HW4

☑ Done	☐
🗓 Due date	@December 22, 2022
⊙ Subject	DLCV

Problem 1:

1. Please explain:

(a) **The NeRF idea in your own words**

NeRF 會把物體跟場景的資訊 encode 後丟進 MLP，再把 3D 影像投影成 2D 圖片。

(b) **Which part of NeRF do you think is the most important**

我覺得是 **position encoding** 的部份，透過這資訊能讓模型對位置有更好的理解。

(c) **Compare NeRF's pros/cons w.r.t. other novel view synthesis work**

NeRF 可以繪製出精細的結果，結構也比較完整，SRN 在細節方面就容易產生模糊，而 NV 常在結構表現上失真。

2. Describe the implementation details of Direct Voxel Grid Optimization(DVGO) for the given dataset.

DVGO is an optimization method that is used to find the minimum value of a given function using a voxel grid representation of the search space.

3. Please evaluate your generated images and ground truth images with the following three metrics

(a) **Please report the PSNR/SSIM/LPIPS on the validation set.**

PSNR: 35.185 SSIM: 0.974 LPIPS: 0.419

(b) **You also need to explain the meaning of these metrics.**

- PSNR

  Peak signal-to-noise ratio 為訊號最大可能功率和影響它的表示精度的破壞性雜訊功率的比值。PSNR的單位為dB，其值越大，圖片失真越少。一般来说，PSNR高於40dB说明圖片畫質已經和原圖差不多。

- SSIM

  Structural similarity index, 是一種用於量化圖片間的結構相似性指標。SSIM仿照人類的視覺系統，從亮度、對比度以及結構量化圖片的資訊。SSIM值的範圍為0至1，越大代表圖像越相似。

- LPIPS

  Learned Perceptual Image Patch Similarity, 是學習生成的含noise圖像到原本ground truth的reverse mapping，強制generator學習 fake image重新建構ground truth的reverse mapping。在比較上，LPIPS又比前兩項指標更貼近人類的感知情況。值越低也表示圖像相似度越高。

(c) **Different configuration settings such as iteration number/number of voxel/stepsize … lead to different performance.**

Setting	PSNR	SSIM	LPIPS
Setting 1	35.161	0.971	0.421
Setting 2	35.091	0.968	0.425

Setting 1 把 stepsize調高 2 倍

Setting 2 為調高 *num_voxels*

Problem 2:

1. Describe the implementation details of your SSL method for pre-training the ResNet50 backbone.

在這題我用了 BYOL 來做 model pretrain，lr = 3e-4，optimizer = adam, batch_size = 256，epoch 一開使用100，但效果不理想所以調整為 500

2. Please complete the following Table, which contains different image classification setting, and discuss/analyze the results.

Setting	Pre-training (Mini-ImageNet)	Fine-tuning (Office-Home dataset)	Validation accuracy (Office-Home dataset)
A	-	Train full model (backbone + classifier)	**25.6%**
B	w/ label (TAs have provided this backbone)	Train full model (backbone + classifier)	**37.1%**
C	w/o label (Your SSL pre-trained	Train full model (backbone +	**41.87%**

	backbone)	classifier)	
D	w/ label (TAs have provided this backbone)	Fix the backbone. Train classifier only	**28.07%**
E	w/o label (Your SSL pre-trained backbone)	Fix the backbone. Train classifier only	**26.84%**