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

中華民國一一零年十一月

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單純居住，而之後開始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。後來室內設計逐漸從因應需求，考量環境，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3D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建模工具也讓我們能夠了解現今三維模型在市場中有多麼重要。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。

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。

**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

圖渲染的時間需花費大量時間以及設計師的視覺化思考與技術需求，但在繪製過程中若參數調整有些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，需耗費大量時間精力及技術需求。

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，其中將草圖轉換為真實照片的部分，已經有可以做到將建築立面圖像轉換為真實照片的網路，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，也滿足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。

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，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，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全景分割來進行訓練，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，解決的時間、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**

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 Studio Max(3ds Max)、SketchUp(SKP)、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。

**2.2 生成對抗網路(GANs)**

生成對抗網路(Generative Adversarial Network)由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，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。

**2.3 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，

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。

**2.4 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。

**2.5 引導影像濾波器(Guided Image Filter)**

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，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。

**2.6 全景分割(Panoramic Segmentation)**

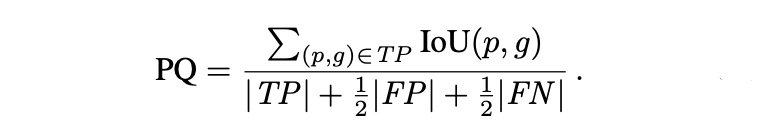
全景分割來自於由Kaiming He等人所提出的，主要任務為將語意分割(Semantic Segmentation)與實例分割(Instance Segmentation)統一結合起來。

語意分割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，需先將圖像感興趣的物件偵測並提取出來再將每種不同物件內的像素點進行語意分割，歸類出來，因此即使是相同類別也會被分割成不同的物件。

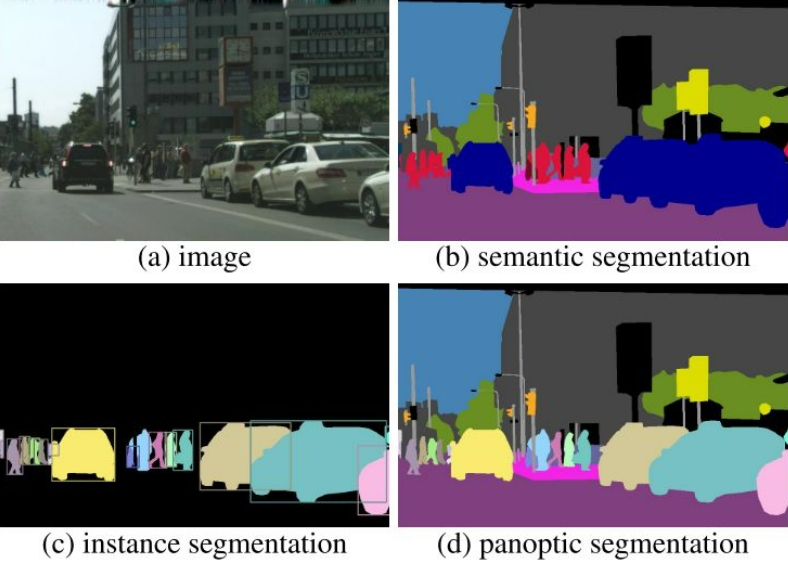
全景分割即是融合了語意分割與實例分割對圖片中的所有物體進行偵測和分割(包含背景)，現今全景分割也有許多應用，例如自動駕駛。早期就有人提出過全景分割的概念，但當時尚未有一個合適的評估指標導致當時全景分割未受到語意分割與實例分割同等的關注，而在2018年Kaiming He等人提出來一個評估指標Panoptic Quality(PQ)，透過該指標將能同時考量分割遮罩的IoU(Intersection Over Union)品質以及類別預測的準確性，其中PQ考量主要分為三點:

1. 完整性(Completeness):能同時評估Stuff Class與Thing Class的完整性問題。
2. 可解釋性(Interpretability):有實際的意義並且可以理解模型的訓練方式。
3. 簡單性(Simplicity): 設計簡單，方便於其他研究重現或使用。

之後所設計的公式如下:



其中TP(True Positive)為大於0.5的IoU，意義即為只將IoU大於0.5預測物件進行計算。雖說全景分割近似於語義分割與實例分割結合而成，但Kaiming He撰寫文中特別強調 ”*PQ is not a combination of semantic and instance segmentation metrics*.”，主要還是因為計算PQ時是設定 IoU=0.5作為區分正確與否的重要依據。因為不論是語意分割或是實例分割皆會將IoU平均在進行運算。而EfficientPS為常見的全景分割工具，不論是風景、街景還是自動駕駛的預測準確度皆位於榜首，其作法是在語意分割上使用可分離式卷積，更好的捕捉精細的特徵及上下文訊息;實例分割上使用了Mask RCNN，在檢測及分類目標的同時也將資料分為前景與背景。



(圖一)a為原圖片、b為語意分割結果、c為實例分割結果、d為全景分割結果

參考文獻:

EfficientPS 出自论文 EfficientPS: Efficient Panoptic Segmentation，该文作者来自德国弗莱堡大学

Panoptic Segmentation

ICCV 2017的best paper

Image-to-Image Translation with Conditional Adversarial Networks

Conditional Generative Adversarial Nets

An Enhanced pix2pix Dehazing Network with Guided Filter Layer

Guided Image Filtering