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 GAN

研究生:李承諺

指導教授:廖秀莉 教授

中華民國一一零年O月

1. **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。

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，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。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(3D草圖)，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3D Warehouse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效果圖，3D效果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效果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。

人工智慧的發展…

**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效果圖，但渲染時需添加不同的參數，若想渲染出非常完美的3D效果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或是3D效果圖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(圖一)，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效果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效果圖突破為真實照片的樣貌。

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;判別器必須分辨出圖片是否為原圖片還是生成器生成出來的假圖片，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，設計照片絕對會強調照片能夠越清晰越好，因此本研究會在訓練時加入影像濾波器，希望在訓練後能將影像平滑化去除雜訊，產生出高清晰的圖像。另外因室內設計的房間照片中會存在多個家俱，屬於多類別的照片，因此此研究將嘗試對圖片做實例分割來進行訓練，目的希望將具有多家俱的室內房間照片能夠保留房間的所有物件，並在訓練的過程還原所有前景，最後比較兩種案例實驗後的成果。

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。

1. **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一。

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，而早期在電腦還沒普及的時候，室內設計往往會運用手繪 2D 設計圖、透視圖等方式進行設計的展現。雖然能夠正確的表達設計師的設計構想，但在客戶端時常需要靠著自身的想像力，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手繪的方式及2D圖的呈現，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圖像，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3D Warehouse 、SketchUp(SKP)、3D Studio Max(3ds Max)、Rhinoceros 3D(Rhino 3D)等都是常見的3D建模軟體，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，而在渲染軟體這部分以V-Ray為主流，常用的3D建模軟體都能與V-Ray做結合，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。

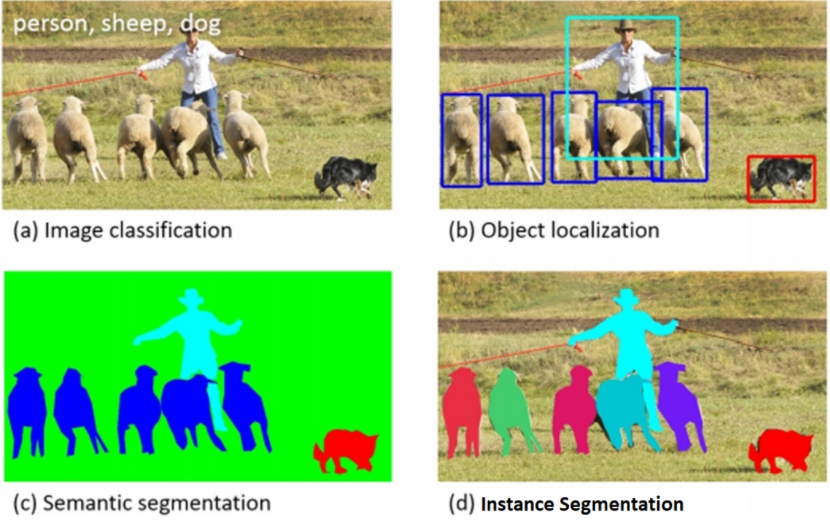


圖一為SketchUp所建立的三維模型

**2.2 實例分割(Instance Segmentation)**

實例分割為基於物件偵測與語義分割的一種方法。物件偵測是指對於圖像上去查找物體的類別並檢測它們在圖像上的位置，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。語義分割是現今計算機視覺的領域廣泛流行的技術，是指將圖像中的每一個像素進行分類，歸類每格像素的種類接著進行區域劃分。

實例分割為物件偵測與語意分割的結合，其任務較為困難，需先將圖片的ROI區域(感興趣區域)偵測並提取出來再將每種不同ROI區域內的像素點進行語意分割，進行像素分類，因此即使是相同類別也會被分割成不同的物件如圖二。



圖二 物件分割示意圖

**2.3 Mask-RCNN**

RCNN(Region-based Convolutional Neural Networks)，為區域的卷積神經網路，是一種結合區域提名(Region Proposal)和卷積神經網路(CNN)的方法。

卷積神經網路主要的目的是在進行特徵萃取，而傳統的卷積神經網路是透過一個或多個濾波器(filter)逐一掃過整張圖片，並對每個濾波器進行特徵偵測。但CNN更適合用來做影像的分類，因為一旦有不同大小的物體存在還需要使用不同大小的濾波器來進行掃描，且每次掃描都需要進行CNN的計算，需要耗費大量的運算資源，也因此有了RCNN的出現。

比起傳統CNN，RCNN多了Region Proposals的機制，RCNN會先將圖片分割為2000個小影像，稱為ROI(Region of Interest)，再將每一個ROI輸入至預訓練好的CNN模型進行特徵萃取，再以SVM(Support Vector Machine)分類器將每一組ROI區分前景與背景，最後再由邊界框(Bounding Box)抓取物件位置。雖然RCNN先將圖片切割成2000塊小影像在進行CNN的計算，但仍要計算到2000個左右的影像，速度仍是不夠快，也因此有了之後RCNN改良模型的出現如Fast R-CNN、Faster R-CNN、Mask R-CNN等。

Mask R-CNN改良了傳統的RCNN，傳統的RCNN會先以搜索的方式分割2000個ROI，而 Mask R-CNN基於Faster R-CNN的方法，與其先分割成2000個小影像，不如先進行CNN進行特徵篩選，再將有用的影像進行Region Proposals，以避免訓練到不重要的資料，以節省運算時間。而Mask R-CNN也將傳統RCNN以邊界框來抓取物體位置改為抓取圖片的像素對物體進行圖像分割(Image segmentation)，利用FCN(Fully Convolutional Networks)進行計算取得圖像的遮罩分割。而傳統的CNN在過程中會有池化(Pooling)的動作，池化是使用最近插值法(Nearest Neighbor Interpolation)取得數值，但此方式會對原影像與輸出影像產生距離的誤差，導致最後圖像分割後可能對原始位置有偏差，因此Mask R-CNN使用雙線性插值法(Bilinear Interpolation)來改善傳統池化，解決位置偏差的問題，稱為ROIAlign，讓遮罩位置更佳準確。

**2.4 生成對抗網路(GANs)**

生成對抗網路(Generative Adversarial Network)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，判別器會回傳一個純量(Scalar)，純量越大代表輸入的假圖片越接近真實的照片，然後將其結果回傳至生成器。生成器接續用此回饋結果調整參數繼續改良生成圖片，再輸入至判別器回傳結果，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假，兩邊最終達到奈許均衡。奈許均衡是John Nash(1950)提出的一種非合作博弈論，指雙方都在有對手條件的情況下達到的最佳平衡狀態。

圖?? 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 Mehdi(2014)等人所提出。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純量就會越高。

圖??cGAN基本架構

**2.6 Pix2Pix**

Pix2Pix是以cGAN為基礎做的改進，來自於UC Berkeley團隊所提出的Image-to-image translation with conditional adversarial networks，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skip-connections，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。

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Scalar，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。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。



(圖二)Pix2Pix實際應用的網路範例圖

圖?? Unet基本架構

圖??PatchGAN基本架構

**2.7 引導影像濾波器(Guided Image Filter)**

引導影像濾波器(Guided Image Filter)為Kaiming He等人(2012)所提出。引導影像濾波器為一種能將影像保持平滑的濾波器，能夠清楚抓出影像的邊界，達到讓影像平滑的效果。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。不同於一般的濾波器，引導影像濾波器沒有較複雜的計算，是以線性組合為基礎出發，且引導圖片與輸出圖片的方向一致，不會出現梯度反轉的問題。

而引導影像濾波器除了能將影像平滑化之外，也有著不同的延伸。當引導圖同時為輸入圖片時，最終的結果就會是保留輸入圖片的邊界，並通過濾波器還原回輸出圖像，達到細節加強的效果。

**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。分為系統架構、資料蒐集、實例分割、生成對抗網路、影像濾波層。

系統架構為本研究的整體研究架構流程，資料蒐集為三維模型與對應真實照片的蒐集方式，實例分割為對增強資料的運用，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。

**3.1系統架構**

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實例分割，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。其中架構如圖3-1所示。

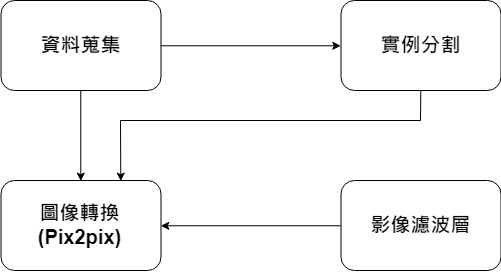


圖3-1 系統架構圖

**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。因為室內房間照片屬於多物件的照片，因此本研究將採用實例分割的方式將房間的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Mask-RCNN來對圖片進行實例分割，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，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。

**3.3 以Mask-RCNN進行實例分割**



為了更精確的訓練到影像中每個傢俱物件的細節，在訓練前先將每張影像進行實例分割，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Mask-RCNN來做實例分割，先抓出前景的輪廓，接著再利用OpenCV把木目標圖中的背景作為遮罩，產出保留前景的遮罩圖檔，再丟入模型進行訓練。

**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，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，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。

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。

**第四章 參考文獻:**

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.

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 (pp. 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 (pp. 1125-1134).

Morgenstern, O., & Von Neumann, J. (1953). Theory of games and economic behavior. Princeton university press.

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

Weng, W., & Zhu, X. (2015). Convolutional Networks for Biomedical Image Segmentation. IEEE Access.

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.

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

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。

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

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。