中原大學

以生成對抗網路為基礎將室內設計三維模型轉換之研究

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。

## 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真實的照片如圖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。

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有上色的風格。最後訓練Pix2pix與CycleGAN兩種方法，並比較哪種方法更適合運用在室內設計的領域

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**

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)。

## 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)。

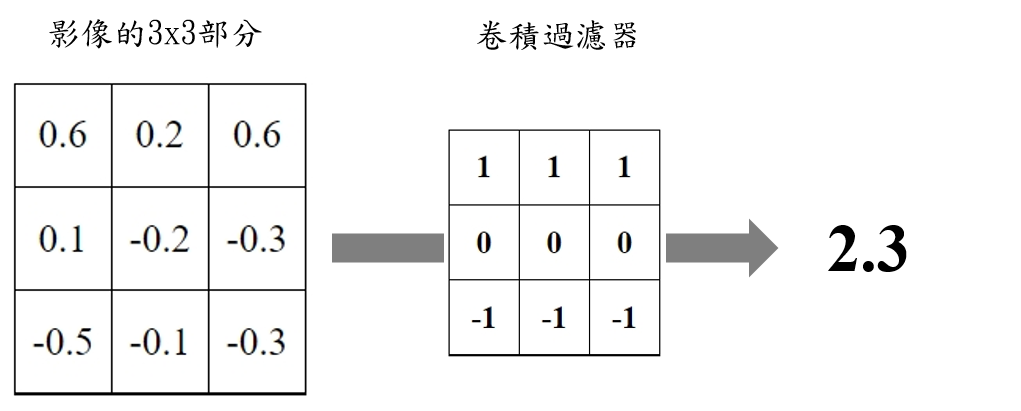


圖2‑2卷積層操作簡易架構

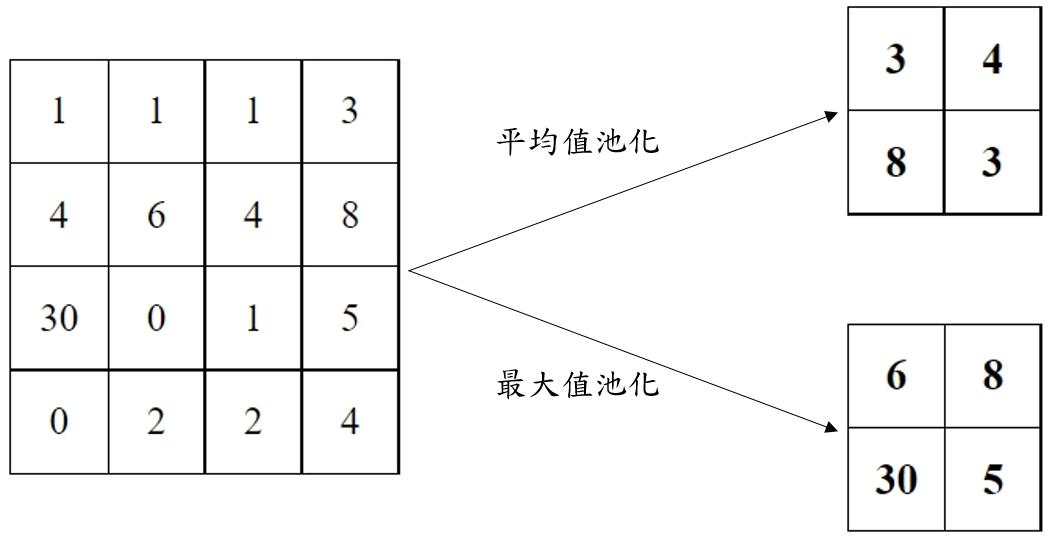


圖2‑3為2x2平均池化層與最大值池化層

## 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2-5。

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2-6為整個生成對抗網路的基本架構。



圖2‑5自動編碼器簡易網路架構示意圖



圖2‑6生成對抗網路的基本架構

## 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2-7為條件式生成對抗網路的基本架構。



圖2‑7 cGAN基本架構

## 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層的影像是一致的，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2-8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2-9。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2-10。



圖2‑8 U-net基本架構



圖2‑9 PatchGAN基本架構



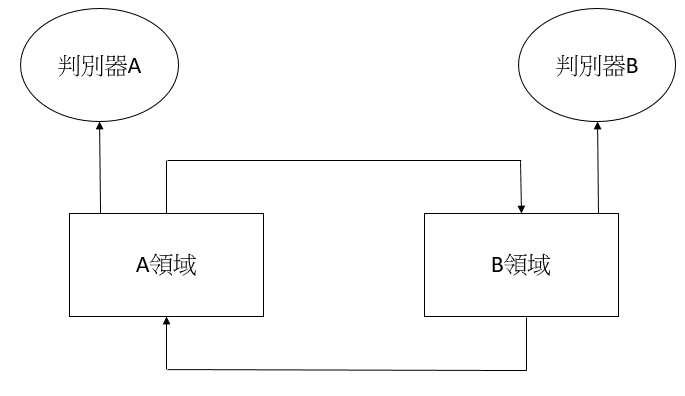
圖2‑10 Pix2Pix實際應用的範例圖

(資料來源：Isola et al., 2017)

2.3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??所示，而CycleGAN也運用在許多地方，如圖片的風格轉換等。



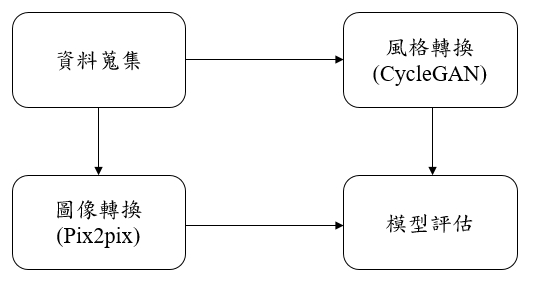
第三章 研究方法

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、生成對抗網路、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

本論文主要分為三個部分，第一部分為室內設計三維模型與真實照片的資料蒐集。第二部分為使用Pix2pix進行圖片轉換將三維模型轉換為擬真照片。第三部分為使用CycleGAN為將三維模型轉換為擬真照片，第四部份進行模型評估。其中流程如圖3-1所示。



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，轉換為真實照片的風格，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，此研究先以google搜尋的方式蒐集圖片，利用”室內設計裝潢”、”室內房間照片”等之類的關鍵字進行圖片的擷取，以及與桃園市某設計公司合作，使用此公司的作品集當作訓練資料，並挑選客廳、臥室等較相似的照片當成主要的資料集，以利於神經網路的學習。接著再利用SketchUp繪製三維素模圖像，按照蒐集到的真實室內設計照片進行繪製，繪製出相互對應的圖片以利於pix2pix的訓練。且室內照片是一個擁有多種物件的圖像，因此也會萃取出室內設計裝潢的前景物件如傢俱獨立進行繪製，並連帶完整房間照片一同丟進模型進行訓練，以加強物件輪廓。

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，此研究採用左右翻轉、逆時針旋轉以及順時針旋轉的方式來增強資料集，以達到所需的訓練量。

## 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3D模型對真實照片的轉換，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 生成器架構的編碼器由4x4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。

解碼器則由4x4轉置卷積層所組成，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。

判別器我們採用PatchGAN架構，一系列的4x4卷積層，步長為2，並在每個卷積層後接一個批次標準化與斜率為0.2的激活函式Leaky-ReLU，在最後一層使用softmax激活函式，判斷70x70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。

## 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 )) ≈ x* 和 *G(F(y)) ≈ y*，目標函式如圖3-?。

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，本次實驗的學習率設定為0.0002，batch size設定為1。圖片輸入的大小設定為256x256，最後並比較不同的epoch產生的結果。

## 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，並這兩種模型的生成結果進行比較，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。