

DLCV-Hw4 Report

R11942180 電信二 黃湛元

Problem: 3D Novel View Synthesis

1. (15%) Please explain:

a. the NeRF idea in your own words

NeRF利用深度學習建立一個光輻射場（radiance field）的表示，能夠對三維空間中光線的顏色和密度進行建模。它透過神經網絡學習空間中每一點的顏色和密度，並結合光線追蹤技術來渲染新的視角圖像。這種方法能夠使用少量的2D圖像來重建出高質量的3D場景。

b. which part of NeRF do you think is the most important

NeRF的核心在於其對於連續空間的建模能力。通過學習一個從空間座標和觀察方向到顏色和密度的映射，NeRF 能夠精確地捕捉到光線與物體表面的交互，產生高質量的新視角圖像。這種對於空間的深度理解是其生成逼真圖像的關鍵。

c. compare NeRF's pros/cons w.r.t. other novel view synthesis work

i. **NeRF** :

- **優點**：原始 NeRF 透過神經網絡建立從 2D 圖像到 3D 場景的映射，能夠產生高質量的新視角圖像。
- **缺點**：計算成本高，訓練和渲染時間長。
- **方法**：利用 MLP 和體積渲染技術來學習場景的輻射場。

ii. **"Direct Voxel Grid Optimization"** :

- **優點**：相比於原始 NeRF，此方法在輻射場重建方面具有更快的收斂速度。
- **缺點**：在處理極其複雜的場景時，表現不如基於 MLP 的方法。

- **方法**：直接優化一個可微分的 voxel grid，而非像 NeRF 那樣透過 MLP 來學習。

iii. "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding"：

- **優點**：提供即時的神經圖形基元訓練，即便在複雜的應用中也能快速收斂。
- **缺點**：可能在保持細節的精確度方面略遜於原始 NeRF。
- **方法**：使用 **Multiresolution Hash Encoding** 來增強神經網絡，允許使用較小的 network 而不犧牲品質。

2. (15%) Describe the implementation details of **your NeRF model** for the given dataset. You need to explain your ideas completely.

我沿用了原始 NeRF 的框架，並沒有對模型架構進行重大修改。只調整了層數量和 embedding 維度，改變模型的性能和對細節的捕捉能力。

1. **Embedding**:

- 使用正弦和餘弦函數對位置 (xyz) 和方向向量進行 encode，這有助於模型學習更複雜的幾何形狀和照明效果。
- **Embedding** 將每個輸入通道 (xyz 和方向) 映射到一個高維空間，提高了模型表達不同頻率特徵的能力。

2. **NeRF 模型**:

- 使用 Fully-connected network (D 層，每層 W 個神經元) 來對場景進行建模。
- 模型分為兩部分：用於預測密度 (sigma) 的部分和用於預測顏色的部分。這兩部分共享前幾層網絡，使得顏色預測可以利用空間信息。

3. **Forward 方法**:

- 在 **forward** 方法中，根據輸入的光線對 NeRF 模型進行推理。這涉及到對每個光線的位置和方向進行編碼，然後將它們輸入模型以預測顏色和密度。

3. (15%) Given novel view camera pose from **metadata.json**, your model should render novel view images. Please evaluate your generated images and ground truth images with the following three metrics (mentioned in the NeRF paper). Try to use at least **three** different hyperparameter settings and discuss/analyze the results.

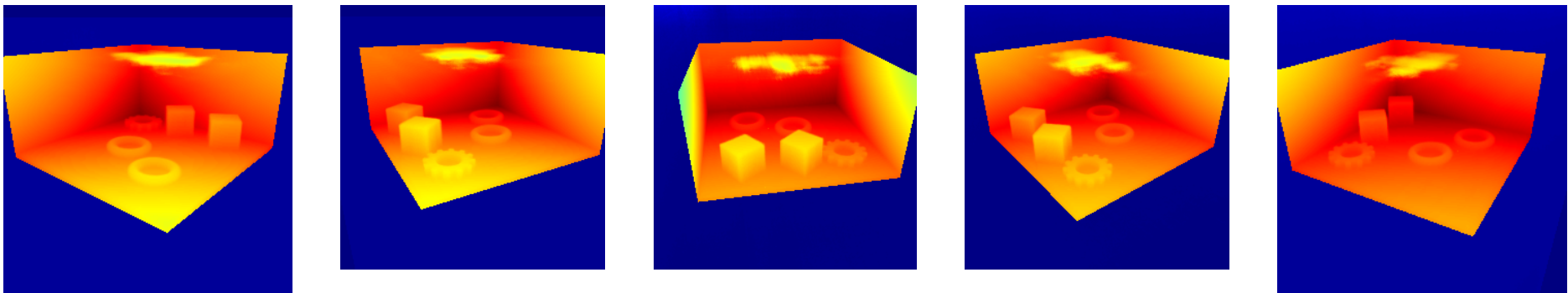
- **Please report the PSNR/SSIM/LPIPS on the validation set.**

- You also need to explain the meaning of these metrics.
- Different configuration settings such as MLP and embedding size, etc.

	Embed. xyz	Embed. dir	Hidden units	Epoch	N_importance	Optimizer	Learning rate
Set A	10	4	256	16	64	Adam	5e-4
Set B	15	6	256	16	64	Adam	5e-4
Set C	15	6	384	16	64	Adam	5e-4

	PSNR	SSIM	LPIPS
Set A	44.22	0.99	0.00
Set B	43.79	0.99	0.00
Set C	44.42	0.99	0.00

4. (15%) With your trained NeRF, please implement depth rendering in your own way and visualize your results.



Reference

1. https://github.com/kwea123/nerf_pl
2. <https://arxiv.org/abs/2003.08934>
3. <https://arxiv.org/pdf/2111.11215.pdf>
4. <https://nvlabs.github.io/instant-ngp/assets/mueller2022instant.pdf>
5. ChatGPT