ABLE-NeRF: Attention-Based Rendering with Learnable Embeddings for Neural Radiance Field

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Zhe Jun Tang[1], Tat-Jen Cham[2], Haiyu Zhao[3]

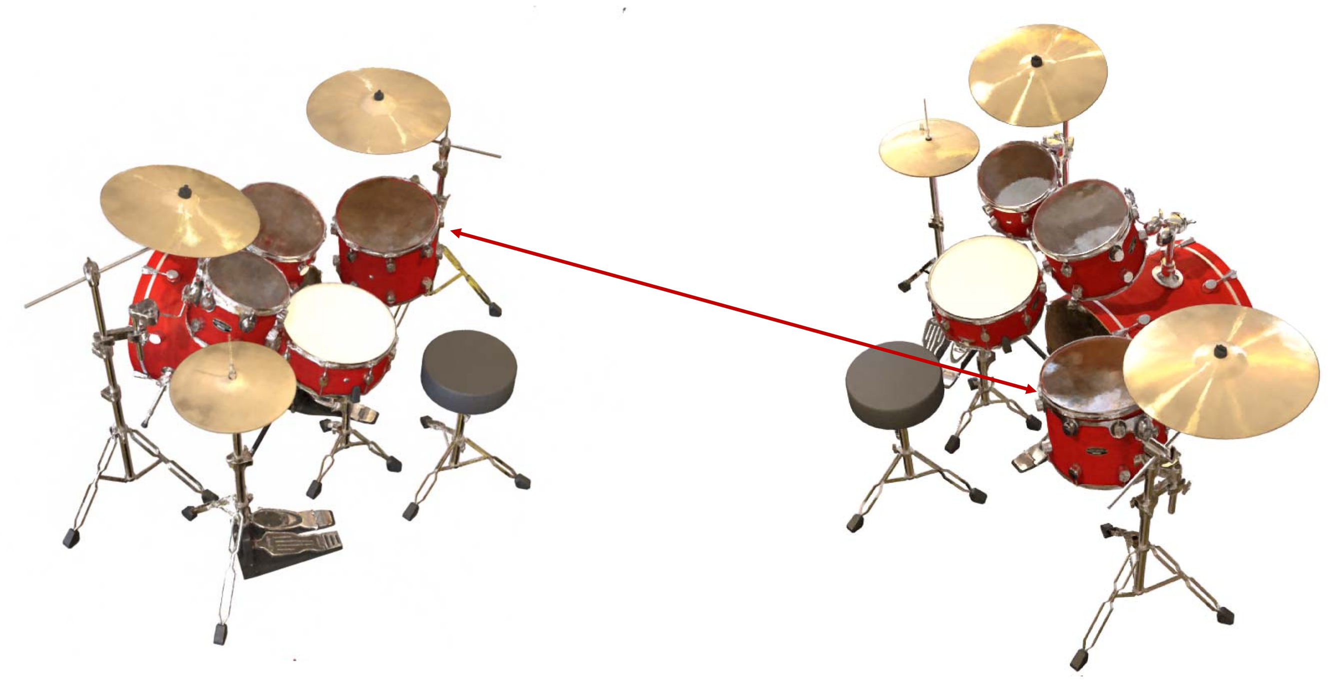
[1]S-Lab Nanyang Technological University, [2]Nanyang Technological University, [3]SenseTime Research

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE
S-LAB FOR ADVANCED INTELLIGENCE

Render accurate view dependent effects for view synthesis

- Standard volume rendering method with NeRF accumulates features in color space result in rendering blurry transparent and glossy surfaces
- Prior Volume Rendering works include a view dependent scattering term or an opacity hull/ 'alpha sphere' which stores an opacity value viewed at direction ω
- NeRF computes σ based on spatial coordinates ignore dependency on viewing angles, resulting in 'blurry' surfaces

$$\hat{\boldsymbol{C}}(r) = \sum_{i=1}^{N} T_i\left(1 - \exp(-\sigma_i \delta_i)\right)\boldsymbol{c}_i \qquad T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right)$$

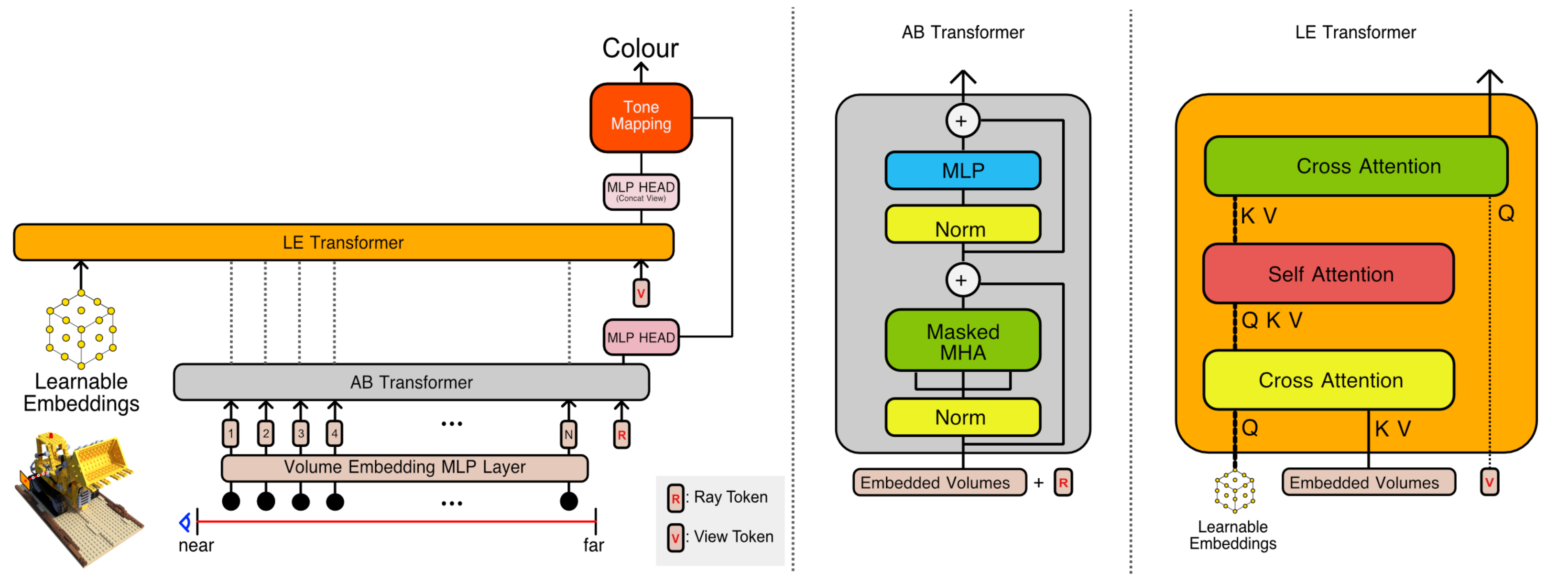


NeRF based approaches render surfaces exhibiting both translucency and specularity poorly

Our Contributions

- A new approach demonstrating the capability and superiority of transformers modelling a physics-based volumetric rendering approach
- A novel memorization-based framework with Learnable Embeddings (LE) to capture and render detailed view-dependent effects with a cross-attention network

ABLE-NeRF



Attention Based Rendering

- Sample conic frustums along ray and append a Ray Token
- Constrain sequence of Ray Token and ray samples with masking to encode a uni-directional ray information
- Each sample only attends to Ray Token and samples lying in-front of it
- Decode Ray Token to obtain Diffused Color

	Ray Token	Ray Sample 1	Ray Sample 2	...	Ray Sample N
Ray Token	0	0	0	0	0
Ray Sample 1	0	0	-∞	-∞	-∞
Ray Sample 2	0	0	0	-∞	-∞
...			...		
Ray Sample N	0	0	0	0	0

Ray Masking in AB Transformer

Learnable Embeddings

- Include Embeddings as learnable network parameters to capture static scene illumination
- Query Embeddings with View Token to recover view-dependent color

Without LE | Complete Model | Ground Truth | Ours | Standard Volume Rendering

- Our model can recover intra-scene surface reflections by querying from Embeddings for static scene illumination compared to strict ray casting volume rendering methods

Experimental Results

	Chair	Lego	Materials	Mic	Hotdog	Ficus	Drums	Ship
PhySG	24.00	20.19	18.86	22.33	24.08	19.02	20.99	15.35
VolSDF	30.57	29.46	29.13	30.53	35.11	22.91	20.43	22.51
Mip-NeRF	35.12	35.92	30.62	36.76	37.34	33.19	25.36	30.52
Ref-NeRF	35.83	36.25	35.41	36.76	37.72	33.91	25.79	30.28
Ours (no LE)	35.76	36.62	34.57	35.90	38.68	34.28	25.98	30.60
Ours	36.25	38.03	35.46	37.11	39.07	35.69	26.84	31.75

PSNR Comparison on Blender Dataset. For additional results, refer to paper.

Conclusion

- Our end-to-end deep learning-based approach with transformers is capable of learning a physics-based volumetric rendering method without an explicit formulation
- Inclusion of Learnable Embeddings to capture static scene illumination allows ABLE-NeRF to produce convincing view-dependent lighting effects

More Information Visit

References

Wojciech Matusik, Hanspeter Pfister, Addy Ngan, Paul Beardsley, Remo Ziegler, and Leonard McMillan. Image based 3d photography using opacity hulls. ACM Transactions on Graphics (TOG), 21(3):427–437, 2002

Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 5481–5490. IEEE, 2022.

Max, N.: Optical models for direct volume rendering. IEEE Transactions on Visualization and Computer Graphics (1995)