GNeRF: GAN-based Neural Radiance Field without Posed Camera

Quan Meng[1] Anpei Chen[1] Haimin Luo[1] Minye Wu[1]
Hao Su[2] Lan Xu[1] Xuming He[1] Jingyi Yu[1]
[1] Shanghai Engineering Research Center of Intelligent Vision and Imaging
School of Information Science and Technology,
ShanghaiTech University [2] University of California, San Diego
{mengquan,chenap,luohm,wumy,xulan1,hexm,yujingyi}@shanghaitech.edu.cn {haosu}@eng.ucsd.edu

A. Pose Distribution Analysis

We analyze mismatch of pose distribution on the Synthetic-NeRF dataset in Tab. 1. We change the range of pose sampling space and report the novel view synthesis quality on the chair scene in the Synthetic-NeRF dataset. The pose sampling space is represented by camera position and camera rotation. The camera position is determined by three parameters: radius, elevation, and azimuth. The camera rotation is calculated from the camera position, camera lookat point, and camera up vector. As is shown in the table, changing the radius or lookat points does not depreciate the performance greatly. For the third pose distribution of changing the elevation range from $[0, 90]$ to $[-90, 90]$, we note that this setting will generate views that are not covered by any training image which will disturb the inversion network. It demonstrates that our method relies on a reasonable camera sampling space but not necessarily an accurate one.

Radius	Ele(deg)	Azi(deg)	Lookat	PSNR
4	$[0, 90]$	$[0, 360]$	(0, 0, 0)	31.30
$[3, 5]$	$[0, 90]$	$[0, 360]$	(0, 0, 0)	29.68
4	$[-90, 90]$	$[0, 360]$	(0, 0, 0)	23.08
4	$[0, 90]$	$[0, 360]$	$\mathcal{N}(0, 0.01^2)$	30.48

Table 1. **Pose Distribution Analysis.** We change the camera sampling space by singly change Radius, Elevation(Ele), Azimuth(Azi), and lookat point, and report the novel view synthesis quality on the chair scene in Synthetic-NeRF dataset.

Size	Chair	Drums	Hotdog	Lego	Mic	Ship
400	108	87	98	108	19	82
800	108	100	104	108	84	108

Table 2. **Image numbers Analysis.** We reduce the training image numbers and compare with the COLMAP-based NeRF.

B. Additional Results

Image Size. We provide the test on how the input image size influences the COLMAP on the Synthetic-NeRF dataset. The input images are 108 images with the size of 400×400 or 800×800, we report the registered images concerning image size in Tab. 2. As illustrated, the COLMAP is sensitive to the image size. It works satisfactorily with the original image size 800×800 but struggles to register the downsampled images. For this reason, we provide only results of COLMAP-based NeRF with registered poses from images with the size of 800×800.

C. Applications

3D Reconstruction from Unposed Masks In Fig. 1, we learn 3D representation and camera poses from a collection of unposed masks by optimizing the radiance fields and camera poses simultaneously. Specifically, we treat the mask as a 1-channel image and render it with volume rendering as RGB images. With the trained NeRF model, we then extract the 3D representation using the marching cubes algorithm [1] following the original NeRF script [1]. This case further demonstrates that our architecture can estimate camera poses from high-level features of an image without reliance on keypoints or texture, which is fundamentally different from conventional pose estimation methods. This ability of our method can be applied to other applications, such as the task of reconstructing transparent objects whose visual appearance is too complex for image-based reconstruction. Since it is much easier to obtain the masks of their shapes either by semantic segmentation tools or other sensors.

Image Noise Analysis In Fig 2, we test our method on images with intense noise. The COLMAP-based NeRF methods completely fail to estimate the camera poses of images with Gaussian noise $\mathcal{N}(0, 0.5^2)$, leading to failure of learning the radiance fields. In contrast, our method is not sen-

[1]https://github.com/kwea123/nerf_pl

Figure 1. **3D reconstruction and camera pose estimation from a collection of masks without pose information.**

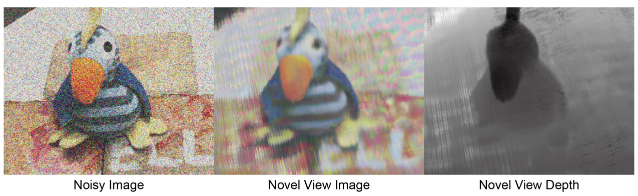

Figure 2. **Image Noise Analysis.** Despite adding intense noise on training images, our method is able to learn accurate radiance fields and camera poses of the noisy images while COLMAP-based NeRF methods completely fail.

sitive to noise and still able to render novel view and depth map with less noise. We demonstrate the accuracy of pose estimation qualitatively by a rendered image (the middle image) with the estimated pose of the left noisy image.

References

[1] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *ACM Transactions on Graphics*, 1987. 1