2021 年臺灣國際科學展覽會
優勝作品專輯

作品編號　**190038**

參展科別　電腦科學與資訊工程

作品名稱　利用 **VAE-pix2pix** 生成擬真的山脈模型

得獎獎項　大會獎 四等獎

就讀學校　桃園市立武陵高級中等學校

指導教師　蘇木春、劉思德

作者姓名　程品奕、李杰穎

關鍵詞　類神經網路、生成對抗網路、山脈地形

作者簡介

李杰穎、程品奕為桃園市立武陵高中科學班學生，目前就讀高三。於高二時於中央大學蘇木春教授的指導下進行專題研究，並以「利用生成對抗網路生成擬真的山脈地形」參加第六十屆全國中小學科學展覽會榮獲大會獎第二名，作者李杰穎以此題目參加第十九屆旺宏科學獎榮獲最高榮譽「旺宏獎」。改進原作品後，以「利用 VAE-pix2pix 生成擬真的山脈模型」參加 2021 台灣國際科展。

摘要

本研究利用 NASA 的 SRTM 1 Arc-Second 資料集來收集全球各地的地形高度圖(heightmap)，也利用 MapTiler 網站收集相對應的衛星空照圖，用這些收集的圖像，訓練我們建構的 VAE-pix2pix 模型。VAE-pix2pix 為 Variational Autoencoder (VAE)及 pix2pix (為一個 Conditional Generative Adversarial Network)結合的模型，能將人工繪製的高度圖加上真實山脈應有的細節(包含尖銳的山脊、山壁上的紋路、連續的河流網路等……)，並生成出相對應的擬真衛星空照圖。相較於原 pix2pix 模型，VAE-pix2pix 所生成的高度圖及衛星空照圖會更接近於真實世界的地形高度圖及衛星空照圖，同時 VAE-pix2pix 模型也能透過改變 latent code 的數值來生成出不同風格的高度圖及空照圖，如地貌的顏色或雪線的高度等，這些都增加模型生成圖像的多樣性。為了使我們建構的模型能更廣泛的被應用，我們在 Unity 上開發了 Unity 客戶端，其生成的 mesh 可以讓使用者直接應用於遊戲的場景，簡化了遊戲中生成擬真山脈模型的任務。

ABSTRACT

In this study, we use NASA's SRTM 1 Arc-Second dataset to collect altitude maps from around the world, and we also use MapTiler to collect corresponding satellite images. Using these collected images, we trained our VAE-pix2pix model, which is a Variational Autoencoder (VAE) combined with pix2pix (a Conditional Generative Adversarial Network). VAE-pix2pix can add details of the real-world mountain should have (including sharp ridges, mountain wall textures, continuous river networks, etc.) to the heightmap, users draw which. Our model can generate the corresponding satellite images as well. Compared with the original pix2pix model, our model can generate heightmap and satellite images that are more realistic. It can also generate different styles of heightmap and satellite images by changing the value of the latent code, such as the color of the landform or the height of the snow line. This increases the diversity of the images generated by the model. To make our model can be better used, we have developed a client on Unity, which can generate a mesh that allows users to directly use it when developing games in Unity. In conclusion, our work has simplified generating a realistic mountain model in the game or other fields as well.

壹、前言

一、研究動機

隨著 3C 的普及，遊戲已經成為現代人打發時間、舒壓及社交的必需品；隨著科技技術的進步，對於遊戲畫質的要求也越高，而在製作各種遊戲時，常常會需要生成擬真的地形作為遊戲的場景。

傳統上，遊戲的擬真山脈地形是透過人工繪製。在將大致的架構畫出來後，還需花費不少時間捏出山脊和挖出河流等細節部分。近年來，人工智慧演算法在圖像的生成上有重大的突破，不論是生成圖片或是將影像的風格提取出來，並轉換到另一張影像，都已經是可行的方法，因此本研究希望簡化人工繪製的過程，透過訓練生成對抗網路來達到生成擬真山脈地形的成果。

二、研究目的

本研究期望能簡化遊戲製作者在生成擬真山脈地形模型的流程，同時確保生成擬真山脈的效果。在製作 3D 地形模型時，需要**高度圖(heightmap)**來指定地表的形狀，以及**紋理貼圖(texture)**來指定地表的顏色，傳統的方法是需要設計師一筆一畫來畫出山脈的細節及風格。本研究期望能建構出一個類神經網路模型，自動畫出山脈的細節及風格。在輸入圖像部分，使用者不需繪製山脈的細節及風格，只需要繪製一張人工手繪的**大致地形架構**並提供一組**風格參數**，將其輸入訓練好的 VAE-pix2pix 模型，即可生成出細節豐富的地形高度圖和紋理貼圖，供建立 3D 地形模型時使用。

貳、研究過程與方法

一、文獻探討

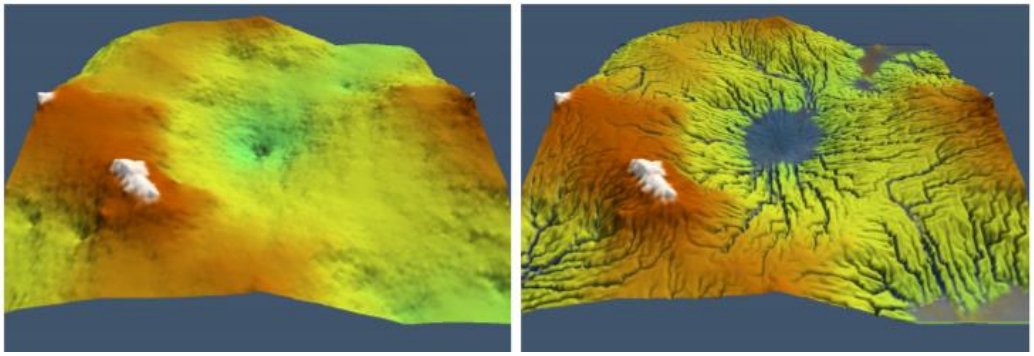
(一) 物理侵蝕模型

一般來說，若要提升遊戲中山脈地形的真實度，其中一種方法為使用物理的侵蝕模型，如論文[6]，其方式為建構一個物理模型，模擬水流在地形上侵蝕與堆積，論文作者藉由此模型結合 GPU 的運算，實現改變地形樣貌的效果。

根據論文[6]內容所敘述的物理模型，本研究利用 PyTorch 來實現論文[6]所述之水流侵蝕模型，之後會將其與經過訓練的 VAE-pix2pix 模型進行真實度及實用性的比較。

此物理侵蝕模型將地表分成正方形的網格，使用歐拉法求地表上每格的水深、含沙量和網格間的水流速，並根據水量和流速進行侵蝕和堆積，疊代多次後可得出侵蝕一段時間後的地面和水面高度圖。一次疊代的步驟如下：

1. 在每個網格加上等量的水，模擬均勻的降雨
2. 更新流速(加速度受坡度和阻力影響)
3. 根據流速讓水流到鄰近的格子，同時搬運等比例的砂土
4. 根據水量和流速進行侵蝕，增加水中含沙量，降低地面高度
5. 將水中一定比例的沙土堆積到地面
6. 移除每格一定比例的水，模擬蒸發



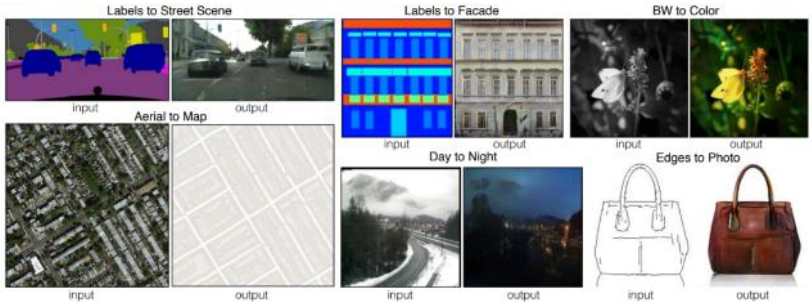
侵蝕前　　侵蝕後

圖一、物理侵蝕模型的效果 (取自[6])

(二) pix2pix 模型

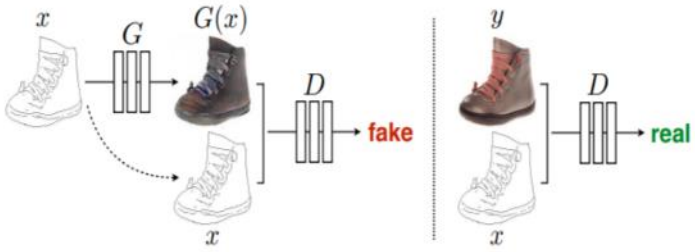
論文[5]則為 pix2pix，是一個 Conditional generative adversarial network (又稱 Conditional GAN)。pix2pix 的用途是把圖像轉換成一種特定的風格，或是依據邊緣或色塊等標籤生成擬真圖像。

pix2pix 訓練時需要成對的影像資料，而模型的目標則是將第一張圖片轉換為第二張圖片。例如圖二，即為 pix2pix 可做到的各種應用，而 pix2pix 模型的目標是將左圖做為輸入，輸出右邊的圖像。



圖二、pix2pix 可做到的圖像風格轉換(取自[5])

pix2pix 模型由 generator 和 discriminator 兩個部分組成。generator 的結構為 U-Net，訓練時會嘗試把輸入圖像轉換為目標圖像。discriminator 則是一個分類器，訓練時會嘗試分辨哪些圖是 generator 生成的圖，哪些是真的目標圖像。兩者會同時訓練，generator 生成的圖越不容易被 discriminator 分辨出來，就代表 generator 表現得越好。利用這點來訓練 generator，就能讓它的輸出盡可能的真實。

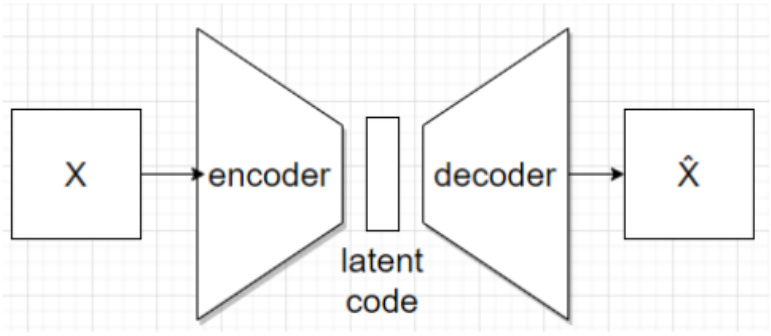


圖三、訓練 Conditional GAN 將鞋子的邊緣圖生成實際鞋子的圖像。(取自[5])

如圖三中，discriminator 的任務是辨認出哪些圖片是由 generator 所生成(如左)，哪些是原始圖像(如右)。我們的研究將使用 pix2pix 作為基礎模型。

(三) VAE (Variational Autoencoder)

Autoencoder 模型為 Encoder-decoder 結構。訓練時，輸入圖像 x 會被 encoder 編碼成 latent code，再由 decoder 依照 latent code 的資訊嘗試還原出 x。Latent code 為整個模型結構的瓶頸，所以 encoder 的目標是把輸入 x 以最少資訊損失的方式壓縮成維數相對很小的 latent code，以供 decoder 使用。也就是說，encoder 做的是非線性降維，它會萃取輸入圖片的高階特徵。訓練完成後，藉由調整 latent code，decoder 就能用來生成各種不同風格的圖像。



圖四、VAE 的基本模型結構

而 VAE(Variational Autoencoder)[7]類似 Autoencoder，但其中 encoder 的輸出為平均及標準差，這兩個參數代表著一個高斯分布，訓練時 latent code 會從該分布中隨機取出，傳給 decoder。且 latent code 的先驗分布會被額外的 latent loss 拉成接近標準高斯分布的形狀，這項限制使 VAE 能學到更有意義的 latent space，也方便應用。其基本模型結構如圖四。

生成擬真的人臉圖像即為 VAE 典型的應用，而 VAE 學習到的 latent space 中，各維度的意義可能是臉的方向、膚色、頭髮長度或眼睛大小。

我們的研究將以 VAE 與 pix2pix 結合，成為一種新的類神經網路架構，接著會以這個架構訓練能生成地形高度圖和紋理貼圖的模型，最後會與基礎的 pix2pix 模型與物理侵蝕模型進行生成品質的比較。

二、收集訓練模型所需之圖像資料

本研究主要會收集五個地區的地形高度圖及衛星空照圖。這五個地區分別為橫斷山脈、喜馬拉雅山、祕魯安地斯山脈、阿根廷及加拿大的冰河地形。透過收集不同區域的地形，使 VAE-pix2pix 能學到不同地區的高度圖及空照圖的特徵。

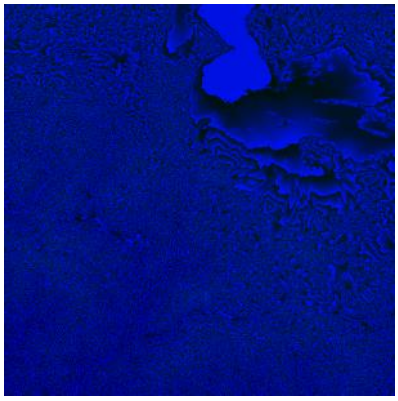
(一) 地形高度圖

地形高度圖的資料來源為 NASA 的 SRTM 1 Arc-Second 資料集[3]。此資料集將 1 經/緯度範圍的高度資料儲存為一張高度圖，每張高度圖的編號方式是按照其高度圖左下角的座標作為檔名，若此張高度圖的收集範圍為 25°N, 98°E、25°N, 99°E、24°N, 99°E、24°N, 98°E 四個座標點所圍成的範圍，則此張高度圖的檔名即為 N24E98，我們在附錄中會使用這種方式來表示各地區的收集範圍。

因為 SRTM 資料集採用特殊的 HGT 格式，不能使用一般的圖像軟體讀取，也使得生成訓練資料集的工作變得麻煩，所以我們利用 gmalthgtparser 來將 HGT 格式轉為可以直接以圖像軟體讀取的 PNG 格式。gmalthgtparser 為一個 Python module，可以讀取 HGT 檔案中特定地點的高度值(單位為公尺)，讀取到高度值後，我們利用式一將高度值轉為 RGB 值。

$$(R,G,B) = \left(\left\lfloor \frac{height}{256^2} \right\rfloor, \left\lfloor \frac{height \,\%\, 256^2}{256^1} \right\rfloor, \left\lfloor \frac{height \,\%\, 256}{1} \right\rfloor\right) \quad \text{(式一)}$$

對高度值進行轉換後，我們即可以將一張 HGT 檔案的高度圖轉換為方便易用的 PNG 高度圖，轉換後的高度圖如圖五。



圖五、由 hgt 檔案轉換的 PNG 圖檔

(二) 衛星空照圖

本研究中，地形高度圖需要與衛星空照圖相互對應，所以使用 MapTiler 所提供的 XYZ tiles map 來收集衛星空照圖。XYZ tiles map 是一種儲存地圖資料的方式，其方式為將大圖切割成許多張小圖，可以使地圖加載的速度變快，亦可節省網路資源。我們利用 MapTiler 所提供的衛星空照圖 tiles map 服務，收集橫斷山脈範圍內的多張衛星空照圖，再以 EPSG:4326 (WGS 84)座標系統將各張小圖(tiles)組合成一張與地形高度圖互相對應的衛星

空照圖。

因為衛星空照圖需與地形高度圖相互對應，所以空照圖的收集數量要與高度圖相同。表一為收集五個地區的高度圖及空照圖數量，具體的收集範圍列於附錄 A：

表一、五個地區的高度圖及空照圖收集總數 (單位：張)

地區	高度圖數量	空照圖數量
橫斷山脈	16	16
喜馬拉雅山	10	10
祕魯安地斯山	15	15
阿根廷冰河地形	9	9
加拿大冰河地形	5	5

三、本研究建構的模型結構 — VAE-pix2pix：

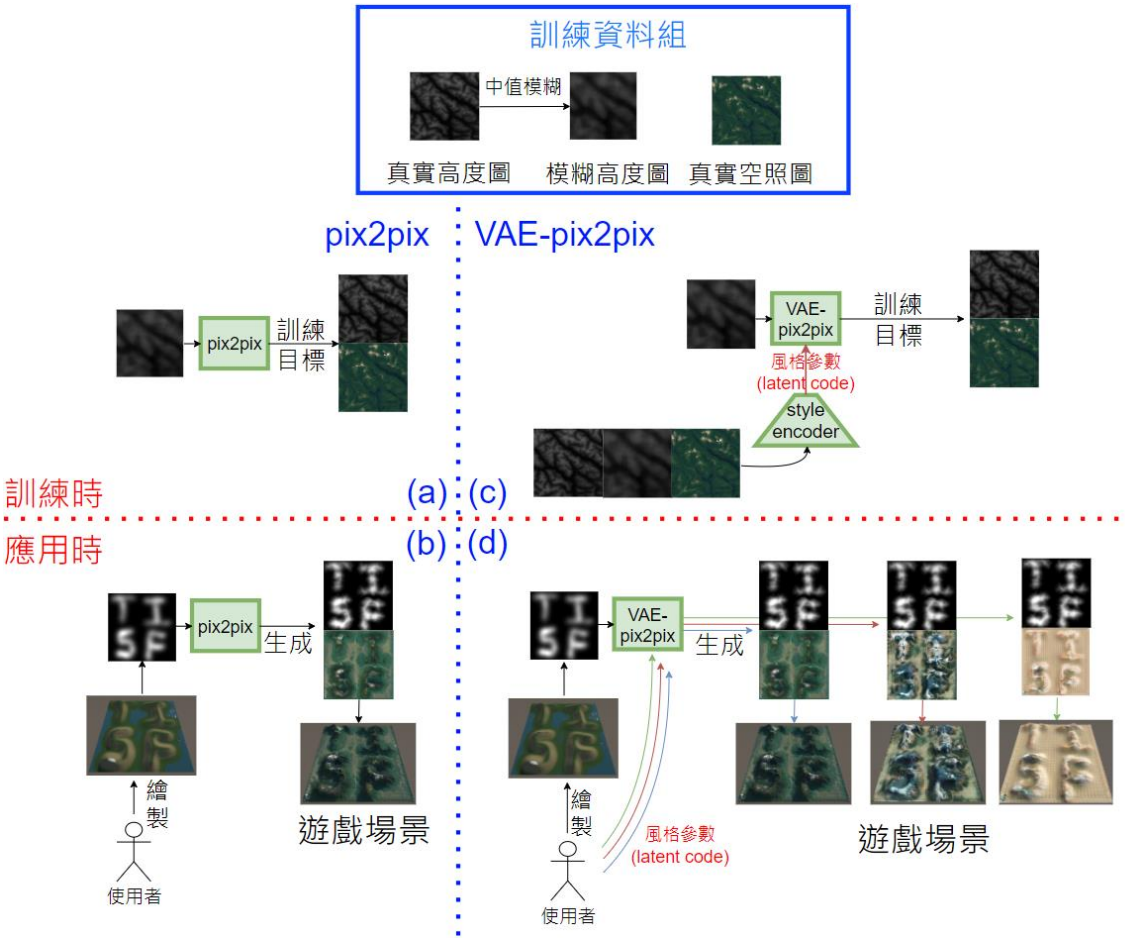
本研究的目的是簡化人工繪製地形的過程。我們建構一個類神經網路模型，以人工繪製的大致地形架構作為輸入，生成擬真的高度圖和紋理貼圖供遊戲開發者使用。

我們將蒐集到的真實高度圖用中值模糊處理，把山脊上的小河谷和大河谷上的小凸起物抹除，以模擬手繪大致山脈的架構。訓練時，這張**模糊高度圖**就做為 pix2pix 模型的輸入，未經模糊處理的**真實高度圖**和與之對應的**衛星空照圖**則做為 pix2pix 模型的目標輸出，如圖六(a)。

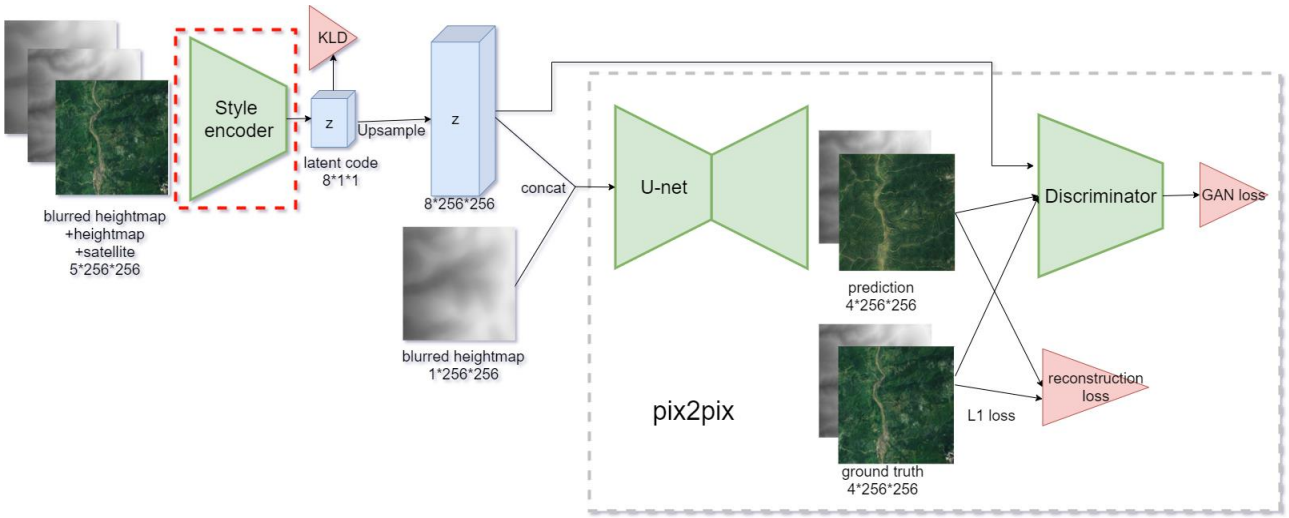
如果僅使用原 pix2pix 模型來訓練，因 pix2pix 本身缺乏調整生成風格的機制，所以應用時無法任意調整生成地形的風格，限制了應用的範圍。為了讓模型能調整生成風格，我們建構了 **VAE-pix2pix 模型**，在 pix2pix 的 U-Net 前增加了 style encoder (為 VAE 的 encoder)。訓練時，style encoder 負責從高度圖和空照圖提取出這些風格資訊（latent code）輸入 U-Net，再由 U-Net 以正確的風格生成高度圖和空照圖，如圖六(c)。latent code 也會提供給 discriminator（此路徑不接收反向傳播的梯度），以利判斷 U-Net 是否生成正確的風格。而應用時，使用者只要輸入給 U-Net 不同風格資訊，就能控制它生成不同風格的地形，如圖六(d)。

從 VAE 的角度來看，此模型相當於利用 U-Net 作為 Decoder，並在瓶頸處額外輸入高度

圖作為空間資訊的 VAE。



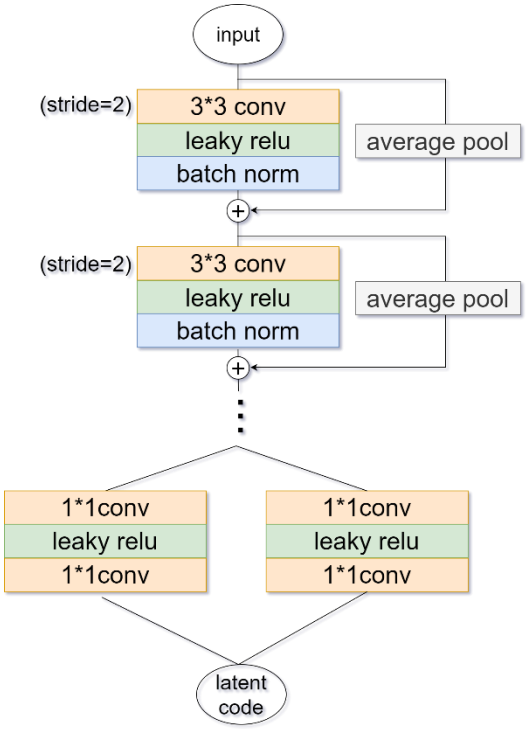
圖六、pix2pix 和 VAE-pix2pix 的訓練及應用流程簡圖



圖七、VAE-pix2pix 整體模型結構

為了不限制能處理的圖片大小，並允許使用者在不同位置指定不同風格，本研究的 style encoder 不會像典型的 VAE 一樣使用 fully connected layer 來產生 latent code，而是用 convolution layer，以保留空間維度。

因為 style encoder 的輸入中含有目標空照圖(ground truth) 的資訊，所以作為瓶頸的 latent code 必須足夠窄，以防止 U-Net 輕易的將目標高度圖和空照圖直接輸出。如果輸入圖的長寬為 h, w，則 style encoder 會將 latent code 縮小到一個有 8 channels，長為 (h/256)，寬為(w/256)的 tensor。



圖八、Style encoder 的具體結構

設計模型結構時，一開始的 style encoder 是無跳躍連結(skip-connections)的多層 CNN，但我們發現如果不在 style encoder 中使用 batch normalization (BN)，訓練過程中會造成梯度爆炸 (gradient exploding)；但如果使用 BN，則會造成 latent code 梯度消失和 KLD vanishing，也就是 style encoder 退化成只會輸出標準高斯分布的 latent code。我們認為這個現象很有可能是因為 style encoder 中，接近後端的 BN 使產生的 latent code 過於不穩定，導致 U-Net 不採納 latent code，只參考模糊高度圖。為了解決這個問題，我們用類似 ResNet 的方式，把每一組[conv, ReLU, bn] block 的旁邊加上一條路徑，使資料可以選擇路徑，不一定要經過導致不穩定的 BN，同時，因為這條路徑上沒有可訓練的 bias 或 weight，所以不會像直接去除 BN 的時候一樣發生梯度爆炸。

四、對 U-Net 的修改

為了讓 VAE-pix2pix 模型能更符合我們的需求，我們對 U-Net 進行修改，修改的項目如下：

(一) Convolution Layer 的 stride

原本 pix2pix 的每個 convolution layer 的 stride 都是設為 2，這樣會使 feature map 每往下一層都縮小成 0.5 倍，以處裡更大範圍的特徵。但是在生成地形這項工作中，只需要產生較小範圍的細節，且全局特徵已經能從模糊高度圖和 style encoder 得到，所以我們把 U-Net 改成每兩層 convolution layer 才會有一個是 stride 為 2 (另一個是 stride 為 1)，使 U-Net 更專注於精

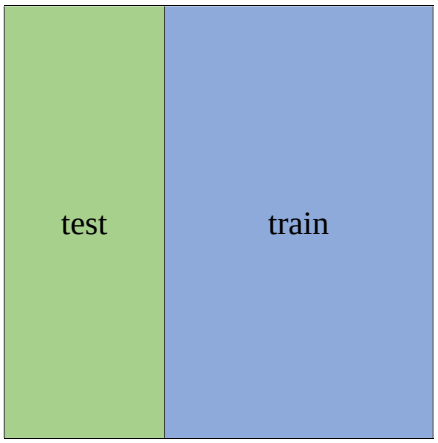
細特徵。不過此設定增加了 feature map 大小，使訓練時間變成原本的約 4 倍。

(二) 激勵函數 tanh

U-Net 會輸出一張高度圖和一張紋理貼圖，其中高度圖的像素值不應被 U-Net 尾端的 tanh 運算限制，所以我們把 tanh 改成只會對空照圖運算。

五、生成訓練資料集

我們將收集的高度圖和空照圖分為 train、test 兩個區域 (train 用於訓練模型，而 test 則用來在訓練完成後用於測試模型)，並分別從這兩個區域中切割出訓練資料集的圖像。這樣可以使個兩個部分的圖像不重複，以測試模型的精準度。



圖九、train、test 在大圖的位置

將收集的空照圖及高度圖分隔成上述兩個區域後，我們隨機在其上切割出多個大小為 256 x 256 像素的高度圖及空照圖。並將高度圖透過線性變換的方式由 24 bits 高度圖轉換為 8 bits 的高度圖。後再利用中值模糊 (median blur) 將真實高度圖模糊為模糊高度圖，以模擬人工手繪的高度圖。其中中值模糊的 kernel size 設為 29。

最後，我們再將模糊高度圖、衛星空照圖及真實高度圖併排組成訓練資料組，如圖十。訓練時，這三張圖片會輸入到 style encoder 中，而 style encoder 提取出 latent code 後，latent code 再與模糊高度圖一同輸入到 U-Net 中，最後會一起生成出擬真的衛星空照圖及高度圖。

此外，訓練資料集共有 3158 組圖像，train 資料集的資料對數為 2526 組，而 test 資料集為 632 組。

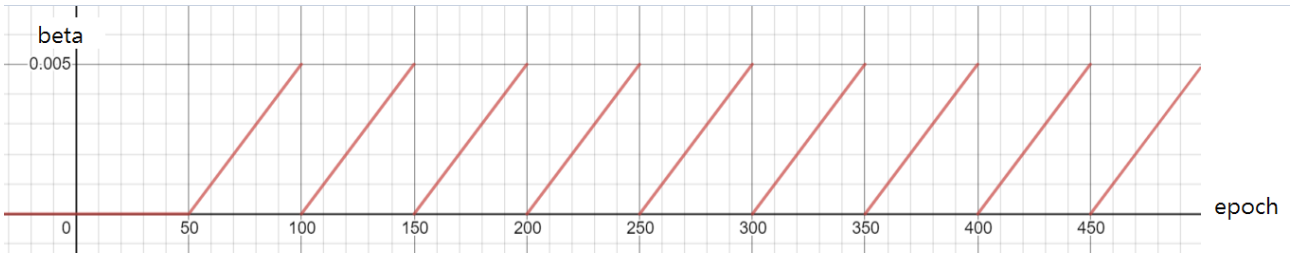
圖十、訓練資料組，由左至右分別為模糊高度圖、衛星空照圖及真實高度圖

六、訓練模型

本研究的模型結構是在 pix2pix 模型的基礎上修改，額外加上 style encoder，所以 pix2pix 的部分仍然沿用論文作者的程式碼，進行訓練及測試時也是使用 pix2pix 原作者所提供的程式。

訓練參數如下：

- Epoch：500
- Batch size：18
- Learning rate：0.0002 (前 300epoch)，由 0.0002 線性下降至 0 (後 200epoch)
- Beta (為 KLD 的乘數，用來限制 latent code 分布的 loss)：隨時間的變化如圖十



圖十一、beta 隨 epoch 的變化方式

訓練時對資料組做的 data augmentation (資料增強) 如下：

(一) 對模糊高度圖、衛星空照圖及真實高度圖同時作用的 data augmentation：

- 50% 的機率上下翻轉
- 50% 的機率水平翻轉
- 50% 的機率轉置 XY 座標(斜向翻轉)

(二) 對模糊高度圖及真實高度圖同時作用的 data augmentation：

- 隨機縮放：將整張圖的像素值都乘以e^x，$x \sim N(0,0.0225)$ (μ 為 0，σ^2為 0.0225 的高斯分布)
- 隨機偏值：將整張圖的像素值都增加 x，$x \sim N(0,4)$

參、研究結果與討論

一、評斷生成對抗網路模型的方法

一般來說，很難找到一個好的方法來評斷生成對抗網路的表現，因為它不像一般的分類器能利用分類的精確度來評斷神經網路的優劣。

在本研究中，我們將會利用比較輸出與目標輸出的 L1 Loss、L2 Loss、Perceptual Loss、FID、SSIM Index (structural similarity index)及生成每張圖像所需之平均時間來探討物理侵蝕模型、基礎的 pix2pix 模型與 VAE-pix2pix 的差異。

(一) L1 Loss

L1 Loss 的計算方式是將兩張大小相同圖像所相對應的像素值相減後取絕對值，再相加在一起，最後取平均。具體公式如式二：

$$\text{L1 Loss}(X,Y) = \frac{1}{h \times w}\left(\sum_{i=1}^{h}\sum_{j=1}^{w}\left|X_{i,j} - Y_{i,j}\right|\right) \text{ (式二)}$$

其中，h, w 分別為圖片的高度及寬度。**可以發現 L1 Loss 的數值越小，代表兩張圖越相近**。藉由 L1 Loss 我們可以看出兩張圖的相似程度。

(二) L2 Loss

L2 Loss 與 L1 Loss 的計算方式相似，只是將絕對值替換成平方，具體計算公式如式三：

$$\text{L2 Loss}(X,Y) = \frac{1}{h \times w}\left(\sum_{i=1}^{h}\sum_{j=1}^{w}\left(X_{i,j} - Y_{i,j}\right)^2\right) \text{ (式三)}$$

與 L1 Loss 相似，**L2 Loss 同樣是數值越小，代表兩張圖越相近**，我們也可以透過 L2 Loss

來看出兩張圖的相似程度。但是 L2 Loss 對於偏離越多的值，對 Loss 值的影響更大。

(三) Perceptual Loss

在計算 Perceptual Loss [11]時，會利用到一個已經訓練好(pre-trained)的 VGG 16 模型。Perceptual Loss 的計算方式是計算兩張圖像在 VGG 16 各層 activation 的 L1 Loss，最後再將各層計算出的 L1 Loss 相加。從 Perceptual Loss 可以看出兩張圖的風格是否相似，且 Perceptual Loss 同樣是數值越小，代表兩張圖越相近。

(四) FID (Fréchet Inception Distance)

在計算 FID [2] 時，會利用到一個已經訓練好(pre-trained)的 inception network v3 神經網路來提取兩張圖片的特徵(feature)。圖片的特徵(為一個 2048 維的高階特徵)主要可以從 inception network 輸出層的前一層提取到。對於目標輸出，我們可以假設這個 2048 維向量是服從高斯分布。那由神經網路輸出的特徵應該也要服從高斯分布。所以我們知道生成對抗網路的目標是使這兩個分布的距離盡量接近。

而計算這兩個分布的距離等同於求目標輸出和輸出的 2048 維特徵的距離。數學上，如果想要計算兩個分布的距離，我們可以使用 Fréchet distance 來進行計算。

在計算上，我們會假設這兩個分布是服從高斯分布，且我們知道若一個隨機變數服從於高斯分布，則這個隨機變數可以使用高斯分布的標準差與平均表示，只要兩個分布的標準差和平均皆相同，則兩個分布相同。標準差和平均就是用來計算 FID。但因為這裡我們要計算的是多維的向量，所以我們會使用平均和共變異數(covariance)矩陣來計算兩個分布的距離。而平均的維度是 2048 維，而共變異數矩陣就是一個 2048 x 2048 維的矩陣。有了以上的定義後，我們就可以使用式四來計算輸出與目標輸出的 FID。

$$\mathrm{FID}(X,Y) = \|\mu_X - \mu_Y\|_2^2 + \mathrm{Tr}\left(\Sigma_X + \Sigma_Y - 2(\Sigma_X\Sigma_Y)^{\frac{1}{2}}\right) \quad \text{(式四)}$$

其中μ_X、μ_Y為 X、Y 的平均，Σ_X、Σ_Y代表 X、Y 的共變異數矩陣。**可以發現當 FID 值越小時，代表輸出與目標輸出的分布越接近**。但是 FID 並不會看出輸出與目標輸出的空間對應關係。所以我們上述所提之 L1 Loss 及 L2 Loss 即是為了評斷模型在空間上的準確性。

(五) SSIM index (structural similarity index)

SSIM 指標 [10] 是一種用來評斷兩張圖像相似程度的方法，相較於其他種方法，SSIM 指標能更好的符合人眼對圖像品質的判斷。

SSIM 指標主要透過比較兩張圖片的亮度、對比度及結構(structure)來評斷兩張圖片的相似程度，具體計算方式如式五：

$$\text{SSIM}(\mathbf{X},\mathbf{Y}) = [l(\mathbf{X},\mathbf{Y})]^{\alpha}[c(\mathbf{X},\mathbf{Y})]^{\beta}[s(\mathbf{X},\mathbf{Y})]^{\gamma},$$
$$l(\mathbf{X},\mathbf{Y}) = \frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}, c(\mathbf{X},\mathbf{Y}) = \frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}, s(\mathbf{X},\mathbf{Y}) = \frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \quad \text{(式五)}$$

$l(\mathbf{X},\mathbf{Y})$是用來比較 X, Y 兩張圖的亮度，$c(\mathbf{X},\mathbf{Y})$則是用來比較對比度，而$s(\mathbf{X},\mathbf{Y})$用來比較兩張圖的結構。μ_x,μ_y代表兩張圖像素值的平均，σ_x,σ_y代表兩張圖像素值的標準差，σ_{xy}為兩張圖的共變異數(covariance)，C_1,C_2,C_3是三個常數，以避免出現分母為 0 的情況。另外，在本研究中，我們設定$\alpha = \beta = \gamma = 1$。

根據以上的公式，我們可以發現 SSIM index 滿足對稱性($\text{SSIM}(X,Y) = \text{SSIM}(Y,X)$)、有界性($-1 \le \text{SSIM}(X,Y) \le 1$)及極限值唯一($\text{SSIM}(X,Y) = 1 \Leftrightarrow X = Y$)

(六) 生成每張圖像所需的平均時間

我們利用 PyTorch 內建的 torch.cuda.Event() 來計算生成圖像所需的時間。計算完生成總時間後，就可以算出生成單張圖像所需的平均時間。在本研究中，我們是利用 NVIDIA GTX 1080 Ti 12 GB 的顯示卡來進行圖像的生成。

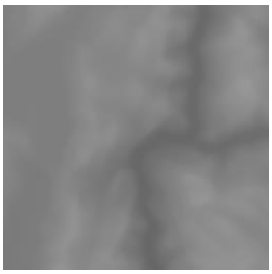
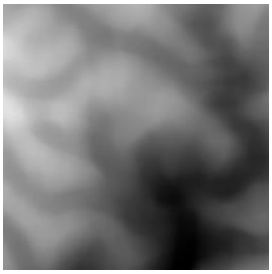
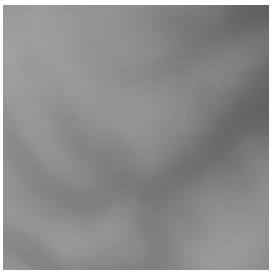
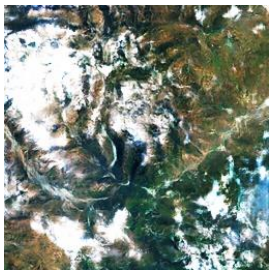
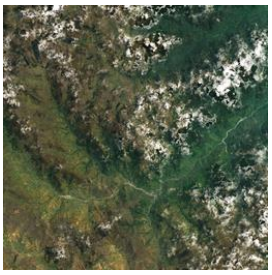
二、VAE-pix2pix 模型之訓練結果

(一) 衛星空照圖的生成結果

表二為原 pix2pix 結構與 VAE-pix2pix 結構生成衛星空照圖的測試結果，其圖像皆為來自 test 資料集之圖像，也就是說模型並沒有在訓練過程中“看過”這些圖像。藉由觀察這些圖像，我們可以更好的評斷模型的學習程度。表三則為兩個模型在 test 資料集的平均指標值。

表二、原 pix2pix 及 VAE-pix2pix 的生成衛星空照圖的測試結果

地區	輸入	模型輸出		目標輸出
		原 pix2pix	VAE-pix2pix	

橫斷山脈				
喜馬拉雅山				
祕魯安地斯山				
阿根廷冰河地形				
加拿大冰河地形				

表三、原 pix2pix 及 VAE-pix2pix 生成衛星空照圖的平均指標值

模型	L1 Loss	L2 Loss	Perceptual Loss	FID	SSIM	平均生成時間 (毫秒)
原 pix2pix	61.716	6987.011	4.358	**117.008**	0.1076	**13.925**

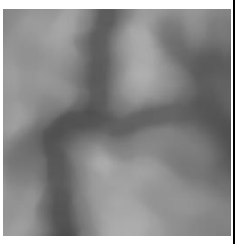
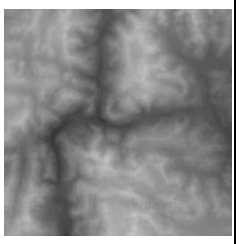
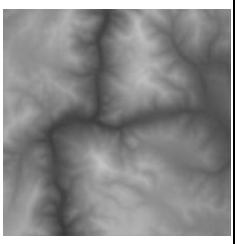
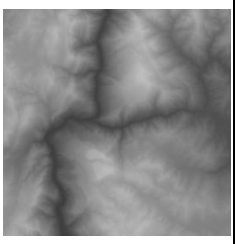
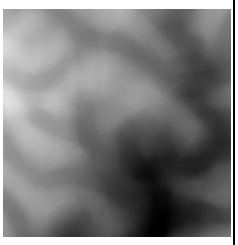
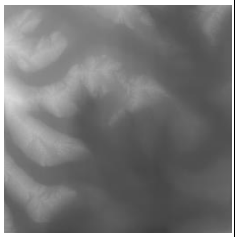
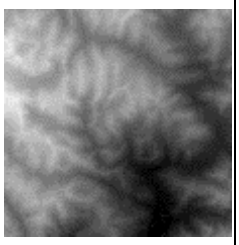
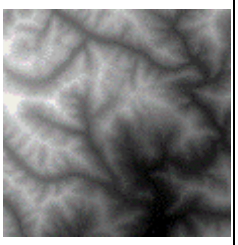
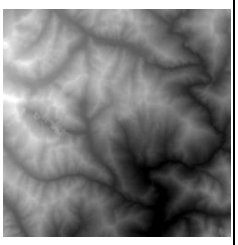
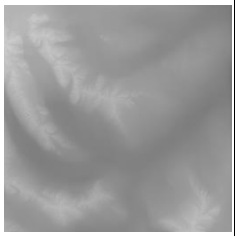
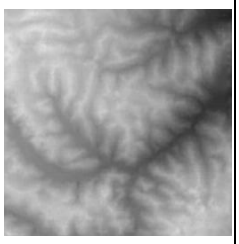
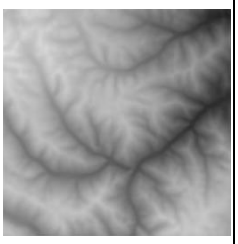
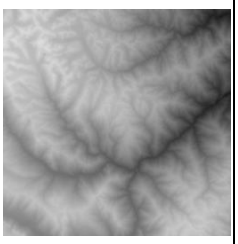
VAE-pix2pix	**29.969**	**2134.224**	**2.897**	121.742	**0.2816**	54.253

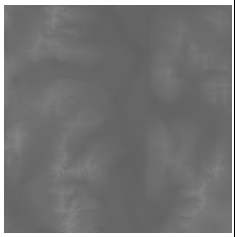
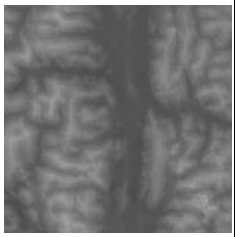
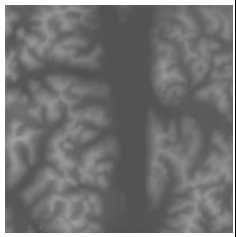
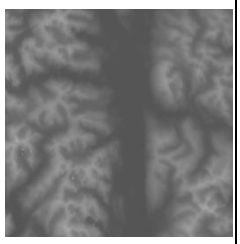
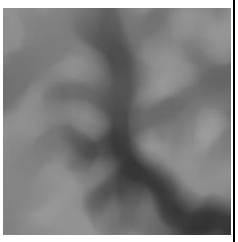
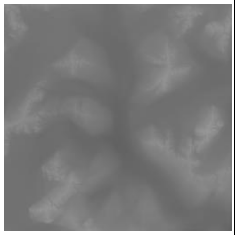
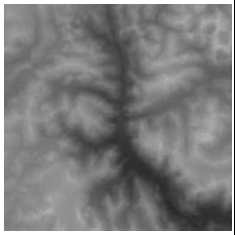
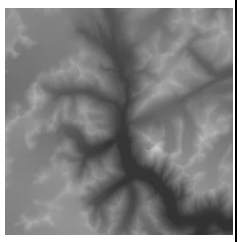
可以發現我們建構的 VAE-pix2pix 架構相較於原 pix2pix 架構生成出的衛星空照圖較接近於目標輸出，不僅山脈的顏色更接近於目標輸出，且輸出的空照圖相當符合輸入高度圖的結構。這點也反映在各指標上，L1 Loss、L2 Loss、Perceptual Loss 及 SSIM 都表現較好。

(二) 地形高度圖的生成結果

同樣的，以下的測試圖像都是來自 test 資料集。測試完成後，測試結果如表四，我們會將模型輸出與目標輸出計算出各指標，如表五，藉此來判斷每個模型的表現。

表四、物理侵蝕模型、原 pix2pix 結構及 VAE-pix2pix 結構的生成地形高度圖的測試結果

地區	輸入	模型輸出			目標輸出
		物理侵蝕模型	原 pix2pix	VAE-pix2pix	
橫斷山脈					
喜馬拉雅山					
祕魯安地斯山					

阿根廷冰河地形					
加拿大冰河地形					

表五、物理侵蝕模型、原 pix2pix 及 VAE-pix2pix 生成地形高度圖的平均指標值

模型	L1 Loss	L2 Loss	Perceptual Loss	FID	SSIM	平均生成時間 (毫秒)
物理侵蝕模型	14.335	335.69	1.3168	168.260	0.7499	340.199
原 pix2pix	7.09	107.478	1.5114	184.077	0.7411	**13.925**
VAE-pix2pix	**4.311**	**48.567**	**1.1382**	**89.971**	**0.882**	54.253

從表五，我們可以發現 VAE-pix2pix 生成的高度圖較原 pix2pix 所生成的圖像來說更接近於目標輸出。由各指標也可以看到 VAE-pix2pix 要優於 pix2pix。相較於物理侵蝕模型，本研究的方法約比其生成的速度快 6 倍。

三、各地區的高度圖及空照圖在 latent space 上的分布

latent code 的分布是判斷 VAE 訓練結果好壞的重要觀察項目。研究中在尋找最佳模型參

數的階段，latent code 的分布合理是我們的主要目標之一，因為「不好的」latent code 分布代表模型沒有正常處理風格資訊。

我們從五個地區蒐集的每個資料組(由模糊高度圖、衛星空照圖及真實高度圖組成)，經過已訓練 VAE-pix2pix 模型中 style encoder 的轉換，會被一一映射到 8 維的 latent space 上，如圖十二。每個點代表一個資料組，而點的顏色代表所在地區。因為 latent code 其實是一個高斯分布，圖中資料點顯示的位置只取高斯分布的中心位置。

我們從以下兩點觀察 latent code 分布的好壞：

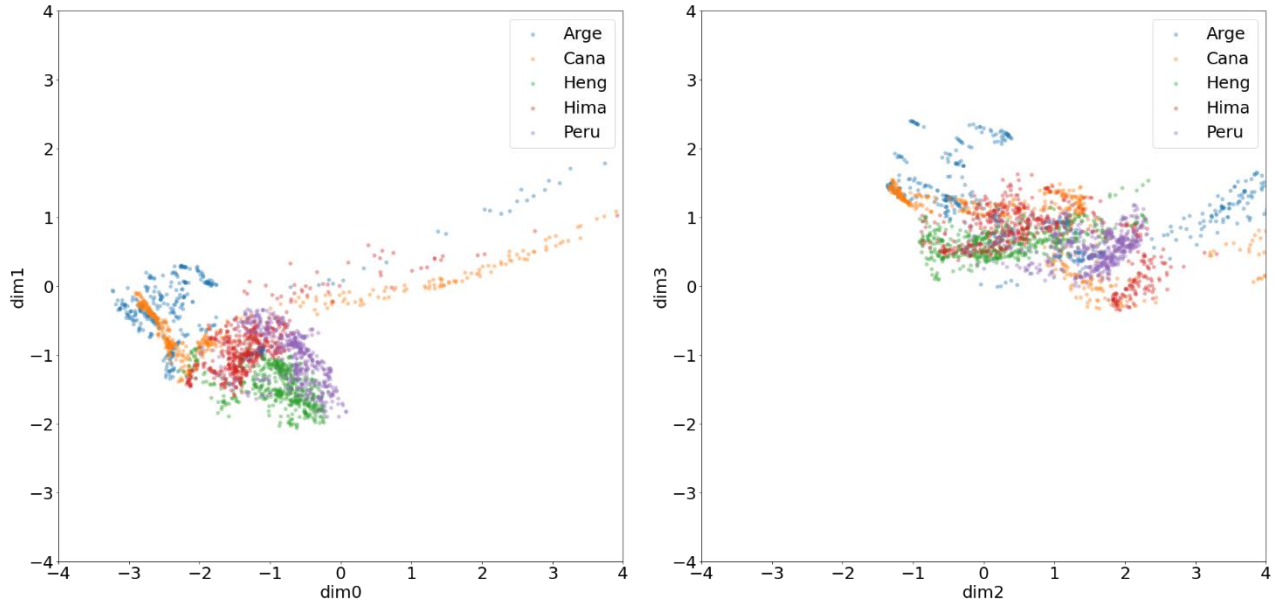
(一) 不同地區的 latent code 是否分離：

由圖十二可以看出，每個地區的 latent code 大致自成一個色塊，表示 style encoder **能很好的分辨每種地區地形的不同特徵**。

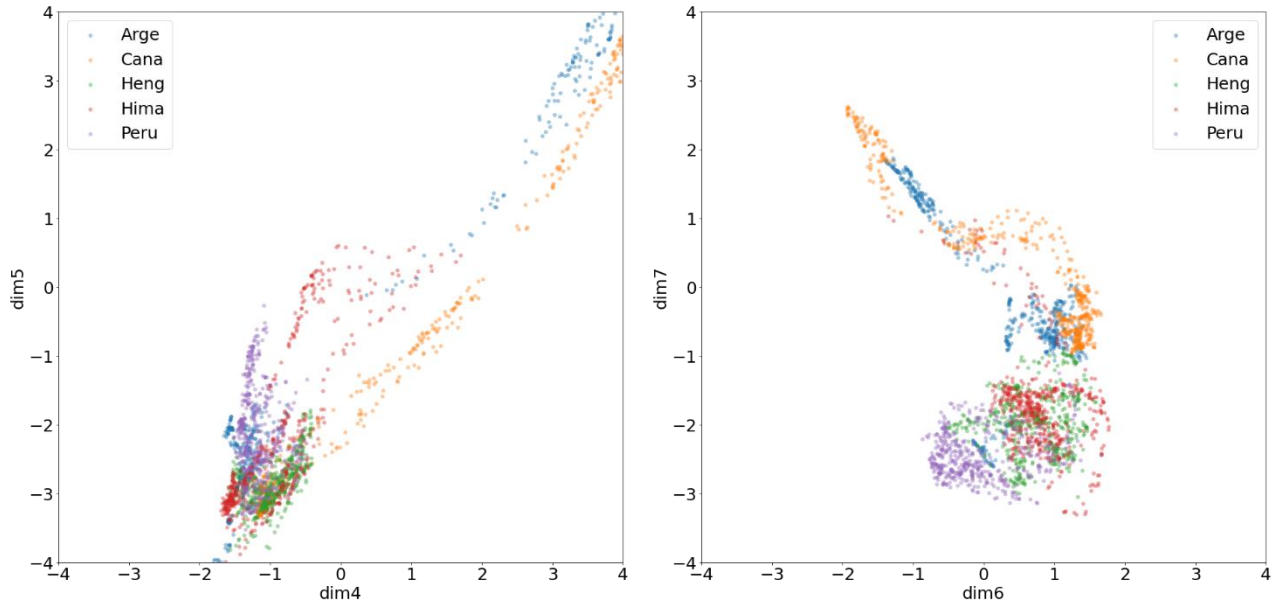
(二) 所有 latent code 是否集中：

由圖十二可以看出，大部分的資料點都集中成同一個連續區塊，表示在應用時，沿著連續的路徑調整輸入的 latent code，**能連續轉換**生成出的地形風格，且藉由插值兩種地區的 latent code，我們的模型可以生成出**「兩種風格的中間」**的地形，這點也反映在圖十七的測試上。

為了使 latent code 能符合以上兩點，我們多次調整 beta 參數，但一部分的阿根廷和加拿大地形的 latetnt code 跑到遠處，可能要再調整 beta 等參數來解決此問題。



(a) 第一維與第二維作圖　　(b) 第三維與第四維作圖



(c) 第五維與第六維作圖　　(d) 第七維與第八維作圖

圖十二、將資料組透過 style encoder 轉換後映射到 8 維 latent space 上

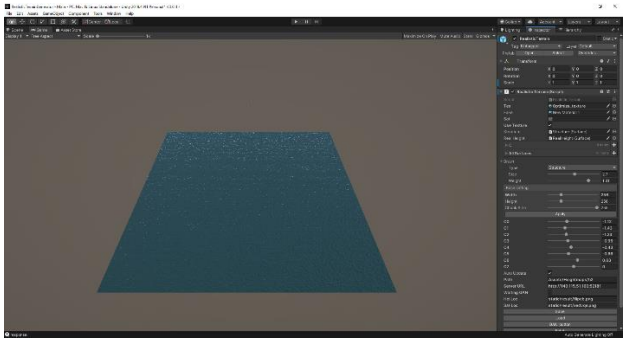
(Arge 為阿根廷、Cana 為加拿大、Heng 為橫斷山脈、Hima 為喜馬拉雅山、Peru 為祕魯)

四、建構模型的 API 伺服器

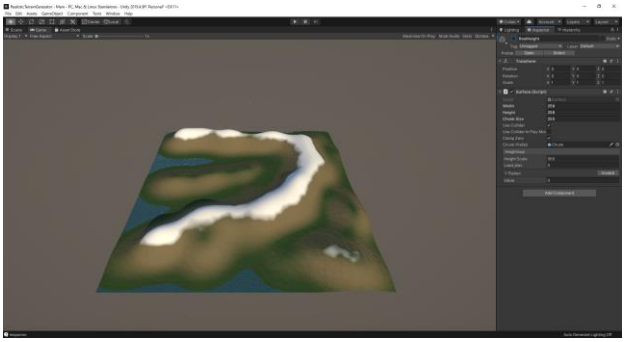
為了避免每次需要應用各模型處理高度圖時都要重新載入模型，我們利用 Python 的 Flask 套件建立了一個 API 伺服器，並將其建置於工作站上。用戶可以直接上傳手繪的高度圖，並在伺服器上用訓練好的各模型進行處理，後再回傳回用戶端。不僅運算快速，且也省去了架設軟體環境和下載模型檔案的時間。這個 API 可以快速的在模型之間切換，且支援本研究中所有訓練的模型。

五、Unity 用戶端

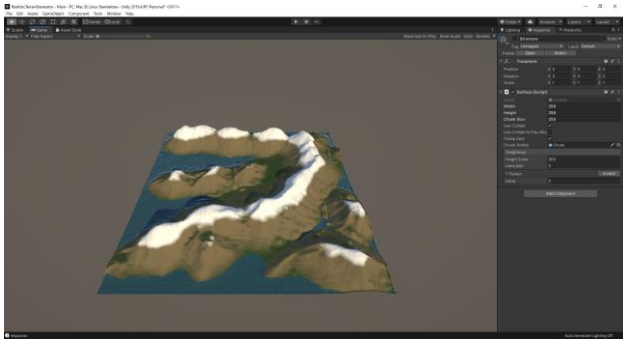
本研究自行開發了一個 Unity 客戶端，使用戶可以在 Unity Editor 的編輯模式中直接對地形進行操作。使用者可以匯入一張高度圖或直接在地形上繪製 3D 的模型。每當畫完一筆畫後，Unity 會直接對 API 伺服器送出請求，並在一秒內更新出擬真的山脈地形。也可以將貼圖模型所生成的擬真衛星空照圖貼在 3D 模型上，具體功能如圖十三、圖十四、圖十五、圖十六：



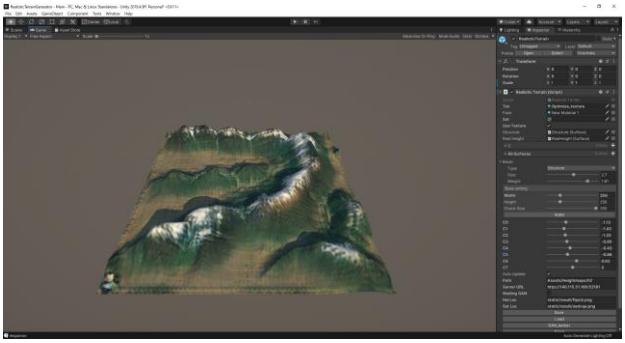
圖十三、Unity 客戶端的初始畫面



圖十四、用戶手繪之大致地形



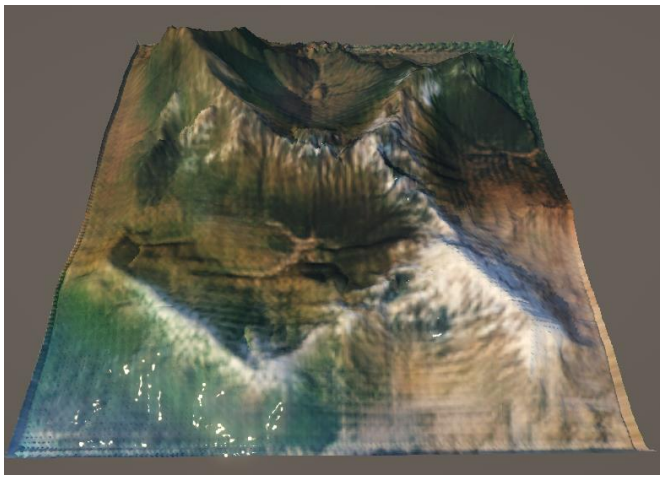
圖十五、VAE-pix2pix 生成的擬真地形



圖十六、將生成的空照圖貼在擬真地形

六、由用戶端調整 latent code 對於輸出風格的效果

圖十七展示了以 256 x 256 的手繪圖作為輸入高度圖，並在圖上不同區域使用不同的地區的風格，分別為阿根廷、加拿大、喜馬拉雅和秘魯，此四個地區的衛星空製圖如圖十八。圖十七的中央為四個地區在 latent space 上的平均，而往左上為秘魯、右上為喜馬拉雅、左下為加拿大、右下為阿根廷。可以發現四個方向的生成風格都有不同，且與真實地區的空照圖相似。



圖十七、在同一張圖中展示不同 latent code 的效果

祕魯　　喜馬拉雅　　加拿大　　阿根廷

圖十八、各地區的真實空照圖

準確來說，圖十七中的 latent code 設置為：

定左下角座標為(0,0)，右上角座標為(1,1)，加拿大、祕魯、阿根廷、喜馬拉雅的地區的平均 latent code 分別為 a, b, c, d ，則(x, y)處 latent code = $3\big((1-x)(1-y)a + (1-x)yb + x(1-y)c + xyd\big) - 2\frac{(a+b+c+d)}{4}$。所以在接近中心處的 latent code 為四個地區 latent code 之

間的內插值；在靠近邊緣及角落處的 latent code 則為平均 latent code 往四個地區 latent code 的外推（extrapolation）。

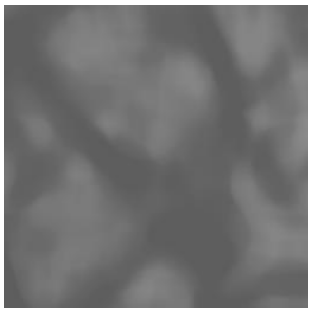
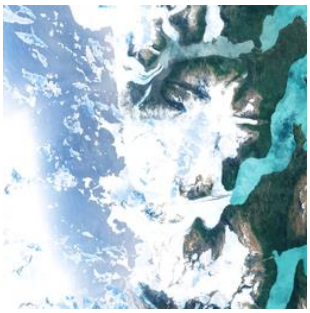
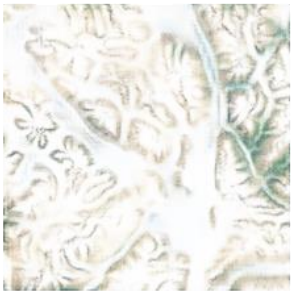
圖十七顯示了本研究模型將 pix2pix 與 style encoder 結合後，生成的地形不侷限於訓練資料的風格，能**生成出其他未出現在訓練資料集，甚至不是真實存在的地形風格**，這項特性大大增加了模型的應用價值。

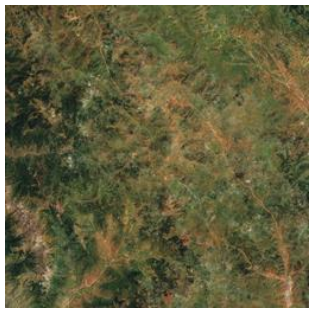
在應用上，使用者也可以利用 VAE-pix2pix 這項能在空間上設定不同 latent code 的特性，增加設計地形的自由度。

七、透過改變 style encoder 的輸入改變輸出圖像的風格

為了測試以上的討論，我們將一張高度圖及多張與不同地區的空照圖一起輸入到 VAE-pix2pix 的 style encoder 中，驗證能否改變輸出圖像的風格。表六為測試的結果：

表六、將同一張高度圖與不同的空照圖作為 style encoder 的輸入，並比較其輸出

輸入的高度圖	**輸入的空照圖(style)**	**VAE-pix2pix 的輸出**
		
		

經測試，可以發現 VAE-pix2pix 模型可以很好的將輸入空照圖的特徵提取出來，且也可以保留高度圖上河流及山脊等特徵。

八、討論物理侵蝕模型、pix2pix 模型及 VAE-pix2pix 之間的差異

在此節中，我們會比較物理侵蝕模型、原 pix2pix 模型及本研究建構的 VAE-pix2pix 模型之間的差異，比較如表七：

表七、物理侵蝕模型、原 pix2pix 模型及 VAE-pix2pix 的比較

模型	功能	生成速度	生成品質	是否可以改變生成圖像的風格
物理侵蝕模型	將大致高度圖經過侵蝕後變得較為擬真	慢(約 300 毫秒)	最差	否

原 pix2pix 模型	將大致高度圖變得更為擬真、根據高度圖生成相對應的衛星空照圖	最快(約 13 毫秒)	其次	否
VAE-pix2pix 模型	將大致高度圖變得更為擬真、根據高度圖生成相對應的衛星空照圖、改變生成圖像的風格	快 (約 50 毫秒)	最好	是

肆、結論與應用

在本研究，我們利用 NASA 的 SRTM 1 Arc-Second 及 MapTiler 收集了全球五個地區的高度圖及相對應的空照圖。利用這些收集的圖像，訓練了自行建構的 VAE-pix2pix 模型。VAE-pix2pix 為 Variational Autoencoder (VAE)及 pix2pix 結合的模型，可以將人工繪製的高度圖自動加上真實山脈應有的細節(包含尖銳的山脊、山壁上的紋路、連續的河流網路等……)，也能生成相對應的擬真衛星空照圖。

經過實測，相較於原 pix2pix 模型，VAE-pix2pix 所生成的高度圖及空照圖會更接近於真實世界的山脈高度圖，且 VAE-pix2pix 模型也可以透過改變其 latent code 的數值來生成出不同風格的高度圖及衛星空照圖，如地貌的顏色或雪線的高度等，這些都能增加模型生成圖像的多樣性，甚至能以可能不真實存在的風格生成地形，讓應用更為廣泛。與物理侵蝕模型進行比較，不僅生成速度遠快於物理侵蝕模型，生成品質也更為擬真，這些都是優於傳統模型的地方。

為了使模型的使用更加簡單，不用在終端機上打入許多複雜的指令，我們將模型的使用包裝成 Unity 客戶端，Unity 客戶端可以在圖形使用者介面完成在模糊高度圖加上山脈細節的工作，並能直接在畫面上顯示 3D 模型，也可以將生成出的衛星空照圖貼在擬真地形的 3D 模

型中，使空照圖及高度圖能更好的呈現。

綜合以上，本研究的 VAE-pix2pix 模型可以生成出更為擬真的高度圖及空照圖，且能自由調整生成圖像的風格。而我們開發的 Unity 客戶端，可以使我們的模型直接應用於遊戲的開發中，使得原先需要分成兩個步驟生成擬真的山脈地形與生成相對應的空照圖整合為一個步驟。簡化遊戲開發生成擬真山脈模型的任務。

伍、參考文獻

[1] 程品奕、李杰穎(2020)。利用生成對抗網路生成擬真的山脈地形。中華民國第六十屆中小學科學展覽會。

[2] Dowson, D. C., & Landau, B. V. (1982). The Fréchet distance between multivariate normal distributions. *Journal of multivariate analysis, 12*(3), 450-455.

[3] Earth Resources Observation And Science (EROS) Center. *Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global*. 2017.

[4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).

[5] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).

[6] Jákó, B., & Tóth, B. (2011, November). Fast Hydraulic and Thermal Erosion on GPU. In *Eurographics (Short Papers)* (pp. 57-60).

[7] Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114.*

[8] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

[9] Maximilian Seitzer. *pytorch-fid: FID Score for PyTorch.* https://github.com/mseitzer/pytorch-

fid. Version 0.1.1. Aug. 2020.

[10] Zhou Wang et al. “Image quality assessment: from error visibility to structural similarity”. In: *IEEE transactions on image processing* 13.4 (2004), pp. 600–612.

[11] Richard Zhang et al. “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric”. In: *CVPR*. 2018.

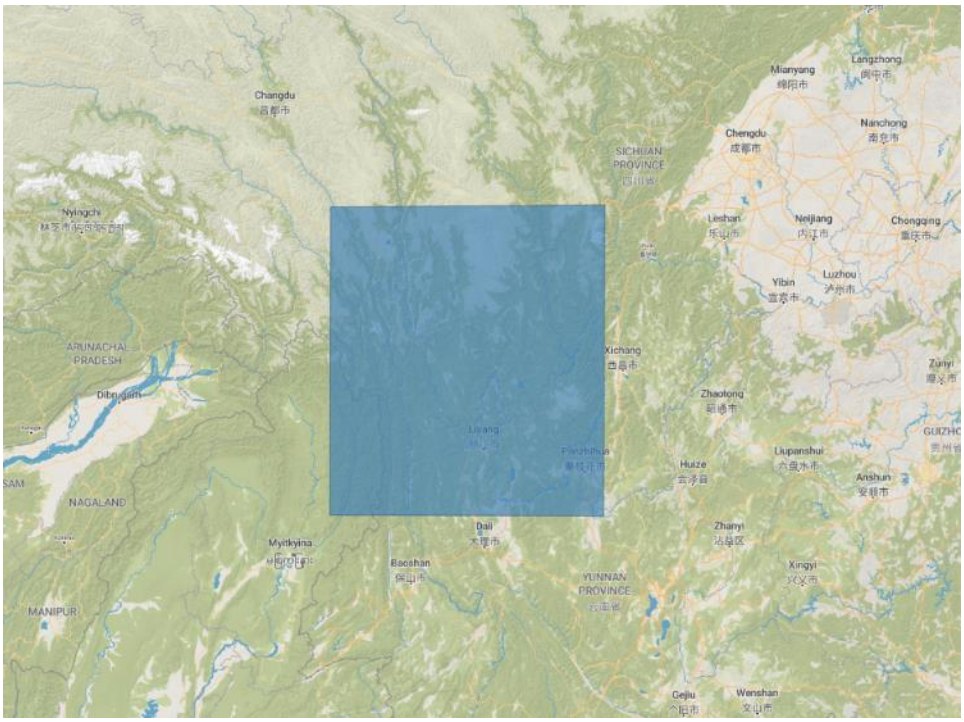
附錄

A. 高度圖及空照圖的具體收集範圍

如內文所提，我們會利用高度圖左下角的經緯度座標表示一張高度/空照圖，如一張高度圖包含 25°N, 98°E、25°N, 99°E、24°N, 99°E、24°N, 98°E 四個座標點所圍成的範圍，則此張高度圖的檔名即為 N24E98。除了會在下面列出各個座標點，我們也有將具體的收集範圍在地圖上框出。

a、 中國橫斷山脈：

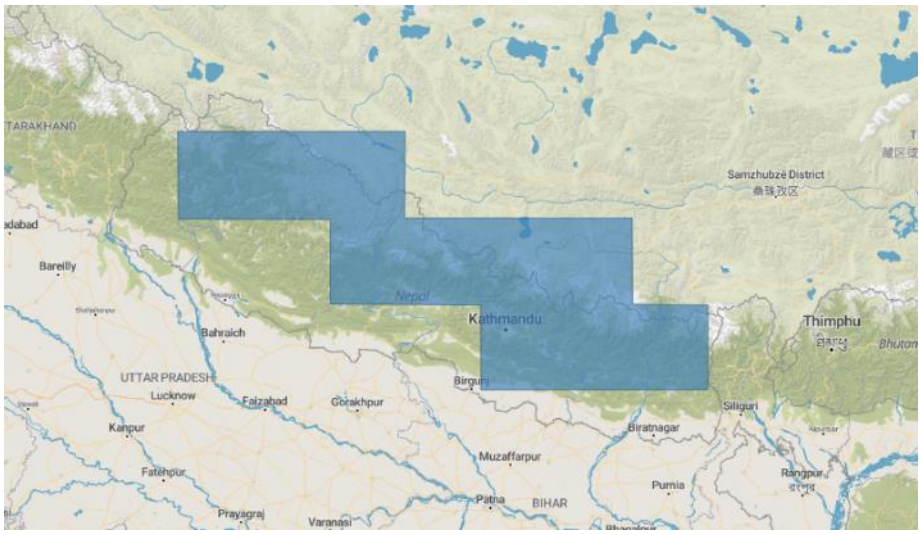
i.	N27E099	vii.	N29E101	xiii.	N28E099
ii.	N27E098	viii.	N29E100	xiv.	N28E098
iii.	N26E101	ix.	N29E099	xv.	N27E101
iv.	N26E100	x.	N29E098	xvi.	N27E100
v.	N26E099	xi.	N28E101		
vi.	N26E098	xii.	N28E100		



b、 喜馬拉雅山脈：

i.	N29E081	iii.	N29E083	v.	N28E084
ii.	N29E082	iv.	N28E083		

vi.	N28E085	viii.	N27E086	x.	N27E087



vii.	N27E085	ix.	N28E086		

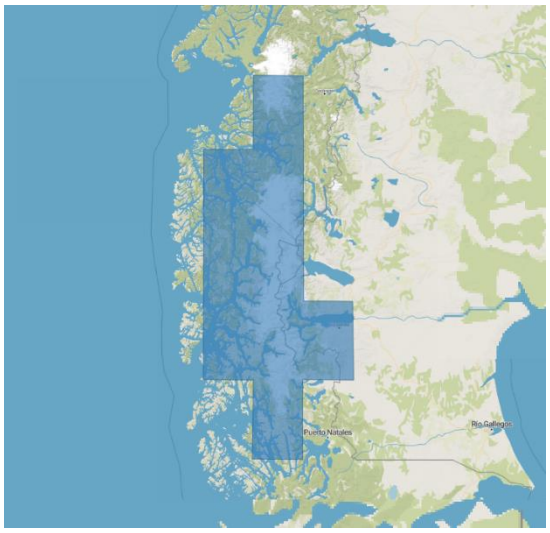
c、 祕魯安地斯山脈：

i.	S07W079	vi.	S10W077	xi.	S14W076
ii.	S08W079	vii.	S11W077	xii.	S14W075
iii.	S08W078	viii.	S12W077	xiii.	S15W075
iv.	S09W078	ix.	S12W076	xiv.	S15W074
v.	S10W078	x.	S13W076	xv.	S16W073

d、 阿根廷冰河：

i.	S48W074	ii.	S49W074

iii. S49W075 vi. S51W073 ix. S52W074



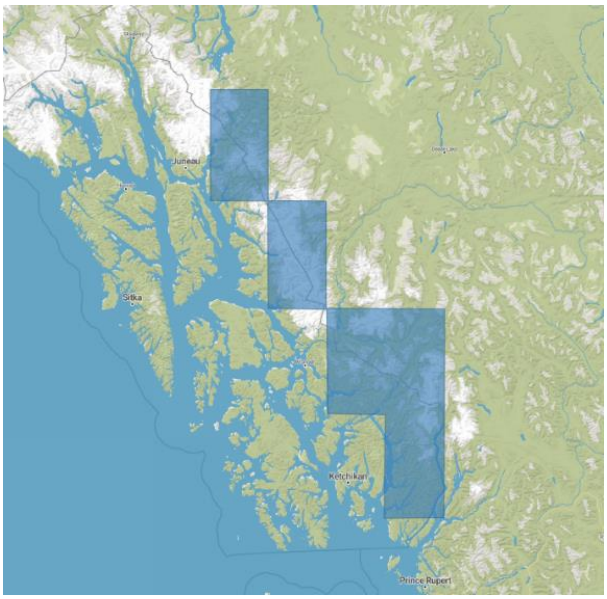
iv. S50W074 vii. S51W074

v. S50W075 viii. S51W075

e、 加拿大冰河：

i. N58W134 iii. N56W132 v. N55W131

ii. N57W133 iv. N56W131



【評語】190038

本研究利用 NASA 的 SRTM 1 Arc-Second 資料集來收集全球各地的地形高度圖(height-map)，也利用 MapTiler 網站收集相對應的衛星空照圖。利用這些收集的圖像，建構的 VAE-pix2pix 模型。該作品完成度非常的高。可以實際上線服務。若能多做些文獻探討，針對這樣的應用，找出關鍵議題然後進一步改善將更好。