

# PEP-GS: Perceptually-Enhanced Precise Structured 3D Gaussians for View-Adaptive Rendering

Junxi Jin, Xiulai Li, Haiping Huang, Lianjun Liu, Yujie Sun

Hainan University, Haikou, China

{junxijin, lixiulai01, huanghp, lianjun\_320 sunyujie}@hainanu.edu.cn

## Abstract

Recent advances in structured 3D Gaussians for view-adaptive rendering, particularly through methods like Scaffold-GS, have demonstrated promising results in neural scene representation. However, existing approaches still face challenges in perceptual consistency and precise view-dependent effects. We present PEP-GS, a novel framework that enhances structured 3D Gaussians through three key innovations: (1) a Local-Enhanced Multi-head Self-Attention (LEMSA) mechanism that replaces spherical harmonics for more accurate view-dependent color decoding, and (2) Kolmogorov-Arnold Networks (KAN) that optimize Gaussian opacity and covariance functions for enhanced interpretability and splatting precision.(3) a Neural Laplacian Pyramid Decomposition (NLPD) that improves perceptual similarity across views. Our comprehensive evaluation across multiple datasets indicates that, compared to the current state-of-the-art methods, these improvements are particularly evident in challenging scenarios such as view-dependent effects, specular reflections, fine-scale details and false geometry generation.

## 1. Introduction

Neural rendering of 3D scenes has emerged as a transformative technology in computer vision and graphics, enabling photorealistic visualization with applications spanning virtual reality, digital media production, and architectural visualization. Recent advances in 3D Gaussian Splatting (3DGS) have marked a significant milestone in this field, achieving real-time rendering while maintaining high visual fidelity. The subsequent development of structured 3D Gaussians, as demonstrated by Scaffold-GS, has further enhanced this approach by introducing a hierarchical scene representation that respects underlying geometric structures.

However, despite these advances, several critical challenges persist. First, while existing methods achieve high



Figure 1. Comparison of PEP-GS against SOTA methods. Our method achieves better perceptual consistency and more accurate view-dependent effects across different viewing angles, particularly in scenes with complex lighting, textures, and false geometry generation.

numerical accuracy in metrics like PSNR, they often struggle to maintain perceptual consistency across different viewing angles, particularly in scenes with complex lighting and reflections. Second, the prevalent use of spherical harmonics for encoding view-dependent effects limits the accurate representation of local lighting variations and specular highlights. Third, current approaches to Gaussian opacity and covariance estimation lack interpretability and precise control, often resulting in suboptimal splatting quality in challenging scenarios.

To address these limitations, we present PEP-GS, a perceptually-enhanced framework for structured 3D Gaussian splatting. Figure 1 demonstrates the superior performance of our approach compared to existing methods, particularly in maintaining perceptual consistency and handling complex view-dependent effects. Our approach builds upon the hierarchical structure of Scaffold-GS while intro-

ducing three key technical innovations:

- A Local-Enhanced Multi-head Self-Attention (LEMSA) mechanism that replaces traditional spherical harmonics, capturing sophisticated view-dependent effects through learned attention patterns over local geometric features.
- Kolmogorov-Arnold Networks (KAN) that provide a mathematically rigorous framework for opacity and covariance estimation, improving both the interpretability and accuracy of Gaussian splatting.
- Introducing a Normalized Laplacian Pyramid Distance (NLPD) as a perceptual loss metric enhances perceptual consistency in image quality assessment and image generation by systematically evaluating and aligning multi-scale visual features.

Through extensive experimentation, we demonstrate that PEP-GS achieves significant improvements over previous state-of-the-art methods. Our approach shows particular strength in challenging scenarios such as specular reflections, view-dependent lighting effects, and fine geometric details. Notably, we maintain real-time rendering capabilities while delivering superior visual quality and consistency across novel views.

The primary contributions of this work are:

- A perceptually-enhanced framework for structured 3D Gaussian splatting that significantly improves view-dependent rendering quality while maintaining real-time performance.
- Novel neural architectures (LEMSA, KAN, and NLPD) that address fundamental limitations in current Gaussian-based neural rendering approaches.
- Comprehensive experimental validation demonstrating substantial improvements over existing methods, particularly in challenging scenarios involving complex lighting and geometric details.

## 2. Related work

### 2.1. Neural Scene Representations and 3D Gaussians

Neural scene representation has evolved significantly from early MLP-based approaches to current state-of-the-art methods. NeRF [17] pioneered volumetric representation using coordinate-based networks, achieving high quality but suffering from slow rendering speeds. Subsequent works explored various acceleration techniques [1, 4] and alternative representations like sparse voxel grids [2] and hash-based structures [18]. A breakthrough came with 3D Gaussian Splatting [5], which achieved real-time performance through efficient rasterization of optimizable 3D Gaussians. However, this approach often leads to redundant Gaussians trying to fit every training view, resulting in less robust performance under significant view changes and challenging lighting conditions.

Most recently, Scaffold-GS [16] made a significant advance by introducing a structured hierarchy to 3D Gaussian representation. It utilizes anchor points initialized from Structure-from-Motion to establish a region-aware scene representation, where each anchor spawns a set of neural Gaussians with learnable offsets. The physical properties of these Gaussians are dynamically predicted based on anchor features and viewing conditions. This structured approach effectively reduces redundant Gaussians while maintaining high-quality rendering, showing enhanced capability in handling scenes with varying levels-of-detail and view-dependent observations. However, its reliance on spherical harmonics for color encoding and basic opacity estimation still limits its performance in complex specular effects and precise geometric details.

### 2.2. View-dependent Effects and Quality Enhancement

Current methods for handling view-dependent effects predominantly rely on spherical harmonics [2, 5] for their computational efficiency. While point-based approaches [20] offer alternative solutions, they often struggle with scene coverage and consistency. The challenge of achieving perceptually consistent results across different viewing angles remains particularly acute in scenes with complex lighting and reflections. Previous attempts at quality enhancement have focused primarily on improved sampling strategies [1] or feature encoding [18], but have not fully addressed the fundamental limitations in handling view-dependent effects and perceptual consistency. Our work builds upon Scaffold-GS’s structured representation while introducing novel components specifically designed to overcome these limitations through enhanced perceptual decomposition, sophisticated attention mechanisms, and rigorous opacity estimation.

## 3. Methods

PEP-GS builds upon the structured 3D Gaussian framework introduced in Scaffold-GS [16], while introducing three key technical innovations to address its limitations in perceptual consistency, view-dependent effect rendering, and splatting precision. As illustrated in Figure 2, our method comprises three main components: Local-Enhanced Multi-head Self-Attention (LEMSA), Kolmogorov-Arnold Networks (KAN) and Neural Laplacian Pyramid Decomposition (NLPD).

**Local-Enhanced Multi-head Self-Attention (LEMSA):** To capture rich view-dependent effects such as specular reflections and highlights, we propose LEMSA as an attention-based alternative to spherical harmonics in traditional MLP decoders. LEMSA leverages the viewing direction to achieve dynamic feature aggregation, enabling adaptive information fusion within the local neighborhood

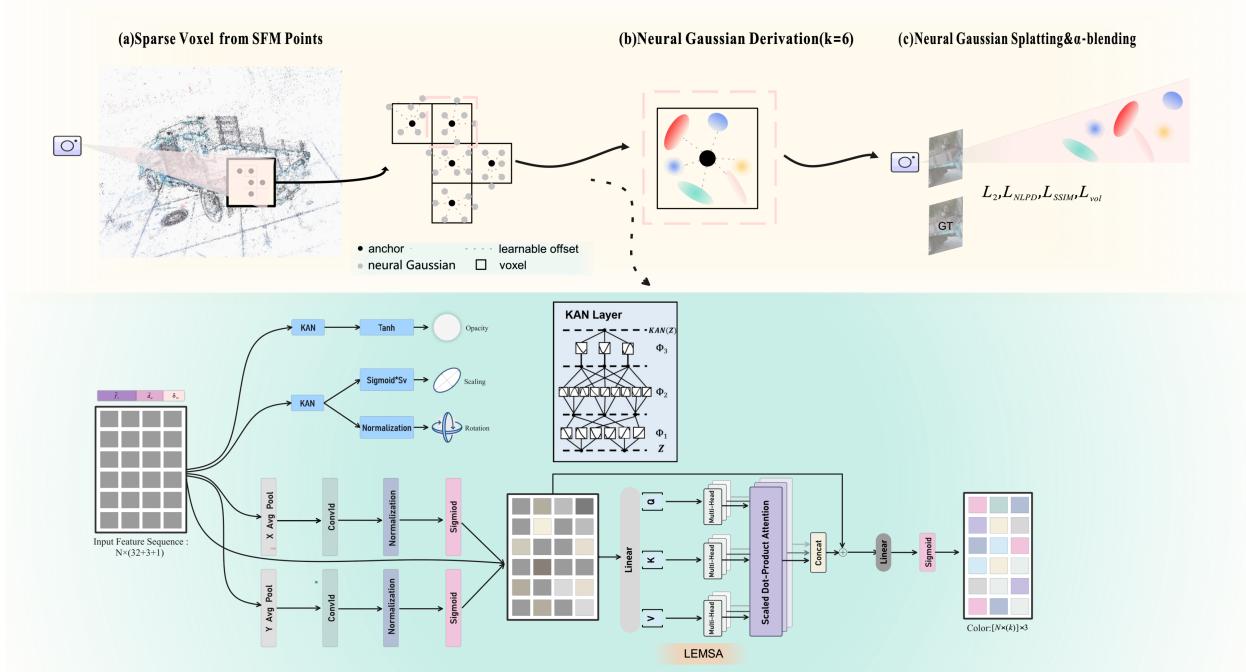


Figure 2. Overview of PEP-GS

of each Gaussian and optimizing color representation at both local and global levels.

Specifically, we first concatenate the anchor feature  $\hat{f}_v$ , the relative viewing distance  $\delta_{vc}$  and direction  $\vec{d}_{vc}$  between the camera and the anchor point into a feature sequence. This sequence undergoes average pooling along the X and Y directions, followed by 1D convolution and custom normalization. The pooled features are then activated by a Sigmoid function to produce adaptive weights. These weights are multiplied element-wise with the original input to enhance local features, allowing for adaptive adjustment based on directional changes.

After local optimization, we further capture global information by projecting the features into query (Q), key (K), and value (V) vectors. Attention weights are computed by taking the dot product between the query vector of the center Gaussian's color and the key vectors of its neighbors, followed by a Softmax function. The final output is obtained as a weighted sum of the value vectors. This mechanism combines local geometry and viewpoint information, learning dynamic feature fusion patterns that improve view-dependent color estimation.

**Kolmogorov-Arnold Networks (KAN):** The traditional multilayer perceptron (MLP) is a fundamental component in deep learning models, typically used to approximate complex multivariable functions through multiple layers of nonlinear mappings. An MLP consists of a series of weight matrices  $W$  and activation functions  $\sigma$  applied in alterna-

tion, and its structure can be expressed as:

$$\text{MLP}(Z) = (W_{K-1} \circ \sigma \circ W_{K-2} \circ \sigma \circ \cdots \circ W_1 \circ \sigma \circ W_0)Z$$

However, in high-fidelity 3D reconstruction tasks, the high computational complexity and limited interpretability of MLPs constrain their effectiveness in handling high-dimensional features, making it challenging for the model to capture intricate details and fine structures.

In contrast, the Kolmogorov-Arnold Network (KAN) design, based on the Kolmogorov-Arnold representation theorem, provides a modular and physically consistent alternative to MLPs. Like MLPs, KANs employ a fully connected structure, but unlike MLPs, which use fixed activation functions, KANs utilize learnable activation functions on edges, allowing adaptive feature mapping. A  $K$ -layer KAN network can be represented as a nested composition of multiple KAN layers:

$$\text{KAN}(Z) = (\Phi_{K-1} \circ \Phi_{K-2} \circ \cdots \circ \Phi_1 \circ \Phi_0)Z$$

where each  $\Phi_i$  denotes the  $i$ -th layer of the network, consisting of multiple learnable activation functions. These activation functions operate on the edges and adjust dynamically according to input feature variations, significantly enhancing the model's expressiveness across different feature dimensions.

Based on this, we replaced the traditional MLP with KAN to process the two-dimensional feature sequence

formed by concatenating the anchor feature  $\hat{f}_v$ , the relative viewing distance  $\delta_{vc}$ , and the direction vector  $\vec{d}_{vc}$  between the camera and the anchor point. In opacity prediction, we applied a Tanh activation function to the KAN output and set a threshold to retain effective neural Gaussian distributions. KAN demonstrates superior stability in opacity prediction, effectively preventing the premature removal of critical Gaussian distributions due to low opacity thresholds. Through KAN’s precise scaling predictions, the model dynamically adjusts the scale of each Gaussian distribution according to the anchor feature, viewing angle, and position, allowing it to flexibly adapt to the complex geometric structures within the scene and to preserve fine textures and structural details in the reconstruction. Similarly, KAN offers precise control over rotation parameters, enabling each Gaussian distribution to adapt its orientation based on the viewing angle, thus achieving a more realistic and accurate capture of spatial relationships and view-dependent occlusions. By finely controlling opacity, scaling, and rotation, KAN effectively retains complex textures, thin structures, and subtle details within the scene, significantly enhancing the resolution and quality of the reconstruction results. The integration of KAN substantially improves the model’s adaptability to view-dependent changes, aiding in the accurate capture of dynamic occlusions in the scene.

#### **Neural Laplacian Pyramid Decomposition (NLPD):**

In image rendering and 3D reconstruction, traditional loss functions such as  $L_1$  loss and SSIM often have limitations in capturing detail and texture consistency, particularly in terms of multi-scale perceptual sensitivity. To address this issue, we introduce the Normalized Laplacian Pyramid Distance (NLPD) as a perceptual loss metric to capture multi-scale detail discrepancies between rendered and real images, thereby improving the model’s performance in texture fidelity and visual consistency.

Specifically, we construct a multi-scale Laplacian pyramid for both the rendered and real images. By decomposing the images layer by layer into different spatial frequencies, we can extract edge and texture information at multiple scales, allowing us to measure perceptual differences with greater accuracy. At each level of the pyramid, we apply a Laplacian filter to extract scale-specific features, generating residual signals that reflect fine-grained detail differences. To mitigate the effects of local brightness and contrast variations, we further apply contrast normalization to the residual signals, stabilizing them across scales and preventing error amplification.

After contrast normalization, we compute the root mean square (RMS) of the residual signals at each scale to quantify perceptual differences. We then average the RMS values across all pyramid levels to obtain the overall NLPD loss. This multi-scale loss measure effectively captures

detail discrepancies across spatial frequencies, enhancing both global structural similarity and local detail consistency, making the rendered images visually closer to real scenes.

## 4. Experiments

### 4.1. Setups

#### 4.1.1 Datasets and Metrics

We conduct a comprehensive evaluation of our proposed model on seven scenes from Mip-NeRF360 [1], two scenes from Tanks&Temples [6], two scenes from DeepBlending [3]. Consistent with methodologies outlined in previous works [1, 5], we adopted a train/test split approach where every 8th photo was selected for testing. To assess the rendering quality, we further measured the average Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) [19], and Learned Perceptual Image Patch Similarity (LPIPS) [23].

#### 4.1.2 Implementation Details

Our method is built upon the open-source Scaffold-GS codebase [16] and we assesse the quality of our approach by comparing it to current state-of-the-art(SOTA) baselines, including Instant-NGP [18], Plenoxels [2], Mip-NeRF360 [1], 3D-GS [5], Mip-splatting [22] and Scaffold-GS [16]. In line with the approach described in 3D-GS [5], our models undergo training for 30K iterations across all scenes. All experiments were conducted on an RTX 4090 GPU with 24GB of memory.

## 4.2. Comparisons

### 4.2.1 Quantitative Comparisons.

Table 1 presents the results on three real-world datasets [1, 3, 6], highlighting that our approach achieves competitive performance in rendering quality compared to state-of-the-art methods and stands out in terms of image quality and perceived similarity. Although Scaffold-GS [16] achieves the highest PSNR performance on the Deep Blending dataset, our method demonstrates more stable and comprehensive performance across other datasets. It can be noticed that our approach achieves comparable results with 3DGs on Mip-NeRF360 dataset, and achieves superior results in Tanks&Templates datasets, exhibiting notable competitiveness in terms of SSIM and LPIPS, which captures more challenging environments with the presence of *e.g.* specular reflections, saturation and brightness under the bright light and thin structures.

### 4.2.2 Qualitative Comparisons.

Qualitative results across various datasets are provided in Figure 3. We also conduct a more comprehensive and de-

Table 1. Quantitative comparison to previous methods on real-world datasets. Competing metrics are extracted from respective papers.

Dataset Method	Metrics	Mipnerf-360			Tanks&Templates			Deep Blending		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Instant-NGP		26.43	0.725	0.339	21.72	0.723	0.330	23.62	0.797	0.423
Plenoxels		23.62	0.670	0.443	21.08	0.719	0.379	23.06	0.795	0.510
Mip-NeRF360		29.23	0.844	0.207	22.22	0.759	0.257	29.40	0.901	0.245
3D-GS		28.69	0.870	0.182	23.14	0.841	0.183	29.41	0.903	0.243
Mip-Splatting		29.12	0.869	0.184	23.79	0.848	0.176	29.57	0.900	0.245
Scaffold-GS		29.35	0.870	0.188	24.14	0.853	0.175	30.26	0.909	0.252
Ours		29.80	0.870	0.183	24.79	0.862	0.163	30.00	0.910	0.249

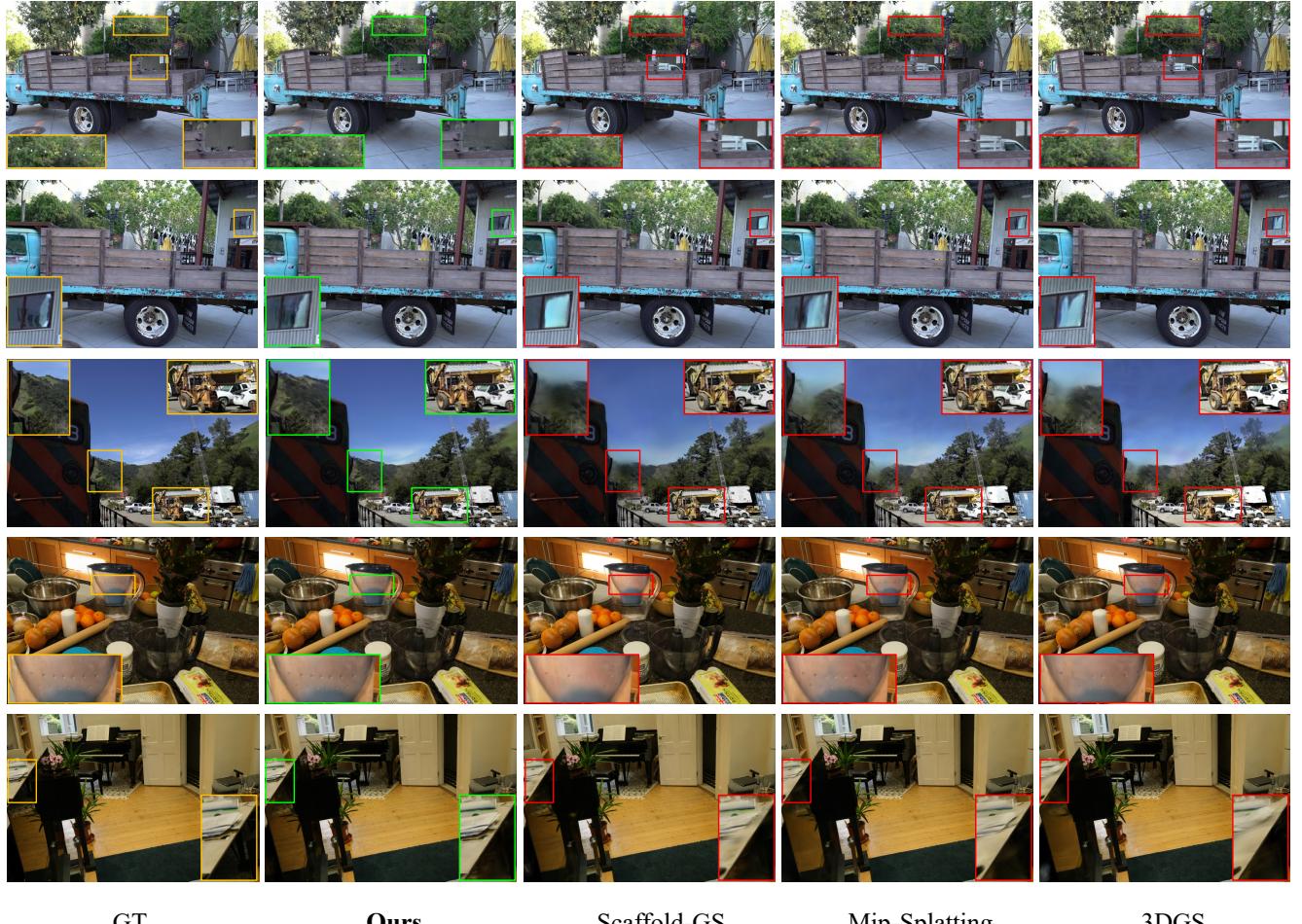


Figure 3. Qualitative comparison of PEP-GS , Scaffold-GS [16], Mip-splatting [22], 3D-GS [5] across diverse datasets [1, 3, 6]. Patches that highlight the visual differences are emphasized with arrows and green & yellow insets for clearer visibility. Compared to existing base-lines, Our method demonstrates superior detail preservation, reduced artifacts, improved color consistency, and better handling of specular reflections across various scenes.

tailed comparison of rendering results against the Scaffold method. As shown in Figure 4, it is evident that our method demonstrates sharper details and textures, effectively preserving local structures and fine-scale details during image reconstruction. As depicted in Figure 5, under challeng-

ing sunlight environments, the Scaffold-GS method exhibits slightly excessive brightness, leading to the loss of certain details. In contrast, our method not only accurately maintains clearer textures and structural features but also avoids overexposure or shadow detail loss. Figure 6 and 1 further

illustrate that in challenging scenes like grass fields, our method better suppresses artifacts and unnatural textures while adeptly capturing complex textures and thin structures, demonstrating superior stability. In terms of metric analysis, our method achieves lower LPIPS values and higher SSIM scores, confirming its significant advantage in overall visual quality and perceptual consistency.



Figure 4. Comparison of fine-scale details in the DrJohnson scene within the Deep Blending dataset.



Figure 5. Comparison of brightness and saturation in the train scene within the Tanks&Templates dataset.

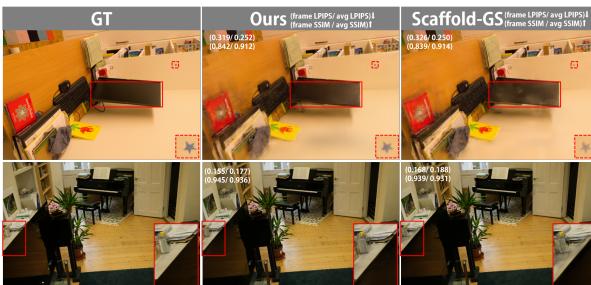


Figure 6. Comparison of texture-less regions and thin structures in the playroom scene within the Deep Blending dataset.

## 5. Conclusion and Future Work

In this paper, we presented PEP-GS, a perceptually-enhanced framework for structured 3D Gaussian splatting that advances the state-of-the-art in neural scene representation and view-adaptive rendering. Our approach introduces three key innovations: Local-Enhanced Multi-head

Self-Attention, Kolmogorov-Arnold Networks and Neural Laplacian Pyramid Decomposition. These components work together to address critical limitations in existing methods regarding perceptual consistency, view-dependent effects, and splatting precision. Experimental results across multiple datasets demonstrate significant improvements, compared to the current state-of-the-art methods, these improvements are particularly evident in challenging scenarios such as view-dependent effects, specular reflections, fine-scale details and false geometry generation.

Looking forward, several promising research directions emerge. Drawing inspiration from federated learning approaches [8, 9] and peer-assisted learning systems [11], we plan to explore collaborative learning frameworks for multi-scene optimization. Integration with edge computing and elastic systems [12, 15] could enable more efficient distributed rendering, while secure data sharing mechanisms [24] could facilitate large-scale scene collection and sharing. The success of contribution measurement methods [21] and dynamic network optimization [13] suggests potential improvements in adaptive parameter tuning. Furthermore, insights from traffic flow prediction [7], medical imaging [10], and smart city applications [25] could inform the development of domain-specific optimizations. We also envision integrating large language models [14] for semantic scene understanding and exploring real-time optimization techniques to further enhance rendering quality while maintaining computational efficiency.

## References

- [1] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5460–5469, 2022. 2, 4, 5
- [2] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5501–5510, 2022. 2, 4
- [3] Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics (ToG)*, 37(6):1–15, 2018. 4, 5
- [4] Peter Hedman, Pratul P Srinivasan, Ben Mildenhall, Jonathan T Barron, and Paul Debevec. Baking neural radiance fields for real-time view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5875–5884, 2021. 2
- [5] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. 2, 4, 5
- [6] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene

- reconstruction. *ACM Transactions on Graphics*, 36(4), 2017. 4, 5
- [7] Boyi Liu, Jieren Cheng, Kuanqi Cai, Pengchao Shi, and Xiangyan Tang. Singular point probability improve lstm network performance for long-term traffic flow prediction. In *Theoretical Computer Science: 35th National Conference, NCTCS 2017, Wuhan, China, October 14-15, 2017, Proceedings*, pages 328–340. Springer, 2017. 6
- [8] Boyi Liu, Lujia Wang, Ming Liu, and Cheng-Zhong Xu. Federated imitation learning: A novel framework for cloud robotic systems with heterogeneous sensor data. *IEEE Robotics and Automation Letters*, 5(2):3509–3516, 2019. 6
- [9] Boyi Liu, Lujia Wang, Ming Liu, and Cheng-Zhong Xu. Lifelong federated reinforcement learning: a learning architecture for navigation in cloud robotic systems. *IEEE Robotics and Automation Letters*, 4(4):4555–4562, 2019. 6
- [10] Boyi Liu, Bingjie Yan, Yize Zhou, Yifan Yang, and Yixian Zhang. Experiments of federated learning for covid-19 chest x-ray images. *arXiv preprint arXiv:2007.05592*, 2020. 6
- [11] Boyi Liu, Lujia Wang, Xinquan Chen, Lexiong Huang, Dong Han, and Cheng-Zhong Xu. Peer-assisted robotic learning: a data-driven collaborative learning approach for cloud robotic systems. In *2021 IEEE international conference on robotics and automation (ICRA)*, pages 4062–4070. IEEE, 2021. 6
- [12] Boyi Liu, Lujia Wang, and Ming Liu. Elasticros: An elastically collaborative robot operation system for fog and cloud robotics. *arXiv preprint arXiv:2209.01774*, 2022. 6
- [13] Boyi Liu, Lujia Wang, and Ming Liu. Roboec2: A novel cloud robotic system with dynamic network offloading assisted by amazon ec2. *IEEE Transactions on Automation Science and Engineering*, 2023. 6
- [14] Boyi Liu, Jingwen Tong, and Jun Zhang. Llm-slice: Dedicated wireless network slicing for large language models. *arXiv preprint arXiv:2410.18499*, 2024. 6
- [15] Boyi Liu, Jingwen Tong, and Yufan Zhuang. Edgeloc: A communication-adaptive parallel system for real-time localization in infrastructure-assisted autonomous driving. *arXiv preprint arXiv:2405.12120*, 2024. 6
- [16] Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20654–20664, 2024. 2, 4, 5
- [17] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, pages 405–421. Springer, 2020. 2
- [18] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4):1–15, 2022. 2, 4
- [19] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 4
- [20] Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. Point-nerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5438–5448, 2022. 2
- [21] Bingjie Yan, Boyi Liu, Lujia Wang, Yize Zhou, Zhixuan Liang, Ming Liu, and Cheng-Zhong Xu. Fedcm: A real-time contribution measurement method for participants in federated learning. In *2021 International joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2021. 6
- [22] Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3d gaussian splatting. *arXiv preprint arXiv:2311.16493*, 2023. 4, 5
- [23] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 4
- [24] Shenhui Zhang, Wenkai Li, Xiaoqi Li, and Boyi Liu. Authros: Secure data sharing among robot operating systems based on ethereum. In *2022 IEEE 22nd International Conference on Software Quality, Reliability and Security (QRS)*, pages 147–156. IEEE, 2022. 6
- [25] Zhaohua Zheng, Yize Zhou, Yilong Sun, Zhang Wang, Boyi Liu, and Keqiu Li. Applications of federated learning in smart cities: recent advances, taxonomy, and open challenges. *Connection Science*, 34(1):1–28, 2022. 6