中華民國第 63 屆中小學科學展覽會

作品說明書

高中組　電腦與資訊學科

052503

GAN 圖像生成模型之畫質與效能強化

學校名稱：國立臺南第一高級中學

作者：	指導老師：
高二 郭育愷 高二 張丞漢 高二 邱宇晨	顏永進

關鍵詞：深度學習、生成對抗網路、圖像生成

摘要

近年來，名為生成對抗網路的非監督式學習方法蓬勃發展，透過使產生假資料的生成器與辨別資料真偽的辨別器互相學習，只要提供資料集便可學習其特徵而生成出能以假亂真的資料。本研究提出一套針對現有生成圖像的生成對抗網路模型的改良方法，透過進行實驗探討不同變因對生成品質與效率的影響，調整原有的優化器設置、卷積層參數、模型架構等，並以客觀指標評估實驗結果，證實經過本研究提出的方法改良有更好的效果。另外改進的方式應用在各種資料集訓練的模型及更高解析度的模型，數據表明也有不錯的成果。而本研究希望能提供更明確的模型改進方向給研究人員，並減少嘗試改良模型所花費的時間與能源成本，以此減少訓練龐大的模型所造成的環境影響。

壹、前言

一、研究動機

偶然看到名為 StyleGAN (Karras et al., 2018)的人臉生成模型，居然只需要提供一組人臉照片的資料集，就可以透過電腦進行學習，生成出一張不存在於世界上的人臉。但在查詢更多相關資料後，我們發現 StyleGAN 等等近期提出的模型架構很龐大，需要所費不貲的設備進行大量運算，耗費較多的金錢、時間以及能源成本，進而造成環境負擔甚至加劇氣候變遷。因此我們希望能提出一套改良現有模型的方法，能得到較好的生成品質，且有較高的訓練效率，可以用較普遍的設備進行訓練，同時節省時間和能源成本，並且以此達到聯合國提出的永續發展目標（SGDs）中提升能源使用效率的目標。

圖 1-1、StyleGAN 成果(Karras et al., 2018)

二、研究目的

本研究旨在提出一套強化 GAN 圖像生成模型之品質與效率的方法，分析不同變因對模型生成品質與效率的影響，藉以討論出改善模型的方向與流程。研究目標如下：

（一）探討影響 GAN 生成 64*64 解析度圖片品質、效率之因素並改良模型

（二）探討前項之方法用於較高解析度之圖像生成的品質與效果

（三）將前兩項之方法總結成提升 GAN 圖像生成模型之品質與效率的流程

三、文獻回顧

（一）圖像生成相關技術與模型

1.卷積神經網路（Convolutional Neural Network，CNN）(O'Shea & Nash, 2015)

CNN 於上世紀末期提出，是一種前饋神經網路，即內部訊號由輸入層單向傳播至輸出層。其靈感來自動物的視覺神經由不同的神經元，每個對應到一部份視覺區域稱作感受範圍，而相鄰的細胞會有相似且重疊的感受範圍，由此構成一張完整的影像。CNN 對於圖像處理有優秀表現，其構造通常包含卷積層、池化層、全連接層。

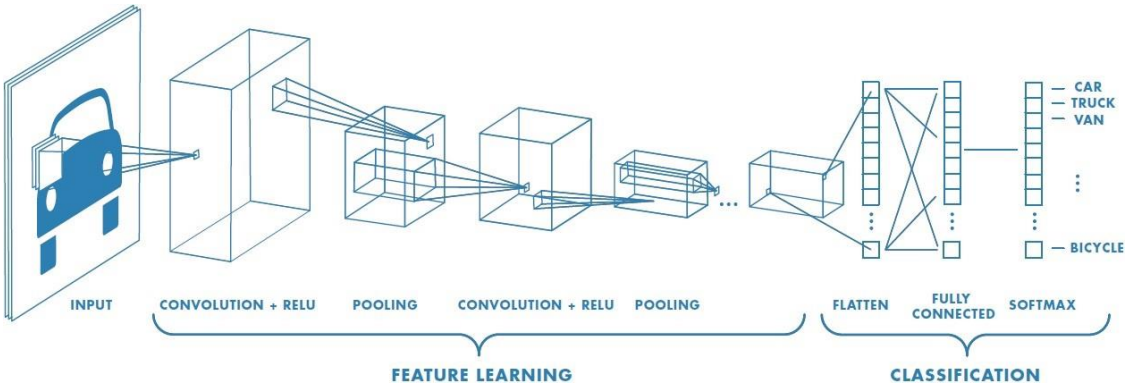


圖 1-2、CNN 架構(Prabhu, 2019)

(1)卷積層（Convolution Layer，Conv）

透過卷積核（Kernel）對輸入圖片上的區塊進行卷積運算，再移動一個步幅（Stride）到下個區域，並循環此操作以將輸入圖片投影至另一矩陣上，稱作特徵圖（Feature Map）。而卷積層輸出的特徵圖可能會有很多張，可以將輸出特徵圖的數量理解為通道數（Filter size，原指卷積核數量，但通道數較易於理解），例如 64*64 的 RGB 彩色圖片 ，就是有紅、綠、藍三個通道。而輸出特

徵圖的方式為，把輸入特徵圖的每個通道用不同的卷積核進行計算，然後將這些特徵圖的值進行相加，作為輸出的其中一個通道。也就是說若輸出有多個通道，就會有「輸入通道數*輸出通道數」個不同的卷積核。綜上所述，每個卷積層的參數數量大約為「輸入通道數*輸出通道數*卷積核長度*卷積核寬度」。

(2)池化層（Pooling Layer）

將輸入圖片以指定函式進行下採樣，以降低資料量，亦可以模糊邊緣。通常使用「最大池化（Max Pooling）」進行運算，即將固定大小的區塊，取出最大值投影至新的矩陣，但亦有平均池化與最小池化等。由於池化的過程會快速降低資料量，因此近年來有減少使用甚至捨棄的趨勢。

圖 1-3、最大池化圖示

(3)全連接層（Fully Connected Layer，FC）

將進行卷積與池化的資料攤平（Flatten）至一維，並利用全連接層，將輸入的每個數值乘上其權重，對應至新的全連接層，最後收斂至一個或多個最終輸出。

2.生成對抗網路（Generative Adversarial Network，GAN）(Goodfellow et al., 2014)

GAN 是一種非監督式學習的方法，由 Ian J. Goodfellow 等人於 2014 年提出。其核心概念為訓練辨別器（Discriminator）與生成器（Generator）兩個模型，辨別器之目標為分辨資料集中的真實資料與生成器生成的假資料；生成器的目標則是輸入一個隨機分布的雜訊 z（通常採用高斯分布），生成出以假亂真的圖片，使得辨別器能將其識別為真實資料。透過兩者互相學習、進步，學習結束後只取生成器來進行生成仿真資料的工作。

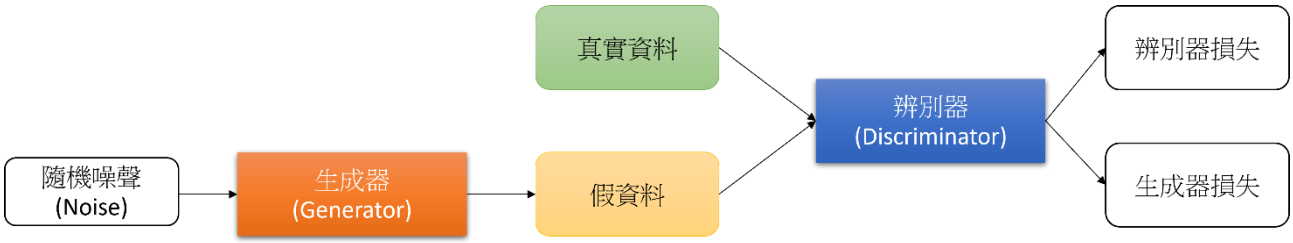


圖 1-4、GAN 架構

3.深度卷積生成對抗網路（Deep Convolutional GAN，DCGAN）(Radford et al., 2016)

將上述能夠生成仿真資料的 GAN 與善於處理圖像的 CNN 進行結合即是 DCGAN。其中辨別器之結構為 CNN，輸出經過 Sigmoid 函數轉換為 0~1 之間的數值，目標為將真實圖片辨識為盡可能接近 1、生成的假圖片辨識為盡可能接近 0；生成器結構則是將 CNN 之卷積層替換為轉置卷積層（Transposed Convolution Layer），與卷積層不同之處在於，轉置卷積核中的數值是與輸入矩陣的每一個數字相乘並映射在輸出圖片上，最後將重疊部分進行相加，而生成器的目標則是使其生成的圖片能夠被辨別器標示為 1。

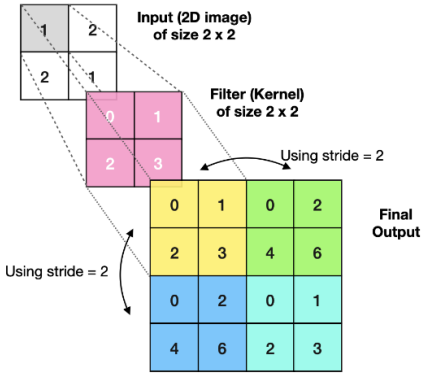


圖 1-5、轉置卷積圖解(Dobilas, 2022)

另外 DCGAN 還做出一些改變，如捨棄池化層、加入批次標準化層、生成器使用 ReLU 激活函數、輸出使用 tanh 激活函數、辨別器使用 LeakyReLU 激活函數。

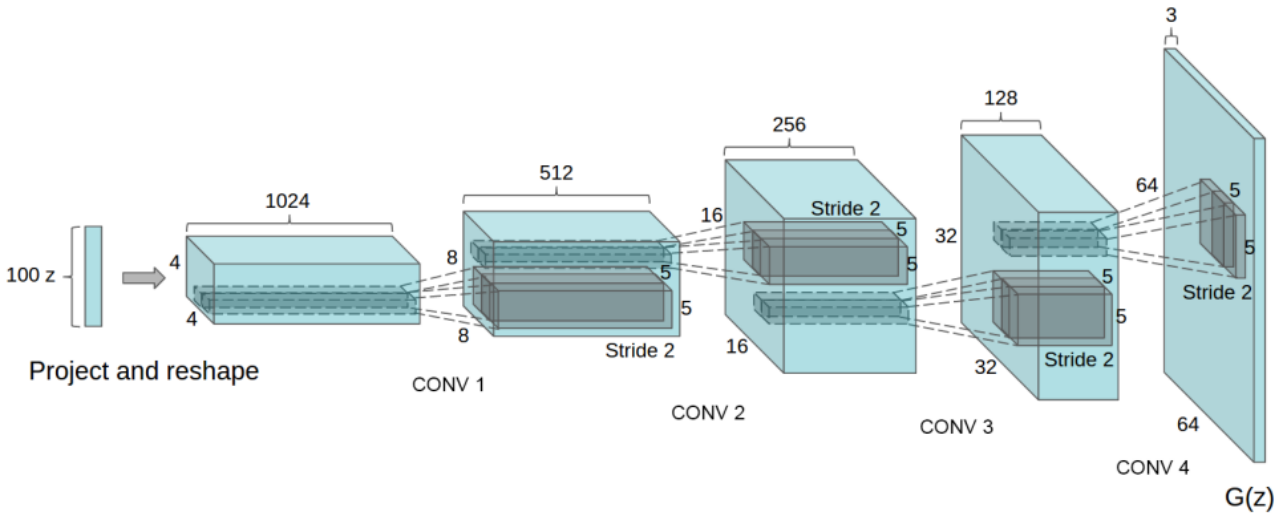


圖 1-6、DCGAN 之生成器架構(Radford et al., 2016)

4.殘差神經網路（Residual Neural Network，ResNet）(He et al., 2015)

ResNet 是一種前饋神經網路，由 Kaiming He 等人於 2015 年提出，透過跳躍連接（Skip Connections）與捷徑（Shortcuts）來越過某些層，相較於傳統的模型學習 x→F(x)的方法，ResNet 學習 x→x+F(x)，也就是在 F(x)沒有學習到任何特徵時，可以保留原始的 x 作為輸出，利用該技巧可以解決傳統模型在層數過多、模型過深時所產生的退化（Degradation）問題。下圖為 ResNet 的基本單位：殘差單元。

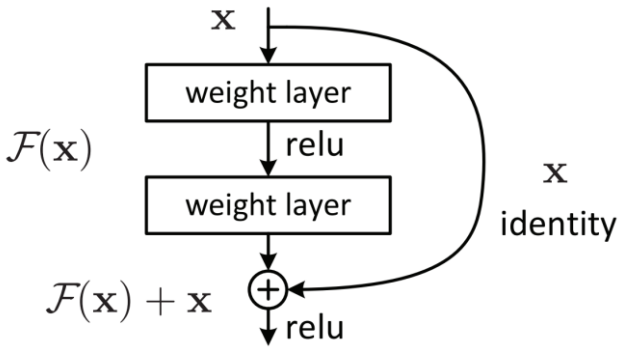


圖 1-7、殘差單元圖示(He et al., 2015)

（二）評估模型之方法

模型生成之效果大致有解析度、圖像品質、效率等。其中解析度是指圖片的長寬各由多少像素組成，由於不同的解析度涉及到不同的模型架構（模型 output size、層數、參數量等）、訓練時長、圖片資訊含量，因此不適合將不同解析度之圖片進行比較，我們將每個解析度分開討論。本研究由以下角度來分析模型成效：

1.品質

圖像品質是指在同樣解析度下，圖像的清晰度、雜訊、與真實人臉的相似度等，需要注意的是，解析度並不能直接衡量品質的好壞，例如下圖 1-2 中為同一張圖片，左圖為 64*64 解析度，右圖為 128*128 解析度但經過模糊處理，儘管 128*128 有更多的像素格保存圖片資料，但整體品質仍肉眼可見的較右圖差，因此 64*64 的圖像可能比 128*128 的圖像品質更好。

圖 1-8、CelebA-HQ 資料集中的圖片（左）64*64 解析度（右）128*128 解析度模糊處理

我們使用 Fréchet inception distance（FID）(Heusel et al., 2018)作為主要的客觀評價標準。FID 是一種用以評估 GAN 生成圖像的指標，在大多數的實際應用中，將真實圖片以及由欲評估模型生成的仿真圖片輸入在 ImageNet 資料集上訓練的 InceptionV3 模型，取出其中的特徵向量，並進行數學運算以求得兩者分布之距離，越小的距離代表模型生成之圖片與真實資料分布越接近，生成品質越好。該指標為目前最廣泛被用以評估圖像生成模型的指標，在近期重要的圖像生成模型論文接有採用，如 StyleGAN (Karras et al., 2018)、BigGAN (Brock et al., 2018)等，被證實比先前提出的其他指標更能反映人類視覺上的感受。

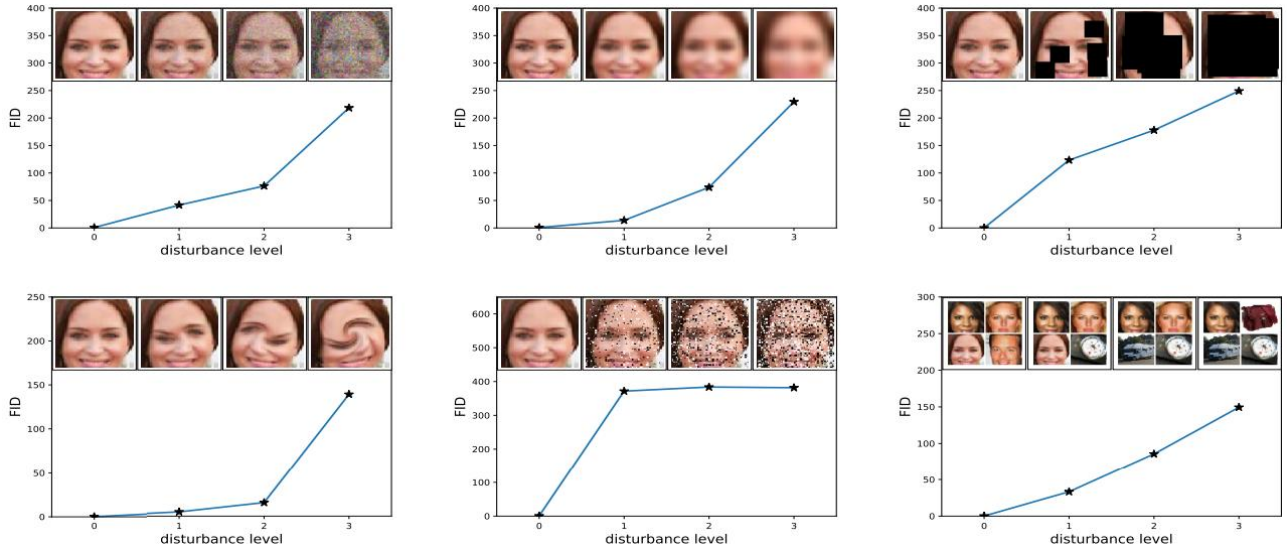


圖 1-9、不同處理方式下的 FID（左上）高斯噪聲（中上）高斯模糊（右上）插入黑色塊（左下）旋轉（中下）椒鹽噪聲（右下）資料集受汙染(Heusel et al., 2018)

2.效率

最後模型效率的部分，可以分做許多部份進行討論，如訓練效率、圖像生成效率、模型檔案大小。其中我們較著重於訓練效率，也就是訓練所需要花費的時間，亦決定能源的消耗量，在同樣硬體下訓練效率會受到許多因素如模型架構、深度、參數量等影響；圖像生成效率因為相較於訓練過程來說較短，通常不特別討論；模型檔案大小則受模型參數量影響，決定了模型是否容易進行保存、傳輸，但由於現今通訊與儲存硬體的發展，也較少特別進行討論。

因此對於每一次訓練，我們使用相同的硬體，並記錄訓練所花費之時間、參數量等資訊，並就模型架構、深度進行討論，以此評估模型訓練之效率。

貳、研究設備及器材

一、硬體設備：AI 運算伺服器

CPU：Intel(R) Xeon(R) Gold 5118 CPU @ 2.30GHz

RAM：128GB

GPU：Quadro RTX 5000 16GB（伺服器有兩個該型號 GPU，但研究過程訓練模型若未特別註明皆是只使用一個 GPU）

二、軟體環境

（一）Python 3.9.12

（二）環境管理工具：Conda 4.12.0

（三）機器學習套件：TensorFlow 2.9.1

（四）其他套件：

NumPy 1.22.4

Pillow 9.1.1

Matplotlib 3.5.2

參、研究過程或方法

一、研究架構

在開始研究前，我們擬定了完整的研究架構如下圖 3-1，確定研究主題與目標後，查詢與研究主題相關的文獻，而後對可能影響結果的變因設計實驗，並在進行實驗後進行分析、評估、比較，最後進行討論並做成結論。

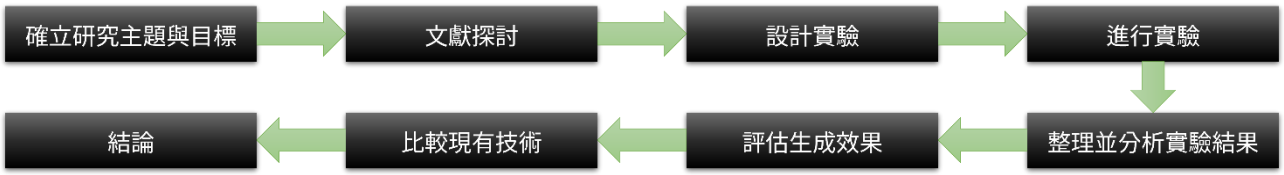


圖 3-1、研究架構圖

二、實驗流程與設計

為了順暢進行實驗，我們先將實驗流程確定，在蒐集要使用的資料集後，將其圖片縮放（Resize）至不同解析度，並且先利用解析度較低之圖片，訓練出可以生成低解析度圖片的模型，並且測試與討論不同的變因對其結果的影響，保留好的改變並繼續進行下一項實驗。然後利用高解析度的圖片與低解析度的模型，訓練出生成高解析度圖片之模型，同樣測試與討論不同變因造成的影響，並按此進行多項實驗，最後再將這些不同解析度的模型進行評估並討論。下圖 3-2 為我們的實驗流程圖。

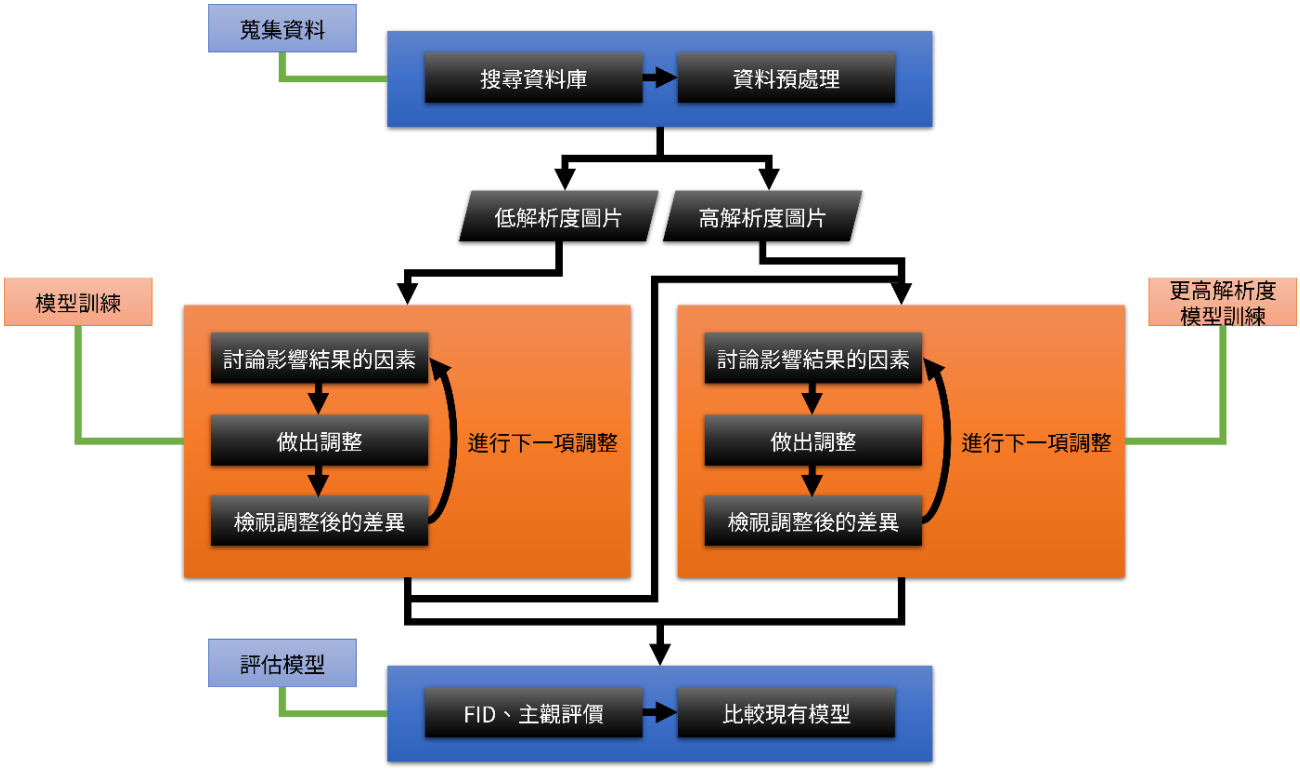


圖 3-2、實驗流程圖

（一）蒐集資料集與預處理

我們訓練所使用的資料集為 CelebA-HQ (Karras et al., 2017)，該資料集是 CelebA 資料集的高解析度版本，有 3 萬張 1024*1024 解析度的人臉圖片，在 2017 年與 PGGAN 一起提出。為了訓練出生成不同解析度圖片的模型，我們將該資料集之圖片，進行縮放至 64*64、128*128、256*256、512*512 這幾種解析度。

除此之外，為了驗證模型的效能，我們還使用 Stanford Cars 資料集(Krause et al., 2013)與 UT Zappos50K 資料集(Yu & Grauman, 2014, 2017)。Stanford Cars 資料集包含 16185 張汽車圖像；UT Zappos50K 資料集包含 50025 張鞋子圖像。

（二）模型訓練

我們採用架構簡單的 DCGAN 為基礎進行改進，利用 TensorFlow 套件實作出 DCGAN 後，設定批次（batch，可理解為一次對一疊資料同時進行運算）大小為 32，進行 5 萬步（step，利用一個批次運算、修改一次模型內參數）學習，每一步會先從資料集中隨機取 32 張真實圖片對辨別器進行訓練，再讓生成器生成 32 張假圖片對辨別器進行訓練，然後再將生成器與辨別器組合，使生成器的輸出進入辨別器，此時辨別器不會被訓練（即 trainable 參數為 False，模型內參數不做修改），接著對整個模型輸入 32 組雜訊，每組由 100 個介於 0~1 之間的浮點數構成，藉此訓練生成器，使生成器能生成出被辨識為真實圖片之假圖片，盡可能混淆辨別器。

而在訓練過程中，每步會記錄辨別器與生成器模型的損失（使用二元交叉熵 binary cross entropy），而每隔 100 步會生成一張大圖片，其中包含 25 張當下生成器生成的圖片，另外每 1000 步會保存一次辨別器與生成器模型。

（三）評估模型

利用訓練時每隔 1000 步保存的模型，我們對第 0 步的模型（完全未經訓練）到第 50000 步的模型都進行 FID 分數計算，計算方式為從原始資料集取 10000 張圖片以及讓生成器生成 1000 張圖片，並利用 InceptionV3 預訓練模型進行特徵提取，最後進行分數計算，然後將結果以圖表進行呈現，並對其中最好的結果進行記錄。

肆、研究結果

一、64*64 解析度圖像生成模型

由於影響模型效能的因素非常多，為了能夠一步步的探討影響模型效能的因素，我們將修改不同變因的實驗做以下安排：

實驗 1-1：先從決定模型如何調整內部參數以學習特徵的優化器開始。

實驗 1-2：在不修改模型架構的情況下，調整模型卷積層、轉置卷積層的參數。

實驗 1-3：利用調整好的優化器與卷積層參數，對模型架構進行一步的修改。

實驗 1-4：將得到的模型進行不同資料的生成以進一步討論模型的生成效果與限制。

（一）優化器設置

實驗 1-1a：改變學習率

學習率代表了模型修改參數的時候，一次所需要修正的量，直接影響到了模型的學習速度與穩定度，由於 GAN 的訓練過程是辨別器與生成器互相學習的動態過程，因此選用錯誤的學習率可能導致訓練失敗。在該實驗我們使用 DCGAN 論文中使用的 Adam 優化器，測試了以下不同的學習率設置。

表 4-1-1、實驗 1-1a 各組別說明與結果

組別	辨別器學習率	生成器學習率	最小 FID 值	訓練時長
1	0.0002	0.0002	302.21	10001 秒
2	0.0005	0.0002	66.30	10119 秒
3	0.001	0.0002	136.54	10162 秒
4	0.001	0.0004	100.09	11964 秒

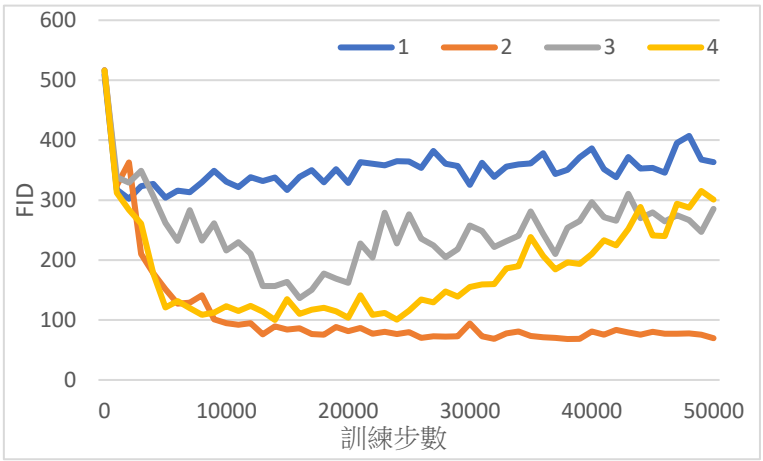


圖 4-1-1、實驗 1-1a 各組 FID 對訓練步數變化情形

可以發現整體而言組別 2 的表現最好，除了可以正常收斂外，整體的穩定性也還不錯，而組別 1 是無法收斂，組別 3、4 則是可以收斂但整體穩定性不佳。另外由於不同組別訓練過程的圖像生成效果變化、辨別器與生成器損失變化有顯著差異，因此下圖 4-1-2 展示了對於相同的輸入雜訊，隨訓練步數的生成效果變化，而下圖 4-1-3 為各組的辨別器與生成器損失變化（每 100 步之損失取平均）。

步數	1000	2000	3000	4000	5000	10000	20000	30000	40000	50000
組別 1										
組別 2										
組別 3										
組別 4										

圖 4-1-2、實驗 1-1a 各組以相同雜訊生成之圖片對訓練步數變化情形

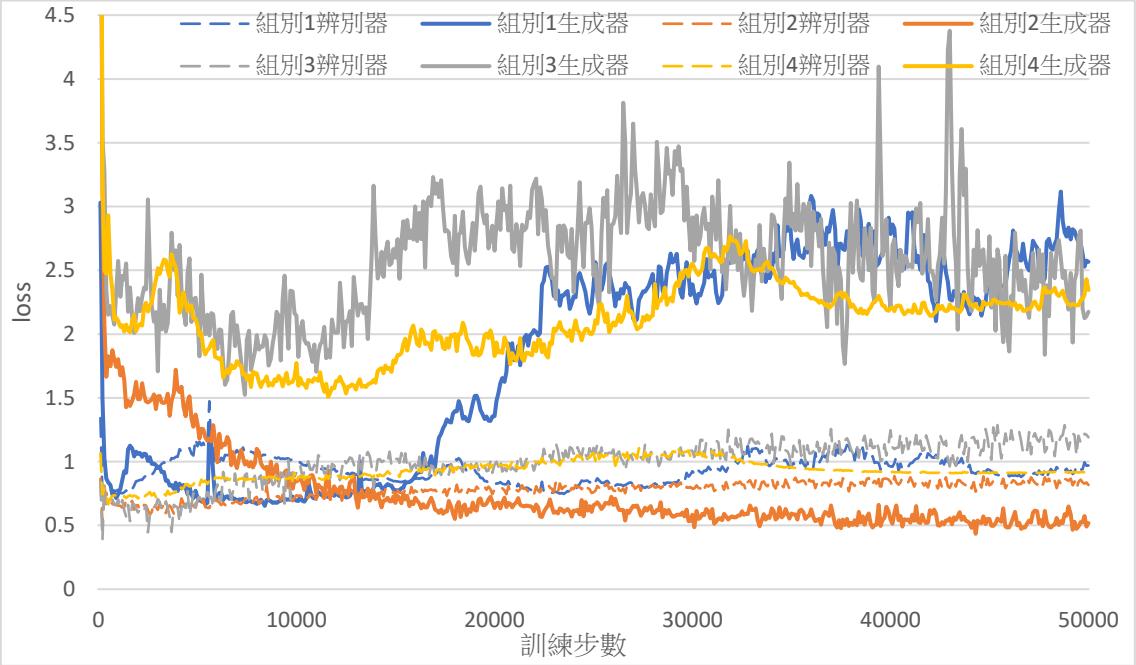


圖 4-1-3、實驗 1-1a 各組辨別器與生成器損失變化

在圖 4-1-2 中可以看到，組別 2 相較之下效果較為良好且穩定，而組別 3、4 過程中也一度收斂到較好的結果，但穩定性與最終效果並不好；而在圖 4-1-3 則可以看到組別 1 生成器很快收斂到低點，但後期則變高，配合圖 4-1-2 可發現應是生成器難以更好的欺騙辨別器，開始生成雜訊並造成訓練失敗，而組別 2 生成器一開始收斂較慢，但後期穩定的下降且辨別器亦十分穩定，組別 3、4 則是生成器自始至終皆無收斂至較低損失。以上結果將在伍、討論部分進一步說明，而鑒於本實驗結果，後續實驗採用組別 2 辨別器與生成器學習率分別為 0.0005、0.0002 的設置。

實驗 1-1b：不同優化器

一個模型的最終目標是透過調整內部參數在損失函數找到局部最佳解，而優化器則是調整參數尋找最佳解的方式，因此優化器的不同會影響收斂的過程，進而影響最後模型的成效。我們採用了 3 種常用的優化器來進行比較，分別有：每次調整所有參數的 SGD、加入自適應學習率機制的 RMSprop、以及兼具自適應學習率和動量機制的 Adam。透過這三個較常見的優化器來探討優化器對模型的影響。

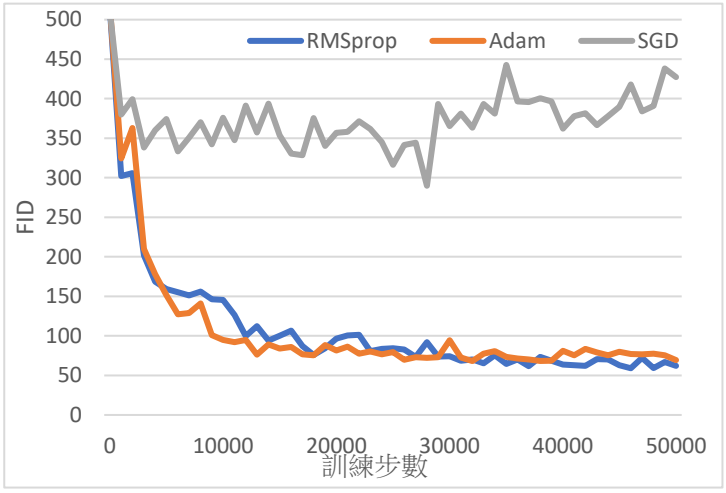


圖 4-1-4、實驗 1-1b 各組 FID 對訓練步數變化情形

表 4-1-2、實驗 1-1b 各組別說明與結果

優化器	最小 FID 值	訓練時長
Adam	68.30	10119 秒
RMSprop	59.09	10150 秒
SGD	289.90	9901 秒

在圖 4-1-4 可以看到使用 SGD 的模型難以收斂，而使用 Adam 與 RMSprop 的模型有較好的收斂結果，另外雖然 Adam 一開始收斂速度較快，但後期沒有繼續收斂的跡象，而 RMSprop 則是持續緩慢的收斂，最終達到較 Adam 良好的結果。而因為不同組別訓練過程的圖像生成效果有不同的變化，下圖展示了該過程。

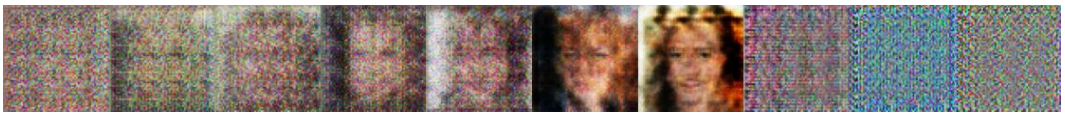
步數	1000	2000	3000	4000	5000	10000	20000	30000	40000	50000
Adam										
RMSprop										
SGD										

圖 4-1-5、實驗 1-1b 各組以相同雜訊生成之圖片對訓練步數變化情形

上圖可以看到 RMSprop 的變化較小，代表訓練過程的變動較穩定，也更容易收斂至更好結果。因該實驗結果，後續實驗都採用 RMSprop 優化器。

（二）卷積層參數

實驗 1-2a：改變 filter 數量

在文獻回顧中卷積層的部分，我們提到可將 filter 理解為通道數的概念，而在網路中雖然擁有較多的通道代表同樣解析度可以儲存更多資料，但過多的通道數除了容易使一部份通道並沒有記錄到資訊而浪費掉，也會使得參數量與訓練難度增加；反之，過少的通道數則是讓資訊量不足以生成出品質良好的圖片。生成器在原始 DCGAN 中 4*4 的特徵圖有 1024 通道，而後每提升一次解析度通道數會較前一層減半，最後從 128 通道的 32*32 特徵圖直接轉為 RGB 圖片；辨別器在原始 DCGAN 中則是直接將 RGB 圖片轉換為 64 通道的 32*32 特徵圖，而後每降低一次解析度，通道數就變為上一層的兩倍，直至 512 通道的 4*4 特徵圖，後面直接 flatten 並接上 dense 層。因此我們將原始 DCGAN 中辨別器與生成器的通道數量進行調整（除了輸出保持 3 通道外，將所有通道同乘表格中之比值，如下表組別 2 之生成器 4*4 特徵圖有 512 通道，以此類推），以下是實驗結果。

表 4-2-1、實驗 1-2a 各組別說明與結果

組別	辨別器 filter 與原始設置之比值	生成器 filter 與原始設置之比值	辨別器參數量	生成器參數量	最小 FID 值	訓練時長
1	1	1	431.8 萬	1886 萬	59.09	10150 秒
2	1	1/2	431.8 萬	513.0 萬	49.44	8548 秒
3	1	1/4	431.8 萬	149.0 萬	91.77	8711 秒
4	1/2	1/2	108.4 萬	513.0 萬	67.49	6776 秒

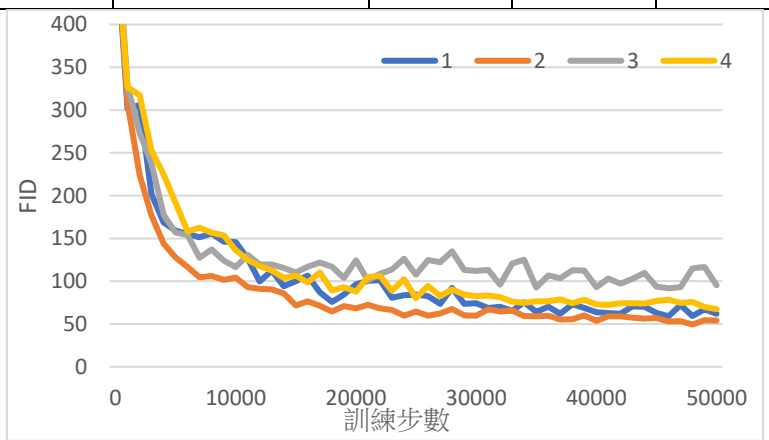


圖 4-2-1、實驗 1-2a 各組 FID 對訓練步數變化情形

由實驗數據可以看出，辨別器 filter 數量與 DCGAN 一樣、生成器 filter 減少為原本的一半可以在 50000 步內達到更快的收斂速度與更好的結果，且可顯著降低訓練時間。由於實驗結果，後續實驗都是採用組別 2 減半生成器 filter 的做法。

實驗 1-2b：改變 kernel size

卷積核大小相當於模型中單個神經元的感知範圍，過大的卷積核儘管可以涵蓋更大的範圍，但可能造成網路難以學習卷積核所需要匹配的特徵；過小的卷積核同樣會使感知範圍之間重疊過少而無法生成或辨識圖片。因此我們將原始 DCGAN 中辨別器與生成器的 kernel size 進行調整（組別 1 為原始設置），以下是實驗結果。

表 4-2-2、實驗 1-2b 各組別說明與結果

組別	辨別器 kernel size	生成器 kernel size	辨別器參數量	生成器參數量	最小 FID 值	訓練時長
1	5x5	5x5	431.8 萬	513.0 萬	49.44	8548 秒
2	5x5	4x4	431.8 萬	358.0 萬	52.95	7830 秒
3	5x5	3x3	431.8 萬	237.4 萬	47.27	8367 秒
4	5x5	2x2	431.8 萬	151.3 萬	91.42	7538 秒
5	4x4	3x3	276.8 萬	237.4 萬	44.11	7404 秒
6	3x3	3x3	157.0 萬	237.4 萬	60.64	6982 秒

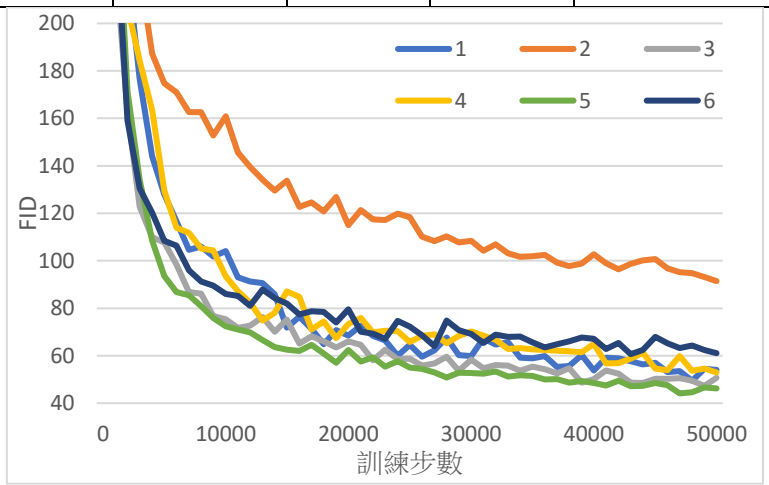


圖 4-2-2、實驗 1-2b 各組 FID 對訓練步數變化情形

由實驗數據可看出對於辨別器與生成器分別採用 4x4 與 3x3 的卷積核大小有最好的收斂速度與結果，且可稍微降低訓練時間。因此後續實驗皆採用這樣的設置。

（三）改變模型架構

實驗 1-3a：生成器網路強化

觀察目前的網路架構，可以發現對於網路較前面的層，有較多參數、通道來計算出足夠的資訊量，如 16*16*128 的特徵圖到 32*32*64 的特徵圖，卷積層有 8192（輸入通道數 128 乘上輸出通道數 64）個不同卷積核，但最後直接由 32*32*64 的特徵圖提升至 64*64*3 的 RGB 圖片，只會有 192 個不同的卷積核，也就是說處理 32*32 到 64*64 解析度資料的參數量非常少，因此該部分可能是影響模型效能的一個關鍵。我們把生成器最後輸出為 64*64 的層數增加，讓生成器在最後輸出前能加入更多資訊，另外我們將卷積核大小為 1x1、步幅為 1 的卷積層應用在網路中，此層的用途僅是對上一層的輸出進行線性組合，且為了得到更複雜的組合，該層的通道數量通常較多。不過在增加層數的同時，可能會造成網路較難以訓練，因此我們也嘗試將殘差的概念應用於此。下圖即是我們的生成器架構，其中若沒有特別註明，卷積層（conv）與轉置卷積層（convT）後皆有批次標準化層（bn）以及使用 ReLU 激活函數（輸出層直接使用 tanh，不使用 ReLU）。

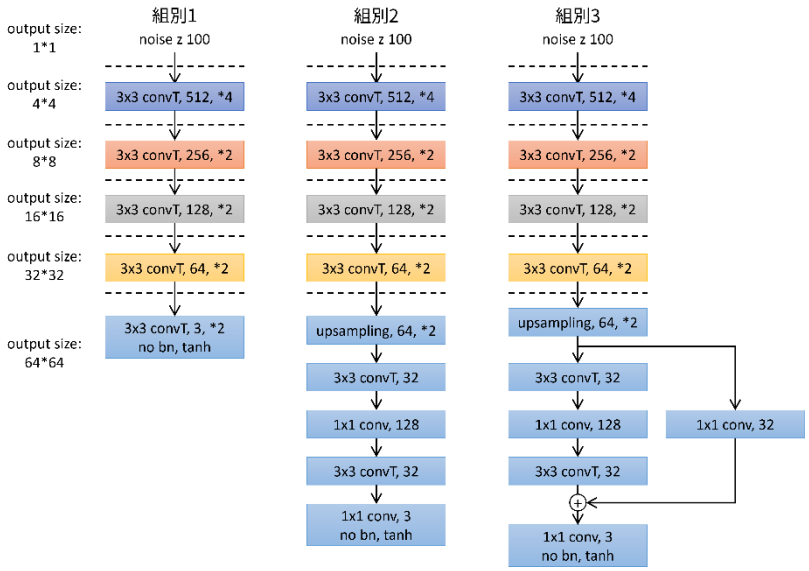


圖 4-3-1、實驗 1-3a 各組別模型架構圖示

而為了方便表達，後續類似的模型架構會使用下表 4-3-1 的形式表達，其中若未特別註明皆是卷積層（加註 T 者為轉置卷積層）且使用 bn 與 ReLU，另外殘差單元部分使用中括號表示，而跳躍連接的部分，若輸出殘差單元之形狀與輸入形狀不同，統一使用通道數與殘差單元輸出相同的 1x1 conv 進行線性組合。

表 4-3-1、實驗 1-3a 各組別生成器架構說明

組別	1	2	3
output size:4	3x3T, 512, *4	3x3T, 512, *4	3x3T, 512, *4
output size:8	3x3T, 256, *2	3x3T, 256, *2	3x3T, 256, *2
output size:16	3x3T, 128, *2	3x3T, 128, *2	3x3T, 128, *2
output size:32	3x3T, 64, *2	3x3T, 64, *2	3x3T, 64, *2
output size:64	3x3T, 3, *2, no bn, tanh	upsampling, 64, *2 3x3T, 32 1x1, 128 3x3T, 32 1x1, 3, no bn, tanh	upsampling, 64, *2 [3x3T, 32 1x1, 128 3x3T, 32] 1x1, 3, no bn, tanh
生成器參數量	237.4 萬	243.3 萬	243.5 萬

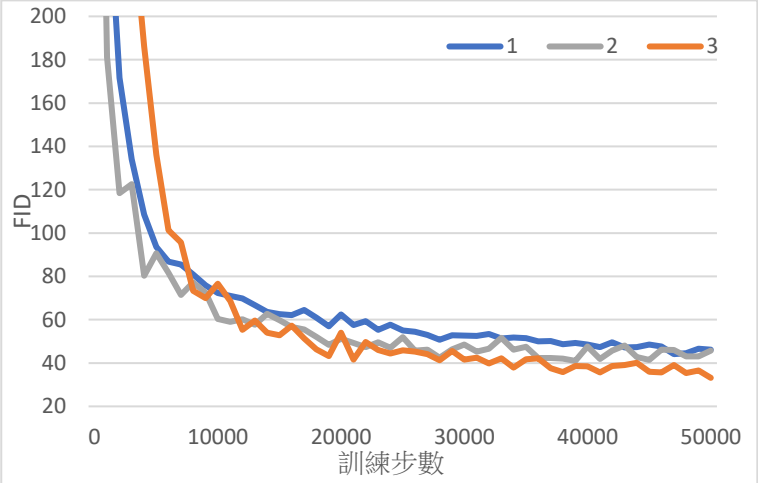


圖 4-3-2、實驗 1-3a 各組 FID 對訓練步數變化情形

表 4-3-2、實驗 1-3a 各組別結果

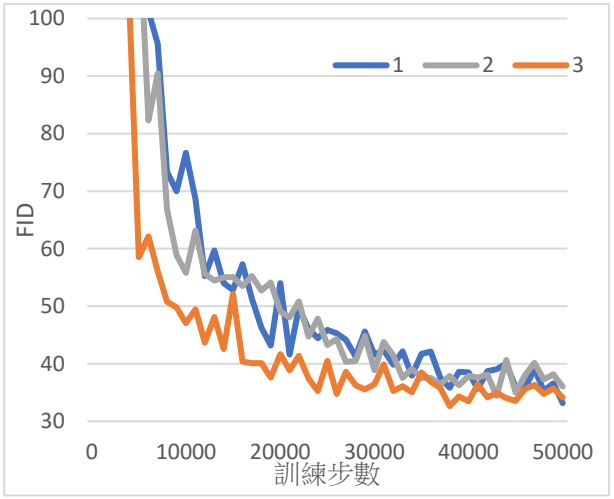
組別	最小 FID 值	訓練時長
1	44.11	7404 秒
2	41.00	7848 秒
3	33.16	8094 秒

由實驗數據可以看出組別 2 對 64*64 特徵圖進行更多處理確實對模型有幫助，提供更快的收斂速度與更好的結果，而組別 3 將最後 64*64 特徵圖處理的層應用殘差單元，雖然一開始收斂速度較慢一些，但最後結果有非常明顯的進步且整體較組別 2 穩定。基於該結果，後續實驗採用組別 3 結合殘差的更多的尾端反卷積層。

實驗 1-3b：辨別器網路強化與優化

由上個實驗中的經驗可以發現，在辨別器網路中，從 64*64*3 原始圖片轉換到 32*32*64 特徵圖，處理數據的參數量較少，因此針對該部分進行加強。但由於我們還是採用 DCGAN 的訓練模式，該方法如果辨別器網路過於強大，使得生成器無法對抗的情況下，會造成訓練失敗，因此在強化辨別器的部分，只簡單地加上了一層通道數為 32、步幅為 1 的卷積層以處理輸入數據。不過經實驗 1-3a 的生成器網

路改進，如今的生成器網路有更多的內部參數，在最一開始需要略多一點的時間學習，而現在增加辨別器網路的強度，可能導致訓練一開始辨別器過強而生成器難以學習的問題。我們採用一種常見的優化方式，於最初辨別器模型學習的真實資料標籤 1 加入一些噪音（減去高斯分布 0~0.2 間的數，使其成為 0.8~1 間的高斯分布），而後逐漸減小噪音，於 2000 步後恢復正常，以此削弱一開始辨別器的性能，使生成器能更好的與之對抗，等到生成器學習一定的特徵後，再恢復辨別器的強度。以下為實驗結果。（註：組別 1 辨別器參數量為 276.8 萬，組別 2、3 為 280.0 萬）



↑圖 4-3-3、實驗 1-3b 各組 FID 對訓練步數變化情形

表 4-3-3、實驗 1-3b 各組別說明與結果

組別	說明	最小 FID	訓練時長
1	對照組	33.16	8094 秒
2	輸入後加步幅 1 的 3x3 conv	34.39	8130 秒
3	同第 2 組但一開始加入噪音	32.65	8173 秒

可以發現組別 1 與 2 訓練過程與結果的差異不大，但組別 3 的收斂速度快上許多，另不同組別的辨別器與生成器損失隨訓練過程變化有顯著差異，如下圖。

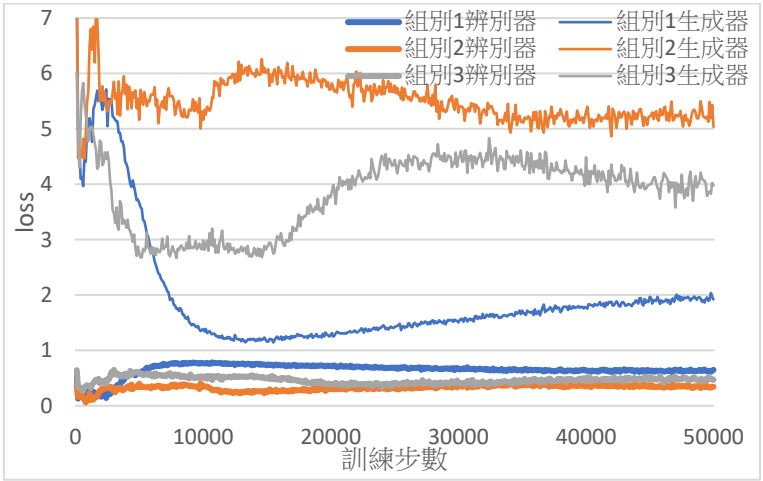


圖 4-3-4、實驗 1-3b 各組辨別器與生成器損失變化

由上圖可以發現，組別 2、3 由於辨別器增強，因此生成器損失變大，而組別 3 相較組別 2 生成器更能收斂。鑒於實驗結果，後續實驗皆採組別 3 之模式。

實驗 1-3c：加入條件批次標準化（Conditional Batch Normalization）機制

原本的 DCGAN 模型中，除了生成器的最後一個與辨別器的第一個卷積層外，其餘辨別器與生成器中的每個卷積層後皆有批次標準化層（Batch Normalization），目的是將卷積後的資料乘上 gamma 進行縮放，並加上 beta 進行平移，其中 gamma 與 beta 為模型內參數，透過學習每次輸入的批次分布情況，學習如何縮放與平移資料，使得其分布不會過度極端，可以避免過擬合（overfitting）與模式坍塌（mode collapse，指 GAN 生成器對於所有輸入雜訊皆生成相同或極度相似內容，雖然可能可以騙過辨別器並取得好分數，但仍屬訓練失敗）的狀況。而條件批次標準化則是透過輸入不同條件來運算進行縮放與平移的 gamma 與 beta，可以被用於生成不同種物品（即生成條件）的 GAN，而在我們的模型中，由於希望一開始的雜訊能夠代表人臉不同的特徵，因此也可以將雜訊輸入條件批次標準化層。（註：加入條件批次標準化層後，生成器參數量由 243.5 萬增加至 283.3 萬）

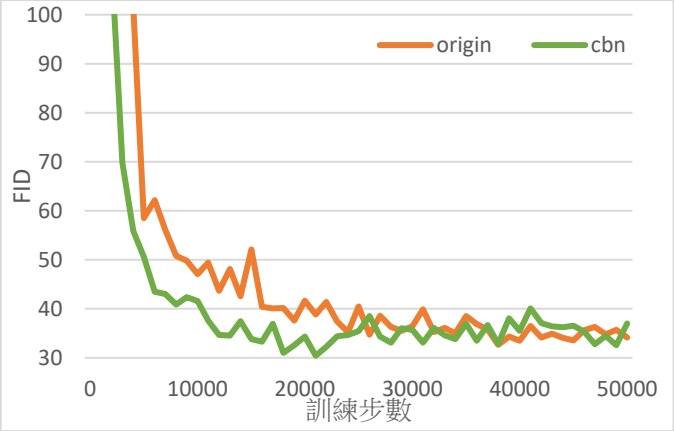


圖 4-3-5、實驗 1-3c 各組 FID 對訓練步數變化情形

表 4-3-4、實驗 1-3c 各組別說明與結果

組別	最小 FID 值	訓練時長
原始(origin)	32.65	8173 秒
加入條件批次標準化(cbn)	30.43	10017 秒

由實驗數據可以看出，加入條件批次標準化後，前期的收斂速度較原本快上不少，且雖然訓練 50000 步的時間明顯變長，但原始模型的 FID 最小值是在第 38000 步出現，但加入 cbn 在第 18000 步就達到比其更低的 FID 分數，並在 21000 步取得最小 FID，所以考慮訓練步數，cbn 的效率還是較高，因此後續實驗皆採用。

（四）不同資料集的生成效果

實驗 1-4：

為了瞭解我們的模型對不同資料集的生成效果或者限制，因此在本實驗測試不同資料集，其中包含 UT Zappos50K，該資料集擁有稍微多一點的圖片，且背景為

純白、皆為相同角度的鞋子，構圖較單純，但鞋子的紋理較人臉多樣且複雜；另外還有 Stanford cars 資料集，其圖片雖然主體都是汽車，但由於不同的角度以及較複雜的背景，加上較小的圖片量，使得該資料集生成的難度較大。以下為實驗結果。

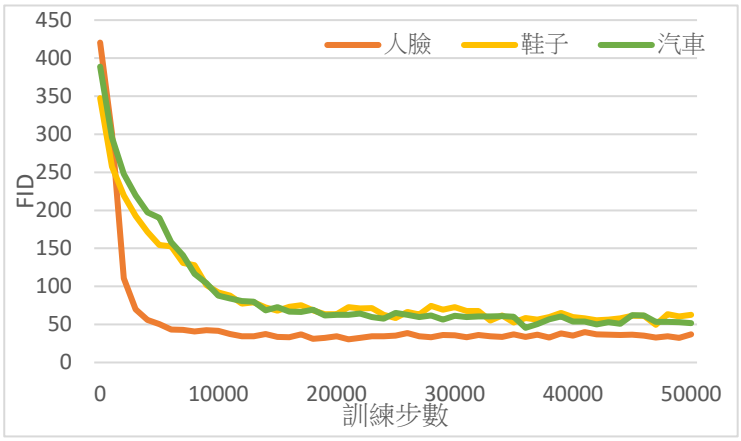


表 4-4-1、實驗 1-4 各組別說明與結果

組別	最小 FID 值	訓練時長
人臉	30.43	10017 秒
鞋子	49.51	8745 秒
汽車	45.61	9746 秒

圖 4-4-1、實驗 1-4 各組 FID 對訓練步數變化情形

由以上實驗結果可以看出，生成鞋子與汽車的收斂速度皆較慢一些，且可能由於上述的資料集圖片特點，造成較難訓練到較好結果，但經不同資料集訓練的 FID 並不適合拿來直接做比較，該結果僅用以評估收斂過程。以下為不同資料集的生成結果。（註：採用紀錄的模型中 FID 最低的模型）

圖 4-4-2、CelebA-HQ 資料集與其生成結果（左）資料集（右）生成結果

圖 4-4-3、UT Zappos50K 資料集與其生成結果（左）資料集（右）生成結果

圖 4-4-4、Stanford Cars 資料集與其生成結果（左）資料集（右）生成結果

由以上實驗結果可以發現在人臉的生成中，模型生成的圖片大致上看起來沒問題，但仔細看即可發現瑕疵；而生成的鞋子圖片比起原本的資料集，較不會有太過複雜的圖樣，偶爾會出現不自然的色塊；至於生成汽車的圖片，可以發現模型有成功生成出汽車的部分樣貌，但多數圖片難以將這些小細節完美的組合，且輪廓也較粗糙，因此整體效果雖然堪用但並不特別理想。但由以上的結果已經可以看出，我們的模型在其他資料也可以正常訓練。

二、更高解析度圖像生成模型

為了探討影響更高解析度模型的不同因素，我們將進行以下實驗：

實驗 2-1：就 64*64 解析度的不同模型架構，將其提升至 128*128 解析度以討論不同架構對於較高解析度圖像的生成能力，並且進行不同資料集的效果檢視。

實驗 2-2：在研究的最後，採用先前得出的模型架構與拓展至較高解析度之方法，嘗試生成 256*256 或更高解析度圖片，並探討其生成效果或限制。

（一）提升至 128*128 解析度

實驗 2-1a：不同低解析度模型架構提升至高解析度

為了確認上一部分生成 64*64 圖像之不同架構生成高解析度的表現，因此將實驗 1-1、1-2、1-3 最終效果最好的模型皆增加一層使輸出提升至 128*128。其中架構大致與 DCGAN 相同的 1-1、1-2，直接利用原本建構模型的邏輯，生成器分別直接在輸出前增加一層 filter 為 64、32 的轉置卷積層，使其多提升一次解析度，辨別器則是在輸入後加一層 filter 為 32 的卷積層，使其多一次的降低解析度。而將實驗 1-3 的模型擴展到 128*128，生成器的部分，為了延續 1-3 實驗中對高解析度特徵

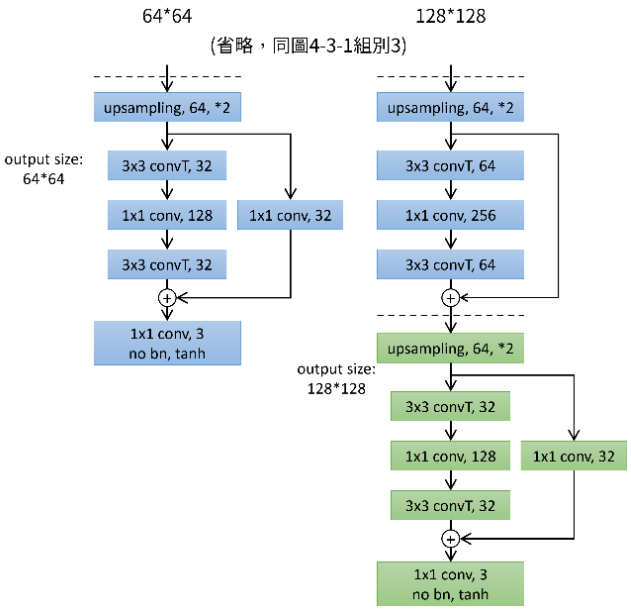
圖進行更多處理的精神，我們將殘差單元用以提升 64*64 特徵圖至 128*128，且為了避免過少的資訊量，採用通道數為 32，並將前一層通道數提升至 64，而因為前一層的殘差單元輸入與輸出皆為 64 通道，因此跳躍連接不使用 1x1 conv；而辨別器部分，同樣需要增加一層卷積層來降低特徵圖解析度，同樣為了避免向後傳遞的資訊量不夠，因此採用 1-3 中輸入後不降低解析度、通道數 32 的卷積層，而後重複兩次通道數為 64 的卷積層。以下為本實驗的模型架構表與結果表。

表 4-5-1、實驗 2-1a 各組別生成器架構說明

組別	1(擴展 1-1 設計)	2(擴展 1-2 設計)	3(擴展 1-3 設計)
output size:4	5x5T, 1024, *4	3x3T, 512, *4	3x3T, 512, *4
output size:8	5x5T, 512, *2	3x3T, 256, *2	3x3T, 256, *2
output size:16	5x5T, 256, *2	3x3T, 128, *2	3x3T, 128, *2
output size:32	5x5T, 128, *2	3x3T, 64, *2	3x3T, 64, *2
output size:64	5x5T, 64, *2	3x3T, 32, *2	upsampling, 64, *2 [3x3T, 64 1x1, 256 3x3T, 64]
output size:128	5x5T, 3, *2, no bn, tanh	3x3T, 3, *2, no bn, tanh	upsampling, 64, *2 [3x3T, 32 1x1, 128 3x3T, 32] 1x1, 3, no bn, tanh
生成器參數量	1906.1 萬	239.2 萬	317.3 萬

表 4-5-2、實驗 2-1a 各組別辨別器架構說明

組別	1(擴展 1-1 設計)	2(擴展 1-2 設計)	3(擴展 1-3 設計)
output size:128	input	input	input 4x4, 32, no bn
output size:64	5x5, 32, /2, no bn	4x4, 32, /2, no bn	4x4, 64, /2
output size:32	5x5, 64, /2	4x4, 64, /2	4x4, 64, /2
output size:16	5x5, 128, /2	4x4, 128, /2	4x4, 128, /2
output size:8	5x5, 256, /2	4x4, 256, /2	4x4, 256, /2
output size:4	5x5, 512, /2	4x4, 512, /2	4x4, 512, /2
output size:1	Flatten+FC	Flatten+FC	Flatten+FC
辨別器參數量	436.7 萬	280.0 萬	286.6 萬



↑圖 4-5-1、

實驗 2-1a 組別 3 提升解析度之架構圖示

表 4-5-3、實驗 2-1a 各組別結果

組別	最小 FID 值	訓練時長
1	110.80	14819 秒
2	220.01	10422 秒
3	43.11	18565 秒

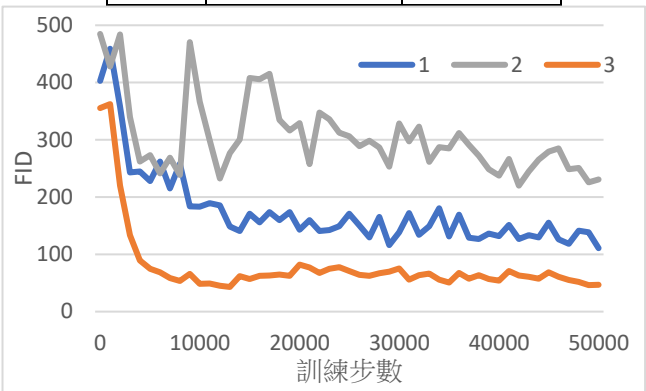


圖 4-5-2、實驗 2-1a 各組 FID 對訓練步數變化情形

由以上數據可以看出，組別 1、2 生成效果皆不佳，但組別 1 收斂較為成功，而組別 3 雖然相較在 64*64 解析度的實驗 1-3 FID 分數稍微沒那麼好，仍能正常的生成圖片。組別 3 生成人臉的結果與生成其他資料共同呈現在以下實驗 2-1b。

實驗 2-1b：提升至高解析度後在不同資料集的生成效果

在提升至 128*128 解析度後，我們一樣測試其在不同資料集上的生成效果，同樣包含 UT Zappos50K 與 Stanford cars 資料集。以下為實驗結果。

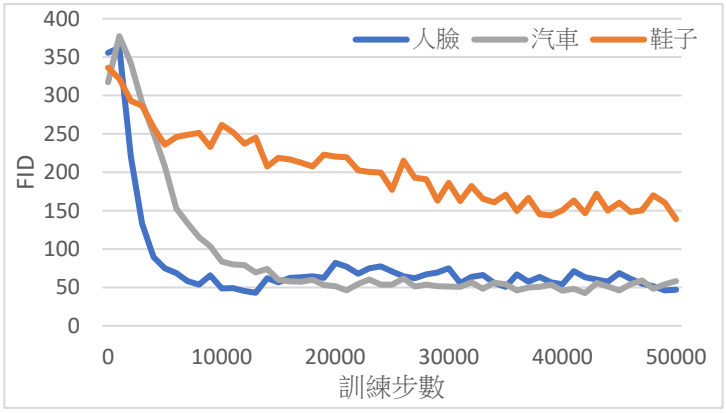


表 4-5-4、實驗 2-1b 各組別說明與結果

組別	最小 FID 值	訓練時長
人臉	43.11	18565 秒
鞋子	139.00	8745 秒
汽車	42.71	17999 秒

圖 4-5-3、實驗 2-1b 各組 FID 對訓練步數變化情形

以上結果可以看出，人臉的收斂速度最快，汽車其次，而鞋子則十分緩慢，但到最後仍還在收斂。因在不同資料上訓練的 FID 不適合直接比較，因此該結果僅用於評估收斂情形。以下為生成結果。（註：採用儲存的模型中 FID 最低的模型）

圖 4-5-4、CelebA-HQ 資料集與其生成結果（左）資料集（右）生成結果

圖 4-5-5、UT Zappos50K 資料集與其生成結果（左）資料集（右）生成結果

圖 4-5-6、Stanford Cars 資料集與其生成結果（左）資料集（右）生成結果

可以發現生成的圖片的缺點與實驗 1-4 生成 64*64 圖像的結果類似，在人臉部分細看即有瑕疵，且較低解析度的圖像要多；而鞋子部分輪廓尚可接受，但細節較難以呈現，且色彩明顯較資料集單調；汽車則是主體同樣容易扭曲，且有部分不自然的色塊，但主體尚可辨認出是汽車。另外由於鞋子與汽車的資料集並非針對高解析度圖像生成，因此部分圖片解析度不到 200 像素，若繼續提升至 256*256 解析度可能影響生成結果，因此後面更高解析度只討論 CelebA-HQ 資料集。

（二）提升至 256*256 與更高解析度

實驗 2-2：

利用實驗 1-1 到 1-3 的模型架構，結合實驗 2-1 的提升至較高解析度方法，我們將 256*256、512*512 解析度的模型都建構出來，並在 CelebA-HQ 資料集上測試其生成效果。但在運行 512*512 解析度的模型訓練時，由於處理時的數據量較大，GPU 的記憶體無法儲存 32 的批次大小，因此改用 8 作為批次大小，如此可能導致每步的訓練更難抓到收斂的方向而導致學習更慢或不穩定，而最後結果是模型訓練失敗，無法正常的建構出人臉且發生模式坍塌，因此該實驗只好在 256*256 止步。雖然有一些對空間使用的優化方式，如降低浮點數精度、多 GPU 並行訓練等，但若做出如此的優化理應討論其對結果的影響，而礙於篇幅與時間便不針對這些方式進行實作、測試與比較。以下為實驗結果。

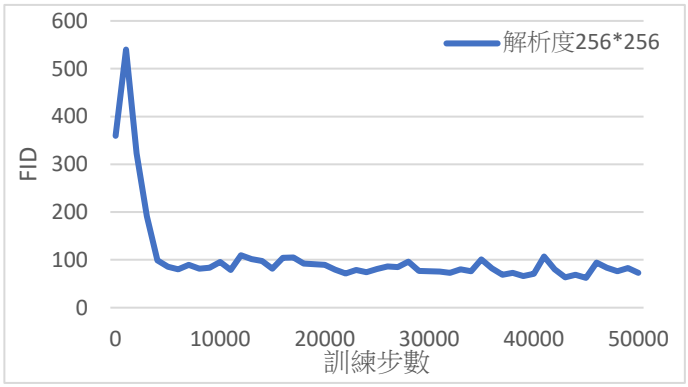


圖 4-6-1、實驗 2-2 之 FID 對訓練步數變化情形

表 4-6-1、實驗 2-2 說明與結果

說明	解析度 256*256
辨別器參數量	293.2 萬
生成器參數量	351.3 萬
最小 FID 值	62.37
訓練時長	51128 秒

同樣取最小 FID 值之模型生成圖像如下圖。

圖 4-6-2、CelebA-HQ 資料集與其生成結果（左）資料集（右）生成結果

由上圖可以很明顯的看出，生成出來的圖片瑕疵更多了，且眼睛與嘴巴部分出現十分嚴重的不自然、皮膚顏色也有些違和，導致整體看起來並沒有很像真人。可能是因為該解析度可以保留更多細節，如牙齒縫隙、眼白與眼黑清楚的分界、髮線等等，這使得模型需要學習的特徵大幅增加，而目前架構難以記錄如此多細節。

伍、討論

一、64*64 解析度圖像生成模型

（一）實驗 1-1（優化器）

學習率部分，經由實驗結果，我們發現對於 DCGAN 較好的學習率組合是辨別器與生成器分別為 0.0005 與 0.0002，推測是稍微提升辨別器學習率，可以使其辨別能力較不易被生成器的學習速度直接追上而導致訓練失敗，但將辨別器學習率提升過多，或者同時提高生成器學習率，都容易導致訓練不穩定。該結果與分別給予辨別器與生成器不同學習率的 TTUR (Heusel et al., 2018)更新規則相似。至於不同種類的優化器，可以發現使用 RMSprop 能在穩定性與收斂速度上取得平衡，我們猜測這與其適合處理複雜的 error surface 有關，相比之下 Adam 有著差不多的收斂速度，但在後期較為不穩定，難以收斂至更好的結果，SGD 則是難以順利收斂。

由以上結果，優化器與學習率確實對圖像生成模型的學習過程與收斂結果有所影響，因此要改良圖像生成模型，首先應選擇較穩定且容易收斂的優化器設置。

（二）實驗 1-2（卷積層參數）

在我們的實驗中，可以發現減少生成器通道數，以及降低生成器與辨別器的卷積核大小，有助於模型表現。在通道數的部分，我們對除了輸出層以外的卷積層減少一半的通道數，使得生成器參數量降至約原本的四分之一，讓模型的訓練較為高效，但可以發現減少通道數雖然很快就得到較好的結果，但後期沒有太大的進步，反而是原始的模型直到最後都還有收斂，因此若提供較多的步數進行訓練，原本的通道數可能會學到更多的特徵，達到相近或更好的效能，但考量到在效率與品質上能達到平衡，依然使用減少通道數的做法。另外卷積核大小部分，我們的結果與網路上許多 DCGAN 的實作案例相似，會降低卷積核大小，但其實在實驗中可以發現生成器的卷積核大小為 3、4、5 結果上並無太大差異，只不過 3x3 的卷積核大小有更短的訓練時間，也在我們設定的 5 萬步內達到了最好結果，因此採用該方案。

而由以上討論，卷積層的通道數與卷積核大小對模型訓練的效率有極大影響，適度調整卷積層參數將有助於圖像生成的品質與效率，也是改進的重要方向。

（三）實驗 1-3（更改模型架構）

透過實驗 1-3a、1-3b，我們發現原本 DCGAN 模型中有較高解析度特徵圖可學習參數較少的缺點，因此利用在生成器加入更多反卷積層與 1x1 conv 線性組合層並結合殘差作法，以及在辨別器加入更多卷積層並在一開始加入雜訊優化，以改進模型效能，數據顯示模型生成品質得到明顯提升並只略微增加訓練時間。另外我們有嘗試將生成器中增加層數與殘差單元應用在更前端的層上，對於最後結果並沒有改進，礙於篇幅沒有將該數據呈現出來，對於該現象我們認為由於較低解析度特徵圖本來就無法包含太多資訊，對於原本的架構其資訊量已經飽和，前端的層再增加層數並無助於增加其資訊量，原本的單層轉置卷積層就足以提供其足夠的資訊。

而實驗 1-3c 中，加入條件批次標準化可以讓生成器在對特徵圖進行標準化時，依據初始的雜訊進行縮放與平移，由結果可以發現這樣的做法確實能增強模型生成品質，另外儘管該方法會增加訓練時間，但考量到更少的訓練步數就能收斂至更好的結果，整體而言依然能提高訓練效率。

由以上討論可以知道，改進 GAN 模型的架構，尤其是生成器，並加入讓模型更易於學習、更能提取資料特徵的機制進行改進，儘管不見得能提升訓練效率，但圖像品質會有明顯的提升，這也是目前 GAN 圖像生成模型主要的改進方向。

（四）實驗 1-4（不同資料集）與 64*64 解析度圖像生成結果總結

透過進行以上改進，我們的模型在 CelebA-HQ 資料集可以達到 30.43 的 FID 分數，在 UT Zappos50K 與 Stanford cars 也分別達到 49.51、45.61。而為了確保模型能夠生成多樣的圖像，並非對於任何輸入皆產生相同圖像的模式坍塌，以在 CelebA-HQ 資料集訓練為例，我們輸入隨機訊號產生 10 張圖像如下圖 5-1。（註：採用過程中記錄到 FID 最佳的模型）

圖 5-1、64*64 解析度圖像生成多樣性展示

而為了確認模型有將雜訊用以產生圖像，我們對其進行漸變（或稱插值，interpolation）如下圖 5-2，先隨機產生兩雜訊 a 與 b，接著將 a、b 和其係數和為 1

的組合，如 0.1a+0.9b、0.2a+0.8b 等等，輸入模型進行計算，若得到連續的圖片則代表模型對於兩個雜訊的組合可以生成出特徵介於兩者的圖像，反之得到不連續圖片則代表是從真實資料中直接複製圖像，並不符合生成的用意，屬於訓練失敗。

圖 5-2、64*64 解析度圖像生成漸變展示

可以看到我們的模型在多樣性與漸變過程上，皆給出了符合期待的結果。

二、更高解析度圖像生成模型

（一）實驗 2-1（提升至 128*128 解析度）

在實驗 2-1a 中，可以發現實驗 1-2 的模型提升解析度之效果最差，因此我們可以推論，若沒有經過實驗 1-3 的改進，單純以 1-2 縮減通道數與卷積核大小的模型提升解析度，會因為後期所保留的資訊量過少而生成失敗，儘管礙於篇幅沒有展示生成結果，但該組還發生了模式坍塌，也應歸咎於該原因；而實驗 1-1 的模型雖然因為較多的通道數而取得較 1-2 好的結果，但由於架構上與 DCGAN 完全相同，在增加模型深度以生成更高解析度之圖片仍然容易遇到瓶頸；而我們在實驗 1-3 改進出的模型，透過將 64*64 以上的特徵圖，除了最後輸出之特徵圖採用 32 通道殘差單元，其餘利用 64 通道殘差單元擴展提高解析度，取得了最好的結果，推測與其它結果差異之大的原因是我們雖然降低了前段的通道數，但後段仍保持較多的通道與較多的參數以學習特徵，而該部份我們也測試了將 1-3 模型擴展的其他方式，但大多都只取得 50~70 的 FID 分數，而礙於篇幅並沒有呈現該部分數據。

在實驗 2-1b 中，可以看到提升解析度後的模型一樣在不同資料集上都給出堪用的生成結果，即主體輪廓都算清楚，但細節的紋理較難被還原。而 FID 部分，分別在 CelebA-HQ、UT Zappos50K 與 Stanford cars 上得到 43.11、139.00、42.71，其中 UT Zappos50K 之 FID 結果明顯較差，我們認為可能更資料集中的鞋子有各種樣式、圖樣，主體更複雜所致。而我們對在人臉資料集上訓練的模型中取 FID 最好的，進行多樣性與漸變的測試如下圖 5-3、5-4，可以看到多樣性符合我們期待，另外儘管漸變過程出現有瑕疵的圖片，但仍然有吻合我們所需要的連續圖片。

圖 5-3、128*128 解析度圖像生成多樣性展示

圖 5-4、128*128 解析度圖像生成漸變展示

透過這些討論，可以證實在 64*64 解析度模型中使用的改進方法，在較高解析度上也能較原本的模型有所提升，且架構上的改進對高解析度的特徵圖進行更多處理，也較能解決如 DCGAN 等架構簡單的模型，在提升解析度時難以訓練的問題。

（二）實驗 2-2（提升至更高解析度）

如實驗中所述，由於提升解析度時會增加運算時所使用的記憶體，因此硬體上的限制是提升解析度時所需要克服的，儘管如此，仍有軟體的優化方式可以緩解該問題，如降低運算時使用的浮點數精度或使用多個 GPU 並行訓練。

而在 256*256 解析度的模型，我們可以看到其在訓練時依然可以穩定收斂，但若看到生成出的結果，可以發現圖像中的瑕疵明顯變多，且使得整體圖片變得極度不自然，推測是因為所需要學習的資訊量大幅增加，而目前模型難以學習、處理或記錄這些細節，而相比之下，近期能夠生成較高解析度圖像的模型，通常都有更多架構或學習方法上的許多重大改進，使這些細節得以被學習。也因此，雖然我們並沒有訓練更高解析度的模型，但推測由於運用殘差單元，模型應該可以收斂，但後期要開始產生人臉之細節時，可能會無法學習或者輸出怪異的圖片 。

三、與現有技術進行比較

經過以上對於 64*64 解析度模型進行改進的實驗，以及將改進後模型擴展至更高解析度的實驗，我們透過 FID 可以證實較原本的 DCGAN 有更好的生成品質，我們的模型在同樣步數下花費差不多的訓練時間，但以更少的步數即可收斂至更好結果，且更容易擴展至高解析度圖像生成，證實我們所提出改進 GAN 圖像生成模型的品質與效率之方法，使模型效果有顯著的提升。而查詢網路上相關提升 GAN 圖像生成模型效能的研究，絕大多數都專注於架構或訓練方式之提升，並未找到針對各項變因提供調整與改進

方向的研究，因此我們認為提供一套改進方法對於改良相關模型的工作有所幫助，除了提供造成影響的變因供參考，亦減少盲目調整與測試所花費的時間與能源成本。儘管相較於現今最新的圖像生成模型，如 StyleGAN 等已經可以取得個位數的 FID 分數，研究中的模型仍有很大的進步空間，但考量我們仍然採用最原始的 DCGAN 訓練模式，在其之後陸續有更好的訓練方式、可以改善效能的架構相繼出現，因此若能結合這些新的技術，研究中的模型可能也會得到更好的結果。不過本研究的核心概念，是對於圖像生成模型進行改良的方式、思路，從優化器、卷積層參數、模型深度與架構等著手，改良生成圖像品質並盡可能地提升效能，因此也適用於改進除了 DCGAN 以外的其他模型。

陸、結論

一、結論

本研究提出一套改良 GAN 生成圖像之品質與效率的流程，並以客觀指標證明生成品質與訓練速率優於原先的模型。實驗結果表明，對於 DCGAN，採用辨別器學習率略高的 RMSprop，稍微減少特徵圖通道數與卷積核大小，並將辨別器與生成器中尺寸較大的特徵圖進行更多處理，再利用殘差單元替換單層轉置卷積層、生成器改用條件批次標準化層，皆會提高模型穩定度、生成品質與訓練效率，而將這些改進應用在更高解析度上的模型同樣能達到不錯的效果。儘管改良後的 DCGAN 與更近代的模型相比，仍有許多進步空間，但本研究所提供逐步改良模型效能的思路與方式亦能應用於其他 GAN 圖像生成模型上。

二、未來展望

GAN 被提出至今已經接近十年，而在 DCGAN 後也經歷了七、八年的時間，在這段期間不同的技術相繼被提出並累積起來，才促成如今進步的技術。在蒐集資料時，其實有發現一些可能會增加品質與效率的技術，例如進行漸進式成長的 PGGAN (Karras et al., 2018)，或利用預訓練模型改進辨別器並引入跨通道、跨尺寸特徵的 Projected GANs (Sauer et al., 2021)。其實我們最初的目標是開發出一款能兼顧品質（Quality）且快速（Quick）的圖像生成模型 QQGAN，想要在模型中加入更多訓練方式與模型架構上的改變，但礙於時間與篇幅未能實現出來，期許未來能再繼續改進我們的模型。

柒、參考文獻資料

Dobilas, S. (2022). Transposed convolutional neural networks—How to increase the resolution of your image. Medium. https://towardsdatascience.com/transposed-convolutional-neural-networks-how-to-increase-the-resolution-of-your-image-d1ec27700c6a

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. https://doi.org/10.48550/ARXIV.1406.2661

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. arXiv. https://doi.org/10.48550/arXiv.1512.03385

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2018). Gans trained by a two time-scale update rule converge to a local nash equilibrium. arXiv. https://doi.org/10.48550/arXiv.1706.08500

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive growing of gans for improved quality, stability, and variation. arXiv. https://doi.org/10.48550/arXiv.1710.10196

Karras, T., Laine, S., & Aila, T. (2018). A style-based generator architecture for generative adversarial networks. https://doi.org/10.48550/ARXIV.1812.04948

Krause, J., Stark, M., Deng, J., & Fei-Fei, L. (2013). 3d object representations for fine-grained categorization. 2013 IEEE International Conference on Computer Vision Workshops, 554–561. https://doi.org/10.1109/ICCVW.2013.77

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. https://doi.org/10.48550/ARXIV.1511.08458

Prabhu. (2019). Understanding of convolutional neural network (Cnn)—Deep learning. Medium. https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv. https://doi.org/10.48550/arXiv.1511.06434

Sauer, A., Chitta, K., Müller, J., & Geiger, A. (2021). Projected gans converge faster. arXiv. https://doi.org/10.48550/arXiv.2111.01007

Yu, A., & Grauman, K. (2014). Fine-grained visual comparisons with local learning. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 192–199. https://doi.org/10.1109/CVPR.2014.32

Yu, A., & Grauman, K. (2017). Semantic jitter: Dense supervision for visual comparisons via synthetic images. 2017 IEEE International Conference on Computer Vision (ICCV), 5571–5580. https://doi.org/10.1109/ICCV.2017.594

【評語】052503

此作品提出一套針對現有生成圖像的生成對抗網路模型的改良方法，透過進行實驗探討不同變因對生成品質與效率的影響。在實驗中作者調整原有的優化器設置、卷積層參數、模型架構等變因，並以一些客觀指標評估實驗結果，實驗結果顯示此作品提出的一些改良方法有較好的圖像生成效果。建議未來除了變化這些常用的設定和變數之外，能有自己創新的內容。

作品海報

GAN 圖像生成模型之畫質與效能強化

摘要

近年來，名為生成對抗網路的非監督式學習方法蓬勃發展，透過使產生假資料的生成器與辨別資料真偽的辨別器互相學習，只要提供資料集便可學習其特徵而生成出能以假亂真的資料。本研究提出一套針對現有生成圖像的生成對抗網路模型的改良方法，透過進行實驗探討不同變因對生成品質與效率的影響，調整原有的優化器設置、卷積層參數、模型架構等，並以客觀指標評估實驗結果，證實經過本研究提出的方法改良有更好的效果。另外改進的方式應用在各種資料集訓練的模型及更高解析度的模型，數據表明也有不錯的成果。而本研究希望能提供更明確的模型改進方向給研究人員，並減少嘗試改良模型所花費的時間與能源成本，以此減少訓練龐大的模型所造成的環境影響。

研究動機

偶然看到名為 StyleGAN 的人臉生成模型，只需要提供一組人臉照片，就可以透過電腦進行學習，生成出一張不存在於世界上的人臉。 查詢更多相關資料後，我們發現其架構很龐大，需要進行大量運算，耗費較多金錢、時間以及能源，進而造成環境負擔甚至加劇氣候變遷。因此我們希望能提出一套改良現有模型的方法，能得到較好的生成品質，且有較高的訓練效率，可以用較普遍的設備訓練，同時節省時間和能源成本，並且以此達到聯合國提出的永續發展目標中提升能源效率的目標。

StyleGAN 成果

研究目的

本研究旨在提出一套強化 GAN 圖像生成模型之品質與效率的方法，分析不同變因對模型的影響，藉以討論出改善模型的方向與流程。研究目標如下：

（一）探討影響 GAN 生成 64*64 解析度圖片品質、效率之因素並改良模型

（二）探討前項之方法用於較高解析度之圖像生成的品質與效果

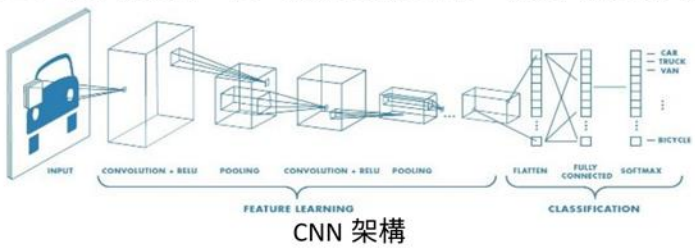
（三）將前兩項之方法總結成提升 GAN 圖像生成模型之品質與效率的流程

文獻回顧

（一）圖像生成相關技術與模型

1.卷積神經網路（Convolutional Neural Network，CNN）(O'Shea & Nash, 2015)

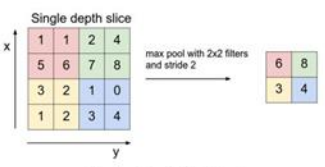
CNN 於上世紀末期提出，是一種前饋神經網路，訊號由輸入層單向傳播至輸出層。其靈感來自動物的視覺神經由不同的神經元，相鄰的細胞會有相似且重疊的視覺區域，由此構成完整的影像。CNN 對圖像處理有優秀表現，構造通常包含卷積層、池化層、全連接層。



CNN 架構

(1)卷積層（Convolution Layer，Conv）：透過卷積核（Kernel）對輸入圖片上的區塊進行卷積運算，再移動一個步幅（Stride）到下個區域，並循環此操作投影至另一矩陣上，稱作特徵圖（Feature Map）。而特徵圖可能會有很多通道，如 RGB 彩色圖片就是有紅、綠、藍三個通道。

(2)池化層（Pooling Layer）：將輸入圖片進行下採樣，以降低資料量，亦可以模糊邊緣。通常使用「最大池化（Max Pooling）」進行運算，即將固定大小的區塊，取出最大值投影至新的矩陣。由於池化的過程會快速降低資料量，近年來較少使用。



最大池化圖示

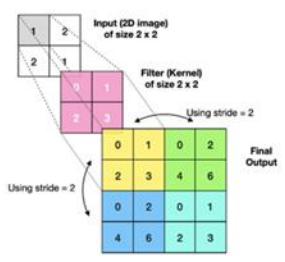
(3)全連接層（Fully Connected Layer，FC）：將資料攤平（Flatten）至一維，並利用全連接層，將每個數值乘上對應的權重並輸出至下一層。

2.生成對抗網路（Generative Adversarial Network，GAN）(Goodfellow et al., 2014)

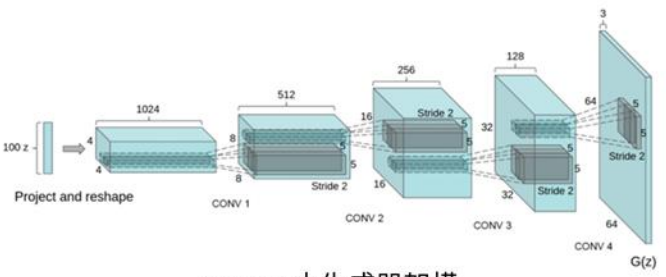
GAN 是一種非監督式學習的方法，概念為訓練辨別器（Discriminator）與生成器（Generator）兩個模型，辨別器分辨真實資料與生成器的假資料，生成器則要生成出以假亂真的圖片。透過兩者互相學習、進步，最後即可生成仿真資料。

3.深度卷積生成對抗網路（Deep Convolutional GAN，DCGAN）(Radford et al., 2016)

將 GAN 與 CNN 進行結合即是 DCGAN。其中辨別器之結構為 CNN；生成器結構則將 CNN 之卷積層替換為轉置卷積層（Transposed Convolution Layer），不同之處為轉置卷積核中的數值是與輸入矩陣的相乘並映射在輸出圖片，而生成器的目標是使生成的圖片被辨識為真實資料。



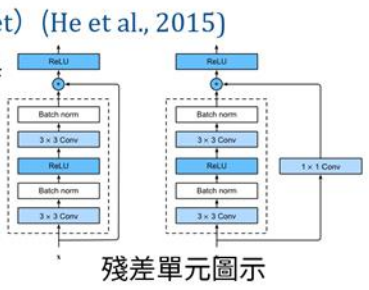
轉置卷積圖解



DCGAN 之生成器架構

4.殘差神經網路（Residual Neural Network，ResNet）(He et al., 2015)

ResNet 是一種前饋神經網路，透過跳躍連接與捷徑越過某些層，相較傳統的學習 x→F(x)，ResNet 學習 x→x+F(x)，利用該技巧可以解決傳統模型的退化（Degradation）問題。右圖為 ResNet 的基本單位：殘差單元。



殘差單元圖示

（二）評估模型之方法

模型生成之效果大致有解析度、圖像品質、效率等。其中解析度是指圖片的長寬各由多少像素組成，不同的解析度涉及到不同的模型架構，不適合將不同解析度進行比較，因此將每個解析度分開討論。本研究由以下角度分析模型成效：

1.品質

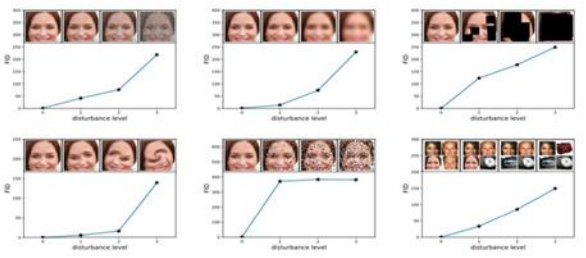
圖像品質是指在同樣解析度下，圖像的清晰度、雜訊、與真實人臉的相似度等，而解析度並不代表品質的好壞，例如右圖為同一張圖片，其上為 64*64 解析度，其下為 128*128 解析度但經過模糊處理，儘管 128*128 有更多像素格保存圖片資料，整體品質仍較 64*64 解析度差。

不同處理方式下的 FID

(左上)高斯噪聲　(中上)高斯模糊

(右上)插入黑色塊　(左下)旋轉

(中下)椒鹽噪聲　(右下)資料集受汙染



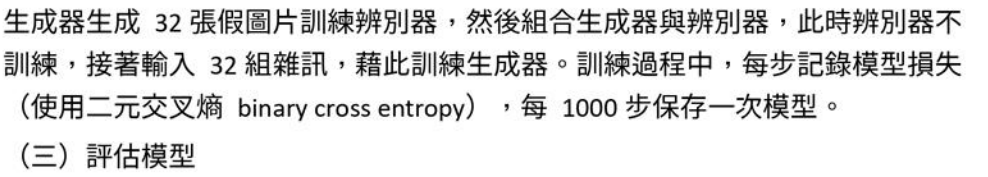
我們使用 Fréchet inception distance（FID）(Heusel et al., 2018)作為主要的客觀評價標準。FID 是一種用以評估生成圖像的指標，將真實圖片以及模型生成的圖片輸入 InceptionV3 模型，取出特徵向量並進行數學運算以求得兩者分布之距離，越小的距離代表模型生成之圖片與真實資料分布越接近，生成品質越好。該指標為目前最廣泛被用以評估圖像生成模型的指標，被證實比先前提出的其他指標更能反映人類視覺上的感受。

2.效率

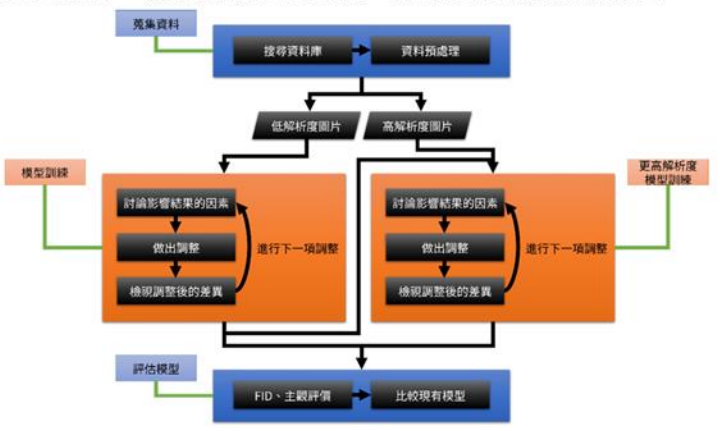
最後模型效率的部分，可以分做許多部份進行討論，如訓練效率、圖像生成效率、模型檔案大小等。其中我們較著重於訓練效率，在同樣硬體下訓練效率會受到許多因素如模型架構、深度、參數量等影響。對於所有訓練，我們使用相同的硬體，並記錄訓練資訊，並進行討論以評估模型效率。

研究過程或方法

一、研究架構：在開始研究前，我們擬定了完整的研究架構如下圖。



二、實驗流程與設計：為了順暢進行實驗，先確定實驗流程如下圖。



（一）蒐集資料集與預處理

訓練使用的資料集為 CelebA-HQ (Karras et al., 2017)，有 3 萬張 1024*1024 解析度的人臉圖片。為了訓練不同解析度的模型，將圖片縮放至 64*64、128*128、256*256 等解析度。另外為了驗證模型效能，還使用 Stanford Cars (Krause et al., 2013)，包含 16185 張汽車圖像；與 UT Zappos50K(Yu & Grauman, 2014, 2017)，包含 50025 張鞋子圖像。

（二）模型訓練

我們採用架構簡單的 DCGAN 為基礎進行改進，實作出 DCGAN 後，設定批次（batch，一次對一疊資料同時進行運算）大小為 32，進行 5 萬步（step，利用一個批次運算、修改一次模型內參數）學習，每一步先取 32 張真實圖片、讓生成器生成 32 張假圖片訓練辨別器，然後組合生成器與辨別器，此時辨別器不訓練，接著輸入 32 組雜訊，藉此訓練生成器。訓練過程中，每步記錄模型損失（使用二元交叉熵 binary cross entropy），每 1000 步保存一次模型。

（三）評估模型

利用訓練時保存的模型進行 FID 分數計算，然後將結果以圖表呈現。

研究結果與討論

一、64*64 解析度圖像生成模型

由於影響模型效能的因素非常多，為了能夠一步步的探討影響模型效能的因素，我們將修改不同變因的實驗做以下安排：

實驗 1-1 ：先從決定模型如何調整內部參數以學習特徵的優化器開始。

實驗 1-2 ：在不修改模型架構的情況下，調整模型卷積層、轉置卷積層的參數。

實驗 1-3 ：利用調整好的優化器與卷積層參數，對模型架構進行一步的修改。

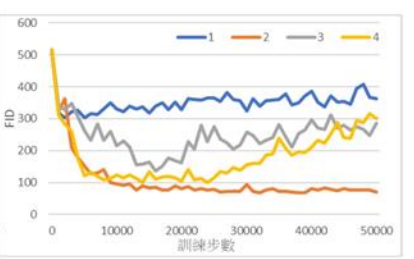
實驗 1-4 ：將得到的模型進行不同資料的生成以討論模型的生成效果與限制。

（一）優化器設置

實驗 1-1a ：改變學習率

學習率代表了模型修改參數的時候，一次所需要修正的量，影響模型的學習速度與穩定度，由於 GAN 的訓練過程是辨別器與生成器互相學習的動態過程，因此選用錯誤的學習率可能導致訓練失敗。該實驗我們使用 Adam 優化器，測試了以下不同的學習率設置。

組別	辨別器學習率	生成器學習率	最小 FID 值	訓練時長
1	0.0002	0.0002	302.21	10001 秒
2	0.0005	0.0002	66.30	10119 秒
3	0.001	0.0002	136.54	10162 秒
4	0.001	0.0004	100.09	11964 秒

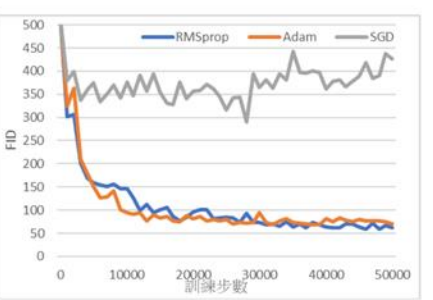


整體而言組別 2 表現最好，可以正常收斂、整體的穩定性也還不錯，組別 1 無法收斂，組別 3、4 可以收斂但穩定性不佳。 鑒於本實驗結果，後續實驗採用組別 2 的設置。

實驗 1-1b ：不同優化器

優化器是模型尋找最佳解的方式，因此會影響收斂的過程，進而影響最後成效。我們採用了 3 種常用的優化器來進行比較，分別有：每次調整所有參數的 SGD、加入自適應學習率機制的 RMSprop、以及兼具自適應學習率和動量機制的 Adam。

優化器	最小 FID 值	訓練時長
Adam	68.30	10119 秒
RMSprop	59.09	10150 秒
SGD	289.90	9901 秒



可以看到 SGD 難以收斂，Adam 與 RMSprop 有較好結果，雖然 Adam 一開始收斂較快，但後期沒有繼續收斂，而 RMSprop 則是持續緩慢收斂，最終較 Adam 良好。因 RMSprop 的訓練過程較穩定，後續採用該優化器。

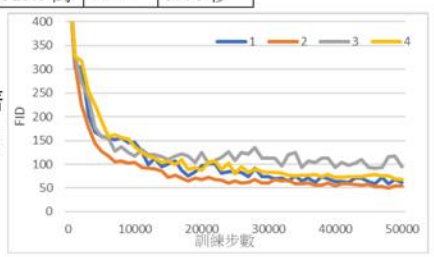
實驗 1-1 討論：經由實驗結果，較好的學習率組合是辨別器與生成器分別為 0.0005 與 0.0002，推測是稍微提升辨別器學習率，可以使其辨別能力較不易被生成器的學習速度直接追上而導致訓練失敗。該結果與分別給予辨別器與生成器不同學習率的 TTUR(Heusel et al., 2018)更新規則相似。至於優化器的種類，使用 RMSprop 能在穩定性與收斂速度上取得平衡，相比之下 Adam 在後期較為不穩定，SGD 則難以收斂。

（二）卷積層參數

實驗 1-2a ：改變 filter 數量

卷積層的 filter 可以理解為通道數，雖然擁有較多通道代表更多資料，但過多會使訓練難度增加；反之過少則不足以生成出良好的圖片。原始 DCGAN 中生成器在 4*4 特徵圖有 1024 通道，後提升一次解析度通道數減半；辨別器則是將 RGB 圖片轉為 64 通道的 32*32 特徵圖，後每降低一次解析度通道數變兩倍。我們將原始 DCGAN 中通道數進行調整，除了輸出通道外，其餘乘上表格中比值，如表中組別 2 之生成器 4*4 特徵圖有 512 通道。

組別	辨別器 filter 比值	生成器 filter 比值	辨別器參數量	生成器參數量	最小 FID 值	訓練時長
1	1	1	431.8 萬	1886 萬	59.09	10150 秒
2	1	1/2	431.8 萬	513.0 萬	49.44	8548 秒
3	1	1/4	431.8 萬	149.0 萬	91.77	8711 秒
4	1/2	1/2	108.4 萬	513.0 萬	67.49	6776 秒

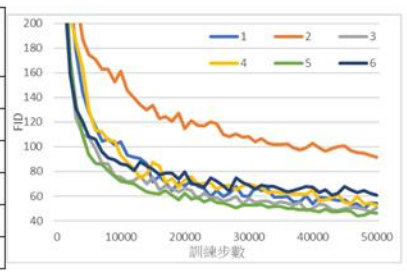


可以看出，組別 2 辨別器 filter 不變、生成器 filter 減半有更快的收斂速度與更好的結果，且可顯著降低訓練時間。由於實驗結果，後續實驗都是採用該做法。

實驗 1-2b ：改變 kernel size

卷積核大小相當於神經元的感知範圍，過大的卷積核可能造成網路難以學習卷積核所需要匹配的特徵，過小的卷積核會使感知範圍之間重疊過少而無法生成或辨識圖片。因此我們將原始 DCGAN 中 kernel size 進行調整（組別 1 為原始設置），以下是實驗結果。

組別	辨別器 kernel size	生成器 kernel size	辨別器參數量	生成器參數量	最小 FID 值	訓練時長
1	5x5	5x5	431.8 萬	513.0 萬	49.44	8548 秒
2	5x5	4x4	431.8 萬	358.0 萬	52.95	7830 秒
3	5x5	3x3	431.8 萬	237.4 萬	47.27	8367 秒
4	5x5	2x2	431.8 萬	151.3 萬	91.42	7538 秒
5	4x4	3x3	276.8 萬	237.4 萬	44.11	7404 秒
6	3x3	3x3	157.0 萬	237.4 萬	60.64	6982 秒



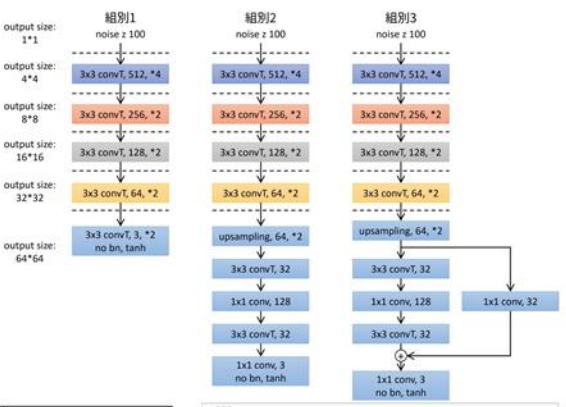
由實驗數據可看出對於辨別器與生成器分別採用 4x4 與 3x3 的卷積核大小有最好的收斂速度與結果，且可稍微降低訓練時間。因此後續實驗皆採用這樣的設置。

實驗 1-2 討論：減少通道數讓模型的訓練較為高效，但理論上原本的通道數可能會學到更多的特徵，達到更好的效能，考量到在品質與效率間達到平衡，依然使用減少通道數的做法。減少卷積核大小部分，我們的結果與許多實作相似，會降低卷積核大小，實驗亦表明 3x3 的卷積核大小可以提供更快的收斂速度與更短的訓練時間，並在實驗中達到了最好效果。

（三）改變模型架構

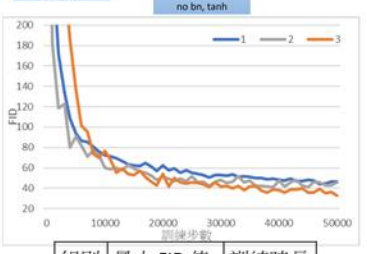
實驗 1-3a ：生成器網路強化

目前網路架構前面的層有較多參數、通道，如 16*16 特徵圖到 32*32 有 128*64=8192 個不同卷積核（兩者通道數相乘），但最後由 32*32 特徵圖提升至 RGB 圖片，只會有 64*3=192 個不同卷積核，該部分可能是影響模型效能的關鍵。我們把生成器輸出為 64*64 的層數增加，輸出前能有更多資訊，且將核大小 1x1、步幅為 1 的卷積層應用在網路中，用途是對上一層的輸出進行線性組合。不過在增加層數的同時，可能會造成網路較難以訓練，因此嘗試應用殘差概念。下圖是我們的生成器架構，若未特別註明，卷積層（conv）與轉置卷積層（convT）後皆有批次標準化層（bn）以及使用 ReLU 激活函數（輸出層直接使用 tanh，不使用 ReLU）。



為了方便表達，後續類似的模型架構會使用下表的形式表達，其中若未特別註明皆是卷積層（加註 T 者為轉置卷積層）且使用 bn 與 ReLU，另外殘差單元部分使用中括號表示，而跳躍連接的部分，若輸出殘差單元之形狀與輸入形狀不同，統一使用通道數與殘差單元輸出相同的 1x1 conv 進行線性組合。

組別	1	2	3
output size:4	3x3T, 512, *4	3x3T, 512, *4	3x3T, 512, *4
output size:8	3x3T, 256, *2	3x3T, 256, *2	3x3T, 256, *2
output size:16	3x3T, 128, *2	3x3T, 128, *2	3x3T, 128, *2
output size:32	3x3T, 64, *2	3x3T, 64, *2	3x3T, 64, *2
output size:64	3x3T, 3, *2, no bn, tanh	upsampling, 64, *2 3x3T, 32 1x1, 128 3x3T, 32 1x1, 3, no bn, tanh	upsampling, 64, *2 [3x3T, 32 1x1, 128 3x3T, 32] 1x1, 3, no bn, tanh
生成器參數量	237.4 萬	243.3 萬	243.5 萬



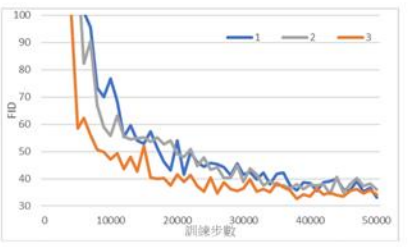
組別	最小 FID 值	訓練時長
1	44.11	7404 秒
2	41.00	7848 秒
3	33.16	8094 秒

實驗數據可看出組別 2 對 64*64 特徵圖進行更多處理確實對模型有幫助，提供更快的收斂速度與更好的結果，而組別 3 將最後的層應用殘差單元，最後結果有非常明顯的進步且較組別 2 穩定。基於該結果，後續採用結合殘差的更多尾端反卷積層。

實驗 1-3b ：辨別器網路強化與優化

由上個實驗的經驗，辨別器網路中從原始圖片轉換到 32*32 特徵圖，處理數據的參數量較少，因此針對該部分加強。為避免辨別器網路過於強大，使生成器無法對抗 DCGAN 會訓練失敗，因此辨別器只加一層通道數 32、步幅為 1 的卷積層。另外生成器經改進後，一開始需要略多一點時間學習，而現在增加辨別器強度，可能導致訓練一開始辨別器過強而生成器難以學習。我們採用一種常見的優化方式，於最初辨別器的真實標籤加入噪音，而後逐漸減小，於 2000 步恢復正常，以此削弱一開始的辨別器，使生成器更好對抗，待學習一定特徵再恢復。以下為實驗結果。

組別	說明	最小	訓練時長
1	對照組	33.16	8094 秒
2	輸入後加步幅 1 的 3x3 conv	34.39	8130 秒
3	同第 2 組但一開始加入噪音	32.65	8173 秒

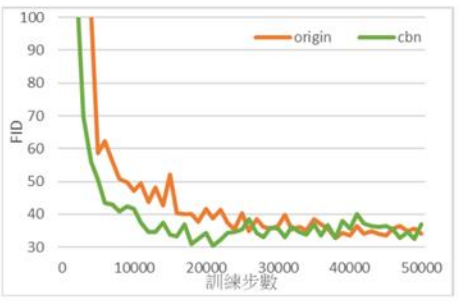


可以發現組別 1 與 2 訓練過程與結果的差異不大，但組別 3 收斂速度快上許多。鑒於結果，後續實驗皆採組別 3 之模式。

實驗 1-3c ：加入條件批次標準化（Conditional Batch Normalization）機制

原本的 DCGAN 模型中，多數卷積層後皆有批次標準化層（Batch Normalization），將卷積後的資料乘上 gamma 縮放、加上 beta 平移，其中 gamma 與 beta 為可透過每次輸入的批次分布情況學習的參數，該層可以避免過擬合（overfitting）與模式坍塌（mode collapse）。而條件批次標準化則是透過輸入條件來運算 gamma 與 beta，可用於生成不同種物品的 GAN，而在我們的模型中，由於一開始的雜訊能夠代表人臉不同的特徵，因此也可以將雜訊輸入該層。

組別	最小 FID 值	訓練時長
原始(origin)	32.65	8173 秒
加入條件批次	30.43	10017 秒



由數據可以看出，加入 cbn 後，前期的收斂速度快上不少，雖然訓練 50000 步的時間明顯變長，但加入 cbn 在第 18000 步就達到比原始模型最低點更低的 FID 分數，並在 21000 步取得最小 FID，所以考慮更少的訓練步數，cbn 的效率還是較高，因此後續實驗皆採用。

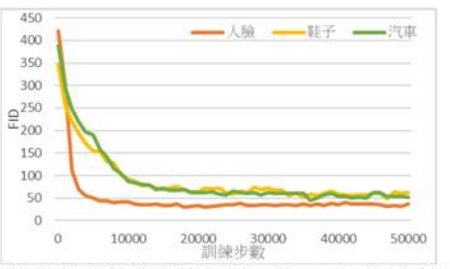
實驗 1-3 討論：我們發現原本 DCGAN 有較高解析度特徵圖可學習參數較少的缺點，因此在生成器加入更多反卷積層與線性組合層並結合殘差，以及在辨別器加入卷積層並用雜訊優化，數據顯示品質得到明顯提升並只略增加訓練時間。而實驗 1-3c 中，加入條件批次標準化可以讓生成器在對特徵圖進行標準化時，依據初始的雜訊進行縮放與平移，儘管該方法會增加訓練時間，但更少的訓練步數能收斂至更好結果，整體而言能提高效率。

（四）不同資料集的生成效果

實驗 1-4 ：

為了瞭解我們的模型對不同資料集的效果，在本實驗測試 UT Zappos50K 與 Stanford cars 資料集。以下為實驗結果。

組別	最小 FID 值	訓練時長
人臉	30.43	10017 秒
鞋子	49.51	8745 秒
汽車	45.61	9746 秒



由以上實驗結果可以看出，生成鞋子與汽車的收斂速度皆較慢一些，且可能由於不同的資料集圖片特點，如構圖、紋理較複雜，造成較難訓練到較好結果，且經不同資料集訓練的 FID 並不適合拿來直接做比較，該結果僅用以評估收斂過程。以下為不同資料集的生成結果。（採用紀錄的模型中 FID 最低的模型）

資料集圖片　　模型生成圖片

由以上實驗結果可以發現在人臉的生成中，模型生成的圖片大致上看起來沒問題，但有些許瑕疵；而生成的鞋子圖片比起原本的資料集，較不會有太過複雜的紋理；至於生成汽車的圖片，儘管有生成出汽車的部分樣貌，但難以將這些小細節完美的組合，輪廓也較粗糙，因此整體效果僅還算堪用。但由以上的結果已經可以看出，我們的模型在其他資料也可以正常訓練。

實驗 1-4 討論：透過進行以上改進，在 CelebA-HQ 達到 30.43 的 FID 分數，UT Zappos50K 與 Stanford cars 也分別達到 49.51、45.61。而為了確保模型能夠生成多樣的圖像，並非對於任何輸入皆產生相同圖像的模式坍塌，我們輸入隨機訊號產生 10 張圖像如下圖左。而為了確認模型有將雜訊用以產生圖像，我們對其進行漸變（或稱插值，interpolation）如下圖右，若得到連續的圖片則代表模型對於兩個雜訊的組合可以生成出介於兩者的圖像，反之不連續圖片則代表是從真實資料中直接複製圖像，並不符合生成的用意，屬於訓練失敗。可以看到我們的模型在多樣性與漸變過程上，皆給出了符合期待的結果。

二、更高解析度圖像生成模型

為了探討影響更高解析度模型的不同因素，我們將進行以下實驗：

實驗 2-1 ：就 64*64 解析度的不同模型架構，將其提升至 128*128 解析度。

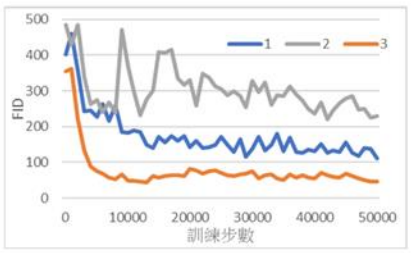
實驗 2-2 ：嘗試生成 256*256 或更高解析度圖片，並探討其生成效果或限制。

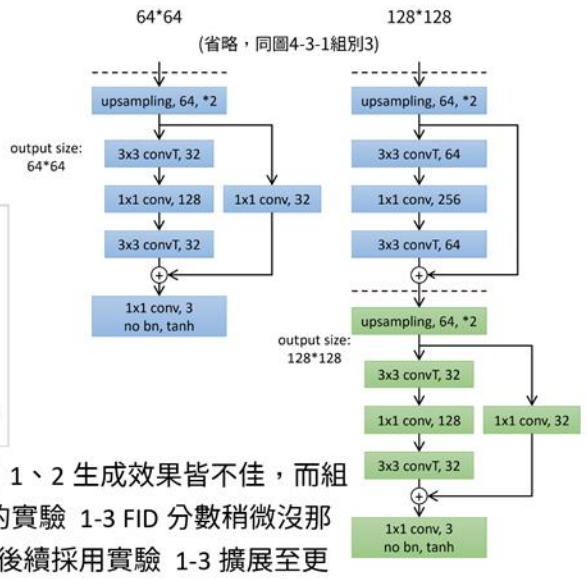
（一）提升至 128*128 解析度

實驗 2-1a ：不同低解析度模型架構提升至高解析度

將實驗 1-1、1-2、1-3 皆增加一層使輸出提升至 128*128。其中 1-1、1-2 直接利用原本建構模型的邏輯，生成器在輸出前增加一層轉置卷積層，辨別器在輸入後加一層卷積層。而實驗 1-3 的模型為了延續對高解析度特徵圖進行更多處理的精神，我們將殘差單元用以提升 64*64 特徵圖至 128*128，如下圖；而辨別器部分重複兩次通道數為 64 的卷積層。

組別	最小 FID 值	訓練時長
1	110.80	14819 秒
2	220.01	10422 秒
3	43.11	18565 秒



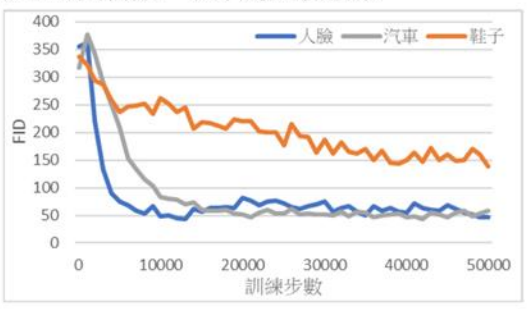


由以上數據可以看出，組別 1、2 生成效果皆不佳，而組別 3 雖然相較在 64*64 解析度的實驗 1-3 FID 分數稍微沒那麼好，但仍能正常的生成圖片。後續採用實驗 1-3 擴展至更高解析度的做法。

實驗 2-1b ：提升至高解析度後在不同資料集的生成效果

在提升至 128*128 解析度後，我們一樣測試其在不同資料集上的生成效果，同樣包含 UT Zappos50K 與 Stanford cars 資料集。以下為實驗結果。

組別	最小 FID 值	訓練時長
人臉	43.11	18565 秒
鞋子	139.00	8745 秒
汽車	42.71	17999 秒



資料集圖片

模型生成圖片

可以發現生成的圖片細看即有瑕疵，且較低解析度的圖像要多。

實驗 2-1 討論：可以發現將實驗 1-3 的模型提高解析度，透過將 64*64 特徵圖利用殘差單元再進行一次處理，取得了最好的結果，推測該模型與其它結果差異之大的原因是我們雖然降低了較低解析度特徵圖的資訊量，但較高解析度的特徵圖仍保持較多的通道以學習特徵。

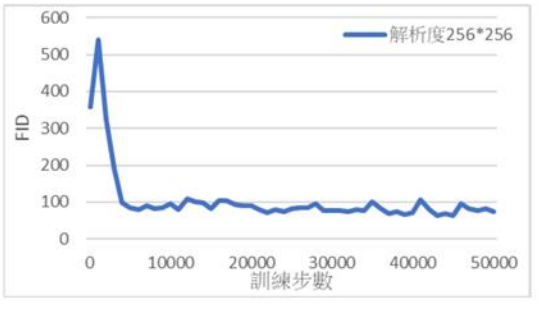
（二）提升至 256*256 與更高解析度

實驗 2-2 ：

利用實驗 1-1 到 1-3 的低解析度模型架構，結合實驗 2-1 提升至較高解析度，我們得以將更高解析度的模型建構出來。但在運行 512*512 解析度的模型訓練時，由於 GPU 記憶體不足，因此實驗在 256*256 止步。以下為實驗結果。

說明	解析度 256*256
辨別器參數量	293.2 萬
生成器參數量	351.3 萬
最小 FID 值	62.37
訓練時長	51128 秒

模型生成圖片



由左圖可以很明顯的看出，生成出來的圖片瑕疵更多，且出現十分嚴重的不自然，導致整體看起來並沒有很像真人。可能是因為該解析度可以保留更多細節，使得模型需要學習的特徵大幅增加，而目前架構難以記錄如此多細節。

實驗 2-2 討論：在 256*256 解析度的模型，我們可以看到其在訓練時依然可以穩定收斂，但生成出的結果瑕疵明顯變多，使得整體圖片變得極度不自然，推測是因為所需要學習的資訊量大幅增加，而目前模型難以學習、處理或記錄這些細節。雖然我們並沒有訓練更高解析度的模型，但可以推測其可能會較難學習或者輸出怪異的圖片。

三、總結討論：與現有技術進行比較

經過以上實驗，我們透過 FID 可以證實改進後的模型較原本的 DCGAN 擁有更好的生成品質，改進後的模型在同樣步數下花費差不多的訓練時間，但以更少的步數即可收斂至更好結果，且更容易擴展至高解析度圖像生成，證實我們所提出改進 GAN 圖像生成模型的品質與效率之方法，使模型效果有顯著的提升。

而查詢網路上相關提升 GAN 圖像生成模型效能的資料，現有研究大多數都專注於架構或訓練方式之提升，並未找到針對各項變因提供調整與改進方向的研究，因此我們認為提供一套改進方法對於改良相關模型的工作有所幫助，除了提供造成影響的變因供參考，亦減少盲目調整與測試所花費的時間與能源成本。

儘管相較於現今最新的圖像生成模型，如 StyleGAN 等已經可以取得個位數的 FID 分數，研究中的模型仍有很大的進步空間，但考量我們仍然採用最原始的 DCGAN 訓練模式，在其之後陸續有更好的訓練方式、可以改善效能的架構相繼出現，因此若能結合這些新的技術，研究中的模型可能也會得到更好的結果。不過本研究的核心概念，是對於圖像生成模型進行改良的方式、思路，從優化器、卷積層參數、模型深度與架構等著手，改良生成圖像品質並盡可能地提升效能，因此也適用於改進除了 DCGAN 以外的其他模型。

結論

一、結論

本研究提出一套改良 GAN 生成圖像之品質與效率的流程，並以客觀指標證明生成品質與訓練速率優於原先的模型。實驗結果表明，對於 DCGAN，採用辨別器學習率略高的 RMSprop，稍微減少特徵圖通道數與卷積核大小，並將辨別器與生成器中尺寸較大的特徵圖進行更多處理，再利用殘差單元替換單層轉置卷積層、生成器改用條件批次標準化層，皆會提高模型穩定度、生成品質與訓練效率，而將這些改進應用在更高解析度上的模型同樣能達到不錯的效果。儘管改良後的 DCGAN 與更近代的模型相比，仍有許多進步空間，但本研究所提供逐步改良模型效能的思路與方式亦能應用於其他 GAN 圖像生成模型上。

二、未來展望

GAN 被提出至今已經接近十年，而在 DCGAN 後也經歷了七、八年的時間，在這段期間不同的技術相繼被提出並累積起來，才促成如今進步的技術。在蒐集資料時，其實有發現一些可能會增加品質與效率的技術，例如進行漸進式成長的 PGGAN (Karras et al., 2018)，或利用預訓練模型改進辨別器並引入跨通道、跨尺寸特徵的 Projected GANs (Sauer et al., 2021)。其實我們最初的目標是開發出一款能兼顧品質（Quality）且快速（Quick）的圖像生成模型 QQGAN，想要在模型中加入更多訓練方式與模型架構上的改變，但礙於時間與篇幅未能實現出來，期許未來能再繼續改進我們的模型。

參考文獻資料

Dobilas, S. (2022). Transposed convolutional neural networks—How to increase the resolution of your image. Medium. https://towardsdatascience.com/transposed-convolutional-neural-networks-how-to-increase-the-resolution-of-your-image-d1ec27700c6a

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. https://doi.org/10.48550/ARXIV.1406.2661

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. arXiv. https://doi.org/10.48550/arXiv.1512.03385

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2018). Gans trained by a two time-scale update rule converge to a local nash equilibrium. arXiv. https://doi.org/10.48550/arXiv.1706.08500

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive growing of gans for improved quality, stability, and variation. arXiv. https://doi.org/10.48550/arXiv.1710.10196

Karras, T., Laine, S., & Aila, T. (2018). A style-based generator architecture for generative adversarial networks. https://doi.org/10.48550/ARXIV.1812.04948

Krause, J., Stark, M., Deng, J., & Fei-Fei, L. (2013). 3d object representations for fine-grained categorization. 2013 IEEE International Conference on Computer Vision Workshops, 554–561. https://doi.org/10.1109/ICCVW.2013.77

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. https://doi.org/10.48550/ARXIV.1511.08458

Prabhu. (2019). Understanding of convolutional neural network (Cnn)—Deep learning. Medium. https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv. https://doi.org/10.48550/arXiv.1511.06434

Sauer, A., Chitta, K., Müller, J., & Geiger, A. (2021). Projected gans converge faster. arXiv. https://doi.org/10.48550/arXiv.2111.01007

Yu, A., & Grauman, K. (2014). Fine-grained visual comparisons with local learning. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 192–199. https://doi.org/10.1109/CVPR.2014.32

Yu, A., & Grauman, K. (2017). Semantic jitter: Dense supervision for visual comparisons via synthetic images. 2017 IEEE International Conference on Computer Vision (ICCV), 5571–5580. https://doi.org/10.1109/ICCV.2017.594