



中国人工智能学会
Chinese Association for Artificial Intelligence



余杭人才
YUHANG TALENTS



HANGZHOU FUTURE SCI-TECH CITY
杭州未来科技城



科创中国
INNOVATION CHINA



青春
潮
创
杭州
HANGZHOU

GIGAVISION

2023 GLOBAL AI
INNOVATION CONTEST

全球人工智能技术 创新大赛

▷ 算法挑战赛



2023 GLOBAL AI
INNOVATION CONTEST

全球人工智能技术 创新大赛

▷ 算法挑战赛

赛道二：GigaRendering-基于十亿像素稀疏图像的新视角渲染

队伍名称：NeuWorld



目录

NeuWorld团队介绍

榜单实验结果

Rendering赛题解析

应用demo展示

解决方案及创新点

总结与展望

NeuWorld团队介绍



队长：魏子庄，华为技术有限公司全球技术服务部AI算法研究员，本科博士均毕业于北京大学计算机学院。主要研究方向为三维视觉、深度学习、神经渲染等，曾获中国博士研究生国家奖学金，入选华为“天才少年”计划。



队员：朱擎天，北京大学计算机学院硕士三年级研究生，主要研究方向为基于深度学习的三维重建、语义分割、神经渲染等，曾获北京大学奔驰奖学金，毕业后拟赴东京大学攻读博士学位。

指导老师：



张昕，华为技术有限公司
算法与技术开发部首席技术专家



李凌云，华为技术有限公司
主任工程师、数字孪生lab PL



汪国平，北京大学教授
国家杰出青年、CCF专委会主任

Rendering赛题解析

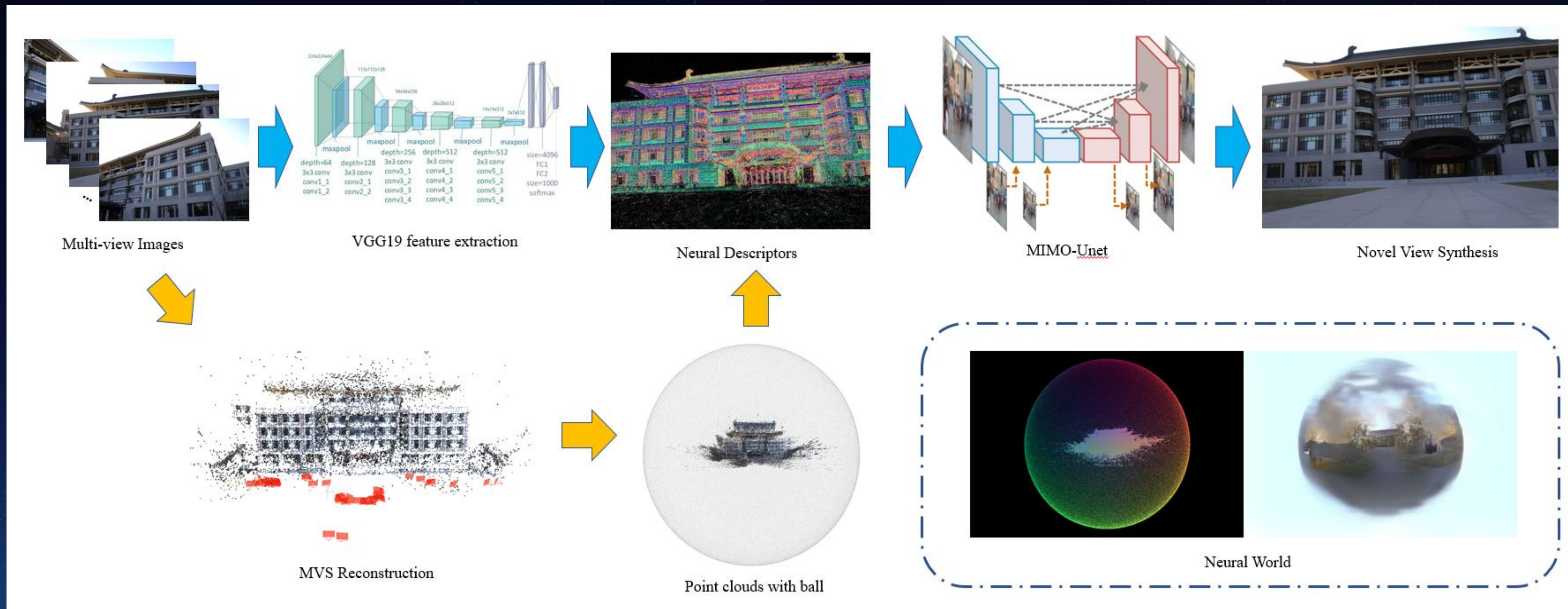


- **稀疏视角**：每个场景均只包含了不超过100张的训练视角，且场景均具有一定的规模，这些视角无法完美覆盖场景的每个角度。
- **复杂场景**：8个场景均为室外开放场景，且视角变化大，场景中的物体属性特征存在差异，不同图像之间的光照条件也存在变化。
- **超高分辨率**：场景图像的分辨率极高，以DayaTemple为例，图像的分辨率为8688x5792，需要实现对于高分辨率图像的高质量渲染。
- **位姿不准确**：Rendering数据集所提供的相机位姿由COLMAP方法得到，其本身存在一定的误差，这要求算法具有一定的鲁棒性。
- **算法效率**：主办方限制了计算资源与代码复现的总时长，在给定8块24G显存的3090显卡的前提下，整个网络训练与推理在3天之内全部完成，这也对算法的性能和效率提出了要求。

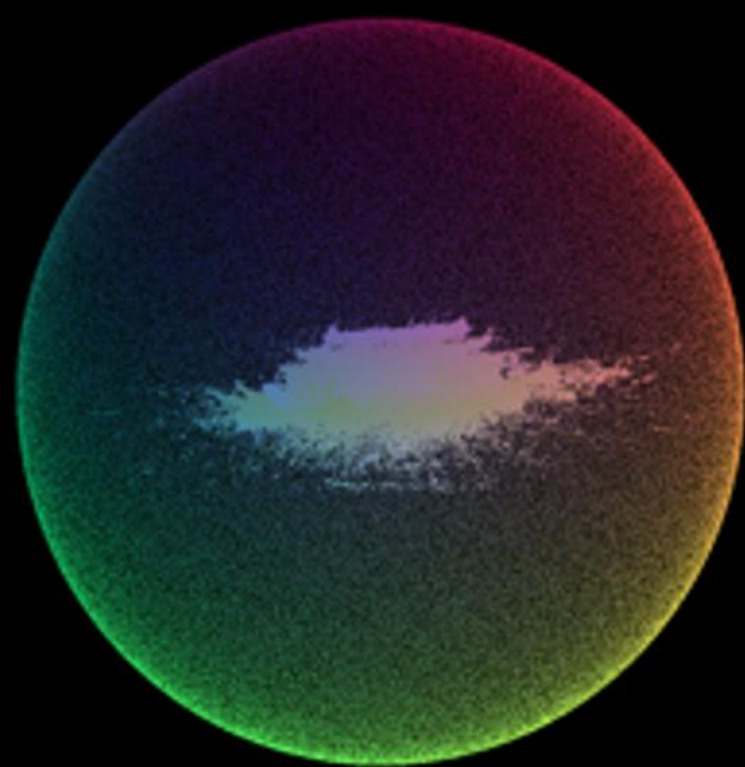
场景名称	是否室外开放场景	图像分辨率	训练视角数量	测试视角数量
Daya Temple	✓	8688x5792	83	17
HaiyanHall	✓	9504x6336	25	5
Library	✓	9568x6376	57	11
MemorialHall	✓	8688x5792	17	13
Museum	✓	8688x5792	54	32
PeonyGarden	✓	9568x6376	42	9
ScienceSquare	✓	9568x6376	32	7
theOldGate	✓	8688x5792	47	28

解决方案及创新点

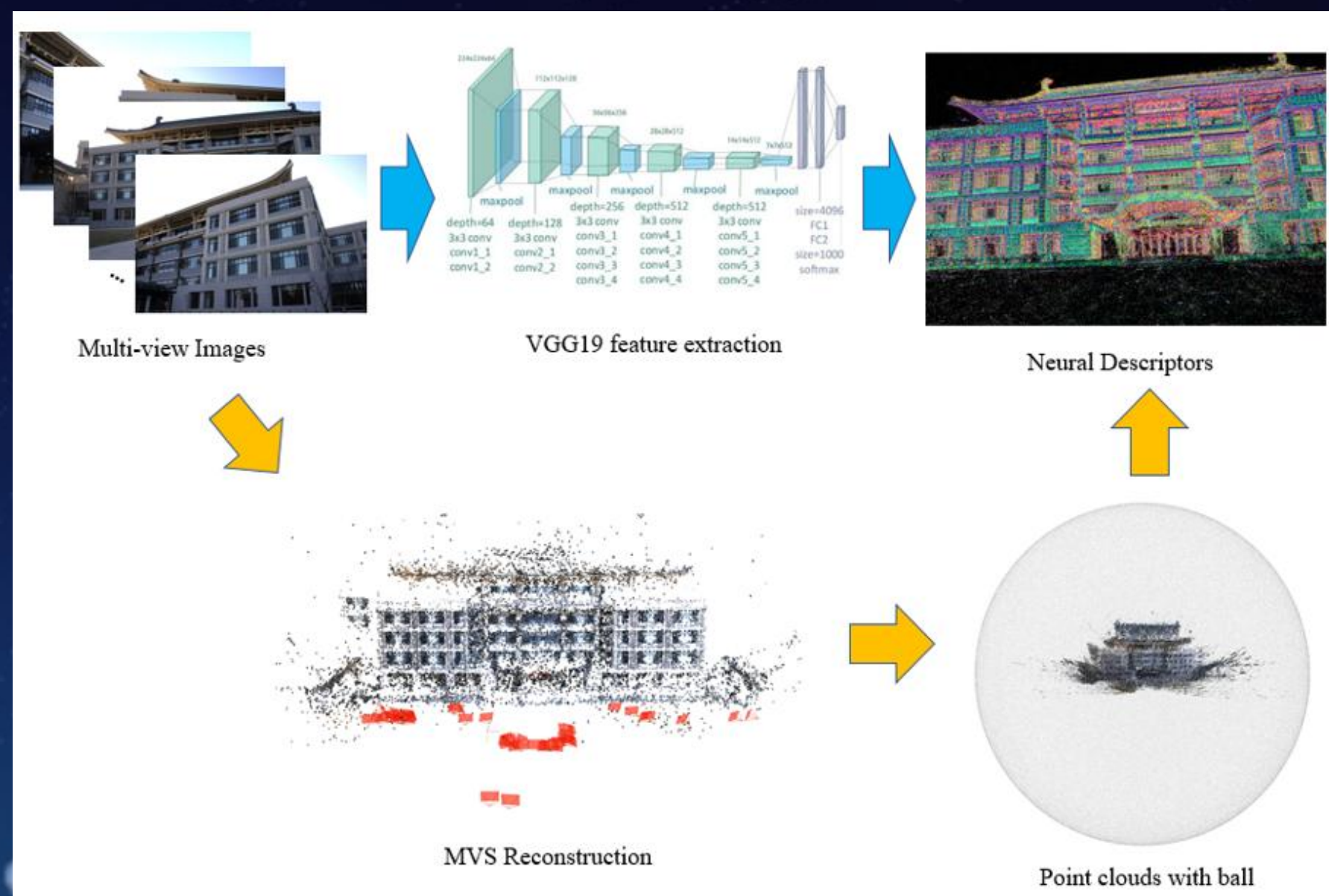
NeuWorld方法整体流程



解决方案及创新点



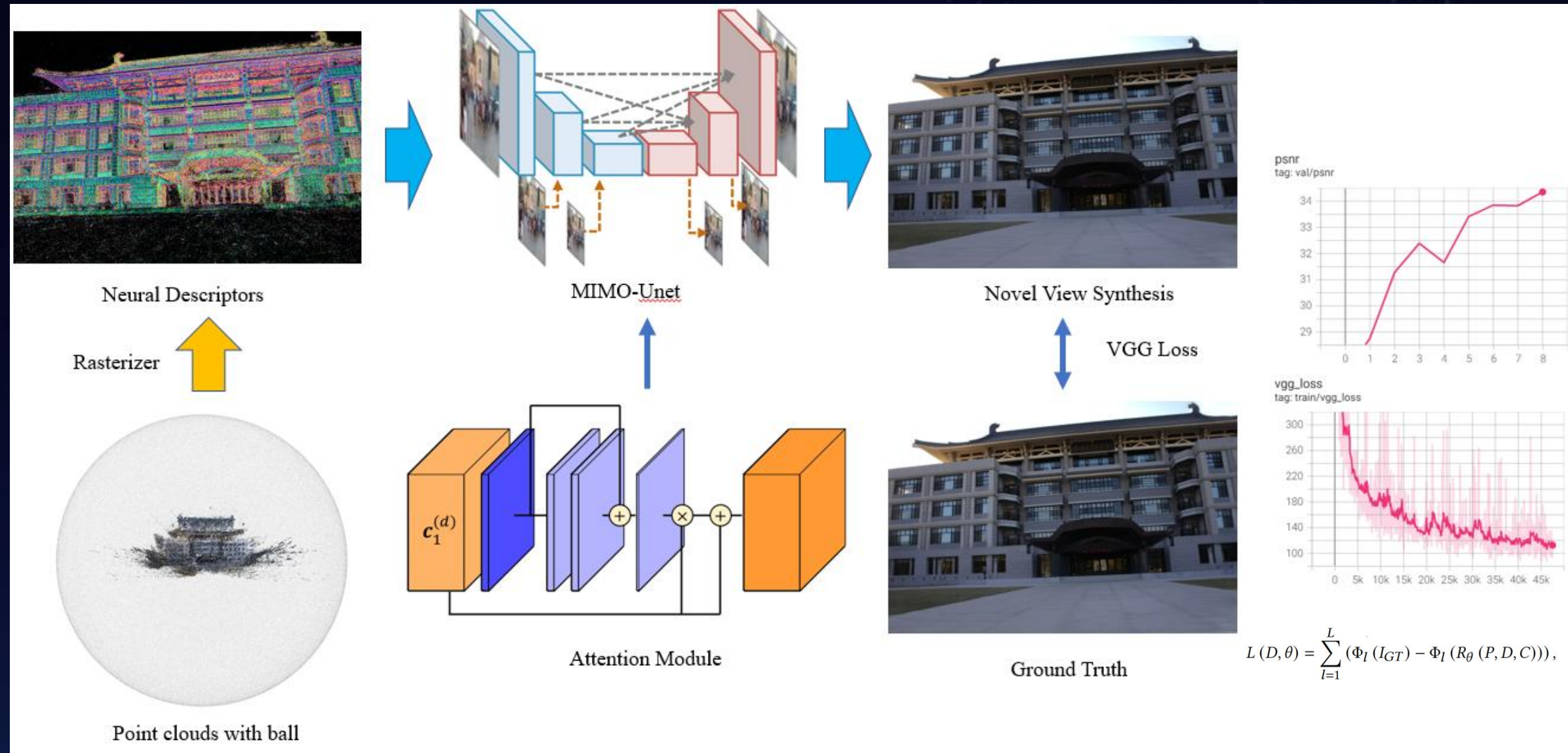
- 我们针对Giga Vision数据集有针对性地提出了一套基于**Neural Point-based Rendering**的**神经渲染**流程，该方法能够有效地处理这些视角稀疏的欠约束场景，且能够对于不同的场景实现自监督端到端网络训练，不需要额外的训练数据。整个网络能够在比赛规定的时间和显存资源开销内完成整个训练和渲染流程，**在比赛榜单上取得了第1名。**



- 针对开放场景中存在大片天空区域、深度跨度广等问题，我们提出了**基于空间包围球的环境Environment Map建模方法**，将天空等各种环境相关信息记录在包围球的点云上，**再通过包围球上映射到的神经描述子构建出新视角的环境信息**。该方法大幅提升了我们算法对于室外开放场景的表现，使得对于这些场景的环境渲染结果更加逼真。

特征点云编码网络

解决方案及创新点



特征点云解码网络

在得到新视角渲染得到的特征图像后，我们通过多尺度UNet网络对特征图像进行解码和补全，最终得到完整的新视角渲染图像。我们通过计算网络生成结果和Ground Truth图像之间的VGG Loss对整个网络进行端到端的自监督训练

➤ 我们参照AAAI2023 READ文章中的做法，在神经点云渲染阶段引入了MIMO-UNet的网络结构，并将其扩展到了4层。对于每个网络层级，我们引入了**注意力机制Gated Network**用于减少异常值，并参照**MIMO-UNet++**的思想扩展了Res Block的数量，使得网络能够表述大规模的复杂场景。

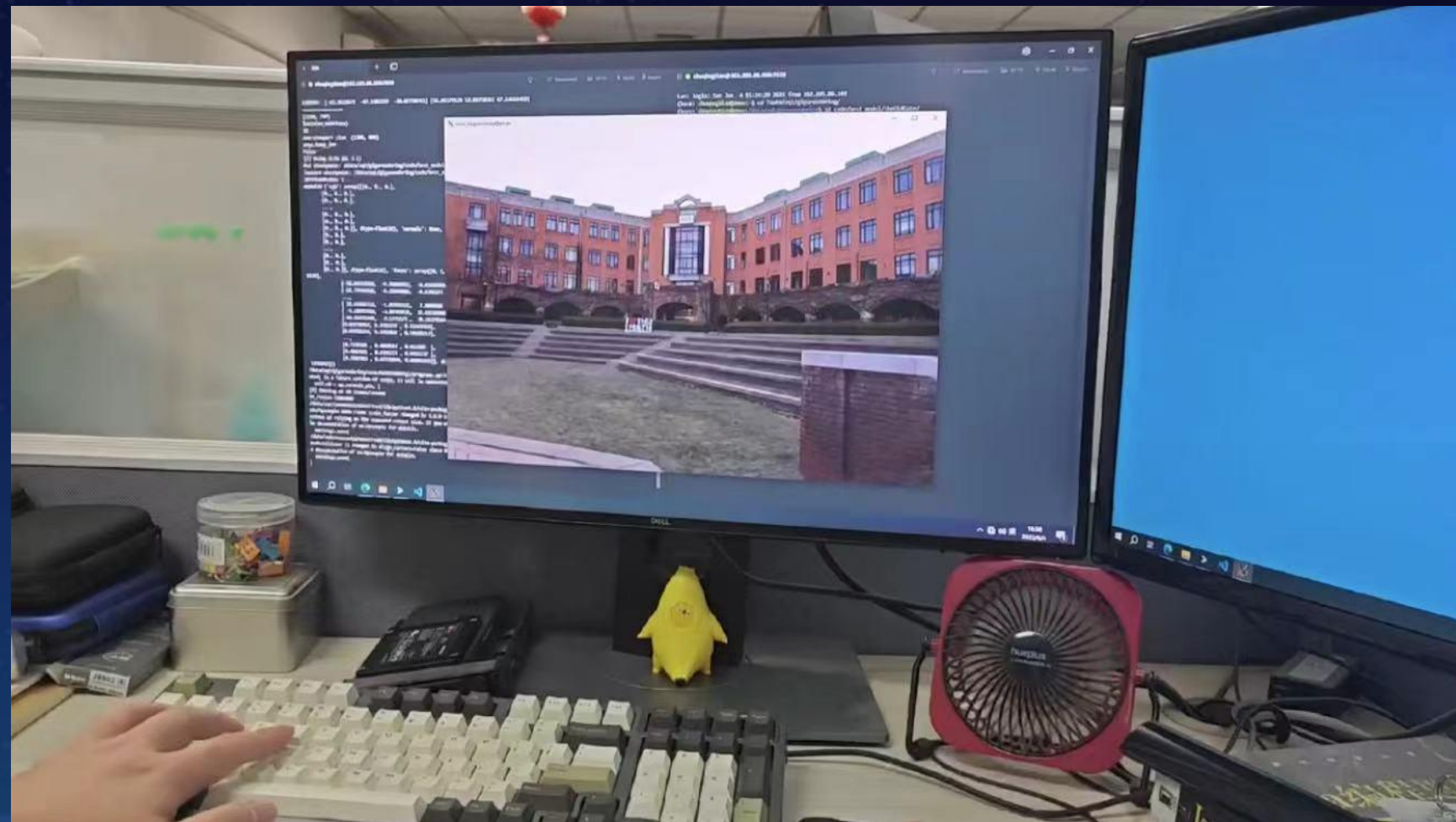
➤ 针对训练数据中位姿不准、光照不一致等问题，我们对训练数据进行了数据增强，即每次采样的数据在原先图像的基础上进行了Crop、Resized、Shift等操作，从而扩展出了大量的额外训练数据。这种**数据增强的机制**综合考虑到了多视角图像从不同方向上的观测信息，避免了网络过拟合到某个局部解上，**增强了网络的泛化性和生成新视角图像的能力。**

解决方案及创新点

NVIDIA-SMI 470.182.03 Driver Version: 470.182.03 CUDA Version: 11.4									
GPU	Name	Persistence-M	Bus-Id	Disp.A	Volatile	Uncorr.	ECC		
Fan	Temp	Perf	Pwr:Usage/Cap	Memory-Usage	GPU-Util	Compute	M.		
						MIG	M.		
0	Tesla V100-SXM2 ...	On	00000000:20:00.0	Off	68%	Default	0		
N/A	47C	P0	228W / 300W	19315MiB / 32510MiB		N/A			
1	Tesla V100-SXM2 ...	On	00000000:32:00.0	Off	55%	Default	0		
N/A	49C	P0	260W / 300W	17442MiB / 32510MiB		N/A			
2	Tesla V100-SXM2 ...	On	00000000:5B:00.0	Off	100%	Default	0		
N/A	49C	P0	260W / 300W	17772MiB / 32510MiB		N/A			
3	Tesla V100-SXM2 ...	On	00000000:5F:00.0	Off	88%	Default	0		
N/A	69C	P0	227W / 300W	20660MiB / 32510MiB		N/A			
4	Tesla V100-SXM2 ...	On	00000000:B5:00.0	Off	100%	Default	0		
N/A	53C	P0	249W / 300W	18546MiB / 32510MiB		N/A			
5	Tesla V100-SXM2 ...	On	00000000:BE:00.0	Off	100%	Default	0		
N/A	49C	P0	260W / 300W	17838MiB / 32510MiB		N/A			
6	Tesla V100-SXM2 ...	On	00000000:DF:00.0	Off	100%	Default	0		
N/A	50C	P0	243W / 300W	17350MiB / 32510MiB		N/A			
7	Tesla V100-SXM2 ...	On	00000000:E7:00.0	Off	100%	Default	0		
N/A	51C	P0	260W / 300W	17346MiB / 32510MiB		N/A			

上半区4卡Batchsize=8
并发训练

下半区4卡Batchsize=8
并发训练



- 为了满足比赛对于计算资源的需求，我们在工程实现当中引入了**多GPU并行加速的机制**。而为了更好地调度GPU集群的资源，减少显卡间通信开销，我们对8块**GPU分为了上下半区调度**，并将同一显卡分为两个Batch进行训练。上述显存调度加速策略**充分利用了8块RTX3090显卡的算力**，使得我们的方法在同样计算资源下计算效率提升80%。
- 除了针对比赛的设计外，为了方法能够拥有更好的实用价值，我们还采用了基于OpenGL Shader光栅化渲染的异步调度机制实现**场景的实时渲染交互**；而基于神经点云的渲染方法天然具有**场景的可编辑性**，通过改动这些点云能够对于场景中的各种物体进行修改。这些扩展也使得我们的方法具有了**良好的应用性**。

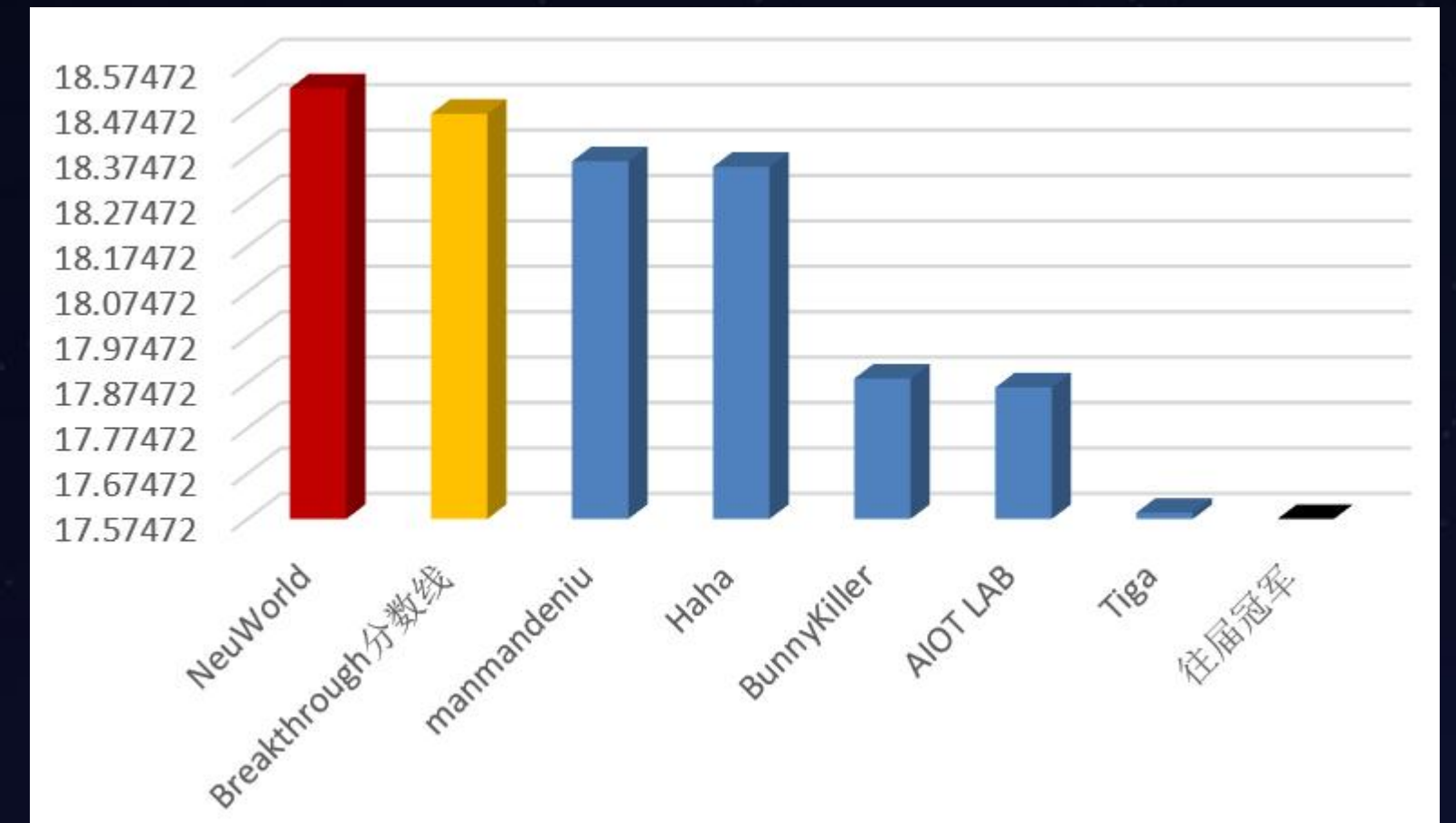
榜单实验结果

Leaderboard

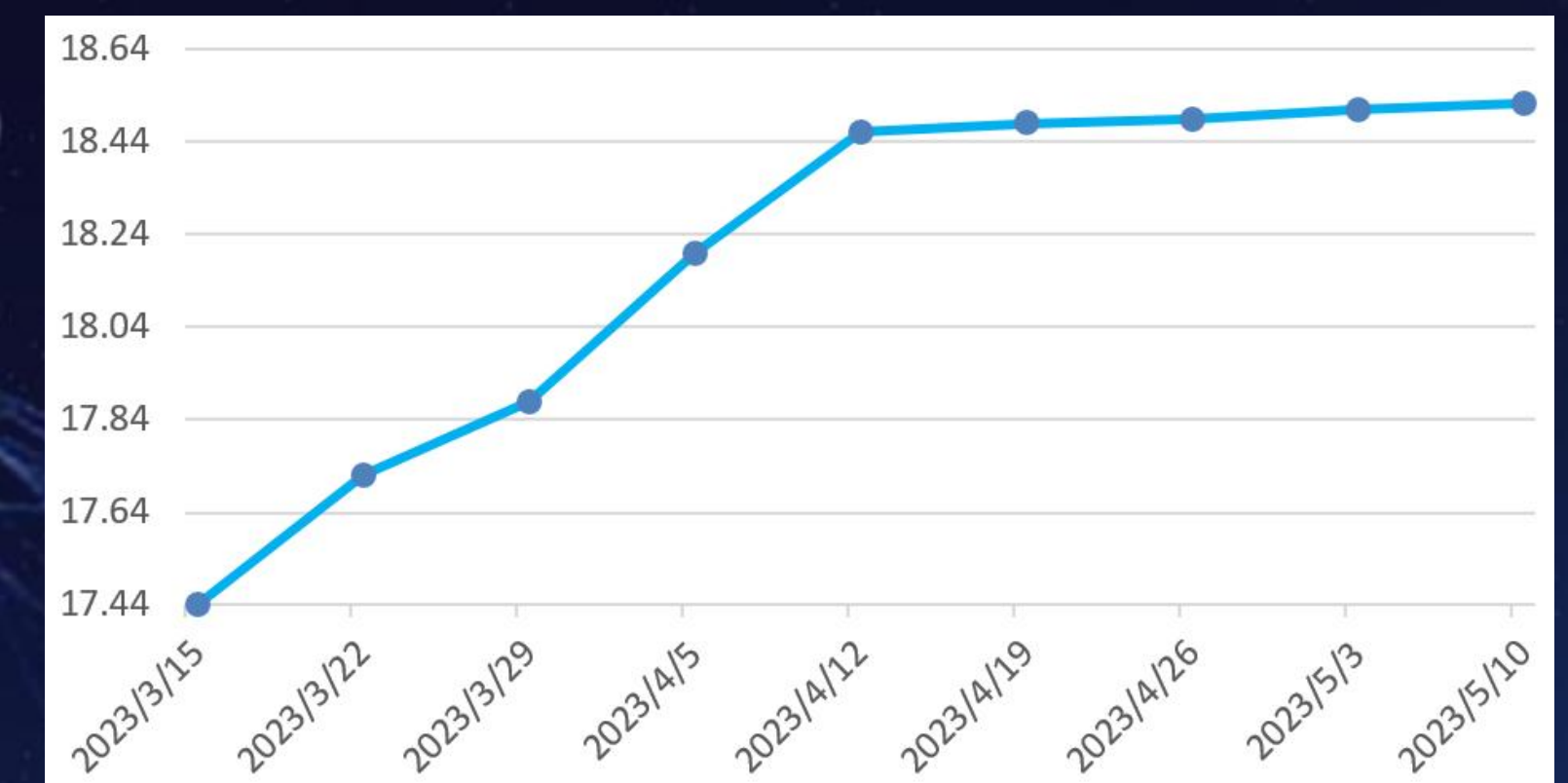
Preliminary Overall Ranking

Breakthrough Award score: 18.46682(PSNR).

#	Team	Members	PSNR	SSIM	LPIPS	Method	Code	Paper
1	NeuWorld		18.52301	0.630421	0.50278			
2	manmandeniu		18.36235	0.644217	0.32751			
3	Haha		18.35047	0.629117	0.42156			
4	BunnyKiller		17.88386	0.6108	0.4323			
5	AIOT LAB		17.86436	0.604607	0.5718			
6	Tiga		17.58869	0.567661	0.55641			
7	Contraction		17.55792	0.635199	0.48498			
8	湘北高中		17.53878	0.60103	0.47963			
9	407		17.4813	0.611828	0.49654			
10	cct小白菜		17.42156	0.577901	0.50958			



前六名队伍分数分布



NeuWorld每周分数变化

2023 GAIC-Rendering比赛榜单最终结果

榜单实验结果

新视角渲染结果展示



Daya Temple



Haiyan Hall



Library

从定性的角度分析，NeuWorld方法能够生成出高保真度的新视角图像，渲染出来的图像非常接近于真实拍照的效果。

榜单实验结果

新视角渲染结果展示



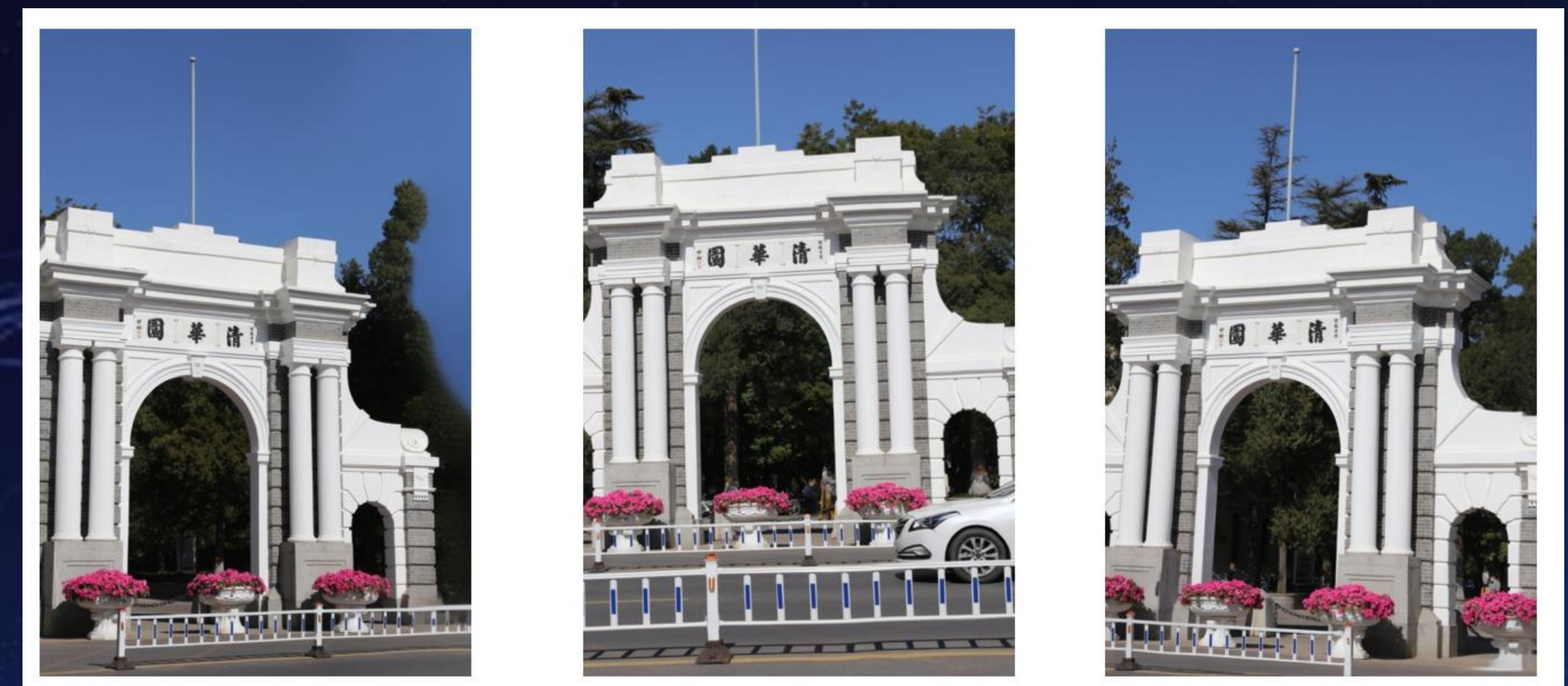
MemorialHall



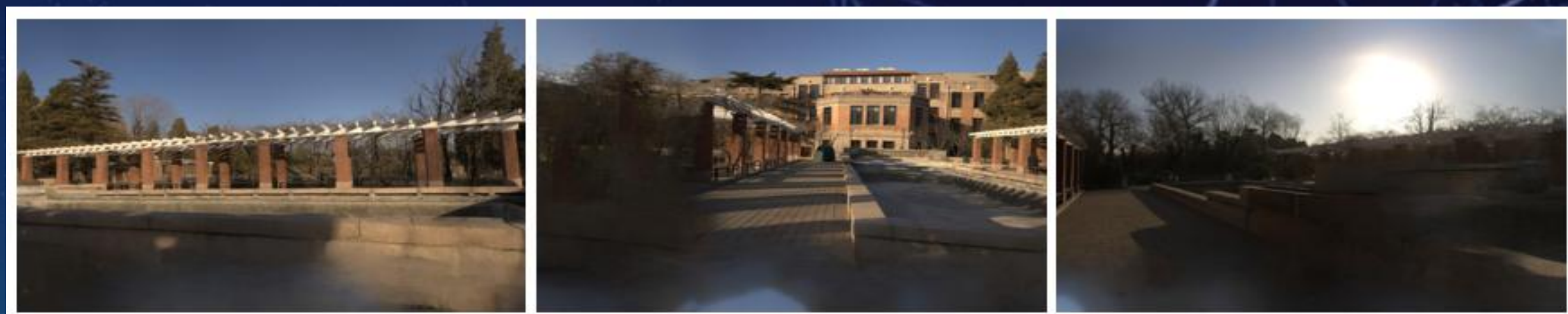
ScienceSquare



Museum



theOldGate



PeonyGarden

榜单实验结果



NeRF++

Ours

我们的方法和NeRF++方法相比具有明显的优势，**网络收敛速度达到NeRF++方法的30倍以上，推理速度提升1000倍以上。**训练过程中我们网络仅需要10分钟PSNR就能达到20，而NeRF++通常达到相同的PSNR需要超过10小时。测试阶段我们的方法能够通过点云快速光栅化实现自由视角的实时渲染，而NeRF++通常需要10分钟以上。从最终的渲染效果上来看，由于几何约束的引入，我们的方法能够基于稀疏视角图像合成出高保真度自由视角，其**真实感远远超过NeRF++方法。**

我们方法与NeRF++方法的直观对比

应用demo展示

场景实时漫游渲染



Library场景实时交互漫游



MemorialHall场景实时交互漫游

应用demo展示

场景编辑实例展示



移除华表



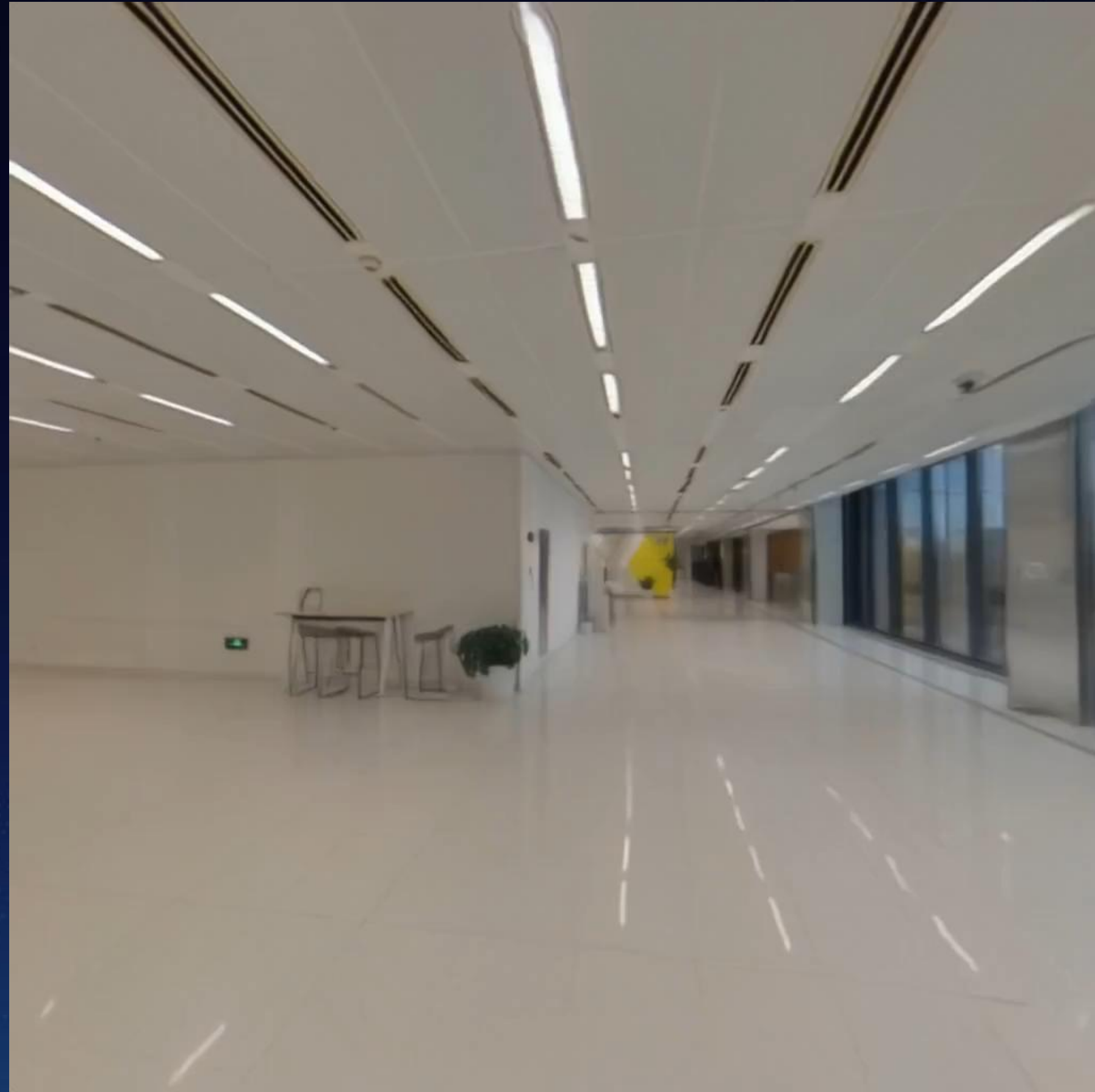
移除帐篷



移除护栏

应用demo展示

自采数据应用实例展示



百张图像实现室内沉浸式漫游



大规模城市场景漫游 (对比传统mesh建模)

总结与展望

- 我们提出了一种基于点云的神经渲染方法，能够得到高保真度新视角渲染图像，在**2023 GAIIIC-Rendering**赛道中取得了**第1名**。
- 针对Giga Vision数据集中视角稀疏的欠约束场景，我们在渲染网络中引入**三维点云的几何约束**，能够对于不同的场景实现自监督端到端网络训练，不需要额外的训练数据。针对开放场景中存在大片天空区域、深度跨度广等问题，我们提出了基于空间包围球的环境**Environment Map**建模方法，将天空等各种环境相关信息记录在包围球的点云上，再通过包围球上映射到的神经描述子构建出新视角的环境信息。为了进一步地提高算法指标上的性能，我们对特征解码网络UNet进行了扩展，在神经点云渲染阶段引入了**MIMO-UNet**的网络结构。对于每个网络层级，我们引入了注意力机制**Gated Network**用于减少异常值，并参照MIMO-UNet++的思想扩展了Res Block的数量，使得网络能够表述大规模的复杂场景。除了这些网络上的改动以外，我们还引入了**数据增强、多GPU分区并发、光栅化渲染**等机制，使得方法能够在规定的时间和显存开销下完成训练和渲染。
- 实验结果显示，我们的方法大幅领先于其他参赛队伍，得到了非常接近于真实拍照的新视角渲染图像，且**具有实时渲染、场景编辑等能力**，具有很好的实用价值。
- 稀疏视角场景的三维建模和渲染始终是三维视觉领域的重点研究方向，**视角数量不足所带来的解的歧义性**始终是实际应用当中需要克服和解决的问题。而想要解决欠约束这一问题势必需要引入一些先验信息加以辅助，相信未来**几何信息与三维渲染的结合**仍是解决这一问题的关键，也肯定会有涌现出色的方法带来更为逼真的渲染效果，本方法作为NeuWorld团队的一次探索和尝试，希望能够给广大三维视觉领域研究重建和渲染问题的学者们带来启发。

参考文献

- [1] Cho S J, Ji S W, Hong J P, et al. Rethinking coarse-to-fine approach in single image deblurring[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 4641-4650.
- [2] Zhang K, Riegler G, Snavely N, et al. Nerf++: Analyzing and improving neural radiance fields[J]. arXiv preprint arXiv:2010.07492, 2020.
- [3] Zhang J, Zhang J, Mao S, et al. GigaMVS: A Benchmark for Ultra-Large-Scale Gigapixel-Level 3D Reconstruction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11): 7534-7550.
- [4] Schonberger J L, Frahm J M. Structure-from-motion revisited[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 4104-4113.
- [5] Aliev K A, Sevastopolsky A, Kolos M, et al. Neural point-based graphics[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16. Springer International Publishing, 2020: 696-712.
- [6] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.
- [7] Li Z, Li L, Ma Z, et al. READ: Large-Scale Neural Scene Rendering for Autonomous Driving. Proceedings of AAAI, 2023.
- [8] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer International Publishing, 2016: 694-711.

感谢观看