Optica Imaging Congress (3D, AO, COSI, IS, pcAOP) © Optica Publishing Group 2024

Permutation Transient Attention Encoder for None-Line-of-Sight Imaging

Wenjie Yue,[1,2,#] Xiaxu Chen,[1,2,#] Gangping Liu,[1,2] Chutian Wang,[3] Edmund Y. Lam[3] and Jun Ke[1,2,*]

[1] *School of Optics and Photonics, Beijing Institute of Technology, Beijing, China 100081*
[2] *Key Laboratory of Photo-electronic Imaging Technology and System, Ministry of Education of China, Beijing, China 100081*
[3] *Department of Electrical and Electronic Engineering, University of Hong Kong, Hong Kong, China*
[#] *These authors contributed to the work equally and should be regarded as co-first authors.*

[*]*jke@bit.edu.cn*

Abstract: This paper introduces Permutation Transient Attention Encoders as a novel and efficient approach for NLOS reconstruction. Our method, utilizing Transient Attention Blocks and Permute-MLP, enhances transient feature extraction and offers more discriminative feature representations than existing approaches.

1. Introduction

Non-Line-of-Sight (NLOS) imaging aims to collect transient measurements to reveal scenes hidden from the camera's direct view. A typical NLOS imaging system utilizes external light sources such as ultrafast lasers to illuminate a diffuse relay surface. Simultaneously, a single-photon avalanche diode (SPAD) or streak camera was used to detect the light reflected sequentially on the relay surface and hidden objects. However, reconstructing complex scenarios remains challenging for the existing methods because of the multiple diffuse reflections experienced by the measured light. Inspired by neural radiance fields (NeRF) and physical embedding-based deep learning methods [1,2], we propose an end-to-end model for efficient NLOS reconstruction, as shown in Fig 1. The experimental results demonstrate that our model achieves leading performance in objective evaluation metrics.

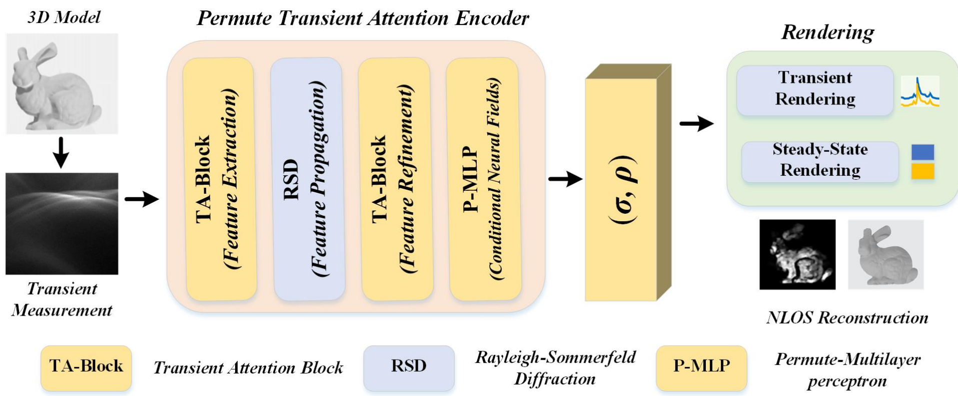

Fig. 1. The schematic diagram of our proposed Permutation Transient Attention Encoder for NLOS reconstruction.

2. Method

As shown in Fig. 1, our model consists of a Permute Transient Attention Encoder (PTAE) and a rendering module. In Fig. 1, the modules with a yellow background represent the parts of the model that are optimizable during backpropagation. The modules with a blue background indicate the components in which the parameters remain constant during backpropagation. Nonetheless, these blue modules enable differentiable computations, allowing the calculation of gradients through backpropagation.

Inspired by CBAM [3], a simple yet effective attention module for feed-forward convolutional neural networks, we propose a Transient Attention Block (TA-Block) specifically designed for transient measurements. As shown in Fig. 2.(a), The TA-Block comprises a Conv-3d, Local-3d Channel Attention Module, and Local-3d Spatial

Attention Module. The computational processes of the Local-3d Channel Attention Module and Local-3d Spatial Attention Module are illustrated in Fig. 2.(b) and Fig. 2.(c), respectively. In the ablation experiments, we demonstrated the effectiveness of Transient Attention Blocks.

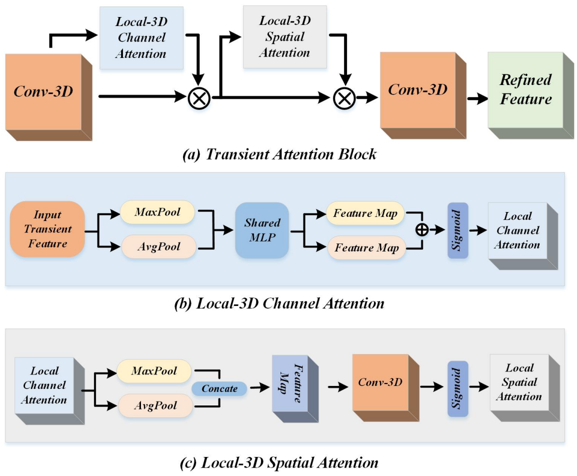

Fig. 2. NLOS imaging system

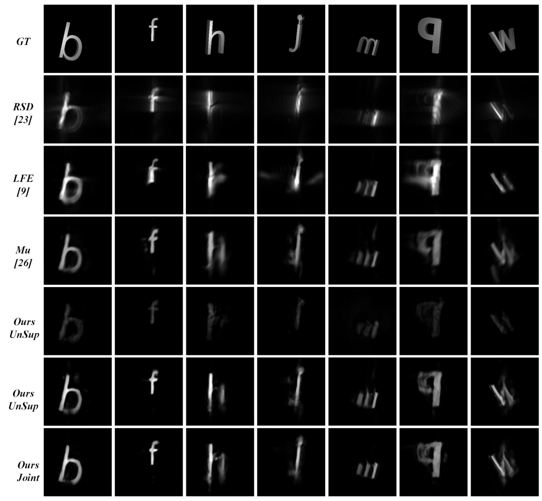

Fig. 3. NLOS Reconstruction results

In our end-to-end NLOS model, TA-Blocks were used to extract shallow features from the NLOS measurements. Then, physical prior knowledge named Rayleigh-Sommerfeld Diffraction (RSD) is utilized for subsequent feature processing. After Permute-MLP, the tailored module to accommodate 3-dimensional token formats. In contrast to traditional MLP models that use two-dimensional inputs, features from the previous module are encoded into the desired reflectivity ρ and volume density σ. They are serviceable for steady-state rendering of 2D images and transient rendering of temporal histograms. Crucially, these forward models are entirely differentiable, allowing end-to-end training of our deep learning model.

3. Results

We compared our model with other remarkable feed-forward reconstruction methods such as RSD [4], LFE [1] and Mu's model using supervised and unsupervised joint training strategy [2]. RSD represents the state-of-the-art (SOTA) physics-based approaches, while LFE and Mu's model are pioneering learning-based methods. To ensure a fair comparison, we retrained the LFE and Mu's model on the alphanumeric dataset until convergence. The test results for the numerical metrics are shown in Fig. 3 and Table 1 indicate that our supervised learning method achieved the best performance. However, our unsupervised learning approach did not perform well in terms of numerical evaluation metrics. This is because the optimization of our unsupervised method is not specifically designed for the PSNR and SSIM metrics. In addition, as shown in Fig. 3, the proposed method also exhibited visual quality.

Table 1. Quantitative evaluation on alphanumerics dataset

EVA	RSD	LFE	Mu's model	Ours (sup)	Ours(unsup)	Ours(joint)
PSNR	22.02	22.34	24.64	24.94	20.14	24.92
SSIM	0.395	0.886	0.845	0.859	0.751	0.856
RMSE	0.084	0.064	0.063	0.059	0.102	0.061

References

1. W. Chen, F. Wei, K. N. Kutulakos, S. Rusinkiewicz, and F. Heide, "Learned feature embeddings for non-line-of-sight imaging and recognition," *ACM Transactions on Graphics (ToG)*, vol. 39, no. 6, pp. 1–18, 2020.
2. F. Mu, S. Mo, J. Peng, X. Liu, J. H. Nam, S. Raghavan, A. Velten, and Y. Li, "Physics to the rescue: Deep non-line-of-sight reconstruction for high-speed imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
3. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
4. X. Liu, I. Guillén, M. La Manna, J. H. Nam, S. A. Reza, T. Huu Le, A. Jarabo, D. Gutierrez, and A. Velten, "Non-line-of-sight imaging using phasor-field virtual wave optics," *Nature*, vol. 572, no. 7771, pp. 620–623, 2019.