MomentsNeRF: Incorporating Orthogonal Moments in Convolutional Neural Networks for One or Few-Shot Neural Rendering

Ahmad AlMughrabi[1] Ricardo Marques[1] Petia Radeva[1]

UNIVERSITAT DE BARCELONA

BCN-AIM ARTIFICIAL INTELLIGENCE IN MEDICINE

[1]Universitat de Barcelona

Overview

Neural Radiance Fields (NeRF) are powerful for learning 3D scene representations for photo-realistic view synthesis.

Why is one/few-shot neural rendering **difficult**?

- **Rendering problems**: blurry rendering; missing textures; showing artifacts; missing scene information; incorrect colors; reflection issues.

Why MomentsNeRF (**Our Motivation**)?

- Prioritizing the **robustness of the learned feature** representations in NeRF.
- Improving the ability of NeRF to **generalize** across multiple scenes.

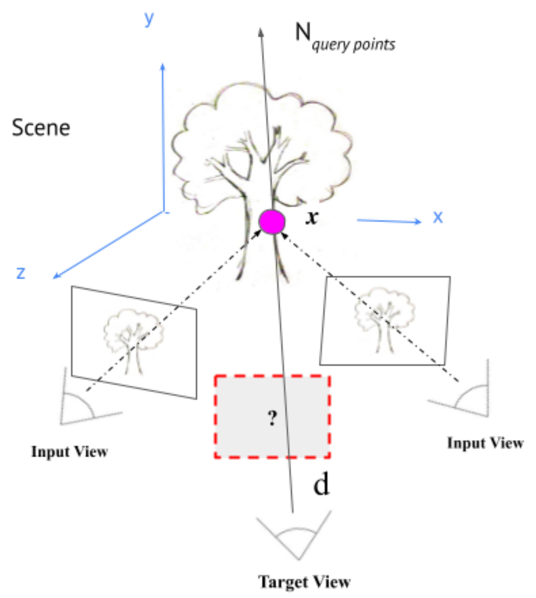
What is MomentsNeRF?

- A framework combines multi-view stereo cost volumes with physically based volume rendering for neural radiance field reconstruction. The cost volumes use orthogonal moments to extract robust features for MLP learning. Our framework is trained and tested on DTU's real object dataset, producing realistic view synthesis with just one input image. It outperforms concurrent techniques in generalizable few-shot neural rendering, offering superior rendering quality and reduced optimization time compared to other models. Our framework comprises three distinct phases: 1. **Cost Volume**, 2. **Moments Neural encoding Volume**, and 3. **Volume Rendering**

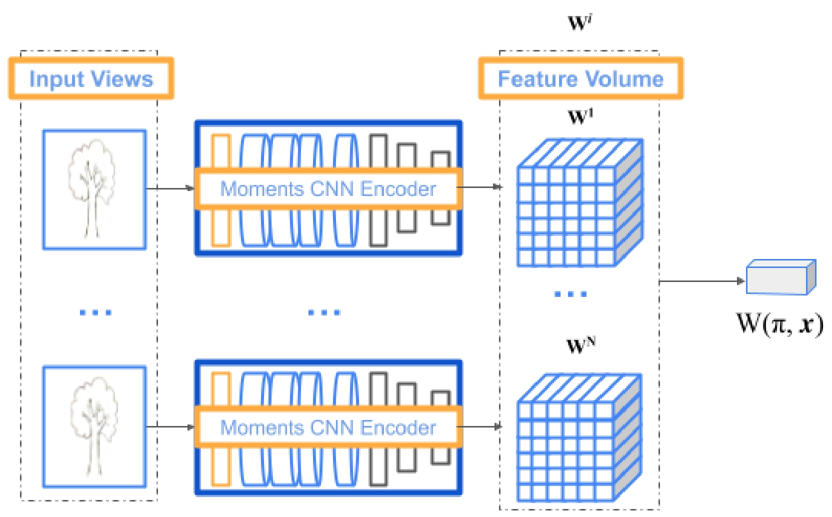
Applications: Digital twins; Augmented and Virtual Reality; Gaming; 3D reconstruction, and more.

1. Cost Volume

- Performs ray casting.

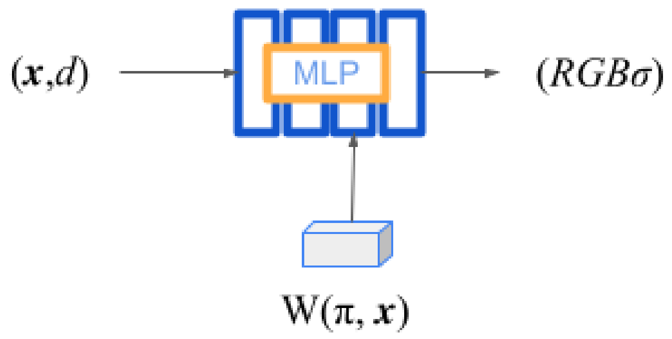


- Applies bilinear interpolation on pixel-wise features to extract a moment feature vector.



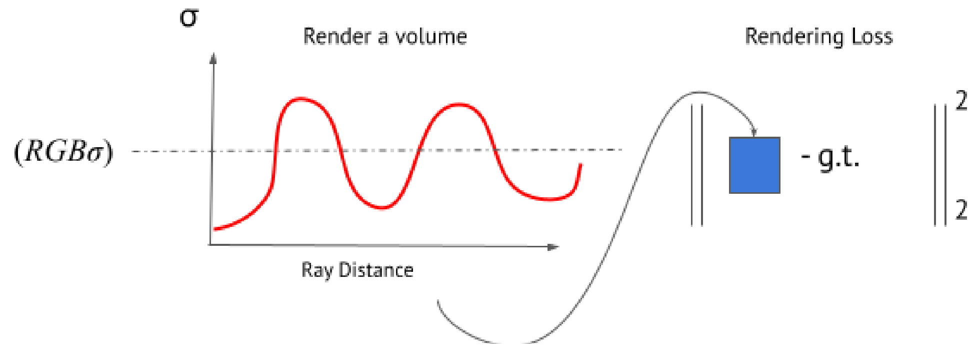
2. Moments Neural Encoding Volume

- Accepts a query point x, a viewing direction d, and the projected features from different feature volumes W that is fed to a MLP.



3. Volume Rendering

- Accepts the output of the MLP: $c = RGB$ color and density values σ, which results in a synthesized image for the target view.



Results

	1 View				3 Views				6 Views				9 Views			
	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	PSNR↑	SSIM↑	LPIPS↓	DISTS↓
SRF	-	-	-	-	16.06	0.55	0.431	-	16.060	0.657	0.353	-	19.970	0.678	0.325	-
MVSNeRF	-	-	-	-	16.26	0.601	0.384	-	18.220	0.694	0.319	-	20.320	0.736	0.278	-
mip-NeRF	-	-	-	-	7.640	0.227	0.655	-	14.330	0.568	0.394	-	20.710	0.799	0.209	-
DietNeRF	-	-	-	-	10.01	0.354	0.574	-	18.700	0.668	0.336	-	22.160	0.740	0.277	-
RegNeRF	-	-	-	-	15.33	0.621	0.341	-	19.100	0.757	0.233	-	22.300	0.823	0.184	-
FreeNeRF	-	-	-	-	18.02	0.68	0.318	-	22.390	0.779	0.24	-	24.200	0.833	0.187	-
PixelNeRF	15.311	0.523	0.555	0.339	18.990	0.678	0.395	0.251	19.962	0.713	0.347	0.233	20.471	0.734	0.307	0.210
Ours	21.543	0.729	0.186	0.178	23.810	0.828	0.138	0.169	24.443	0.847	0.131	0.173	24.655	0.855	0.127	0.174

Table 1:A quantitative comparison of our model with the SOTA on the DTU dataset. The best results are marked with red, the second best results marked with orange, while the third best results marked by yellow.

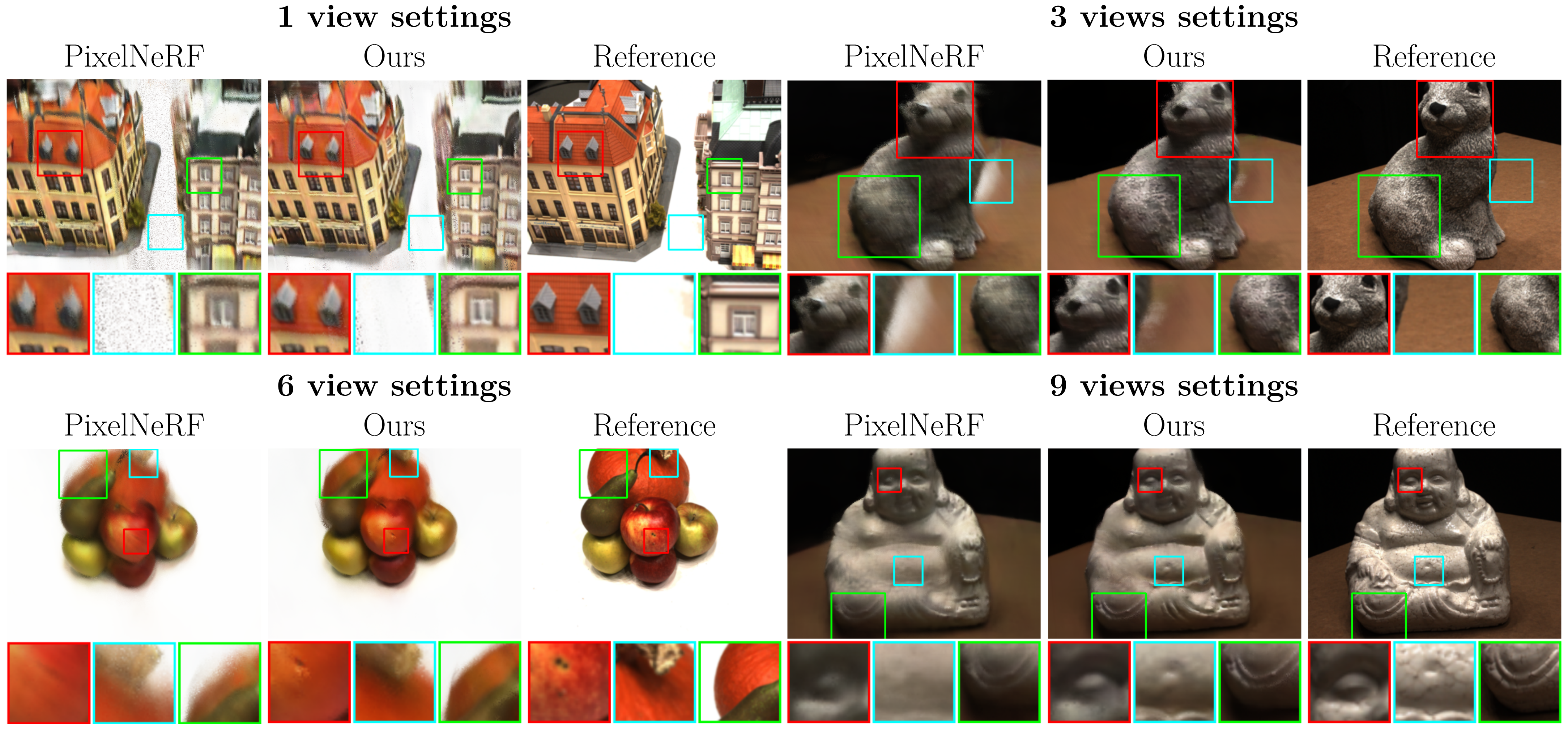


Figure 1:Qualitative comparisons for scan21, scan55, scan63, and scan114 scene in 1, 3, 6, and 9 views settings.

	1 View				3 Views				6 Views				9 Views			
	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	PSNR↑	SSIM↑	LPIPS↓	DISTS↓	PSNR↑	SSIM↑	LPIPS↓	DISTS↓
Ours	21.543	0.729	0.186	0.178	23.810	0.828	0.138	0.169	24.443	0.847	0.131	0.173	24.655	0.855	0.127	0.174
w/o Zernike	17.337	0.386	0.733	0.459	23.794	0.808	0.511	0.312	24.445	0.836	0.435	0.280	24.565	0.842	0.385	0.252
w/o Gabor	17.107	0.370	0.366	0.321	23.419	0.795	0.156	0.181	24.214	0.834	0.137	0.182	24.597	0.846	0.132	0.183
w/o PE&AF	17.263	0.376	0.356	0.317	23.992	0.818	0.145	0.174	24.540	0.845	0.134	0.181	24.641	0.850	0.131	0.179

Table 2:An ablation quantitative comparison of our model with removing different components on the DTU dataset. The best results are marked with red, the second best results marked with orange, while the third best results marked by yellow.

Cont. Results

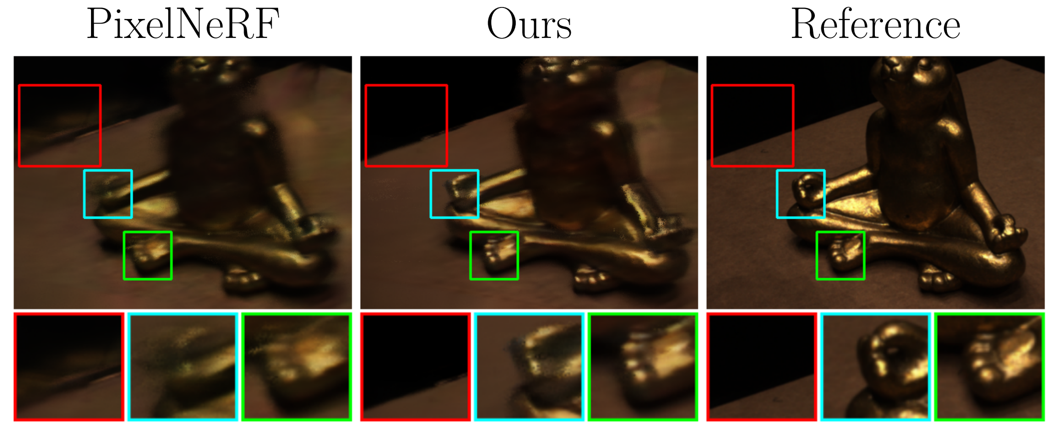


Figure 2:Additional Qualitative comparisons for scan110 scene in 3 views settings.

Conclusions

MomentsNeRF improves the existing approaches by **3.39 dB PSNR**, **11.1% SSIM**, **17.9% LPIPS**, and **8.3% DISTS** metrics (Table 1 and Fig. 2). Moreover, MomentsNeRF excels in texture details, artifact correction, missing data handling, and color adjustment across different scene parts, surpassing other models. An essential and interesting question —*how the MomentsNeRF for a 360 scene impact the robustness of NeRF*— is still open. Moreover, more feature extraction methods can be integrated to extend our framework.

References

[1] A. Chen, Z. Xu, F. Zhao, X. Zhang, F. Xiang, J. Yu, and H. Su. Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In *Proceedings of the IEEE/CVF in ICCV2023*, pages 14124–14133, 2021.

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

[3] J. Yang, M. Pavone, and Y. Wang. Freenerf: Improving few-shot neural rendering with free frequency regularization. In *Proceedings of the IEEE/CVF in CVPR2023*, pages 8254–8263, 2023.

[4] A. Yu, V. Ye, M. Tancik, and A. Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF in CVPR2023*, pages 4578–4587, 2021.

Acknowledgements

This work is partially funded by the Horizon EU project MUSAE (No. 01070-421), 2021-SGR-01094 (AGAUR), DeepFoodVol (AEI-MICINN, PDC2022-133642-I00), PID2022-141566NB-I00 (AEI-MICINN).

Check out our project page for more details and discussions!

Code & Model also available!!!

https://amughrabi.github.io/momentsnerf