

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.

指導教授：廖秀莉 博士

研 究 生：李承諺

中華民國 111 年 6 月

摘要

室內設計產業在現今來說是不可或缺的一塊產業，許多時候客戶在裝潢住宅前會想預先看到自己房子內的真實樣貌，此時室內設計業就會提供擬真圖提供給客戶作為參考，擬真圖是利用建模軟體所繪製出來的。而建模的步驟為先設計出一個平面，接著拉高成為一個素模，再將素模繪製出有材質、有顏色、有燈光的 3D 模型，最後再利用渲染軟體將 3D 模型渲染為 3D 擬真圖。但從素模到 3D 擬真圖的繪製過程中會花費大量的時間以及視覺化思考。隨著時間的發展，人工智慧的技術也已經非常成熟，若擬真圖的設計可以利用人工智慧來製作將會省下許多時間以及視覺化思考等技術需求。本論文中使用生成對抗網路為基礎來學習室內設計的建模，將素模直接轉換為 3D 擬真圖，我們先利用 SketchUp 製圖蒐集大量的室內圖片，並使用 Pix2pix 以及 CycleGAN 對圖片進行訓練，視情況調整不同的訓練次數，從中分析出最佳結果接著與 V-Ray 渲染出來的圖片進行比較。實驗結果顯示，Pix2pix 對於素模與擬真圖的轉換有較不錯的表現，相比之下 CycleGAN 並不適合對素模與擬真圖進行轉換。最後生成出來的最佳結果與 V-Ray 渲染出來的擬真圖進行比較後能夠發現，生成對抗網路模型產生的圖片的確能自動設計好房間，生成速度也非常快，整體的架構也非常清楚，但相比之下利用 V-Ray 渲染出來的圖片，生成對抗網路產生的圖片細節並沒有那麼明顯。最後我們希望透過本論文的實驗，能夠快速地将素模轉換為擬真圖，解決時間、技術、以及金錢問題，讓沒有設計能力的人也可以獲得所需要的圖片。

關鍵字:生成對抗網路、室內設計、圖像翻譯、循環式生成對抗網路、三維模型、擬真圖、素模

Abstract

The interior design industry is an indispensable part of the industry today. Many times, customers want to see the real appearance of their houses before decorating their houses. At this time, the interior design industry will provide realistic pictures for customers as a reference, the immersive map is drawn using modeling software. The modeling steps are to first design a plane, then pull it up to become a prime model, and then draw a 3D model with materials, colors and lights from the prime model, and finally use the rendering software to render the 3D model into a 3D simulation. real picture. However, it takes a lot of time and visual thinking in the process of drawing from the original model to the 3D realistic drawing. With the development of time, the technology of artificial intelligence has also become very mature. If the design of the simulation map can be made by artificial intelligence, it will save a lot of time and technical requirements such as visual thinking. In this paper, we use the generative adversarial network as the basis to learn the modeling of interior design, and directly convert the original model into a 3D realistic image. We first use SketchUp to collect a large number of indoor images, and use Pix2pix and CycleGAN to train the images. Adjust different training times according to the situation, analyze the best results and compare them with the pictures rendered by V-Ray. The experimental results show that Pix2pix has a good performance for the conversion of pixel simulation and simulation images. In contrast, CycleGAN is not suitable for the conversion of pixel simulation and simulation images. After comparing the best result generated with the realistic image rendered by V-Ray, it can be found that the image generated by the generative adversarial network model can indeed automatically design the room, the generation speed is also very fast, and the overall architecture is also very clear , but in contrast

to the pictures rendered by V-Ray, the details of the pictures generated by the generative adversarial network are not so obvious. Finally, we hope that through the experiments of this paper, we can quickly convert the original model into a realistic image, solve the problems of time, technology, and money, so that people who have no design ability can also obtain the required images.

Keywords: Generative Adversarial Network(GAN) · Interior design · Pix2pix · CycleGAN · 3D model · 3D dynamic simulation drawing · Blank model



中 原 大 學

致謝

回想起碩士的這兩年，時光飛逝，能如期地完成學業，著實感動。同時感謝這兩年來一直在我身邊的人，陪伴著我完成這篇論文，要感謝的人實在太多，在此獻上最大的謝意。

首先要感謝我的指導老師廖秀莉老師的教誨，在課業與研究上給予我相當多的建議以及照顧。在老師細心栽培的教導之下，讓我在這兩年的碩士生涯獲益匪淺，也讓我了解到什麼是論文研究的正確態度，同時更幫助我拓展了不少人生視野。此外，還有感謝兩位口試委員劉士豪教授以及皮世明教授，給予了本篇論文許多寶貴的建議以及方向，使得本篇論文更加精進以及完善。

另外也要感謝所有研究室同學以及學弟妹，在研究的路途上彼此互相砥礪互相陪伴，總是一起討論並提供許多寶貴的意見，彼此之間也會相互照顧。在這兩年的光陰，不管是一起吃飯、聊天、打球、或是學術上的討論，都讓我的身心獲得充分的舒壓以及成長，所有悲歡離合、酸甜苦辣都跟著你們一起度過了，也將成為了我內心世界最肥沃的養分，這些透過同學以及學弟妹們所編織出來的美好回憶將會永遠留在我的心底。

最後，謝謝我的爸爸，因為人歸根究底還是需要一個溫暖的家庭，謝謝我的爸爸讓我擁有一個溫暖的家庭，爸爸也在無形之中成為了我最堅強的依靠。每次回到家裡爸爸的陪伴與鼓勵都能夠讓我再次振作，讓我能無後顧之憂學習並完成學業，衷心感謝家人無所求的付出與支持。

最後的最後，向曾經幫助過我的人致上滿滿的敬意與感謝。

目錄

摘要	I
Abstract.....	II
致謝	IV
目錄	V
圖目錄	VII
表目錄	IX
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	6
2.1 室內設計三維模型	6
2.2 卷積神經網路(Convolutional Neural Network).....	7
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)	10
2.5 Pix2pix(圖像翻譯).....	11
2.6 循環式生成對抗網路(CycleGAN)	14
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	20
3.5 以 CycleGAN 進行圖像風格轉換	23

3.6	參數設定	27
3.7	模型評估	28
第四章	實驗結果與分析	29
4.1	實驗環境	29
4.2	以 Pix2pix 為基礎的三維模型轉換結果	30
4.3	以 CycleGAN 為基礎的三維模型轉換結果	42
4.4	實驗結果與訓練資料進行比較	48
第五章	結論與未來展望	52
5.1	結論	52
5.2	未來展望	55
參考文獻	57



中原大學

圖目錄

圖 1-1 室內設計素模.....	2
圖 1-2 室內設計 3D 模型.....	3
圖 1-3 室內設計 3D 擬真圖.....	3
圖 1-4 三維立面圖像轉換為擬真的照片.....	4
圖 2-1 為 SketchUp 所建立的三維模型.....	7
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.....	9
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.....	11
圖 2-7 U-net 基本架構.....	13
圖 2-8 PatchGAN 基本架構.....	13
圖 2-9 Pix2Pix 實際應用的範例圖.....	13
圖 2-10 CycleGAN 基本架構圖.....	15
圖 2-11 CycleGAN 應用範例.....	15
圖 3-1 研究流程圖.....	17
圖 3-2 素模與相同角度的擬真圖片.....	18
圖 3-3 水平翻轉過後的圖片.....	19
圖 3-4 將成對資料合併為 256x512 的新圖片.....	20
圖 3-5 CycleGAN 目標函式.....	23
圖 4-1 實驗 1 生成結果細節處.....	34

圖 4-2 實驗 1 生成結果產生雜訊處.....	34
圖 4-3 epoch400 次時牆壁紋路已成形	37
圖 4-4 epoch400 次時門的邊緣較為銳化.....	37
圖 4-5 epoch400 次時能夠生成出門把	38
圖 4-6 第 500 次 epoch 生成出第三個門把	41
圖 4-7 書桌 epoch400 次與 epoch500 次對比.....	41
圖 4-8 臥室 epoch400 次與 epoch500 次對比.....	42



中 原 大 學

表目錄

表 1 編碼器詳細架構.....	21
表 2 解碼器詳細架構.....	21
表 3 判別器詳細架構.....	22
表 4 生成器 G 編碼器詳細架構.....	24
表 5 生成器 F 編碼器詳細架構.....	24
表 6 生成器 G 解碼器詳細架構.....	24
表 7 生成器 F 解碼器架構.....	25
表 8 判別器詳細架構.....	26
表 9 參數設定.....	27
表 10 硬體設備及實驗環境.....	29
表 11 Pix2pix 實驗數據.....	30
表 12 實驗 1 結果.....	31
表 13 epoch300 次與 epoch400 次的比較.....	34
表 14 epoch400 次與 epoch500 次的比較.....	38
表 15 CycleGAN 實驗數據.....	42
表 16 實驗 2 結果.....	43
表 17 epoch300 次與 epoch400 次的結果比較.....	46
表 18 Pix2pix 生成結果與 V-Ray 渲染結果比較.....	49
表 19 對 Pix2pix 最佳結果進行分析.....	52

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，之後慢慢延伸出使用者的需求或為了更改內部環境從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，現今絕大部分的裝潢，在動工前都會透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角色，因此室內設計在現今的社會上已成為了無法缺少的一塊產業。

早期電腦尚未普及的時候，室內設計會運用手繪平面圖、透視圖等方式進行設計的展現。但時常在工程完成後，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，室內設計業也一直不斷的進步，不斷尋求新的設計表達方式，過程中有在平面圖中加入合成影像，或者將平面拉高建立素模，到現在三維模型的展現，也稱作 3D 模型。而至今已有大量的繪圖軟體能夠輕鬆地建立三維模型，俗稱建模，現今網路上有許多非常火紅的建模工具例如 3ds-Max 以及 SketchUp，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，從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。而在 3D 建模的出現，也幫助室內設計業解決了許多在裝潢前後設計師與客戶之間的認知差異，以避免商業糾紛。

之後在渲染引擎的出現，室內設計業又有更一步的突破，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 擬真圖繪製的完整過程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，客戶總有想預先看到自己房屋內真實樣貌之需求，3D 擬真圖能幫助設計人員清楚的表達出設計的構想，而現今已有許多工具能夠將平面圖繪製成 3D 擬真圖，但作為一個設計師，想要繪製出一張 3D 擬真圖困難的並不是在於使用繪圖軟體的技術需求，更重要的是設計師的設計構想以及創意，因此想繪製出非常完美的 3D 擬真圖除了需花費時間利用建模軟體也非常重視設計師的視覺化思考。

而建模大部分的步驟都是先拉出一個平面，並將此平面拉高成一個立體，此時會是一個素模(圖 1-1)，接著將材質貼在素模中並選擇顏色進行上色以及調整燈光等，來完成一個完整的 3D 模型(圖 1-2)，最後再將 3D 模型配合渲染引擎將模型渲染為接近真實照片的 3D 擬真圖(圖 1-3)。

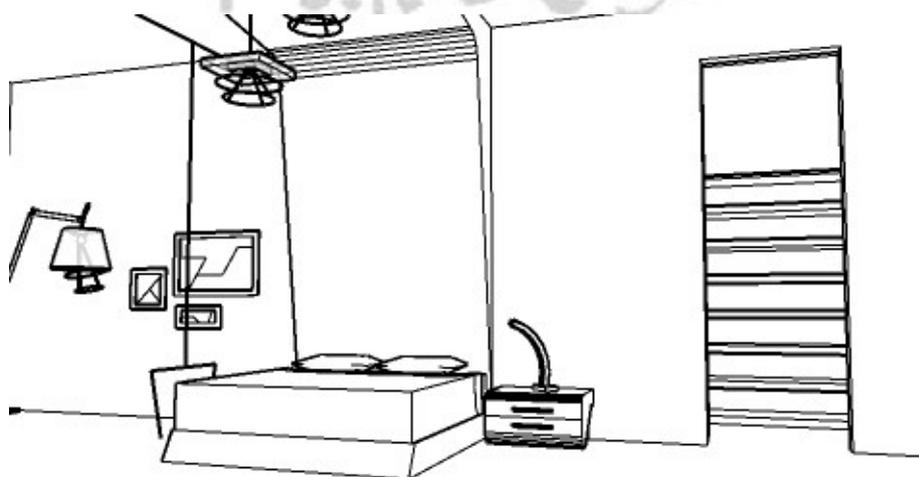


圖 1-1 室內設計素模



圖 1-2 室內設計 3D 模型



圖 1-3 室內設計 3D 擬真圖

但在素模繪製成 3D 模型的過程中，必須考慮到顏色、材質、風格等，並將整體房間設計的非常和諧，此時需要大量的視覺化思考需求，此步驟最花設計師的時間以及精力。而在繪製出 3D 模型後，將 3D 模型渲染成 3D 擬真圖的過程中，若是參數調整稍有誤差也會造成渲染時間過長或是渲染出來的效果不盡理想，而此時如果想看到比 3D 擬真圖更逼真圖片卻只能透過 Photoshop 等合成軟體來完成，需耗費更多的時間與精力。

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圖片(圖 1-4)，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，將室內設計中的素模轉換

為 3D 擬真圖，利用人工智慧進行上色、貼材質、渲染，不僅能解決繪製 3D 擬真圖中所需要花費的時間成本及設計師技術需求，也減少了設計師在創作時的視覺化思考。

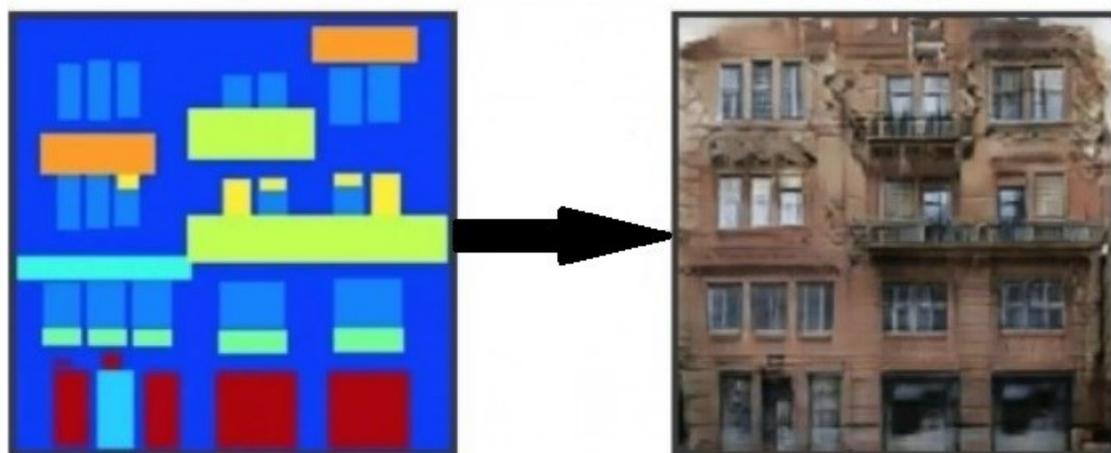


圖 1-4 三維立面圖像轉換為擬真的照片

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 3D 擬真圖，而這些工具也都簡單容易上手，但之所以需要設計師，並不是需要他們操作軟體的能力，而是設計師們設計的思考創作能力，因此最後呈現出來擬真圖絕對是經過設計師視覺化思考後所出來的，而在將素模設計成 3D 擬真圖是整個設計流程最花費時間與精力的步驟，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 擬真圖提供給客戶觀賞。

因此本研究想以生成對抗網路(Generative Adversarial Network)來解決上述問題，分別蒐集大量的素模圖片以及 3D 擬真圖進行網路的訓練。Ian J. Goodfellow 等人於 2014 年提出生成對抗網路，生成對抗網路為兩個神經網路之間的戰爭，分別為生成器與判別器，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，直到生成器無法生成出更逼真的圖，判別器無法區分出該圖片是真是假。

而素模以及擬真圖最大的差異就在於是否有材質、顏色、燈光，還有經過渲染引擎的渲染步驟，因此本研究會使用生成對抗網路中的變形 Pix2pix(圖像翻譯)來進行圖像的翻譯，將素模轉換成有材質、有顏色、有燈光的擬真圖。

再來本研究會將擬真圖當作是一種圖片風格，素模也作為一種風格，然後採用另一個生成對抗網路中的變形 CycleGAN 來進行風格轉換，將未上色的素模風格轉換為擬真圖片的風格。最後去比較 Pix2pix 與 CycleGAN 兩種方法，哪種生成對抗網路的訓練方法更適合運用在室內設計的領域。

Pix2pix 是一種用於圖像翻譯的條件式對抗網路，主要的想法就是能夠將某一領域的圖片轉換為另一種領域的圖片例如將三維立面圖轉換為建築物的照片。而 CycleGAN 為循環式生成對抗網路，也同樣能夠將某一領域的圖片轉換為另一領域的圖片，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，但對於室內設計領域的訓練的何者成效較佳，並沒有真正的答案，因此本研究想利用室內設計擬真圖的生成，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。接著再評估利用生成對抗網路生成出來的擬真圖以及利用渲染軟體渲染出來的擬真圖兩者的優缺點並進行比較。

最後希望透過本研究訓練出來的模型，能夠快速地將素模圖片快速地轉換為 3D 擬真圖，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圖片。

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。所以往往在工程完成之後，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，嚴重的話甚至會引發法律糾紛。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，而三維模型主要以建模軟體來繪製，在台灣市面上已有許多成熟的建模軟體，例如 SketchUp(SKP)、3D Studio Max(3ds Max)等都是常見的 3D 建模軟體(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、設計、貼材質、上色、調整燈光、最後渲染。而在設計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、上色、調整燈光完之後再交給渲染軟體去做渲染完成 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(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(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)。

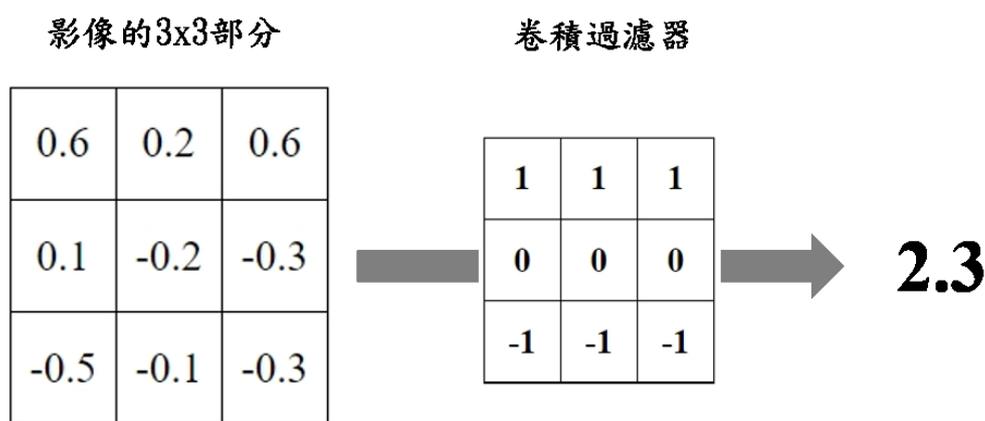


圖 2-2 卷積層操作簡易架構

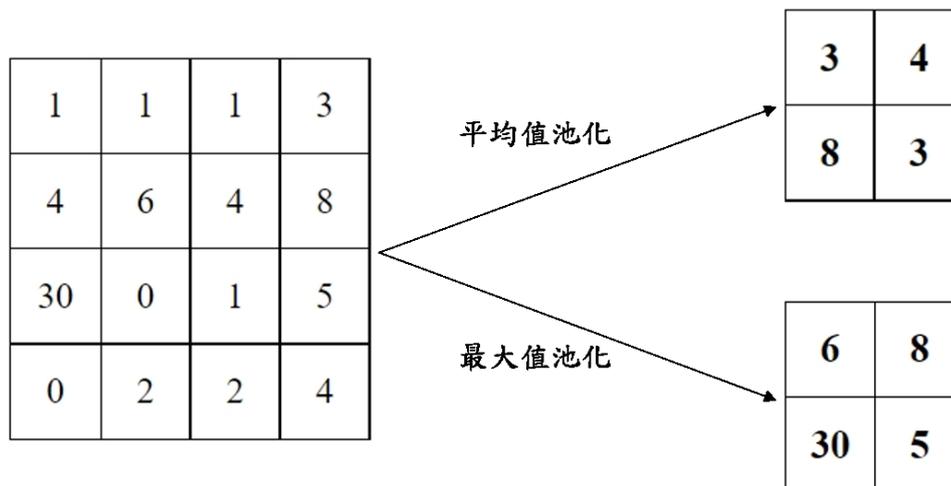


圖 2-3 為 2x2 平均池化層與最大值池化層

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(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。

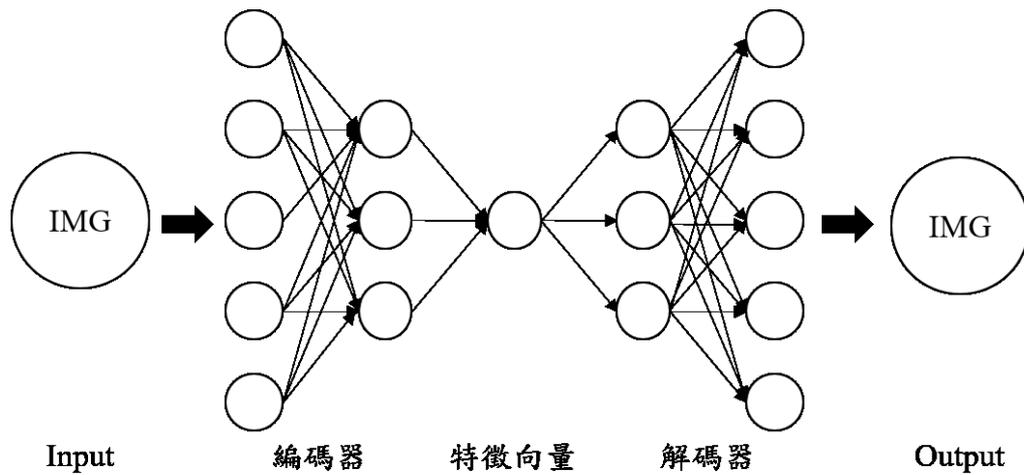


圖 2-4 自動編碼器簡易網路架構示意圖

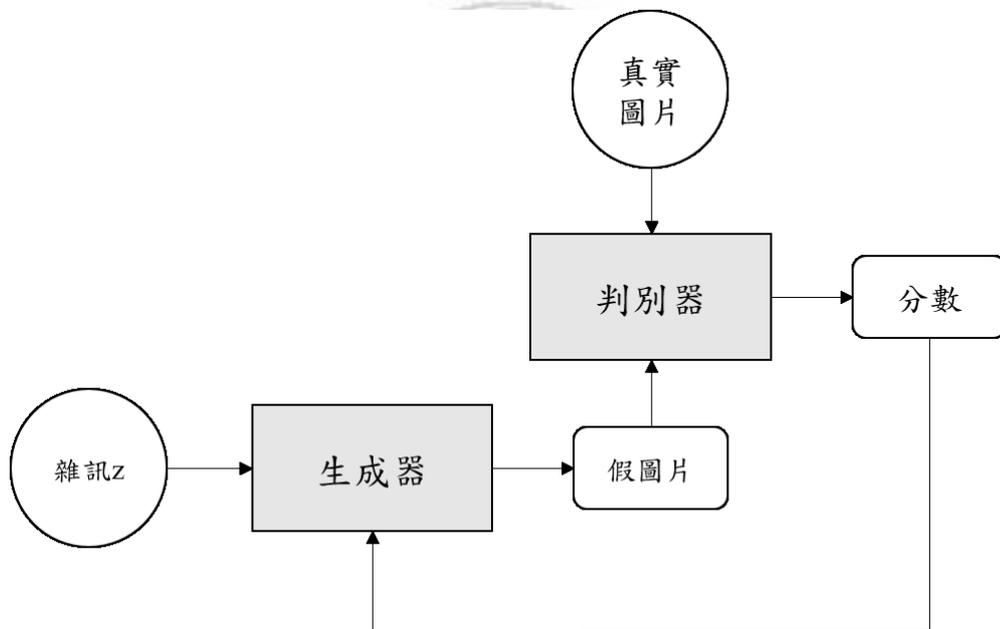


圖 2-5 生成對抗網路的基本架構

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，傳統的生成對抗網路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訓練時加上一組條件，使得模型變得更佳客製化。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，而條件式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。

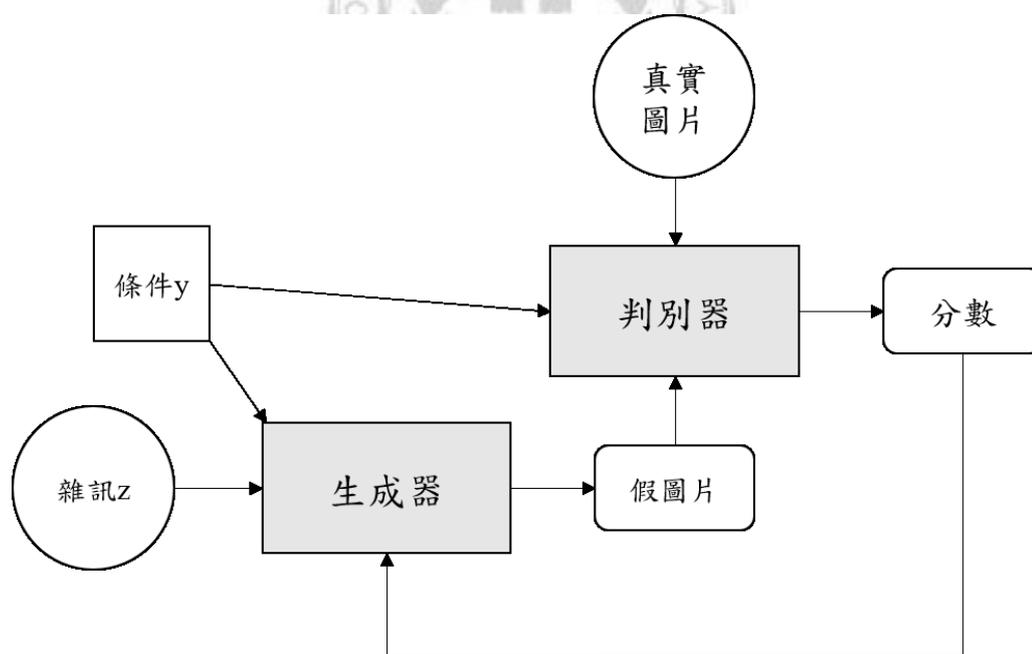


圖 2-6 cGAN 基本架構

2.5 Pix2pix(圖像翻譯)

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 8 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 $n-i$ 層的影響是一致的，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 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(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(圖 2-9)。

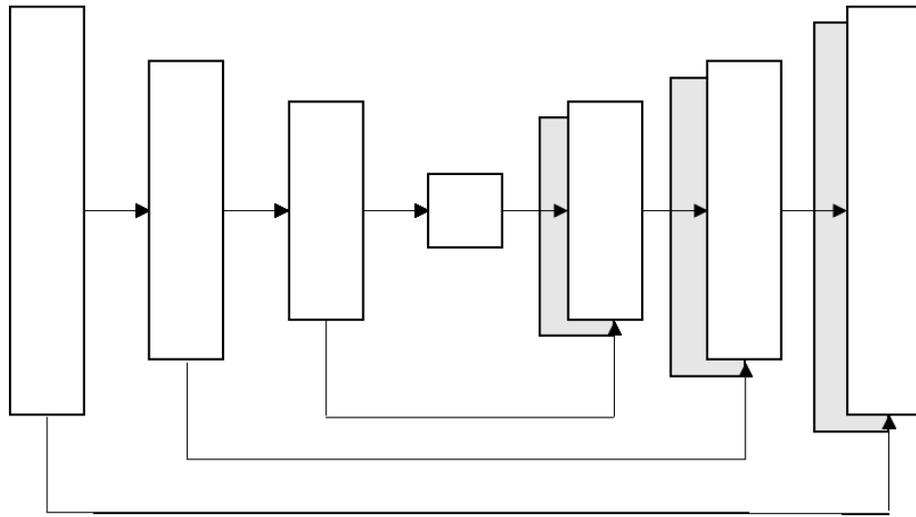


圖 2-7 U-net 基本架構

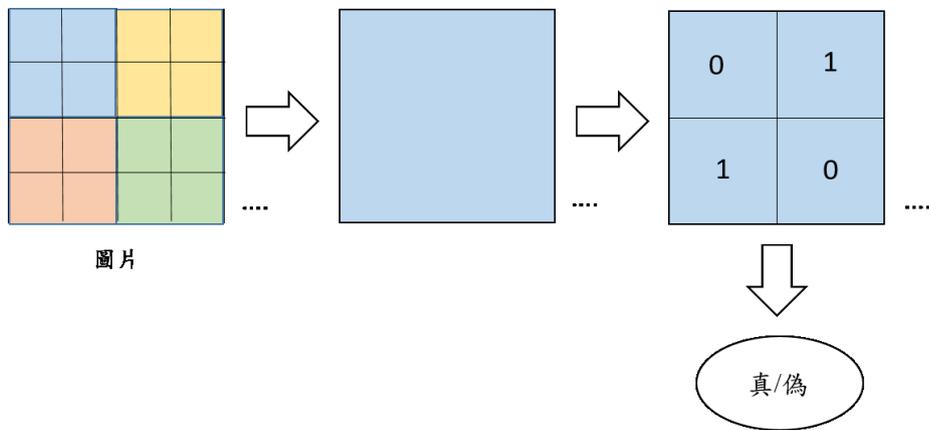


圖 2-8 PatchGAN 基本架構

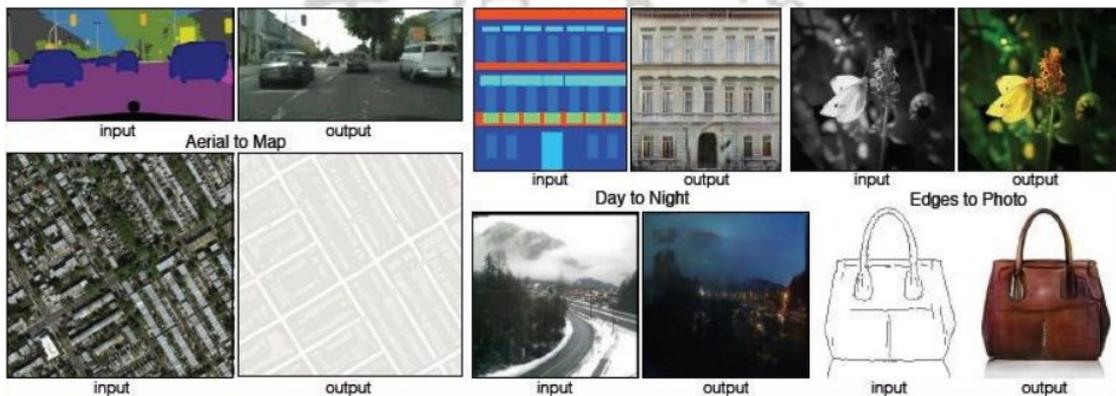


圖 2-9 Pix2Pix 實際應用的範例圖

(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組生成器與判別器的模型，其中一組生成器與判別器的任務為將 A 領域轉換為 B 領域的圖片並判斷生成出來的圖片與 B 領域圖片的相似程度，而 CycleGAN 的判別器採用的是 PatchGAN 的架構。

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，因此我們需要第二組生成器與判別器，此組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，這樣生成出來的圖片才會是有意義的，圖 2-10 為 CycleGAN 的基本網路架構。

CycleGAN 也運用在許多地方，能夠將圖片的風格轉換，如將馬的照片轉換為斑馬、將風景照轉換為畫像(圖 2-11)。

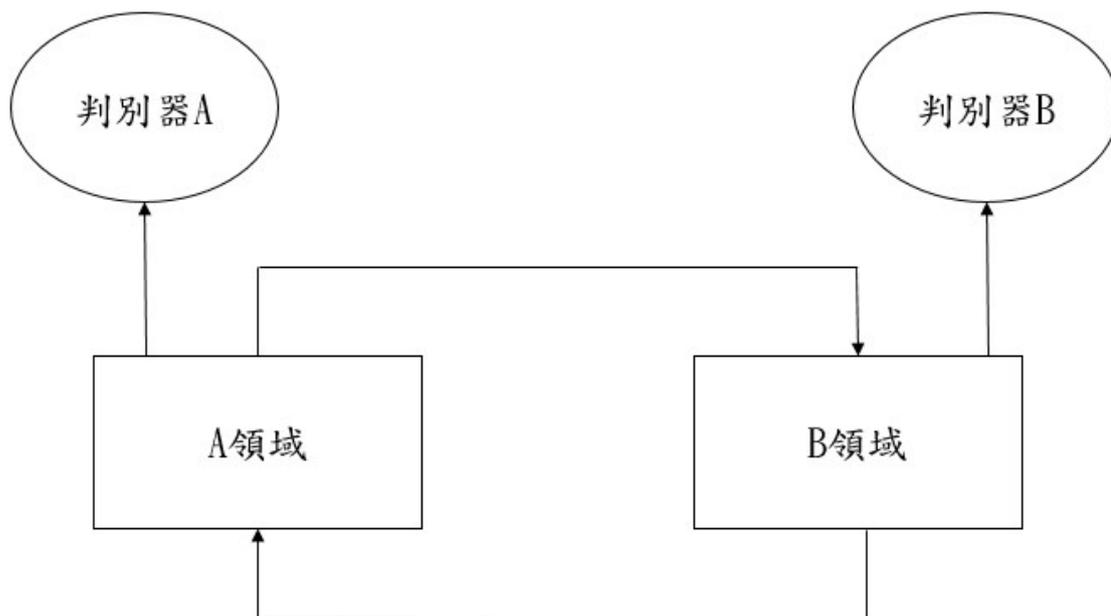


圖 2-10 CycleGAN 基本架構圖



圖 2-11 CycleGAN 應用範例

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，資料蒐集為素模與對應擬真圖片的蒐集方式，生成對抗網路為素模轉換為擬真圖片的基本網路運用，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

本研究主要分為六個部分，第一部分為室內設計素模與擬真照片的資料蒐集。第二部分為資料的前處理，先將所有圖像轉換為 256x256，接著再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 進行圖片轉換將素模轉換為擬真圖片。第四部分為使用 CycleGAN 為將素模轉換為擬真圖片。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。其中研究流程如圖 3-1 所示。

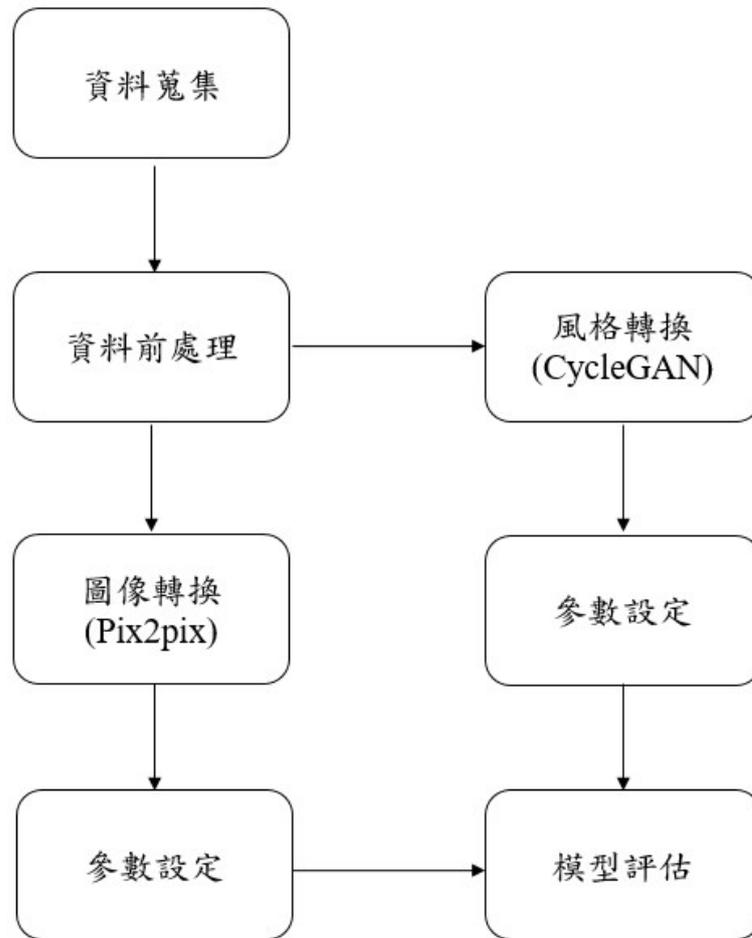


圖 3-1 研究流程圖

3.2 資料蒐集

本研究進行三維素模與擬真圖片的轉換，需要使用到數量龐大的圖片當作資料集。使用 Pix2pix 進行圖像翻譯前須找出有相互對應的室內圖像，同樣的場景各有一張素模圖片以及擬真圖片作為對應當成訓練資料。在 CycleGAN 進行風格轉換，本研究將素模當成一種風格，擬真圖作為另一種風格，並將素模的風格轉換為擬真圖的風格，而用 CycleGAN 進行訓練不需要相互對應的資料，只需要蒐集大量的素模圖片以及擬真圖進行訓練。在資料的蒐集上，本研究選擇以臥室來當成主要的資料集，並且坪數皆在 20 坪以內，以利於神經網路的學習。接著再利用 SketchUp 進行臥室的建模，並繪製數種不同臥室空間，建立完素模後將房間擷取下來，接著貼材質、上色、調整燈光，最後使用

V-Ray 進行渲染，渲染成擬真圖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，如圖 3-2 所示。

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，因此本研究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，但過度的變形可能會導致模型過擬合，因此為了考慮室內設計的特性，本研究採用水平翻轉的方式來增強資料集(圖 3-3)，以達到所需的訓練量。

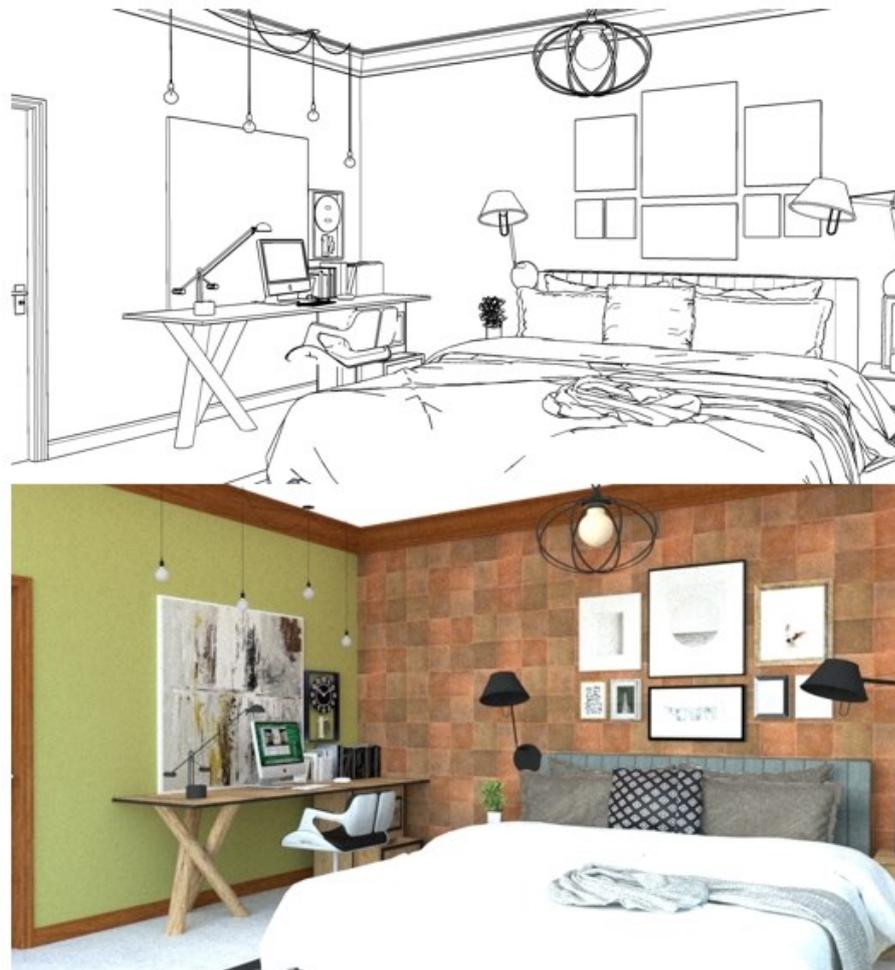


圖 3-2 素模與相同角度的擬真圖片



圖 3-3 水平翻轉過後的圖片

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，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。且本研究將使用 Pix2pix 進行訓練，Pix2pix 需要成對的資料，因此會先將所有成對的圖片合併為一張 256x512 的圖片，並另存成新圖片(圖 3-4)，以方便訓練時的標注，且圖片的檔案皆為 JPG 檔案。最後再將資料輸入至模型進行訓練。



圖 3-4 將成對資料合併為 256x512 的新圖片

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 來進行素模與擬真圖片的轉換，會先將已經合併好的成對資料 256x512 的圖片進行裁切，裁切成 256x256，並標注為同一組資料，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 的圖片。

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，並在每層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 進行素模與擬真圖片的轉換。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 負責將素模轉換為擬真圖片，再利用另一個生成器 F 將擬真圖片轉換為素模，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 3-5。

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

圖 3-5 CycleGAN 目標函式

CycleGAN 使用的生成器可以分成兩個部分，編碼器、解碼器。原圖會輸入由 4 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。接下來，再由 4 層反卷積層(Deconvolution 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，如表 4 至表 7。

表 4 生成器 G 編碼器詳細架構

Layer	Feature Size	Number of Filters
Input_1	256x256	3
Conv2d_1	128x128	64
Conv2d_2	64x64	128
Conv2d_3	32x32	256
Conv2d_4	16x16	512
Conv2d_5	16x16	1

表 5 生成器 F 編碼器詳細架構

Layer	Feature Size	Number of Filters
Input_2	256x256	3
Conv2d_6	128x128	64
Conv2d_7	64x64	128
Conv2d_8	32x32	256
Conv2d_9	16x16	512
Conv2d_10	16x16	1

表 6 生成器 G 解碼器詳細架構

Layer	Feature Size	Number of Filters
DeConv_1	16x16	512

DeConv_2	32x32	128
DeConv_3	64x64	64
DeConv_4	128x128	32
DeConv_5	256x256	32
DeConv_6	256x256	3

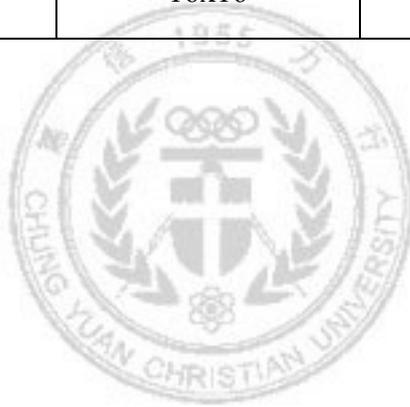
表 7 生成器 F 解碼器架構

Layer	Feature Size	Number of Filters
DeConv_7	16x16	512
DeConv_8	32x32	128
DeConv_9	64x64	64
DeConv_10	128x128	32
DeConv_10	256x256	32
DeConv_10	256x256	3

判別器使用的是 PatchGAN 的架構，由 4 層的卷積層，並在每一層卷積層後面接上激活函式 Leaky-ReLU 以及 Instance Normalization 代替 Batch Normalization，最後一層則使用 PatchGAN 的架構，將每張圖片切成多個 16x16 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。

表 8 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



中 原 大 學

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，太小則會花過多時間學習。本研究使用的 Learning rate 為 0.0002，Batch size 為 1，LeakyReLU 的負斜率設定為 0.2。訓練次數方面，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，因此本研究會先以 300 次 epoch 作為訓練，並先觀察 300 次 epoch 的結果，再適度地去調整 epoch 次數，找出最合適的 epoch。

表 9 參數設定

訓練參數名稱	參數值
Batch size	1
Leaky-ReLU 負斜率	0.2
Optimizer	Adam
Learning rate	0.0002
epoch	300

3.7 模型評估

本研究最後會訓練出一套 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。Pix2pix 與 CycleGAN 的模型皆會先以 epoch300 進行訓練，並將圖片列出來進行比較，再適度的調整 epoch 次數。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，並將生成出來的最佳結果與 V-Ray 渲染出來的擬真圖進行比較。



中原大學

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 進行室內房間的繪製，一共蒐集到了 8000 張圖片，素模與擬真圖片各占一半，也就是各 4000 張圖片來進行實驗。共包含 20 間不同的臥室，並將每間擷取出 200 張不同角度的圖片，其中有一部分的圖片是由水平翻轉的方式蒐集而來。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。

4.1 實驗環境

本研究進行實驗所使用的硬體設備，如表 10 所示，顯示卡為 RTX3050，記憶體共 16GB，處理器使用 Intel i7-11800H CPU2.30GHz，軟體方面使用 Python3.9、Tensorflow2.5.0、pillow8.3.1。資料總訓練時間為 260 小時。

表 10 硬體設備及實驗環境

CPU	Intel i7-11800H
GPU	RTX 3050
RAM	16GB
作業系統	Windows10
CUDA	11.1
cuDNN	8.2.1
Python	3.9
Tensorflow	2.5.0

pillow	8.3.1
--------	-------

4.2 以 Pix2pix 為基礎的三維模型轉換結果

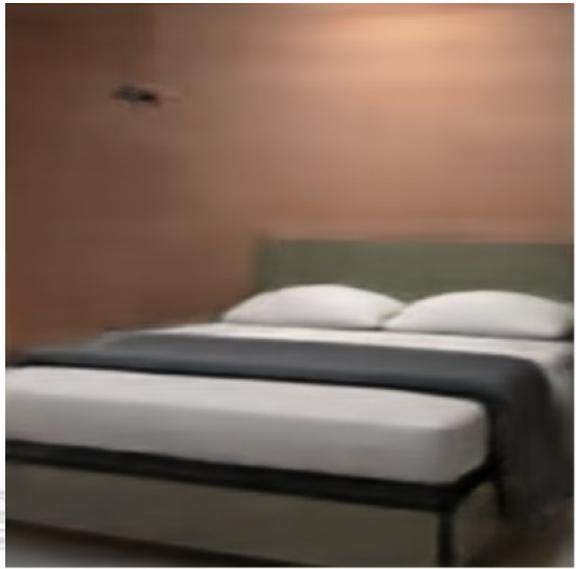
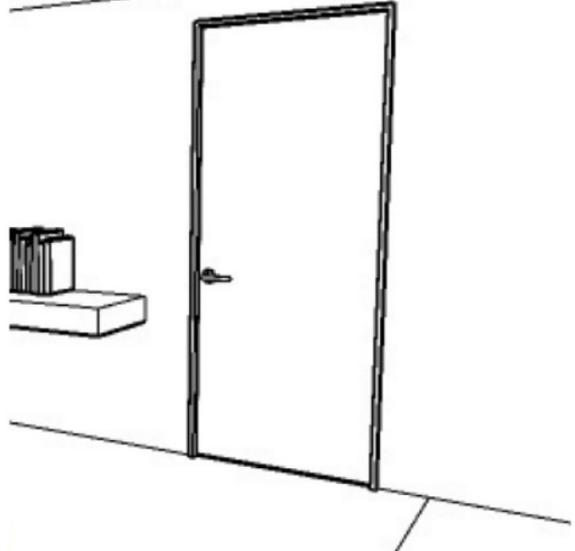
本階段實驗的訓練資料為素模與擬真圖片的成對資料，共蒐集了 8000 張，素模與擬真圖片各為 4000 張，也就是共有 4000 組成對資料。而本階段實驗會抓取 4000 組圖片進行訓練。且以 Pix2pix 作為訓練的實驗定為實驗 1，並以 300 次 epoch 且學習率為 0.0002 為參數作為設定。

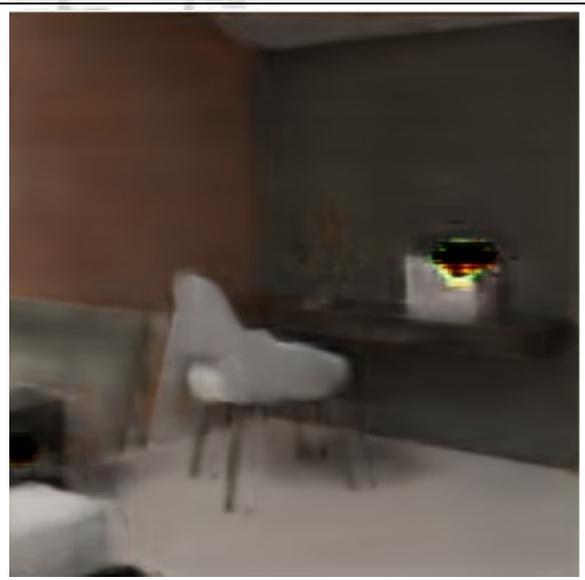
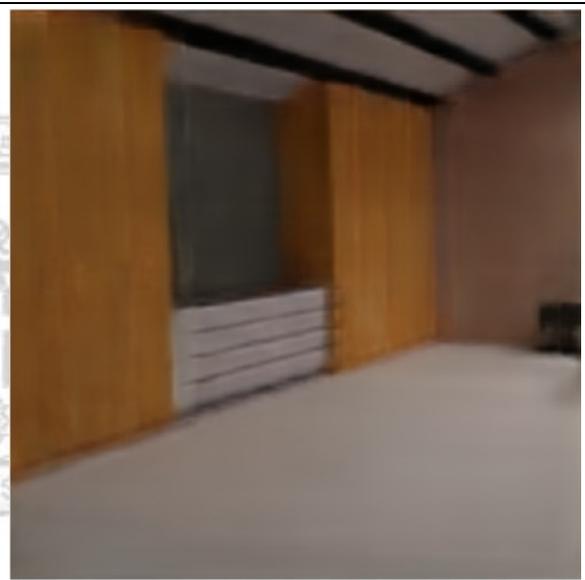
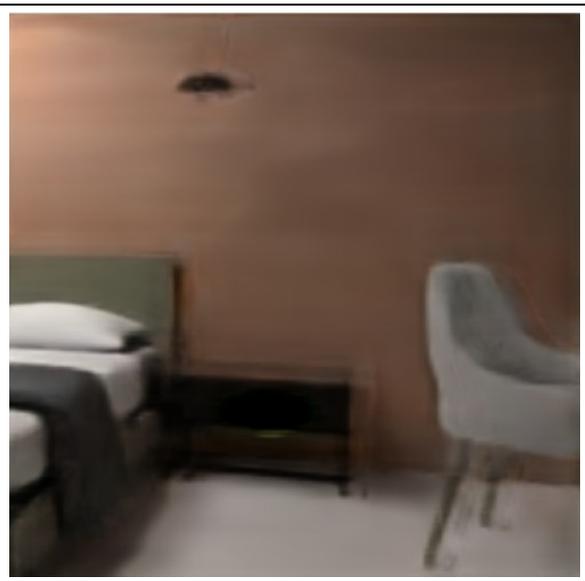
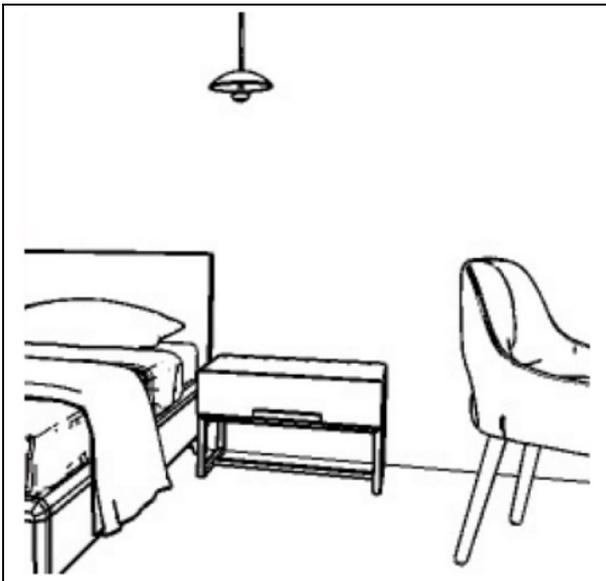
表 11 Pix2pix 實驗數據

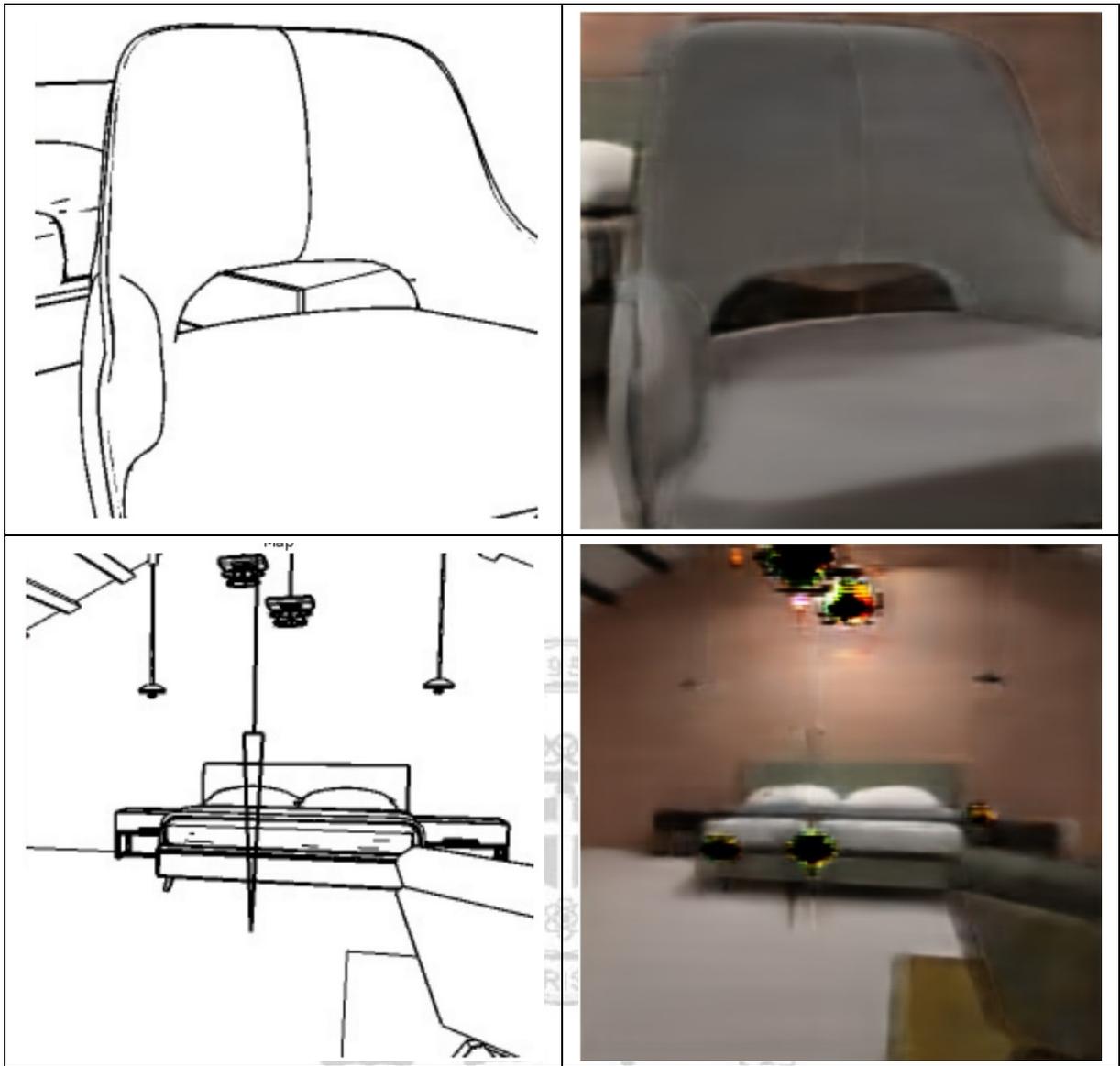
	訓練資料數	epoch	Learning rata
實驗 1	4000 組	300	0.0002

中 原 大 學

表 12 實驗 1 結果

素模	生成圖
 A black and white line drawing of a bed with a headboard, two pillows, and a blanket. A simple pendant lamp hangs from the ceiling above the bed.	 A 3D rendered image of a bed with a green headboard, white pillows, and a dark blue blanket. A small, dark pendant lamp hangs from the ceiling.
 A black and white line drawing of a door with a handle. To the left of the door is a small shelf with three books on it.	 A 3D rendered image of a door with a handle. To the left of the door is a small shelf with three books on it. The scene is dimly lit.





本次實驗結束之後，隨機挑選出了 7 張測試資料集，也就是沒有參與訓練的資料，並進行結果觀察。可以看出模型能夠將素模轉換為擬真圖片的樣貌，整體架構都生成的非常清楚，整體的生成結果是非常成功的。但也看得出來圖片普遍都較為模糊，且許多細節並沒有被生成出來，像是牆壁的紋路以及地板的紋路如圖 4-1 的紅框處，甚至還有出現雜訊的現象發生如圖 4-2 的紅框處，我們初步判斷是訓練次數不夠多，因此再往上加 100 次 epoch，共 400 次 epoch，並與 epoch300 次的實驗結果進行比較。

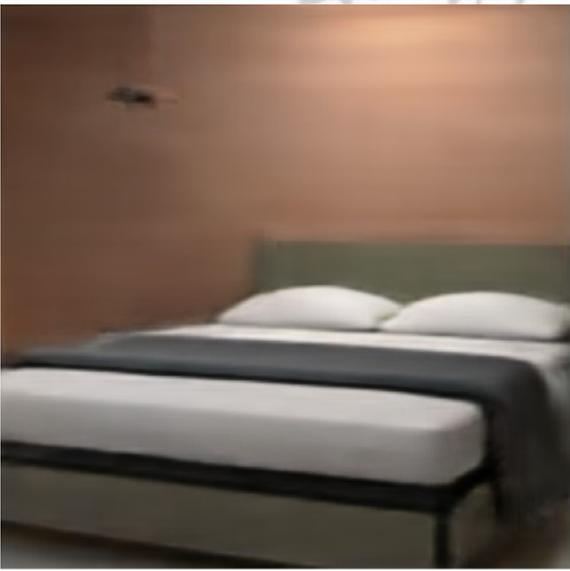
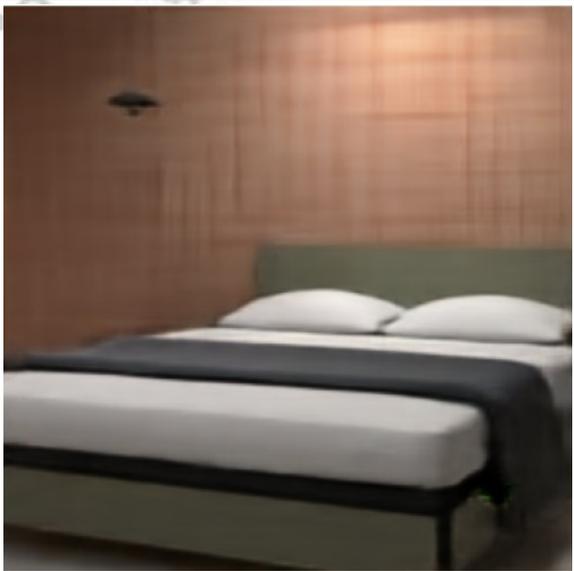


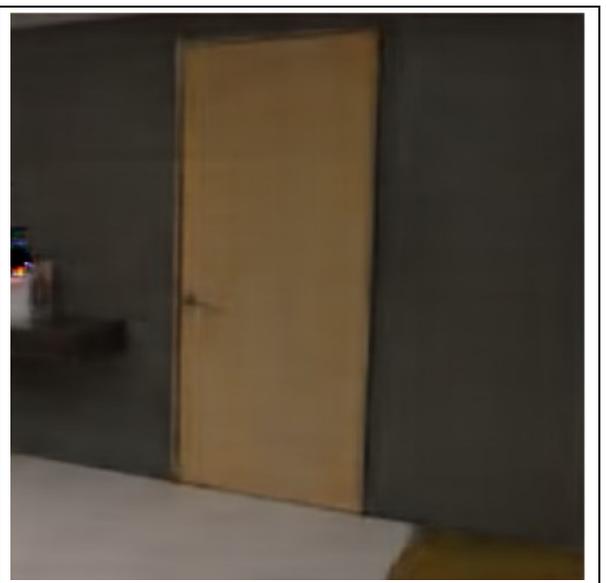
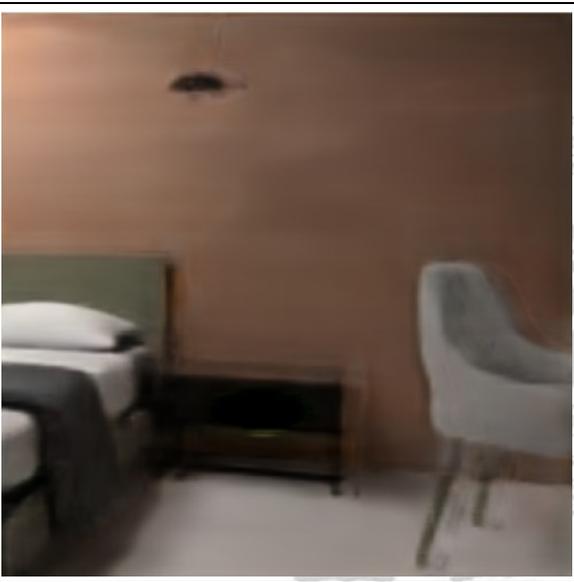
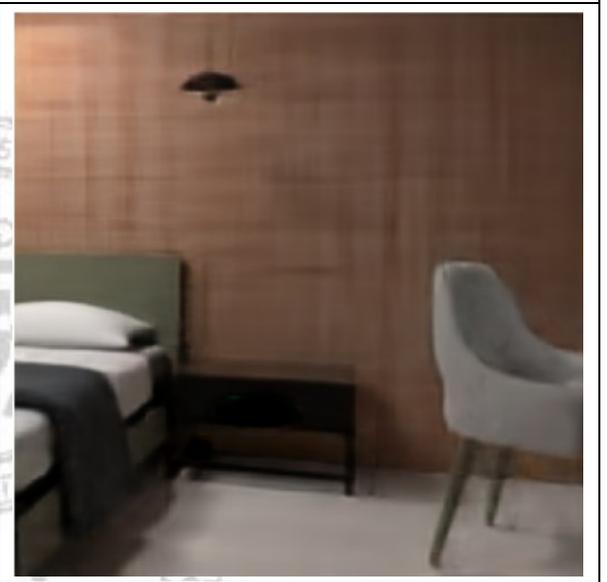
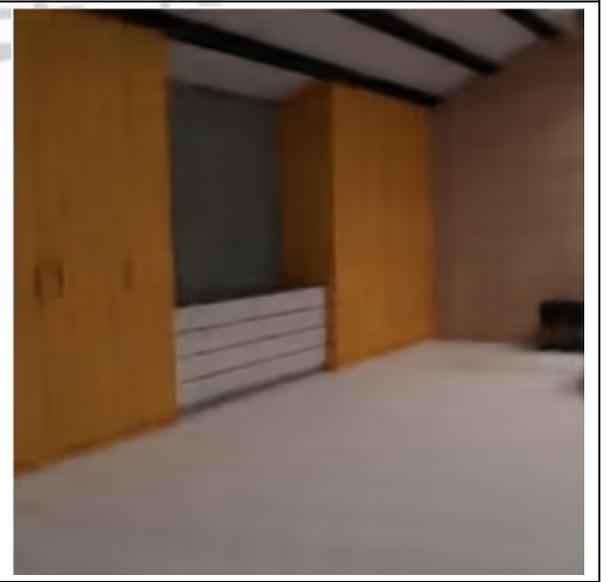
圖 4-1 實驗 1 生成結果細節處

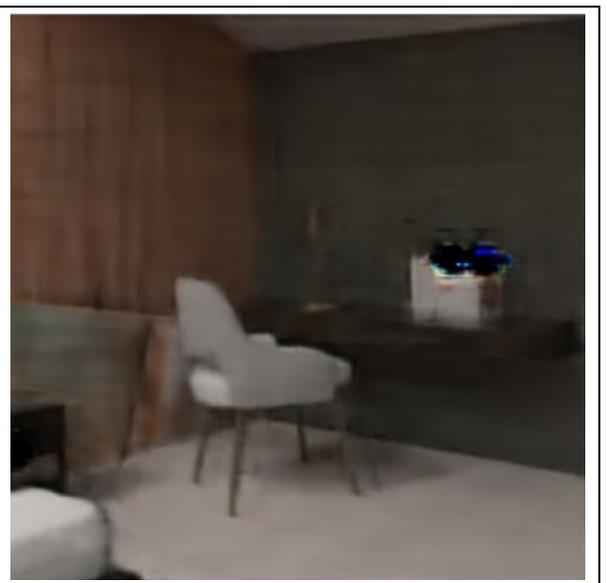
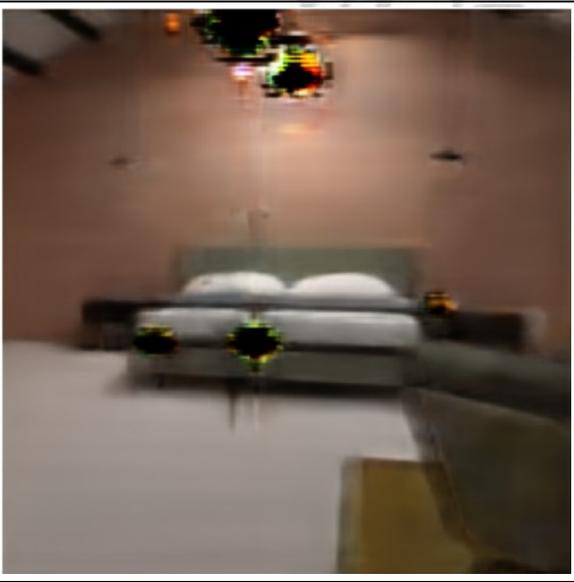
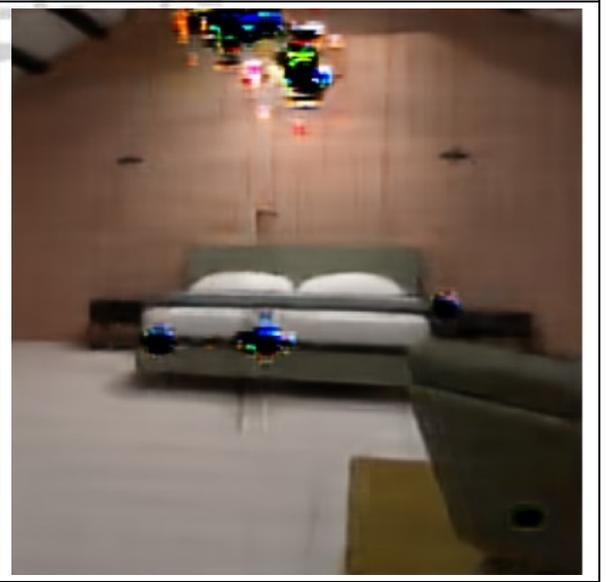


圖 4-2 實驗 1 生成結果產生雜訊處

表 13 epoch300 次與 epoch400 次的比較

	epoch300	epoch400
床		

門		
床頭櫃		
衣櫃		

書 桌		
椅 子		
臥 室		

由本次實驗過後可以觀察出，在第 300epoch 的時候，生成出來的圖片變得更精緻，且許多細節也都能生成出來。像是床、床頭櫃以及書桌的牆壁可以發現紋路已經逐漸成形如圖 4-3。門的邊緣也由模糊變為較為銳化如圖 4-4。衣櫃在第 300 次 epoch 的時候沒有門把，但在 400 次 epoch 之後門把已經可以生成出來如圖 4-5。但在雜訊方面，依然沒有改善，因此我們將訓練次數繼續往上調整，增加至 500 次，繼續觀察實驗結果。



圖 4-3 epoch400 次時牆壁紋路已成形

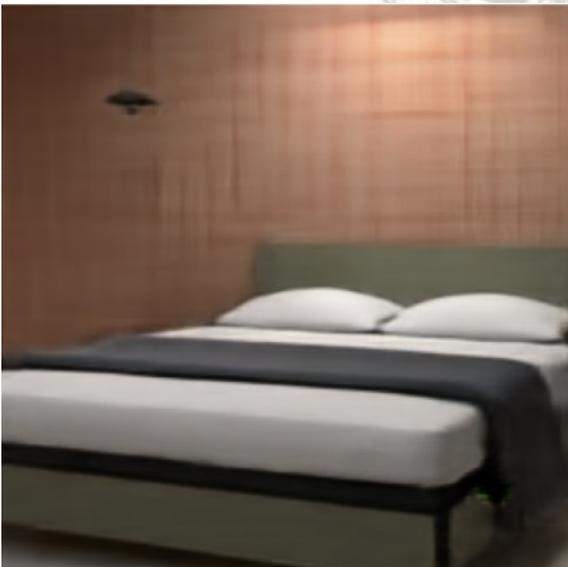
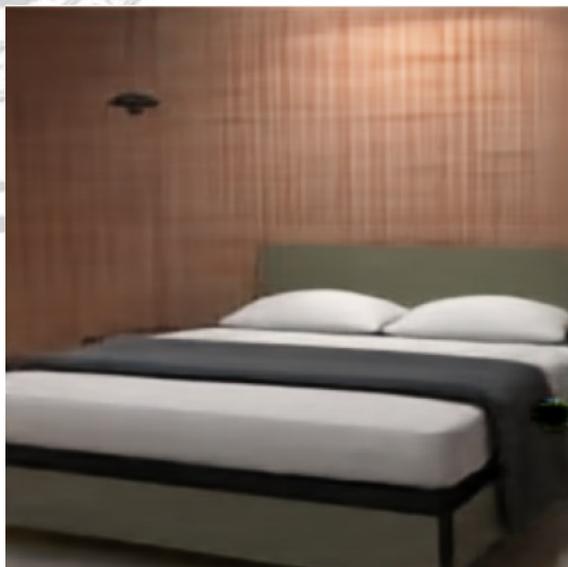


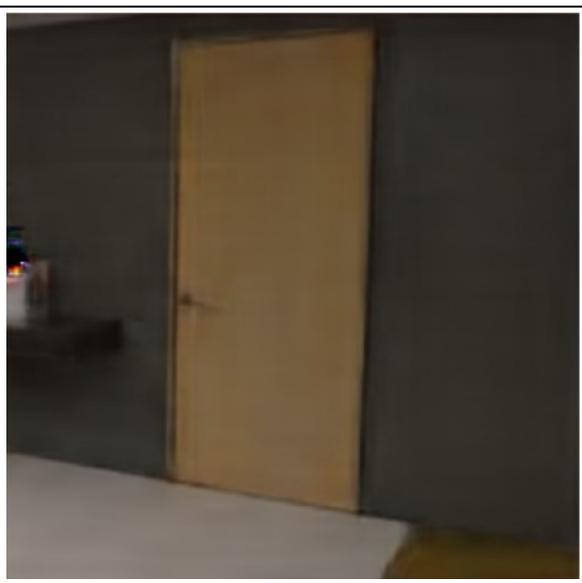
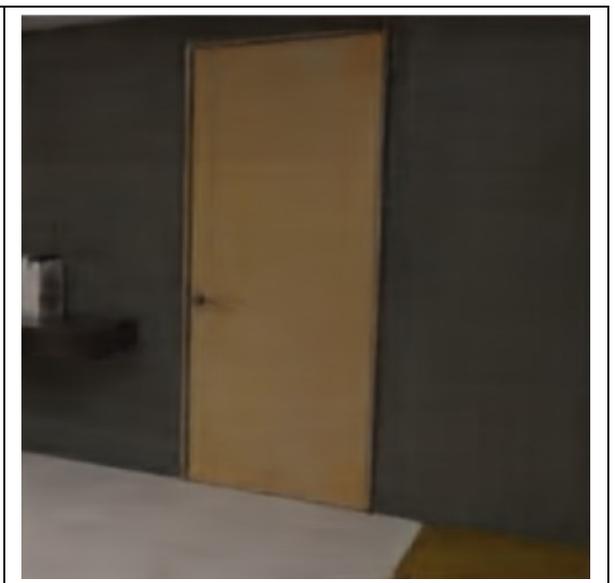
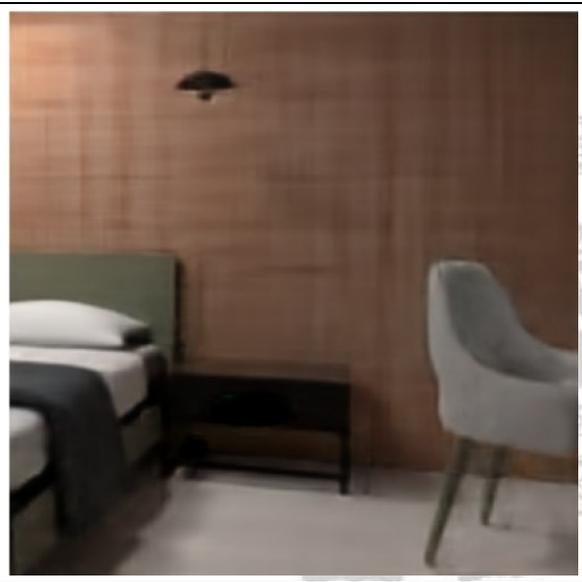
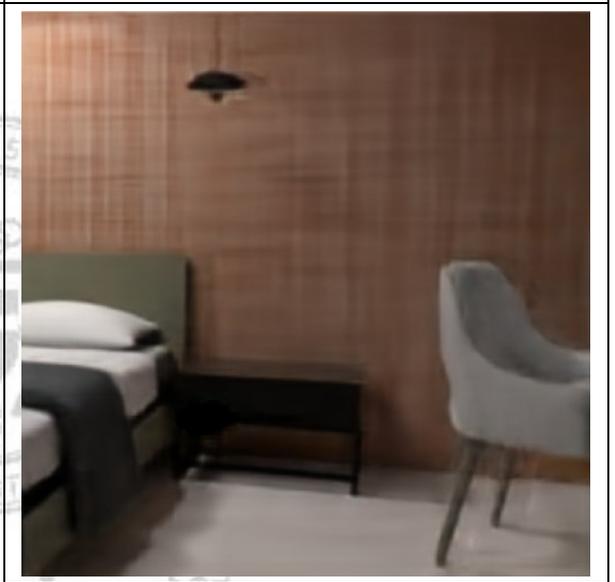
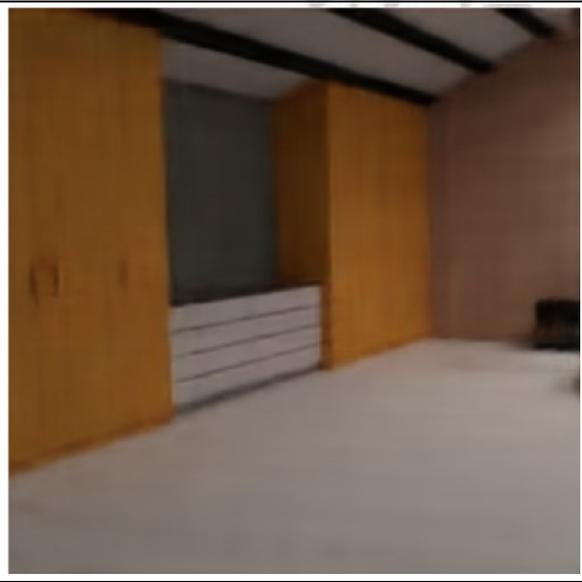
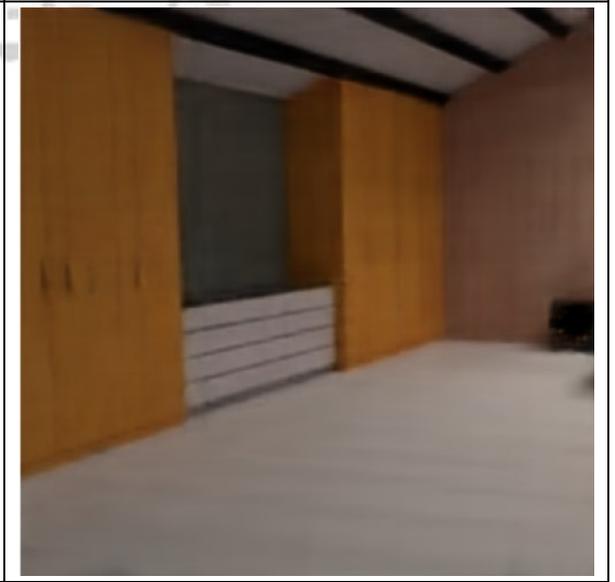
圖 4-4 epoch400 次時門的邊緣較為銳化



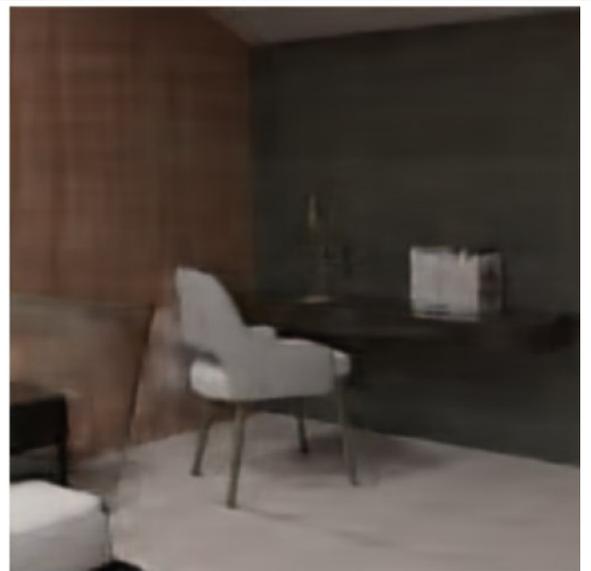
圖 4-5 epoch400 次時能夠生成出門把

表 14 epoch400 次與 epoch500 次的比較

	epoch400	epoch500
床		

門		
床頭櫃		
衣櫃		

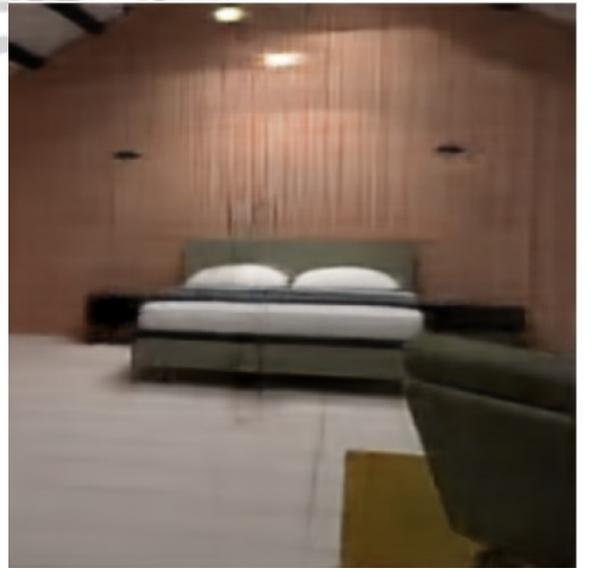
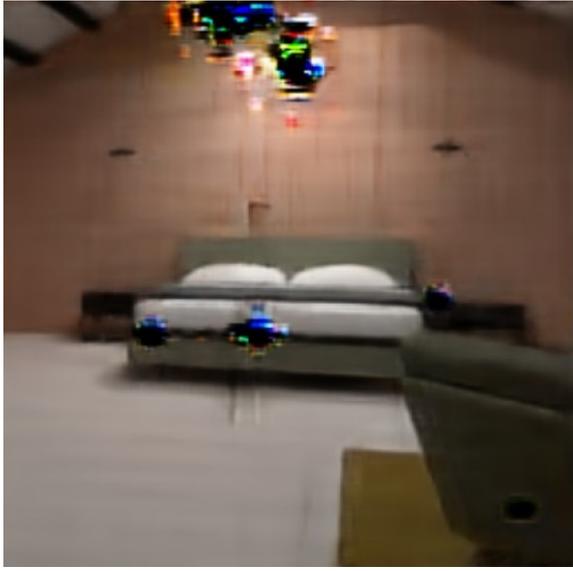
書
桌



椅
子



臥
室



由本次實驗過後可以觀察出，到了第 500 次 epoch 過後，可以發現衣櫃的部分出現了過擬合的現象，衣櫃的門把逐漸形成第三個門把如圖 4-6 紅框處。除此之外，床與床頭櫃的牆壁的紋路特徵相比 400epoch 更為明顯，且門的邊緣也更為精緻，相差最大的是原本書桌以及臥室出現雜訊的部分皆獲得改善如圖 4-7 與圖 4-8，而椅子與 400 次 epoch 的變化差異不大。



圖 4-6 第 500 次 epoch 生成出第三個門把



圖 4-7 書桌 epoch400 次與 epoch500 次對比



圖 4-8 臥室 epoch400 次與 epoch500 次對比

綜合上述 300 次至 500 次 epoch 可以發現到第 500 次 epoch 的時候，相比前幾次 epoch，除了衣櫃的門把出現過擬合之外，其他圖片皆表現得更佳，因此本研究判斷第 500 次 epoch 時為最佳表現。

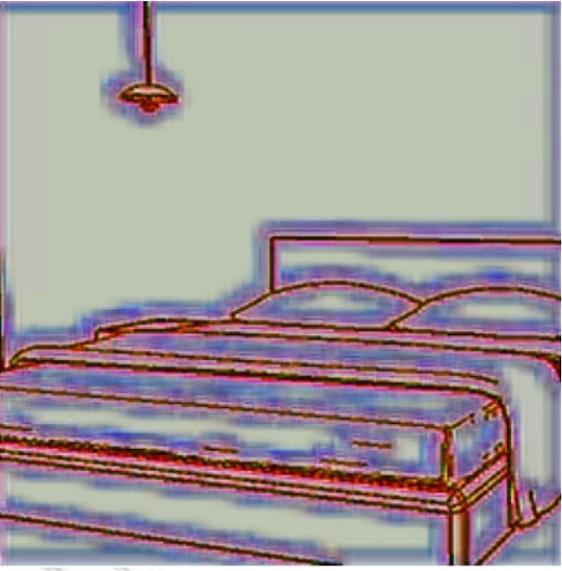
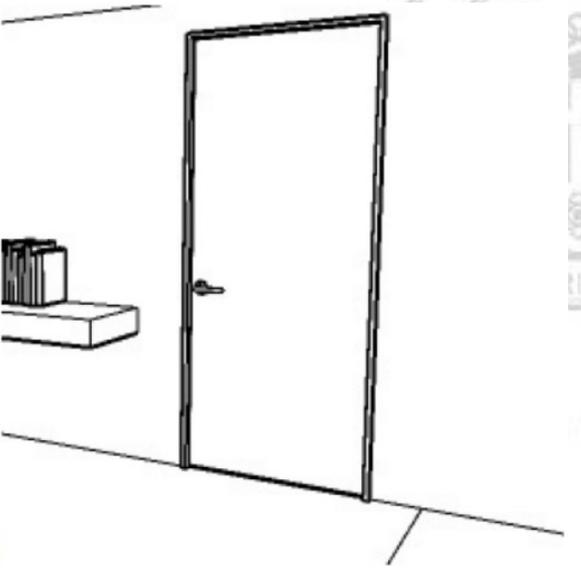
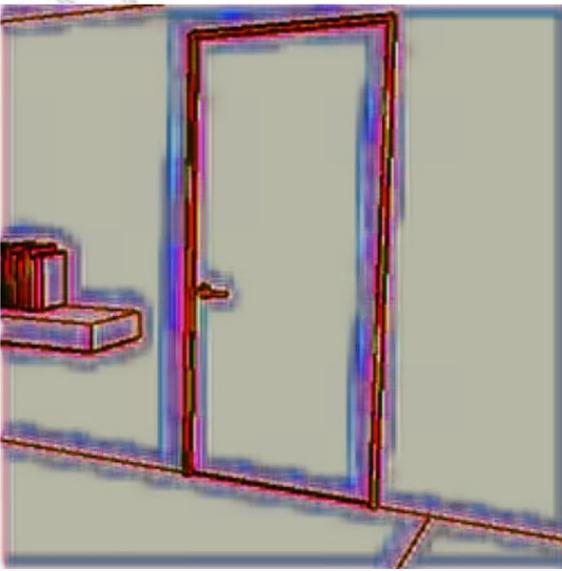
4.3 以 CycleGAN 為基礎的三維模型轉換結果

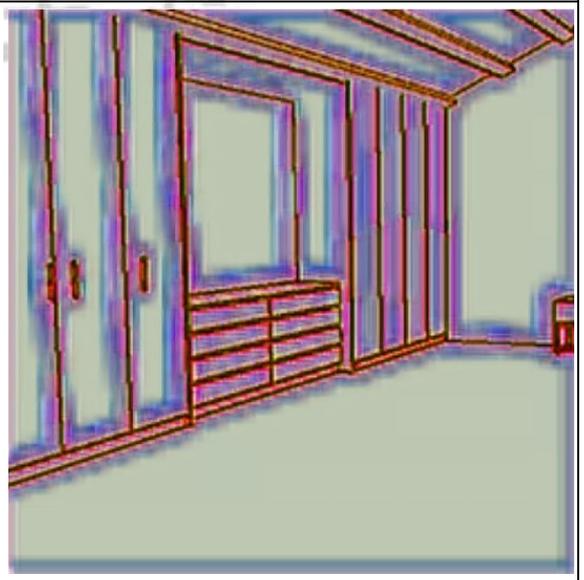
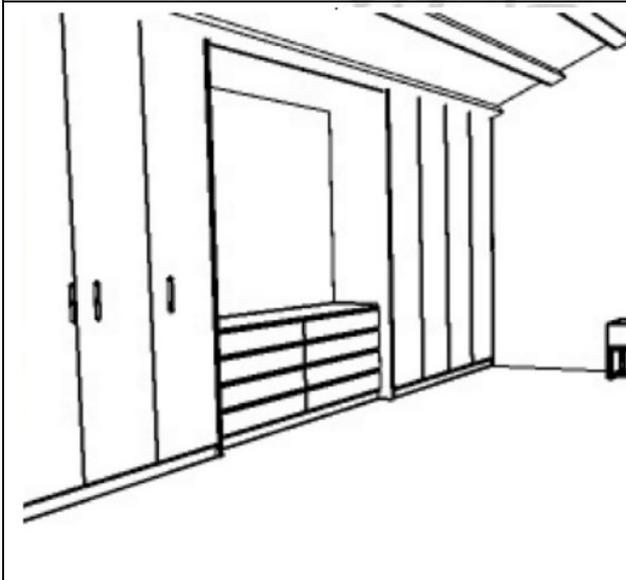
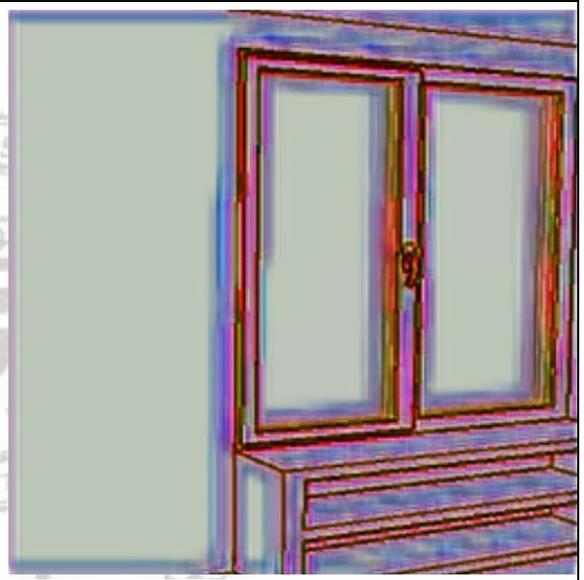
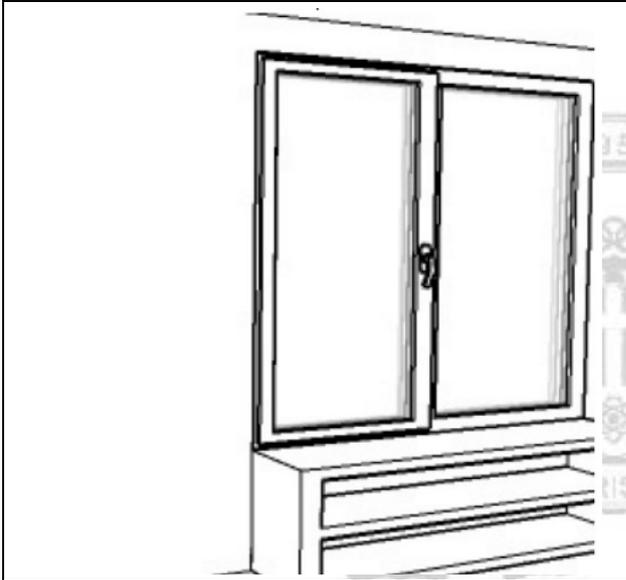
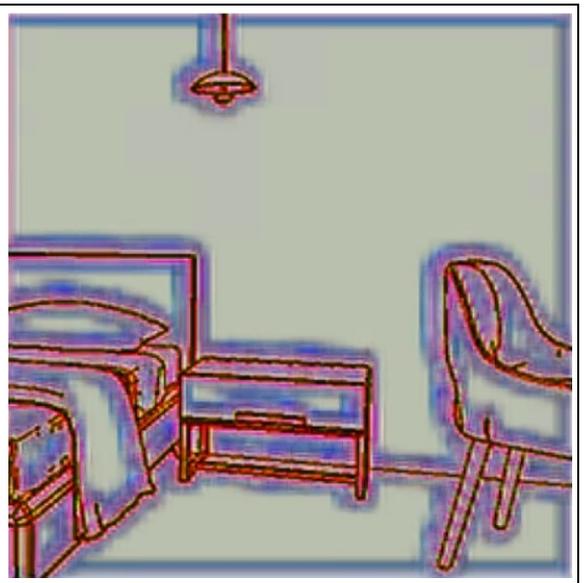
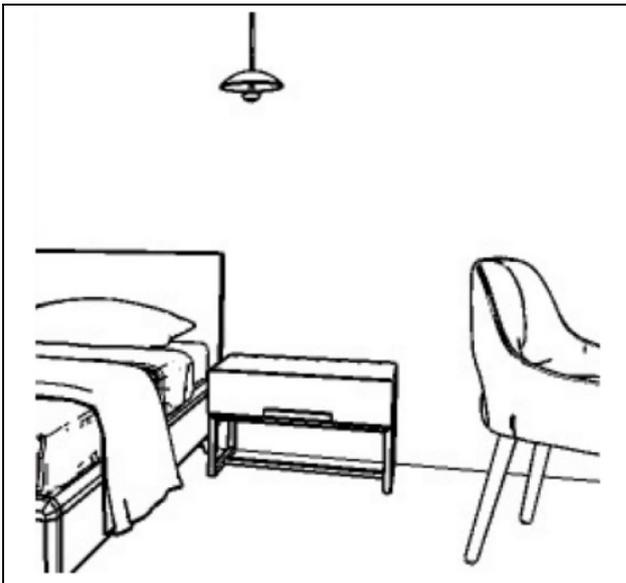
本次實驗使用循環式對抗網路進行訓練，因此不需要成對的資料，只需要蒐集大量的素模的圖片以及擬真圖片進行訓練，本階段實驗的資料有 8000 張蒐集出來的圖片，素模與擬真圖片各占一半，而為了與實驗 1 進行對照，本階段實驗一樣會進行 300 次的 epochs 進行訓練，學習率一樣設定為 0.0002，並將實驗定為實驗 2。

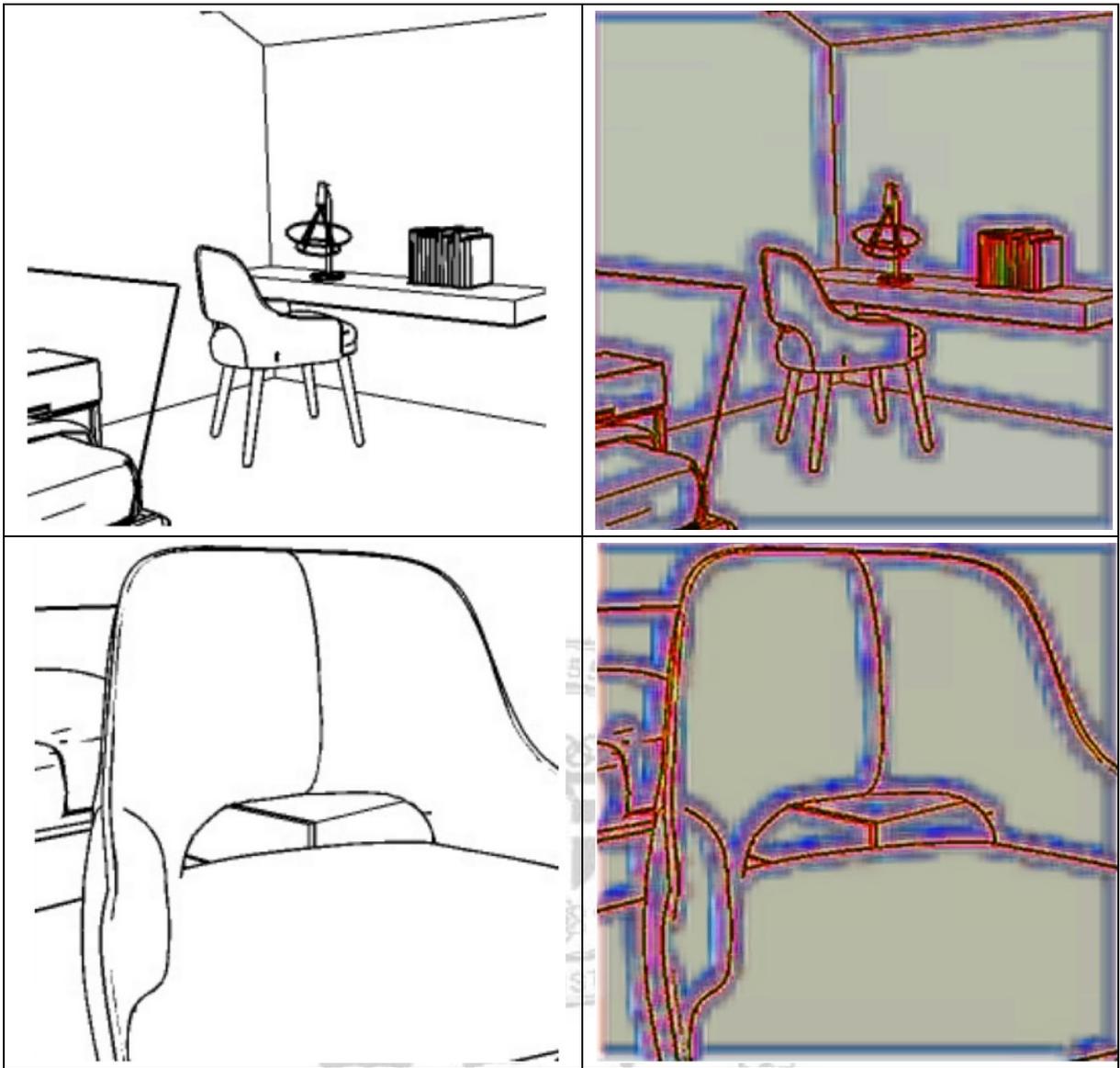
表 15 CycleGAN 實驗數據

	訓練資料數	epoch	Learning rata
實驗 2	8000 張	300	0.0002

表 16 實驗 2 結果

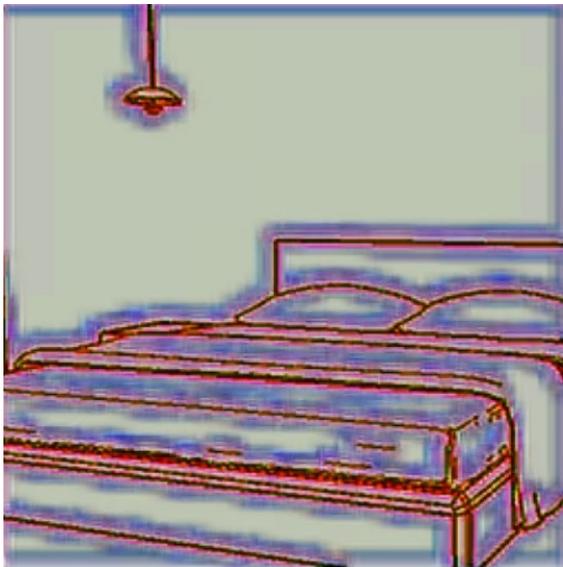
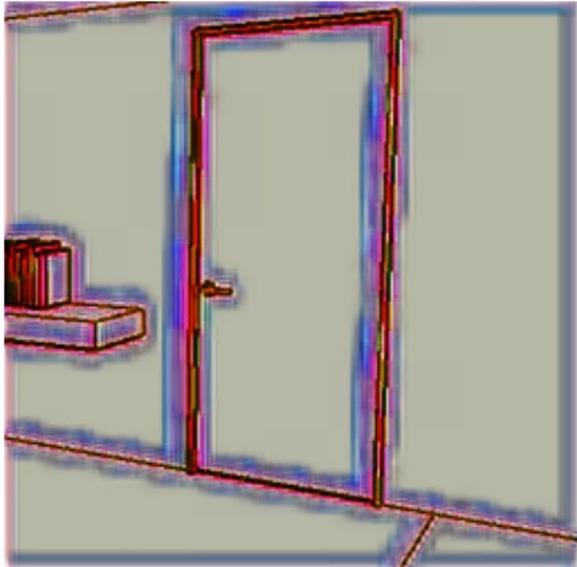
素模	生成圖
 A black and white line drawing of a bedroom. It features a bed with a headboard, two pillows, and a blanket. A simple hanging lamp is positioned above the bed. The drawing is minimalist, focusing on the outlines of the objects.	 A colorized and shaded version of the bed and lamp from the line drawing. The bed is rendered in shades of blue and purple, while the lamp is a reddish-brown color. The background is a light gray, and the overall style is more realistic and textured than the original line drawing.
 A black and white line drawing of a doorway. To the left of the door is a small shelf with three books on it. The door is closed and has a handle. The drawing is minimalist, focusing on the outlines of the objects.	 A colorized and shaded version of the door and shelf from the line drawing. The door is a reddish-brown color, and the shelf and books are in shades of blue and purple. The background is a light gray, and the overall style is more realistic and textured than the original line drawing.



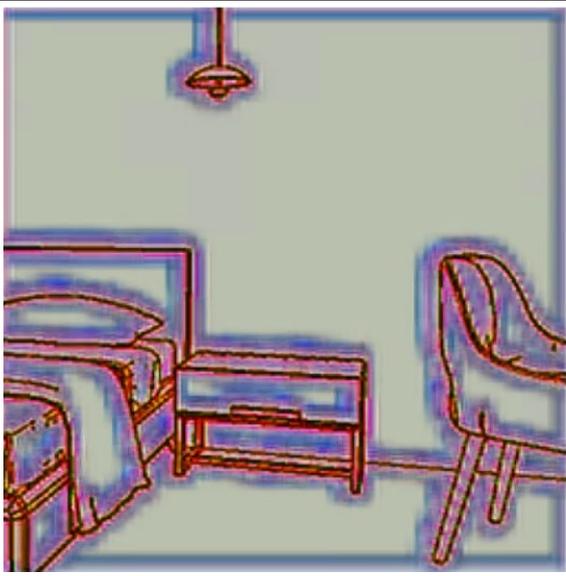


由實驗 2 的結果可以觀察出，CycleGAN 作為素模與擬真圖片的轉換效果並不如預期，不管是材質、顏色、燈光都沒有被生成出來，僅有線條邊緣有稍微有顏色上來。我們初步判斷是因為訓練次數不夠而沒有生成出任何效果出來，因此我們在往上增加 100 次 epoch，並與 epoch300 次進行比較，繼續觀察實驗結果。

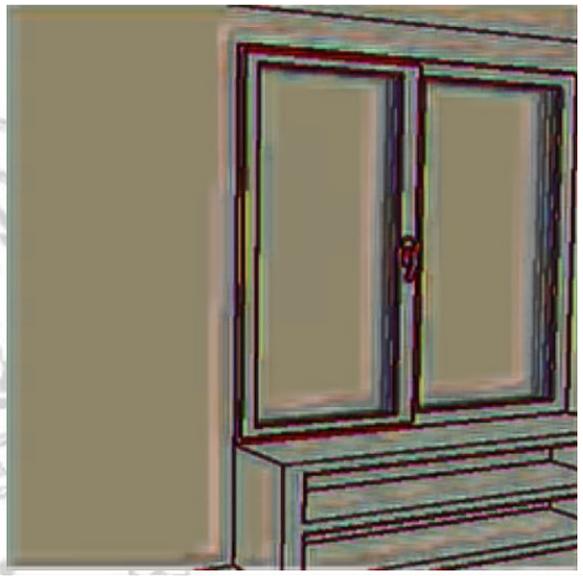
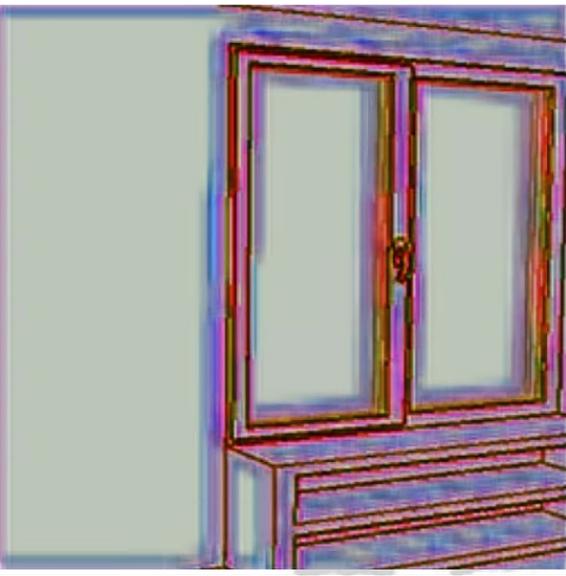
表 17 epoch300 次與 epoch400 次的結果比較

	epoch300	epoch400
床		
門		

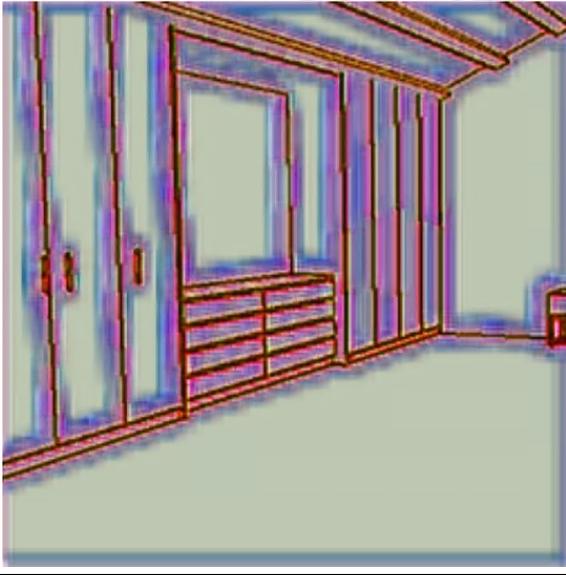
床頭櫃

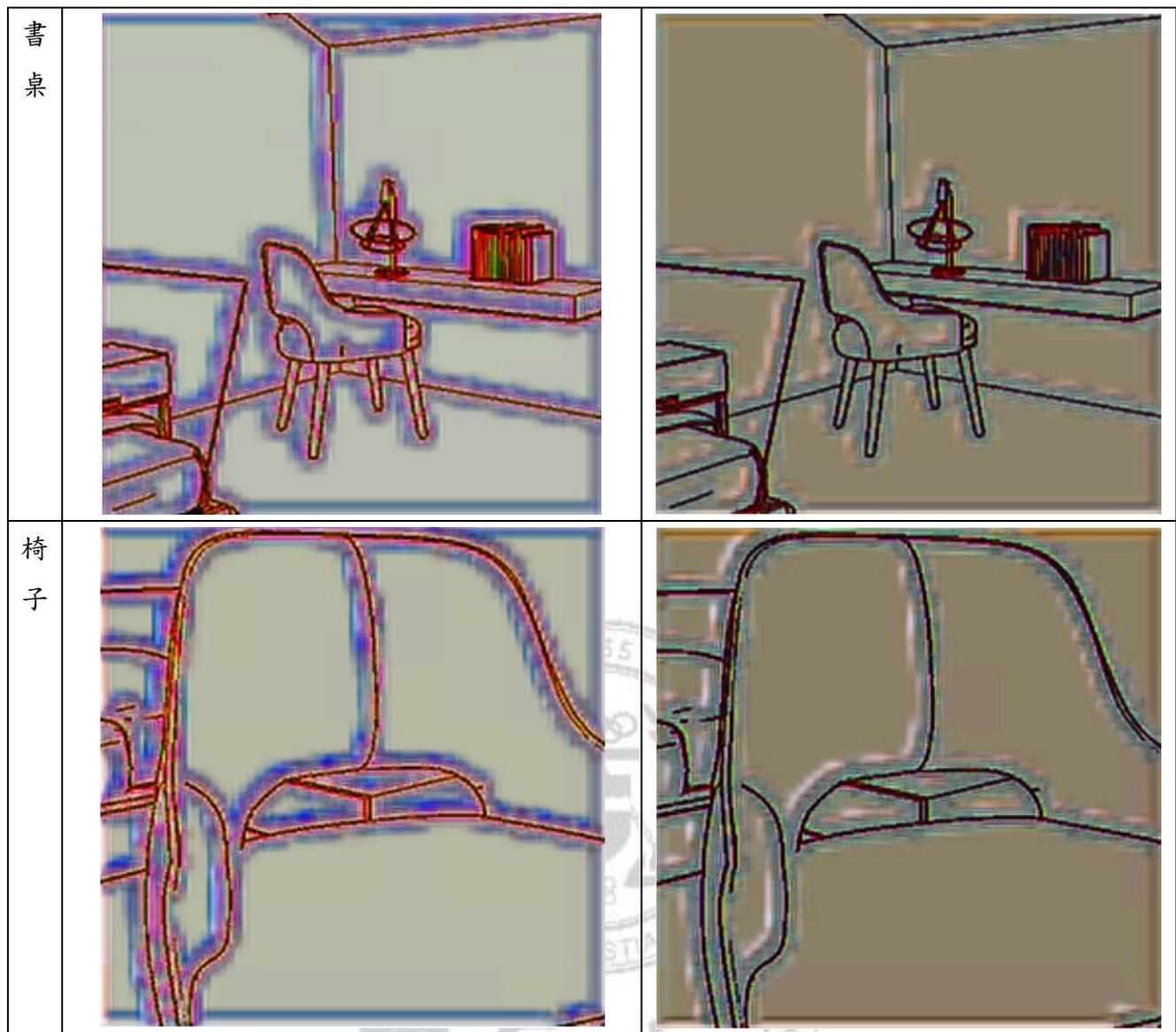


窗戶



衣櫃





在訓練 400 次 epoch 之後可以觀察出，材質與顏色依然沒有被生成出來，僅有線條上有一些改變，由此可知 epoch 的次數並無法改變 CycleGAN 的生成效果。因此可以推斷 CycleGAN 不適合用來做為素模與擬真圖片的轉換。本實驗結束後可以發現 Pix2pix 的效果對於素模與擬真圖片的轉換具有更良好的效果。

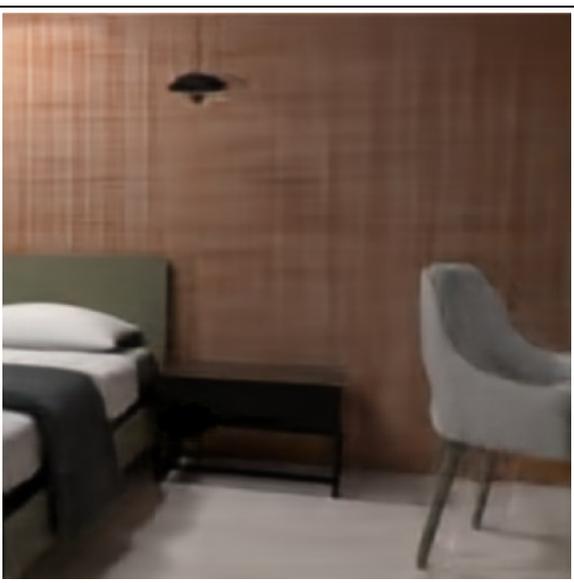
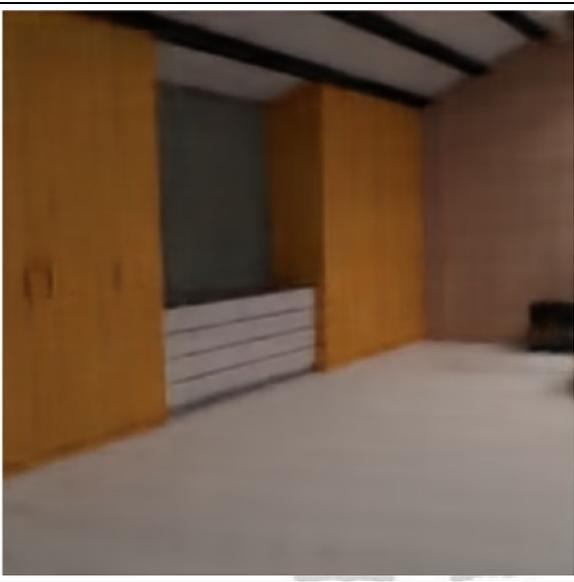
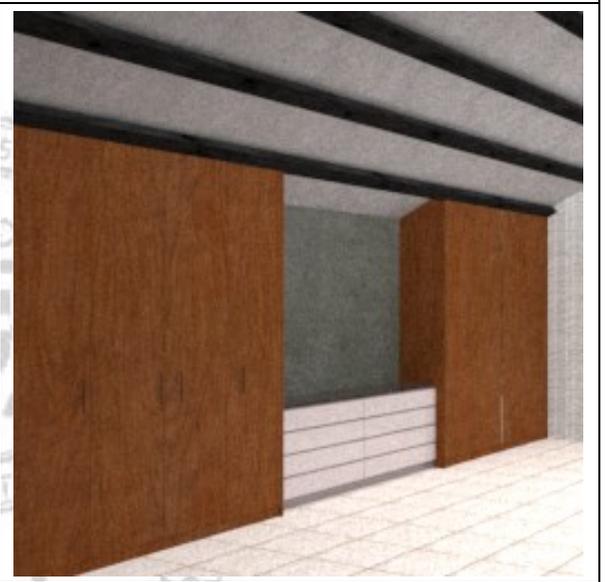
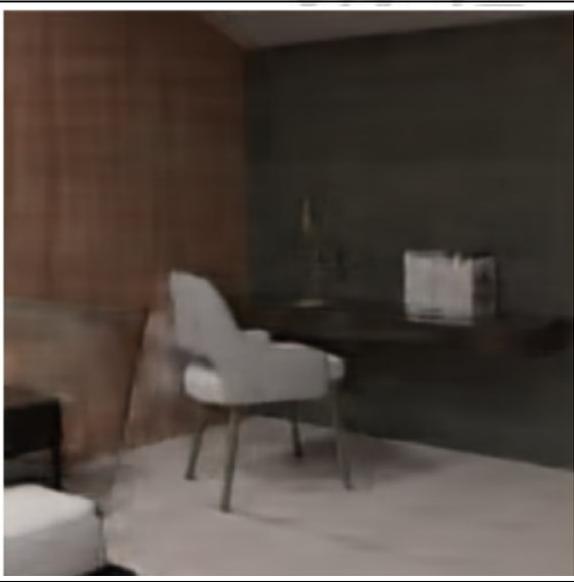
4.4 實驗結果與訓練資料進行比較

由實驗 1 與實驗 2 的結果可以發現，Pix2pix 對於素模與擬真圖片的轉換效果更優於 CycleGAN，因此我們將拿出 Pix2pix 的生成出來的最佳結果與使用 V-Ray 渲染出來

的擬真圖片作為比較。

表 18 Pix2pix 生成結果與 V-Ray 渲染結果比較

	Pix2pix 生成結果	V-Ray 渲染結果
床		
門		

<p>床 頭 櫃</p>		
<p>衣 櫃</p>		
<p>書 桌</p>		



經過對比後可以發現，V-Ray 渲染出來的圖片有許多細節是 Pix2pix 模型無法完整的生成出來，例如衣櫃紋路、地板紋路等等，而 Pix2pix 生成出來的圖片仔細看可以發現有些地方較為模糊，燈光的部分也沒有被很明顯的生成出來，圖片整體的感覺也較暗，但在擬真圖的製作過程當中，Pix2pix 可以跳過貼材質、上色、渲染等步驟就能快速地将素模轉換為擬真圖片，對比製圖的過程，GAN 更為輕鬆方便，更能省下設計師視覺化思考以及創作時間。

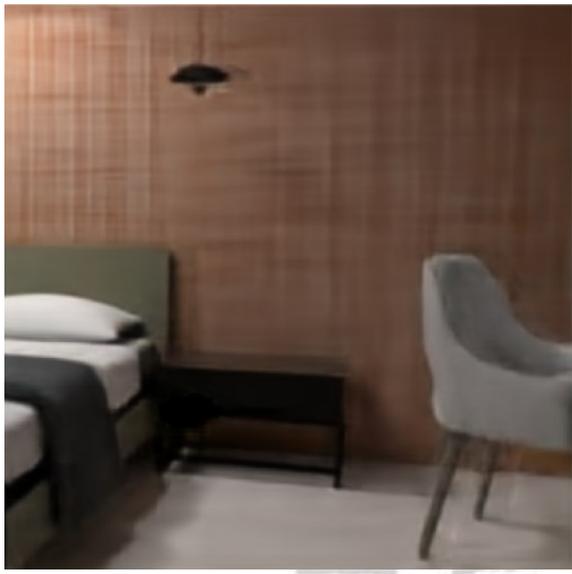
第五章 結論與未來展望

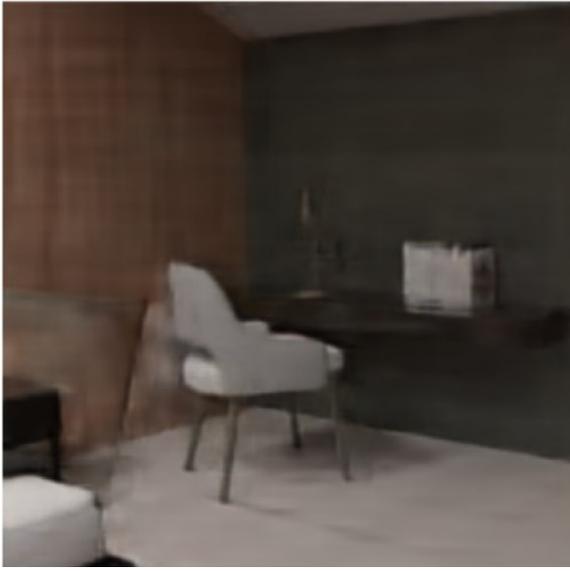
5.1 結論

本研究實驗過後，Pix2pix 分別訓練了 300 次 epoch、400 次 epoch、500 次 epoch，CycleGAN 分別訓練了 300 次 epoch、400 次 epoch，實驗觀察後發現 Pix2pix 較適合運用於室內設計的素模與擬真圖片的轉換，CycleGAN 表現的不如預期。而在本研究中，Pix2pix 也能夠在沒有非常大量資料的情況下將擬真圖片生成出來。因此以下將 Pix2pix 所生成出來的最佳結果進行分析。

表 19 對 Pix2pix 最佳結果進行分析

	Pix2pix 生成結果	結果分析
床		能夠生成出整個床的架構，包含床頭板、枕頭、棉被、都能夠完整的生成出來，燈光方面也有些微的燈光生成出來。牆壁上的紋路也非常明顯，顏色也都自動生成上正常的配色。

門		<p>門的輪廓、門把、牆壁皆有生成出來，輪廓也非常明顯，但燈光在這張圖片上表現的不明顯，使得整體圖片的感覺較暗，地板也生成出兩種不同的顏色，為美中不足的地方。</p>
床頭櫃		<p>生成出來的床頭櫃為黑色的，而旁邊的椅子也有生成出來，牆壁上的紋路也非常明顯。但整理來說燈光較暗，導致圖片也較暗，為美中不足的地方。</p>
衣櫃		<p>衣櫃方面，整體都有生成出來，天花板、地板也都有與衣櫃明顯的區別，而前方的衣櫃門把雖然較為模糊但也有被生成出來。</p>

書 桌		<p>書桌方面雖然圖片有些地方較為模糊，如左下方的位置，但整體的架構還是非常清晰，書桌、椅子、桌上的書都有明顯的生成出來。</p>
椅 子		<p>椅子生成出來的非常相似，連陰影也都呈現的出來，椅子的輪廓、顏色等也都非常完整。</p>
臥 室		<p>臥室整體本身也能夠完整的呈現出來，牆壁、床、天花板、地板、，燈光、前方的椅子都非常相近於擬真圖，整體來說非常成功。</p>

但 CycleGAN 表現的就不盡理想，表現得並不如預期，可能的原因有以下三種。第一，訓練資料不足導致無法有效地進行風格轉換，因此未來希望能夠增加更多的資料集來嘗試。第二，素模與擬真照片並不能算是兩種不同風格的圖片，因此模型在生成的過程中無法有效地分辨出效果是否良好，也導致對抗網路的效果在這邊不起作用。第三，模型無法分辨僅有以線條為主的素模，抓不到圖片的特徵，因此沒辦法將圖片進行風格的轉換。

最後在 Pix2pix 與 V-Ray 渲染結果的擬真圖相比過後，模型確實能自動將素模繪製出一間完整的擬真圖，不管是貼材質、上顏色、或是燈光都能自動的生成上去，在製圖中能夠省略視覺化思考等設計思維，省下了技術需求以及繪製時間。只不過現階段實驗因為時間的考量，雖能實驗模型能夠快速的生成出圖片，但圖片最終表現的並沒有比運用 V-Ray 的渲染效果還要更佳，對於重視細節的室內設計產業，這顯然是不足夠的。因此未來希望能夠運用更大量的資料來進行擬真照片的生成，或是持續的調整參數並進行測試，將模型的生成效果能提升至更佳。

5.2 未來展望

本研究提出了運用生成對抗網路快速的將素模轉換為擬真圖片的方式，藉此來節省設計師構想、上色、貼材質、調整燈光以及渲染的時間。若能在渲染前先重複生成多種不同顏色、材質的圖片，對於室內設計產業來說，設計業者能夠利用此模型先大量生成不同種風格，接著再交給客戶去挑選滿意的或喜好的風格，對於室內設計業的行銷是一種創新，期望吸引更多客戶。

在應用方面，我們希望未來模型空間的種類也能由僅僅的臥室增加其他空間，例如客廳、廚房、廁所，甚至是公司、商場等，不管任何空間都能夠一鍵生成，期望未來室

內設計能夠運用人工智慧節省設計的構想、設計的時間以及金錢，也希望人工智慧能夠逐漸取代室內設計產業，將世界推向完全人工智慧的時代。



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., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (2961-2969).
- He, K., Sun, J., & Tang, X. (2012). Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, 35(6), 1397-1409.
- 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。

連禹睿 (2021)。基於生成對抗網路 GAN 模型之書法字體生成系統。

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

張峻璋 (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。



中原大學