Problem1.

1.

a. 藉由 MLP $F_\theta: (x, d) \rightarrow (c, \sigma)$，去預測給定位置 x 與觀測角度 d 時，其對應顏色與 volume density 的數值，最後透過 volume rendering 重建場景

b. 透過預測(c, σ)來使用 volume rendering 重建場景

c. pros : 渲染品質非常好, cons : 需要大量的訓練時間，且在渲染時也沒有效率，無法從事 real time 任務

2.

Training 分成 coarse 與 fine stage，並且使用 post-activated density voxel grid : $inter(x, V)$去加速取得場景中的 3D 結構

(a) Coarse geometry searching : 首先建立出包含整個場景的 bounding box，接著就可以針對每個 bbox 中的 voxel grid, $V^{(density)(c)}, V^{(rgb)(c)}$使用 post-activated 的方法得到 volume density 與 color，最後即可透過 volume rendering 重建，並與 ground truth 計算L_2 loss。為了確保在訓練初期，沿著光線方向的所有取樣點都不被遮蔽，因此作者將$V^{(density)(c)}$內所有的值均初始化成 0，並在 density activation 加上 bias，使得 accumulated transmittance 以 1/per voxel size 遞減，此外 voxel grid 中每一點的 learning rate 會以自身可視的格子點數做調整。

(b) Fine detail reconstruction : 首先利用$V^{(density)(c)}$區分 known free space/unknown space 並找出 fine stage 的 bbox，針對每個 bbox 中更高解析度的 voxel grid $V^{(density)(f)}$，透過 post-activated 得到 volume density，color 則是經過 MLP 得到，且內部加入 positional encoding，另外在 query 時，會跳過 known free space 或者地於 threshold 的點來加速，最後計算L_2 loss

3.

PSNR: 衡量訊號最大可能功率和影響它表示精度的破壞性雜訊功率的比值

SSIM: 衡量圖片間的結構相似程度

LPIPS: 透過神經網路的提取特徵，衡量圖片間的感知相似程度

Setting	PSNR	SSIM	LPIPS
Default	35.176	0.974	0.023
Step size : 0.5 -> 0.1	35.282	0.975	0.021
Number of voxel : 1024000->136^3	35.153	0.974	0.022

增加 sample 頻率以及 voxel 的密度有助於表現提升，但是 training/inference 時間也會些微變長。

Problem 2.

1.

使用 BYOL 作為 SSL 的訓練 backbone 的方法，data augmentation 包含隨機的顏色增強，水平翻轉，轉灰階，加入 Gaussian Noise。

	setting
Batch size	512
optimizer	Adam, lr= 10^{-3}, weight decay= 1.5×10^{-6}
scheduler	Cosine Anneling
Training epochs	1524

2.

Setting	Pre-training (Mini-ImageNet)	Fine-tuning (Office-Home dataset)	Validation accuracy (Office-Home dataset)
A	-	Train full model (backbone + classifier)	31.28%
B	w/ label (TAs have provided this backbone)	Train full model (backbone + classifier)	35.71%
C	w/o label (Your SSL pre-trained backbone)	Train full model (backbone + classifier)	54.19%
D	w/ label (TAs have provided this backbone)	Fix the backbone. Train classifier only	28.33%
E	w/o label (Your SSL pre-trained backbone)	Fix the backbone. Train classifier only	37.68%

使用 Byol backbone 且對整個 model finetune 的 setting C，有非常好的表現，而只 train classifier 的 setting E 則位居第二，綜合反映出 Byol 的可行性。剩下比較值得注意的是 setting D，他的表現竟然不如 setting A，可能的原因是 backbone 內的大量參數無法被調整，且它 supervise learning 在過小的訓練集(Mini-image

Net)上，導致轉移到 office home 訓練集上就出現了 overfitting，這點換做在 setting B 上就有改善。