大家好，我是李承諺，我要報告的主題是以生成對抗網路為基礎探討室內設計三維模型轉換之研究，指導老師是廖秀莉，這是我的目錄，分為緒論，文獻探討，參考文獻。

首先是研究背景: 室內設計的概念起源於美國，於經濟發展的需求演化而來，原本蓋建築物只是單純的把它蓋好就可以住了，而之後開始延伸出使用者的需求或為了更改內部環境或風水等等而開始考量室內設計。

後來室內設計逐漸從因應需求，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，沒有家就沒有我們。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。

再來是三維模型的介紹，早期室內設計會運用手繪 2D 設計圖、透視圖、等方式進行設計表現時，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業主自身的想像力，來綜合判斷完工後的實際樣態。但往往在工程完成後，發現與當初的想像有相當大的落差，造成設計師及業主雙方的困擾，嚴重的話甚至產生法律糾紛。

因此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，過程中有加入2D的合成影像，到現在三維模型的展現，也稱作3D模型(3D草圖)，而至今也有大量的繪圖軟體能讓輕鬆地生成3D模型，像是網路上非常火紅的線上建模工具3D Warehouse、SketchUp等等，而產生出大量的3D建模工具也讓我們能夠了解三維模型有多麼重要，才會陸續出現多種建模工具。

也因為3D建模的出現，也解決了許多在裝潢前設計師與客戶之間的認知差異，這些認知差異也能順利地在裝潢前順利解決，避免商業糾紛。

再來室內設計風格，一個成功的設計，除了品質跟功能，在視覺上一定會具有一定的吸引力，而設計者在設計過程中，處理各種設計要素，室內環境、構造、特徵，就會產生出不同的室內設計風格。

而影響風格的設計要素主要有空間、裝飾、色彩、光影、陳設、綠化等要素，通常設計者會依照這些因素來設計或判斷房間的設計風格。

根據 學者Simon, HA, (1975) 的說法，設計風格包括多個方面，例如設計作品中呈現的共同特徵和設計事物的方式 。因為定義明確的樣式關鍵字可以幫助用戶在設計過程中理解不同的設計並有效地相互交流。基於不同的特徵組合，可以通過多種方式定義設計風格。在建築和室內設計中，設計風格通常與文化、時間、地區、哲學或個人有關。但標籤一個設計的風格實在太主觀，因此隨著人工智慧的發展，也陸陸續續出現辨識風格的模型出現。

而在2020 Jinsung Kim等人就做出了一套模型將真實房間的照片進行設計風格的辨識，以現代、自然、古典和休閒等四種風格為基礎來做辨識，其準確率為68.8%，而模型辨識的皆為真實照片。

接下來介紹生成對抗網路，簡稱GAN，是一種非監督式學習的方法。由ian goodfellow再2014年提出來，傳說中他在酒吧喝酒，不知道她在酒加了什麼，喝一喝就想到GAN這個點子

而GAN是由兩個神經網路所組成，一個叫生成器原理為反向神經網路，他是將神經數目組合擴充，原本輸入0或1，最後將神經數目擴充還原輸出整張圖，再來是判別器就是一般的正向神經網路，將生成後的圖片進行辨識，最後輸出0或1，由這兩組神經網路不斷去對抗，生成器要生成與真實圖片最像的圖片，判別器要判別出圖片的真假，一直重複輪迴，就像正派與邪派的關係，正派越強，邪派也要變強，邪派越強，正派也要跟著變強，所以這兩個會互相競爭，互相變強，這也是為什麼叫做GAN的原因，最後雙方會找到一個平衡，一個有對手情況的最好，這就是大家經濟學熟悉的奈許均衡，達到奈許均衡也代表GAN網路訓練到最好。

再來是我的研究動機，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，而一般而言，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擬真圖渲染的時間也需花費大量時間以及技術需求，且渲染出一張完美的擬真圖可能就需要數小時以上，或者參數調整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，若想看到真實照片卻只能透過photoshop等合成軟體來完成，需花費大量時間精力。

再來是GAN鮮少對室內設計進行研究， GAN目前已對許多圖片操作有很成熟的技術，例如生成人臉、生成文字、生成音樂等，也延伸出許多GAN變種，

利如風格轉換技術，將一張普通的獅子變成梵谷畫的如下面這張圖，或是將白天照片變成晚上的照片，而GAN已存在6年多，但卻很少有人對他進行室內設計的研究，所以我將以GAN為基礎，對室內設計產業進行研究

我想透過GAN的變種pix2pix(圖像翻譯)將三維模型轉換為真實照片，pix2pix是一種無監督式學習進行圖片對圖片的變換，而三維模型轉換為真實照片可以想像成將三維模型的風格轉換成真實照片的風格，雖然普通的CNN也可以調整像素進行對圖片的轉換，但CNN的池化會壓縮圖片造成圖片不清晰，而pix2pix能解決圖片不清晰的問題

可以看到左邊是三維模型，右邊是真實照片，我想用GAN的技術達成將三維模型轉換成真實照片的目的，讓圖片更加逼真。

然後這是我初步構想的研究設計:

首先訓練前會進行語意分割，因為室內設計的房間照片會存在著許多物件

，如家具，所以我想在訓練前先透過語意分割抓取家具的輪廓避免在訓練的過程丟失細節，而現在著名語意分割架構也有許多例如FCN、U-Net等都是有名的語意分割工具。

接著在訓練時在pix2pix加入影像濾波層，影像濾波層是由許多影像濾波器所組成，他能夠將影像平滑化，減少雜訊出現，且透過增強內核的方式也能達到增強細節的效果。

而室內設計的本質在於設計，還沒居住的情況下，最能滿足客戶的設計因素不是居住的舒不舒服，而是在於視覺也就是設計的風格，因此我想透過GAN生成出特定風格的照片，因GAN辨識器只能辨識與真實照片的loss無法辨識你與某風格的距離，因此我要將生成後的照片丟進Jinsung Kim的辨識風格模型進行辨識，若辨識出來的準確率能到達68.8%左右，就代表GAN生成出來的圖片的水準有到達真實照片的水準，不同的設計風格也能夠準確地生成出來。

上面是語意分割的效果，能夠抓取物件輪廓，避免在訓練過程丟失物件

下面是影像濾波器將影像平滑化的效果，效果是能將那些胡椒鹽雜訊去除，將影像平滑化

步驟一將圖片進行與意分割，步驟二在pix2pix加入影像濾波層，步驟三訓練真實照片，步驟四將成品輸入風格辨識模型進行風格辨識。

再來是我的研究目的: 此研究使用生成對抗網路將室內設計三維模型轉換為真實照片，並利用JK模型對轉換後照片進行模型評估，最後希望透過本研究訓練出來的模型，能夠快速地將3D草圖轉換為真實照片，以解決的時間、技術以及金錢問題以滿足客戶需求，也希望產出來的結果能夠幫助室內設計產業。

我的研究問題是: A.訓練前加入語意分割，增強輪廓，觀察訓練結果

B.在pix2pix加入影像濾波層，將影像平滑化，觀察訓練成果

C. 透過不同的epoch挑選出效果最佳的照片，產生出特定風格之照片，並丟入Jinsung Kim模型進行風格辨識

論文一:

研究方法:

此研究方法的步驟分為三點:1.先準備室內設計風格的資料集，2.使用CNN訓練識別設計風格的模型，3.對驗證資料進行風格識別

首先是設計風格資料的準備: 此研究的模型資料集用韓國住宅常見設計風格的客廳照片。

蒐集來自Daum房地產、Ohouse、Ggumim、Houzz、Zipdoc等室內設計共享平台的資料，蒐集時間為2020/2-2020/6，每個平台都提供帶有風格標籤的圖片，並且保留了具有良好視野的圖片，最後蒐集了480張圖片，每種風格有120張，分別為現代、自然、古典、休閒風。

接著是資料的預處理:

第一步是標註圖片的風格，將每張圖片標註他是哪種風格

再來是調整圖片的大小，這在訓練時是必要的，因為這個研究為遷移學習使用VGG-16當主要骨幹網路來進行訓練，而VGG-16的輸入須將圖片調整至224\*224，因此在資料的預處理須先把每張圖片的尺寸改為224\*224

再來是數據增強，數據增強在訓練模型是很有用的，而數據增強的方法有將圖片做旋轉、平移、縮放、翻轉、剪切等，但過度的變形會擔心訓練模型過擬合，因此修改時須考慮到數據集的特性，所以說左右翻轉跟縮放、旋轉是比較合理的方法。

因此最後將480張先分為400張訓練集跟80張驗證集，再將400張左右翻轉，得到800張，再將800張進行縮放跟旋轉，最後得到4000張，最後將訓練的資料集分為100張、200張、400張、800、4000張，以不同的數量進行訓練，用(80%當訓練集，20%驗證集)，比較這些數量訓練後的成果。

為了加快訓練，此研究使用了遷移學習，遷移學習可以將類似問題的訓練模型被重新用於類似的新問題，以節省訓練的時間及計算性能，而此研究用預訓練過的模型Place365來當作骨幹網路，Places365 是用於各種類型場景（例如，室內、室外、辦公室和自助餐廳）的流行數據集的模型辨識，而Place365是使用VGG-16來當骨幹網路，接著在Places365的後面添加新的神經網路，這個神經網路有256個輸出全連接層，一個Dropout(0.5)及一個softmax激活函數所組成。

再來是研究結果: 首先是800張訓練集訓練後，用80張測試集預測後的混淆矩陣結果如下，

從上到下分別為休閒、古典、現代、自然風，從左至右也一樣，

可以看到休閒風預測正確機率為90%，古典風65%，現代風85%，自然風35%，

總準確率為68.8%(top-1)，可以看出前面三個都訓練得不錯，只有自然風不太理想，透過混淆矩陣發現，自然風時常與古典風搞混。

這些分別是不同資料集數量去訓練的，100張是32.4%，200張是54.3%，400張是59.8%，800張68.8%，4000張反而下降了變成61.5%，可以看到800章訓練的效果是最佳的高達68.8%，且這些都是Top-1的預測結果，68.8%已經很接近人類了。

再來是論文二:

以多尺度加權融合空間注意力生成對抗網絡的圖像除雨之研究，作者是Xiaojun Bi、Junyao Xing(2020)出處IEEE，頁數為69838-69848

為何挑選此論文

研究問題:

解決下雨天照片雨痕的問題，要將下雨天照片的雨水去除，而除雨的工作在深度學習的技術上早已達到，雖然效果不錯，但仍存在一些問題，像是在使用深度學習中訓練圖片的雨痕尺寸是有限的，但現實雨天的情況下中雨痕尺寸繁多，無法準確分類，所以在真正除雨的過程中可能會造成過度除雨或沒除到的情況，造成該被除的地方沒除到或是不該被除的地方被除到，導致圖片對比度或飽和度發生改變產生模糊的情況，因此此研究使用多尺度加權module 來加強特徵的提取，再用GAN利用他的鑑別器和生成器對抗的特點更有效的去除雨水，保留更逼真的圖片，因此此研究訓練出一個新的GAN叫做MWA-GAN來與其他著名的除雨模型進行比較。

研究方法:

簡單來說就是GAN分成生成器跟鑑別器，生成器的工作是將含有雨水的圖片生成出沒有雨水的圖片，鑑別器的工作就是判斷創造出來的假圖與真實無雨圖片的真假，生成器包含多尺度加權Module、改進空間Module、自編碼器

鑑別器包含卷積層、batch-normalization、ReLu層，將GAN加入那些Module之後提出新的GAN叫做:MWA-GANet再將下雨的照片的雨痕去除

而生成器多加幾個Module比較複雜所以我提出來解釋，生成器比較複雜，我提出來解釋

多尺度加權Module的使用:

一般在提取圖片特徵用卷積層跟池化層，卷積用來提取特徵，池化用來特徵融合，可是這樣感受野會不夠大，要增加感受野第一個是對圖片下採樣，但怕圖片過小會抓不到小特徵，或是增加卷積次數增加感受野，但增加卷積次數會增加網路的計算量，增加網路負擔，因此增加感受野的方法使用擴張卷積，他能夠不增加參數及不損失圖片分辨率來增加感受野，但擴張卷會有間隔，有間隔會產生網格偽影，偽影是指原本被掃描物體並不存在而在圖像上卻出現的各種形態的影像，為解決此問題而使用三路擴張卷積，三條路徑使用三個不同的擴張卷積，最後融合解決網格偽影的問題，然後三個擴張卷積分別使用1、2、3的擴張因子，達到多尺寸特徵的抓取，一方面解決了網格偽影問題，另一方面可以使最終的特徵圖包含不同尺度的訊息，使網絡能夠提取更豐富的雨線特徵(最後感受野為70\*70)。

空間注意力Module:

這個就像RNN的提取上下文的特徵，忽略不相關的資訊，能更準確抓到雨條痕的位置

上下文自動編碼:

由前兩個方式，先抓取不同雨的特徵，在準確抓取雨的位置，能正確的抓取位置及雨痕大小去除雨水，生成沒有雨水的照片。

再來是資料的蒐集，此研究參考前人的資料集，之前有人已經用深度學習研究過去除雨痕的照片，他將雨痕分為輕、中、重三個等級且每個等級含有4000張圖片，但現實生活中雨痕不可能只有這三種，雖然模型訓練不差但問題還是存在著，而這個研究要參考她的資料集且只抓取中等密度的雨水4000張進行訓練，測試集抓取1200張不同雨條文密度及方向的照片，為了是要證明此模型能用較少的資料集以及單一密度的雨痕資料集就能夠完美的去除雨痕。

研究結果:直接看圖片，左上角為原本沒有雨的真實照片，右邊三張是三個不同雨水密度的驗證資料集，下面是分別將三張去雨痕的效果，可以看出只訓練一種中密度雨痕的資料卻可以將三種不同雨痕密度的照片去除雨水且看起來非常成功。

再來是與其他除雨研究進行比較:a為真實下雨圖片b為真實沒下雨圖片，c為用純GAN除雨的效果，d為GAN+空間注意力Module的效果，e為GAN+多尺度加權Module的效果，f為MWA-GANet的效果，可以看出其他三種去雨水的訓練去完都還存在著雨水，沒除乾淨，而MWA-GANet卻把雨水儲得一乾二淨，也就是說經過與其他訓練方式進行比較能看出MWA-GANet效果最佳。

論文三:具有引導過濾層的pix2pix去霧網路

研究問題: 1.隨著GAN風格轉換的發展，將除霧的工作想像為風格轉換的工作2.除霧效果因霧的濃度分布不均，在GAN中受到限制3.GAN中加入影像濾波層，以最大限度保留模糊圖像的邊緣細節

早期深度學習可以除霧，利用DRL將模糊圖片轉成無霧圖片，DRL使用13層ResNet(殘差網路)來學習。

隨著GAN風格轉換的發展，我們可以把除霧的工作想像成把一個有霧風格的圖片轉換成一個無霧風格的圖片

但除霧效果是不能用風格轉換直接去訓練的，因為每張圖片霧的濃度會隨著照片內容的遠近而有所不同，可能我拍近的地方這邊霧比較少，遠的霧會更濃，導致霧的濃度在圖片分布不均，如右下角的圖片，左下角的椅子沒什麼霧，但遠的地方看得出霧的濃度非常濃

因此此研究在網路中再使用GAN變種pix2pix圖像翻譯的功能進行風格轉換，再加入影像濾波層，利用去除雜訊的特徵，把霧當成雜訊去除，以最大限度地保留模糊圖像的邊緣和細節。這樣，過濾器能夠有效的去訓練出每一個抓出來的特徵，抑制去霧過程產生的圖片模糊的副作用。

整個網路架構: 因為在霧霾的圖片中，整個圖片的輪廓會因為霧的分布不均而受到影響，因此此研究加入了引導濾波層來去除雜訊獲取模糊圖像的輪廓，將輪廓萃取出來再拼到彩色圖像去進行訓練，防止訓練時模糊影像沒被獲取到

透過影像濾波層抓取圖片特徵的效果如右圖，可以看到設置不同的平滑內核跟正則化參數，能抓取出來的特徵也不一樣大，平滑內核越大能捕捉更多細節特徵，解決圖片被霧影響的問題，正則化參數則是防止過擬合

這個GAN模型一樣分為生成器跟鑑別器，生成器用於通過保留圖像的結構和細節並消除霧度來生成清晰的圖像，編碼器在卷積層下操作，利用下採樣增加感受野，解碼器再進行上採樣，還原圖片大小。

鑑別器接受生成器的輸出，判斷生成的圖像是否是真實清晰的圖像。基本操作:卷積層、 batch normalization、 LeakyReLU(防止神經元死亡)

訓練資料:1449對訓練資料進行訓練(真實圖片及合成後的加霧圖片)

測試資料:80對測試資料

引導濾波層:影像濾波器的平滑內核半徑分別設置為{2, 4, 8, 16, 32}，正則化係數設置為{0.001, 0.0001}

最後是研究結果:

（1）我們的方法有效地去除了真實霧霾圖像的霧霾，即使在合成數據集的訓練期間，也證明了該方法的成功性；(2) 可以看到DCP的訓練結果導致天空區域的顏色失真，但我們的方法沒有這個問題，解決了DCP帶來的負面影響；(3) DehazeNet和 AOD-Net是較差的去霧方法，而 DCPDN 和 cGAN無法有效消除濃霧圖像中的霧霾。此研究提出的方法展示了更好的視覺效果

論文四:以生成對抗網路為基礎之閩式建築風格轉換之研究，此為碩士論文

研究問題:由於閩式文化逐漸沒落，此研究想透過人工智慧看能否保存敏是文化，另外在設計遊戲場景，需要閔式建築時不須花費時間採景，利用風格轉換將美式風格之建築轉換為閔式建築，降低開發遊戲之門檻，讓沒有繪圖能力的遊戲開發者也能夠取得其所需的遊戲場景。

左邊為閩式建築示意圖，右邊為美式建築示意圖

再來會透過兩個變種GAN，CycleGAN跟InstaGAN進行風格轉換並比較兩者成效差異。

研究方法

1.準備訓練用的圖片資料集，使用cycleGAN對資料進行訓練

2.使用InstaGAN進行風格轉換，訓練前先使用FastFCN進行語意分割

3.轉換後對兩者不同GAN的成品進行比較

CycleGAN為兩個GAN所組成，不須對英的圖片就可以進行訓練，因為他會先產生出另一種風格的照片，接著再將轉換後的風格在轉換為原本風格的照片，透過此循環，訓練出A風格與B風格的照片在輸入鑑別器做辨識，因為不須對應圖片因此解決了資料集不足的問題。

FastGAN進行語意分割分割出建築物已集背景在萃取出建築物得特徵進行訓練，InstaGAN須先進行語意分割，且會一次轉換部分的特徵，然後多做幾次九ˋ可以把所有特徵都轉換完成，

它的優點就是能萃取較多的特徵並減少網路計算量。

再來是資料蒐集的方法:

為了能夠轉換較完整的閩式建築，選擇了同樣有屋頂跟屋簷的美式建築。

閩式建築資料及的蒐集方式用google爬蟲的方式，在估狗下關鍵字:閩式建築、金門傳統建築、台灣傳統建築，把搜尋結果全部下載下來，在用以圖搜圖的方式進行擷取，在手動過濾掉不符合的圖片。

美式建築的資料擷取，跟閩式建築一模一樣，爬蟲的關鍵字為別墅、美式別墅建築，最後在進行資料增強，將每張圖片垂直翻轉、逆時針、順時針旋轉以達到所需的資料量。

而CycleGAN生成器的架構分為編碼器、變換器、解碼器:

編碼器:3個卷積層組成，會縮小影像。

變換器:殘差連結，為9個殘差塊組成，為防止梯度消失或梯度爆炸

解碼器:兩層轉置卷積，將圖片尺寸放大，輸出影像

CycleGAN判別器的架構為:PatchGAN:

會先將所有影像分割為70\*70，在一個一個判定個塊影像的真實率。

再來是語意分割，此研究使用FastFCN模型進行語意分割，他使用預訓練模型ade20k，25000張圖片(訓練集20000，驗證集2000，測試集3000)其中建築類占約10.7%，使用0.001學習率及140epoch，將建築進行語意分割

分割完再用openCV將圖片轉為遮罩

遮罩就是把影像遮住，步驟如圖所示，先進行與意分割，在進行遮罩遮掉背景圖

InstaGAN進行風格轉換的步驟:

1. 將原資料集進行處理後生成遮罩
2. InstaGAN需連同遮罩圖及原圖一起輸入，生成也會連同遮罩一同生成
3. 先分別萃取遮罩特徵及原圖特徵，之後先將遮罩特徵結合，再將原圖特徵與遮罩特徵結合，再送入辨別器，辨別器先提取特徵再進行辨別，並使用不同的epoch進行比較

研究結果:

CycleGAN由不同的epoch進行比較，分為實驗一200、實驗二300、實驗三400，學習率皆為0.0002，如圖為loss變化

由本次實驗結果可以觀察出來，到了第 200 epoch 時，閩式建築的燕尾特徵，如 下圖 4-1 左圖黃色圓圈處，已經有部分生成出來了，只是還有些殘缺，到了 4-1 中圖 實驗 2 時，黃色圓圈處生成出來的特徵將近完整，但是依舊殘缺不全。於是我們將訓 練繼續增加 epoch，再到了 4-1 右圖實驗 3，結果仍然沒有太大的差異。如下圖 4-2 右 圖黃色圓圈處所示，可以觀察到先前的燕尾形狀已經漸漸開始產生多餘的形狀。因此 我們認為在 epoch 到第 300 次的實驗 2，是比較理想的狀態。

可以發現到實驗 4 的實驗結果已將閩式建築的特徵 呈現得相當到位，而此次生成的遮罩中，在建築物的屋頂也有了燕尾的特徵。在實驗 5 的時候，中上圖黃色圓圈處和中下圖的遮罩上的屋脊已經開始生成出多餘的燕尾了，右 圖實驗 6 時也是如此。由本結果來看，我認為這次的訓練只要達到實驗 4 的 200 epochs， 就已足夠將一些明顯的閩式建築特徵呈現於結果當中。

InstaGAN(左) VS CycleGAN(右)

兩個都表現不錯，但各有優缺點，InstaGAN產生出的圖片較為模糊，但較能夠把建築特徵生成出來，CycleGAN較為清晰，但紅圈的特徵卻完全沒有出現

綜上所述，如果是在需要清晰 影像的情況下，實驗 2 的方法會是較為推薦的做法；而在需要完整的將特徵呈現且清晰 度的需求又沒那麼高的情形之下，實驗 4 的方法將能夠達到更為符合需求的結果

論文五:

基於生成對抗網路的繪畫風格轉換

首先將原圖片切成數小塊，也將風格圖切成數小塊，再將原圖的數小塊丟入產生器訓練，生成出的圖片再與風格圖一同丟入鑑別器做訓練，經重複訓練後，再用拼圖的方式完成影像。

概念為分治法:此概念為將一個大困難問題簡化成兩個或多個相同的子問題來作解決，就如同將大張原圖裁切成多張小塊

研究結果:

觀察再研究前先進行分群會發生什麼事情，因為擔心GAN中是透過亂數去隨機抓取進行學習，但風格樣試圖與原圖若相差太遠，產生的結果就會崩壞，因此再訓練前額外加入分群方法，再進行隨機抓取的時候都抓到相近的圖片來改善此問題。

但實驗結果發現事前分群會讓圖片出現一格一格的小塊圖樣，導致結構線條消失，無法讓結構保留下來，因此研究不採用分群的方式

線條等細節非常流暢，沒有特別突兀的地方，反觀其他人提出的方法紋理線條不連貫，右下角顏色也與原圖相差許多