資料集的真假人臉辨識任務中，使用的人臉資料集為 AT&T，由 40 個不同類別且每一個類別都有十張不同的圖像所組成，可做為 few-shot 學習的例子。基本網路架構使用 LeNet-5，並用 L_1 距離度量輸入對的特徵向量間的距離，且提出了新的損失函數：對比損失（Contrastive Loss），可以驅動模型正確的辨識人臉的相似與不相似，最終在 AT&T 數據集中達到不錯的準確率，該文獻提出的方法很適合應用於人臉分類或驗證場景。在文獻[1]和[26]中都使用孿生網路作為訓練網路，應用在圖像辨識與分類中，2021 年 Shorfuzzaman 與 Hossain [1]使用孿生網路辨識並分類 COVID-19 病例的胸腔 X 光圖片，提出了一種損失函數，結合對比損失（contrastive loss）處理圖像辨識任務，二元交叉熵（Binary cross entropy）處理圖像分類任務，即使應用在元學習架構下（訓練資料很少），在從胸腔 X 光圖像診斷 COVID-19 時也達到了 95.6% 的準確率，在醫學應用的領域中貢獻很大。而 2022 年 Chao 等人[26]將孿生網路用於預測並分類糖尿病視網膜病變的嚴重程度，採用 ResNet 做為基本網路，並提出混和均方根損失（Mean square error loss, MSE）處

理回歸預測問題與交叉熵 (cross-entropy loss) 處理圖像分類問題作為損失函數，因為 MSE 還有正則化的功能，可以幫助模型避免過擬合，最終在 DeepDRiD 資料集中，取得了比 Auto-sklearn、AutoKeras 和 Google AutoML Vision 等成熟方法更好的結果。

(五)研究主題：模型與損失函數設計

本計畫的研究主題將探討透過設計一個孿生網路架構，並適當地設計網路的損失函數，期望提高在元學習模式下人臉識別的準確率，並嘗試深入探討孿生網路中各項模型參數與超參數設定對辨識率的影響。

5.1 模型設計

在文獻[27][28]中，作者皆混合使用不同種類的損失函數作孿生網路的聯合訓練，參考這兩篇文獻的網路架構，本計畫中使用的孿生網路架構如圖 5 所示。

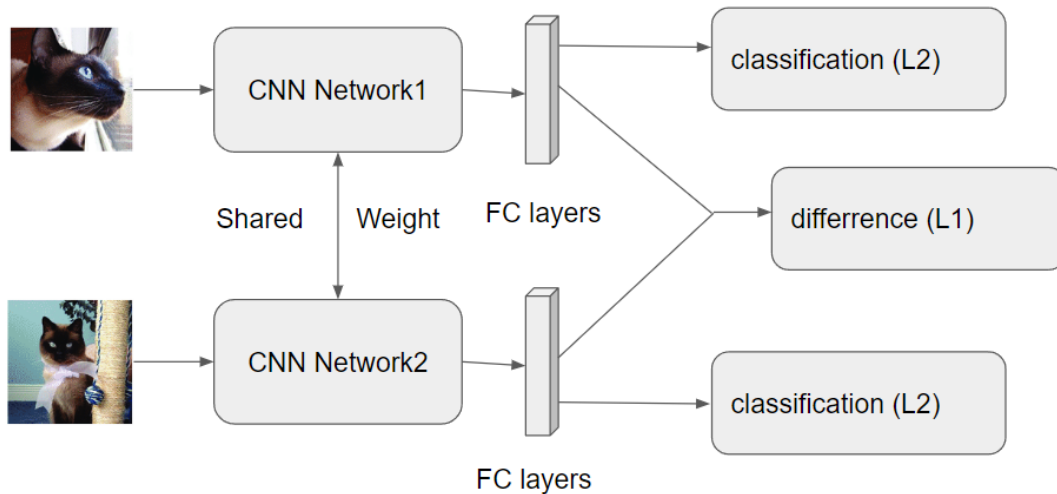


圖 5：本計畫所使用之孿生網路架構

相較於常見的用於圖像分類問題中的孿生網路架構及使用的損失函數，通常是利用對比損失 (contrastive loss) 計算輸入圖像與該類別圖像的相似度後，再加入 Softmax 層 (或全連接層) 判斷輸入圖像與各類別的相似度。但本研究計畫中使用之孿生網路，除了採用相似度損失函數，另外還考慮加入多元分類損失函數，而且權重更新方向也同時考慮多元分類損失函數與相似度損失函數的結果。

5.2 損失函數設計[1][26]

在元學習的架構下，訓練資料很少 (一個類別可能只有少量資料，甚或只有一筆資料)，若訓練中控制優化方向的損失函數對訓練資料量大小不太敏感，可以避免梯度消失的問題，因此本計畫選擇對比損失 (contrastive loss) 作為新損失函數構成之一。此外，在人臉多元分類的任務中，損失函數的優化方向要可以明確地分辨各類別，將同類別資料在特徵空間中的距離縮小，增大不同類別資料的特徵向量距離，達到更高的分類準確率，因此本計畫選擇 L -Softmax Loss 作為新損失函數的構成之二，並加入 Focal Loss 以防止訓練網路時發生過擬合。綜上所述，本研究計畫中採用的損失函數將考慮結合兩種損失函數： $Loss = L_1(\text{相似度損失函數}) + L_2(\text{多元分類損失函數})$ ，其中 L_1 為對比損失 (contrastive loss)，而 L_2 為 L -Softmax Loss 和 Focal Loss 的加總。接下來再更深入的解釋各項損失函數。

5.2.1 L_1 -相似度損失函數[17]

對於相似度辨識問題，通常使用到的是對比損失 (contrastive loss)，這是基於預測輸出和真實輸出之間相似性度量的函數，是一種度量學習損失，函數表示如下：

$$L_1 = L(W, Y, X_1, X_2) = (1 - Y) \frac{1}{2} (D_w)^2 + \frac{1}{2} Y \{\max(0, m - D_w)\}^2$$

$$D_W = ||X_1 - X_2||^2$$

W 是網路參數權重， Y 是成對標籤，如果 X_1 、 X_2 這對樣本屬於同一類，則 $Y=0$ ； $Y=1$ 代表樣本屬於不同類。 D 是樣本對的 L_1 距離。當 $Y=0$ ，調整參數，最小化 X_1 、 X_2 之間的距離。當 $Y=1$ ，且 X_1 、 X_2 之間的距離大於 m ，則不做優化（省時省力）；當 X_1 、 X_2 之間的距離小於 m ，則增大兩者距離到 m 。

對比損失函數可以很好的辨別輸入圖片對為相似或者是不相似，但對於少量資料訓練下，卻可能造成訓練過度而出現過擬合的結果，且會讓網路對相似度過於敏感，這裡的相似度可能是顏色相似或是位置相似。因此幼年的獅子與橘貓的相似度可能會高於橘貓與暹羅貓的相似度，所以還需要加上其他可以互補對比損失函數缺點的函數。

5.2.2 L_2 -多元分類損失函數-1[29]

本計畫使用到的第一個用於多元分類任務的損失函數為 L -Softmax (Large-Margin Softmax Loss)，在原 Softmax Loss 下使用餘弦距離代替原 L_1 距離作為距離度量的函數，表示如下：

$$L_S = (-\log \frac{e^{\|W_{yi}\| \|X_i\| \varphi(\theta_{yi})}}{e^{\|W_{yi}\| \|X_i\| \varphi(\theta_{yi})} + \sum_{j \neq yi} (e^{\|W_j\| \|X_i\| \varphi(\theta_j)})})$$

$$\varphi(\theta) = \begin{cases} \cos(m\theta) & 0 \leq \theta \leq \frac{\pi}{m} \\ D(\theta) & \frac{\pi}{m} \leq \theta \leq \pi \end{cases}$$

由於使用餘弦距離作為度量工具，根據圖 6 所示，可以減少類別內距離並增大類別外距離，加強分類的準確度，還可以學習更多的判別特徵，但缺點是對於小量資料可能會分類過度而產生過擬合問題。

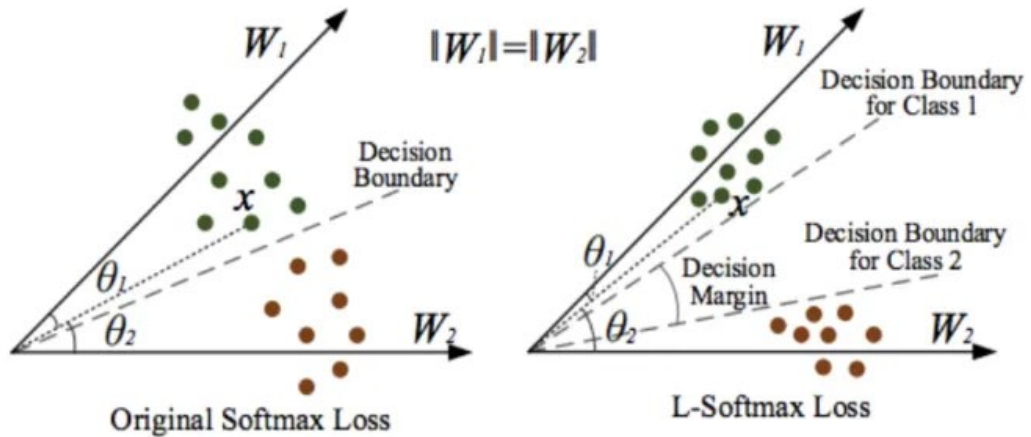


圖 6：L-Softmax 示意圖[31]

5.2.3 L_2 -多元分類損失函數-2[30]

本計畫使用的第二個用於多元分類任務的損失函數為 Focal Loss，除用以解決樣本類別不均的問題外，也加以限制了樣品中易分類樣品的比重，函數表示如下：

$$L_f = (-\alpha_t (1 - p_t)^\gamma \log(p_t)) , p_t = \begin{cases} p & y = 1 \\ 1 - p & \text{else} \end{cases}$$

當 $\alpha_t = 0.5, \gamma = 0$ ，此函數就變為常見的交叉熵損失。當 γ 增加時，調製係數 $1 - p_t$ 也會增加。專注參數 γ 平滑地調節了易分類樣本調低權值的比例。調製係數 $1 - p_t$ ，目的是通過減少易分類樣本的權重，使模型在訓練時更專注於難分類的樣本。 p 表示預測樣本屬於 1 的機率（範圍為 0~1）。

與交叉熵損失函數相似，解決“簡單”樣本主導訓練過程並關注學習對“困難”樣本（難以分類或屬於少數類別的樣本）進行分類的問題。優點為：(a) 降低了“簡單”樣本的損失權重，並更加強調“困難”樣本；(b) 可能可以解決對比損失造成的過擬合問題，

因為不是把關注都放在易分類的資料對上，且降低對於錯誤方類別資料的懲罰；(c)因為對於少量資料的情境下，分類錯誤的資料也有可學習的特徵。缺點為很難用於人臉辨識任務中，因為它不會對給定資料的類型約束，即模型可以學習任何特徵，無論相似的資料點在轉換後是否彼此靠近。所以還需要搭配其他類型的損失函數以應用在人臉辨識任務中。

5.3 執行步驟

根據研究問題，本計畫的執行步驟如下：

- (a) 實作基於少量訓練資料下的文獻[3]，從觀察準確率受資料集大小的影響，分析使用少量資料學習具體會造成訓練結果產生甚麼缺點。
- (b) 實作本計畫提出的更改後的孿生網路架構與損失函數應用於相同任務中。
- (c) 比較更改前後的孿生網路的準確率查看能否解決第一步產生的問題。
- (d) 將更改前後的孿生網路應用在不同資料集中，分析兩者的表現。

(六)預期結果

本計畫的預期結果列述如下：學生期望透過此計畫的進行

- (a) 可深入理解 AI 與 Deep Learning 的基本原理與技術細節。
- (b) 熟悉 Python 語言、OpenCV 開發環境與 TensorFlow Framework，並學會如何利用這些資源與技術來實作深度學習系統，熟悉深度學習程式編碼技巧。
- (c) 理解設計 CNN 的基本原理與技術細節，並且利用 CNN 模型，提取影像特徵。
- (d) 了解孿生網路中各項因素對辨識率的影響。
- (e) 本研究預期設計一個損失函數，在基於元學習架構下人臉多元辨識任務中提高準確率。
- (f) 了解如何蒐集相關資料，如何閱讀學術文獻，如何整理文獻資料。
- (g) 學會如何思考問題，如何解決問題。
- (h) 學會如何進行實驗，如何分析結果，如何呈現結果。
- (i) 最後學會如何撰寫學術報告，如何發表論文。

(七)需要指導教授指導內容

學生現就讀大三，做研究與撰寫研究計畫之經驗不多，第一次參與國科會的專題寫作計畫，要開始學習如何作研究，需要指導教授指導的內容包括：

- 一、如何蒐集文獻與資料。
- 二、如何閱讀文獻與有系統的整理資料。
- 三、如何思考指導教授提出的問題並學習提出解決方案。
- 四、如何撰寫Python程式，並學習使用OpenCV提供的功能有效地解決影像處理的問題與執行統計分析的工作。
- 五、學習使用TensorFlow框架提供的功能有效地完成訓練神經網路的任務。
- 六、學習如何進行實驗、分析實驗結果與整理實驗數據。
- 七、學習如何撰寫結案報告與文獻。
- 八、學習如何撰寫學術論文，並嘗試投稿研討會或期刊發表研究成果。

(八)參考文獻

- [1] M. Shorfuzzaman and M. S. Hossain, "MetaCOVID: A Siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients," *Pattern recognition*, 113: 107700, 2021.

- [2] E. Khvedchenya and T. Gabruseva, "Fully convolutional Siamese neural networks for buildings damage assessment from satellite images," *arXiv:2111.00508*, 2021.
- [3] Y. Taigman, *et al.*, "Deepface: Closing the gap to human-level performance in face verification," *CVPR*, 2014.
- [4] J. Bromley, *et al.*, "Signature verification using a siamese time delay neural network," *Advances in Neural Information Processing Systems*, 6, 1993.
- [5] J. Snell, K. Swersky and R. Zemel, "Prototypical networks for few-shot learning," *Advances in Neural Information Processing Systems*, 30, 2017.
- [6] A. Santoro, *et al.*, "A simple neural network module for relational reasoning," *Advances in Neural Information Processing Systems*, 30, 2017.
- [7] O. Vinyals, *et al.*, "Matching networks for one shot learning," *Advances in Neural Information Processing Systems*, 29, 2016.
- [8] C. Finn, P. Abbeel and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," *ICML*, 2017.
- [9] A. Nichol, J. Achiam and J. Schulman, "On first-order meta-learning algorithms," *arXiv:1803.02999*, 2018.
- [10] Z. Li, *et al.*, "Meta-SGD: Learning to learn quickly for few-shot learning," *arXiv:1707.09835*, 2017.
- [11] M. Andrychowicz, *et al.*, "Learning to learn by gradient descent by gradient descent," *Advances in Neural Information Processing Systems*, 29, 2016.
- [12] Ravichandiran and Sudharsan, *Hands-On Meta Learning with Python*, Packt Publishing, 2018.
- [13] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov., 1998. **【LeNet】**
- [14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *ICLR*, 2015. **【VGGNet】**
- [15] S. Szegedy, *et al.*, "Going deep with convolutions," *CVPR*, 2015. **【GoogLeNet/Inception】**
- [16] K. He, *et al.*, "Deep residual learning for image recognition," *CVPR*, pp. 770-778, 2016. **【ResNet】**
- [17] S. Chopra, R. Hadsell and Y. LeCun. "Learning a similarity metric discriminatively, with application to face verification," *CVPR*, 2005.
- [18] P.-T. d. Boer, *et al.*, "A tutorial on the cross-entropy method," *Annals of operations research*, 134.1: 19-67, 2005.
- [19] F. Schroff, D. Kalenichenko and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *CVPR*, 2015.
- [20] F. Ren and S. Xue, "Intention detection based on siamese neural network with triplet loss," *IEEE Access*, vol. 8, pp. 82242-82254, 2020.
- [21] E. Hoffer and N. Ailon, "Deep metric learning using triplet network," *ICLR*, 2015.
- [22] M. Osadchy, M. Miller and Y. LeCun, "Synergistic face detection and pose estimation with energy-based models," *Journal of Machine Learning Research*, 8:5, 2007.
- [23] Y. Sun, X. Wang and X. Tang, "Deep convolutional network cascade for facial point detection," *CVPR*, 2013.
- [24] G. B. Huang, H. Lee and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," *CVPR*, 2012.
- [25] G. Koch, R. Zemel and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," *Deep Learning Workshop on ICML*, vol. 2, 2015.
- [26] C. Feng, *et al.*, "MTCSNN: Multi-task clinical siamese neural network for diabetic retinopathy severity prediction," *arXiv:2208.06917*, 2022.
- [27] I. A. Lungu, *et al.*, "Siamese networks for few-shot learning on edge embedded devices," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 10, no. 4, pp. 488-497, 2020.
- [28] I. A. Lungu, Y. Hu and S.-C. Liu, "Multi-resolution siamese networks for one-shot learning," *2nd IEEE Int. Conf. on Artificial Intelligence Circuits and Systems (AICAS)*, 2020.
- [29] H. Wang, *et al.*, "Cosface: Large margin cosine loss for deep face recognition," *CVPR*, 2018.
- [30] Y. Zhao, *et al.*, "Paralleled attention modules and adaptive focal loss for Siamese visual tracking," *IET Image Processing*, vol. 15, no. 6, pp. 1345-1358, 2021.