GigaVision Challenge

When Gigapixel Videography Meets Computer Vision

Track: Rendering

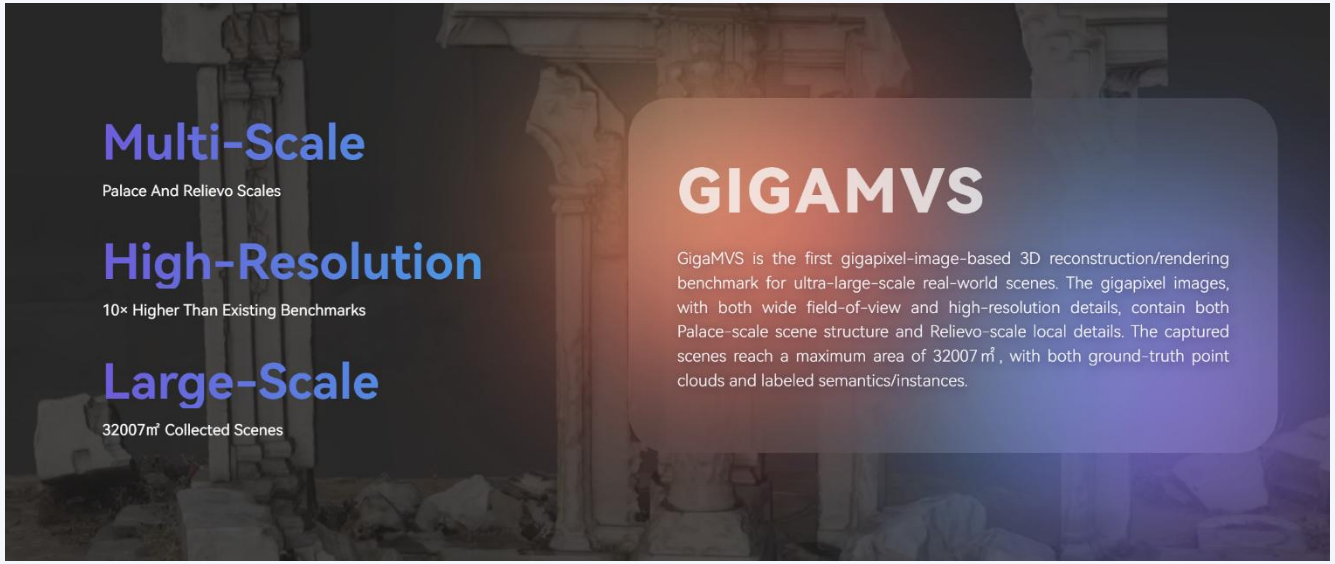
Team name: DTM 3D

Team Introduction

Zizhuang Wei is currently an **AI algorithm researcher** in **Digital Twin Lab, Huawei**. He received the Ph.D degree from Graphics and Interaction Lab, Dept. of EECS, Peking University. His research interests focus on 3D reconstruction and deep learning.

Qingtian Zhu is currently a **master student** at Graphics and Interaction Lab (GIL) of **Peking University**. His research interests include 3D reconstruction and computational photogrammetry.

Task Analysis



Original Images

Camera poses

Challenges

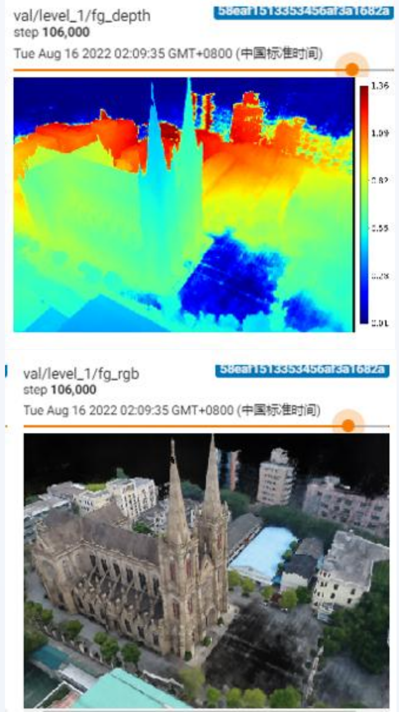
- Very high resolution
- Large scale scenes
- Unbounded scenario
- Sparse view reconstruction
- Inaccurate camera poses
- Large area of sky
- Complex lighting conditions
- ……

GIGAVISION

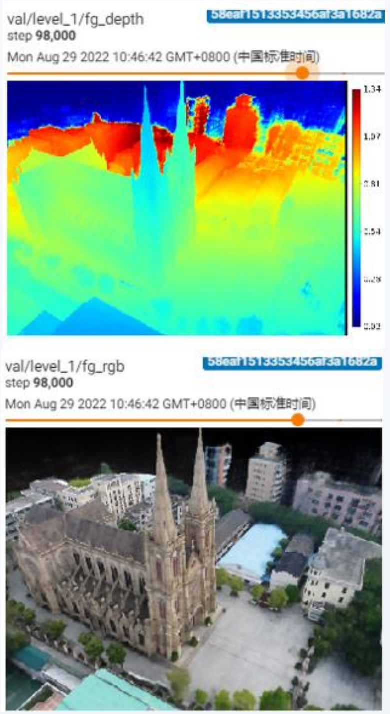
Task Analysis

	PSNR ↑	SSIM ↑	LPIPS ↓	Time (hrs)	# Params
NeRF [12, 30]	23.85	0.605	0.451	4.16	1.5M
NeRF w/ DONeRF [31] param.	24.03	0.607	0.455	4.59	1.4M
mip-NeRF [3]	24.04	0.616	0.441	3.17	0.7M
NeRF++ [46]	25.11	0.676	0.375	9.45	2.4M
Deep Blending [15]	23.70	0.666	0.318	-	-
Point-Based Neural Rendering [23]	23.71	0.735	0.252	-	-
Stable View Synthesis [38]	25.33	0.771	0.211	-	-
mip-NeRF [3] w/bigger MLP	26.19	0.748	0.285	22.71	9.0M
NeRF++ [46] w/bigger MLPs	26.39	0.750	0.293	19.88	9.0M
Our Model	27.69	0.792	0.237	6.89	9.9M
Our Model w/GLO	26.26	0.786	0.237	6.90	9.9M

Table 1. A quantitative comparison of our model with several prior works using the dataset presented in this paper.



MLP with 256 hidden units



MLP with 1024 hidden units

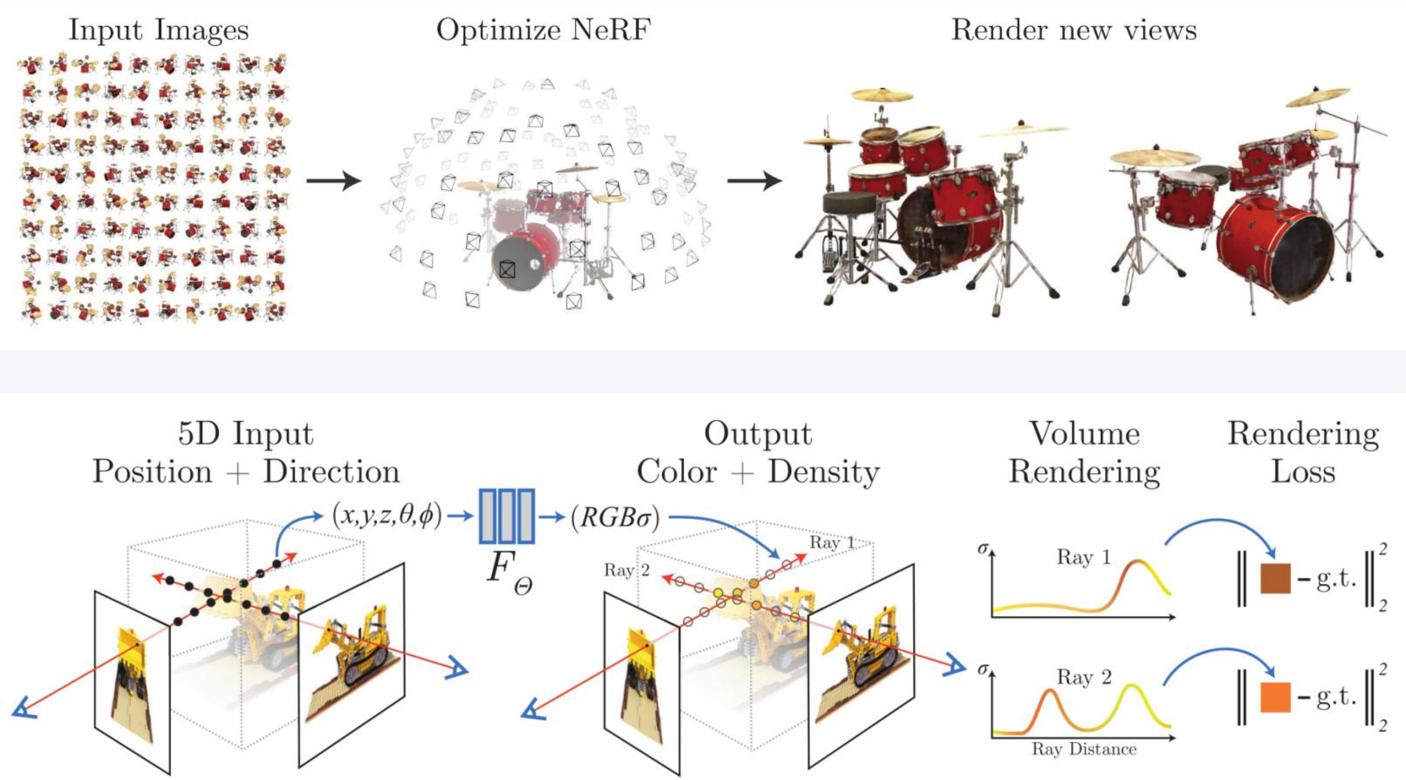
From Mip NeRF 360[1] (CVPR 2022 Oral)

PSNR=20.6 PSNR=22.8

- Parameterization in Unbounded Scenarios
- Using bigger MLPs

[1] Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., & Hedman, P. (2022). Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5470-5479).

Solution and Innovation



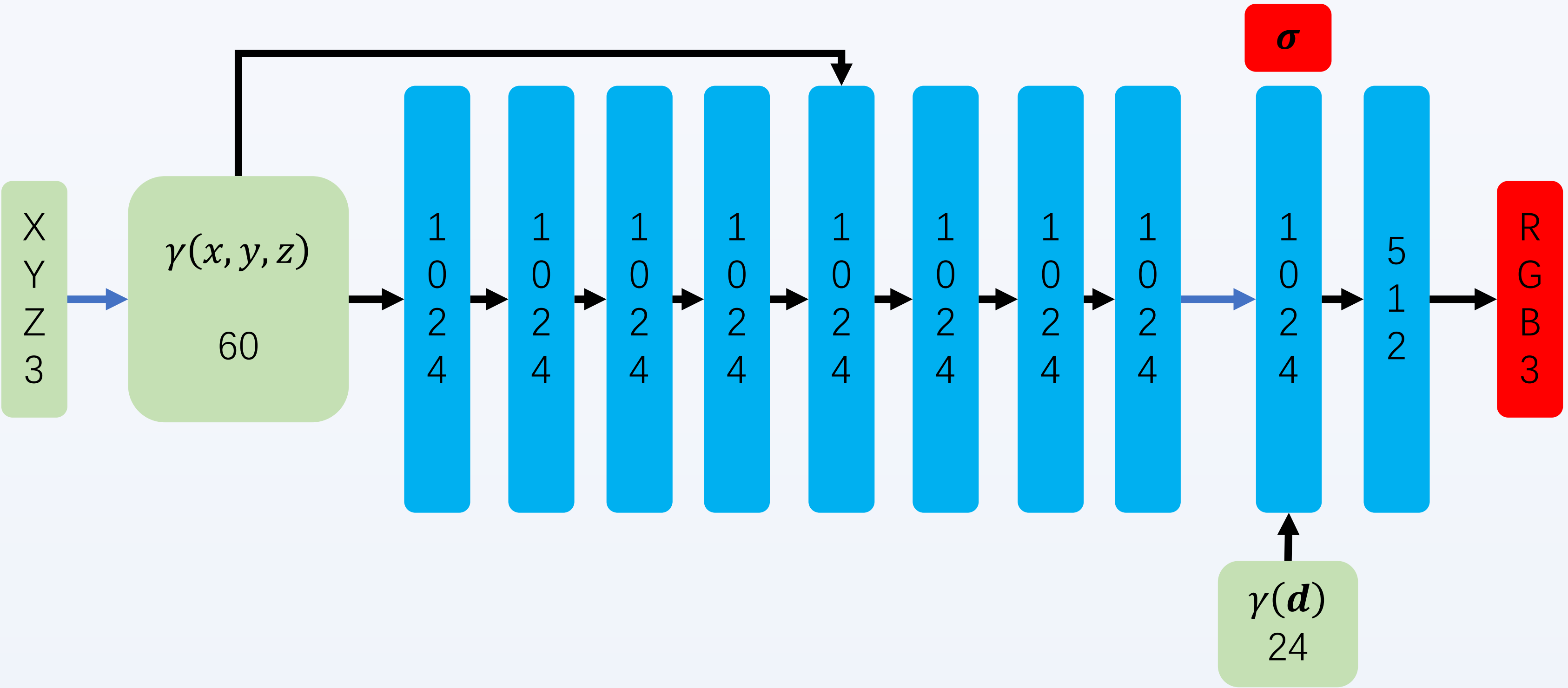
Neural Radiance Field[2] framework

We synthesize views by querying 5D coordinates along camera rays and use volume rendering techniques to project the output colors into an image. A fully-connected deep network is used to represent the scenes by **Neural Radiance Field**.

GIGAVISION

[2] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, *65*(1), 99-106.

Solution and Innovation



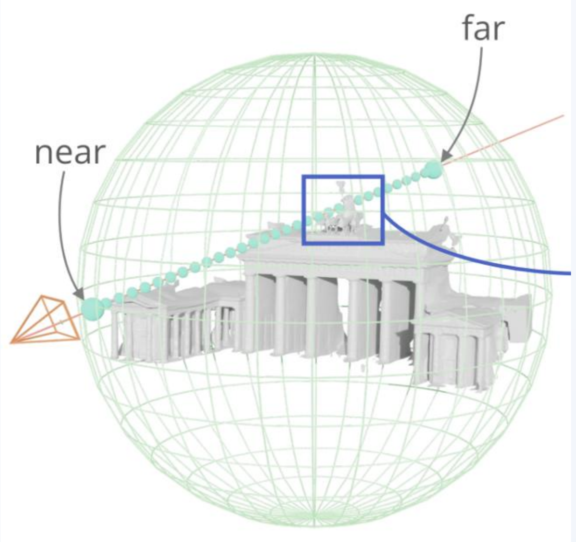
8-layer fully-connected MLP Network

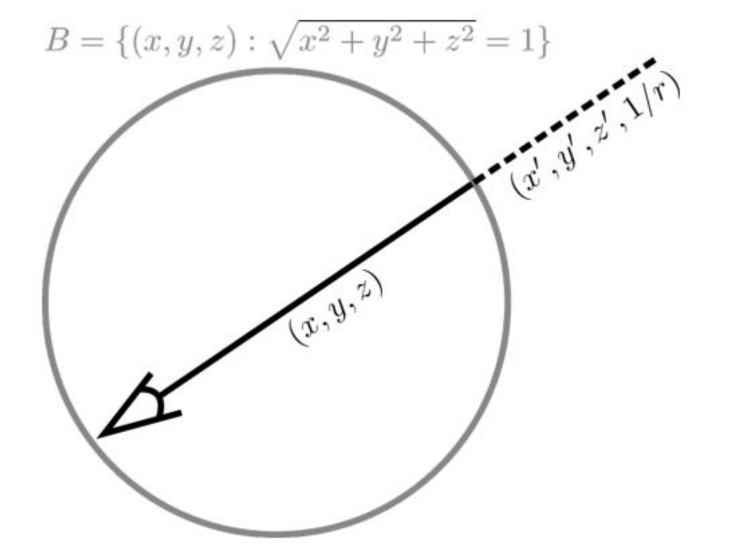
$$\mathcal{L} = \sum_{\mathbf{r}\in\mathcal{R}} \left[\left\| \hat{C}_c(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 + \left\| \hat{C}_f(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 \right]$$

We use an **8-layer mlp network** to train our model, with **1024 hidden units**, which is able improve the network's ability to represent large-scale scenes.

GIGAVISION

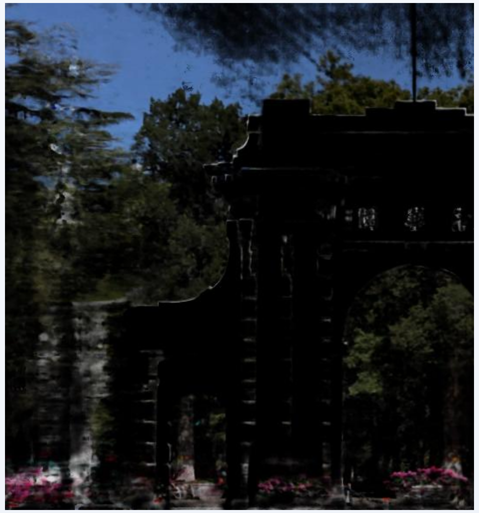
Solution and Innovation





$$\mathbf{C}(\mathbf{r}) = \underbrace{\int_{t=0}^{t'} \sigma(\mathbf{o}+t\mathbf{d}) \cdot \mathbf{c}(\mathbf{o}+t\mathbf{d}, \mathbf{d}) \cdot e^{-\int_{s=0}^{t} \sigma(\mathbf{o}+s\mathbf{d})ds} dt}_{\text{(i)}}$$

$$+ \underbrace{e^{-\int_{s=0}^{t'} \sigma(\mathbf{o}+s\mathbf{d})ds}}_{\text{(ii)}} \cdot \underbrace{\int_{t=t'}^{\infty} \sigma(\mathbf{o}+t\mathbf{d}) \cdot \mathbf{c}(\mathbf{o}+t\mathbf{d}, \mathbf{d}) \cdot e^{-\int_{s=t'}^{t} \sigma(\mathbf{o}+s\mathbf{d})ds} dt}_{\text{(iii)}}.$$



Foreground **Background**

Inside: original depth; outside: inverse depth.

We apply different parameterizations[3] for scene contents inside and outside the unit sphere.

[3] Zhang, K., Riegler, G., Snavely, N., & Koltun, V. (2020). Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*.

Preparation

File organization

```
——nerf
————data
——————DayaTemple
————————camera_path
——————————intrinsics
——————————pose
————————test
——————————intrinsics
——————————pose
——————————rgb
————————train
——————————intrinsics
——————————pose
——————————rgb
——————......
```

Requirements

Subject	Requirement
OS	Linux Ubuntu 20.04.5
GPU	32G Tesla V100 at least
CPU	Intel Xeon Gold 6132
Memory	256G+
Disk	4T
Cuda	11.4
OpenCV -python	4.4
Python	3.6.13

Settings

Subject	Requirement
Resolution for training	1086 X 724
Resolution for testing	8688 X 5792
Cascade stages	2
Cascade samples	128 , 64
Learining rate	0.0005
Iterations	>=500000

It takes about 30 days for training and about 10 days for testing with 8 X Tesla V100 GPUs.

GIGAVISION

Results

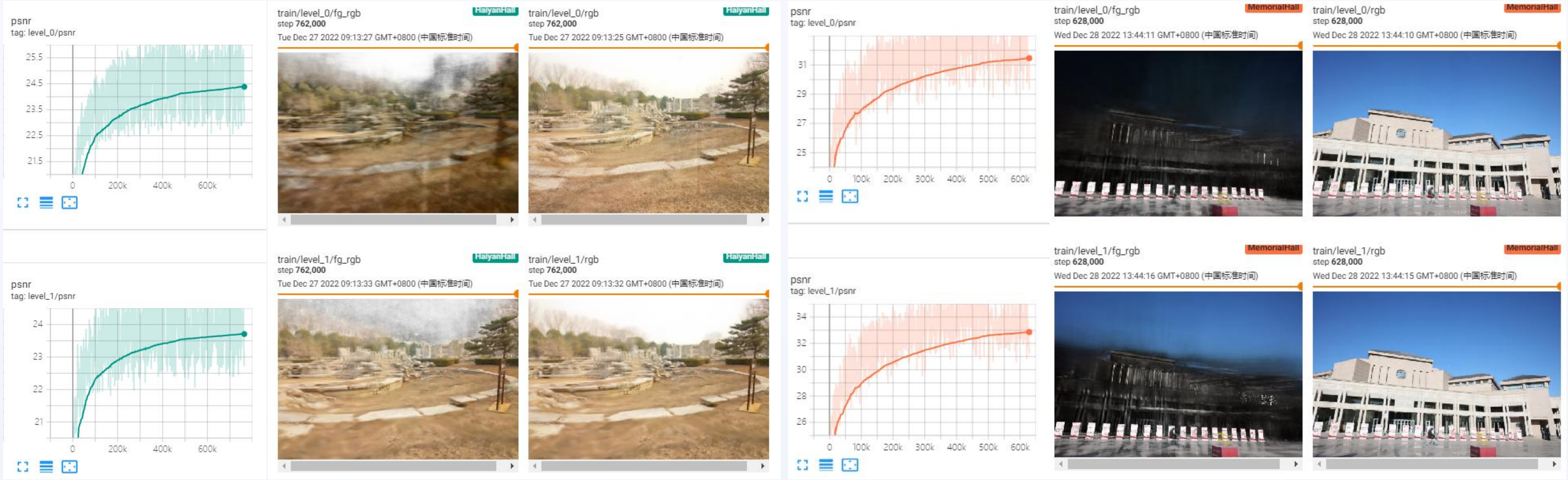
Leaderboard

#	Team	Members	PSNR	SSIM	LPIPS	Method	Code
	Ben Mildenhall et al. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis .ECCV 2020.						
	PDMYR		17.57472	0.640418	0.47436		
	DTM 3D		17.41261	0.654217	0.56762		
	Unrendered		17.40951	0.629034	0.46825		
	cs271		17.07217	0.657387	0.56029		
	算法cj		16.04978	0.586286	0.55697		
	1080Ti		12.75886	0.499533	0.5962		
	CNU		12.65009	0.563411	0.6728		
	try 1 try		10.01236	0.443952	0.72387		
	GGBoy		9.72022	0.430803	0.63517		
	CUR		4.34982	0.433014	0.37134		

Our method ranks 2nd on Track Rendering (Except the baseline methods)

GIGAVISION

Intermediate Results



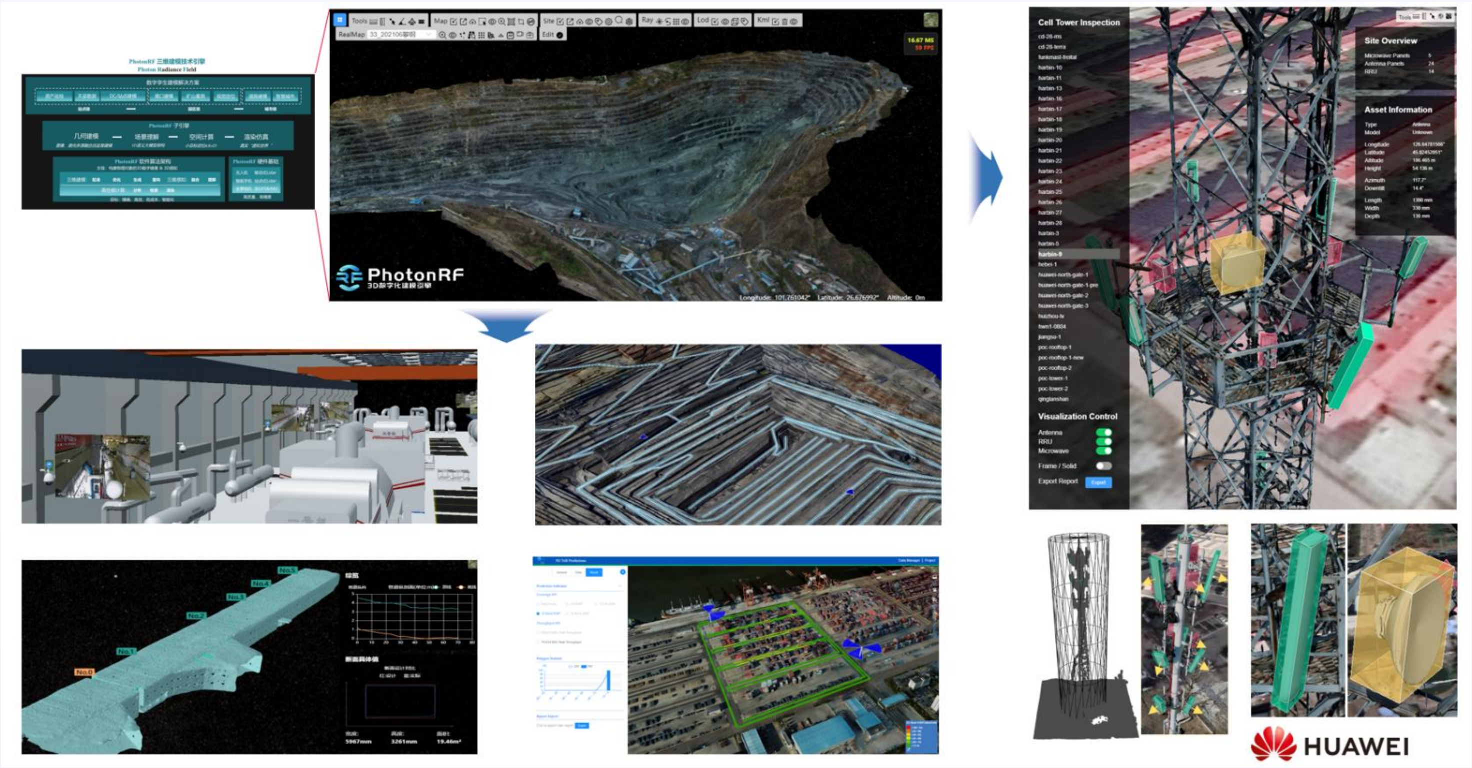
The average PSNR during training is **24.7**, while the second cascade level is not always better than the first level. However, we still keep the finer level images for evaluation. It's worth mentioning that we only submitted **once** rendering results (limited by computing resources).

Conclusion

- Our method took the **second place** using a Neural Radiance Field framework, in which **different parameterizations** for scene contents inside and outside the unit sphere and **larger MLPs** play a key role.
- Because of sparse views and large scenes, it is difficult for nerf networks to obtain perfect results. The methods which introduce **geometric constraints**, are expected to achieve better results.

GIGAVISION

Invitation Digital Twin Lab, Huawei



Our team focuses on cutting-edge technology research and engine development of **image/LiDAR 3D reconstruction** and **2/3D semantic understanding** for solving technical problems such as **environment 3D modeling** and perception in **5G network simulation**.

Contact: Hong.Shen233@huawei.com

Welcome to join us!

GIGAVISION

THANKS !