

中原大學
資訊管理學系
碩士學位論文

以生成對抗網路為基礎將室內設計三維模
型轉換之研究

Research on Converting 3D Model of Interior Design based on Generative Adversarial
Network.

中原大學

指導教授：廖秀莉 博士

研 究 生：李承諺

中華民國 111 年 6 月

目錄

目錄	II
圖目錄	IV
表目錄	V
第一章 緒論	1
1.1 研究背景	1
1.2 研究動機與目的	2
第二章 文獻探討	5
2.1 室內設計三維模型	5
2.2 卷積神經網路(Convolutional Neural Network)	6
2.3 生成對抗網路(Generative Adversarial Network, GAN)	9
2.4 條件式生成對抗網路(conditional GAN, cGAN)	11
2.5 Pix2pix	12
2.6 循環式生成對抗網路(CycleGAN)	15
第三章 研究方法	16
3.1 研究架構	16
3.2 資料蒐集	17
3.3 資料前處理	19
3.4 以 Pix2Pix 為基礎進行三維素模的轉換	19
3.5 以 CycleGAN 進行圖像風格轉換	22
3.6 參數設定	23
3.7 模型評估	24
第四章 實驗結果與分析	25
4.1 實驗環境	25
4.2 以 Pix2pix 為基礎的三維模型轉換結果	26
4.3 以 CylceGAN 為基礎的三維模型轉換結果	26

4.4 實驗結果與訓練資料進行比較	27
第五章 結論與未來展望	28
參考文獻	29



中原大學

圖目錄

圖 1-1 三維立面圖像轉換為擬真的照片	3
圖 2-1 為 SketchUp 所建立的三維模型	6
圖 2-2 卷積層操作簡易架構	8
圖 2-3 為 2x2 平均池化層與最大值池化層	8
圖 2-4 自動編碼器簡易網路架構示意圖	10
圖 2-5 生成對抗網路的基本架構	10
圖 2-6 cGAN 基本架構	12
圖 2-7 U-net 基本架構	14
圖 2-8 PatchGAN 基本架構	14
圖 2-9 Pix2Pix 實際應用的範例圖	14
圖 2-10 CycleGAN 基本架構圖	15
圖 3-1 系統流程圖	17
圖 3-2 三維素模與擬真照片對比圖	18
圖 3-3 將成對資料合併為新圖片	19
圖 3-4 CycleGAN 目標函式	23

中原大學

表目錄

表 1 編碼器詳細架構	20
表 2 解碼器詳細架構	21
表 3 判別器詳細架構	22
表 4 參數設定	24
表 5 硬體設備及實驗環境	25
表 6 Pix2pix 實驗數據	26
表 7 CycleGAN 實驗數據	27



中原大學

第一章 緒論

1.1 研究背景

室內設計的概念起源於美國，於經濟發展的需求演化而來，原本蓋建築物的目的只是為求居住，而之後慢慢延伸出使用者的需求或為了更改內部環境而開始考量室內設計。後來室內裝潢逐漸延伸出考慮到生活品質、居住品質、心理層面、視覺等因素，使人類能在生活、起居、心理、視覺等各方面得到無比的滿足，現在每個人住的家裡，裝潢前都有通過室內設計來設計出滿足各種層面需求的家，而家的安定，對每個人的生活或是成長過程都扮演著重要的腳色，因此室內設計在現今的社會上已成了無法缺少的一塊產業。

而早期在電腦還沒普及的時候，室內設計會運用手繪 2D 設計圖、透視圖等方式進行設計的展現。但往往在工程完成後，發現與當初的想像有相當大的落差，造成設計師及客戶雙方的困擾，嚴重的話甚至會引發法律糾紛。而電腦普及後，室內設計業也一直在不斷的進步，不斷尋求新的設計表達方式，過程中有在 2D 設計圖中加入合成影像，到現在三維模型的展現，也稱作 3D 模型。而至今也有大量的繪圖軟體能讓輕鬆地生成三維模型，建模的步驟大部分都是先拉出一個平面，接著拉高，將平面拉成一個立體，此時會是一個素模，接著就要貼材質將三維立體圖上色，就能將完整的三維模型製作出來。而現今網路上已有非常多火紅的線上建模工具例如 3ds-Max 以及 SketchUp 等等，這些工具都能清楚的將設計師所要表達的構想給表現出來，而市面上生產出大量的 3D 建模工具也讓我們了解現今三維模型在市場中有多麼重要。因為 3D 建模的出現，解決了許多在裝潢前設計師與客戶之間的認知差異，這些認知差異也能在裝潢前順利解決，以避

免商業糾紛。

之後在渲染引擎的出現，室內設計又有更一步的突破，渲染引擎能透過調整不同的參數，搭配各式各樣的材質及燈光設計等將 3D 建模出來的三維模型渲染為接近真實照片的 3D 擬真圖，3D 擬真圖的繪製需要設計師敏銳的視覺感官經驗，進行視覺化思考，並透過經驗及技術不斷地進行視覺的判斷及思考加以調整得到出來的結果，在現今 3D 擬真圖的呈現已是室內設計上最能夠清楚表達設計師構想的方式。

1.2 研究動機與目的

許多時候，在房子裝潢之前，客戶總有想要預先看到自己房子的真實樣貌之需求，一般而言，在裝潢前能看到最真實的房間照片能透過 3D 模型來幫助，或者將 3D 模型渲染成 3D 擬真圖，而現今已有許多工具能夠將模型渲染成 3D 擬真圖，但作為一個設計師，製作一張設計圖困難的並不是只在於使用工具的技術需求，更重要的是設計師的設計構想以及創意，因此想渲染出非常完美的 3D 擬真圖除了需花費時間利用軟體渲染也非常重視設計師的視覺化思考，且在繪製過程中若參數調整稍有誤差也會導致渲染時間過長或是渲染出來的效果不盡理想，而此時若想看到比三維模型更真實照片卻只能透過 Photoshop 等合成軟體來完成，需耗費大量時間及精力。

隨著時間的發展，近年來人工智慧的技術可以做到的事情已經越來越多，其中在圖像轉換這部分，已經有可以做到將建築三維立面圖像轉換為擬真的照片如圖 1-1，但市面上卻鮮少有對室內設計進行的研究，如果可以運用這項技術，將室內設計中的三維模型轉換為擬真照片，並利用人工智慧進行上色、貼材質，不僅能解決三維模型渲染成 3D 擬真圖所需要花費的時間成本及設計師技術需求，也減少了設計師在貼材質時的視覺化

思考。

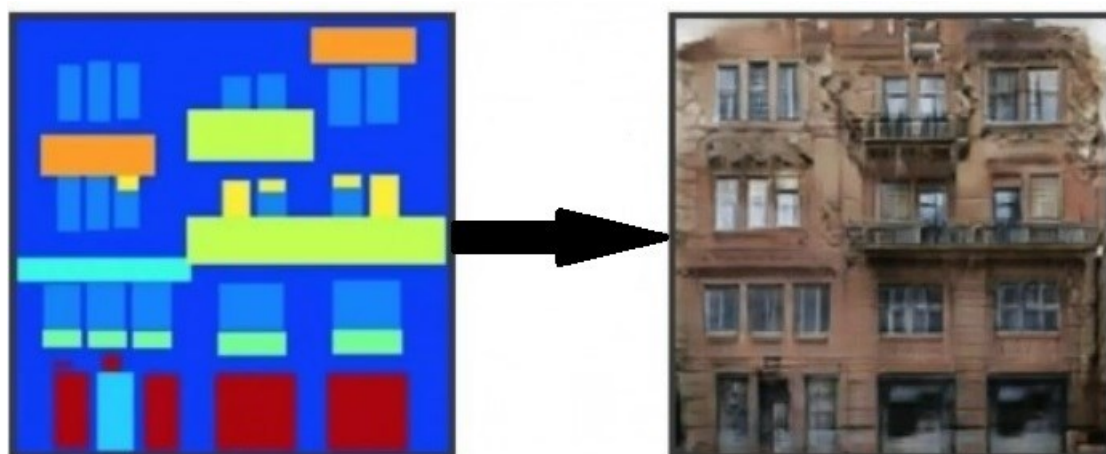


圖 1-1 三維立面圖像轉換為擬真的照片

因此本研究想以生成對抗網路(Generative Adversarial Network)來解決上述問題，分別蒐集三維模型以及對應的擬真照片進行網路的訓練。生成對抗網路由 Ian J. Goodfellow 等人於 2014 年所提出。為兩個神經網路之間的戰爭，分別為生成器(Generator)與判別器(Discriminator)，生成器利用現有圖片來生成新的假圖片，以欺騙判別器；判別器必須分辨出輸入圖片是否為原圖片還是生成器生成出來的假圖片，並給予回饋。接著不斷循環此過程，在過程中兩邊皆會改進並增強自己的方法，直到判別器無法區分出該圖片是真還是假。

其中因為對象為室內設計房間照片，雖然網路上已經有許多工具能輕鬆讓我們渲染出室內設計的擬真圖，而這些工具也都簡單容易上手，而我們之所以需要設計師，並不是需要他們使用工具的能力，而是需要設計師們設計的思考能力，因此最後呈現的房間擬真圖絕對是經過設計師精心設計過的，而設計的過程在整個製圖流程是最花費時間與精力的，因此如果能讓人工智慧自動幫我們設計，就能減輕設計師的設計負擔，也將能更快速地完成 3D 擬真圖提供客戶觀賞，而將未設計前的素模經過設計後差了就是顏色

的轉換，因此本研究會使用生成對抗網路中的變形 Pix2pix 來進行圖像的翻譯，將 3D 素模轉換成已貼完材質的 3D 擬真圖。

而每種設計就好比是一種設計風格，因此本研究也會採用生成對抗網路中的變形 CycleGAN 來進行風格轉換，將未上色的風格轉換為上色後的風格。最後訓練 Pix2pix 與 CycleGAN 兩種方法，並比較哪種訓練方法更適合運用在室內設計的領域。

Pix2pix 是一種用於圖像翻義的條件式對抗網路，主要的想法就是能夠將某一類的圖片轉換為另一類的圖片，而 CycleGAN 也同樣能夠將某一類的圖片轉換為另一類的圖片，例如能夠將一張白天的照片轉換為夜晚的照片。而這兩種方法皆是條件式對抗網路的延伸。雖然 Pix2pix 與 CycleGAN 都能夠將圖片進行轉換，但這兩種對抗網路的架構卻非常不一樣，最大的區別就是 Pix2pix 模型必須要成對的數據資料(Paired data)，而 CycleGAN 利用非成對的資料(unpaired data)也能夠進行訓練，因此以目前來說 CycleGAN 的實用性較高，使用程度也較為廣泛，但訓練出來的成果何者較佳，並沒有真正的答案，因此本研究想利用室內設計圖片的生成，對 Pix2pix 與 CycleGAN 來比較何種模型對生成圖片的成效較佳，並以室內空間圖片為例。

最後希望透過本研究訓練出來的模型，能夠快速地將三維模型轉換為擬真照片，解決時間、技術以及金錢問題以滿足客戶需求，讓沒有設計能力的人也能得到所需的照片。

第二章 文獻探討

2.1 室內設計三維模型

三維模型為一個物體以三維的方式表現，通常使用電腦或其他影像裝置進行顯示，顯示的物體可以是現實的物體，也可以是虛構的東西。現今三維模型已運用於許多產業，例如醫療行業製作器官的模型、電子遊戲業用於虛擬場景等，當然室內設計業也不例外，且三維模型的出現也大大的幫助室內設計業解決許多問題如圖 2-1。

許多時候在裝潢房子之前，客戶大部分會有預先看到自己房間樣貌的需求，冷翊(2018)提到早期室內設計會運用 2D 平面圖、透視圖、等方式進行設計表現，雖然能夠正確的表達設計師的設計構想，但仍有相當大的部分需要靠客戶自身的想像力，來綜合判斷完工後的實際樣態。而很多時候在工程完成後，可能會發現與當初的想像有相當大的落差，造成設計師及客戶雙方的困擾，嚴重的話甚至產生法律糾紛。隨著時間的發展，加上電腦的普及應用，三維模型也逐漸取代以往 2D 平面圖的方式呈現，作為記錄物體的方式。三維模型能夠具有多個特徵，在視覺上的呈現明顯優於 2D 平面圖，而三維模型主要以 3D 建模軟體來繪製，在台灣市面上已有許多成熟的 3D 建模軟體，例如 SketchUp(SKP)、3D Studio Max(3ds Max)等都是常見的 3D 建模軟體如圖 2-1，而建置 3D 模型的步驟通常為建置 2D 平面圖、建立素模、貼材質、最後渲染。而在貼材質這塊是對設計師來說最花費時間的一道流程，最後貼完材質再交給軟體去做渲染完成 3D 擬真圖。

渲染 3D 擬真圖為建模的最後一道程序，將原先建置的三維模型透過渲染引擎渲染成更接近真實 3D 場景的圖片，稱為 3D 擬真圖(張峻偉，2019)。而在渲染引擎這部分以

V-Ray 為主流，為 Chaos Group 公司於 1997 年所開發，常用的 3D 建模軟體都能與 V-Ray 進行結合，渲染的過程能夠將模型進行著色，也能因應不同材質做調整，更能顯示出建築的紋理、光源等，使得照片更為真實(謝淳鈺，2019)。



圖 2-1 為 SketchUp 所建立的三維模型

(資料來源：<https://reurl.cc/rQ150y>)

2.2 卷積神經網路(Convolutional Neural Network)

卷積神經網路是深度學習中極為重要的一塊領域，近年來電腦視覺能夠發展起來很大的原因都是因為卷積神經網路的出現，而此神經網路最大的特點就是「卷積」。

卷積是一種數學運算，可以對資料做特徵的萃取，增加資料的特徵，萃取出來的特徵再傳送到下一層卷積做特徵萃取。而卷積神經網路常用於圖片的處理，如圖片的識別，識別圖片的圖像是貓還是狗，或是英文字母的識別等，因此卷積層的輸出層一般採用多分類神經網路，而本研究採用的是對圖像的處理，因此以下會針對圖像的卷積層來做說明。

卷積神經網路是由成堆的卷積層和池化層所組成。卷積層就是利用卷積操作對圖像進行轉換，它是一個數字濾波器的集合，它會為特徵圖生成新圖像，而特徵圖萃取出了原始圖的獨特特徵。與其他神經網路相比，卷積層以非常不同的方式運行，該層不使用連接權重和加權和，但它包含了轉換圖像的過濾器，稱為卷積過濾器，而原始圖像會通過卷積過濾器輸進行過濾再輸入新的圖像，輸入圖像的過程產生新的特徵圖。卷積過濾器為一個二維矩陣，通常會是 3x3、4x4 矩陣等，卷積過濾器的操作過程是把過濾器與部分影像進行像素乘法，並加總其結果如圖 2-2。當影像過濾的那部分很接近過濾器時，輸出會是較大的正值，而當該部分影像與過濾器相反時，該數值會是較大的負值。而將過濾器由左至右、由上到下掃過整張影像並記錄每一部份的卷積輸出，就能根據過濾器的數值來取得某個特徵的新陣列。

由於卷積會將整張圖片掃過一遍並對每格像素進行計算，但此方式會造成計算耗能消耗過大，為了解決此問題也因此有池化層的出現。池化層則是將相鄰像素合併為一個像素，因此池化會降低圖片的維度，達到計算量降低的效果。池化和卷積一樣，也可選擇池化的大小，通常為 2x2 或 3x3，而池化合併出來的值通常為合併像素的平均值或最大值如圖 2-3。由於卷積神經網路的主要關注點是圖像，所以卷積層與池化層的概念是在二維平面上來操作(Kim, 2017)。

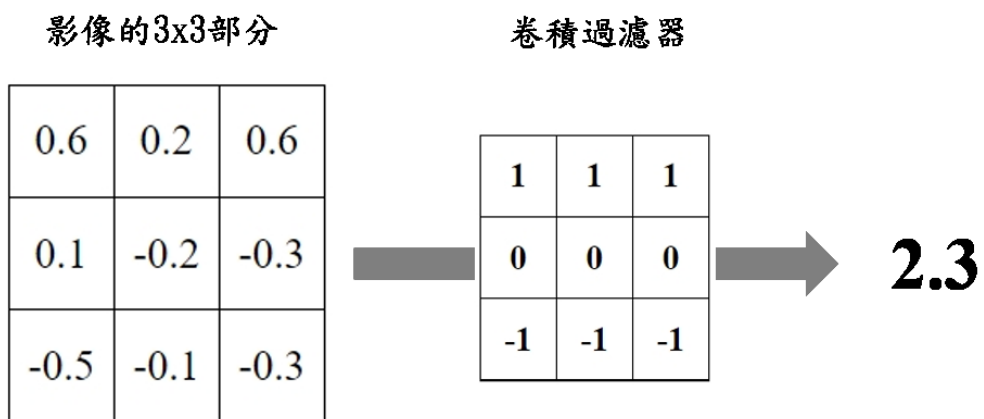


圖 2-2 卷積層操作簡易架構

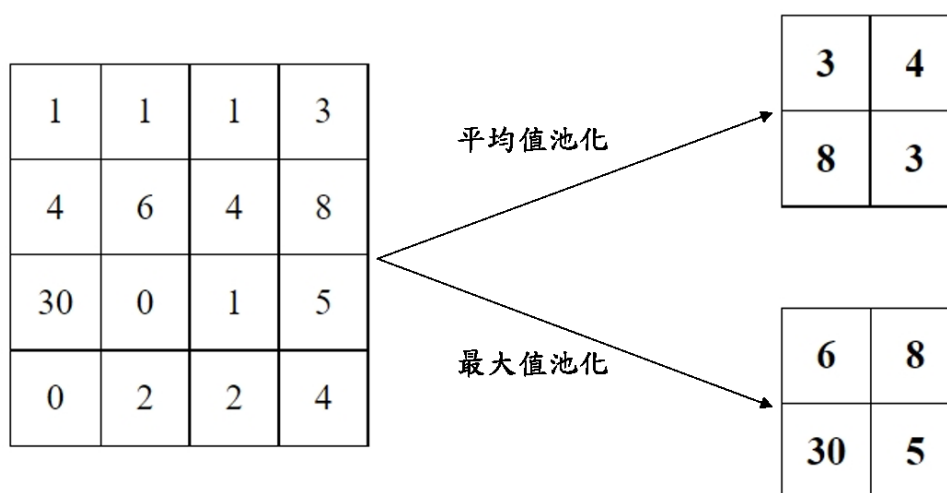


圖 2-3 為 2x2 平均池化層與最大值池化層

2.3 生成對抗網路(Generative Adversarial Network, GAN)

生成對抗網路(Generative Adversarial Network)簡稱 GAN，由 Ian J. Goodfellow(2014) 所提出。生成對抗網路為兩個神經網路所組成，分別為生成器(Generator)與判別器(Discriminator)。生成器由潛在空間取樣隨機數列 z 做為第一次的輸入接著生成圖片，再將生成出來的圖片輸入到判別器。

其中生成器使用的是自動編碼器(autoencoder)，自動編碼器是由編碼器(encoder)與解碼器(decoder)所組成的神經網路，編碼器用於將高維度的輸入資料壓縮成一個低維度的向量，此時的特徵向量為整個輸入資料最具代表性的精華，解碼器則是將特徵向量解壓縮回原始領域如圖 2-4。

判別器為一個任意的神經網路，對於圖片通常會使用卷積神經網路(CNN)。判別器會回傳一個分數(Scalar)給生成器，分數越大代表輸入的假圖片越接近真實的照片，生成器接收到分數之後調整參數繼續改良生成圖片，再輸入至判別器回傳結果，一直循環此步驟，經由兩個網路不斷互相對抗，直到生成器無法在生成出更真實的照片，判別器無法判斷出圖片的真或假，雙方最終達到一個在有對手條件的情況下的最佳平衡狀態，圖 2-5 為整個生成對抗網路的基本架構。

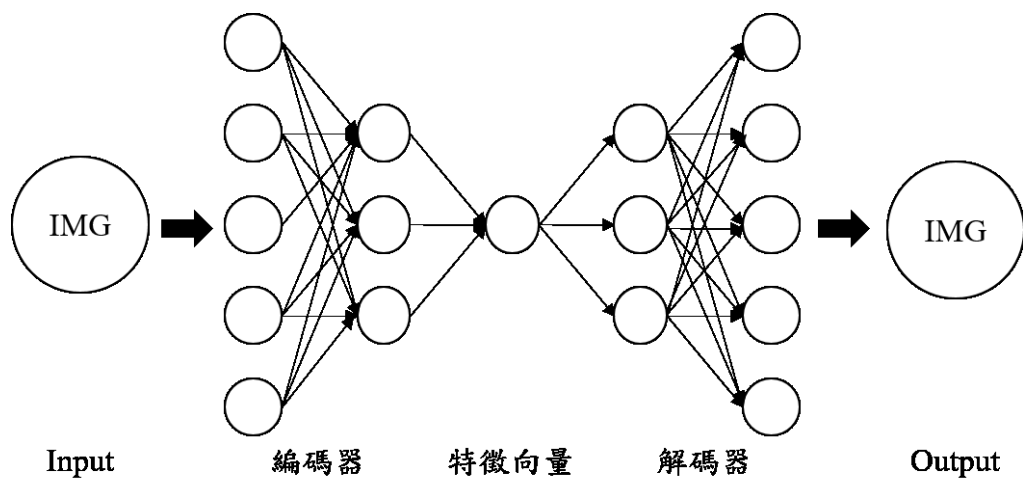


圖 2-4 自動編碼器簡易網路架構示意圖

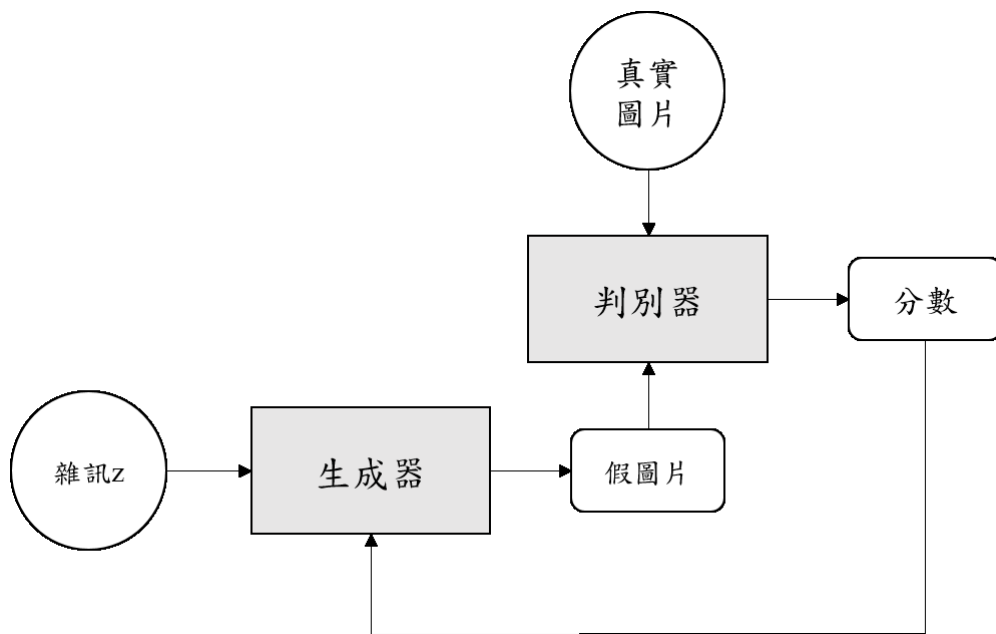


圖 2-5 生成對抗網路的基本架構

2.4 條件式生成對抗網路(conditional GAN，cGAN)

傳統的生成對抗網路在生成器訓練之前，會使用事前機率分布，隨機的挑選數據分布輸入至生成器中，再進行與判別器的相互競爭最終生成出接近真實世界的數據。此方式雖能生成出接近真實世界的圖片，但也存在著某些問題，因為數列都是隨機取樣的，我們不能控制它要生成出哪種圖片，GAN 只在乎生成出來的圖片與真實圖片相近與否，無法限制出生成後為哪種圖片。

為了改善此缺點，條件式生成對抗網路(conditional GAN)便誕生出來了，為 Mirza et al.(2014)所提出。相較於傳統的生成對抗網路，條件式生成對抗網路能夠在 GAN 訓練時加上一組條件，使得 GAN 變得更佳客製化。條件式生成對抗網路的生成器會在生成圖片之前加上一組隨機的數列 z 之外，也加上了一個輔助條件 c ，輔助條件 c 可以是文字、類別、圖片等各式的資料，而條件是生成對抗網路的判別器要輸入生成器所產生的圖片之外，也要輸入生成器的輔助條件 c ，此時判別器的工作變成要判別輸入圖片的真實度以及圖片與輔助條件 c 是否湊成一對，如果同時達到此目標那判別器的回傳分數就會越高，圖 2-6 為條件式生成對抗網路的基本架構。

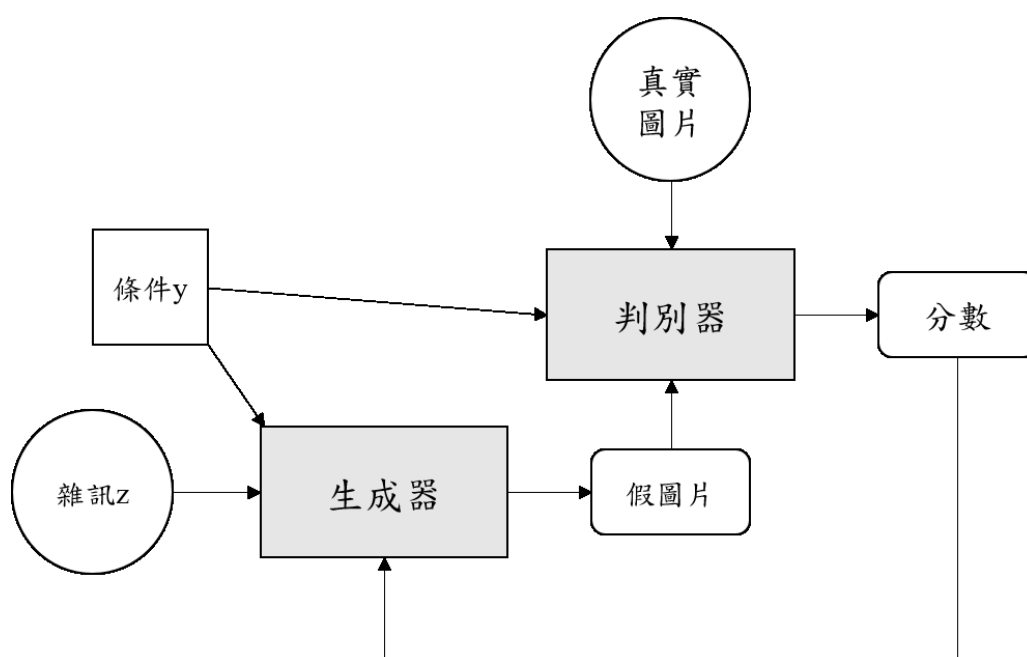


圖 2-6 cGAN 基本架構

2.5 Pix2pix

Pix2pix 是以 cGAN 為基礎做的改進，由 Isola et al.(2017)所提出，以 cGAN 能夠加入輔助條件的特性，將圖片當作輔助條件，跟著隨機數列一同輸入至生成器，再將生成出來的假圖片輸入至判別器去做出回饋。

生成器是採用 U-Net 的結構，其結構與傳統的 autoencoder 較為類似，分別有 8 層卷積層做為 encoder(編碼器)與 7 層反卷積層作為 decoder(解碼器)，但傳統的 autoencoder 存在著一個缺點，它的卷積層在提取特徵時會不斷地將圖像尺寸變小，在由變小後的特徵經由反卷積重組回與原圖尺寸相同大小的新圖像，這樣使得原圖許多細節資訊無法傳遞至 decoder，因此與傳統的 autoencoder 不同的是 U-Net 加入了拼接的思想，將第 i 層拼接至第 $n-i$ 層，其中 n 為總網路層數，這樣做是因為第 i 層與第 $n-i$ 層의 影像是一致的，方便他們同時輸入一樣維度的資訊，此時每一層反卷積的輸入都為前一層的輸出加上與該層對稱的卷積層的輸出，從而保證 encoder(編碼器)的資訊在 decoder 的資訊能夠

不斷地被重新記憶，使得生成的影像盡可能保留原影像的資訊，圖 2-7 為 U-Net 的基本架構。

判別器則引入了 PatchGAN 的思想，傳統的判別器是對整張影像去輸出一個分數，而 PatchGAN 的思想是對整張圖像切分成 $N*N$ 多張小圖像分別去做出判斷，分別判斷每一個區域的真假，最後再取其平均值作為判別器最後的輸出如圖 2-8。由於此方式對於判別器而言能更精確的判斷出每一個局部區域的真假，因此可以判斷出更加清晰的細節。此外 PatchGAN 相較於傳統的判別器每次判別的區域更小，判別所需的參數也更少，因此運行的速度也更加快速。目前 Pix2pix 也應用於許多地方，如將白天影像轉換成晚上，將素描畫轉換為真實照片、將 GPS 街景照轉換為真實街景如圖 2-9。



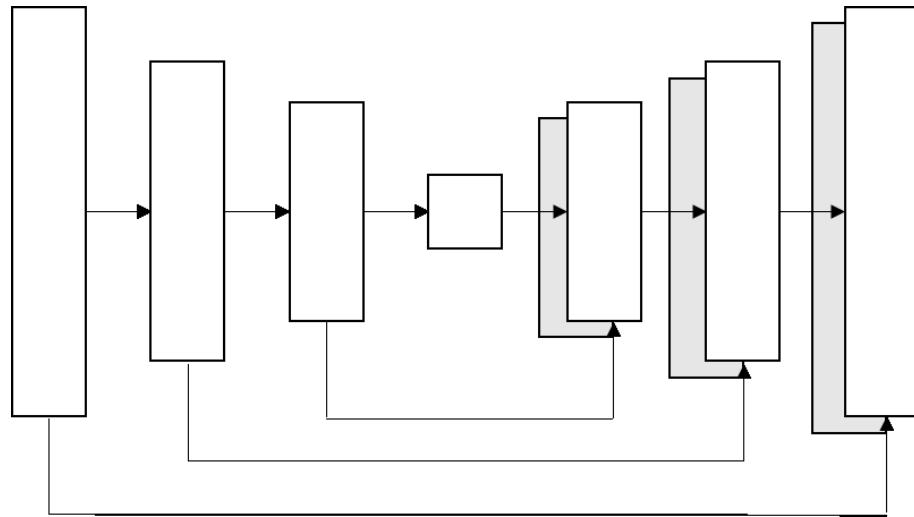


圖 2-7 U-net 基本架構

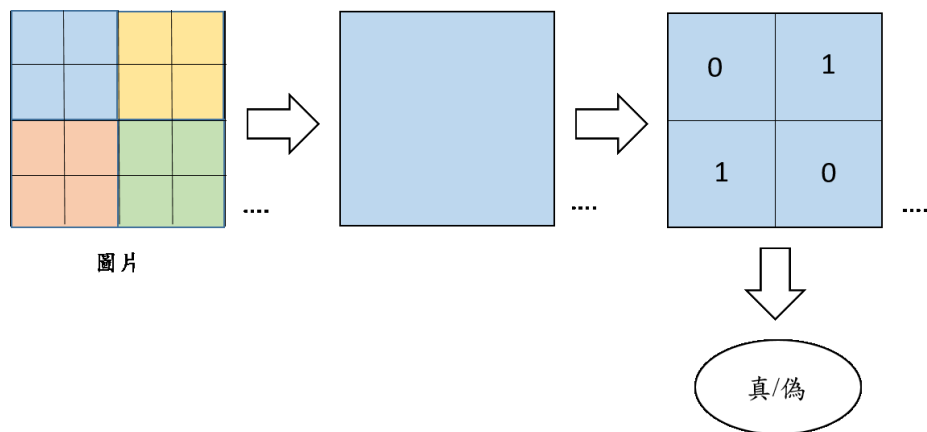


圖 2-8 PatchGAN 基本架構

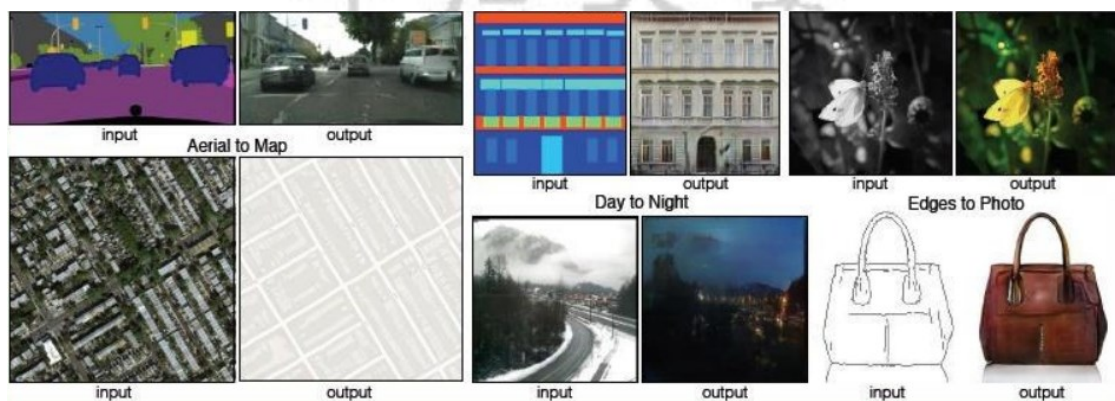


圖 2-9 Pix2Pix 實際應用的範例圖

(資料來源：Isola et al., 2017)

2.6 循環式生成對抗網路(CycleGAN)

循環式生成對抗網路(CycleGAN)由 Jun-Yan Zhu et al.(2017)提出。循環式生成對抗網路是基於條件式生成對抗網路的概念，而他最大的特點就是採用無監督式學習，也就是說它可以不需要成對的資料就可以進行訓練。循環式生成對抗網路的主要概念是分別訓練兩對生成器與判別器的模型，其中一對生成器與鑑別器要生成的任務為將 A 領域轉換為 B 領域的圖片。

但由於 CycleGAN 是一個沒有成對資料的訓練網路，因此往往在生成圖片時可能無法預期轉換為 B 領域的圖片與 A 領域有關聯，因此我們需要第二對生成器與判別器，此生成器與判別器的任務就是將 B 領域的圖片再轉換回 A 領域的圖片，並且轉換回去的圖片要與原圖片有所關聯，這樣才可能生成出來的圖片才會是有意義的，如圖 2-10 所示，而 CycleGAN 也運用在許多地方，如圖片的風格轉換等。

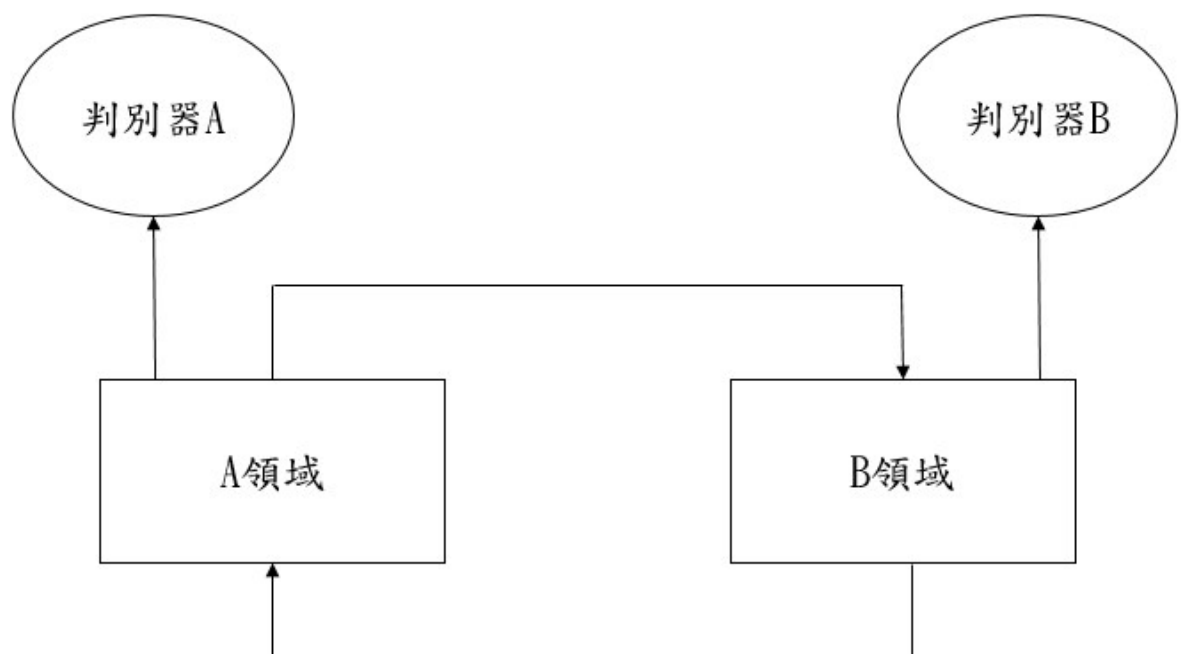


圖 2-10 CycleGAN 基本架構圖

第三章 研究方法

本章節將對室內設計三維模型轉換研究做出詳細介紹。分為研究架構、資料蒐集、資料前處理、生成對抗網路、參數設定、模型評估。

研究架構為本研究的整體研究流程，資料蒐集為三維模型與對應真實照片的蒐集方式，生成對抗網路為三維模型轉換為擬真照片的基本網路運用，模型評估為 Pix2pix 與 CycleGAN 兩個生成網路生成的結果進行評估比較。

3.1 研究架構

本論文主要分為六個部分，第一部分為室內設計三維模型與真實照片的資料蒐集。第二部份為進行資料的前處理，將所有圖像轉換為 256x256，並將所有成對的照片合併為 256x512 並另存為新照片以供 Pix2pix 模型進行訓練。第三部分為使用 Pix2pix 進行圖片轉換將三維模型轉換為擬真照片。第四部分為使用 CycleGAN 為將三維模型轉換為擬真照片。第五部分為調整 Pix2pix 與 CycleGAN 的模型參數。第六部份進行模型評估。其中流程如圖 3-1 所示。

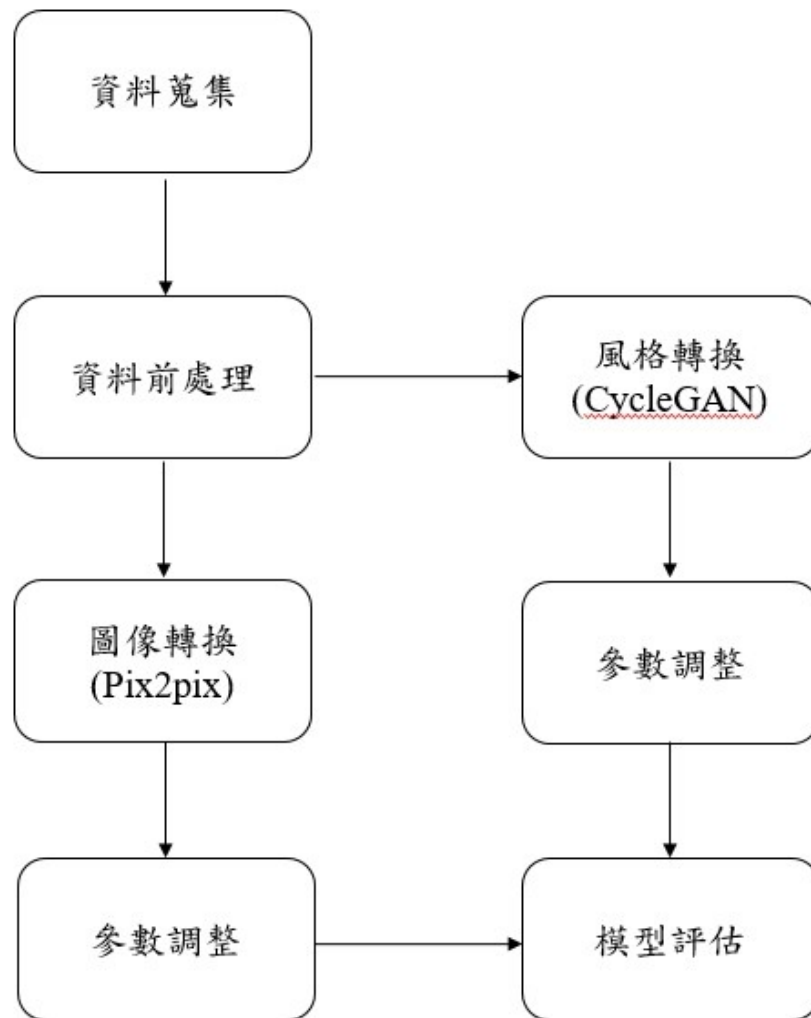


圖 3-1 系統流程圖

3.2 資料蒐集

本研究進行三維素模與擬真照片的轉換，需要使用到數量龐大的圖片當資料集，且使用 Pix2pix 進行圖像翻譯，須找出有相互對應的室內圖像，同樣的場景各有一張三維素模以及真實照片作為對應當成訓練資料。在 CycleGAN 進行風格轉換，將三維素模當成一種風格，轉換為擬真照片的風格，而用 CycleGAN 進行訓練不需要相互對應的資料，只需要蒐集大量的三維素模的圖像以及真實照片的圖像進行訓練。在資料的蒐集上，本研究選擇臥室的照片來當成主要的資料集，以利於神經網路的學習。接著再利用

SketchUp 進行臥室的建模，並繪製風格相近的不同臥室空間，且選擇坪數皆在 10 坪以內的套房為主，建立完素模後將房間擷取下來，接著貼材質並使用 V-Ray 進行渲染，渲染成擬真照片後再擷取出與素模相同角度的照片當成對應照片，以利於 Pix2pix 的學習。

然而在蒐集圖片之後，擔心所需的圖片與實際所需的訓練資料還有一段差距，因此我們採用資料增強的方式，來增加資料集。常見的資料增強方法有將圖片進行旋轉、平移、縮放、翻轉、拉伸等，但過度的變形會導致模型過擬合，因此為了考慮室內照片的特性，此研究採用水平翻轉的方式來增強資料集，以達到所需的訓練量。



圖 3-2 三維素模與擬真照片對比圖

3.3 資料前處理

資料蒐集後，擔心每張圖片的大小均不相同，為了模型方便訓練，在訓練之前會先將每張圖片的大小進行調整，調整成大小均相同的圖片，因此本研究將擷取下來的房間圖片等比例縮小至 256x256，方便模型統一訓練。且本研究會使用到 Pix2pix 進行訓練，Pix2pix 需要成對的資料，因此本研究會先將所有的成對圖片合併為一張 256x512 的圖片，並另存成新圖片，以方便訓練時的標註如圖，且圖片的檔案皆為 JPG 檔案。最後再將此成對的圖片分為訓練資料與測試資料，比例為 9:1，最後再將資料輸入至模型進行訓練。



圖 3-3 將成對資料合併為新圖片

3.4 以 Pix2Pix 為基礎進行三維素模的轉換

本次實驗我們採用 Pix2Pix 來進行三維素模對擬真照片的轉換，我們參考 Kamyar Nazeri 於 GitHub 中使用的 places365 的資料集，同樣採用 256x256 的圖像作為輸入及輸

出，因此會先將已經合併好的成對資料 256x512 的圖片進行裁切，再輸入至模型進行訓練。生成器架構的編碼器(Encoder)由 8 層的卷積層組成，下採樣步長為 2，卷積後接著批次標準化以及負斜率為 0.2 的激活函式 Leaky-ReLU，來避免神經元死亡的問題。

解碼器(Decoder)一樣由 8 層的卷積層所組成，上採樣步長為 2，並且將每 i 層的編碼器與 n-i 層的解碼器做連接，接著批次標準化以及激活函式 Leaky-ReLU，在最後一層使用 tanh 激活函式，使網路生成的值限制在-1 到 1 之間，最後輸出的維度為 256x256x3 的圖像。

表 1 編碼器詳細架構

Layer	Feature Size	Number of Filters
Input	256x256	3
Conv2d_1	128x128	64
Conv2d_2	64x64	128
Conv2d_3	32x32	256
Conv2d_4	16x16	512
Conv2d_5	8x8	512
Conv2d_6	4x4	512
Conv2d_7	2x2	512
Conv2d_8	1x1	512

表 2 解碼器詳細架構

Layer	Feature Size	Number of Filters
DeConv_1	2x2	512
concatenate_1	2x2	1024
DeConv_2	4x4	512
concatenate_2	4x4	1024
DeConv_3	8x8	512
concatenate_3	8x8	1024
DeConv_4	16x16	512
concatenate_4	16x16	1024
DeConv_5	32x32	256
concatenate_5	32x32	512
DeConv_6	64x64	128
concatenate_6	64x64	256
DeConv_7	128x128	64
concatenate_7	256x256	128
DeConv_8	256x256	3

判別器我們採用 PatchGAN 架構，卷積層由 4 層的卷積層所組成，步長為 2，並在每個卷積層後接一個批次標準化與激活函式 Leaky-ReLU，在最後一層使用 PatchGAN，

判斷 16x16 的各個區域為真或假的機率，最後將機率的平均值作為判別器的輸出，如表 3。

表 3 判別器詳細架構

Layer	Feature Size	Number of Filters
Input	256x256	3
Conv2d	128x128	64
Conv2d_1	64x64	128
Conv2d_2	32x32	256
Conv2d_3	16x16	512
Conv2d_4	16x16	1

3.5 以 CycleGAN 進行圖像風格轉換

本次實驗使用的是 CycleGAN 進行三維素模與擬真照片的轉換。為了保持良好的生成效果 CycleGAN 會使用兩個生成器與兩個鑑別器，其中一個生成器 G 負責將三維素模轉換為擬真照片，再利用另一個生成器 F 將擬真照片轉換為三維素模，而每個生成器轉換之後都會有一個對應的判別器，它會負責判斷生成器生成出來的圖片真偽，最後會比較原圖與復原過後的圖片的相似程度，使其滿足 $F(G(x)) \approx x$ 和 $G(F(y)) \approx y$ ，目標函式如圖 3-3。

$$Loss_{cyc}(G, F, X, Y) = \frac{1}{m} \sum_{i=1}^m [F(G(x_i)) - x_i] + [G(F(y_i)) - y_i]$$

圖 3-4 CycleGAN 目標函式

CycleGAN 使用的生成器可以分成三個部分；編碼器、變換器與解碼器。原圖會輸入由 3 個卷積層組成的編碼器，並對圖片進行池化增加通道數並縮小圖片尺寸。產生出來的活化函式會送入由數個殘差塊(residual blocks)組成的變換器，本實驗使用 9 個。接下來，再由 2 層轉置卷積層(Transposed Convolution layer)組成的解碼器將影像尺寸放大還原，最後再由輸出層輸出最終的圖片。

判別器使用的是 PatchGAN 的架構，將每張圖片切成多個 70x70 的小圖片，並判斷各小塊影像的真實機率最後進行平均。圖片輸入的大小設定為 256x256，最後產出擬真風格的室內設計圖片。

3.6 參數設定

本研究使用 Pix2pix 與 CycleGAN 兩種對抗網路的模型進行訓練，這兩種模型使用的 Optimizer 是 Adam 優化器。一般來說，優化器的學習率(learning rate)會影響模型訓練的好壞，太高的學習率會造成模型的過擬合，太小則會花費過多時間學習。本研究使用的 learning rate 為 0.0002，Batch size 為 4，LeakyReLU 的負斜率設定為 0.2。而兩種模型皆會嘗試以不同的 epoch 來訓練模型，太少的 epoch 會造成模型的欠擬合，過多的 epoch 則會造成模型的過擬合，因此本研究要找出最合適的 epoch，使得模型剛好擬合，兩種模型一開始皆會使用 200epoch、300epoch、400epoch 來進行訓練，生成出不同 epoch 產生出來的圖片，接著進行比較，找出最合適的 epoch。

表 4 參數設定

訓練參數名稱	參數值
Batch size	4
Leaky ReLu、ReLU 斜率	0.2
Optimizer	Adam
Learning rate	0.0002
epoch	200~400

3.7 模型評估

最後會訓練出一套 Pix2pix 圖像翻譯的訓練模型以及 CycleGAN 風格轉換的訓練模型，並對這兩種模型的生成結果進行比較。而 Pix2pix 的模型會先訓練部分資料，由少至多訓練上去，與 CycleGAN 進行比較，觀察以成對資料訓練的方式是否能以少數資料就能與 CycleGAN 訓練出差不多的結果。而 CycleGAN 則會直接由所有的資料進行訓練。最後觀察以 Pix2pix 進行訓練的模型以及 CycleGAN 進行的訓練模型哪種更適合運用於室內設計的三維模型轉換。

第四章 實驗結果與分析

本章節介紹實驗結果，並比較結果的差異。本研究資料集採用 SketchUp 進行圖片的繪製，一共蒐集到了 12000 張，其中有一部分的圖片是由水平翻轉的方式蒐集而來，且三維素模與擬真圖片各占一半，也就是各 6000 張來進行實驗。第一節會介紹本研究使用的硬體設備及實驗環境，第二節會先介紹 Pix2pix 所產生出的結果，第三節是 CycleGAN 所產生出的結果，第四節則會將實驗結果與用 SketchUp 製作的訓練資料來進行比較。

表 5 硬體設備及實驗環境

CPU	Intel i7-11800H
GPU	RTX 3050
RAM	64GB
作業系統	Windows10
CUDA	11.1
cuDNN	8.0.5
Python	3.9
Keras	2.7.0
pillow	8.3.1

4.1 實驗環境

本研究進行實驗所使用的硬體設備，如表 2 所示，顯示卡為 RTX3050，記憶體共 64GB，處理器使用 Intel i7-11800H CPU2.30GHz，軟體方面使用 Python3.9、Keras2.7.0、

pillow8.3.1。資料總預訓練時間為小時。

4.2 以 Pix2pix 為基礎的三維模型轉換結果

本階段實驗的訓練資料為三維素模與擬真照片的成對資料進行訓練，共蒐集了 12000 張，三維素模與擬真照片各為 6000 張，也就是共有 6000 組成對資料。而本階段實驗會抓取 6000 組圖片進行訓練。並將訓練資料與測試資料進行隨機的 9:1 的區分，因 6000 組的資料會有 5400 組的訓練資料，且以下實驗會再由不同的 epoch 數與不同數量的訓練資料分為實驗 1 至實驗 3，且學習率皆為 0.0002。

表 6 Pix2pix 實驗數據

	訓練資料數	epoch	Learning rata
實驗 1	5400 組	200	0.0002
實驗 2	5400 組	300	0.0002
實驗 3	5400 組	400	0.0002

4.3 以 CylceGAN 為基礎的三維模型轉換結果

本實驗使用循環式對抗網路進行訓練，因此不需要成對的資料，只需要蒐集大量的三維素模的圖片以及擬真圖片進行訓練，本階段實驗的資料共有 12000 張蒐集出來的圖片，三維素模與擬真照片各占一半，並將訓練資料與測試資料進行 9:1 的切分，因此訓練資料共有 10800 張，而為了與實驗 1 到實驗 3 進行對照，本階段實驗一樣會進行 200 次、300 次、400 次的 epochs 進行訓練，並分別為實驗 4 至實驗 6。

表 7 CycleGAN 實驗數據

	訓練資料數	epoch	Learning rata
實驗 4	10800 張	200	0.0002
實驗 5	10800 張	300	0.0002
實驗 6	10800 張	400	0.0002

4.4 實驗結果與訓練資料進行比較

實驗 ING...



第五章 結論與未來展望

待補。



中原大學

參考文獻

- Antipov, G., Baccouche, M., & Dugelay, J. L. (2017, September). Face aging with conditional generative adversarial networks. In *2017 IEEE international conference on image processing (ICIP)* (2089-2093). IEEE.
- Bi, X., & Xing, J. (2020). Multi-Scale Weighted Fusion Attentive Generative Adversarial Network for Single Image De-Raining. *IEEE Access*, 8, 69838-69848.
- Bu, Q., Luo, J., Ma, K., Feng, H., & Feng, J. (2020). An enhanced pix2pix dehazing network with guided filter layer. *Applied Sciences*, 10(17), 5898.
- Chang, Y. L., Liu, Z. Y., Lee, K. Y., & Hsu, W. (2019). Free-form video inpainting with 3d gated convolution and temporal patchgan. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (9066-9075).
- Demir, U., & Unal, G. (2018). Patch-based image inpainting with generative adversarial networks. *arXiv preprint arXiv:1803.07422*.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.
- He, K., Sun, J., & Tang, X. (2012). Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, 35(6), 1397-1409.
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (2961-2969).
- 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 (1125-1134).
- Karara, G., Hajji, R., & Poux, F. (2021). 3D Point Cloud Semantic Augmentation: Instance Segmentation of 360° Panoramas by Deep Learning Techniques. *Remote Sensing*, 13(18), 3647.
- Kim, P. (2017). Convolutional neural network. In *MATLAB deep learning* (121-147). Apress, Berkeley, CA.
- Li, G., Ma, B., He, S., Ren, X., & Liu, Q. (2020). Automatic tunnel crack detection based on u-net and a convolutional neural network with alternately updated clique. *Sensors*, 20(3), 717.
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (3431-3440).
- Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
- Morgenstern, O., & Von Neumann, J. (1953). Theory of games and economic behavior. Princeton university press.
- 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* (234-241). Springer, Cham.
- Weng, W., & Zhu, X. (2015). Convolutional Networks for Biomedical Image Segmentation.

IEEE Access.

Wu, H., Zhang, J., Huang, K., Liang, K., & Yu, Y. (2019). Fastfcn: Rethinking dilated convolution in the backbone for semantic segmentation. arXiv preprint arXiv:1903.11816.

Xie, C., Wang, Z., Chen, H., Ma, X., Xing, W., Zhao, L., ... & Lin, Z. (2021). Image Style Transfer Algorithm Based on Semantic Segmentation. *IEEE Access*, 9, 54518-54529.

Xu, Y., Wang, K., Yang, K., Sun, D., & Fu, J. (2019, September). Semantic segmentation of panoramic images using a synthetic dataset. In *Artificial Intelligence and Machine Learning in Defense Applications* (Vol. 11169, p. 111690B). International Society for Optics and Photonics.

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (2223-2232).

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. N. (2017). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (5907-5915).

Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11), 3212-3232.

于佩琴. (2014). 室內設計的本質性: 室內空間居家性之探討. 中原大學室內設計研究所學位論文, 1-83.

宋傑, 肖亮, 練智超, 蔡子贊, & 蔣國平. (2021). 基於深度學習的數字病理圖像分割綜

述與展望. Journal of Software, 32(5).

冷翊(2016)。以三維電腦繪圖為核心的室內設計流程及表現之研究。南華大學藝術與設計學院創意產品設計學系。

林庭生(2021)。以 Pix2Pix 與超解析度成像網路為基礎之金門老照片修復研究。國立金門大學資訊科技與應用碩士班

施旻岳 (2021)。以生成對抗網路為基礎之閩式建築風格轉換研究 (碩士論文)。國立金門大學資訊科技與應用碩士班。

張榮傑(2015)。基於語義分割之影片風格轉換。國立交通大學多媒體工程研究所

張峻瑋. (2019). 3D 效果圖擬真度影響設計發展之視覺思考研究. 中原大學室內設計研究所學位論文, 1-184.

楊詒鈞(2021)。生成對抗網路應用於多角度學習情緒辨識之研究(碩士論文)。國立中興大學資訊管理學系。

簡嘉琳(2021)。基於生成對抗網路的繪畫風格轉換(碩士論文)。國立宜蘭大學資訊工程學系研究所。