

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 1 月

目錄

目錄	II
圖目錄	III
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)	7
2.3 語義分割(Instance Segmentation)	9
2.4 生成對抗網路(GANs)	10
2.5 條件式生成對抗網路(conditional GAN, cGAN)	12
2.6 Pix2pix	13
2.7 引導影像濾波器(Guided Image Filter)	16
第三章 研究方法	18
3.1 研究架構	18
3.2 資料蒐集	19
3.3 以 FastFCN 進行語義分割	23
3.4 Pix2Pix 為基礎加入影像濾波層進行三維模型的轉換	23
3.5 模型評估	24
參考文獻	25

圖目錄

圖 1-1 三維立面圖轉換為真實照片	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 FCN 分割影像概念圖	10
圖 2-5 自動編碼器簡易網路架構示意圖	11
圖 2-6 生成對抗網路的基本架構	11
圖 2-7 cGAN 基本架構	13
圖 2-8 U-net 基本架構	15
圖 2-9 PatchGAN 基本架構	15
圖 2-10 Pix2Pix 實際應用的範例圖	15
圖 2-11 影像濾波器抓取輪廓示意圖	17
圖 3-1 系統流程圖	19
圖 3-2 三維模型與真實照片對比圖	22
圖 3-3 原圖與遮罩圖	23

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，像是網路上非常火紅的線上建模工具 3ds-Max 以及 SketchUp 等等，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，但渲染時需添加不同的參數，若想渲染出非常完美的 3D 擬真圖需花費大量時間以及設計師的視覺化思考與技術需求，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真實照片，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，也讓客戶在裝潢前能夠看到的最真實照片由 3D 擬真圖突破為真實照片的樣貌。

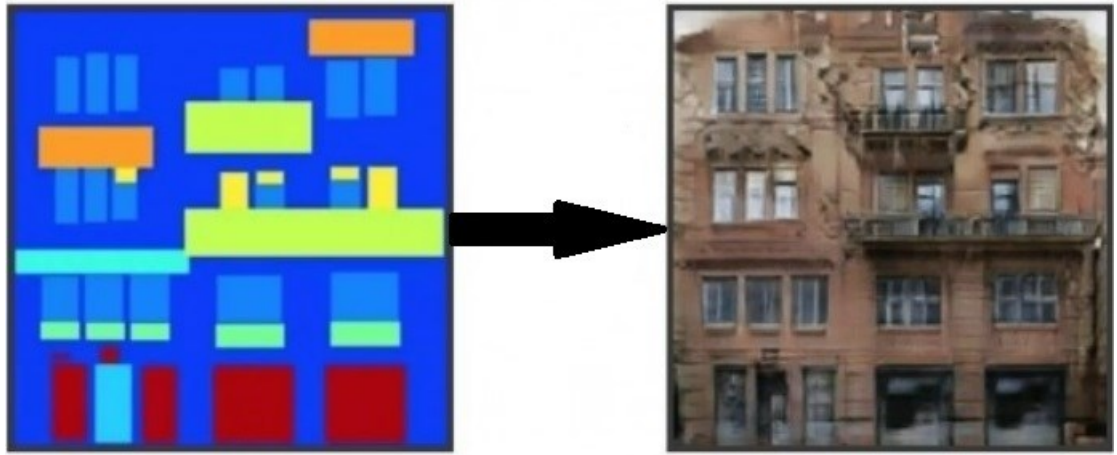


圖 1-1 三維立面圖轉換為真實照片

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真實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，設計照片絕對會強調照片能夠越清晰越好，且避免防止輸出的圖片邊緣是模糊的導致破圖，因此本研究會在訓練時加入影像濾波器。影像濾波器對於圖片的運用已有相關的研究，如 Qirong et al.(2020)提出的去霧網路之架構，以生成對抗網路加上引導影像濾波器對於具有濃霧的圖片進行去霧處理達到不錯的效果。因此此研究也希望能對生成對抗網路生成出來的照片進行影像濾波的處理，希望能強化電腦所生成出來的圖像輪廓，保持圖像物體的結構，生產出高清晰的圖像。

另外因室內設計的房間照片中會存在多個傢俱，屬於多物件類別的照片，因此此研究將嘗試在訓練前對圖片進行語義分割，將圖片中的前景如傢俱獨立分割出來，產生出

新的圖片，目的是希望在訓練的過程連同前景物件也單獨抓出來訓練，將室內設計房間照片能夠保留所有前景物件，在訓練的過程較能夠還原所有前景，最後觀察實驗後的成果。

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真實照片，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，通常在建置完 3D 模型後會再經過一道渲染程序，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)。

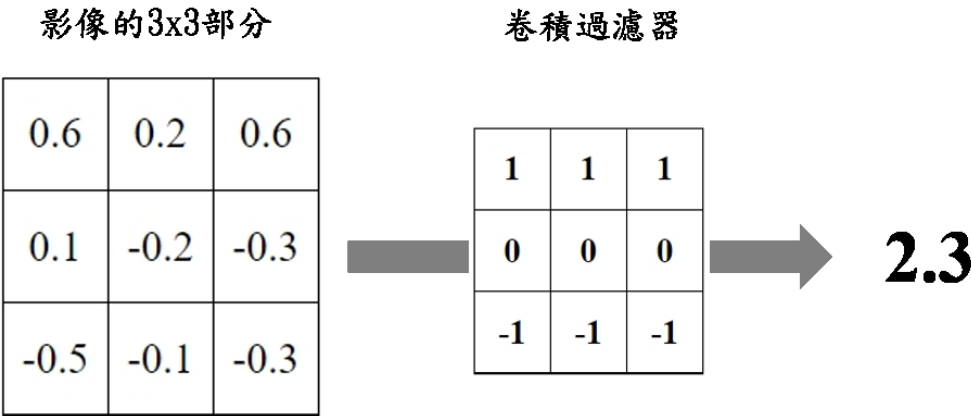


圖 2-2 卷積層操作簡易架構

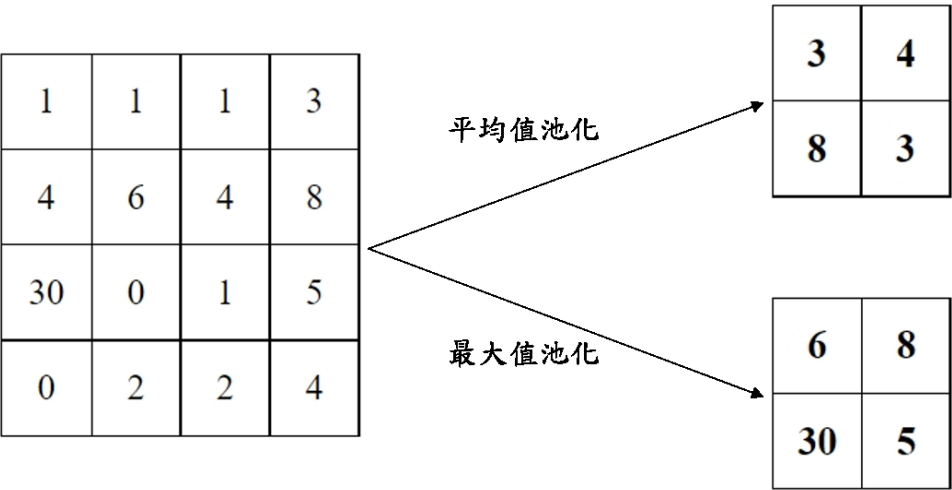


圖 2-3 為 2x2 平均池化層與最大值池化層

2.3 語義分割(Instance Segmentation)

物件偵測是影像辨識中重要的一環，它的目的是在影像中先尋找感興趣的區域，再對這些區域進行分類，以及附帶它是此物件的猜測機率。此技術常用於自動駕駛的應用，因為汽車不僅需要知道周圍物體的類別，還需要檢測它們的位置以避開(Zhao et al., 2019)。

語義分割是基於物件偵測的一個方法，施旻岳(2021)提到語義分割是針對圖片中的每格像素的種類去進行區域的分類，把相同的物件劃分在一起，此技術已應用於非常多產業，如生物醫學、自動駕駛等等。傳統的 CNN 較適合運用於影像分類，因為 CNN 在進行卷積與池化的過程會丟失細節，無法精確地分辨出物體的輪廓，較適合對整張影像作辨識，並標示出圖像類別。針對這種問題，語義分割的架構陸續被提出來，而語義分割主要的思想為 Fully Convolutional Networks(FCN)-全卷積的概念，FCN 由 J Long et al.(2015)提出，主要思想是將全連接層替換為卷積層。由於進行語義分割輸入的影像尺寸與輸出的影像尺寸必須為一樣，但卷積層的池化會使圖像尺寸縮小，於是語義分割使用反卷積的操作，對縮小的圖片進行上採樣(upsampling)，直到與原來的輸入為相同的大小，簡單來說就是把池化後縮小的尺寸再放大回去，以滿足語義分割的需求如圖 2-4。隨著時間的發展，語義分割的架構也陸陸續續地被提出，像是 DeconvNet、U-Net、FastFCN 等。

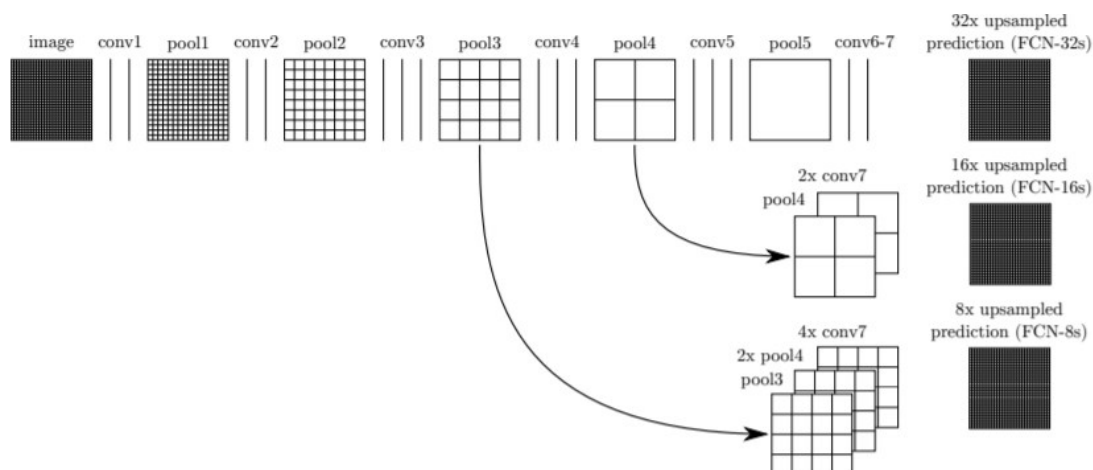


圖 2-4 FCN 分割影像概念圖

(資料來源：Jonathan et al., 2015)

2.4 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。

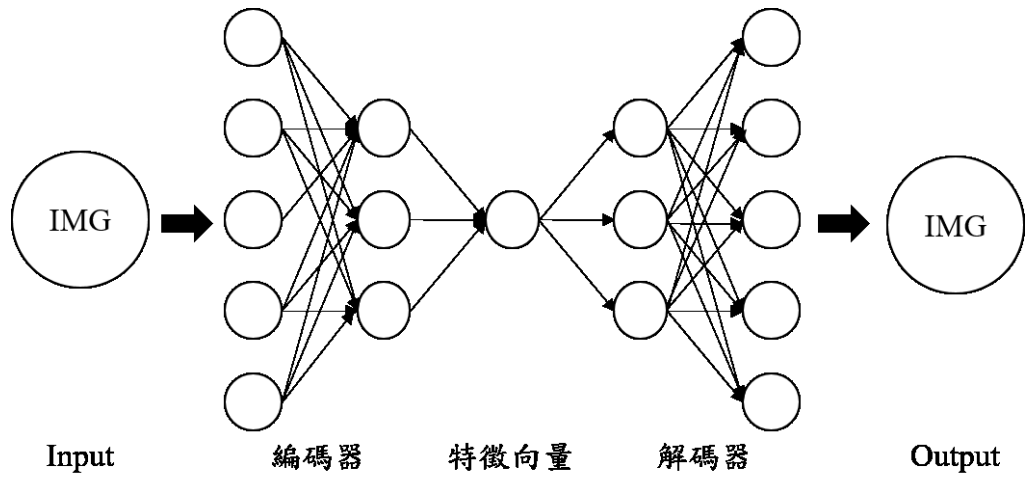


圖 2-5 自動編碼器簡易網路架構示意圖

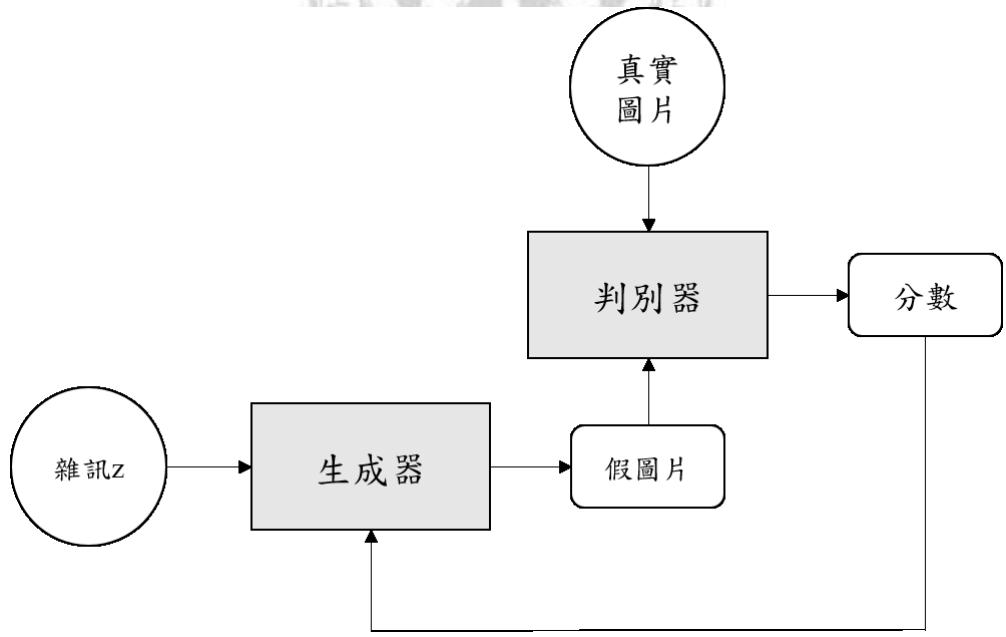


圖 2-6 生成對抗網路的基本架構

2.5 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。

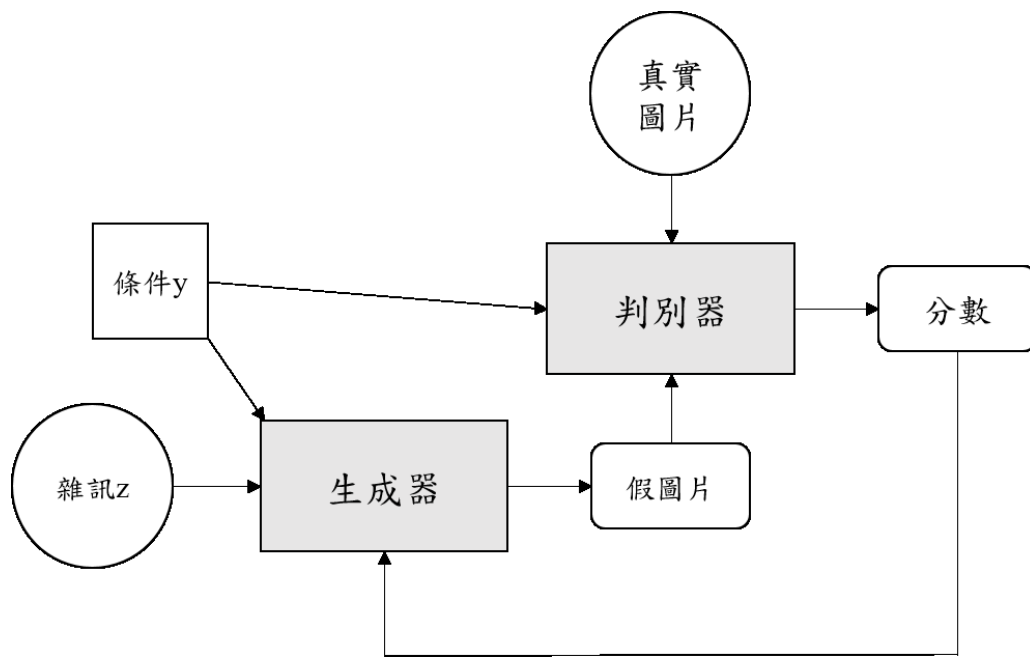


圖 2-7 cGAN 基本架構

2.6 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 2-8 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。



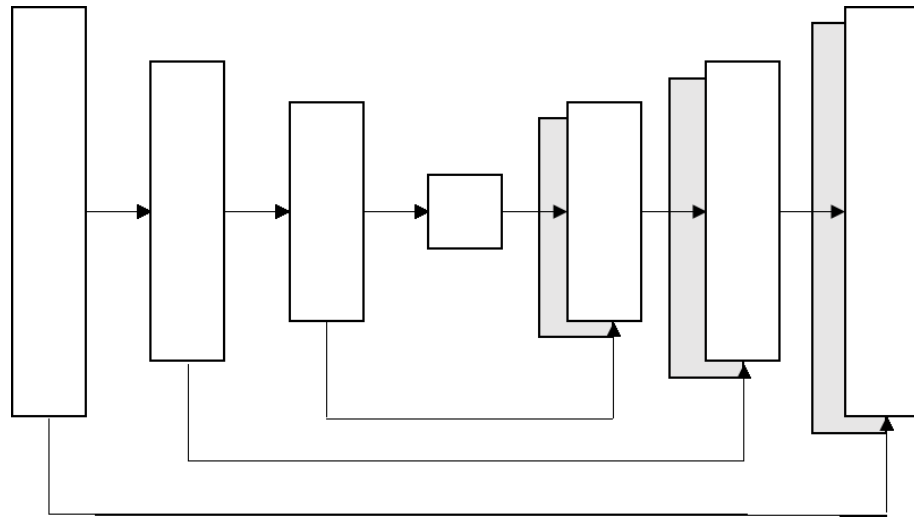


圖 2-8 U-net 基本架構

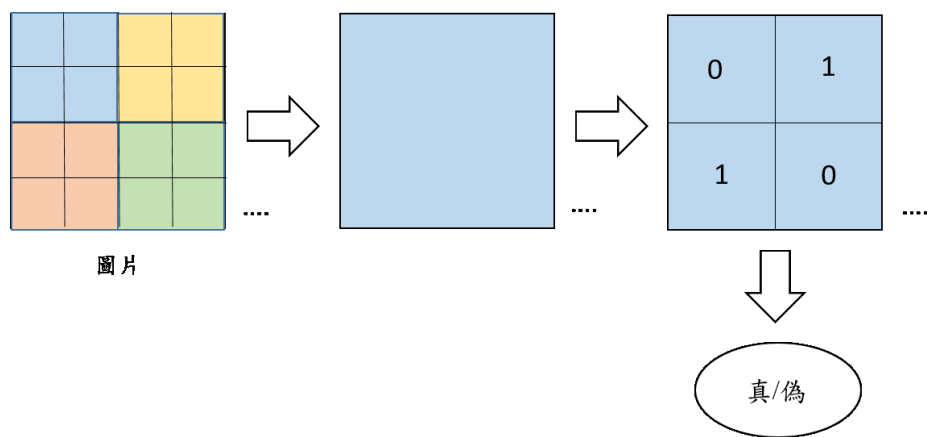


圖 2-9 PatchGAN 基本架構

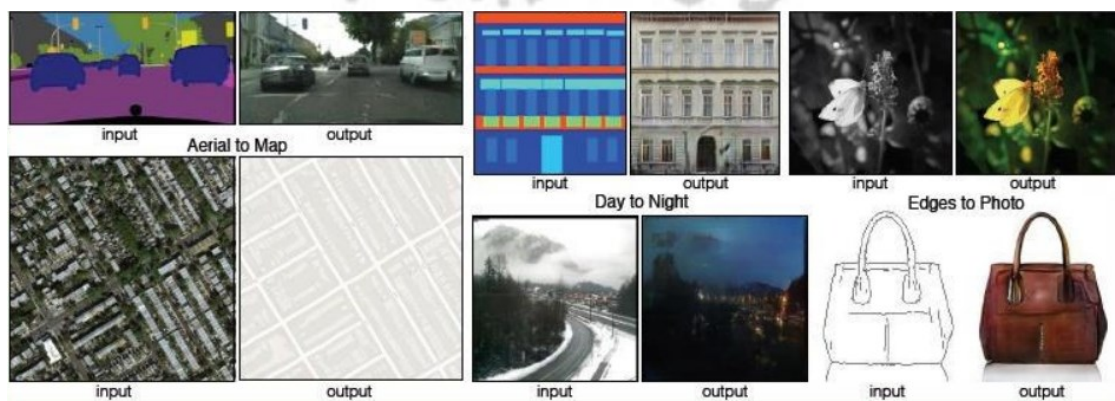


圖 2-10 Pix2Pix 實際應用的範例圖

(資料來源：Isola et al., 2017)

2.7 引導影像濾波器(Guided Image Filter)

引導影像濾波器(Guided Image Filter)為 Kaiming et al.(2012)所提出。引導影像濾波器為一種能將影像保持平滑或是銳化的濾波器。能夠將影像清除雜訊，使得影像平滑化稱作低通濾波器;將影像銳化強化圖像物體輪廓的稱為高通濾波器。引導影像濾波器即為一個需要引導圖的濾波器，引導濾波器的運作方式為通過一張引導圖對初始影像 p (輸入影像)進行濾波的處理，使得最終輸出的影像大致上與初始影像 p 相似，但紋理部分與引導圖相似。

近年來生成對抗網路加上引導影像濾波器也有不錯的應用，如 Qirong et al.(2020)提出的具有引導影像濾波器的 pix2pix 去霧網路，此網路在生成器的編碼器與解碼器之中加入一層引導影像濾波層，在解碼之前加入影像濾波的處理，強化特徵的輪廓，以避免在生成的過程輪廓模糊或消失。而引導影像濾波層在訓練前要設置兩個參數，分別為平滑內核半徑 r 與正則化係數 e ，平滑內核越大越能夠捕捉更多細節，而正則化係數則是防止過擬合，抓到不相干的輪廓。而此網路透過設置不同的平滑內核與正則化係數進行訓練，平滑內核分別為 2、4、6，正則化係數分別為 0.001、0.0001。實驗結果顯示具有引導影像濾波層的 pix2pix 去霧網路與其他去霧網路相比，具有更良好的效果。

由去霧網路可知，引導影像濾波器對模糊圖片的處理效果非常好，因此此研究將對室內設計照片進行影像濾波的處理。因為室內設計圖絕對強調清晰的照片，為了不讓生成出來的圖片太過模糊，因此利用影像濾波器將影像銳化，抓出影像的邊界，通過濾波器還原回輸出圖像，以最大限度地保持模糊圖像的邊緣和細節，達到細節加強的效果，圖 2-11 為影像濾波器抓取輪廓示意圖。



圖 2-11 影像濾波器抓取輪廓示意圖

(資料來源：<https://reurl.cc/3j5xn9>)

中原大學

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、語義分割、生成對抗網路、影像濾波層、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，語義分割為增強資料及強化細節的運用，生成對抗網路與影像濾波層為三維模型轉換為真實照片的基本網路運用，模型評估為與未訓練語義分割模型與未經引導影像濾波器的訓練模型進行比較。

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。第二部分為對真實照片做語義分割，加強資料集。第三部分為使用 Pix2pix 為基礎加上影像濾波層將三維模型轉換為真實照片，第四部份進行模型評估。其中流程如圖 3-1 所示。

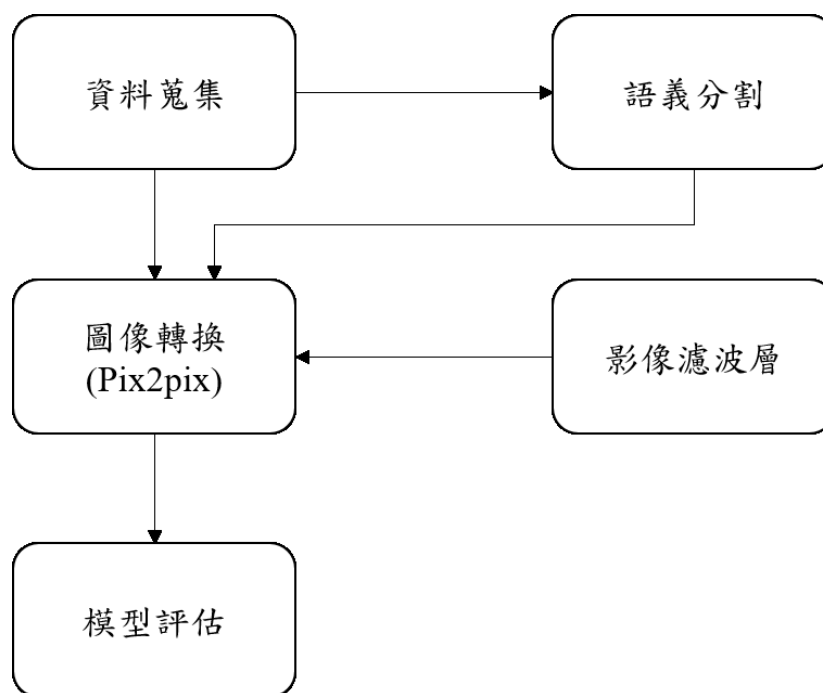


圖 3-1 系統流程圖

3.2 資料蒐集

本研究在進行三維模型與真實照片的轉換，需要使用到數量龐大的圖片當資料集，且找出有互相對應的室內照片，同樣的場景各有一張三維模型與真實照片當成訓練資料。在資料的蒐集上，此研究以 google 搜尋的方式搜尋圖片，利用”三維模型與真實照片對比”、”3D 模型與真實照片”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。因為室內房間照片屬於多物件的照片，因此本研究將採用語義分割的方式將房間的前景物件萃取出來並連帶完整房間圖片一同丟進模型進行訓練，於是此研究採用了 FastFCN 來對圖片進行語義分割，除了萃取出房間個別物件加強物件輪廓外，也由此方式增加訓練的資料集，解決資料集不足的問題。

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。





(a)左圖為三維模型，右圖



(b)上圖為三為模型，下圖為真實照片



(c)上圖為三維模型，下圖為真實照片



(d)上圖為三維模型，下圖為真實照片

圖 3-2 三維模型與真實照片對比圖

3.3 以 FastFCN 進行語義分割

為了更精確的訓練到影像中每個前景的細節，如傢俱，在訓練前先將影像進行語義分割，將影像中所有前景萃取出來一同丟入模型進行訓練。本研究我們採用 FastFCN 來做語義分割，抓出前景的輪廓。本實驗使用了 GitHub 上所提供的 FastFCN 的預訓練模型，且該模型使用的資料集為 ADE20K，ADE20K 的資料集擁有超過 27000 張圖片，其中 25000 張是訓練用的資料，2000 張是驗證用的，其中包含建築的部分有 10.7%。訓練完之後再利用 OpenCV 把目標圖中的背景去掉，產出保留前景的遮罩圖檔如圖 3-3，最後再丟入模型與原照片一同訓練。

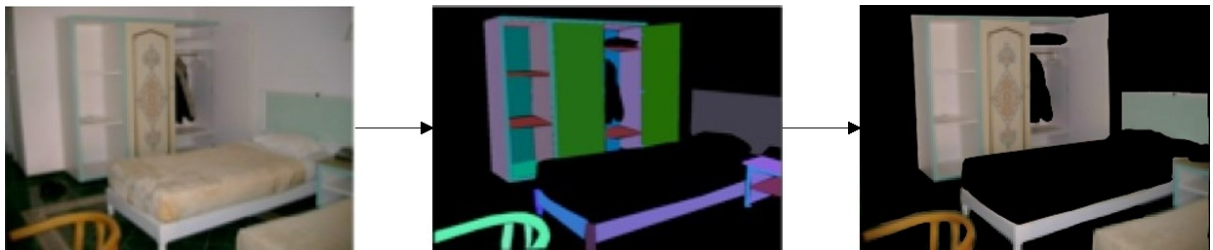


圖 3-3 原圖與遮罩圖

3.4 Pix2Pix 為基礎加入影像濾波層進行三維模型的轉換

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。

且為了在訓練的過程中，損失影像中的傢俱等物件的邊緣，或是造成影像模糊，我們在生成器的中間加入一層影像濾波層對影像進行強化，在模型訓練之餘，另外建立一個影像濾波器抓取物件輪廓，再由 4x4 的卷積層將影像收縮，最後連同原編碼器收縮後的圖片一起輸入至解碼器。而影像濾波器在訓練時須設置兩個參數，分別為濾波器的平滑內核半徑 r 和正則化係數 e 。我們將會以 Qirong et al. (2020) 的去霧網路作為參考，

將平滑內核半徑 r 設 $\{2, 4, 6\}$ ，正則化係數 e 設置 $\{0.001, 0.0001\}$ 做為影像濾波器的訓練。

解碼器則由 4×4 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 $256 \times 256 \times 3$ 的圖像。

判別器我們採用 PatchGAN 架構，一系列的 4×4 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 70×70 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。

3.5 模型評估

最後會訓練出一套具有語義分割資料集以及加入引導影像濾波層的 pix2pix 訓練模型，接著會各訓練一套未訓練語義分割資料集但有加入引導影像濾波層的模型、有訓練語義分割資料集但未加入引導影像濾波層的模型以及未訓練語義分割也未加入引導影像濾波層的模型，並對四種模型的生成結果進行比較，觀察有加入語義分割資料集以及在 pix2pix 中加入引導影像濾波層與其他模型相比，是否會有更好的結果。

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。