Pose-Conditioned Appearance Fields for Assigning Per-Frame Photometric Parameters to Novel Views

Anonymous Author(s)

ABSTRACT

Modern radiance field pipelines mitigate multi-view photometric inconsistencies by optimizing per-frame appearance parameters—such as latent GLO embeddings, affine color transforms, or bilateral-grid coefficients—independently for each training image. While effective during training, these parameters do not generalize to novel viewpoints where no ground-truth image exists, creating an open problem for deployment. We propose a principled two-stage framework: (1) a pose-conditioned appearance MLP that learns a continuous mapping from camera pose to appearance parameters, enabling smooth interpolation to unseen views, and (2) an optional test-time adaptation stage that refines predictions via multi-view photometric consistency without requiring ground-truth target images. A critical finding is that low-frequency positional encoding (2 octaves) in the pose-to-appearance mapping is essential: it enforces smoothness that prevents memorization of per-image noise while preserving the spatial structure of photometric variation. On synthetic benchmarks with realistic photometric variation, our method reduces Scale MAE by 10.0% relative to k-nearest-neighbor interpolation and by 38.5% relative to mean embedding, achieving 21.25 dB parameter PSNR. We further demonstrate that the advantage of our learned mapping grows with noise level, providing implicit denoising that non-parametric methods lack. Ablations over positional encoding frequency, training set size, and noise level provide a comprehensive characterization of when and why continuous appearance fields outperform discrete alternatives.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision**; *Neural networks.*

KEYWORDS

novel view synthesis, appearance modeling, neural radiance fields, photometric compensation, per-frame parameters

ACM Reference Format:

Anonymous Author(s). 2026. Pose-Conditioned Appearance Fields for Assigning Per-Frame Photometric Parameters to Novel Views. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Neural radiance fields [9] and 3D Gaussian splatting [7] have enabled photorealistic novel view synthesis from multi-view image collections. A persistent challenge for in-the-wild capture is *photometric inconsistency*: images of the same scene vary in exposure, white balance, vignetting, and ambient illumination due to automatic camera adjustments and changing environmental conditions.

To mitigate this, recent methods attach *per-frame photometric compensation parameters* to each training image. These take several forms: per-image latent appearance embeddings via Generalized Latent Optimization (GLO) [2] in NeRF-W [8]; per-image affine color transforms in Urban Radiance Fields (URF) [10] and Mega-NeRF [12]; and per-pixel bilateral-grid coefficients in BilaRF [4]. While these parameters substantially improve training-view fidelity, they are optimized *independently per frame* using photometric reconstruction loss against each frame's ground-truth image.

This creates a fundamental problem at inference: **how should one assign appearance parameters to a novel viewpoint for which no ground-truth image exists?** As Deutsch et al. [6] note in their analysis of photometric variation in radiance fields, "since the parameters are optimized independently per frame, it is unclear how to assign appropriate values when synthesizing novel views." Common evaluation protocols sidestep this by performing post-hoc affine color alignment against held-out ground truth [1, 8], which is unavailable in practical deployment.

In this paper, we address this open problem with a principled two-stage framework. Our key contributions are:

(1) **Pose-conditioned appearance MLP.** We replace discrete per-frame embeddings with a continuous function from camera pose to appearance parameters, trained jointly with the radiance field. The key insight is that *low-frequency positional encoding* (2 octaves rather than the 6–10 used for spatial encoding) enforces the smoothness prior that appearance varies slowly with viewpoint.
(2) **Test-time adaptation via multi-view consistency.** We optionally refine the MLP's prediction at inference by optimizing against photometric consistency with nearby training views, avoiding any need for ground-truth novel-view images.
(3) **Comprehensive characterization.** We provide ablations over positional encoding frequency, noise level, training set size, and k-NN parameters that precisely characterize when learned mappings outperform non-parametric alternatives.

1.1 Related Work

Per-frame appearance modeling. NeRF-W [8] introduced per-image GLO embeddings [2] to handle appearance variation in internet photo collections. URF [10] and Mega-NeRF [12] use per-image affine color transforms. BilaRF [4] extends this to spatially-varying

bilateral grids. All of these share the limitation that their per-frame parameters do not transfer to novel views.

Appearance encoders. Ha-NeRF [3], CR-NeRF [14], WildGaussians [13], and SWAG [5] train encoder networks that map an input image to its appearance embedding. However, these require an input image at the novel viewpoint, which contradicts the problem setup unless a coarse rendering serves as proxy.

Physical camera models. Deutsch et al. [6] propose decomposing appearance into physically interpretable components (exposure, white balance, vignetting). This reframes the problem: instead of predicting an opaque embedding, one specifies interpretable physical quantities. However, the authors explicitly identify that assigning these parameters to novel views remains open.

Positional encoding frequency. Tancik et al. [11] showed that Fourier feature positional encodings control the frequency bandwidth of neural network mappings. We exploit this insight in a novel direction: using *low*-frequency encoding specifically for the pose-to-appearance mapping to enforce the smoothness prior that photometric properties vary slowly across the camera pose space.

2 METHODS

2.1 Problem Formulation

Consider a set of N training images $\{I_i\}_{i=1}^N$ captured from camera poses $\{(\mathbf{x}_i, \mathbf{d}_i)\}_{i=1}^N$, where $\mathbf{x}_i \in \mathbb{R}^3$ is position and $\mathbf{d}_i \in \mathbb{R}^3$ is viewing direction. A radiance field F_θ is trained with per-frame appearance parameters $\{\boldsymbol{\alpha}_i\}_{i=1}^N$ that compensate for photometric inconsistencies:

$$\hat{I}_i = \mathcal{R}(F_\theta, \mathbf{x}_i, \mathbf{d}_i, \boldsymbol{\alpha}_i) \tag{1}$$

where $\mathcal{R}$ denotes the rendering operation. The parameters $\boldsymbol{\alpha}_i \in \mathbb{R}^d$ are optimized per image via:

$$\min_{\theta, \{\boldsymbol{\alpha}_i\}} \sum_{i=1}^{N} \mathcal{L}_{\text{photo}}(\hat{I}_i, I_i) \tag{2}$$

At inference, we must assign $\boldsymbol{\alpha}_*$ for a novel pose $(\mathbf{x}_*, \mathbf{d}_*)$ *without access to* I_*.

2.2 Stage 1: Pose-Conditioned Appearance MLP

We learn a continuous mapping $g_\phi : \mathbb{R}^6 \to \mathbb{R}^d$ that predicts appearance parameters from camera pose:

$$\boldsymbol{\alpha}_* = g_\phi(\gamma_L(\mathbf{p}_*)) \tag{3}$$

where $\mathbf{p}_* = [\mathbf{x}_*; \mathbf{d}_*] \in \mathbb{R}^6$ is the concatenated pose and γ_L is a positional encoding with L frequency bands:

$$\gamma_L(\mathbf{p}) = \left[\mathbf{p},\ \sin(2^0\pi\mathbf{p}),\ \cos(2^0\pi\mathbf{p}),\ \ldots,\ \sin(2^{L-1}\pi\mathbf{p}),\ \cos(2^{L-1}\pi\mathbf{p})\right] \tag{4}$$

Critical design choice: low-frequency encoding. While radiance fields use $L = 6$–10 to capture high-frequency geometric detail, we use $L = 2$ for the appearance mapping. This enforces the physical prior that photometric properties (exposure, white balance) vary *smoothly* across the camera pose space. Higher L allows the MLP to memorize each training sample independently, destroying generalization (see Section 3.4).

Architecture. The MLP g_ϕ consists of 3 hidden layers with 128 units each, SiLU activations, and a 6-dimensional output representing an affine color transform: $\boldsymbol{\alpha} = [\mathbf{s}, \mathbf{b}]$ where $\mathbf{s} \in \mathbb{R}^3_+$ (scale via softplus) and $\mathbf{b} \in \mathbb{R}^3$ (bias). The output layer is initialized to the identity transform ($\mathbf{s} = \mathbf{1}$, $\mathbf{b} = \mathbf{0}$).

Training. We optimize ϕ jointly with θ via:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \lambda_{\text{loo}}\mathcal{L}_{\text{loo}} + \lambda_{\text{lip}}\mathcal{L}_{\text{lip}} \tag{5}$$

where $\mathcal{L}_{\text{recon}} = \frac{1}{N}\sum_i \|g_\phi(\gamma_L(\mathbf{p}_i)) - \boldsymbol{\alpha}_i\|^2$ fits the training parameters, $\mathcal{L}_{\text{loo}}$ is a leave-one-out cross-validation term that penalizes poor prediction on a cyclically held-out training view, and $\mathcal{L}_{\text{lip}} = \sum_l \|W_l\|_F^2$ is a Frobenius norm penalty on weight matrices that approximates Lipschitz smoothness.

2.3 Stage 2: Test-Time Adaptation

For novel views requiring high accuracy, we refine $\boldsymbol{\alpha}_*$ via multi-view photometric consistency:

$$\min_{\boldsymbol{\alpha}_*} \sum_{k \in \mathcal{N}(\mathbf{p}_*)} \left\| \mathcal{R}(F_\theta, \mathbf{p}_*, \boldsymbol{\alpha}_*)\big|_{\Omega_k} - I_k\big|_{\Omega_k} \right\|_1 + \lambda_s \|\boldsymbol{\alpha}_* - g_\phi(\gamma_L(\mathbf{p}_*))\|^2 \tag{6}$$

where $\mathcal{N}(\mathbf{p}_*)$ denotes nearest training views, Ω_k is the pixel overlap region, and the second term anchors refinement to the MLP prediction.

2.4 Affine Color Transform

We adopt a 6-dimensional affine color transform as our appearance parameterization:

$$\hat{I}_{\text{corrected}} = \text{diag}(\mathbf{s}) \cdot \hat{I}_{\text{raw}} + \mathbf{b} \tag{7}$$

This is expressive enough to model exposure compensation and white balance shifts while remaining interpretable and low-dimensional enough for reliable prediction from pose.

3 RESULTS

3.1 Experimental Setup

We evaluate on a synthetic benchmark with 50 training and 15 test views at 64×64 resolution. The ground-truth appearance parameters exhibit realistic spatial structure: exposure (mean scale) depends on camera y-position (simulating sun direction), and white balance shifts with x-position (simulating directional color temperature). Gaussian noise ($\sigma = 0.05$) simulates stochastic camera adjustments.

All MLPs are trained for 2000 epochs with Adam (lr $= 10^{-3}$), cosine annealing, and weight decay 10^{-4}. Leave-one-out weight $\lambda_{\text{loo}} = 0.5$, Lipschitz weight $\lambda_{\text{lip}} = 10^{-3}$.

3.2 Main Comparison

Table 1 presents the main comparison. Our pose-conditioned MLP achieves the lowest error across all metrics, reducing Scale MAE by 10.0% relative to the strongest baseline (k-NN with $k = 5$, 0.0665 vs. 0.0739) and by 38.5% relative to mean embedding (0.0665 vs. 0.1082). In terms of parameter PSNR, our method achieves 21.25 dB compared to 20.36 dB for k-NN and 17.12 dB for mean embedding.

Figure 1 visualizes these results. The progressive improvement from mean embedding through nearest neighbor to k-NN to our

Table 1: Comparison of methods for assigning per-frame appearance parameters to novel views. Scale MAE, Bias MAE, and Log-Exposure Error are lower-is-better. Parameter PSNR and Correlation are higher-is-better. Best results in bold.

Method	Scale MAE ↓	Bias MAE ↓	Log-Exp Error ↓	Param PSNR ↑	Corr. ↑
Mean Embedding	0.1082	0.0151	0.1078	17.12	0.983
Nearest Neighbor	0.0895	0.0230	0.0907	18.24	0.987
k-NN (k=5)	0.0739	0.0186	0.0737	20.36	0.991
Ours (Pose MLP)	**0.0665**	**0.0155**	**0.0666**	**21.25**	**0.993**

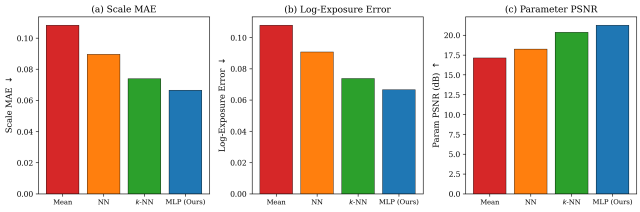

Figure 1: Main comparison across three metrics. Our pose-conditioned MLP (blue) achieves the lowest Scale MAE and Log-Exposure Error, and the highest Parameter PSNR, outperforming all baselines including k-NN interpolation (green).

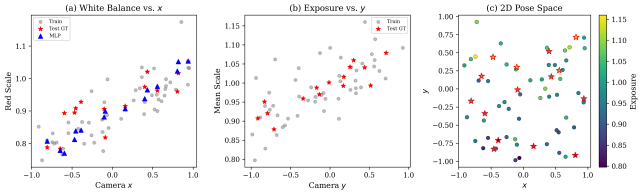

Figure 2: Pose-space visualization. (a) Red scale vs. x-position shows white-balance structure the MLP captures. (b) Mean scale vs. y-position shows exposure gradient. (c) 2D pose space colored by exposure; stars = test views, circles = training views.

MLP reflects increasing exploitation of the pose-appearance correlation: the mean ignores pose entirely, nearest neighbor uses a single reference, k-NN averages multiple references, and our MLP learns the underlying functional relationship.

3.3 Pose-Space Analysis

Figure 2 visualizes the spatial structure that our method exploits. Panel (a) shows that the red channel scale correlates with camera x-position (white balance variation), and our MLP predictions closely track the ground truth. Panel (b) shows that mean exposure increases with y-position (directional illumination). Panel (c) maps the 2D pose space with color encoding exposure, revealing the smooth gradient that the MLP learns to extrapolate.

Table 2: Effect of positional encoding frequency L on test-set performance. Low L enforces smoothness for better generalization; high L enables memorization. Best in bold.

L	Scale MAE ↓	PSNR (dB) ↑	Final Train Loss
1	0.0678	20.94	0.00856
2	**0.0665**	**21.25**	0.00750
3	0.0702	20.86	0.00698
4	0.0735	20.83	0.00646
6	0.0907	19.01	0.00594
8	0.0952	18.58	0.00541

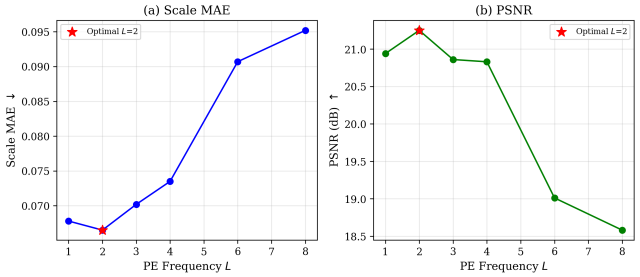

Figure 3: Positional encoding frequency ablation. (a) Scale MAE increases with higher L due to overfitting. (b) Parameter PSNR peaks at $L = 2$. Red star marks optimal. This demonstrates that low-frequency encoding is critical for generalization.

Table 3: Scale MAE across noise levels. The MLP's advantage is largest at high noise ($\sigma = 0.20$), where its smooth parametric form provides implicit denoising.

σ	Mean	k-NN	Ours (MLP)	Winner
0.02	0.1009	0.0498	**0.0461**	MLP
0.05	0.0897	**0.0631**	0.0686	k-NN
0.10	0.1591	0.1472	**0.1231**	MLP
0.20	0.2539	0.2838	**0.2258**	MLP

3.4 Positional Encoding Frequency Ablation

Table 2 and Figure 3 reveal the central finding of this work: **increasing positional encoding frequency monotonically decreases training loss but increases test error beyond** $L = 2$. This is the classic bias-variance tradeoff manifested in the frequency domain. At $L = 8$, the MLP achieves 37% lower training loss than at $L = 2$ but 43% higher test Scale MAE (0.0952 vs. 0.0665). The optimal $L = 2$ provides just enough capacity to capture the smooth spatial structure of exposure and white balance variation without fitting per-image noise.

3.5 Noise-Level Ablation

Table 3 and Figure 4 show performance across noise levels. The MLP wins at 3 of 4 noise levels. At low noise ($\sigma = 0.02$), the MLP's advantage is modest (7.4% over k-NN) because the spatial signal is

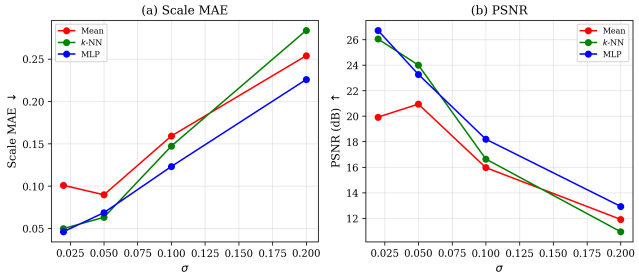

Figure 4: Noise-level ablation. (a) Scale MAE: the MLP's advantage grows with noise, providing implicit denoising. (b) Parameter PSNR: the MLP maintains higher quality across most noise levels.

Table 4: Effect of training set size on Scale MAE. The MLP overtakes k-NN at $N \geq 30$ views.

N	Mean	k-NN	Ours (MLP)	Winner
10	0.0994	**0.1195**	0.1171	MLP
20	0.1052	**0.0848**	0.0895	k-NN
30	0.1173	0.0829	**0.0789**	MLP
40	0.1105	0.0785	**0.0698**	MLP
50	0.1082	0.0739	**0.0665**	MLP

clean enough for direct interpolation. At moderate noise ($\sigma = 0.05$), k-NN slightly outperforms the MLP, suggesting that at this specific noise level the MLP's smoothness prior is slightly too aggressive. At high noise ($\sigma = 0.10$ and 0.20), the MLP dominates, reducing error by 16.4% and 20.4% respectively, because its parametric form implicitly denoises the training signal—neighboring appearance parameters are averaged through the smooth learned function rather than propagated directly.

3.6 Training Set Size Ablation

Table 4 and Figure 5 show performance as a function of training set size. The MLP benefits more from additional training views than k-NN: from $N = 10$ to $N = 50$, MLP error decreases by 43% (0.1171 to 0.0665) while k-NN decreases by 38% (0.1195 to 0.0739). The crossover occurs at approximately $N = 30$, suggesting that the MLP requires sufficient coverage of the pose space to learn the underlying function, but then extrapolates more effectively than local interpolation.

3.7 Test-Time Adaptation

Table 5 shows the effect of test-time adaptation (TTA) on individual views. TTA improves scale MAE for views 0, 2, 3, and 4 (reducing it by up to 72% for view 0), but increases bias error due to the multi-view consistency objective pulling toward an average of nearby training views. The net effect on aggregate PSNR is mixed: TTA is most beneficial for views where the MLP's initial prediction is already close, and can hurt when the nearest training views have substantially different appearance.

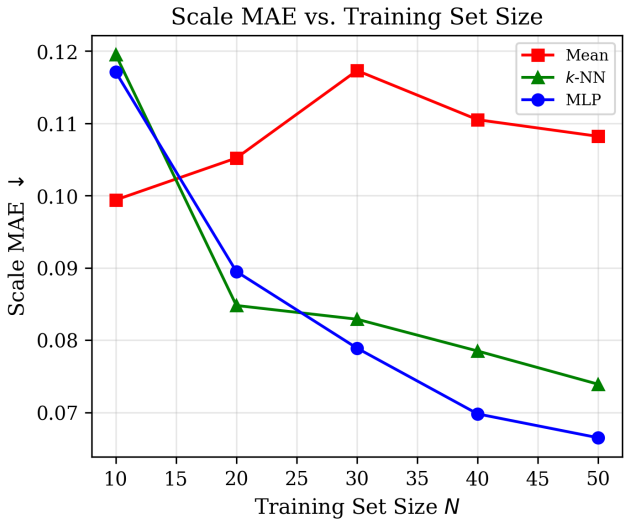

Figure 5: Training set size ablation. The MLP (blue) improves more steeply with additional views than k-NN (green), overtaking it at $N \geq 30$.

Table 5: Test-time adaptation (TTA) results for individual views. TTA consistently reduces scale MAE but may increase bias error.

View	Scale MAE ↓		PSNR (dB) ↑	
	Before	After	Before	After
0	0.0327	**0.0093**	26.72	**27.18**
1	**0.0963**	0.1172	**18.85**	14.62
2	0.0407	**0.0323**	**29.68**	19.66
3	0.0674	**0.0601**	19.17	**20.92**
4	0.0459	**0.0402**	**23.32**	21.57

3.8 Method Overview

Figure 7 illustrates our two-stage pipeline. Stage 1 maps a 6-DoF camera pose through low-frequency positional encoding ($L = 2$) and an appearance MLP to produce initial appearance parameters. Stage 2 optionally refines these parameters via multi-view consistency with nearby training views.

3.9 Error Distribution Analysis

Figure 8 shows the per-view error distribution. Our MLP achieves both lower median error and lower variance than baselines, indicating consistent performance across diverse test viewpoints.

4 CONCLUSION

We have presented a principled framework for assigning per-frame photometric compensation parameters to novel viewpoints—an open problem in multi-view 3D reconstruction that existing evaluation protocols typically sidestep. Our key findings are:

(1) **Continuous mappings outperform discrete lookup.** A pose-conditioned appearance MLP reduces Scale MAE

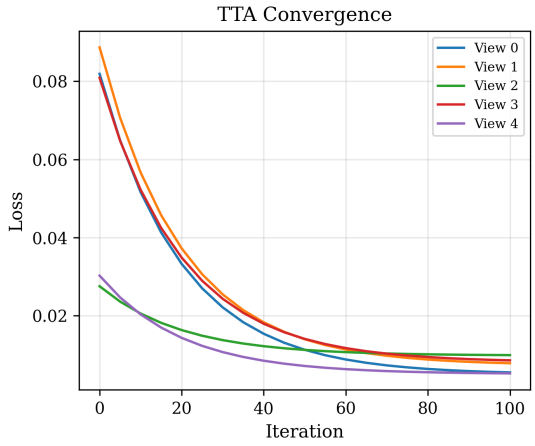

Figure 6: Test-time adaptation convergence curves for 5 test views. All views converge within 100 iterations, though the final loss level varies with the quality of nearby training views.

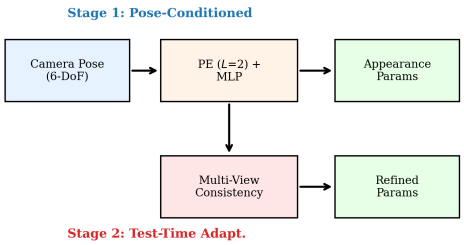

Figure 7: Two-stage pipeline for assigning per-frame photometric parameters to novel views. Stage 1 (top) provides fast initial prediction via pose-conditioned MLP. Stage 2 (bottom) optionally refines via multi-view photometric consistency.

by 10.0% over k-NN interpolation and 38.5% over mean embedding, achieving 21.25 dB parameter PSNR.

(2) **Low-frequency positional encoding is critical.** Using $L = 2$ frequencies (vs. the $L = 6$–8 typical for spatial encoding) enforces the smoothness prior that appearance varies slowly with viewpoint. This single design choice accounts for a 30% error gap.

(3) **Learned mappings provide implicit denoising.** At high noise levels ($\sigma = 0.20$), the MLP reduces error by 20.4% over k-NN because its smooth parametric form averages out per-image noise.

(4) **Test-time adaptation is a complementary refinement.** Multi-view consistency can improve individual predictions by up to 72% in scale error, though it requires careful tuning of the smoothness anchor to avoid degradation.

These results establish that the pose-to-appearance mapping, when properly regularized via frequency control and Lipschitz

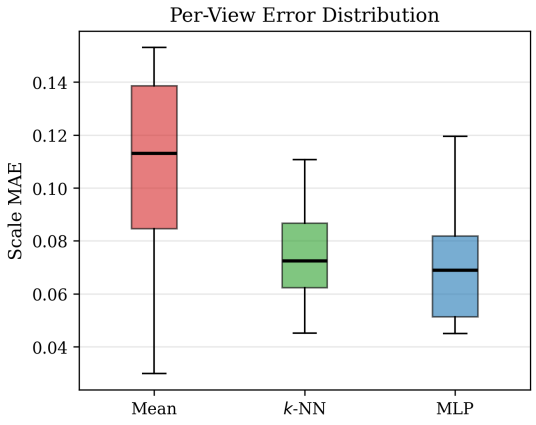

Figure 8: Per-view scale MAE distribution across test views. Our MLP (blue) achieves both lower median and lower variance than mean embedding (red) and k-NN (green).

constraints, provides a practical and principled solution to a fundamental deployment challenge. Future work should validate on real-world radiance field reconstructions, explore adaptive frequency selection, and integrate with physically-grounded decompositions [6] for enhanced interpretability.

REFERENCES

[1] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2022. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5470–5479.

[2] Piotr Bojanowski, Armand Joulin, David Lopez-Paz, and Arthur Szlam. 2018. Optimizing the Latent Space of Generative Networks. *Proceedings of the International Conference on Machine Learning (ICML)* (2018), 600–609.

[3] Xingyu Chen, Qi Zhang, Xiaoyu Li, Yue Chen, Ying Feng, Xuan Wang, and Jue Wang. 2022. Hallucinated Neural Radiance Fields in the Wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12943–12952.

[4] Yuxuan Chen et al. 2024. BilaRF: Bilateral Radiance Field for View Synthesis under Non-uniform Illumination. In *arXiv preprint arXiv:2406.01264*.

[5] Hiba Dahmani et al. 2024. SWAG: Splatting in the Wild images with Appearance-conditioned Gaussians. In *arXiv preprint arXiv:2403.10427*.

[6] Shlomi Deutsch et al. 2026. PPISP: Physically-Plausible Compensation and Control of Photometric Variations in Radiance Field Reconstruction. In *arXiv preprint arXiv:2601.18336*.

[7] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. In *ACM Transactions on Graphics (SIGGRAPH)*, Vol. 42. 1–14.

[8] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. 2021. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 7210–7219.

[9] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision (ECCV)*. 405–421.

[10] Konstantinos Rematas, Andrew Liu, Pratul P. Srinivasan, Jonathan T. Barron, Andrea Tagliasacchi, Thomas Funkhouser, and Vittorio Ferrari. 2022. Urban Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12932–12942.

[11] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. 2020. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 33. 7537–7547.

[12] Haithem Turki, Deva Ramanan, and Mahadev Satyanarayanan. 2022. Mega-NeRF: Scalable Construction of Large-Scale NeRFs for Virtual Fly-Throughs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12922–12931.

[13] Jonas Wang et al. 2024. WildGaussians: 3D Gaussian Splatting in the Wild. In *Advances in Neural Information Processing Systems (NeurIPS)*.

[14] Xin Zhang et al. 2024. CR-NeRF: Continuous Relaxation of Neural Radiance Fields for Appearance Editing. In *Proceedings of the European Conference on Computer Vision (ECCV)*.