臺北市立建國高級中學 111 學年度科學展覽

作品說明書封面

科別：電腦與資訊學科

組別：高級中等學校組

作品名稱：利用GAN實現歷史航空影像之上色

關鍵詞：pix2pix、CycleGAN、CUT

**摘要**

本研究以利用pix2pix及CycleGAN等模型實現歷史航照之上色為動機，並深入探討、評估與修正這些方式，以得出能有最好成效的深度學習模型。最初是進行預處理的動作，此研究擷取中研院的WMTS資料建立資料集，接著建立pix2pix、CycleGAN、CUT等由GAN衍生出的模型，探討改變生成器及判別器權重、生成器架構、GAN損失模式等，是否會對上色成效帶來影響，再透過FID及IS等指標評估。

經過實驗後發現，透過增加生成器、判別器filter數量之比例，可以有效降低CycleGAN、pix2pix中對抗損失上升的現象，也讓二模型的成效提高。而ResNet比起U-Net能幫助pix2pix生成較鮮艷、穩定的圖像，以灰階映射L\*a\*b的方式訓練，可降低生成圖片的一些瑕疵。平均而言，CycleGAN與CUT在為舊航照的上色表現較pix2pix好，而其中CUT對於影像細節的處理又較為精緻、穩定。若欲採用CUT模型，不採用Identity Loss、搭配LSGAN的損失模式會是最佳選擇。

1. **前言**
2. 研究動機

隨著台灣科技發展的進步，網路上可以輕易取得內政部所測製或中研院所徵集的一系列歷史航空照片，藉由數位典藏及GIS圖資發佈技術，讓二戰後期迄今所記錄臺灣歷史航照影像得以完整保存及活用。無論如何，歷年來所統計的資料均對人文科學的研究有著顯著的貢獻與影響。

但在比較過去及現在的地景變遷時，黑白的照片始終會帶來分析上的不方便，且網路上許多模型 (如: CNN、U-Net) 對舊航照的上色效果不甚理想 (問題如: 雜訊過多、鮮豔度不足)，因此此研究的最終目的，便是希望能建立一個上色模型 (運用生成對抗網路所衍伸的 pix2pix、CycleGAN 等模型實作)，並對這些方法進行深入的探討與比較，期許能讓灰階歷史航照的上色達到較佳的效果。

1. 研究目的
2. 利用生成對抗網路所衍伸的分支模型 (pix2pix、CycleGAN、CUT等) 對舊航照進行上色。
3. 以pytorch實作pix2pix、CycleGAN、CUT等模型，並探討以下變因對模型成效之影響:

1. 調整生成器與判別器之權重

2. 改變pix2pix生成器架構 (有無U-Net)

3. 改變pix2pix之訓練方式

4. 調整CUT模型的GAN損失模式

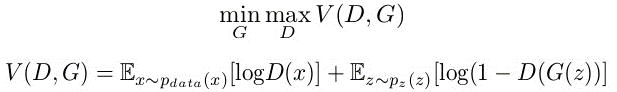
5. 是否採用Identity Loss

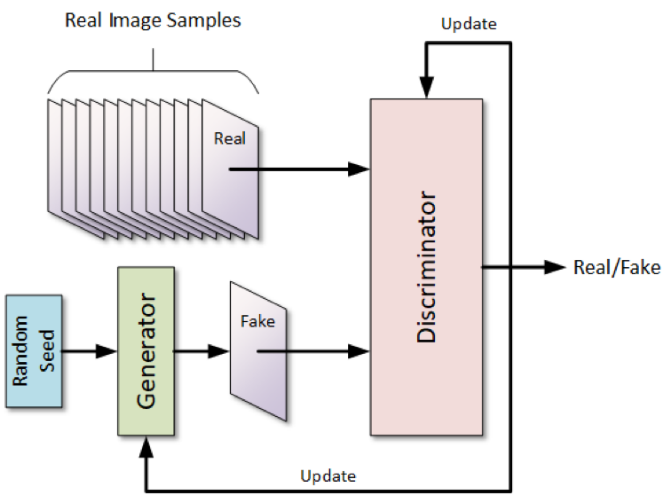
1. 分析各個模型間上色成效的差異，並透過生成對抗模型的FID、IS等指標來評估生成影像的品質與多樣性。
2. 文獻回顧與探討
   1. Pix2pix

pix2pix 是一種用於圖像轉換的深度學習模型，可以將輸入圖像轉換為目標圖像。它是一種生成對抗網絡 (GAN)，而其主要是採用CGAN (Conditional GAN )網路的結構，透過像是增加 U-Net或者 skip-connection 等不同機制，強化GAN。而他所應用的場景像是圖像的還原、修圖、圖像轉譯等。

* + 1. 生成對抗網路 (GAN)

生成對抗網路（Generative Adversarial Network，簡稱GAN）是非監督式學習的一種方法，透過兩個神經網路相互博弈的方式進行學習，該模型由生成器 (Generator) 和判別器 (Discriminator) 組成。生成器的目的是從潛在空間 (latent space) 中隨機取樣，用服從某一分布(高斯分布、常態分布等)的雜訊生成結果，生成結果盡可能模仿訓練集中的真實樣本。判別器的任務則是在輸入真實樣本或生成網路的輸出時，將生成網路的輸出從真實樣本中盡可能分辨出來，若輸入的樣本來自真實資料，則輸出較高的機率，反之則為小機率。透過生成器及判別器的之間的對抗，讓生成器學習產生更佳的結果。公式與模型架構圖如下:





圖一、生成器G生成與真實樣本相似的圖片，鑑別器D學習判斷圖片真偽

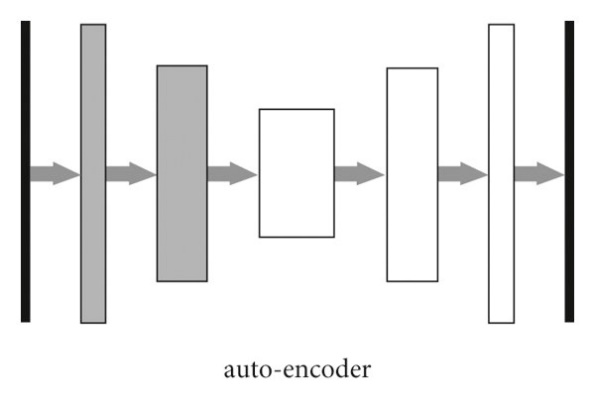
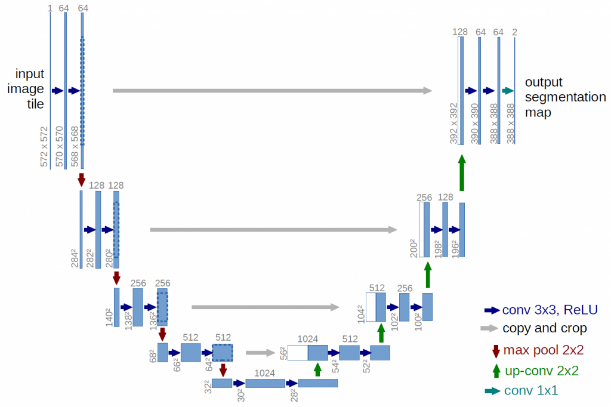
* + 1. CGAN (Conditional Generative Adversarial Network)

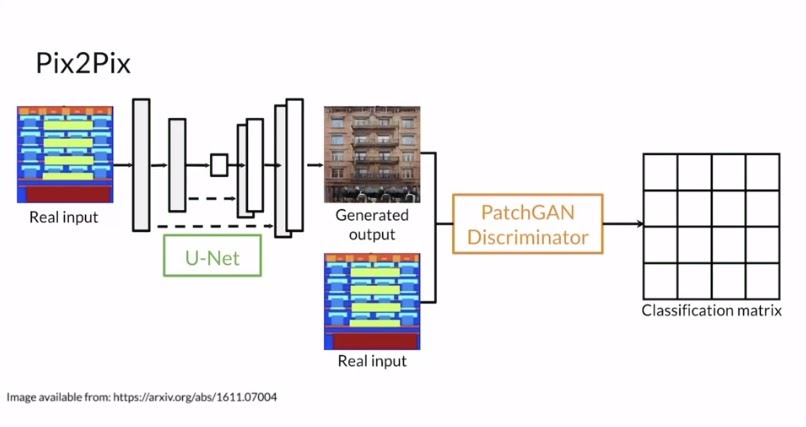
pix2pix的模型是由Conditional GAN衍伸而來，從其中改變了幾處細節。CGAN和一般的 GAN 只有一點微小的差別。一般的 GAN 學習隨機噪聲 (z) 和輸出圖片 (y)之間的映射函數 (G: z → y)，而 CGAN 除了參照GAN的大部分原理外，則是學習輸入圖片 (x) 和隨機噪聲 (z) 和輸出圖片 (y) 之間的映射函數 (G: {x, z} → y)。因此CGAN經常用於有給定條件的生成圖片，而非單純的隨機生成圖片。

* + 1. U-Net Generator:

大部分CGAN 模型的生成器均採用 Autoencoder 的結構 (encoder-decoder network)。傳統的 Autoencoder 的缺點是前半段的編碼器(Encoder)，它萃取特徵的過程，使輸出的尺寸(Size)越變越小，之後，解碼器(Decoder) 再由這些變小的特徵，重建成與原圖一樣大小的新圖像，因此，原圖很多的資訊都會遭到省略。在這個架構中，輸入會經過一系列逐漸縮小的神經層，直到通過瓶頸層，然後再逐漸變大。而在此網路中，每項資訊都會流過所有的神經層，包括最小的瓶頸層。但是，對於圖像風格轉移問題，輸入和輸出的圖片必定有一些資訊是相同的。

因此，如果能讓這些相同的資訊直接流過網路，而不必經過每一個神經層較為合理、有效率。透過將 skip connection (跳躍連結)加入模型中，設該神經網路具有 n 層，在第 i 層和第 n-i 層之間都加入跳躍連結，且每一個連結都可以讓資訊直接流過，即為 U-Net 的形式，也因為前述，U-Net 在圖像的分割與萃取表現的相當不錯。



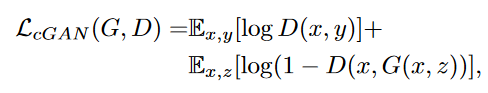
圖二、Auto-Encoder與U-Net的結構 (上)，及pix2pix的完整架構

* + 1. Markovian discriminator (PatchGAN):

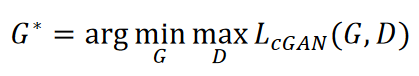
用L1和L2 loss重建的圖像很模糊，也就是說L1和L2並不能很好的恢復圖像的較複雜的高頻部分(圖像中的邊緣等)，但能較好地恢復圖像的簡單的低頻部分(圖像中的色塊)。為了能更好得對圖像的局部做判斷，PatchGAN的結構，也就是把圖像等分成許多patch，在這裡需要 GAN 的判別器去判斷圖案複雜處的真偽，剩下的只要依賴 L1、L2 loss即可。透過設計一會在N\*N大小方格範圍內循環遊走、判別真偽的鑑別器，可以得到較GAN更為理想的圖像轉換效果。

* + 1. 損失函數 (Loss) 的設定

(1) Adversarial Loss



G 的目標是使 G(x)看起來盡可能像 y 訓練集中的資料，而 D 的任務是區分真實樣本 y 和假樣本 G(x)。因此，G 希望最小化這個損失函數，而 D 則希望最大化它。數學表達式如下:



(2) L1 Loss

透過加入 L1 loss，可以幫助生成低複雜圖圖案處的特徵。可以用以下數學式表達:



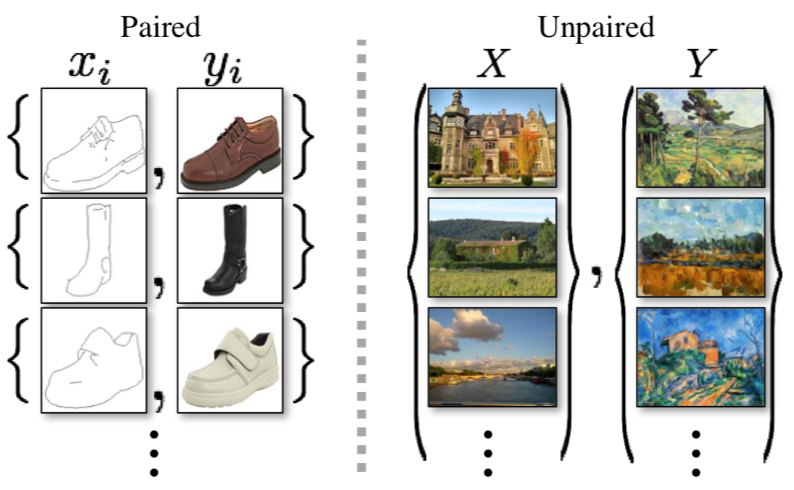
(3) 整體損失函數

pix2pix 整體的損失由 CGAN 的 Adversarial loss 和 L1 loss 組成，而λ為控制兩者比重的係數，而該模型的目標為讓生成器生成越來越擬真的圖片，可以用以下數學式表達如下:



* 1. CycleGAN

CycleGAN 是一種用於圖像轉換的深度學習模型，其模型架構主要由兩個生成器和兩個鑑別器構成。原理是透過兩物件彼此循環生成、被鑑別的過程，故稱之為CycleGAN，而其可以用於許多不同的圖像轉換任務，好處在不需要準備成對的條件圖像與真實圖像，對於成對圖片集不易或無法取得的機器學習頗有幫助。

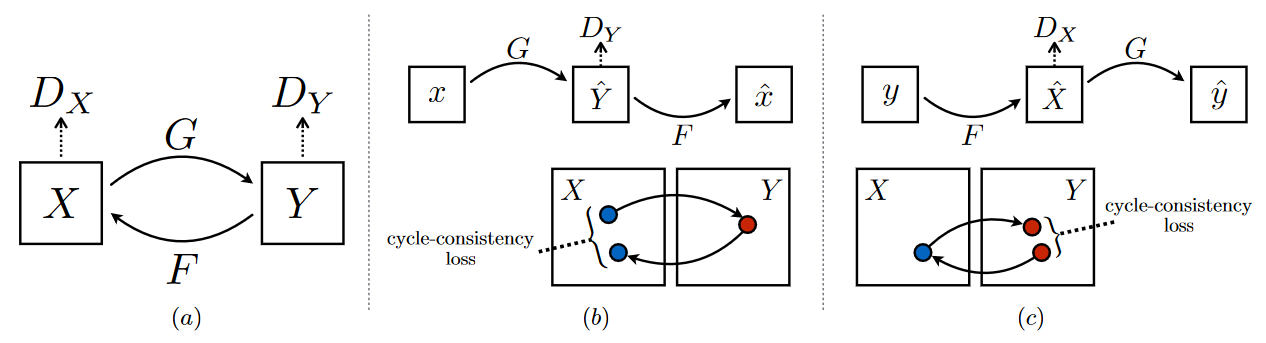


圖三、成對與非成對資料集示意圖

* + 1. Cycle Consistency 的原理

CycleGAN 和 pix2pix 都是由 GAN 衍生而來，但由於訓練圖片集並非成對，因此無法保證 GAN 訓練出的模型會學習到預期中的映射關係。因此，在這些模型中必須加入一些其他架構。CycleGAN使用了 Cycle Consistency 這個概念，代表著，如果有兩個轉換器 G: X→Y 和 F: Y→X，則 G 和 F 互為彼此的逆變換，即 F(G(x))≈x 和 G(F(y))≈y，一張圖片經過一次變換及一次逆變換應要與原本的圖片越接近越好。

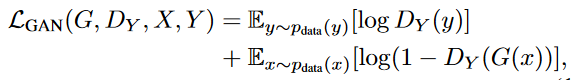
* + 1. CycleGAN模型架構



圖四、CycleGAN模型的原理與訓練流程

1. 假設我們今天的任務是將x domain轉換為y domain，CycleGAN可以看做是2個GAN的融合，一個GAN由生成器G(x→y)和判別器DY(促進G產生更逼真的y domain)構成；另一個GAN由生成器F(y→x)和判別器DX(促進F產生更逼真的x domain)構成，實現從Y域到X域的圖像生成和判別，兩個網絡構成循環（cycle）。
2. 而模型中亦引入透過Cycle consistency的原理設計的Cycle consistency loss。其中可以分為正向(x→G(x)→F(G(x))≈x)及反向(y→F(y)→G(F(y))≈y)的Cycle consistency loss。
   * 1. 損失函數的設定

(1) Adversarial Loss

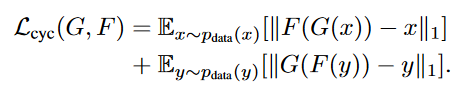


G 的目標是使 G(x)看起來盡可能像 y 訓練集中的資料，而 D 的任務是區分真實樣本 y 和假樣本 G(x)。因此，G 希望最小化這個損失函數，而 D 則希望最大化它。數學表達式如下:



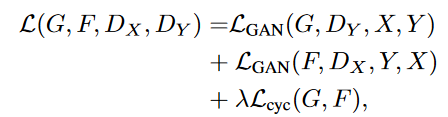
(2) Cycle Consistency Loss

對於 Cycle consistency 的定義，我們可訂出他的損失函數如下:



(3) 整體損失函數

由上述的 adversarial loss 和 cycle consistency loss 組合而成，而λ為控制兩者比重的係數，而該模型的目標為讓生成器G與F生成越來越擬真的圖片，可以用以下數學式表達:





(4) Identity Loss

生成器G用來生成y風格圖像，那麼把y送入G，應該仍然生成y，只有這樣才能證明G具有生成y風格的能力。因此G(y)和y應該盡可能接近。根據論文中的解釋，如果不加該loss，那麼生成器可能會自主地修改圖像的色調，使得整體的顏色產生變化。而 Identity Loss 的數學形式可以表達如下。



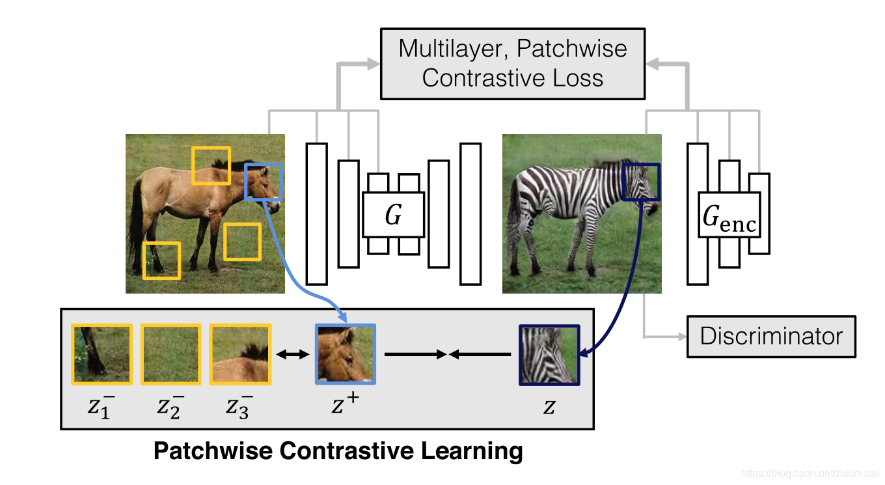
* 1. CUT (Contrastive-Unpaired-Translation)

該框架擺脫以往 Cycle-GAN 需要 2 組 GAN 並使用 Cycle-consistency loss 的架構，改為利用對比學習 (Contrastive Learning) 來鼓勵輸出的圖片相似輸入圖片，因此該模型只需要一組 GAN 即可進行圖像轉換

* + 1. 原理

CycleGAN的假設非常嚴格，要求輸入的圖像域和目標域之間存在雙射關係，這一點在其實是很難滿足的。因此通過最大化輸入輸出圖像塊的共同資訊（mutual information），使用一個對比損失函數 (infoNCE loss)， 來讓encoder將對應的圖像塊之間相互聯繫起來，與其他的圖像塊分離，如此一來encoder可以專注於兩個域之間共性的部分如形狀，而忽略兩個域之間的差異性部分如紋理。

* + 1. CUT 模型架構:



圖五、CUT的訓練流程

如圖所示，CUT使用多層圖像塊的對比損失，最大化相對應的多層圖像塊之間的共同資訊，這樣將生成器和Encoder相結合，取得對應輸入圖像的生成圖像

* + 1. 損失函數的設定

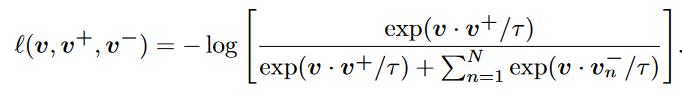
(1) Adversarial Loss

G 的目標是使 G(x)看起來盡可能像 y 訓練集中的資料，而 D 的任務是區分真實樣本 y 和假樣本 G(x)。因此，G 希望最小化這個損失函數，而 D 則希望最大化它。數學表達式如下:



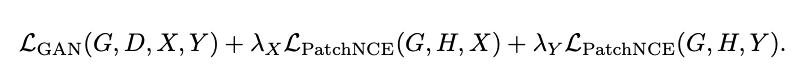
(2) InfoNCE Loss

進行對比訓練(Contrastive learning)時，我們希望模型可以有效的學會某個特徵時（稱做 Query），會採用正樣本(Positive)以及負樣本(Negative)來學習，而在圖像轉換的任務中可以透過提取相同圖像位置的區塊作為正樣本，而其他部分就是負樣本。將輸入圖像以及轉換後的圖像相同區塊的部分我們希望越像越好，而不相關的部分越不像越好。可以數學型式表達如下:



(3) 整體損失

結合前述，總損失公式可以寫成以下的數學形式，其中λx與λy為控制兩者比重的係數，λy同時也代表著CycleGAN中也有的Identity Loss。正常的CUT採用Identity Loss (參數設為λx=1，λy=1)，若未採用 Identity Loss，該CUT架構稱作FastCUT (參數設為λx=10，λy=0)，在一般情況下，FastCUT訓練的速度較快且運用的計算資源較少。



1. **研究設備與器材**

一、硬體:

(一) 筆記型電腦: (CPU: Intel(R) Core(TM) i5-1135G7@2.40GHz，記憶體 8GB)

二、軟體:

(一) Windows 11、Ubuntu 20.04 (作業系統)

(二) Python 3.10 (程式語言)

(三) Pytorch 1.13.1 (深度學習框架)

(四) Visual Stuidio Code、Google Colab (編輯器)

(五) 其他套件與軟體: QGIS (GIS軟體)、torchvision、matplotlib、numpy、WireGuard (連接中研院伺服器的VPN)

三、訓練資料來源:

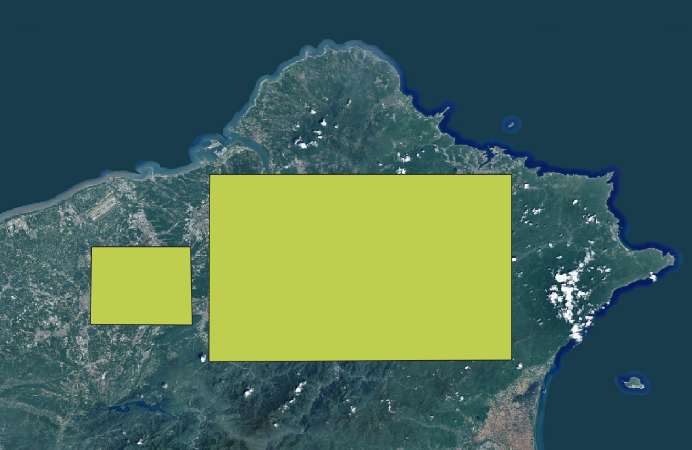
中央研究院臺北、台南百年歷史地圖 WMTS 服務、Google Earth、

1. **研究過程或方法**
2. 研究架構



圖六、研究架構圖

1. 前置處理
   1. 運用QGIS及其外掛程式Qtiles，以及中研院所提供之WMTS，擷取臺北、臺南地區航照。如圖，透過 ESRI shapefile 的格式界定矩形範圍內的區域，使用 Qtiles 套件抓圖 (zoom值設定為15~16)



圖七、透過shapefile的方式界定研究區，以擷取古航照與現代航照

* 1. 將航照透過Python的Pillow套件轉為灰階，存放於另一個資料夾。
  2. 透過shutil套件配置資料集。擷取民初航照的方法與前述相同。

1. 模型的訓練
   1. pix2pix

在 pix2pix 模型的訓練過程中，我們會輸入一對圖像，其中一張圖像是輸入，另一張則是目標輸出。模型的目標是使輸出圖像盡可能接近目標輸出圖像。我們將訓練模型所使用的圖像集分成三個部分，分別為訓練集(train)、驗證集(val)和測試集(test)。

訓練集用於訓練模型，驗證集則是在訓練過程中檢驗模型正確性的工具，而測試集則是在訓練完成後用來測試模型的性能，其中包括灰階處理的現代航照及1970~1980拍攝之臺北、台南地區航照(原始圖像即為灰階)。

而於此實驗當中，會將轉灰階的現代航照與現代航照合成為一張512\*256的圖片放入模型訓練，train中共 1550 張照片，test中共900張照片，val中共900張照片，共訓練200個epoch，而模型的學習率(Learning rate)均設為0.0002，在100個epoch後開始下降。



圖八、pix2pix datasets範例 (尺寸256\*512像素)

在訓練過程中會記錄生成器與判別器之對抗損失，再透過Python的matplotlib繪製成損失曲線，觀察其變化。

* 1. CycleGAN & CUT

在 CycleGAN與CUT模型的訓練過程中，我們會輸入不成對的二組256\*256像素的圖像，一組訓練集為A型態，另一組則為B型態，模型的目標是使A轉B和B轉A的任務越完美越好，進而為黑白影像，而此研究將A定為未上色的灰階影像，而B為有顏色的原影像 (Ground truth)。

可以將資料集分為4個部分，分別為trainA、trainB、testA、testB，train資料夾用來訓練模型，各為1550張，而test資料夾各500張照片，用於測試、生成最終成果，其中包括灰階處理的現代航照及1970~1980拍攝之臺北、台南地區航照(原始圖像即為灰階)。共訓練200個epoch，模型的學習率(Learning rate)均設為0.0002，在100個epoch後開始下降。

在訓練過程中除了記錄生成器與判別器之對抗損失，也會記錄生成彩圖、灰階影像與原datasets的差異，即為Cycle Consistency Loss，再透過Python的matplotlib繪製成損失曲線，觀察其變化。

1. 探討模型細節調整對上色成效的作用
   1. 調整生成器與判別器之權重

改變Generator和Discriminator上的Filter數量。透過改變這項因素，探討生成器與判別器之權重對損失趨勢的作用，在GAN訓練中Generator的Loss常會出現上升的現象，是因為訓練到後期Discriminator outperform Generator，尤其於CycleGAN、pix2pix、CUT等模型更容易發生。

所以此研究中，更動了各模型中的ngf (生成器末卷積層的filter數量)、ndf (判別器首卷積層的filter數量)等參數，再透過FID、IS、對抗損失等指標來衡量權重對模型成效的影響。

* 1. 改變pix2pix生成器架構 (U-Net的有無)

透過將 skip connection (跳躍連結)加入模型中之生成器，設該神經網路具有 n 層，在第 i 層和第 n-i 層之間都加入跳躍連結，且每一個連結都可以讓資訊直接流過，即為 U-Net 的形式。此研究欲探討U-Net對模型成效的影響，故在這邊pix2pix也另外採用ResNet作為該模型的生成器，以前述所提及的評估指標試著比較兩者成效的差異。

ResNet和恆等映射(Identity mapping)有很大的不同，恆等映射的核心思想是如果映射函數為H(x)，模型直接擬合訓練樣本x，輸出為H(x)。而在ResNet中，可以將模型的目標從H(x)改為F(x):=H(x)-x，模型的目標是每個殘差塊(Residual block)學到的「殘差變為0」(等價於恆等映射H(x)=x)，而不是直接學習恆等映射函數。

因此，只要考量殘差和原樣本，就能推出生成的映射函數，在實現上，只需要直接將上層的輸入 x 與經過 non-linear 轉換的輸出 F(x) 相加 (採用shortcut connection)，這樣不僅可以得到 F(x) + x，還不會增加額外的參數增加計算量。仍然可以使用 SGD 進行反向傳播訓練，常見的深度學習框架也可以簡單實現。

* 1. 調整pix2pix訓練方式

pix2pix該監督式學習模型而言，影像上色可以由兩種方式去進行，其一為直接訓練模型針對灰階、RGB模式的圖像的mapping，因為每張圖片是以像素為單位，而每個單位又具有三原色所對應的數值，一旦圖片的尺寸過大，便容易發生上色失準或耗費的運算資源龐大等問題，而另外一種影像上色常用的映射方式，是以L\*a\*b顏色空間的角度去處理圖片。

L\*a\*b顏色空間分為L、a、b三個通道，其中L描述亮度值，a和b兩項數值分別描述綠色至紫紅色與藍色至黃色，而本研究所採用的另一種映射手法，便是由L通道的數值，預測a與b兩圖層之狀況，因此這種映射訓練的方式如下所示，透過模型預測所得的L\*a\*b數據，即可轉換、還原為RGB的圖片。



圖九、使用L\*a\*b特色的pix2pix訓練流程，其中灰階與彩色圖片為成對的datasets。

* 1. GAN損失模式於對比式學習所帶來的影響 (Vanilla GAN / LSGAN / WGAN-GP)

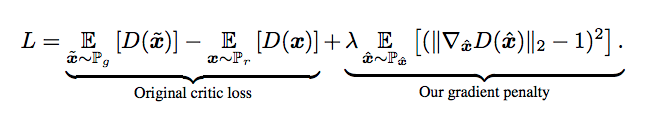
CUT模型主要採用CycleGAN的架構及對比學習的訓練模式，而此研究欲探討加入不同的損失進行評估，是否會影響到模型生成的效能，這邊主要針對Vanilla GAN、LSGAN (最小平方GAN，Least Squares GAN)、WGAN-GP (Wasserstein GAN with Gradient Penalty) 三種生成對抗模型中常見的損失模式進行探討，並評估其生成的效能。

Vanilla GAN是最簡單且最早被提出的生成對抗網路模型，它涉及兩個網絡：生成器（Generator）和判別器（Discriminator）。生成器產生假數據，判別器評估這些數據是否真實，而生成器根據判別器的評估進行改進，除了一般的對抗損失外，LSGAN及WGAN-GP皆採用了不一樣的損失模式。

LSGAN由Vanilla GAN所衍生，也是CUT原始論文所採用的GAN損失模式，其與Vanilla GAN的主要區別在於，它使用Least Squared Loss作為判別器的損失函數。該函數的優點在於，它在處理數據分佈不均的問題時優於Vanilla GAN，也就是透過使用平方損失來代替Cross-Entropy (交叉熵)，以改進判別器的判斷結果，使結果更平滑，可以數學形式表達如下。

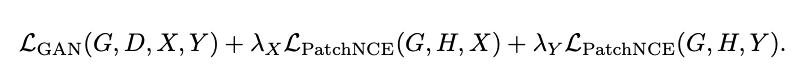


WGAN-GP其與其他生成對抗網絡的主要區別在於，它使用Wasserstein距離作為生成器與判別器之間的評估標準。它還使用了梯度懲罰(Gradient Penalty)作為正則項，以避免判別器在經過預測後輕易抵達飽和狀態，進而改善生成器之性能，可以數學形式表達如下。



* 1. 有無採用Identity Loss 帶來的影響

CUT模型的總損失公式可以寫成以下的數學形式，其中λx與λy為控制兩者比重的係數，λy同時也代表著Identity Loss。正常的CUT採用Identity Loss (參數設為λx=1，λy=1)，若未採用 Identity Loss，該CUT架構稱作FastCUT (參數設為λx=10，λy=0)。

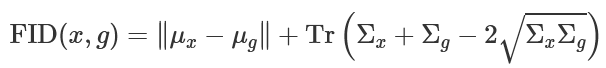


此階段欲透過比較CUT及FastCUT生成影像的表現優劣，以了解Identity Loss對於對比式的非監督式學習之影響。

1. 生成模型指標的採用
   1. FID (Frechet Inception Distance score，FID)

FID的設計源自Inception Net-V3，卻不使用該網路的原輸出，而是刪去了模型的輸出層，因此輸出層變為Inception Net-V3的最後一個池化層。該層的輸出為2048維向量，因此每張圖像都將被預測為2048個特徵。

對於常見的分佈（例如高斯分佈），只要確定了分佈類型和均值和方差，則該分佈就已確定。我們假設生成的圖像和真實圖像也遵循類似的分佈，如果它們之間的均值和方差比較接近，我們就有理由認為生成的圖像是比較真實的。但是直接計算圖像的均值和方差是不可取的，因為協方差矩陣規模太大（像素數\*像素數）。因此，我們首先通過Inception Net-V3映射為2048維的特徵向量，再求特徵向量的均值和協方差矩陣進行比較。可以數學形式表達如下:

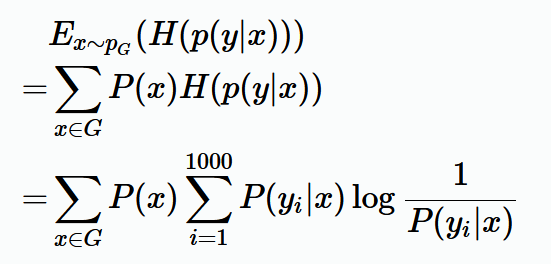


這條算式中μx, Σx是真實圖像的2048維特徵向量集合的均值和協方差矩陣；而μg, Σg是生成圖像的2048維特徵向量集合的均值和協方差矩陣，這些都是通過Inception Net-V3輸出的，較低的FID分數代表生成的分佈更接近真實圖片分佈，若測試用的真實圖片清晰且多樣，也就意味著生成圖像品質高、多樣性佳。

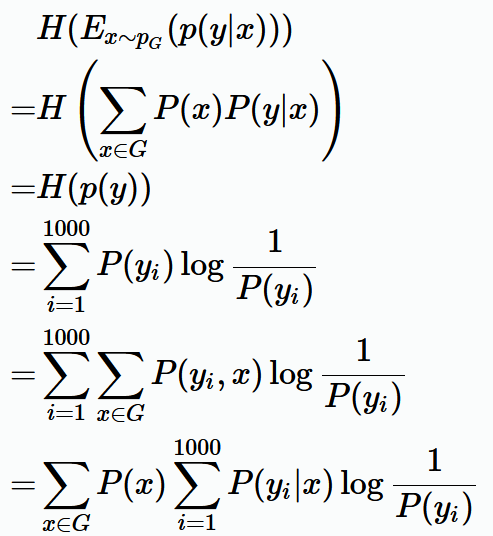
* 1. IS (Inception score)

IS是基於Google的預訓練網絡Inception Net-V3所設計的指標。 Inception Net-V3為一卷積網絡模型，輸入為圖片張亮，輸出為1000維向量。輸出向量的每個維度的值對應圖片屬於某類的概率，因此整個向量可以看做一個概率分佈。以下為IS的兩個定義，IS考慮以下兩個方面來評估生成器的質量：

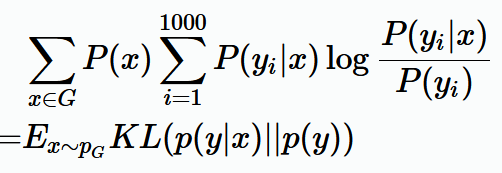
1. 對於單一的生成圖像，Inception輸出的概率分佈熵值應盡量小。越小說明生成圖像越有可能屬於某個類別，圖像質量就越高。可以數學表示如下:

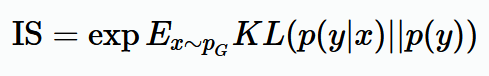


1. 對於生成器生成的一批圖像，Inception輸出的平均概率分佈熵值應盡量大。也就是說，因為生成器應保證生成圖像的多樣性，因此一批圖像在Inception的輸出應盡量平均地“遍歷”所有1000維標籤。可以數學表示如下。



將2個定義合而為一，可以得出以下的公式，其中KL(p(y|x)||p(y))為兩個分布的KL散度(相對熵)，再取指數即可得到IS score，根據定義，IS值越高，生成圖片的質量與穩定度越高:





(三) 評估的方式與流程:

1. FID score的計算需要兩個datasets去進行比較，而在此研究中，pix2pix和CUT模型僅會生成上色後之圖像，因此只需比較其與正確Ground truth的相似度即可，而在跑CycleGAN的模型時，會生成兩組圖像，一組為由現代航照所生成的灰階影像，另一組為由灰階處理影像所生成的上色航照，分別與實際灰階處理影像、實際現代航照進行比較，取生得之兩FID數值平均。

2. IS (Inception Score) 只需要輸入一組dataset即可計算出來，在此研究中，由於目標為生成仿真的古航照上色影像，故僅針對模型生成之上色圖像進行評估。

1. **研究結果**

以下實驗，均以灰階處理航照與實際彩色航照進行訓練，訓練200個epochs ; 測試的部分皆以灰階處理航照與古航照作為模型測試的輸入，上色後的影像為輸出。而CUT、CycleGAN因不具有成對的datasets，與舊航照的實驗相同，都不具有Ground truth可比較。而舊航照皆係1970~1980中研院拍攝之臺北/台南地區航照。

1. 預設模型的上色效果與瑕疵的統整
   1. Pix2pix

表一-實驗結果一-pix2pix

|  |  |  |  |
| --- | --- | --- | --- |
| 類型 | 輸入圖像 | 目標圖像 | 生成圖像 |
| 灰階處理現代航照 |  |  |  |
|  |  |  |
| 舊航照 |  | 無 Ground truth |  |
|  |  |

* 1. CycleGAN

註: CycleGAN是以不成對之資料集進行訓練，故無目標圖像可比對。

表二-實驗結果一-CycleGAN

|  |  |  |
| --- | --- | --- |
|  | 輸入圖像 | 生成圖像 |
| 灰階處理現代航照 |  |  |
|  |  |
| 舊航照 |  |  |
|  |  |

* 1. CUT (Contrastive Unpaired Translation)

表三-實驗結果一-CUT

|  |  |  |  |
| --- | --- | --- | --- |
|  | 輸入圖像 | 生成圖像(CUT) | 生成圖像(FastCUT) |
| 現代航照黑轉白 |  |  |  |
|  |  |  |
| 1970~1980台北航照 |  |  |  |
|  |  |  |

1. 調整生成器與判別器之權重帶來的影響

表四-實驗二結果 (損失曲線)

代號解釋 (將A作為灰階影像，B作為有色彩的影像) :

G\_GAN/A/B: 生成器的對抗損失、G\_L1: 生成器的L1損失、D\_real/fake/A/B: 判別器的對抗損失、cycle\_A/B: Cycle Consistency Loss、NCE\_Y: 對比學習中負樣本的infoNCE Loss、NCE: 平均inforNCE Loss。

ngf:生成器末卷積層的filter數量、ndf: 判別器首卷積層的filter數量。

|  |  |  |
| --- | --- | --- |
|  | ngf=64，ndf=64 | ngf=128，ndf=32 |
| Pix2pix |  |  |
| CycleGAN |  |  |
| CUT |  |  |
| FastCUT |  |  |

表五-實驗二結果 (針對舊航照成效比較)

|  |  |  |  |
| --- | --- | --- | --- |
| ngf、ndf有無調整 | 生成結果一 | 生成結果二 | 生成結果三 |
| Pix2pix 調整前 |  |  |  |
| Pix2pix 調整後 |  |  |  |
| CycleGAN 調整前 |  |  |  |
| CycleGAN 調整後 |  |  |  |
| CUT  調整前 |  |  |  |
| CUT  調整後 |  |  |  |

1. 改變pix2pix生成器之架構 (U-Net的有無)

表六-實驗三

|  |  |  |
| --- | --- | --- |
| 有無U-Net | 灰階處理後的現代航照 | 1970~1980古航照 |
| Ground Truth |  | 無Ground truth |
| 有 |  |  |
| 無 (以ResNet為生成器) |  |  |

1. 改變pix2pix訓練方式

表七-實驗四結果

|  |  |  |  |
| --- | --- | --- | --- |
|  | 圖1 | 圖2 | 圖3 |
| 映射出L\*a\*b後  轉RGB圖片 |  |  |  |
| 灰階映射RGB |  |  |  |

1. GAN損失模式於對比式學習所帶來的影響 (Vanilla GAN / LSGAN / WGAN-GP)

表八-實驗五結果

|  |  |  |  |
| --- | --- | --- | --- |
|  | 圖1 | 圖2 | 圖3 |
| Vanilla |  |  |  |
| LSGAN |  |  |  |
| WGAN-GP |  |  |  |

1. 有無採用Identity Loss 帶來的影響

表九-實驗六結果

|  |  |  |  |
| --- | --- | --- | --- |
| CUT  (灰階處理航照) |  |  |  |
| FastCUT  (灰階處理航照) |  |  |  |
| CUT  (舊航照) |  |  |  |
| FastCUT  (舊航照) |  |  |  |

1. 模型的效能評估

表十-實驗七結果  
此段落中，ngf表示生成器末卷積層的filter數量、ndf表示判別器首卷積層的filter數量。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pix2pix (註: 原pix2pix以U-Net作為生成器架構) | | | | |
| 輸入 | 採用模型 | FID | IS最大值 | IS 平均值 |
| 以灰階處理之航照作輸入 | 原pix2pix | 48.58 | 11.62 | 2.49 |
| pix2pix  (以L\*a\*b映射的方式訓練) | 43.11 | 12.10 | 2.51 |
| ngf=128,ndf=32 | 47.80 | 11.20 | 2.51 |
| ResNet 生成器  (ngf=128,ndf=32) | 46.96 | 12.42 | 2.52 |
| 以舊航照  作輸入 | 原Pix2pix | 139.49 | 13.66 | 3.20 |
| pix2pix  (以L\*a\*b映射的方式訓練) | 138.11 | 13.69 | 3.22 |
| ngf=128,ndf=32 | 145.70 | 13.62 | 3.33 |
| ResNet 生成器  (ngf=128,ndf=32) | 132.90 | 11.75 | 3.38 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CycleGAN | | | | |
| 輸入 | 採用模型 | FID | IS最大值 | IS 平均值 |
| 以灰階處理之航照作輸入 | 原CycleGAN | 169.50 | 11.18 | 3.05 |
| ngf=128,ndf=32 | 145.09 | 15.73 | 3.41 |
| 以舊航照  作輸入 | 原CycleGAN | 144.24 | 11.93 | 3.28 |
| ngf=128,ndf=32 | 128.49 | 12.29 | 3.19 |

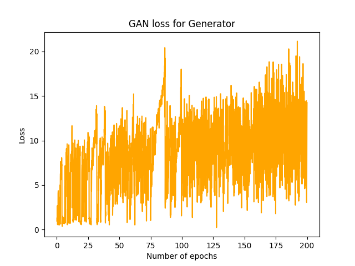
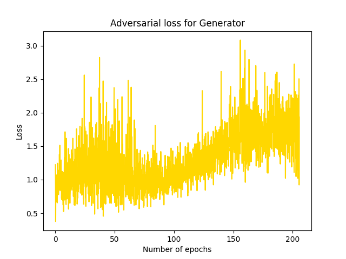
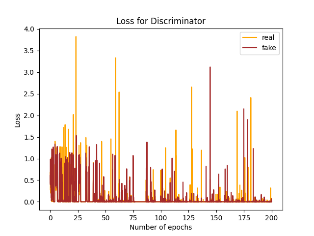
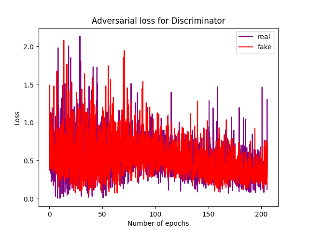
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CUT (註: 原CUT、FastCUT以LSGAN為GAN損失模式) | | | | | |
| CUT模式 |  | 採用模型 | FID | IS最大值 | IS 平均值 |
| 採用Identity Loss | 以灰階處理之航照作輸入 | 原CUT | 145.99 | 13.23 | 3.34 |
| ngf=128,ndf=32 | 165.76 | 16.23 | 3.10 |
| 改變GAN架構  (Vanilla) | 168.22 | 15.22 | 3.03 |
| 以舊航照  作輸入 | 原CUT | 163.55 | 17.64 | 3.41 |
| ngf=128,ndf=32 | 160.14 | 12.25 | 3.07 |
| 改變GAN架構  (Vanilla) | 167.11 | 14.39 | 2.91 |
| 不採用 Identity Loss  (FastCUT) | 以灰階處理之航照作輸入 | 原FastCUT | 97.03 | 15.34 | 3.25 |
| LSGAN with  ngf=128,ndf=32 | 97.11 | 17.25 | 3.33 |
| 改變GAN架構  (Vanilla) | 165.26 | 15.39 | 3.55 |
| 改變GAN架構  (WGAN-GP) | 234.15 | 5.67 | 1.12 |
| 以舊航照  作輸入 | 原FastCUT | 157.04 | 11.27 | 3.60 |
| LSGAN with  ngf=128,ndf=32 | 163.49 | 12.99 | 3.25 |
| 改變GAN架構  (Vanilla) | 176.55 | 12.55 | 3.36 |
| 改變GAN架構  (WGAN-GP) | 331.52 | 3.12 | 1.09 |

1. **討論**
2. 預設模型的上色效果與瑕疵統整

由結果圖片可以看出，pix2pix對於灰階處理航照的上色效果最好，唯獨在光影的處理上較不明顯，CycleGAN與CUT二模型分別有各自的瑕疵，前者在圖片邊緣和建築物邊緣的處理較為模糊，而後者對於建築物的上色不夠精確 (如將建築物的屋頂塗上漸層的藍色)。

對於本研究主要探討舊航照的著色而言，CycleGAN與CUT的著色成效比較優異，但處理多主體的舊航照(同時具有植被與建築物)時，顏色的分界不夠清晰，而pix2pix所生成之圖片幾乎沒有除了綠色與淡棕色以外的顏色，不論是河流、道路、建築物的顏色正確率都偏低。在部分圖片邊緣中，甚至可以見到圖片邊緣扭曲、有雜訊的情況。

1. 調整生成器與判別器權重帶來之影響

圖十、左邊兩張為pix2pix改變權重前後的生成器對抗損失變化，右邊則為

判別器的損失曲線的改變。

由各模型的損失曲線可以看到，增加生成器與判別器的權重比例，儘管生成器的損失依然呈現上升的趨勢，但減緩了許多，且判別器之對抗損失也不再直接歸零，這可以推論透過增加生成器的filter數量 (同時減少判別器的filter數)，能夠改善訓練後期，判別器的表現遠勝生成器成效，導致生成器的對抗損失大幅上升、及判別器的損失降至極低的現象。此外，在經過調整這項參數後，CycleGAN與CUT二模型生成之圖像品質均有不小的提升，在邊緣處理顯得較為不模糊，且套上的色彩較為不濃烈、穩定許多，而pix2pix生成之圖像色彩飽和度有略微提升。

1. 調整pix2pix生成器架構帶來的影響 (U-Net的有無)

使用U-Net的模型所生成之圖像具有顏色飽和度低、邊緣模糊、色調單一等缺點，但在將pxi2pix的生成器改為ResNet後，前述問題均有得到較為顯著的改善。

透過這項結果可以推論，比起使用U-Net進行影像的分割、生成，對於舊航照等分區不明顯 (道路、田多為淡灰色) 的影像，使用以殘差作為映射函數主要根據的ResNet作為生成器較為合適，生成之上色圖像也更接近實際的分布情況。

1. 改變pix2pix訓練方式對模型成效的作用

若訓練pix2pix直接對應灰階圖像與Ground truth，當圖片主體為建築物時，邊緣及圖片的角落會出現紅、藍、綠的單色雜訊，而且會出現單一建築物，卻出現2個顏色以上的不穩定情形。

若使用由L通道的數值，預測a與b兩圖層的方式，可以大幅降低前述的情形，也能提高圖片的質量與穩定度。

1. 調整GAN損失模式於對比式學習之作用

綜合而言使用LSGAN損失模式的CUT具有較好的性能，對於大多數的航照物件如河流、植被均有上到比較正確的顏色，而使用Vanilla進行生成的CUT，所生成的圖像中央有些會出現藍綠色的雜訊。至於若使用WGAN-GP對影像進行上色，則無法正確生成灰階影像原先物件的輪廓，填色也不夠穩定、充滿著雜訊。在訓練後期 (150~200個epoch) 時，LSGAN與Vanilla二種方式的NCE Loss與對抗損失分別呈現穩定的下降和上升。

而WGAN-GP的NCE Loss與對抗損失到後期出現負值，這可能與其訓練的方式有關，Wasserstein distance較常直接用於分析兩個高斯分佈（正態分佈）之間的關係，而非成對datasets很大一步導致了預測的不準確、模糊，由WGAN-GP生成的圖像也可驗證此項結果。

1. 採用Identity Loss對模型效能的影響

圖十一、

CUT與FastCUT

生成圖片細節上的差異

Identity Loss 對於採用對比式學習的CUT模型影響不是很大，在一些影像處理的細節，FastCUT (即未採用Identity Loss的CUT模型)勝過一般的CUT，如上色的多樣性較高、建築物輪廓較為明顯等。

因此，不採用Identity Loss，可以讓CUT模型生成影像的成效更好、更穩定。

1. 模型的效能評估

由前面幾點的結果圖片，及利用FID與IS數值化的影像品質。可以歸納出以下幾點推論:

* 1. CycleGAN與CUT二模型對於舊航照的上色成效較好，而pix2pix對於灰階處理的現代航照有比較穩定、優質的上色。
  2. 透過調整權重、採用 ResNet生成器、L\*a\*b映射訓練的pix2pix由舊航照所生成的上色影像穩定度、多樣性都較原模型來的高。
  3. 不採用Identity Loss可以讓CUT生成影像的品質提升，而其中又以使用LSGAN損失模式最為優異。
  4. 調整權重較能改善pix2pix與CycleGAN的上色效果，而對於CUT模型的成效則無顯著的影響。

1. **結論**

一、 以增加生成器卷積層上Filter的數量，調整生成器與判別器的權重能改善pix2pix與CycleGAN的上色效果，也能有效降低其對抗損失升高之趨勢，而對於CUT模型的成效則較無顯著影響。

二、使用L\*a\*b映射的訓練方式、採用ResNet生成器、增加生成器權重都能顯著改善pix2pix對於灰階歷史航照的上色效果。

三、CycleGAN與CUT在灰階歷史航照上色的各項表現較為優異、穩定。而不使用Identity Loss可以提升CUT模型生成影像的質量，而其中以搭配LSGAN損失模式最為優秀。

1. **參考文獻與資料**

【英文網路資源/論文】

[1] Taesung Park, Alexei A. Efros, Richard Zhang, Jun-Yan Zhu. Contrastive Learning for Unpaired Image-to-Image Translation. In ECCV 2020.

[2] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In CVPR 2020.

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Network. In CVPR 2017

[4] CycleGAN and pix2pix in PyTorch (<https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>)

[5] Contrastive-Unpaired-Translation (CUT) in PyTorch

(<https://github.com/taesungp/contrastive-unpaired-translation>)

[6] FID score for Pytorch (<https://github.com/mseitzer/pytorch-fid>)