

I7

利用GAN實現 歷史航空影像之上色

報告者：225 賴昱錡

指導老師：王鼎中 老師

指導學長：陳哲安 學長

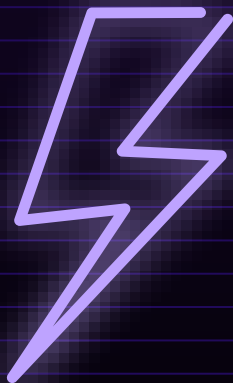
指導教授：中央大學 蔡宗翰 教授

建中數資39屆成果發表會

Luminescen

01

摘要



Abstract

Abstract

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。

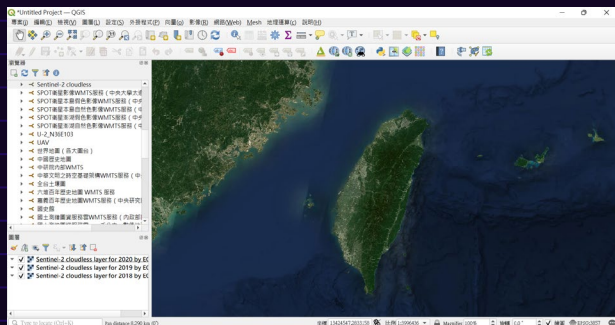
實驗後發現，透過增加生成器、判別器filter數量之比例，可以有效降低CycleGAN、pix2pix中對抗損失上升的現象，也讓二模型的成效提高。而ResNet比起U-Net能幫助pix2pix生成較鮮艷、穩定的圖像，以灰階映射 L^*a^*b 的方式訓練，可降低生成圖片的一些瑕疵。平均而言，CycleGAN與CUT在為舊航照的上色表現較pix2pix好，而其中CUT對於影像細節的處理又較為精緻、穩定。若欲採用CUT模型，不採用Identity Loss、搭配LSGAN的損失模式為佳。

02

研究動機&目的

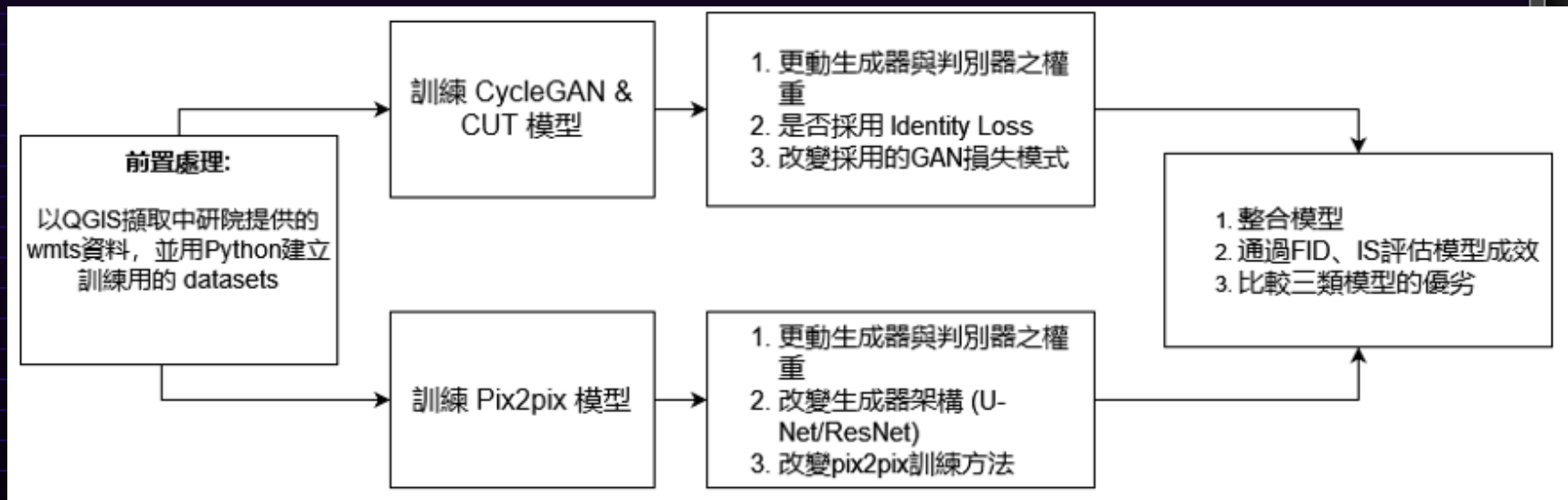


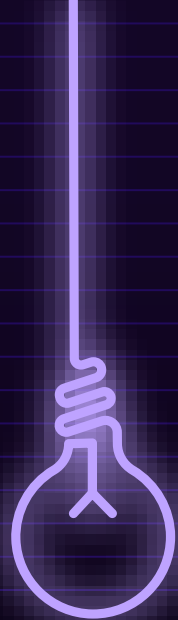
Motivation & Purpose



儘管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。

但灰階的舊航照始終會帶來分析地景時的不方便。因此，此研究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，期許能讓歷史航照的上色達到較佳的效果，並照到最佳的模型設定，讓模型得以更好的運用。





03 研究過程與方法

Method



建中數資39屆成果發表會

Luminescence

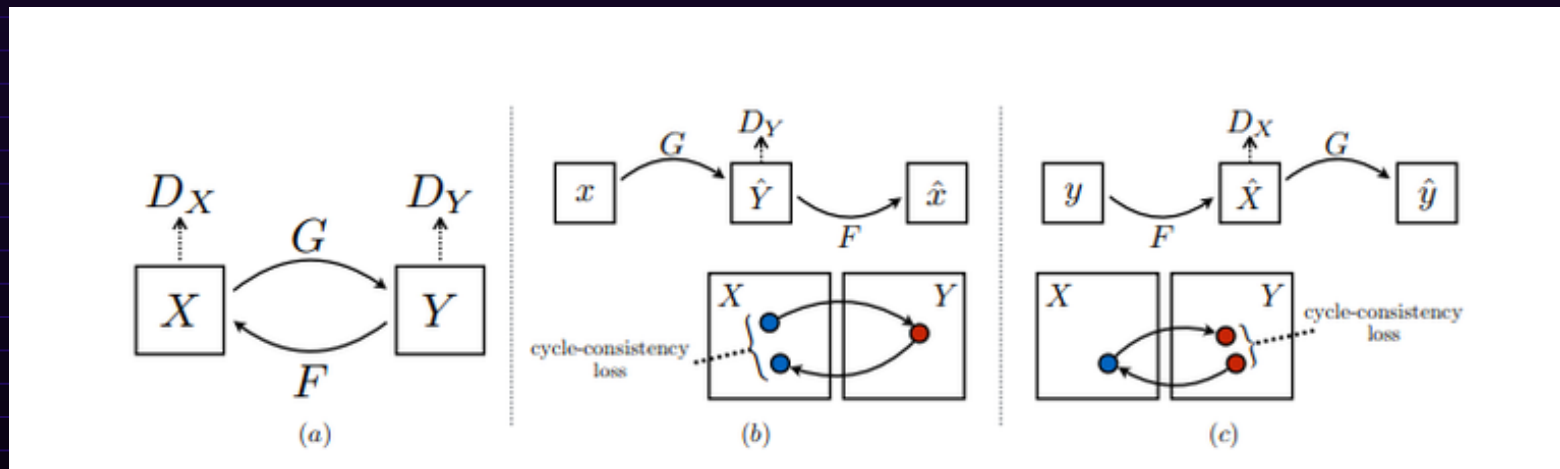


▲ 利用 QGIS 擷取中研院所
提供之正射影像、民初歷史航
照等 wmts 資料 (256*256)
(shapefile格式)



▲ 透過 Pillow、Shutil 等
套件，以原彩色圖像、灰階處
理影像建立訓練用的dataset

CycleGAN 架構 & 訓練方式



以非成對資料集作為輸入

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log D_Y(y)] \\ + \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log(1 - D_Y(G(x)))],$$

$$G^* = \arg \min_G \max_D L_{\text{cGAN}}(G, D)$$

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\|F(G(x)) - x\|_1] \\ + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\|G(F(y)) - y\|_1].$$

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ + \lambda \mathcal{L}_{\text{cyc}}(G, F),$$

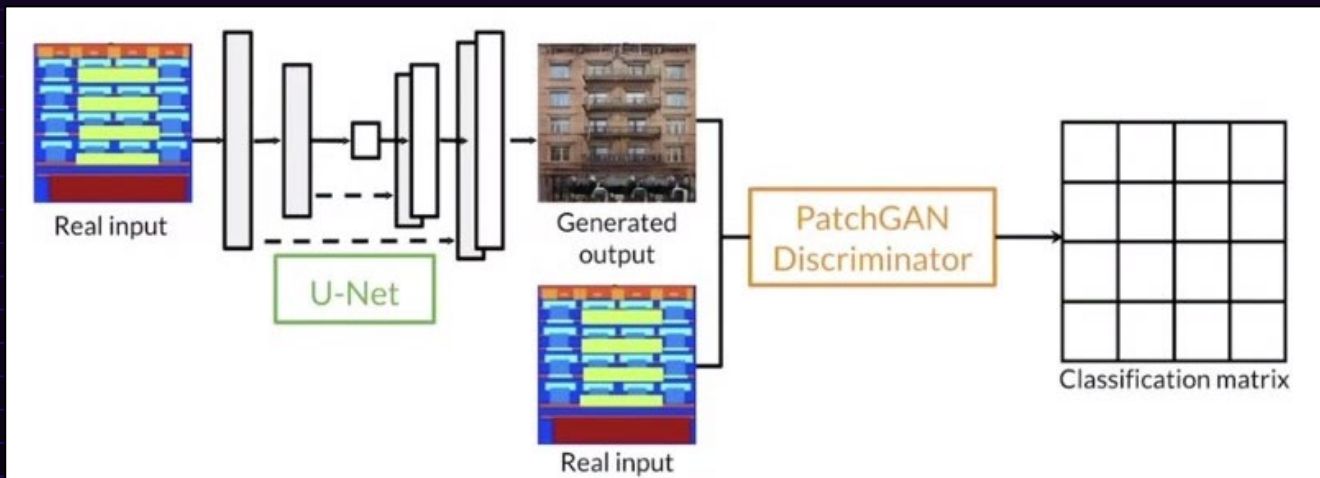
$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} L(G, F, D_X, D_Y)$$

$$L_{\text{Identity}}(G, F) = \mathbb{E}_{y \sim p_{\text{data}}(y)} [\|G(y) - y\|_1] + \mathbb{E}_{x \sim p_{\text{data}}(x)} [\|F(x) - x\|_1]$$

Identity Loss 的計算 (如上圖)

The Identity mapping loss help
preserve the color of input

Pix2pix 架構 & 訓練方式



以成對資料集作為輸入
(先將灰階處理影像、Ground
truth 合併為 256*512 的圖片)

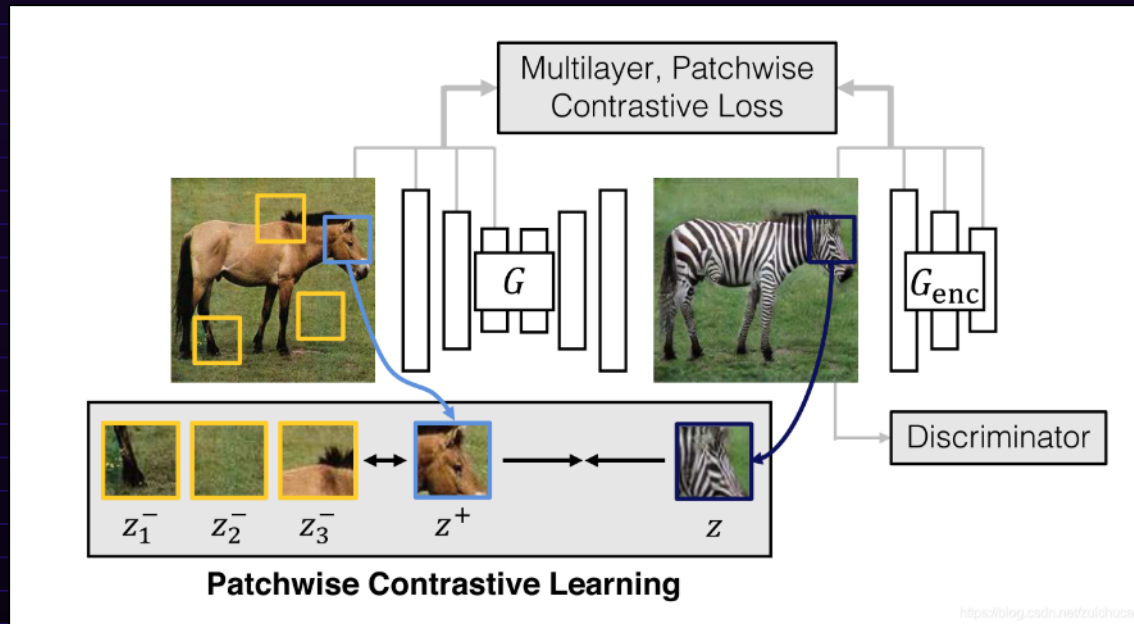
$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))],$$

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1].$$

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G).$$

CUT (Contrastive Unpaired Translation)



*以非成對資料集作為輸入

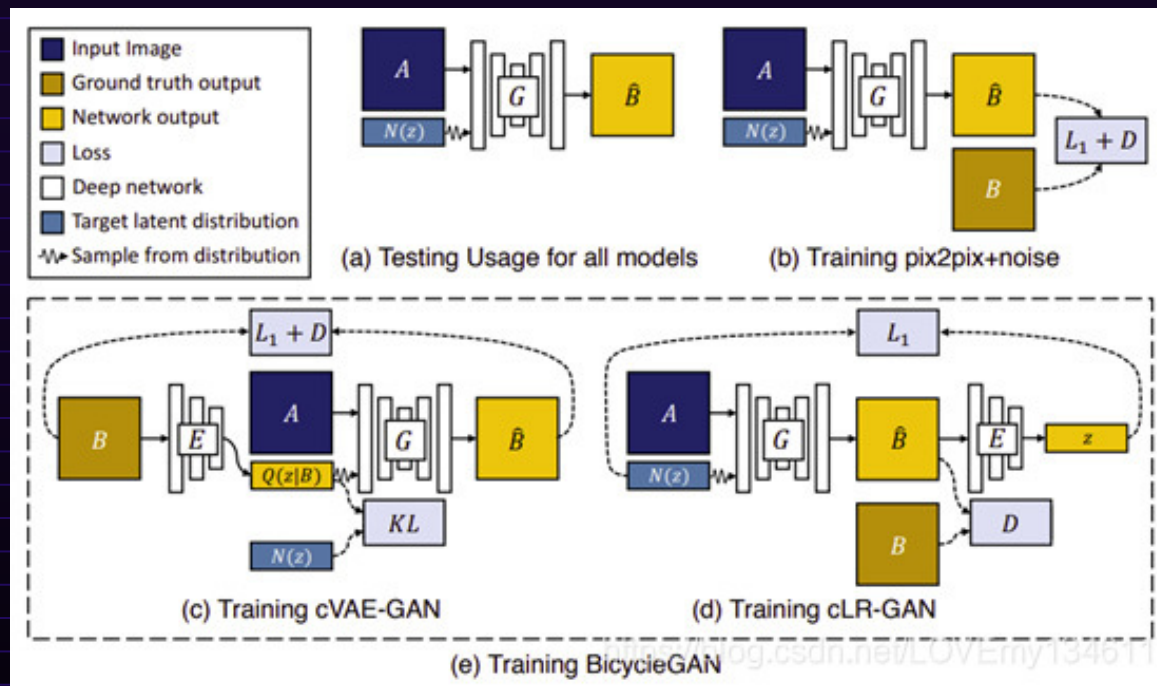
$$\ell(v, v^+, v^-) = -\log \left[\frac{\exp(v \cdot v^+ / \tau)}{\exp(v \cdot v^+ / \tau) + \sum_{n=1}^N \exp(v \cdot v_n^- / \tau)} \right]$$

$$\mathcal{L}_{\text{GAN}}(G, D, X, Y) = \mathbb{E}_{\mathbf{y} \sim Y} \log D(\mathbf{y}) + \mathbb{E}_{\mathbf{x} \sim X} \log(1 - D(G(\mathbf{x}))).$$

$$\ell(\mathbf{v}, \mathbf{v}^+, \mathbf{v}^-) = -\log \left[\frac{\exp(\mathbf{v} \cdot \mathbf{v}^+ / \tau)}{\exp(\mathbf{v} \cdot \mathbf{v}^+ / \tau) + \sum_{n=1}^N \exp(\mathbf{v} \cdot \mathbf{v}_n^- / \tau)} \right].$$

$$\mathcal{L}_{\text{GAN}}(G, D, X, Y) + \lambda_X \mathcal{L}_{\text{PatchNCE}}(G, H, X) + \lambda_Y \mathcal{L}_{\text{PatchNCE}}(G, H, Y).$$

BicycleGAN structure



cLR-GAN:
Conditional Latent Regressor GAN

生成影像的評估指標：FID與IS (參考inception v3模型)

$$\text{FID}(x, g) = \|\mu_x - \mu_g\| + \text{Tr} \left(\Sigma_x + \Sigma_g - 2\sqrt{\Sigma_x \Sigma_g} \right)$$

FID score (Frecht Inception Score)

同時，較低的FID score代表著生成影像的質量、穩定性較佳。

→ 針對生成影像與Ground truth進行比較

生成影像的評估指標：FID與IS (參考inception v3模型)

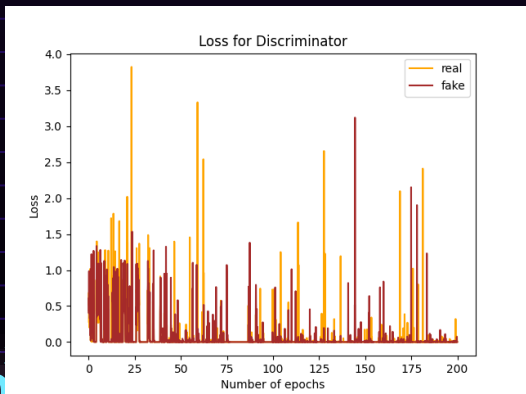
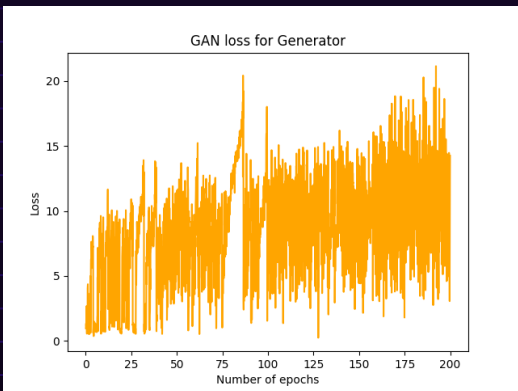
$$\sum_{x \in G} P(x) \sum_{i=1}^{1000} P(y_i|x) \log \frac{P(y_i|x)}{P(y_i)} \\ = E_{x \sim p_G} KL(p(y|x) || p(y))$$

$$IS = \exp E_{x \sim p_G} KL(p(y|x) || p(y))$$

IS (Inception Score): 較高的IS表示生成的
圖片分布與Ground truth 越接近。
→ 直接針對生成影像的分布進行評估

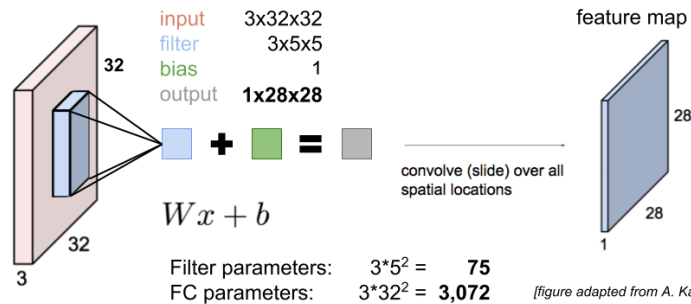
改變一：改變生成器與判別器之權重

- 1. 調整生成器末端(卷積層)與判別器首端的Filter數量。
- 2. 增加生成器的Learning rate



Convolution

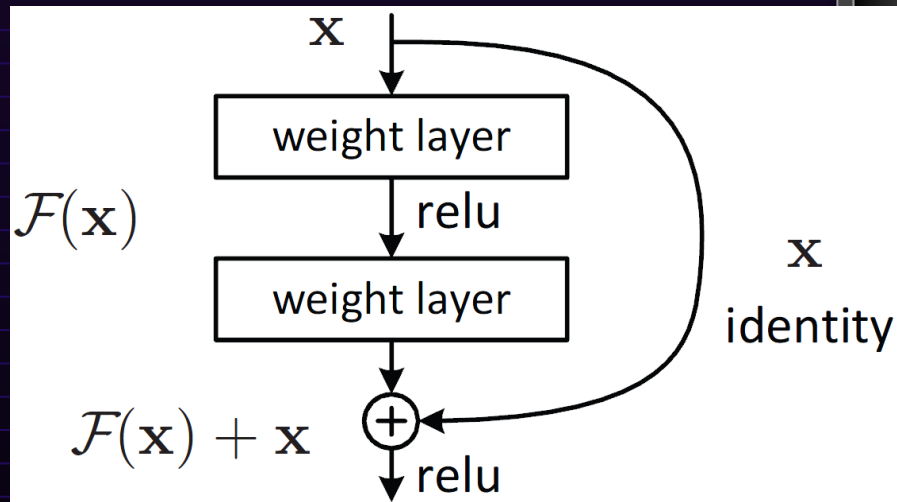
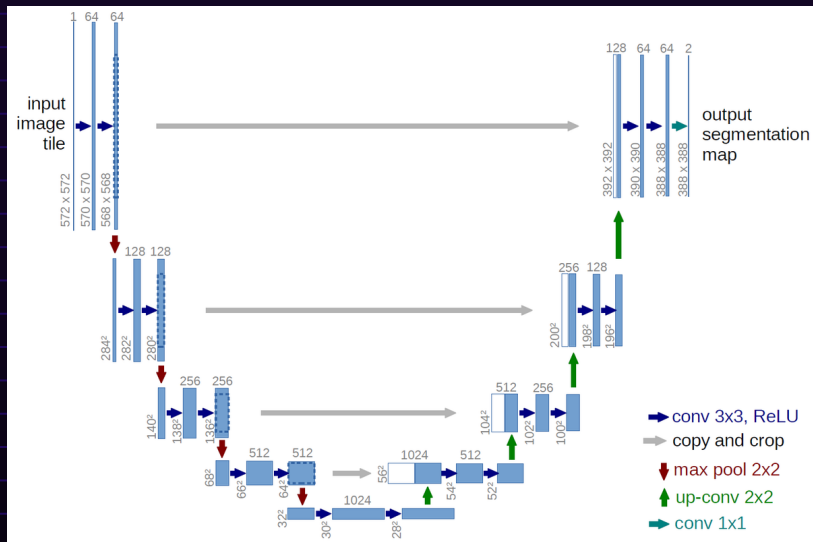
Convoluting the filter with the input gives a **feature map**.



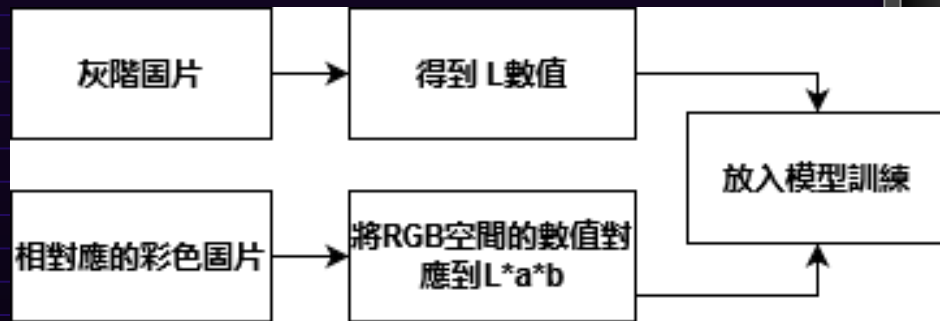
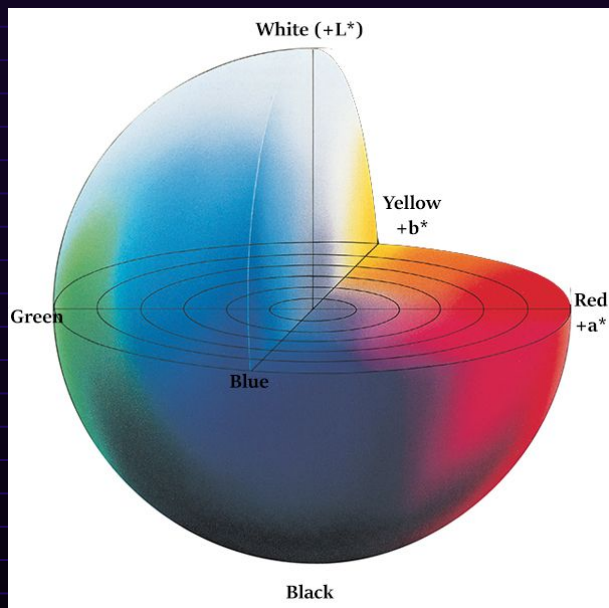
→為解決訓練後期，判別器的性能遠勝生成器，進而使生成器對抗損失大幅上升的現象。

改變二: 改變pix2pix生成器架構 (U-Net/ResNet)

左右兩圖分別為U-Net、ResNet的架構



改變三: 改變pix2pix訓練方式 (映射的模式)



改變四：改變CUT的GAN損失模式：除最原始的Vanilla GAN外，也嘗試了LSGAN、WGAN-GP等損失模式來訓練模型（上/下圖）

$$\begin{aligned}\min_D V_{\text{LSGAN}}(D) &= \frac{1}{2} \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [(D(\mathbf{x}) - b)^2] + \frac{1}{2} \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [(D(G(\mathbf{z})) - a)^2] \\ \min_G V_{\text{LSGAN}}(G) &= \frac{1}{2} \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [(D(G(\mathbf{z})) - c)^2],\end{aligned}$$

$$L = \underbrace{\mathbb{E}_{\tilde{\mathbf{x}} \sim \mathbb{P}_g} [D(\tilde{\mathbf{x}})] - \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_r} [D(\mathbf{x})]}_{\text{Original critic loss}} + \lambda \underbrace{\mathbb{E}_{\hat{\mathbf{x}} \sim \mathbb{P}_{\hat{\mathbf{x}}}} [(\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}})\|_2 - 1)^2]}_{\text{Our gradient penalty}}.$$

→ 討論不同損失模式下對CUT模型帶來的作用

改變五:是否採用Identity Loss

CUT模型的總損失公式可以寫成以下的數學形式，其中 λ_X 與 λ_Y 為控制兩者比重的係數， λ_Y 同時也代表著Identity Loss。

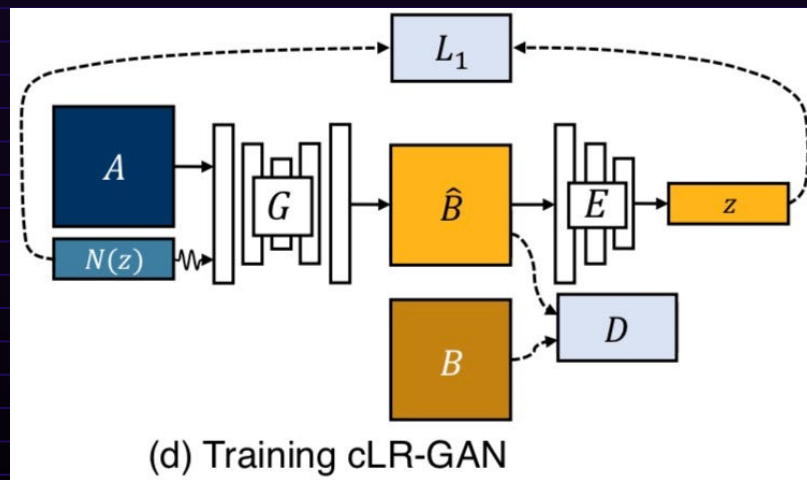
而Identity loss在CycleGAN中的作用是保留輸入和輸出顏色組成的一致性。

$$\mathcal{L}_{\text{GAN}}(G, D, X, Y) + \lambda_X \mathcal{L}_{\text{PatchNCE}}(G, H, X) + \lambda_Y \mathcal{L}_{\text{PatchNCE}}(G, H, Y).$$

改變六：潛在編碼的加入時機

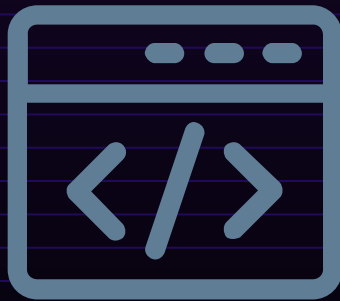
此研究中，我們嘗試改變 z 輸入給Generator的位置(首段、中段…等位置)。

→ 欲探討潛在編碼在何時加入生成器網路能讓歷史航照上色達到較好的效果。



04

研究結果&討論



Results & Discussion

一、預設模型上色成效的評估



Pix2pix

1. pix2pix對於灰階處理
航照的套色效果最好，對
於舊航照的著色而



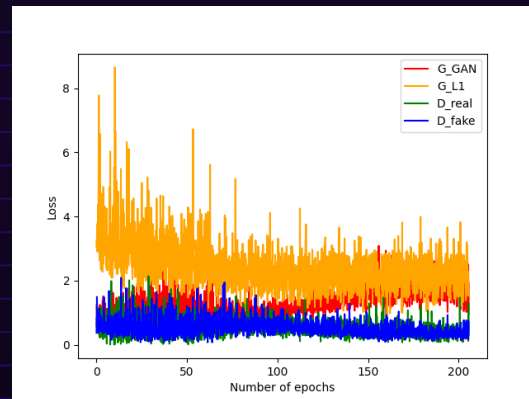
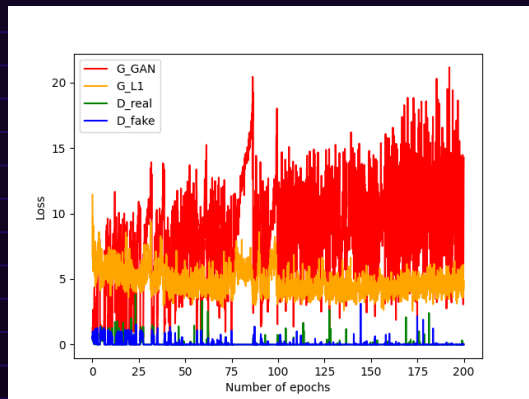
CycleGAN

2. 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。



CUT

二、調整生成器與判別器權重帶來的影響



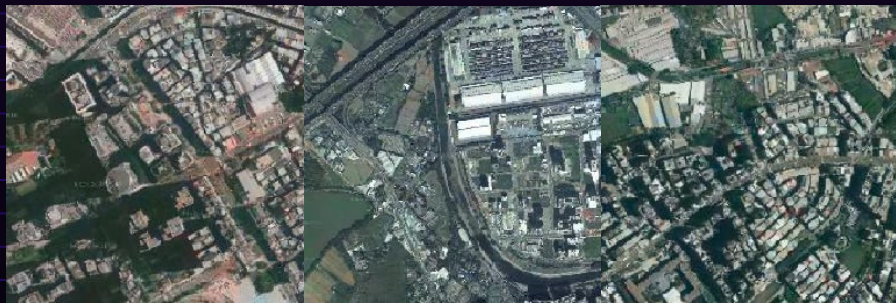
→ 透過增加生成器的filter數量，能夠改善訓pix2pix、CycleGAN練後期，生成器的對抗損失大幅上升的現象

三、改變pix2pix生成器架構帶來的影響



左側為使用U-Net作為生成器架構，右側則使用ResNet
→ 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。

四、改變pix2pix訓練方式帶來的影響



→ 若使用由L通道的數值(即為灰階圖片)，預測 a (藍綠色)與 b (黃紅色)兩圖層的方式，可以大幅降低邊緣出現藍綠雜訊的情形，也能提高圖片的質量與穩定度。

五、改變GAN損失模式為對比式學習帶來的影響



→ 由左至右分別為使用LSGAN、Vanilla、WGAN-GP等GAN損失模式針對歷史航照的上色效果，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

六、Identity Loss對上色成效的影響



CUT(上圖)、FastCUT(下圖)於歷史航照的上色效果比較
→ 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。

七、各生成模型成效之評估

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

七、各生成模型成效之評估

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

七、各生成模型成效之評估

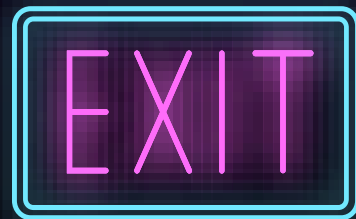
CUT (註：原CUT、FastCUT以LSGAN為GAN損失模式)					
CUT模式		採用模型	FID	IS最大值	IS 平均值
採用Identity Loss	以灰階處理之航照作輸入	原CUT	145.99	13.23	3.34
		ngf=128, ndf=32	165.76	16.23	3.10
		改變GAN架構 (Vanilla)	168.22	15.22	3.03
	以舊航照作輸入	原CUT	163.55	17.64	3.41
		ngf=128, ndf=32	160.14	12.25	3.07
		改變GAN架構 (Vanilla)	167.11	14.39	2.91

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

05.

結論

Conclusion



建中數資39屆成果發表會

Luminescen

Conclusion

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。

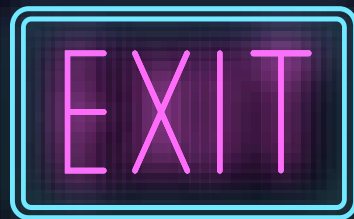
二、使用L*a*b映射的訓練方式、採用ResNet生成器、增加生成器權重都能顯著改善pix2pix對於灰階歷史航照的上色效果。

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。

06.

展望

The Future



建中數資39屆成果發表會

Luminescen

The Future

- 一、設計結合預訓練模型的 Web App，讓歷史航照的上色技術得以更加容易地讓人運用。
- 二、對資料集進行更好的預處理，如對影像中的主體標上 label、去模糊等，供模型生成更加準確的影像
- 三、讓本研究能繼續應用於年代更早的航空影像，並達到修復不全影像的效果。
- 四、能去更了解各改變能改善生成影像質量的原因。
- 五、研究 multimodal image translation 的相關模型。

特別感謝

- 中研院人社中心-廖泫銘研究副技師、實驗室的陳哲安學長
- 中央大學-蔡宗翰教授
- 建中王鼎中老師
- 成發公關網管組的同學們
- 協助這場成發順利完成的每個人和觀眾。

The end of my
presentation.

感謝各位的聆聽!

