EXPLORING METRIC FUSION FOR EVALUATION OF NeRFs

Shreyas Shivakumara, Gabriel Eilertsen, Karljohan Lundin Palmerius

Department of Science and Technology, Linköping University, Norrköping, Sweden

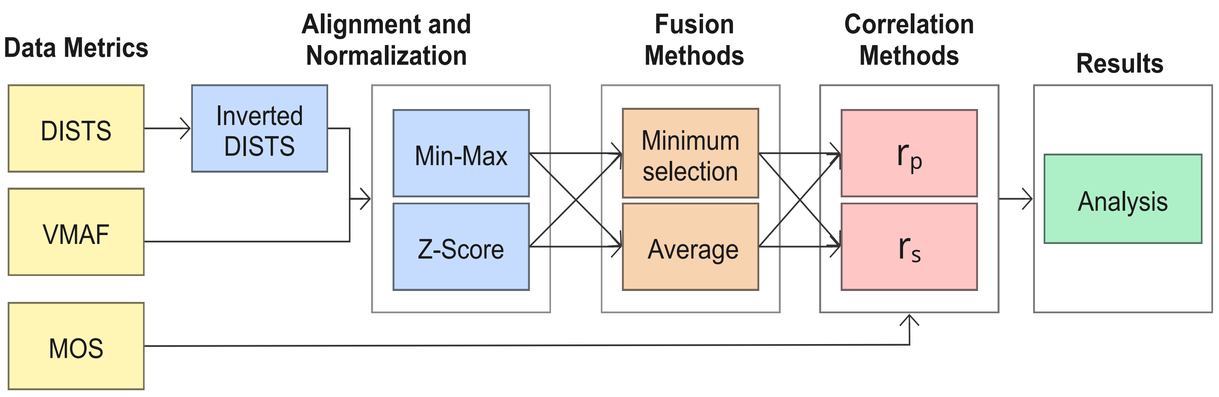
BACKGROUND

- Neural Radiance Fields (NeRFs) enable novel views synthesis by learning a continuous volumetric representation of a scene from a sparse set of input images[1].
- Common quality metrics, such as Peak Signal-to-Noise Ratio and Structural Similarity Index Measure, often fail to accurately reflect perceptual quality, particularly for NeRF specific artifacts such as “floaters”, “ghosting effects“ or view-dependent artifacts.
- Literature highlights that perceptual metrics such as Video Multimethod Assessment Fusion (VMAF) and Deep Image Structure and Texture Similarity (DISTS) correlate better with human subjective judgments[2][3].

MAIN CONTRIBUTION

- We propose a metric fusion framework that combines DISTS and VMAF using normalization and fusion strategies.
- We demonstrate that the fused metrics achieve improved correlation with subjective quality scores compared to individual metrics.

FRAMEWORK



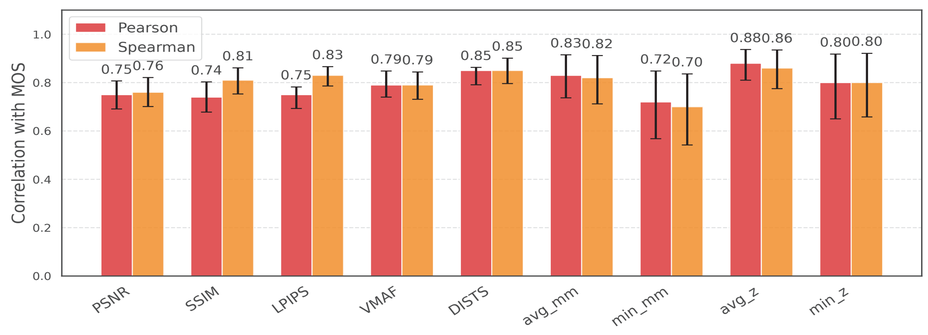
DATASET

Fig 1: Synthetic dataset example - Drum scene from *Explicit-NeRF-QA dataset* [2]

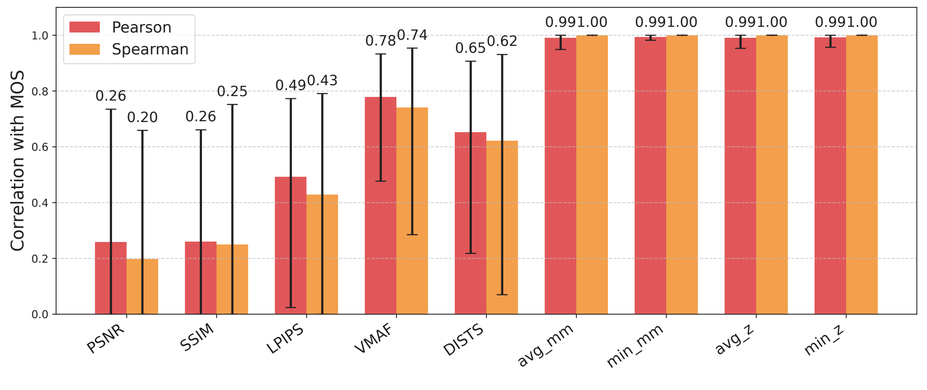
Fig 2: Outdoor dataset example - Truck scene from *NeRF View Synthesis: Subjective Quality Assessment and Objective Metrics Evaluation[3]*

RESULTS

Configuration 1: Normalization scales calibrated and evaluated on the Synthetic dataset



Configuration 2 : Normalization scales calibrated and evaluated on the Outdoor dataset



Configuration 3: Normalization scales calibrated on the Synthetic dataset and evaluated on the Outdoor dataset

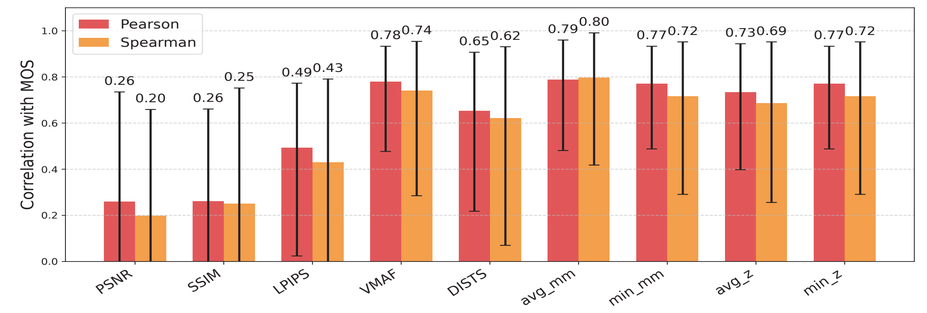


Table 1 : Summary of Fusion Correlation Results

Fusion methods are denoted as avg (average) or min (minimum selection) combined with normalization type (mm = min–max, z = z-score).

Dataset	#Samples	Fusion	r_p	r_s
Configuration 1	320/180	avg_mm	0.837	0.826
		min_mm	0.720	0.704
		avg_z	**0.883**	**0.867**
		min_z	0.800	0.808
Configuration 2	12/4	avg_mm	0.990	1.000
		min_mm	0.994	1.000
		avg_z	0.991	1.000
		min_z	0.992	1.000
Configuraton 3	400/16	avg_mm	**0.788**	**0.797**
		min_mm	0.770	0.716
		avg_z	0.734	0.686

CONCLUSION

- Evaluating NeRF remains challenging as no single metric consistently aligns with human perception across datasets.
- Our experiments show that average Min–Max normalization combined with weighted average fusion of VMAF and DISTS achieves the most reliable correlation with subjective scores, demonstrating robustness and generalizability compared to individual metrics.

FUTURE WORKS

We plan to extend this approach by incorporating additional metrics and using advanced fusion models such as ridge regression and neural networks, to learn optimal fusion weights and improve alignment with subjective quality scores.

Citations

[1] Mildenhall, Ben, et al. “Nerf: Representing scenes as neural radiance fields for view synthesis.” Communications of the ACM 65.1 (2021): 99-106.
[2] Xing, Yuke, et al. “Explicit-NeRF-QA: A quality assessment database for explicit NeRF model compression.” 2024 IEEE International Conference on Visual Communications and Image Processing (VCIP). IEEE, 2024.
[3] Martin, Pedro, et al. “Nerf view synthesis: Subjective quality assessment and objective metrics evaluation.” IEEE Access (2024).

LINKÖPING UNIVERSITY

ELLIIT Excellence Center at Linköping – Lund in Information Technology

QoMEX 2025