

國立陽明交通大學

大學部專題報告

題目：

條件性人臉及漫畫風格轉換

系 別：資訊工程學系

指導教授：邱維辰 教授

專 題 生：葉晨

學 號：0716081

目錄

一、摘要	p.3
二、研究目標	p.3
三、方法與步驟	
3-1 實驗流程	p.4
3-2 資料取得	p.4
3-3 資料前處理	p.5
3-4 訓練模型	
3-4-1生成對抗網路 (GAN)	p.5
3-4-2 U-GAT-IT 原架構	p.5
3-4-3 對 U-GAT-IT 架構的修改	p.6
3-5 訓練方法	p.7
3-6 測試方法	p.8
四、結果與成效分析	
4-1 實驗結果	p.9
4-2 成效分析	p.9
五、結論與未來方向	
5-1 結論	p.11
5-2 未來方向	p.11
六、參考資料	p.11

一、摘要

近年來，機器學習已經被廣泛運用在許多領域，其中非監督式學習(Unsupervised learning)常被運用在圖片到圖片的轉換(image-to-image translation)上。圖片到圖片的轉換的應用包括利用人的照片生成動物的照片、將一幅畫作賦予不同畫家的風格...等。本次專題研究是參考2020年發布的一篇論文及架構[1][2]，論文中利用了一種生成對抗網路(Generative Adversarial Network, 以下簡稱GAN)，作者命名為 U-GAT-IT，來將人臉轉變為漫畫風格的臉。在這次專題研究中，我們對 U-GAT-IT 做優化及改良，除了擴大訓練及測試的 dataset 種類以外，也加入了條件的設定，使 U-GAT-IT 的轉換結果更加自然、功能也更完善。最後我們也利用問卷來評估我們的成效，在許多方面都比原本的 U-GAT-IT 更勝一籌。

二、研究目標

U-GAT-IT 的功能為輸入一張人臉之後，生成出一張相似的漫畫風格的臉，而我們這次的目標為優化 U-GAT-IT 的功能，我們認為可以優化的部分有以下三點：

1. 使 U-GAT-IT 所呈現出來的漫畫臉更貼近真實人臉

我們認為 U-GAT-IT 生成的漫畫臉跟原本的人臉相比，在臉型、髮型上已經有很高的相似度，但是在五官、表情等較細微的部分，仍然不夠貼近原本的真實人臉。因此我們希望能夠改善這個部分，讓生成的結果在五官及表情方面能夠更加相似、自然。

2. 擴充 U-GAT-IT 的 dataset 類型

U-GAT-IT 原作者在訓練及測試模型時，人臉及漫畫臉的訓練資料(training data)及測試資料(test data)皆為年輕的女性。我們希望這個模型可以擴及給更多人使用，因此將數據集(dataset)擴充為四個類別：年輕女性、中年女性、年輕男性、中年男性，分別去訓練及測試模型，使其的運用更加廣泛。

3. 加入條件性轉換

U-GAT-IT 原本生成的漫畫臉是基於原本人臉的五官特徵，因此人臉與漫畫臉的視覺年齡及性別並不會相差太遠。而我們希望加入的功能是，在生成漫畫臉時指定一個條件(性別+年齡)，並且在原本人臉的基礎下，生成一張符合條件的漫畫臉。例如：輸入一張年輕女性的人臉，條件指定為中年男性，模型便會根據該人臉的特徵，生成出一張中年男性的漫畫臉。

以上三點為本次專題主要實驗目標，希望可以讓 U-GAT-IT 的功能更加完善、被更廣泛的使用。

三、方法與步驟

3-1 實驗流程

如下圖所示，本研究大致分為四個步驟。首先，取得一定數量的人臉及漫畫臉。接下來將人臉及漫畫臉調整為同一尺寸，並且進行分類(年輕女性、中年女性、年輕男性、中年男性)。之後將人臉及動漫臉放入修改後的 U-GAT-IT 進行訓練，最後再用不包含在訓練資料裡的人臉進行測試，並以問卷調查作為成效分析的媒介。



3-2 資料取得

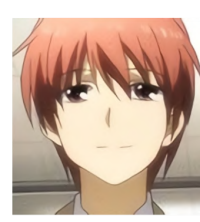
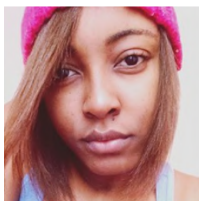
為了使此模型有更廣泛的應用，我們將原本的訓練資料(traning data)分為四個類別：年輕女性、中年女性、年輕男性、中年男性，詳細數量如下表所示：

	年輕女性	中年女性	年輕男性	中年男性
人臉	3400	3400	3400	3400
漫畫臉	5395	455	2182	1101

而測試資料(test data)的詳細數量如下：

	年輕女性	中年女性	年輕男性	中年男性
人臉	50	50	50	50

人臉的資料來源為網路上公開的資源[3]，漫畫臉的取得則是利用爬蟲技巧到動漫網站[4]抓取動漫人物的圖片，並且手動挑選適合的圖片資料。因為動漫人物的圖片取得較不易，因此有數量較不平均的狀況。



3-3 資料前處理

1. 統一尺寸

為確保在訓練時不會因為圖片尺寸而出現誤差，我們希望每張人臉及漫畫臉的尺寸皆為 200*200 pixels。

人臉部分，因為下載時尺寸便為 200*200，故不用調整大小。漫畫臉部分，利用爬蟲技巧將人物圖片從動漫網站抓下來之後，利用 Faster-RCNN 模型[5]自動找出漫畫人物的臉部並切割下來。因為每張切割下來的漫畫臉大小不盡相同，故我們利用了網路上的 Super resolution 程式[6]，在盡量不影響畫質的前提下，將所有漫畫臉的尺寸調整為 200*200。

2. 標記性別及年齡

為了在訓練時可以讓模型清楚知道該圖片的類別，我們將圖片的性別及年齡標註在其檔名。因為所有人臉在下載時都已經標記好了性別及年齡，因此不用再多做處理。漫畫臉部分，則是利用手動方式將圖片分類之後再於檔名標註其類別。

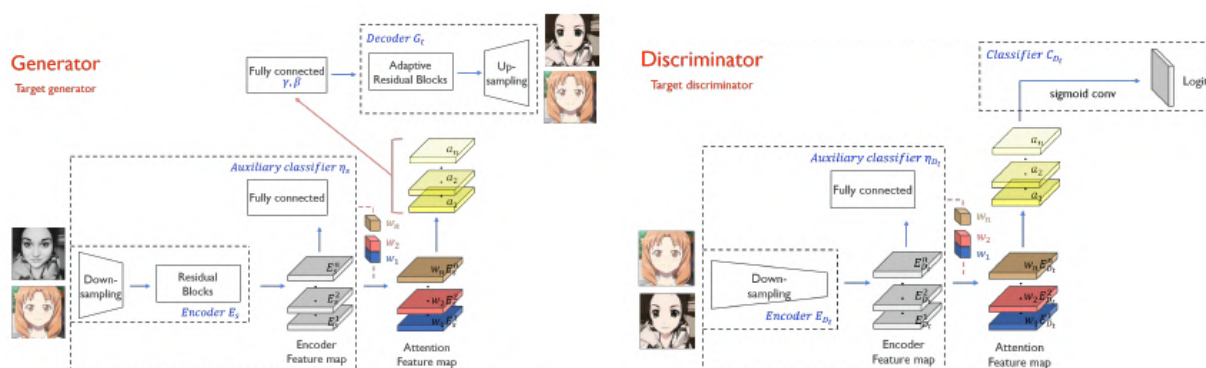
3-4 訓練模型

3-4-1 生成對抗網路 (GAN)

生成對抗網路是一種非監督式學習，其學習方法是利用競爭的方式，整個網路包含生成網路 (Generative Network / Generator) 以及鑑別網路 (Discriminating Network / Discriminator)，透過兩者互相對抗、修正，最後產生的結果便會越來越好。簡單做個比喻，生成網路好比做假鈔的犯人，鑑別網路好比驗鈔機，犯人做出假鈔讓驗鈔機檢驗有多接近真鈔，根據評斷結果再做出更接近真鈔的假鈔，而驗鈔機在一次次的評斷中提稱，提升檢驗水準，讓假鈔更不容易過關。如此一來一往的過程之後，最後比賽的結果就是一張幾可亂真的假鈔了。

3-4-2 U-GAT-IT 原架構

U-GAT-IT 為 2020 發布的一種 GAN 模型，其最主要的目的為 image-to-image translation，以下為 U-GAT-IT 的架構圖[1]：



Generator 部分，包含了 Encoder、Auxiliary classifier、Fully connected layer(γ, β)、Decoder 四個部分。因為原論文中有詳細說明，因此以下簡略的講解各個部分的功能：

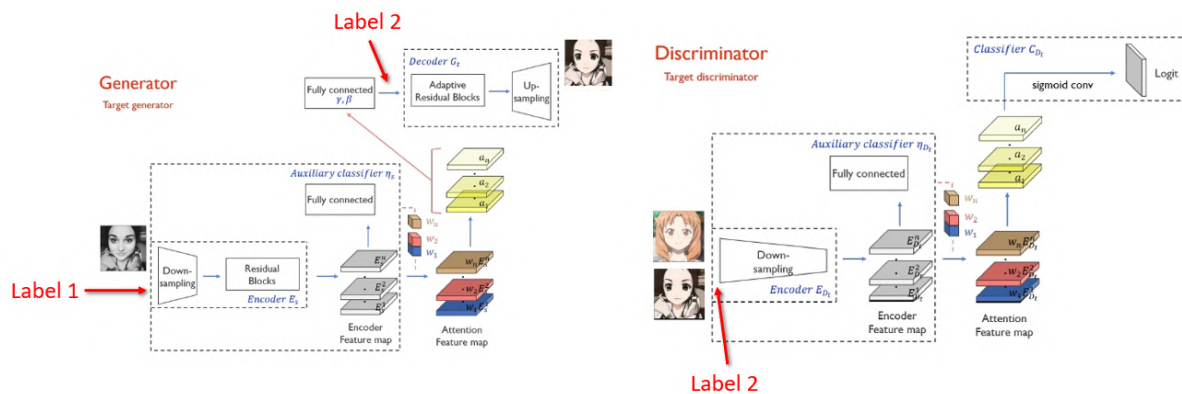
輸入一張人臉之後, Encoder 會從圖片裡面提取各種特徵(Encoder feature map); Auxiliary classifier 會經過學習, 為每個特徵加上權重(w_1, w_2, \dots), 代表每個特徵擁有不同的重要性 (Attention feature map); Fully connected layer 會利用 Attention feature map 決定 γ, β 兩個參數, 再 Normalization 時會扮演重要角色; 最後是 Decoder, 會將處理過的 Attention feature map 轉換為漫畫臉的圖片。

Discriminator 的架構分為三個部分: Encoder、Auxiliary classifier、Classifier, 以下簡單講解各部分的功能:

Discriminator 會輸入兩張漫畫臉(一張為真實的漫畫臉、一張為 Generator 產生的漫畫臉), 先經由 Encoder 提取特徵; 之後經由 Auxiliary classifier 加上權重; 最後 Classifier 再為 Generator 產生的漫畫臉打一個分數(logit), 代表其真實性。

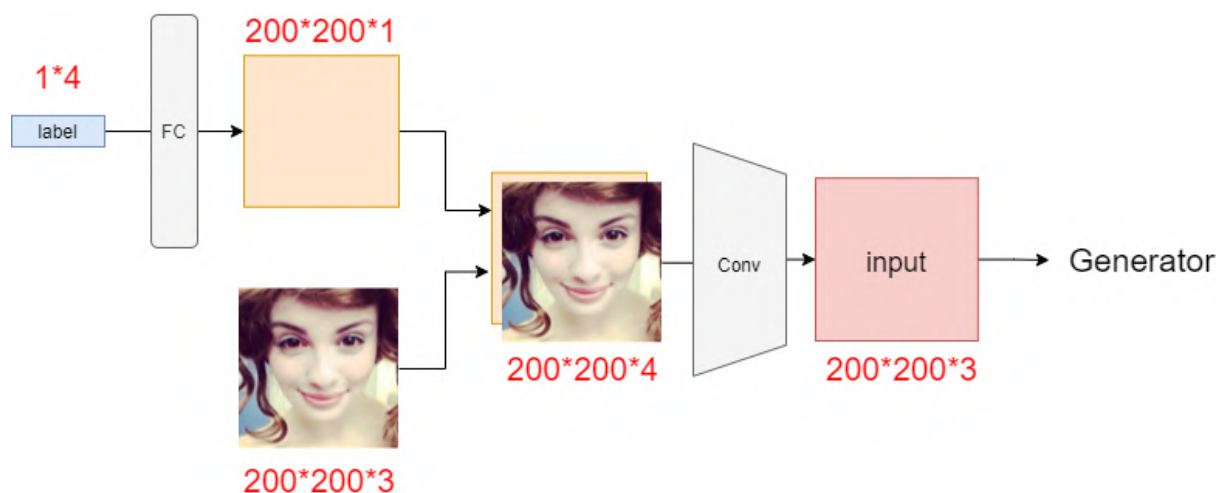
3-4-3 對 U-GAT-IT 架構的修改

為了達到上述的三點研究目標, 我們在原本的 U-GAT-IT 架構加入了圖片的類別:



如圖所示, label 1 為人臉原本的類別, label 2 漫畫臉應該要轉換出來的類別, 而且 label 2 不一定跟 label 1 相同。舉例來說, 若輸入了一張年輕女性的人臉, 希望轉換出中年男性的漫畫臉, 則 label 1=年輕女性, label 2=中年男性。

Generator 部分, 在輸入人臉時, 讀取檔名以確定該人臉的類別, 並將類別轉換成 one-hot vector($1*4$), 利用一層全連接層(Fully connected layer)將其維度轉換為跟人臉的長寬相同 ($200*200*1$), 之後將人臉及轉換過後的 label 堆疊起來($200*200*4$), 再經過一層卷積層 (Convolutional layer), 將維度調整為加入 label 前的人臉的維度($200*200*3$), 以確保之後的架構可以順利運行, 最後再將其輸入 Generator。



一開始就將人臉的類別(label 1)輸入的原因是，希望模型可以提早知道該人臉的類別，在已知更多資訊的情況下，更快的學習到更細緻的五官或表情。

在進入 Decoder 之前，我們利用類似的技巧將希望轉換漫畫臉的類別(label 2)加入網路，並且一起送入 Decoder，希望模型在生成漫畫臉時，會受到 label 2 的影響而產生出符合特定類別的漫畫臉。

Discriminator 部分，在輸入真實與虛假的漫畫臉的同時，也會將 label 2 用跟 Generator 相同的技巧一起輸入 Encoder，以作為 Discriminator 判斷分數的另一項指標。

3-5 訓練方法

對抗生成網路的目的是希望經由 Generator 以及 Discriminator 的對抗，使 Generator 產生的結果越來越真實。因此我們要訓練的不只是 Generator 的參數，希望 Generator 的生成結果更接近真實資料，同時我們也要訓練 Discriminator 的參數，希望其可以更清楚的鑑別真實資料與 Generator 所產生的虛假資料。

每個 epoch 的訓練流程圖如下：



為了確保模型能夠平均學習到每個類別的特徵，我們選擇訓練資料的方法如下：

Generator 部分，我們會在所有人臉之中隨機選擇一張，因為每個類別的人臉數量相同，所以四個人臉的類別(label 1)被選擇的機率相同。同時，我們也是隨機選擇想要轉換的漫畫臉的類別(label 2)，一起輸入 Generator。

Discriminator 部分，則是會根據 label 2 的類別，從訓練資料中選擇相對應的漫畫臉圖片來當作真實資料，並將真實資料與 Generator 產生的虛假資料做比較。例如，假設本次被挑選的 label 2 是代表年輕女性，Discriminator 便會從訓練資料中隨機挑選一張年輕女性的漫畫臉，當作真實資料，並且將其與 Generator 產生的漫畫臉一起輸入，最後產生一個分數，代表真實資料跟虛假資料的相似程度。

將選擇的資料輸入 Generator 之後, Generator 便會根據真實的人臉以及我們指定的目標類別, 轉換出擁有人臉特徵, 同時也符合目標類別的漫畫臉, 此處我們稱為虛假資料。接著 Discriminator 會將虛假資料以及從訓練資料中挑選出來的真實資料一起輸入, 並且沿用原論文的損失函數(loss function), 以均方誤差(MSE)的方法計算四種不同 loss, 最後依據加權之後的 loss 產生一個分數, 代表虛假資料的真實程度。

之後 Discriminator 會根據 loss 優化其參數, 此處我們採用的是 Adam 優化器。因為考慮到 Generator 的負擔較 Discriminator 來的大, 可能造成 Generator 學習速度較慢。因此我們設定 Discriminator 每訓練兩個 epoch 才更新一次其參數, 希望平衡 Generator 以及 Discriminator 的學習速度。

最後 Generator 會計算自己的 loss, 代表虛假資料跟真實資料的差異, 並且根據 loss 以及 Discriminator 所產生的分數優化其參數, 此處一樣是採用 Adam 優化器。

以上為每個 epoch 的訓練流程, 此次實驗總共進行了 160 萬個 epoch。

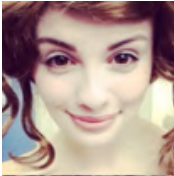
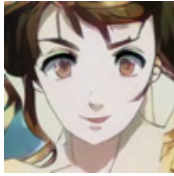
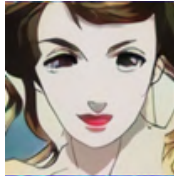
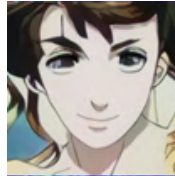
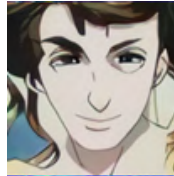



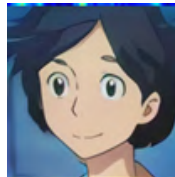



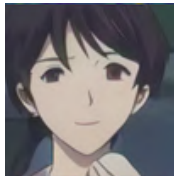
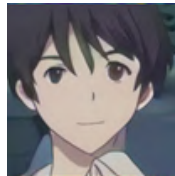
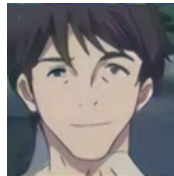

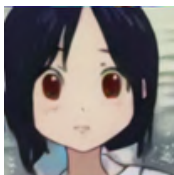
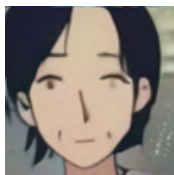
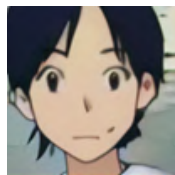
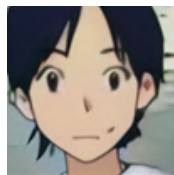
3-6 測試方法

使用一張非訓練資料、已標記類別(label 1)的人臉, 經由使用者指定希望轉換的漫畫臉類別(label 2), 將人臉、label 1、label 2 一起輸入 Generator。Generator 產生結果的漫畫臉之後保存結果圖片, 之後再以問卷調查做進一步的成效分析。

四、結果與成效分析

4-1 實驗結果

以下為各個類別的人臉及漫畫臉的轉換結果：

		年輕女性	中年女性	年輕男性	中年男性
年輕女性					
中年女性					
年輕男性					
中年男性					

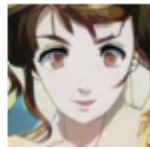
4-2 成效分析

本次實驗的成效分析採問卷調查方式，總共蒐集了 246 份有效問卷。主要目的為調查本次模型是否達到原本我們設立的三個目標。問卷內容分為三個部分：與 U-GAT-IT 的比較、同一類別轉換的相似程度、不同類別轉換的符合程度。

1. 與 U-GAT-IT 的比較

此部分的目的為比較我們的模型與 U-GAT-IT 原本的模型的相似程度。

此部分總共有六道題目，每道題目會出現一張人臉及轉換出來的漫畫臉，如下圖所示：



題幹為「請依照個人主觀感覺，選出相似的程度」，並在每道題目請答題者從1至10分選出一個分數(1表非常不相似，10表非常相似)。六道題目當中，有三道題目是我們的模型所產生的結果，另外三道則是原本的 UGATIT 模型產生的結果。結果顯示，我們的模型得到的平均分數為 6.3，UGATIT 模型得到的平均分數為 4.12。可以得知我們的模型相對原本 U-GAT-IT 的模型產生出更相似的結果。

2. 同一類別轉換的相似程度

此部分的目的為檢驗在人臉及目標漫畫臉類別相同的情況下(label 1 與 label 2 相同)，兩者的相似程度。

此部分一樣有六道題目，每道題目會出現一張人臉及轉換出來的漫畫臉，如下圖所示：



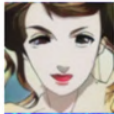
題幹為「請依照個人主觀感覺，選出相似的程度」，並在每道題目請答題者從1至10分選出一個分數(1表非常不相似，10表非常相似)。六道題目當中，「中年女性→中年女性」、「年輕男性→年輕男性」、「中年男性→中年男性」三種類別各有兩道題目。結果顯示，中年女性轉換的平均分數為 5.02，年輕男性轉換的平均分數為 5.11，中年男性轉換的平均分數為 4.11。

3. 不同類別轉換的符合程度

此部分的目的為檢驗在人臉及目標漫畫臉類別不同的情況下(label 1 與 label 2 不同)，漫畫臉符合目標類別(label 2)的程度。

此部分有八道題目，每道題目會出現一張漫畫臉及四個選項，如下圖所示：

1. *



☐ 年輕女性 (約25歲以下)

☐ 中年女性 (非年輕女性)

☐ 年輕男性 (約25歲以下)

☐ 中年男性 (非年輕男性)

題幹為「請依照個人主觀感覺，選出最符合該動漫臉的類別」，並在請答題者選擇最符合該動漫臉的類別。八道題目當中，不論人臉類別為何，轉換為年輕女性、中年女性、年輕男性、中年

男性的漫畫臉各有兩道題目。此處若答題者選擇的答案與轉換漫畫臉時指定的類別(label 2)相同，我們將該題定義為「正確」。結果顯示，漫畫臉類別為年輕女性的正確率為 0.92，中年女性的正確率為 0.71，年輕男性的正確率為 0.66，中年男性的正確率為 0.89。可以發現年輕女性的正確率最高。

五、結論與未來方向

5-1 結論

1. 漫畫臉的多樣性決定了生成結果的多樣性

因為漫畫臉的訓練資料數量較不平均，而在問卷調查的結果之中也發現，數量較少的類別，相似程度的分數就越低，因此我們推測，漫畫臉的多樣性會影響到模型的學習，多樣性越高，模型學習的就越完整，較不容易出現崩潰或是非常不相似的情形。

2. 加入不同的 label 之後，最大的改變在於五官，髮型及臉型並無太大改變

經過觀察，我們發現在選擇不同的目標漫畫臉類別(label 2)時，產生出來的漫畫臉大多都只有五官改變，髮型及臉型則沒有太大的變化。

5-2 未來方向

1. 加入更多訓練資料

訓練資料的多樣性會對結果的穩定性造成影響，因此希望可以加入更多樣化的訓練資料，尤其是漫畫臉的部分，希望可以讓模型對於每個類別都平均的學習。

2. 讓髮型以及臉型一起跟著改變

在人臉類別跟目標漫畫臉類別不相同的情況下，為了讓漫畫臉更符合目標類別，希望未來可以調整模型架構，讓模型更注重圖片中臉型或髮型的區塊，讓產生的漫畫臉的臉型及髮型也能跟著目標類別(label 2)去做改變。

六、參考資料

[1] U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation

<https://arxiv.org/abs/1907.10830>

[2] U-GAT-IT Official PyTorch Implementation

<https://github.com/znxlw/UGATIT-pytorch>

[3] dataset (human face with age & gender)

<https://github.com/JingchunCheng/All-Age-Faces-Dataset>

<https://susanqq.github.io/UTKFace/>

[4] Anime Planet

<https://www.anime-planet.com/>

[5] anime-face-detector

<https://github.com/ghgz2013/anime-face-detector>

[6] waifu2x-caffe

<https://github.com/litcggie/waifu2x-caffe>

多人合作之競賽獲獎與專題貢獻比例證明表

一、多人合作之競賽獲獎與專題皆應提出同一組員貢獻比例(總和為 100%)，並請指導教授簽名。		
二、申請人資料	姓名：葉晨	學號：0716081
三、專題題目/作品題目/獲獎競賽名稱及獎項： 條件性人臉與漫畫風格轉換		
四、同組同學姓名： 羅子涵		
五、個人貢獻百分比：55 %		
六、個人工作項目與貢獻說明： 1. 參考論文架構，利用 pyTorch 將深度學習網路 GAN 修改為 Conditional GAN。 2. 在多次實驗中調整網路參數，改善生成結果。		
申請人簽名：葉晨		日期：110.10.1
指導教授簽名：邱維辰		日期：

註：

1. 如參與不只一項多人合作之競賽獲獎與專題，且指導教授為同一人者，可以條列方式呈現於一張證明表中即可。
2. 若無指導教授，請同組同學於第四點空白處簽名佐證。