1.1
a.透過已知視角的目標物影像預測出未看過的相同目標物體的視角，主要流程是透過輸入射線射到的座標(x,y,z)視角(θ,φ),預測出該點的rgb
b.最主要的是透過相機射線打到物體，並將射線積分起來得到影像顏色，這是最重要的核心，而其他包含position encoding、Hierarchical Volume Sampling等技術大幅提昇效果。
c.nerf相較於dvgo在效果及速度上都輸，尤其在訓練時間上nerf所花的時間為dvgo的80倍以上，而nerf的優點就是模型小，參數較少。
1.2
DVGO使用voxel grid來表示每個scene，並且跟nerf一樣透過射線去預測每個pixel的rgb以及density，並且利用coarse to fine的方法，類似於先了解大概的資訊，再由粗略資訊取得細微資訊，這樣可以加速運算時間以及略過不重要的區域。
1.3
由set1&set3可看出fine train的iteration次數差距四倍但分數上只有些微差距，但是從set1&set2發現，coarse_model_and_render 的num_voxels＆num_voxels_bas差距四倍，在分數上有較大一點的差異。最後在set1&set4比較上，將num_voxels＆num_voxels_base 增加反而在分數上有較差的表現。

Setting	PSNR	SSIM	LPIPS
Setting 1	psnr 35.28490476608276	ssim 0.9747688857963087	lpips (alex) 0.020779788363724946
Setting 2	psnr 35.12535543441773	ssim 0.9743199701940051	lpips (alex) 0.02217728827148676
Setting 3	psnr 35.132656288146975	ssim 0.9743907789647978	lpips (alex) 0.021693064272403716
Setting 4	psnr 35.19235076904297	ssim 0.9748852065469513	lpips (alex) 0.020160079672932624

set1:
coarse_train :N_iters=40000,
fine_train:N_iters=80000,
coarse_model_and_render :num_voxels=4024000,num_voxels_base=4024000,
fine_model_and_render = num_voxels=4096000,num_voxels_base=4096000,
set2:
coarse_train :N_iters=60000,
fine_train:N_iters=40000,
coarse_model_and_render :num_voxels=1024000,num_voxels_base=1024000,
fine_model_and_render = num_voxels=4096000,num_voxels_base=4096000,
set3:
coarse_train :N_iters=40000,
fine_train:N_iters=20000,
coarse_model_and_render :num_voxels=4024000,num_voxels_base=4024000,

fine_model_and_render = num_voxels=4096000,num_voxels_base=4096000,
set4:
coarse_train :N_iters=40000,
fine_train:N_iters=80000,
coarse_model_and_render :num_voxels=4024000,num_voxels_base=4024000,
fine_model_and_render = num_voxels=5832000,num_voxels_base=5832000,

2.1
使用BYOL作為SSL model，image preprocess包含resize to 128*128,center crop,normalize，batch size=90，learning rate=3e-4,optimizer=Adam,train epoch=100。在 BYOL裡對影像做的處理包含random colorjitter,random grayscale,random horizontal filp,random gaussian blur…

2.2
從結果來看，使用我pretrain的backbone進行fine tune，不論是fix backbone或train full model都比其他三者好，理由應該是我backbone 訓練的比較好。
在相同backbone進行fine tune時，train full model都比fix backbone的準確度高。
而在沒有pretrain backbone的情況下直接訓練office-home dataset的準確度是最低的，主要原因應該是資料量太少。

Setting	**Pre-training (Mini-ImageNet)**	**Fine-tuning (Office-Home dataset)**	**Validation accuracy (Office-Home dataset)**
A	-	Train full model (backbone + classifier)	grade: 0.2481572481572481
B	w/ label (TAs have provided this backbone)	Train full model (backbone + classifier)	grade: 0.3488943488943489
C	w/o label (Your SSL pre-trained backbone)	Train full model (backbone + classifier)	grade: 0.4152334152334152
D	w/ label (TAs have provided this backbone)	Fix the backbone. Train classifier only	grade: 0.2874692874692874
E	w/o label (Your SSL pre-trained backbone)	Fix the backbone. Train classifier only	grade: 0.3832923832923833