(一) 摘要

在空間資訊中，影像匹配(Image Matching)為一廣泛應用於數位攝影測量之影像處理技術，藉由半自動或自動化程序匹配影像中對應之目標位置，建立影像間幾何關聯性，其應用面涵蓋幾何校正、影像拼接與模型重建等。以室內定位之傳統方法而言，多半會使用影像匹配的方式，通過匹配點之間的差異獲得其空間關係，透過 2D 影像之間的匹配功能，可以推斷出相機參數，進而由 2D 影像中獲得 3D 訊息，並建立 2D-3D 之對應關係，進而計算出其定位位置。然而不同於影像匹配，在深度學習中，基於類神經網路進行影像辨識，影像辨識除了可用於辨識影像中的物件以進行分類外，也可以利用於在影像的語意標註，其中與空間資訊最為相關的，為基於卷積神經網路所進行的室內定位。

室內定位為將多張影像進行定位，以類神經網路方式訓練之模型，使類神經網路學習影像與定位坐標間的關係，進而得知室內定位成果。進行此項模型訓練時，如同大多深度學習所遇到之困境，資料集要如何取得及擴增資料集，為最大的問題。而在室內定位中，因為要大量的室內影像作為訓練資料，非常耗費人力、時間，且在特殊情形下，有無法大量取得場域內影像之情況，如場域內涉及機密內容，無法進行大量攝影，或場域結構關係，無法照射陽光等情況，無法得到清晰影像。

為了解決此困境，本研究擬以利用 Unity 製作之模型，利用模型內對各位置之影像擷取，視為場域內影像，並使影像透過已訓練之 GauGAN 模型，使影像由插畫形式轉為偽真實影像，以達到與場域內真實影像之仿真效果，再使用此偽真實影像進行室內定位之模型訓練。除此之外，將此模型訓練成果與使用真實場域內影像進行室內定位訓練，兩者進行成果的比較及研究。實驗將以國立成功大學測量及空間資訊學系系館，即為一單棟建築物為例進行探討，建物結構簡單，無特殊機關或設計，因此將建物本身影響因子降至最低。

(二) 研究動機與研究問題

現今因資訊產業的快速發展，室內定位技術成長快速，方法種類繁多，其中透過人工智慧技術影像室內定位，利用所蒐集的影像，透過 AI 人工智慧技術建模，建立資料庫，同時利用手機拍攝現場環境照片，經電腦運算後，準確預測使用者在室內的位置。然而，因建立資料庫需大量高品質影像，且需記錄所有影像之姿態，需要人工的方式進行大量現場環景拍攝，以實務狀況而言成本過高，且容易受到外在因子所影響，如日照強度與日照角度因時間有所變化，因此對於拍攝的影像而言，考驗非常大。除此之外，有許多場域是無法進行拍攝影像，如軍事用地因涉及相關國安機密，因此無法進行室內定位，然而可利用現成物件模型，加上虛擬實境(Virtual Reality，VR)場景模擬，形成 3D 模型，因此若能以模型進行定位，則可解決此困境。

以擁有 3D 模型作為前提，並以模型擷取影像作為室內定位，因模型影像與實際影像差異過大，因此可以推測成果精度表現會較差，因此利用 GauGAN 的訓練模型進行模型影像轉換偽真實影像。形成偽真實影像後，將其定位數據與偽真實影像作為資料庫，做為現場環境影像來源，進行深度學習之訓練。藉由 GauGAN 將模型影像轉換為偽真實影像，使訓練之資料庫影像能夠更接近現場環景影像，進而得到成果精度較高之室內定位模型訓練成果。

本實驗可以與使用現場環景影像訓練之模型成果相比較，因實驗場域為國立成功大學測量及空間資訊學系系館，此實驗場域光線充足，因此基於本實驗之訓練資料來自於轉換後偽影像，相較於真實影像作為訓練資料，在同樣的測試資料下，以偽真實資料作為訓練來源之模型，其成果精度推測會較真實影像的環景影像訓練模型成果差。然而若是相差於兩個標準差內，則可以推論此研究方法雖精度成果表現稍差，但是可以解決現場拍攝環域影像之問題與困境，在人力成本上也能大量的減少。

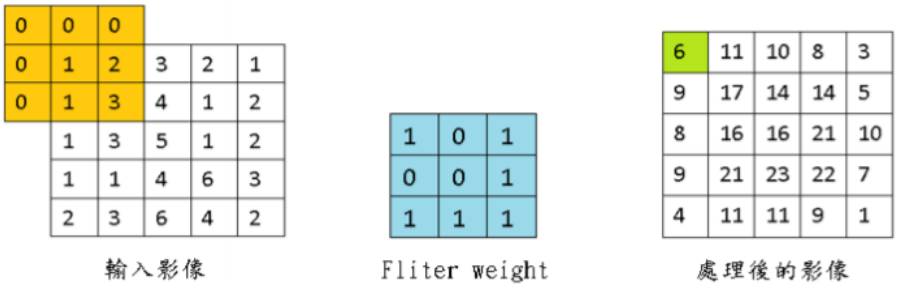
(三) 文獻回顧與探討

在室內定位領域而言，現今有多項方法可使用，可結合物聯網(IoT)之應用(陳麒仁，2017)，具備三種功能之室內定位技術: (1)具備公分等級精準度之室內定位技術，且能偵測使用者臉部方向；(2)利用監視攝影機，以降低額換建立定位環域所需成本；(3)使用物聯網為定位感測應用之創新基礎設施，能夠實現公分等級精準度之個別使用，以智慧裝置互動，進而達到定位效果，然而此研究方法需要額外的裝置，在成本上較高。若能以單一裝置即可進行定位，則可大幅降低定位所需成本。因此若使用 IP-IMP(Indoor Position by Image Matching based on Panorama)的方式(張峻瑋，2018) ，則是以影像辨識與特徵匹配之方法，結合影像比對及距離、方位計算，以達到在室內定位之效果。然而此方法又涉及角度及方位之計算，如何推測其角度、方位，成為整體實驗影響的關鍵。除此之外，也可以使用影像辨識的技術(Jing-Mei Ciou 等人，2019)，以室內定位位置影像及室內定位位置數值作為訓練資料，訓練模型學習透過影像預測定位位置。

卷積神經網路（Convolutional Neural Networks，CNN）於影像辨識而言具備三特點: (1) 輸入的影像能一定程度上匹配網路結構；(2)訓練參數因權重的分配而減少，且使神經網路之結構更加簡單、擬合；(3)特徵的提取和模式的分類可同時進行。

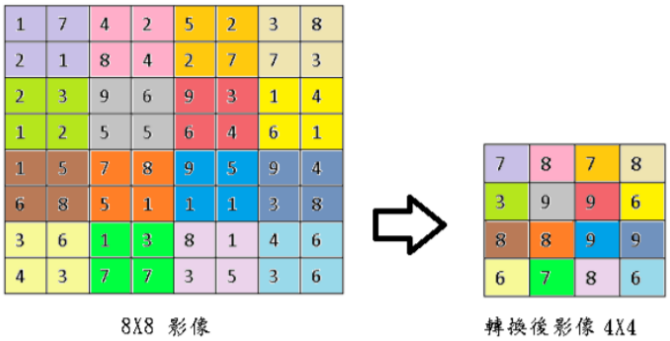
其中，卷積神經網路又分為兩部分(黃星壬，2020)，分別為影像特徵之提取，及全連接層。影像特徵之提取為圖片經過兩層卷積層(Convoltuion layers)，兩層池化層(Pooling layers)，尋找影像圖片的特徵。而全連接層(Fully Connected layers)則是內含平坦層(Flattern)、隱藏層(Hidden layer)、及輸出層(Output layer)。而卷積神經網路正是將上述兩部分結合，即為 CNN 架構。

卷基層(Convolution Layer)中，卷積運算(如圖一)極其重要，經由與 Feature Detector 做卷積運算，產製多張影像，如同將影像卷積，使影像特徵更為突出。



圖一、卷積運算的過程(黃星壬，2020)

而在池化層(Pooling Layer)中，主要方法為採用 Max Pooling(如圖二)。Max Pooling 方式為選擇固定大小矩陣後，將每一矩陣內部中最大數字提出，作為新影像使用，且新影像像素較原影像小，此目的主要在於影像平移的判斷與原影像相符，且有消除雜訊、降低訓練時長、影像位置差異減小、訓練之參數量的降低之功能。



圖二、Max Pooling 之轉換方式(黃星壬，2020)

而以全連接層(Fully Connected layers)而言，則是將前半部分的影像進行平坦化，即為 reshape 為一維向量，經過隱藏層(即模擬內部神經元傳輸過程)，再以輸出層將成果輸出，得到預測結果。

其中，於卷積神經網路中，模型架構影響其訓練過程，而 Jing-Mei Ciou 等人(2019)所提出之室內定位深度學習模型以 PoseNet 作為模型架構。PoseNet (Alex Kendall 等人，2015)為基於 GoogleNet 之改良，PoseNet 為一 23 層深度卷積神經網路，利用轉移學習(Transfer Learning)，將分類問題(Classification)的資料庫應用於解決複雜的影像迴歸(Regression)之問題，其訓練得到之特徵，相較於傳統的區域性視覺特徵，對於不同光照大小及光照角度、運動模糊的情形、及不同相機內參數等，具備更強的穩健性(Robustness)。

除此之外，Alex Kendall 等人(2015)使用了一種生成數據的方法，該方法使用運動中的結構來生成相機姿態的大型迴歸數據集，利用 SfM 生成訓練樣本的標記（相機位姿）能夠單純利用影像來生成資料，用於訓練 PoseNet 模型的訓練資料與標記，此方法得以有效解決樣本不足之情形。除此之外，研究中嘗試了十種影像格式(如表一)，並且依照不同格式進行研究，得到每種格式之精度差異。

Posenet_ori	PoseNet 原先使用的方法，該方法的影像大小調整為 455×256，然後裁剪到 224 X 224 在中心，並且該模型將被加載預訓練的模型。

Posenet_nonpy	PoseNet 原先使用的方法，該方法的影像大小調整為 455×256，然後裁剪到 224 X 224 在中心，但這種模式將不加載預訓練的模型。
Resize224x224_npy	影像直接調整到 224 X 224，該模型將加載預訓練的模型。
調整大小 100x100	將影像直接調整為 100 x 100
調整大小 150x150	將影像直接調整為 150 x 150
調整大小 224x224	將影像直接調整為 224 x 224
調整大小 250x250	將影像直接調整為 250 x 250
調整大小 300x300	將影像直接調整為 300 x 300
調整大小 350x350	將影像直接調整為 350 x 350
調整大小 400x400	將影像直接調整為 400 x 400

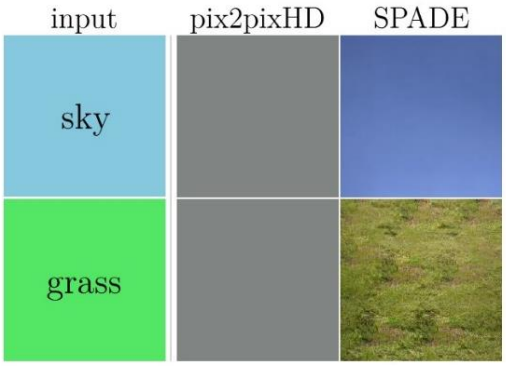
表一、室內定位模型所實驗之不同影像格式(Jing-Mei Ciou 等人，2019)

然而於 Jing-Mei Ciou 等人(2019)所提出之室內定位模型訓練中，使用的訓練資料為拍攝國立成功大學勝利校園工程系大廳後面的地下停車場做為影像來源。相較於室外環境，地下停車場光線暗淡，場景單調，並無明顯特徵，因此實驗成果表現較差，實驗中，於方向誤差和位置誤差的最佳改善率分別為 33.4%和 42%，因此可以觀察到使用卷積神經網路進行室內定位時，實驗場域之環境是影響極大的因子。

而為了改善此情形，本研究將訓練資料改為使用實驗區域之模型的影像擷取，使訓練資料不會因外在環境條件有所影響。為了使實驗區域之模型影像與實際實驗區域之環境影像相似，使用由 Taesung Park 等人(2019)所提出 GauGAN 模型。

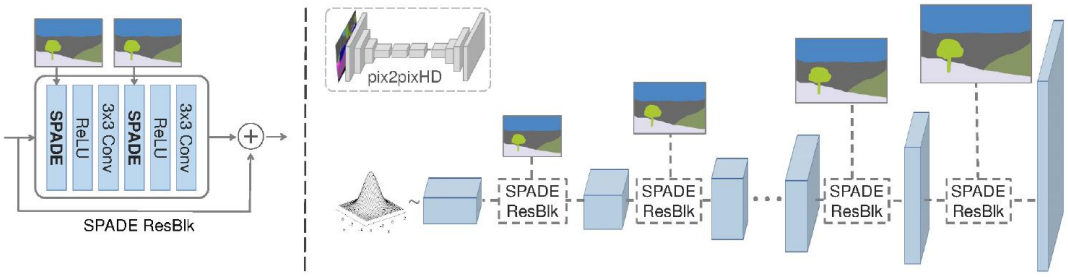
GauGAN 為 NVIDIA 公司所推出的訓練模型，GauGAN 採用生成對抗網路(Generative adversarial Networks，GAN)技術，透過 SPADE 生成影像演算法，及其大量的影像資料庫做為訓練基礎，得以將插圖轉化為影像。

GauGAN 中使用 SPADE 生成影像演算法(Sequential Pattern Discovery using Equivalence Classes)，SPADE(Mohammed J. Zaki，2001)為 GPS 演算法的改進，為了避免多次的對資料庫全掃描的情況，SPADE 增加了一 ID_LIST 的紀錄，ID_LIST 會進行交集運算，利用 lattice 階層式理論使原本的大問題分解為小的子問題，僅需使用較少的記憶體即能處理大量資料。而在 GauGAN 所使用之 SPADE，為 pix2pixHD 之改進(如圖三)。在 pix2pixHD 中是直接將語意分割映像(Semantic Segmentation Map)輸入生成網絡，然而在批次正規化層(Batch Normalization，BN)中，容易失去標號映像(label map)之標號訊息，因此增加一新正規化層，為 Spatially-Adaptive Normalization，以解決此情形。



圖三、pix2pixHD 及 SPADE 的比較(Taesung Park 等人，2019)

而在 SPADE 生成器中，每個正規化層都使用 Segmentation Mask 來調製每一層的激活。如圖四左為帶有 SPADE 的 Residual Block 的結構；圖四右則為生成器中，包含一系列帶有上採樣層的 SPADE Residual Block，此架構透過一個個移除圖像至圖像到圖像轉換網絡的下採樣層，以較少的參數獲得更好的性能。



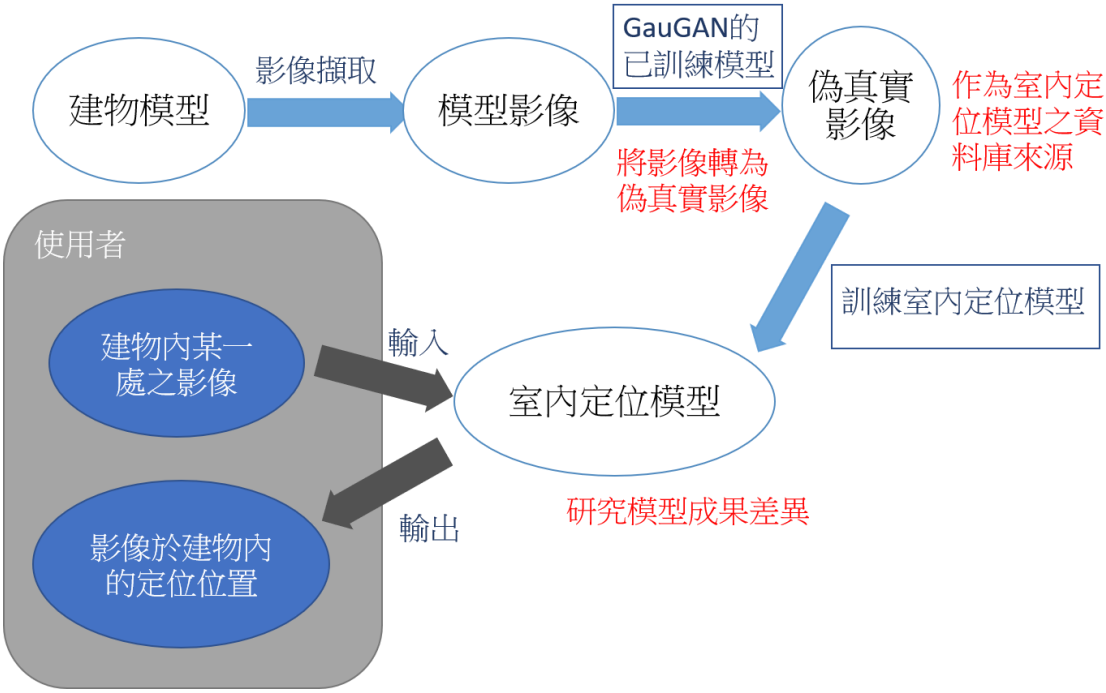
圖四、SPADE

由此可見，GauGAN 為改良後之影像轉換模型，因此能夠使用於將幾何圖像轉換為偽真實影像，然而因為 GauGAN 是以手繪插圖作為訓練資料，因此本實驗使用之建物模型影像在圖形類別上稍與訓練資料的類別不同，然而因原理皆為使用幾何圖像作為辨識方式，因此仍可進行嘗試。

(四) 研究方法及步驟研究流程圖

本研究以國立成功大學之測量及空間資訊學系系館為實驗場域選擇，將建物及建物內部以 Unity 建造模型後，進行大量的影像環域擷取，得到模型影像後，始進行研究。

本研究之研究策略及方法說明如下:

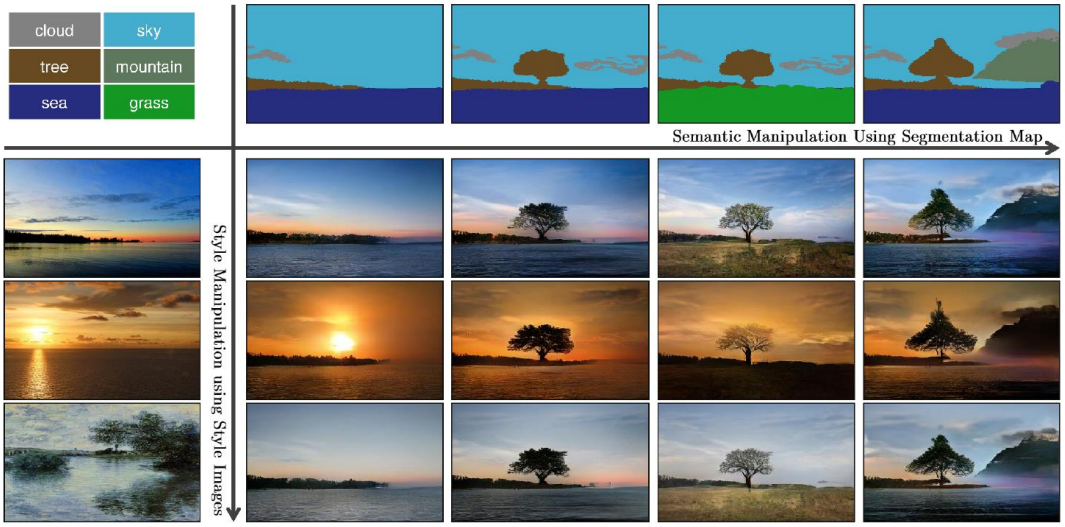


圖五、研究流程圖

1. 利用 GauGAN 將模型影像轉換為偽真實影像

GauGAN 為基於生成對抗網路(Generative adversarial Networks，GAN)技術下所建立的模型。在 GauGAN 中，將圖像區分為標號映像(label map)及樣式圖像(style image)。

將影像作為標號映像進行影像辨識，將圖片區分為各個標號，如海洋、樹木、雲朵等，視為在圖上進行標號，區分物件，再以標號的類型參照資料庫中的樣式，使用者可依照需求選擇需要之圖像。以圖六的第一排為例，影像匯入後，區分為不同物件，並且將不同物件進行對照類別，可以看到咖啡色即對照樹木、綠色即對照草地等。



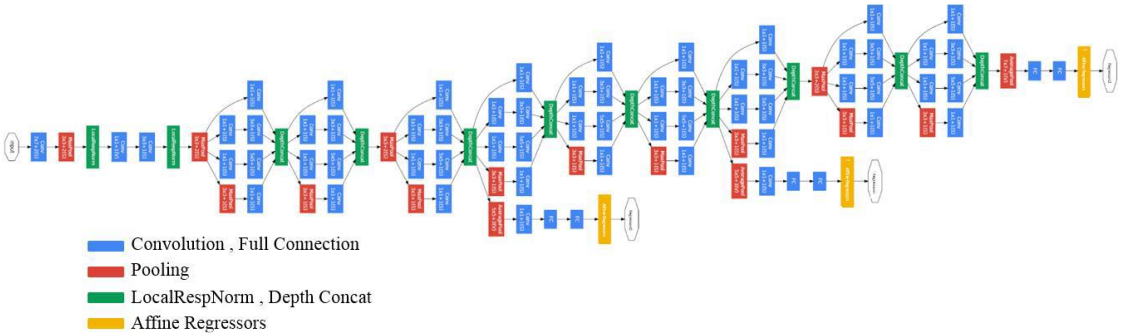
圖六、GauGAN 的標號映像及樣式圖像(Taesung Park 等人，2019)

於本次實驗中，將使用 NVIDI 已訓練成熟的 GauGAN 模型進行影像之轉換。使其先進行辨識，使影像區分為各類標號後，再依據建物之類型與建物材質，進行樣式圖像之選擇。而關於使用面上，由於 Taesung Park 等人(2019)將模型開放、釋出，因此於 github 上可得到其模型資料(https://github.com/NVlabs/SPADE)。而透過 NVIDIA 公司所架設之平台(http://nvidia-research-mingyuliu.com/gaugan/)，也可以進行影像的轉化，因此除了使用 github 上的開放模型，也可以使用自動化機器人來重複進入平台，以達成影像的批次轉換。而經由 GauGAN 的模型，可以將建物模型影像轉換為偽真實影像。

2. 使用偽真實影像進行室內定位模型訓練

將偽真實影像作為模型之資料庫來源，以卷積神經網路（Convolutional Neural Networks，CNN）進行室內定位的模型訓練，將建物模型擷取影像之姿態，以攝影測量中的共線條件計算其定位位置，將其對應之偽真實影像與計算的定位位置作為訓練資料，訓練出室內定位之模型。簡而言之，藉由卷積神經網路學習影像之於定位位置之關係，使其能夠在輸入一陌生場域影像時，能夠比較出此場域影像與哪個影像最為相似，進而推測出最有機會的定位位置。

而本實驗參考 Jing-Mei Ciou 等人(2019)所提出之室內定位模型訓練，參考引用深度神經網絡 GoogleNet 所改良的 PoseNet 架構。GoogleNet 是用於圖像分類的 22 層卷積神經網路(CNN)，其中包含六個 Inception modules 和三個分類器(classifier)，分類器用於測試階段(Testing)。Inception module 為一種將卷積層之過濾器分組的方法，同一層卷積層中透過不同尺度的過濾器來達到更好、更有用的特徵值。PoseNet 的結構上主要是以 GoogleNet 的 23 層卷積神經網路進行一些微調整。



圖七、PoseNet 的結構(Jing-Mei Ciou 等人，2019)

在改良上，主要分為三個部分： (1)將三個多分類器(multi-classifier)改為仿射迴歸(Affine Regression)，將每個 Full Connection Layer 以輸出 7 維姿態向量，包括 3 維的位置向量和 4 維的方向向量；(2)在最後的仿射再生器(Final Affine Regenerator)之前，插入一特徵尺寸(Feature Size)為 2048 的 Full Connection Layer。這將生成一定位向量，使 PoseNet 可以對其進行探索；(3)於測試階段(Testing)時，將四元數方向向量進行正規化之單位向量，因為只有四元數可以表示旋轉量。在進入訓練階段之前，PoseNet 將影像大小調整為 256 像素，並且將其中心裁剪為 224 x 224 像素。

同時，於實驗中有嘗試了十種格式來進行實驗(於文獻回顧與探討中有所提到)，其中以 Resize224x224_npy 之成果精度表現最佳(如表二)，因此本實驗將之做為參考，以影像直接調整為 224 X 224，且模型將加載預訓練的模型，來進行後續之實驗。5

Format of input image	Position + Improvement (%)	Direction + Improvement (%)
posenet_ori	1.291 m	3.622°
Posenet_nonpy	1.400 m	3.139°
Resize224_npy-3	0.749 m (42%)	2.414° (33.4%)
Resize100x100	1.431 m	3.586°
Resize150x150	1.029 m	2.471°
Resize224x224	0.933 m	2.111°
Resize250x250	0.874 m	1.773°
Resize300x300	0.901 m	2.060°
Resize350x350	0.896 m	2.140°
Resize400x400	0.851 m	2.334°

表二、十種格式之成果精度(Jing-Mei Ciou 等人，2019)

3. 進行模型成果差異研究及比較

將室內定位模型以偽真實影像進行訓練後，將測試資料輸入，可以得到其預測精度，即為室內定位模型預測之正確率。然而，若將一般室內模型與本實驗之室內比較，則可見其異同處(如表三):

	一般室內定位模型	本實驗室內定位模型
資料庫	建物內真實影像	建物模型轉換之為真實影像
預期成果精度	高	低
是否需大量人力投入	是	否
訓練資料量	有限	無限，可依需求持續增加

流程		
	建物影像 訓練室內定位模型 室內定位模型	建物模型 影像擷取 模型影像 GauGAN的已訓練模型 偽真實影像 訓練室內定位模型 室內定位模型

表三、室內定位比較

測試時，規劃可拍攝實驗區域內環境影像，並且記錄其定位位置，將影像輸入已訓練好的室內定位模型，使模型輸出預測之定位位置，並與確定為正確答案之定位位置進行比較，進而得到其精度及正確率。

因為實驗場域為陽光充裕場所，因此較容易獲得高品質之真實影像。而本研究因為是使用轉換後的偽真實影像作為訓練資料，真實影像模型則是使用真實影像作為訓練資料，在真實影像模型可獲得高品質影像作為訓練資料的情況下，實驗結果預期上推測會較真實影像模型之成果精度差，然而因本研究為了解決使用真實影像作為室內定位影像來源之困境，因此以 95%信心水準為容許範圍，即為正負兩標準差內為容許範圍。若成果精度差異於兩標準差內，則可以認定本研究方法為可行之室內定位方式。

(五) 預期結果

本研究將探討室內定位之資料影像改進，以解決目前室內定位訓練之困境，預期成果如下：

1. 建物模型影像經GauGAN轉換為偽真實影像後，能與真實影像相近，並保留其特徵部分
2. 探討訓練資料集改為偽真實影像後，能否如一般室內定位模型一般，正確推知其定位位置

(六) 參考文獻

陳麒仁，”應用於物聯網具備臉部方向偵測之公分等級室內定位技術”，碩士

論文，資訊工程學系，逢甲大學，2017.

張峻瑋，”基於全景圖特徵之室內定位設計與實現”，碩士論文，通訊工程研究所，國立中正大學，2018.

黃星壬，”卷積神經網路於識別犬類品種之分析”，碩士論文，統計學研究所，國立中興大學，2020.

Jing-Mei Ciou, Eric Hsueh-Chan Lu ，”INDOOR POSITIONING USING CONVOLUTION NEURAL NETWORK TO REGRESS CAMERA POSE”，ISPRS Geospatial Week(GSW)，2019

Alex Kendall, Matthew Grimes, Roberto Cipolla，”PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization”，Proceedings of the IEEE International Conference on Computer Vision (ICCV)，pp. 2938-2946，2015.

Taesung Park, Ming-Yu Liu, Ting-Chun Wang, Jun-Yan Zhu，“ Semantic Image Synthesis with Spatially-Adaptive Normalization“，Accepted as a CVPR 2019 oral paper，2019.

Mohammed J. Zaki，”SPADE: An Efficient Algorithm for Mining Frequent Sequences”，Machine Learning 42, 31–60，2001.

(七) 需要指導教授指導內容

1. 室內定位之模型訓練操作。
2. 使用 GauGAN 於建物模型影像的轉換上之困境。
3. 室內定位之訓練模型成果分析。
4. GauGAN 之樣式圖樣利用深度學習方式進行判斷。
5. 整體歸納與建議 。