

# AI CUP 2024 春季賽

## 以生成式 AI

### 建構無人機於自然環境偵察時所需之導航資訊

### 競賽 I — 影像資料生成競賽

### 競賽報告

隊伍：TEAM\_5565

隊員：楊棋凱(隊長)、袁杰翔、蔡孟宸、李天佑

Private leaderboard：111.3347 / Rank11

## 壹、環境

作業系統：Windows 10

程式語言：Python 3.10 (開發環境：Visual Studio Code)

使用套件：CUDA11.6, PyTorch1.12.1, Torchvision0.13.1, torchaudio0.12.1, NumPy, Dominate 2.9.1, Visdom 0.2.4, Weights and Biases 0.17.0

作業系統：Windows10

程式語言與開發環境：使用 Python 3.10 作為主要的程式語言，開發環境是 Visual Studio Code，它提供了一個互動式的開發介面，方便進行代碼編寫、測試和可視化。

主要套件和函式庫：PyTorch1.12.1：主要機器學習框架，支援 GPU 加速，使模型訓練更加高效。

Cuda 11.6：用於 GPU 加速的庫，配合 PyTorch 使用，以充分利用硬體資源。

NumPy：用於數據處理和視覺化的基本函式庫。

Weights and Biases：用於創建和操作 HTML 文檔，包括動態生成 HTML 頁面等。

Visdom：視覺化工具，用於創建、組織和共用即時豐富資料的視覺化。

Dominate：用於創建和操作 HTML 文檔，包括動態生成 HTML 頁面等。

預訓練模型：在我們的項目中，我們並沒有使用預訓練模型，而是完全自行訓練我們的模型。

額外資料集：我們有使用 AI cup demo code 自帶的道路資料集，其他的訓練和測試數據都來自於主辦單位提供的標準數據集。

## 貳、演算法與模型架構

因為我們沒有接觸過生成式模型,對於 cGAN[1]也不熟悉,因此我們只使用了官方推薦的 pix2pix[2]來進行了訓練,為了處理不同的環境我們對河流與陸地使用了不同的生成器  $G$  和判別器  $D$ 。

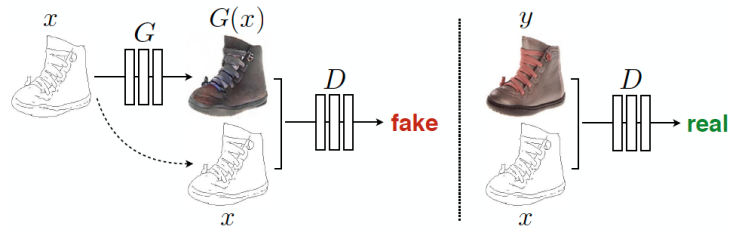
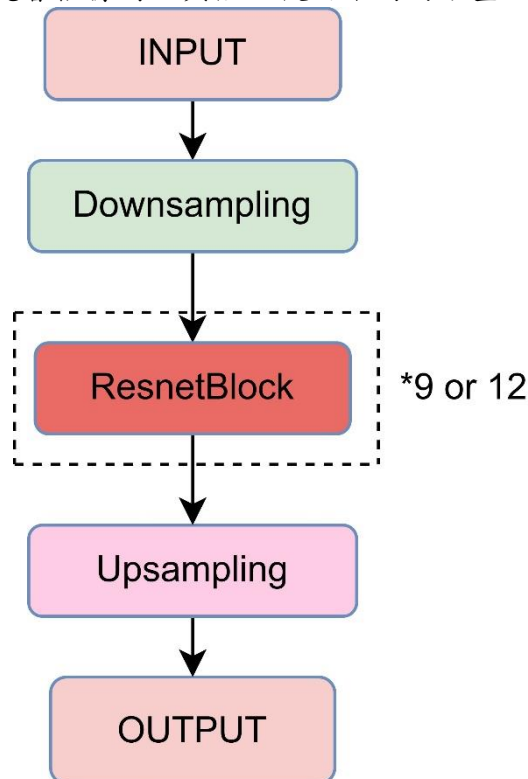


Figure 2: Training a conditional GAN to map edges→photo. The discriminator,  $D$ , learns to classify between fake (synthesized by the generator) and real {edge, photo} tuples. The generator,  $G$ , learns to fool the discriminator. Unlike an unconditional GAN, both the generator and discriminator observe the input edge map.

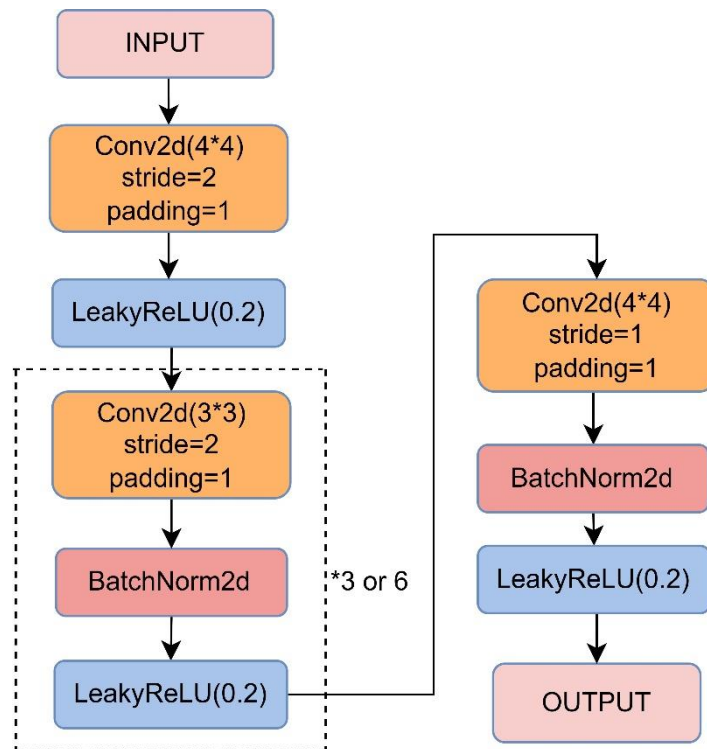
圖一、cGAN 工作原理

生成器模型的整體架構如圖二所示,是一個非常經典的生成器模型,其中的 ResnetBlock 模塊會根據河流與陸地使用不同的數量。



圖二、生成器模型架構

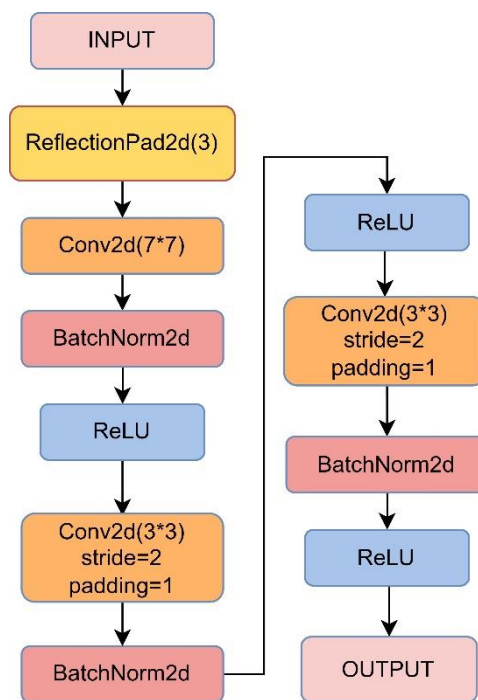
判別器模型的整體架構如圖三所示,其中也有下採樣模塊會根據河流與陸地使用不同的數量。接下來對架構進行進一步的說明：



圖三、判別器模型架構

### 1. Downsampling模塊：

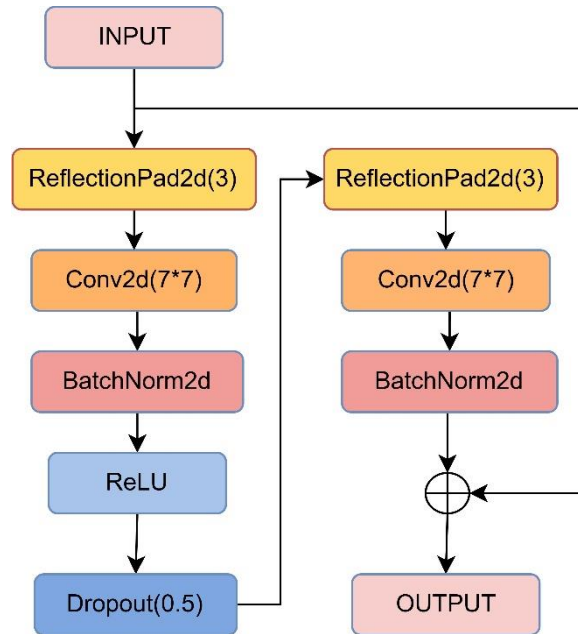
圖四為 Downsampling 架構，這裡通過 Conv2D 進行卷積操作將輸入圖片下採樣縮小了 4 倍，方便提取圖片的特徵值。



圖四、Downsampling架構

## 2. ResnetBlock 模塊：

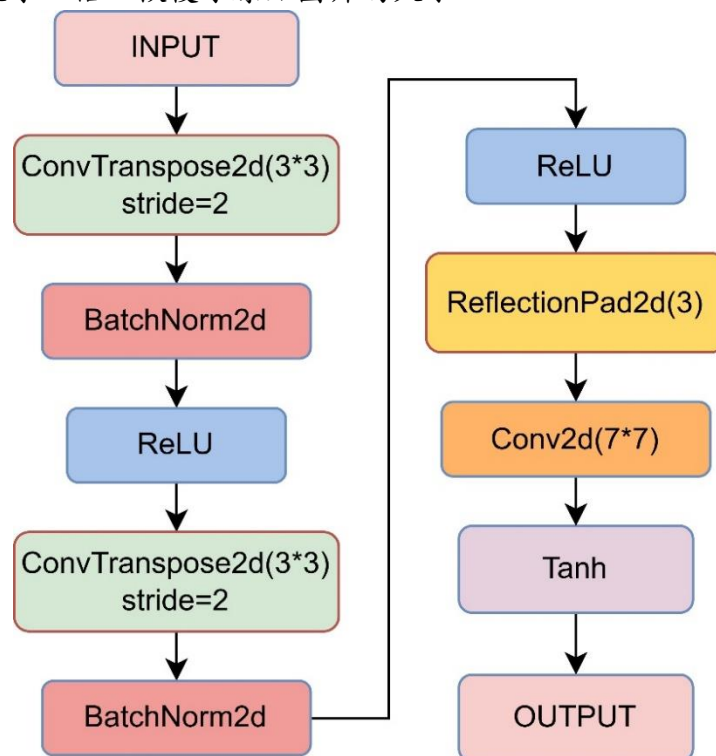
圖五為 ResnetBlock 架構，利用 ReflectionPad2d 與 Conv2D 的組合來提取圖片的特徵值。



圖五、ResnetBlock 架構

## 3. Upsampling 模塊：

圖六為 Upsampling 架構，利用 ConvTranspose2d 進行逆卷積操作將輸入圖片上採樣放大了 4 倍，恢復了原始圖片的大小。



圖六、Upsampling 架構

## 參、技術模型原創性或改良成效

我們在處理圖像大小調整、優化模型時，遇到了一些問題。將模型輸出的 256\*256 大小圖片調整至 428\*240 的標準尺寸後，我們本期待的是模型效能會有所提升，但結果並不如人意，反而讓成績下滑了。我們原猜想，這可能是由於官方評分機制的問題。然而深一層的測試發現，我們在進行圖像大小調整（resize）的時候，引用了不同的方法，導致處理出來的效果不一致。具體來說，我們使用的兩種庫，cv2 和 PIL，處理縮放圖像的的演算法實現方式並不相同，經由實測驗證，結果是 PIL 庫處理出來的效果會讓最終成績有所提升，更適合這一次的比賽。

## 肆、資料分析與處理過程

我們發現 AI cup demo code 程式自帶一份道路資料集（資料集 2），在使用雜湊值確認與原數據集（資料集 1）沒有相同的圖片後，我們將兩份道路資料集合在了一起，並且統一將 label 圖片改為白底黑線的方式來對模型進行訓練。

數據集中有不少模糊的照片原來想要通過 Real-ESRGAN[3]模型來增強圖片，但發現訓練效果反而變差了，我們也有嘗試對模型輸出的圖片直接進行增強依舊評分下滑，猜測可能是因為評分方式是與原圖進行比較，但原圖本身模糊因此反而不適用於本次比賽。



圖七、Real-ESRGAN 模型效果對比

## 伍、訓練方式

優化器：預設使用的是 Adam 優化器，一種廣泛使用的最佳化演算法，特別適用於深度學習。Adam 結合了動量（Momentum）和 RMSprop 的優點，能夠自動調整學習率，因此在許多情況下表現良好。

學習率：使用預設的初始學習率 0.0002 和 linear 線性學習率調整。

損失函數：使用均方誤差（MSE）損失來度量生成圖片與真實圖片之間的差異。

我們也有嘗試使用官方給於的預訓練權重，但是在訓練後發現效果並不如自行訓練出來的分數高，還有嘗試使用不同的生成式網路，例如 wgangp 等但效果都不是很好。因為河流與道路採用了不同的模型，因此需要分開訓練，我們的具體訓練指令如下：

河流模型訓練指令：`python pytorch-CycleGAN-and-pix2pix-master/train.py --dataroot ./pytorch-CycleGAN-and-pix2pix-master/datasets/ROAD_pix2pix --name ROAD_pix2pix --model pix2pix --direction AtoB --netG resnet_12blocks`

道路模型訓練指令：python pytorch-CycleGAN-and-pix2pix-master/train.py --  
 dataroot ./pytorch-CycleGAN-and-pix2pix-master/datasets/ROAD\_pix2pix --  
 name ROAD\_pix2pix --model pix2pix --direction AtoB --netG resnet\_9blocks -  
 -netD n\_layers --n\_layers\_D 6

## 陸、結果分析與結論

我們先使用 unet\_256[4]模型，發現效果不佳，便開始嘗試使用 ResNet[5]模型，發現準確率有明顯提升後便開始更改模型其他細節，成員提出 resize 到標準大小後成績下降不合理，便開始嘗試各種 resize 方式，例如 cv2 的各種插值方式但發現成績還是 PIL 的預設插值方式效果最好。在嘗試過修改圖片後有了圖像增強的想法，開始嘗試使用 Real-ESRGAN 來加強圖片，但效果不好便沒有繼續使用了。

表一、不同模型與 resize 方式的訓練結果比較

河流模型		道路模型	訓練 擴充	圖片大小 resize 方式	Public 分數	Private 分數
G	unet_256	unet_256	無	256*256	131.6411	無
D	basic	basic				
G	unet_256	unet_256	無	428*240(cv2)	132.3565	無
D	basic	basic				
G	unet_256	unet_256	有	256*256	122.0197	無
D	basic	basic				
G	resnet_9blocks	unet_256	有	428*240(PIL)	115.2449	無
D	basic	basic				
G	resnet_9blocks	resnet_9blocks	有	256*256	113.8731	無
D	basic	n_layers_D=6				
G	resnet_9blocks	resnet_9blocks	有	428*240(PIL)	111.7545	無
D	basic	n_layers_D=6				
G	resnet_12blocks	resnet_9blocks	有	428*240(PIL)	110.0217	111.3347
D	basic	n_layers_D=6				

結論：

在整體優化的過程中我們嘗試了各種不同的方式比如使用預訓練權重、使用不同的生成器判別器模型或是使用 Real-ESRGAN 來加強圖片，雖然在此之前沒有使用過 PyTorch 套件以及創建生成式模型的經驗，但我們依舊成功提升了模型的性能，儘管模型已經取得一定的成功，我們仍然認為還有進一步改進的空間，未來的方向可能包括調整 ResNet 的連接方式，嘗試不同的模型，以及探索不同的輸入格式處理方式。

此外，如果未來有機會，我們將透過這次比賽的經驗，進一步改良模型的架構，期待之後不同比賽能夠實現更出色的成績。

## 柒、程式碼

Github 連結：

<https://github.com/YJXsirius/AI-CUP-2024-spring.git>

訓練資料：

<https://drive.google.com/drive/folders/1MRxT6C3Ub8FMBtniSotfoHlMUbsXdvWu?usp=sharing>

訓練權重：

[https://drive.google.com/drive/folders/1TGF4QNO0dQ\\_MQgGqblX3NNQfwpPL7hJg?usp=sharing](https://drive.google.com/drive/folders/1TGF4QNO0dQ_MQgGqblX3NNQfwpPL7hJg?usp=sharing)

## 捌、使用的外部資源與參考文獻

[1]Mirza, M., & Osindero, S. (2014). Conditional Generative Adversarial Nets. Retrieved from <https://doi.org/10.48550/arXiv.1411.1784>

[2]Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2016). Image-to-Image Translation with Conditional Adversarial Networks. Retrieved from <https://doi.org/10.48550/arXiv.1611.07004>

[3]Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. Retrieved from <https://doi.org/10.48550/arXiv.2107.10833>

[4] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Retrieved from <https://doi.org/10.48550/arXiv.1505.04597>

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. Retrieved from <https://doi.org/10.48550/arXiv.1512.03385>



## 報告作者聯絡資料表

隊伍名稱	TEAM_5565	Private Leaderboard 成績	111.3347	Private Leaderboard 名次	11
身分 (隊長/隊員)	姓名 (中英皆需填寫) (英文寫法為名, 姓, 例: Xiao-Ming, Wu, 名須加連字號, 姓前須加逗號)	學校+系所中文全稱 (請填寫完整全名, 勿縮寫)	學校+系所英文中文全稱 (請填寫完整全名, 勿縮寫)	電話	E-mail
隊長	楊棋凱 Qi-Kai, Yang	國立勤益科技大學電機工程系	National Chin-Yi University of Technology Department of Electrical Engineering	0966-353-066	ki9247ki@gmail.com
隊員 1	袁杰翔 Jie-Xiang, Yuan	國立勤益科技大學電機工程系	National Chin-Yi University of Technology Department of Electrical Engineering	0963-067-729	c0963067729@gmail.com
隊員 2	李天佑 Tian-You, Li	國立勤益科技大學電機工程系	National Chin-Yi University of Technology Department of Electrical Engineering	0917-784-777	yoyo20021202@gmail.com
隊員 3	蔡孟宸 Meng-Chen, Cai	國立勤益科技大學電機工程系	National Chin-Yi University of Technology Department of Electrical Engineering	0902-115-520	jack098220207277@gmail.com



指導教授資料					
每隊伍至多可填寫兩名	指導教授 中文姓名	指導教授 英文姓名 (英文寫法為名,姓, 例：Xiao – Ming, Wu，名須加連字號， 姓前須加逗號)	任職學校＋系所 中文全稱 (請填寫完整全名，勿 縮寫)	任職學校＋系 所 英文全稱 (請填寫完整全名， 勿縮寫)	E-mail
教授 1	葉政育	Cheng-Yu, Yeh	國立勤益科技大學 電機工程系	National Chin-Yi University of Technology Department of Electrical Engineering	cy.yeh@ncut.edu.tw

★註 1：請確認上述資料與 AI CUP 報名系統中填寫之內容相同。自 2023 年起，獎狀製作將依據報名系統中填寫內容為準，有特殊狀況需修正者，請主動於報告繳交期限內來信 moe.ai.ncu@gmail.com。報告繳交截止時間後將不予修改。

★註 2：上傳程式碼檔案與報告至程式碼託管平臺或是雲端等網站，將相關連結寄信至：[jamesouo@g.nccu.edu.tw](mailto:jamesouo@g.nccu.edu.tw)，並同時副本至：[t\\_brain@trendmicro.com](mailto:t_brain@trendmicro.com) 與 moe.ai.ncu@gmail.com，缺一不可。