國立交通大學資訊工程學系

資訊專題競賽報告

NeRF應用與校園場景重建

Application of NeRF and campus scene rebuild

專題題目說明、價值與貢獻自評（限100字內）：

我們的專題目標是透過 NeRF 技術對校園中的場景進行重建，並希望透過這次實作來探討其在應用上的潛力與限制

專題隊員：

學號	姓名	手機	E-mail	負責項目說明	專題內貢獻度(%)
109550088	林哲安	0905418476	s125446133@gmail.com	模型訓練、資料集拍攝、成果報告編寫	35%
109550110	陳尚奇	0905253770	chin2839211@gmail.com	資料集拍攝、成果報告編寫	30%
109550175	許登豪	0978272178	za970120604@gmail.com	模型訓練、資料集拍攝、成果報告編寫	35%

本專題如有下列情況則請說明：

1.為累積之成果(含論文及專利)、2.有研究生參與提供成果、3.為大型研究之一部份。

相關研究生資料（無則免填）：

級別年級	姓名	提供之貢獻	專題內貢獻度(%)

【說明】上述二表格之專題內貢獻度累計需等於100%。

指導教授簡述及簡評：

本組嘗試用NeRF技術重建校園場景，從應用端的需求出發，搜尋探討並實際操作3D重建過程中所使用的各種論文技術。包含資料前處理、場景重建、場景的編輯（風格轉換）以及實時展示。本組組員非常積極，最終成功將光復校區的多處景點組合成一個完成度頗高的校園景點網站。

指導教授簽名：

中　華　民　國　一　一　二　年　五　月　二　十　八　日

專題摘要

一、關鍵詞

立體渲染(Volume Rendering)、視點合成(View Synthesis)、神經網路(Neural Network)、深度學習(Deep Learning)

二、專題研究動機與目的

Google街景是一個行之有年、廣受大眾熟悉的服務，但其街景照片時有扭曲的情形且觀賞角度受限而難以擬真地還原實地場景。比如上圖中的雕塑就只能從少數角度觀看。

我們發現近年興起的三維重建技術NeRF(Neural Radiance Field)能夠透過一定數量的照片，還原出逼真的場景，甚至渲染出資料集內未提供的角度。我們認為這項技術有相當大的發展空間，所以決定以此為題，在研究不同NeRF重建的同時，著手將校園中的場景製作成一系列擬真的校園街景，讓大家能夠以更多的角度領略交大風光。

三、專題重要貢獻

此專題專注於近幾年來備受關注的 NeRF 技術，這是目前相當熱門的研究主題，但相較之下其在應用層面上較乏人問津。我們決定以 NeRF 為基底構建一套完整的校園景觀導覽系統，並探討此技術目前在應用上的潛力與可能性。

四、團隊合作方式

我們採每週開一次週會的形式，在週會中報告當前進度並協調之後的工作分配。林哲安和許登豪兩人主導模型訓練，其餘諸如資料集拍攝、網站建構與報告撰寫的部分則是三人通力合作完成。

最後要感謝研究生鄭伯俞學長在專題過程中的指導，其協助監督我們的進度並為我們的專題提供相當多有幫助的建議。

五、研究歷程統整與技術分析

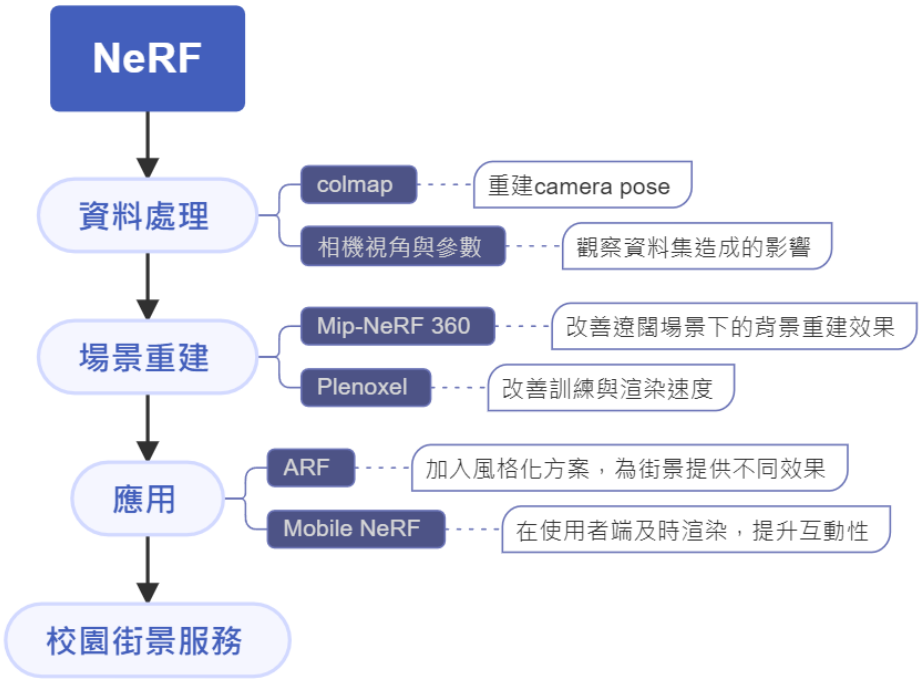
➤ 研究歷程簡述

最一開始，我們使用原版 NeRF 成功以自行收集的資料集還原出複雜的三維物件。然而 NeRF 對場景中背景部分的處理能力較差，對於校園街景重建這個目標而言是相當重大的缺陷。因此，我們開始研究各種延伸論文。

我們將實現校園街景的過程區分成三大步驟。首先為了能夠還原場景，必須先對輸入的圖片資料集進行處理，提取出拍攝時的相機位置等參數。相機參數的正確性與資料集完整性會顯著地影響最終的重建結果。

在場景重建上，NeRF 作為這次專題的基礎技術，我們會先簡單講解它。而後針對重建校園場景這個目標，我們挑選了兩個特化不同特性的架構，展現並比對其各自的重建結果與優缺點。

如今我們已具備還原校園的能力，最後便是擴展應用層面。風格轉換與即時渲染都是很有意義的嘗試，為我們的成果增添了獨特性與互動性。接下來的段落中我們會講解各步驟如何實現，過程中的嘗試與發現及最終達成的結果。

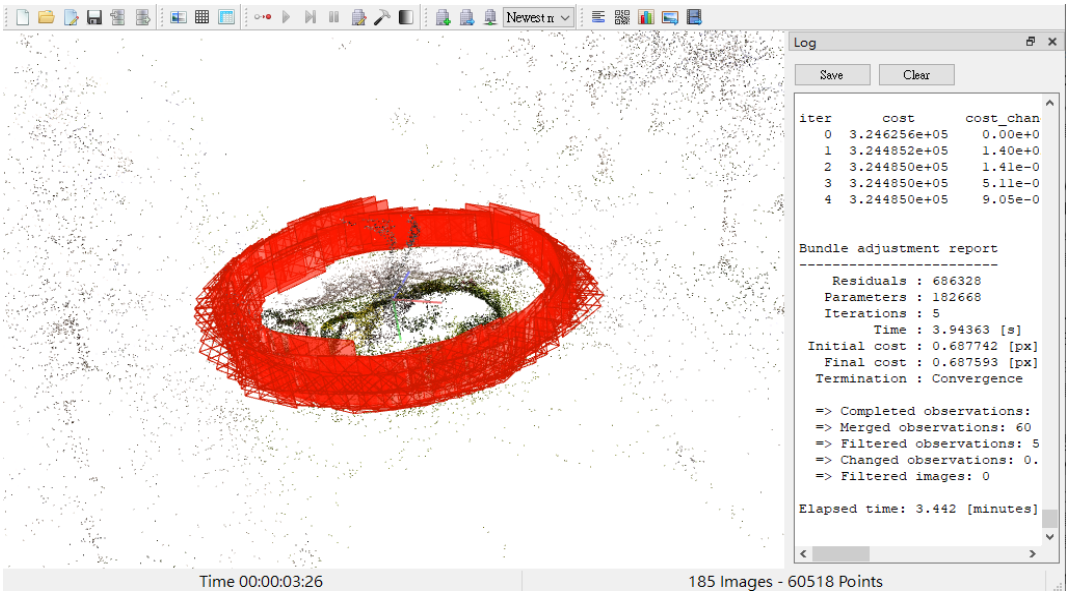


➤ 資料蒐集與前處理

a. 相機參數重建: Colmap

Colmap是一款常見的開源三維重建軟體，它可以利用多個視角圖片中提取到的特徵點及相機參數來進行三維場景的重建。

而在NeRF的流程中，我們主要利用Colmap來提取出資料集中的相機內外參數，所以只需進行到稀疏重建，再利用生成的camera pose作為NeRF網路的輸入即可。



Colmap流程中，先從每張二維影像中提取特徵點，再對數據集中的圖像兩兩進行特徵點的比對。利用不同角度下成功匹配的點位便可進行三維空間中稀疏點雲的重建，最終得出每張圖片的camera pose。

b.資料集拍攝: 相機視角與技巧

相機參數的正確性對於三維重建相當關鍵，一個好的資料集是重現場景的第一步。

➢相機參數的一致性

在試圖對圖書館的鳳凰來儀雕像進行重建時，我們發現利用 Plenoxel 訓練之結果除了空間中的漂浮物以外，物件主體輪廓的周圍也存在如殘影一般的模糊區塊。

多次嘗試後發現是手機相機自動調整焦距所致，在將拍攝的參數固定後重新訓練便成功解決了該問題，能重建出邊緣更為銳利清晰的物件。

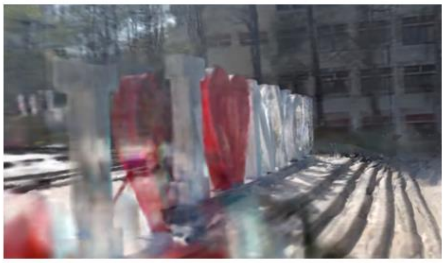
未固定相機焦距

固定相機焦距

➢相機視角的完整性

而在NYCU立牌的重建上則是遇到了環景架構下，正背面的撕裂問題，由於此場景前後存在高低差，且在拍攝時側面視角不足以讓 Colmap 還原出走下階梯時相機確切的相對高度，所以導致重建結果存在正背面形體不連貫的問題。
(如下圖紅色虛線標註之NY部分，形成半透明之漂浮物)

正面


側面


背面

➤ 三維場景重建

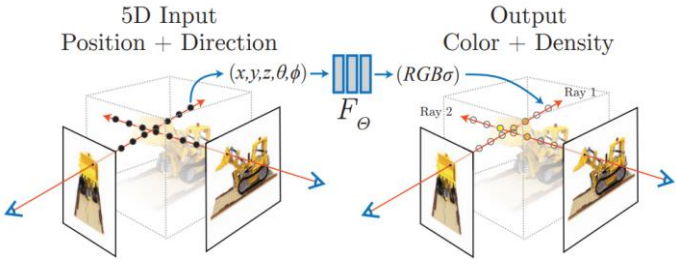
為解決NeRF在背景上的渲染問題，我們在 Mip-NeRF360 上做了許多嘗試。雖成功解決背景細節不足的問題，但由於其受限於MLP架構，渲染時間較長。考量到訓練與渲染所需的時間成本，我們轉而研究速度較快的 voxel based Radiance Field。

a. 三維重建: NeRF

NeRF 是一項新興且熱門的三維重建方法。利用多張不同視角圖片來訓練MLP，並以Volume rendering 建構出三維場景中各點的顏色值與不透明度，最終生成出效果令人驚豔的連續、不存在於原始照片視角的二維畫面。

➢ 核心技術分析

NeRF 將5D coordinate(spatial location(xyz) + viewing direction(θ, φ))作為輸入，表示入射角度位置與採樣點，輸出一組色彩c = (R, G, B)和密度σ，以Volume rendering將這些顏色和不透明度的資訊加總，生成2D圖像。

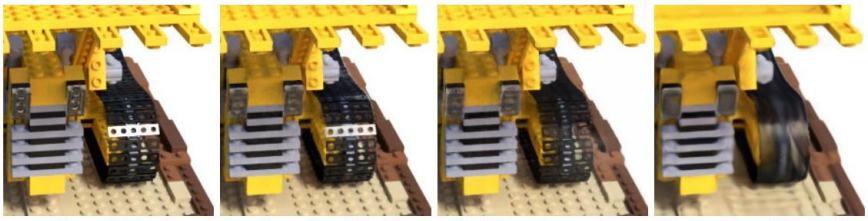


NeRF能夠展現優異成果的關鍵在於其中使用的兩個突破性技巧:

(i). Positional Encoding:

根據先前的研究顯示，直接將位置訊息輸入深度網路時，網路會傾向於學習低頻率的部分，難以fit高頻率部分。而Positional encoding是將輸入利用正餘弦投射到高維空間幫助MLP近似高頻率的函數，能夠成功得到平滑許多的結果。

$$\gamma(p) = \left(\sin\left(2^0\pi p\right), \cos\left(2^0\pi p\right), \cdots, \sin\left(2^{L-1}\pi p\right), \cos\left(2^{L-1}\pi p\right)\right)$$



Ground Truth　Complete Model　No View Dependence　No Positional Encoding

(ii). **Hierarchical volume sampling:**

rendering過程中對空間進行取樣時，大部分的空間中其實是沒有物體的，因此平均取樣的效果並不理想。Hierarchical volume sampling 的想法是訓練兩個架構相同的模型，先用平均取樣的方式訓練其中一個模型，再根據其輸出的密度分布來重新分配取樣點，用以訓練另一個模型，其即為我們的最終結果。

➢ 實驗或成果展示

我們使用思園的交大校徽做為第一個測試 dataset，在 Forward facing 的架構下，僅利用約30張照片進行訓練便取得了不錯的重建成果，這增加了我們對於利用 NeRF 技術來實現街景服務這個構想的信心。

不過NeRF並未對背景進行特別處理，僅適合目標場景上下界可控的情況，諸如無背景的單一物體或小型非開闊場景。為了擴增校園街景服務能提供的場景，我們開始著手進行後續論文的延伸研究。

b. 開闊環境下的背景優化: Mip-NeRF 360

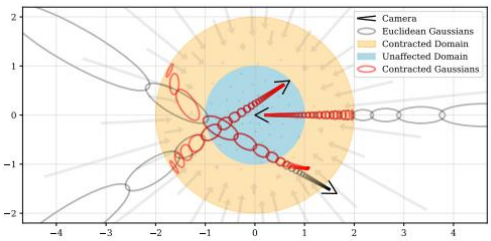
校園中存在許多的戶外場景，為此必須尋找一個能更大程度還原出背景的方法。Mip-NeRF 360 則是此領域中一篇著名的延伸論文，建立在Mip-NeRF的基礎上，除了利用圓錐狀的光線來改善鋸齒問題外，更利用了函數映射來解決場景大小受限的問題。

➢ 核心技術分析

(i). **扭曲函數**:

利用函數將整個場景分割為兩個同心球體，近景保留在小球體中，並將遠景壓縮映射到一個比較大的球體內，解決了原本方法在無界場景的撕裂問題。

$$\text{contract}(\mathbf{x}) = \begin{cases} \mathbf{x} & \|\mathbf{x}\| \leq 1 \\ \left(2 - \frac{1}{\|\mathbf{x}\|}\right)\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) & \|\mathbf{x}\| > 1 \end{cases} \quad (10)$$



(ii).Hierarchical volume samplinge 改良:

額外訓練一個僅輸出密度的 Proposal MLP，計算出密度分布後再進行正常的採樣來訓練輸出色彩的MLP。如此一來既保持低解析度下的效率，也能兼顧高解析度時對於細節與紋理的需求。

(iii). Model Regularization:

進行重建時，時常會在無物件的位置出現飄浮物，由於應該屬於背景的顏色出現在了不正確的位置，連帶導致了背景的缺失。為了解決這樣的問題，Mip-NeRF 360 讓模型的權重盡量地集中，有效避免背景懸浮物的形成。

➢ 實驗或成果展示

以下是 Mip-NeRF 360 與 Plenoxels 在同一個資料集下的渲染結果比較圖。由於當時拍攝過程中相機參數未能統一，結果在Plenoxels 還原出的物體邊緣較為模糊且地面結構出現缺失。兩相比對下， Mip-NeRF 360 在重建上並未受到影響，效果較佳，可見MLP架構對於相機參數誤差的適應性較為優秀。

Mip-NeRF360

PSNR: 28.1

Plenoxels

PSNR: 18.5

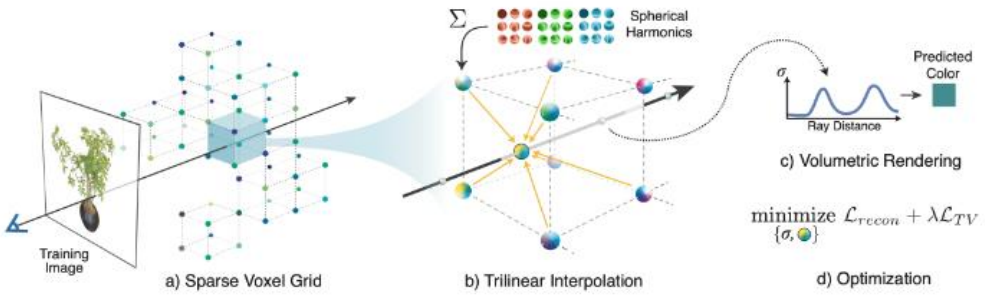
c. 加速訓練與渲染: Plenoxels

在專題進程中段時，實驗室釋出了一些規格更高，VRAM更大的顯卡，使得我們有能力嘗試另一種 NeRF 方案 : 相比以 MLP 表示一個場景，利用 voxel grid 進行三維重建可以更快訓練出模型並加速render流程，代價則是訓練過程使用的VRAM及儲存模型所需的空間遠比前者需求更高。

➢ 核心技術分析

(i). **Voxel Grid :**

摒棄了原始 NeRF 所採用的 MLP，改用 voxel grid 進行場景建模。每個 voxel 各點中儲存不透明度及球諧函數，對於內部任意點的不透明度及顏色則利用對其三線性內差取得，球諧函數則是用來處理不同視角下造成的顏色改變。



➢ 實驗或成果展示

雖說在先前的實驗中 Plenoxel 的表現不如 Mip-NeRF360，但我們必須將訓練與渲染所需的時間成本納入考量。Mip-NeRF360 的訓練需要將近一天且渲染相當費時，而 Plenoxels 則可在一小時以內完成。此外，在我們能穩定控制相機參數之後，也可以在 plenoxels 架構上訓練出效果優異的環景。

➤ 系統應用與成果

解決重建的問題後，我們想再更加拓展這個專題在應用面的發展潛力。首先，我們嘗試結合 ARF，將圖像風格化融合進街景服務中，讓使用者能看見不同風貌的交大校園，提供一些獨特且新奇有趣的體驗。此外，為了避免一般民眾在使用時被硬體所限制，我們藉由 mobileNeRF 直接在網頁上進行渲染，讓使用者可以透過普遍的裝置瀏覽我們的成果。

a. 風格移植: ARF

風格轉換是各式影像領域的熱門研究項目，若將其與 NeRF 結合便能改變重建出的現實場景，為同樣的景物帶來不同的樣貌。在三維重建上亦能對擴增實境或是電腦特效領域有所貢獻。

➢ 核心技術分析

(i). **Nearest Neighbor Feature Matching :**

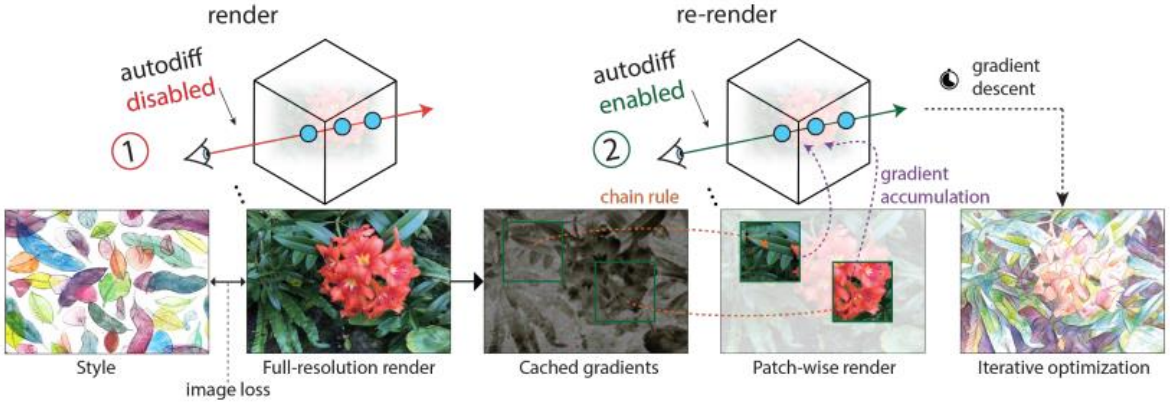
為確保產出的圖片風格符合指定的目標，我們需要計算生成圖的 style loss。 A

RF 其中一項關鍵在於不使用常用的 Gram-matrix-based 方式計算 style loss ，而提出一種新穎算法 NNFM。相較於 Gram-matrix-based loss ，NNFM 的計算範圍更局部，藉此留住更區域性的特徵，在主觀視覺上讓人覺得更貼近原圖。

(ii). deferred back-propagation :

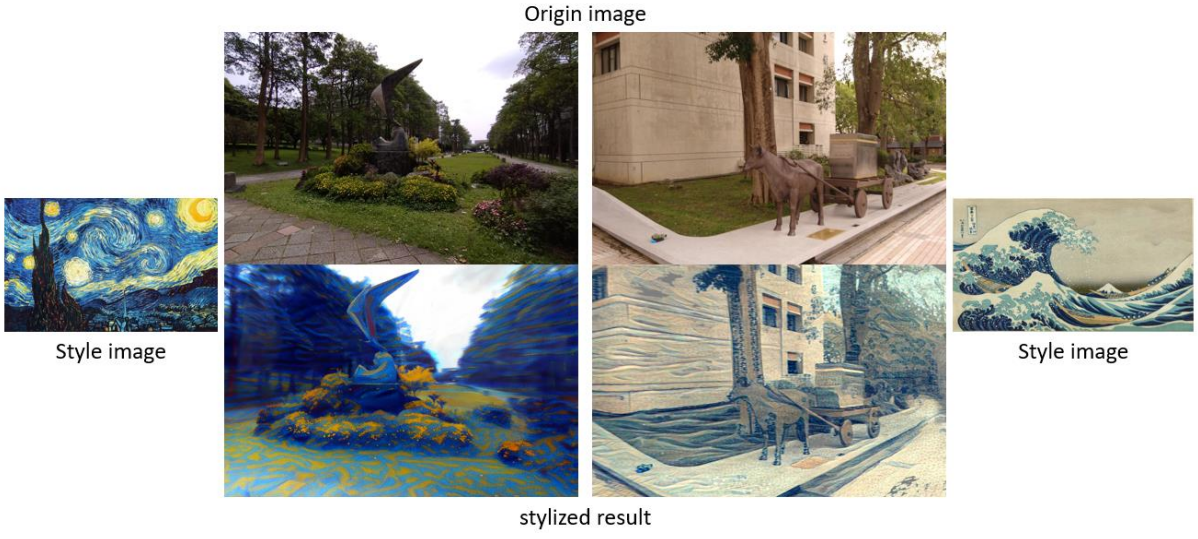
進行渲染時每個像素所需的計算量龐大，optimization 階段相當考驗記憶體的負荷能力。NNFM 需要渲染出完整的圖片來才能計算loss。

為此必須提高利用記憶體的效率，而 deferred back-propagation 可以將原本對整張圖片進行的 back-propagation 範圍縮小到一個 patch的大小 ，減輕對記憶體空間的需求。



➢ 實驗或成果展示

ARF能快速地轉換場景，且產出在主觀的視覺認知上更貼近我們指定的風格，且兼容於plenoxel架構，因此在專題中選擇其作為我們的首要風格化方案。



b. 使用者端的即時渲染: MobileNeRF

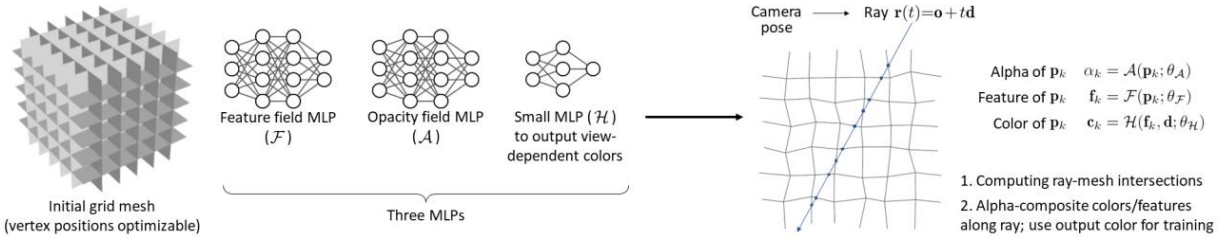
在此時我們注意到一個問題，該如何為使用者呈現我們的場景？訓練好的模型檔案龐大，且坊間普遍的硬體設備計算能力有限，就算提供模型也很可能渲染不出圖片，若僅提供渲染好的影片則顯得與使用者互動性不足。

於是我們將重心轉換到了即時渲染的領域，該如何在更普及的裝置上展現這些場景，並盡量保持畫面的清晰? 而 MobileNeRF 提供了一個合適的做法。

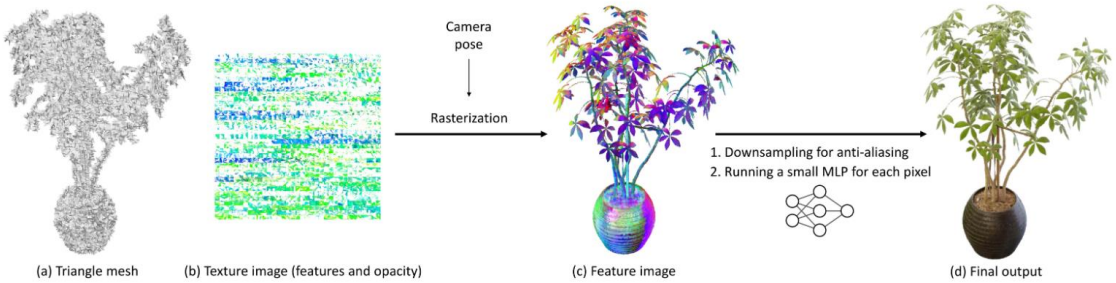
➢ 核心技術分析

(i). **2D-textured triangle mesh:**

MobileNeRF不再使用傳統MLP方式表示一個場景，而改以trinagle mesh來呈現。mesh中各點的特徵、密度以及顏色則藉由三個小型MLP來存儲。將每個網格中儲存的頂點資訊連結成立體表面，便能藉由 differentiable rendering 計算射線與網格交點，再從MLP取值以渲染畫面。

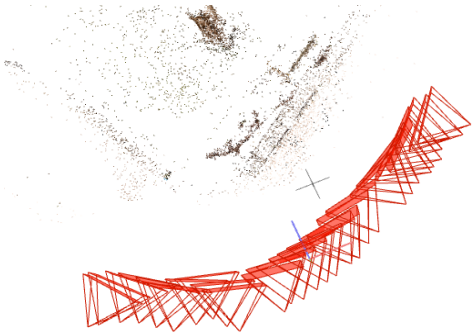


其中不同視角下的顏色差異則是透過存在 2D texture上的8維向量與 fragment shader 中的一個小型 MLP 轉換的。



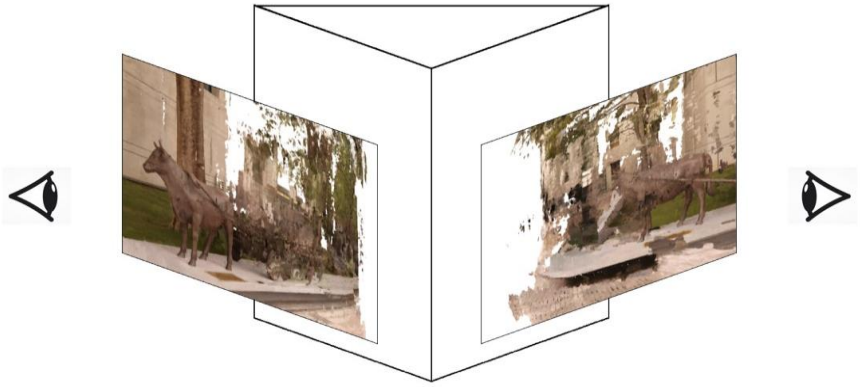
➢ 實驗或成果展示

工程三館前的牛車雕像也被我們納入了所展示的景觀之中，在訓練過程中其帶給我們一些新的發現。下兩張圖是初次重建的結果與對應的 colmap 重建圖，很明顯地可看出場景重建效果不佳，雕像正前方存在多條縱向懸浮物。

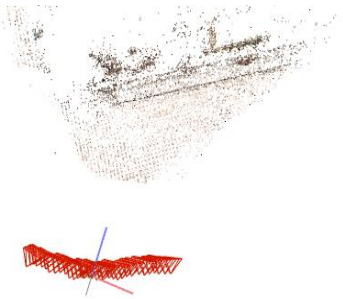


經研究討論後發現，由於我們採用的 forwarding facing model能還原的景象角度有限，而拍攝時所涵蓋的角度太廣，模型為了盡可能的還原兩側景象，導致學習出了錯誤的結構，進而出現懸浮物。

這些懸浮物在空間中呈現近似於三角型的分布，因此從正前方觀看時，視線會被懸浮物遮擋使得圖像重建效果不佳。



為了修正此現象，我們從原本的資料集裡面重新挑選，縮減了資料集中的相片角度，以此新資料集進行訓練，最終成功修正了懸浮物的問題。



MobileNeRF 降低了所需的裝置需求，令我們的成果能在較普遍的裝置上進行即時渲染，讓更多人更方便的看到我們製作的校園場景，並且擁有更多的使用者互動性，我們決定將此項技術融入校園導覽服務之中。

c.校園導覽服務網頁: NeRF in NYCU

最終考量到使用上的方便性，我們決定以網頁的形式提供服務，此段落中將概述我們網頁所具備的功能與設計目的。

(i). 校園景物地圖:

為了讓觀看的使用者能更清楚的了解我們所重建的每個景點在校園內的位置，我們在網站上設置了一張光復校區的地圖，並於其上布置數個定位點，使用者可以透過預覽窗看到景像，或點擊跳轉到有興趣的景點展示頁面。

(ii). 街景展示頁面:

在選擇了展示的地點後，我們使用mobile NeRF來即時渲染呈現出三維重建的成果，使用者可利用滑鼠任意移動或轉動視角，以直觀的操作在範圍內查看不同角度下的場景。

此外，除了實際場景外，我們也利用ARF來添加了不同風格，使用者可以左右轉動視角來在如同畫中世界的場景中移動，一睹熟悉的校園所嶄露的另一面風光。

FPS: 143.3

網頁網址: http://140.113.24.102:8000/
http://140.113.24.205:8080/

(如瀏覽網頁時遇到問題，請洽詢以下email : chin2839211@gmail.com)

六、結論

在經歷了上述研究，並最終完成了目標: 校園街景服務後，我們了解到了 NeRF 技術現存的限制，並見識其他研究團隊如何將其克服，確實此技術具有龐大的發展潛力，但若要將其推入商業領域目前仍有一些障礙有待解決。

我們認為最主要的難題就是費時的訓練與渲染，就現今普及的的圖形設備而言，想要重建大型場景仍是一項不小的工程，擴增實境或遊戲等應用則更需要即時演算的能力。但在眼下諸如 Instant NGP 等嶄新優秀架構的逐步推展下，或許在不久的將來就能翻過硬體限制的高牆，克服 NeRF 的算力需求。

七、參考文獻

[1] Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." *Communications of the ACM* 65.1 (2021): 99-106.

[2] Barron, Jonathan T., et al. "Mip-nerf 360: Unbounded anti-aliased neural radiance fields." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[3] Fridovich-Keil, Sara, et al. "Plenoxels: Radiance Fields without Neural Networks." 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022.

[4] Zhang, Kai, et al. "Arf: Artistic radiance fields." Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXI. Cham: Springer Nature Switzerland, 2022.

[5] Chen, Zhiqin, et al. "Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[6] Müller, Thomas, et al. "Instant neural graphics primitives with a multiresolution hash encoding." *ACM Transactions on Graphics (ToG)* 41.4 (2022): 1-15.