Adaptive Attention Based on Mixture Distribution for Zero-Shot Non-Line-of-Sight Imaging

Qinghua Zhang, Jun Liu, and Yuping Duan

***Abstract*—Non-line-of-sight (NLOS) imaging is an ill-posed problem to reconstruct hidden 3D scenes by leveraging photon time-of-flight information from diffusely reflected light. In the existing regularization models, the spatial residuals were handled by a single distribution, failing to account for the distinct characteristics of background and target objects. In this paper, we propose a novel NLOS reconstruction method that models the non-Gaussian residuals with a mixture distribution. Through a dual method, we derive an adaptive weighted residual model, where the weights generated in the dual space act as a zero-shot attention mechanism to control the contributions of different regions. The corresponding optimization problem can be effectively solved using the alternating minimization algorithm. Numerical experiments on both synthetic and real-world datasets demonstrate that our method surpasses the related existing approaches, achieving state-of-the-art performance.**

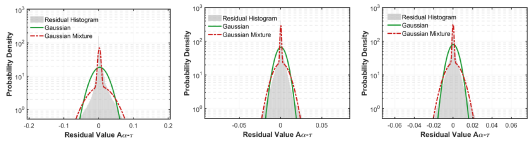
***Index Terms*—Non-line-of-sight imaging, mixture distribution, duality theory, attention mechanism.**

I. Introduction

NON-LINE-OF-SIGHT (NLOS) imaging enables the detection of objects hidden around corners, using time-of-fight (ToF) measurements to reconstruct their structure and appearance. However, these signals often suffer from corruption and low signal-to-noise ratio (SNR), posing significant challenges for reconstruction [1]. In recent years, advanced ToF-based algorithms [2], [3], [4] have been developed, enabling more accurate and robust NLOS reconstruction and expanding its applications in diverse fields such as intelligent driving [5], remote sensing [6], 3D human pose estimation [7], and beyond.

Existing NLOS reconstruction algorithms can be broadly classified into three primary categories: direct approaches, regularization-based methods, and deep learning (DL) based methods. Filtered back-projection is a classical direct method, reconstructing hidden scenes by filtering the measured data and back-projecting them along the corresponding light paths. Alternatively, the light cone transform (LCT) [8], [9] based methods models the physical process as a 3D convolution operation, enabling reconstructions through deconvolution techniques such as Wiener filters. Other direct approaches contain wave-based models [10], [11], [12], geometric algorithms based on Fermat's principles [13] and inverse rendering approaches [14], [15], etc., all of which rely on physical modeling of light transport to recover hidden scenes. Despite their computational efficiency, direct methods in NLOS imaging are often sensitive to noise and have difficulty handling complex scene geometries. Regularization-based methods aim to address these limitations by formulating NLOS imaging as an optimization problem, effectively incorporating prior knowledge of the hidden object. For instance, Wu et al. [16] explored the non-negativity of the reflectivity and the total variation (TV) regularization in the optimization problem of NLOS. Liu et al. [17] collaboratively incorporated sparseness, non-local self-similarity of the hidden object and smoothness of the estimated signal into the regularization term (SOCR). Ding et al. [18] proposed NLOS reconstruction models based on curvature regularization. Recently, DL methods [19], [20], [21], [22], [23] have been introduced into NLOS reconstruction, offering significant improvements by leveraging the powerful feature-learning capabilities of neural network. However, most DL methods suffer from two drawbacks. One is the supervised manner [24], training a deep network often requires a large dataset, which hinders the applicability when NLOS data is difficultly available. The other is the generalization [25], e.g., the network trained on the dataset with a specific distribution may not perform well on datasets with unseen distributions.

During the process of NLOS data acquisition, ToF measurements are inevitably affected by various distortions due to the physical limitations and sensor imperfections. Additionally, the contributions of background and the different regions within the object's surface are commonly inconsistent. However, existing methods often assume single distribution for the residuals, such as Gaussian [2] or Poisson [26] distribution, limiting their ability to handle the complex NLOS scenarios. As illustrated in Fig. 1, during the iterative process of the SPIRAL [16] method, a single Gaussian distribution fails to accurately capture the true residual distribution, whereas a mixture of Gaussian



(a) iterations=10 (b) iterations=50 (c) iterations=100

Fig. 1. Comparison of the single Gaussian model and Gaussian mixture model in fitting the probability density function of residuals during the SPIRAL [16] method at various iterations for the *bowling* scene.

Received 31 January 2025; revised 25 March 2025; accepted 1 April 2025. Date of publication 7 April 2025; date of current version 2 May 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFC3008505, in part by the National Natural Science Foundation of China under Grant 12371527, and Grant 12071345, and in part by Beijing Natural Science Foundation under Grant 1232011. The associate editor coordinating the review of this article and approving it for publication was Prof. Parvaneh Saeedi. *(Corresponding authors: Jun Liu; Yuping Duan.)*

The authors are with the Laboratory of Mathematics and Complex Systems (Ministry of Education of China), School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China (e-mail: qhzhang@mail.bnu.edu.cn; jliu@bnu.edu.cn; ypduan@bnu.edu.cn).

The code is available at: https://github.com/qiuxuanzhizi/AMD-NLOS-V1.

This article has supplementary downloadable material available at https://doi.org/10.1109/LSP.2025.3558458, provided by the authors.

Digital Object Identifier 10.1109/LSP.2025.3558458

distributions provides a much better approximation of the actual residual distribution. Building on this insight, we propose an **A**daptive regularization framework for **NLOS** imaging based on the **M**ixture **D**istribution [27], called **AMD-NLOS**. By characterizing the non-Gaussian residuals in the reconstruction process with a mixture distribution, our method introduces an *attention* mechanism in the dual space to control the contribution of different regions to the overall reconstruction task, enabling *zero-shot* reconstruction on NLOS datasets with previously unseen distributions.

Our main contributions are summarized as follows:

- We are the first to use a mixed distribution to describe the potential differences in NLOS reconstruction that may be caused by different components in transient data.
- Based on the mixture distribution, we propose a method called AMD-NLOS to generate an attention mechanism for zero-shot NLOS Imaging.
- An efficient dual algorithm is proposed to minimize the AMD-NLOS model. Extensive experiments on both synthetic and real-world datasets demonstrate the superior performance of AMD-NLOS over its competitors.

II. Background

A. Forward Model of NLOS

In non-line-of-sight (NLOS) imaging, the forward propagation model describes how light interacts with a hidden scene and returns to the detector via a visible wall, which is assumed to follow a spatial-temporal hypercone defined by the equation $x^2+y^2+z^2-(\frac{ct}{2})^2=0$, where c denotes the speed of light and t is the total time of travel. Additionally, the propagation model is assumed to be isotropic, meaning that the light disperses uniformly in all directions at each interaction point. The general forward model for NLOS imaging with distinct illumination and detection points is [18]:

$$\tau(\mathbf{x}',\mathbf{x}'',t)=\int \alpha(\mathbf{x})\cdot\frac{1}{d(\mathbf{x}',\mathbf{x})^2}\cdot\frac{1}{d(\mathbf{x},\mathbf{x}'')^2}\cdot\delta\left(t-\frac{d(\mathbf{x}',\mathbf{x})+d(\mathbf{x},\mathbf{x}'')}{c}\right)d\mathbf{x}, \tag{1}$$

where $\alpha(\mathbf{x})$ denotes the albedo of the hidden scene at each point $\mathbf{x}$ with $z>0$ in the 3D half-space Ω, $d(\mathbf{x}',\mathbf{x})$ and $d(\mathbf{x},\mathbf{x}'')$ are the Euclidean distances between the hidden point ($\mathbf{x}$) and the illumination ($\mathbf{x}'$) and detection ($\mathbf{x}''$) points, respectively. In the confocal setup, where the illumination and detection points coincide ($\mathbf{x}'=\mathbf{x}''$), the model simplifies to:

$$\tau(\mathbf{x}',t)=\int \alpha(\mathbf{x})\cdot\frac{1}{d(\mathbf{x}',\mathbf{x})^4}\cdot\delta\left(t-\frac{2d(\mathbf{x}',\mathbf{x})}{c}\right)d\mathbf{x}. \tag{2}$$

This confocal model reduces the complexity of the propagation path, assuming symmetric light travel. Considering the impact of background noise and Poisson detection noise, the discretization of NLOS imaging model (2) can be formed as:

$$\boldsymbol{\tau}=\mathcal{P}(\mathbf{A}\boldsymbol{\alpha}+\mathbf{b}), \tag{3}$$

where $\mathbf{A}=\mathbf{R}_t^{-1}\mathbf{H}\mathbf{R}_z\in\mathbb{R}^{n_xn_yn_t\times n_xn_yn_z}$ denotes the light transmission matrix, $\mathbf{H}\in\mathbb{R}^{n_xn_yn_h\times n_xn_yn_h}$ represents the 3D convolution with shift invariance, $\mathbf{R}_t\in\mathbb{R}_+^{n_xn_yn_h\times n_xn_yn_t}$ and $\mathbf{R}_z\in\mathbb{R}_+^{n_xn_yn_h\times n_xn_yn_z}$ correspond to the transformation operators applied to the temporal and spatial domains, respectively, $\boldsymbol{\alpha}\in\mathbb{R}_+^{n_xn_yn_z\times 1}$ denotes the volumetric albedo map of the hidden scene, $\mathbf{b}\in\mathbb{R}_+^{n_xn_yn_t\times 1}$ is the background noise and $\mathcal{P}$ represents Poisson stochastic process.

B. Bayesian Framework for NLOS

The reconstruction of the hidden scene in NLOS imaging can be formulated as a regularization-based inverse problem under the Bayesian framework [28], which is expressed as:

$$p(\boldsymbol{\alpha}\mid\boldsymbol{\tau})\propto p(\boldsymbol{\tau}\mid\boldsymbol{\alpha})p(\boldsymbol{\alpha}), \tag{4}$$

where $p(\boldsymbol{\alpha}\mid\boldsymbol{\tau})$ is the posterior probability of the scene representation $\boldsymbol{\alpha}$ given the transient measurements $\boldsymbol{\tau}$, $p(\boldsymbol{\tau}\mid\boldsymbol{\alpha})$ is the likelihood function describing the measurement process and $p(\boldsymbol{\alpha})$ is the prior distribution that incorporates assumptions about the scene structure. Maximizing the posterior probability leads to the maximum a posterior (MAP) approach, which is typically employed:

$$\boldsymbol{\alpha}^*=\arg\min_{\boldsymbol{\alpha}}-\ln p(\boldsymbol{\tau}\mid\boldsymbol{\alpha})-\ln p(\boldsymbol{\alpha}), \tag{5}$$

where $-\ln p(\boldsymbol{\tau}\mid\boldsymbol{\alpha})$ is the data fidelity term and $-\ln p(\boldsymbol{\alpha})$ denotes the regularization term. In previous works, the fidelity term has been modeled using different assumptions about the residual characteristics, e.g., Poisson distribution [2] and Gaussian distribution [16]. However, real-world scenes often exhibit varying spatial characteristics, making it impractical to accurately model the residuals with a single distribution, which motivates the development of our proposed AMD-NLOS model in the next section.

III. The Proposed Model and Algorithmic Solver

In this section, we introduce the proposed AMD-NLOS model along with its associated algorithmic solver. To facilitate the understanding of our approach, we first present the core dual theory of solving the well-known MAP problem of mixture models.

Definition 1 (smooth max operator [29]): For a vector $\mathbf{y}=(y^1,y^2,\ldots,y^J)^T$ and any $\epsilon>0$, the smooth max operator is defined as follows:

$$\max\nolimits_\epsilon(\mathbf{y}):=\epsilon\ln\sum\nolimits_{j=1}^{J}e^{\frac{y^j}{\epsilon}}. \tag{6}$$

Definition 2 (Fenchel-Legendre transformation [30]): The Fenchel-Legendre transformation of a function $\mathcal{F}$ is denoted by $\mathcal{F}^*$ and defined as:

$$\mathcal{F}^*(\boldsymbol{w}):=\max_{\mathbf{y}}\{\langle\mathbf{y},\boldsymbol{w}\rangle-\mathcal{F}(\mathbf{y})\}. \tag{7}$$

Lemma 1 (Bertsekas et al. [30]): A function $\mathcal{F}:\mathbb{R}^J\to\mathbb{R}\cup\{+\infty\}$ is convex and lower semi-continuous, then $\mathcal{F}=\mathcal{F}^{**}$.

Theorem 1 (Zhao et al. [29]): For any fixed $\epsilon>0$, the Fenchel-Legendre transformation of the smooth max operator $\max_\epsilon$ is

$$\begin{aligned}\max\nolimits_\epsilon^*(\boldsymbol{w})&=\max_{\mathbf{y}}\{\langle\mathbf{y},\boldsymbol{w}\rangle-\max\nolimits_\epsilon(\mathbf{y})\}\\&=\begin{cases}\epsilon\sum_{j=1}^{J}\omega^j\ln\omega^j, & \text{if } \boldsymbol{w}\in\mathbb{W},\\ +\infty, & \text{else.}\end{cases}\end{aligned} \tag{8}$$

Here $\mathbb{W}=\{\boldsymbol{w}=(\omega^1,\omega^2,\ldots,\omega^J)^T|0\le\omega^j\le1,\sum_{j=1}^{J}\omega^j=1\}$ is a probability simplex. Furthermore, since $\max_\epsilon$ is convex and continuous, and thus

$$\max\nolimits_\epsilon(\mathbf{y})=\max\nolimits_\epsilon^{**}(\mathbf{y})=\max_{\boldsymbol{w}\in\mathbb{W}}\left\{\langle\boldsymbol{w},\mathbf{y}\rangle-\epsilon\sum_{j=1}^{J}\omega^j\ln\omega^j\right\}. \tag{9}$$

The likelihood function of a mixed distribution can be viewed as a $\max_\epsilon$ function, thereby we can utilize dual variables to adaptively construct error weights. Our motivation for using dual

theory is to leverage the mathematical properties of the dual variable $\boldsymbol{w}$ belongs to the probability simplex to construct smooth weights (attention mechanism) for each reconstructed voxel point, thereby enhancing the quality of image reconstruction.

Building on the duality theory discussed above, we extended these concepts to the construction of our optimization model. Particularly, the negativity of the fidelity term $\ln p(\boldsymbol{\tau}|\boldsymbol{\alpha})$ in (5) can be further refined by explicitly modeling the residuals $\mathbf{n} = \mathbf{A}\boldsymbol{\alpha} - \boldsymbol{\tau}$ through a distribution $p(\mathbf{n}; \Theta)$. To estimate the parameters of residuals for any given data $\mathbf{n}$, the residuals $n_i, i = 1, 2, \ldots, I\,(I = n_x n_y n_t)$ are assumed to be i.i.d. random variables with probability density function $\sum_{j=1}^{J} \gamma^j p^j(n_i; \sigma^j)$, which leads to the joint probability density function (PDF) of the random vector $\mathbf{n} = (n_1, n_2, \ldots, n_I)^T$:

$$p(\mathbf{n}; \Theta) = \prod\nolimits_{i=1}^{I} p(n_i; \Theta). \tag{10}$$

Then the maximum likelihood estimate of Θ is a parameter Θ^* that maximizes the logarithm of the likelihood function:

$$\begin{aligned}\ln p(\mathbf{n}; \Theta) &= \sum_{i=1}^{I} \ln \sum_{j=1}^{J} \gamma^j p^j(n_i; \sigma^j) \overset{\epsilon=1}{=} \frac{1}{\epsilon} \sum_{i=1}^{I} \epsilon \ln \sum_{j=1}^{J} e^{\frac{y_i^j}{\epsilon}} \\ &\overset{\text{def.1}}{=} \frac{1}{\epsilon} \sum_{i=1}^{I} \max\nolimits_{\epsilon}(\mathbf{y}_i) \overset{\text{theo.1}}{=} \frac{1}{\epsilon} \sum_{i=1}^{I} \max_{\boldsymbol{w}_i \in \mathbb{W}} \{\langle \boldsymbol{w}_i, \mathbf{y}_i \rangle \\ &\quad - \epsilon \langle \boldsymbol{w}_i, \ln \boldsymbol{w}_i \rangle\},\end{aligned} \tag{11}$$

where $\boldsymbol{w}_i = (\omega_i^1, \ldots, \omega_i^J)^T$, $\mathbf{y}_i = (y_i^1, \ldots, y_i^J)^T, y_i^j = \epsilon \ln \gamma^j p^j(n_i; \sigma^j)$, $\Theta = \{\sigma^j, \gamma^j\}_{j=1}^J$, σ^j denotes the standard deviation of the j-th component and ϵ is a positive number.

Based on (5) and (11), the optimization problem can be reformulated as:

$$\begin{aligned}&\arg\min\nolimits_{\boldsymbol{\alpha}, \Theta, \boldsymbol{w} \in \mathbb{W}} \frac{1}{2} \sum\nolimits_{j=1}^{J} \frac{\|\mathbf{A}\boldsymbol{\alpha} - \boldsymbol{\tau}\|_{2,\boldsymbol{w}^j}^2}{(\sigma^j)^2} + \sum\nolimits_{j=1}^{J} \langle \boldsymbol{w}^j, \ln \boldsymbol{w}^j \rangle \\ &+ \frac{1}{2} \sum\nolimits_{j=1}^{J} \langle \boldsymbol{w}^j, \mathbf{1} \rangle (\ln(\sigma^j)^2 - 2 \ln \gamma^j) + \lambda \mathcal{R}(\boldsymbol{\alpha}),\end{aligned} \tag{12}$$

where $\boldsymbol{w}^j = (\omega_1^j, \ldots, \omega_I^j)$ denotes the dual variable, $\|\cdot\|_{2,\boldsymbol{w}^j}^2$ represents the weighted ℓ_2-norm of vector, $\mathcal{R}(\boldsymbol{\alpha})$ is the regularization term and $\lambda > 0$ is the parameter indicating how strict the regularization will be applied. To solve problem (12) efficiently, we introduce an auxiliary variable $\mathbf{u}$, facilitating an approximate formulation [31]:

$$\begin{aligned}&\arg\min\nolimits_{\boldsymbol{\alpha}, \mathbf{u}, \Theta, \boldsymbol{w} \in \mathbb{W}} \frac{1}{2} \sum\nolimits_{j=1}^{J} \frac{\|\mathbf{A}\mathbf{u} - \boldsymbol{\tau}\|_{2,\boldsymbol{w}^j}^2}{(\sigma^j)^2} \\ &+ \frac{\eta}{2} \|\boldsymbol{\alpha} - \mathbf{u}\|_2^2 + \sum\nolimits_{j=1}^{J} \langle \boldsymbol{w}^j, \ln \boldsymbol{w}^j \rangle \\ &+ \frac{1}{2} \sum\nolimits_{j=1}^{J} \langle \boldsymbol{w}^j, \mathbf{1} \rangle (\ln(\sigma^j)^2 - 2 \ln \gamma^j) + \lambda \mathcal{R}(\boldsymbol{\alpha}).\end{aligned} \tag{13}$$

By applying the alternating minimization algorithm, (13) is decomposed into four subproblems, with the iteration index for the variables omitted to avoid notational confusion:

$$\begin{cases} \boldsymbol{\alpha} = \arg\min\limits_{\boldsymbol{\alpha}} \{\frac{\eta}{2} \|\boldsymbol{\alpha} - \mathbf{u}\|_2^2 + \lambda \mathcal{R}(\boldsymbol{\alpha})\}, \\ \mathbf{u} = \arg\min\limits_{\mathbf{u}} \frac{1}{2} \sum_{j=1}^{J} \frac{\|\mathbf{A}\mathbf{u} - \boldsymbol{\tau}\|_{2,\boldsymbol{w}^j}^2}{(\sigma^j)^2} + \frac{\eta}{2} \|\boldsymbol{\alpha} - \mathbf{u}\|_2^2, \\ \Theta = \arg\min\limits_{\Theta} \frac{1}{2} \sum_{j=1}^{J} \frac{\|\mathbf{A}\mathbf{u} - \boldsymbol{\tau}\|_{2,\boldsymbol{w}^j}^2}{(\sigma^j)^2} \\ \quad + \frac{1}{2} \sum_{j=1}^{J} \langle \boldsymbol{w}^j, \mathbf{1} \rangle (\ln(\sigma^j)^2 - 2 \ln \gamma^j), \\ \boldsymbol{w} = \arg\min\limits_{\boldsymbol{w} \in \mathbb{W}} \frac{1}{2} \sum_{j=1}^{J} \frac{\|\mathbf{A}\mathbf{u} - \boldsymbol{\tau}\|_{2,\boldsymbol{w}^j}^2}{(\sigma^j)^2} + \sum_{j=1}^{J} \langle \boldsymbol{w}^j, \ln \boldsymbol{w}^j \rangle \\ \quad + \frac{1}{2} \sum_{j=1}^{J} \langle \boldsymbol{w}^j, \mathbf{1} \rangle (\ln(\sigma^j)^2 - 2 \ln \gamma^j). \end{cases} \tag{14}$$

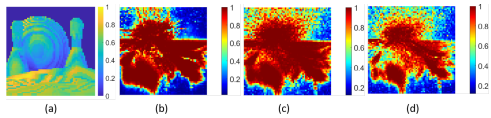


Fig. 2. Visualization of heatmaps from attention mechanism driven by weight matrices. (a) Depth map projection of the albedo of the real hidden object *bowling*. (b)–(d) Heatmaps of the weight matrix after 5, 15, and 25 iterations, respectively.

The first subproblem, in fact, can be interpreted as a denoising problem for $\mathbf{u}$, which can be addressed by incorporating advanced denoising algorithms in a *plug-and-play* manner [32]. The cost function of the second subproblem is quadratic with respect to $\mathbf{u}$, so the associated Euler-Lagrange equation

$$\hat{\mathbf{A}} \left((\mathbf{A}\mathbf{u}) \odot \sum_{j=1}^{J} \frac{\boldsymbol{w}^j}{(\sigma^j)^2} \right) + \eta \mathbf{u} = \hat{\mathbf{A}} \left(\boldsymbol{\tau} \odot \sum_{j=1}^{J} \frac{\boldsymbol{w}^j}{(\sigma^j)^2} \right) + \eta \boldsymbol{\alpha} \tag{15}$$

can be approximately solved using the conjugate gradient (CG) method, where $\hat{\mathbf{A}}$ is the conjugate of $\mathbf{A}$ and $\odot$ stands for element-wise multiplication of vectors. The solution of the third problem can be given explicitly by

$$\gamma^j = \frac{\langle \boldsymbol{w}^j, \mathbf{1} \rangle}{\widetilde{I}}, (\sigma^j)^2 = \frac{\|\mathbf{A}\mathbf{u} - \boldsymbol{\tau}\|_{2,\boldsymbol{w}^j}^2}{\langle \boldsymbol{w}^j, \mathbf{1} \rangle}, \tag{16}$$

where $\widetilde{I} = \sum_{j=1}^{J} \langle w^j, \mathbf{1} \rangle$. The last subproblem also has a closed-form solution:

$$\begin{aligned}\boldsymbol{w}^j &= \frac{\frac{\gamma^j}{\sigma^j} \exp\{-\frac{(\mathbf{A}\mathbf{u} - \boldsymbol{\tau})^2}{2(\sigma^j)^2}\}}{\sum_{\kappa=1}^{J} \frac{\gamma^\kappa}{\sigma^\kappa} \exp\{-\frac{(\mathbf{A}\mathbf{u} - \boldsymbol{\tau})^2}{2(\sigma^\kappa)^2}\}} \\ &= \text{sofmax}\left(-\frac{(\mathbf{A}\mathbf{u} - \boldsymbol{\tau})^2}{2\boldsymbol{\sigma}^2} + \ln \frac{\boldsymbol{\gamma}}{\boldsymbol{\sigma}} \right).\end{aligned} \tag{17}$$

Remark 1: It is worth mentioning that the dual variable $\boldsymbol{w}^j$ can be interpreted as a form of *attention* mechanism, similar to the softmax function [33] commonly used in deep neural networks. While we cannot pre-identify which points in the spatial domain contribute more or less to the reconstruction, the iterative optimization of the dual variable $\boldsymbol{w}^j$ can adaptively identify the most significant points and assign higher weight to them for accurate reconstruction. This adaptive weighting process is intuitively demonstrated in Fig. 2, where the regions with higher weights (darker areas in the heatmaps) correspond to the target object.

IV. Numerical Experiments

In this section, we present the performance of our AMD-NLOS approach on both synthetic and real-world data, comparing it with prominent methods including the direction methods, i.e., FBP [34], LCT [8], F-K migration [11] and phasor-field (PF) [10], and iterative methods, i.e., total variation (TV) based method SPIRAL [16] and curvature regularization based methods CNLOS/CNLOS$*$[1] [18]. In this work, we do not emphasize the regularization term $\mathcal{R}(\boldsymbol{\alpha})$ and adopt the TV [35] penalty in the experiments for simplicity in computation[2]. For more experimental results, please refer to the *supplementary material*.

[1]It is noted that CNLOS and CNLOS$*$ are curvature regularization based methods in the object domain and dual domain, respectively.

[2]In our experiments, the denoising problem in (14) is solved via the fast implementation of BM3D [36].

TABLE I
COMPARISON OF THE RECONSTRUCTION RESULTS IN PSNR(DB), SSIM, AC, RMSE

Method	PSNR↑	SSIM↑	AC↑	RMSE↓
LCT	14.00	0.2391	0.8806	0.1983
PF	14.01	0.2321	0.9366	0.1718
F-K	10.97	0.1827	0.8616	0.2062
SPIRAL	16.41	0.4163	0.9517	0.1045
CNLOS	16.56	0.4672	0.9544	0.1025
CNLOS*	<u>16.71</u>	<u>0.4705</u>	<u>0.9550</u>	<u>0.0989</u>
Ours	**17.40**	**0.4977**	**0.9712**	**0.0906**

The best results are highlighted in bold and the <u>underlined</u> results indicate the second-best performance.

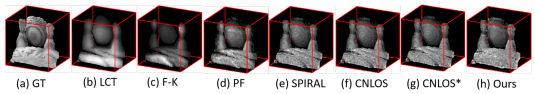


Fig. 3. A visual comparison of the competitive methods in the scene of *bowling*.

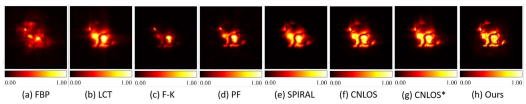


Fig. 4. A visual comparison of the competitive methods in the scene of *bunny*.

A. Experiments on Synthetic Data

In this subsection, two synthetic confocal datasets are utilized to validate the reconstruction performance of our AMD-NLOS method. The first confocal data is the *bowling* scene, generated using 64×64 scanning points to cover a 1m$\times$ 1m square on the wall, with a temporal resolution of 256 bins and a bin width of 32ps. The numerical results are shown in Table I, which includes the following evaluation metrics [37]: the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), the Accuracy (AC) and Relative Mean Absolute Error (RMSE). As expected, it can be observed the proposed AMD-NLOS largely outperforms the direct methods and achieve moderate improvements over the iterative methods. Fig. 3 displays the visual results of competitive methods in the *bowling* scene, which highlights the effectiveness of our approach in restoring continuous albedo across the object's surface and sharp edge, attributing to its ability to accurately identify and adaptively eliminate diverse residuals.

The second confocal data is the *Stanford bunny*, which has the spatial resolution 64$\times$64 and temporal resolution 300 time bins, with each bin corresponding to a photon travel distance of 0.0025 meters. The visual comparison in Fig. 4 reveals that the CNLOS/CNLOS$*$ and our AMD-NLOS can produce promising shape and albedo, in which our AMD-NLOS can preserve sharper edge and more detailed information. To further evaluate the robustness of the proposed method handling the mixed noise, impulse/Gaussian/Rice noise is introduced to the simulation data *bunny*. Fig. 5 demonstrates the superior performance of our proposed method under various mixture noise conditions, highlighting the enhanced robustness of our AMD-NLOS framework compared to existing approaches.

B. Experiments on Real-World Data

To evaluate the practical applicability of our method, we conduct experiments using real-world data provided in [11]. The raw measurements have dimension of 128$\times$128$\times$2048 or 512$\times$512$\times$ 2048, with the time resolution cropped to 512 and each time bin spanning 32ps. Due to memory limitations, for the measurements with 512×512 spatial resolution, we aggregate every 4×4 scanning points, resulting in a total of 128×128 scanning points. Fig. 6 presents the reconstruction results in the scene of *teaser* under the exposure time of 60min, which demonstrates our AMD-NLOS not only effectively suppresses noise but also excels in preserving fine details.

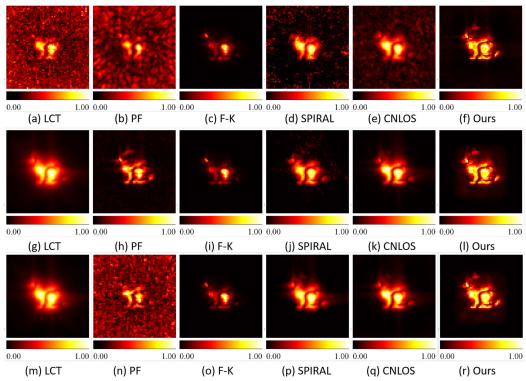


Fig. 5. Visual comparisons of the competitive methods under different mixture noise conditions. **Top**: Poisson noise + salt & pepper noise (r=0.001), **middle**: Poisson + Gaussian noise (intensity value: 0.001), **bottom**: Poisson + Rice noise (intensity value: 0.001).

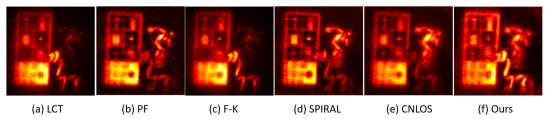


Fig. 6. A visual comparison of the competitive methods in the real scene of *teaser* (60 min, 128 $\times$ 128).

TABLE II
ABLATION STUDY: PSNR (DB), SSIM AND TIME (S) WITH RESPECT TO THE NUMBER OF GAUSSIAN COMPONENTS J IN THE SCENE *BOWLING*

J	2	3	4	5
PSNR	17.40	**17.65**	<u>17.63</u>	<u>17.63</u>
SSIM	**0.4977**	<u>0.4938</u>	0.4887	0.4867
Time	**180.21**	<u>180.45</u>	204.43	210.45

The best results are highlighted in bold and the <u>underlined</u> results indicate the second-best performance.

C. Ablation Study on J

We conduct an ablation study on different values of Gaussian components J. As shown in Table II, it is demonstrated that the performance of our method is not sensitive to J. Therefore, for simplicity and clarity, we choose $J = 2$ in all our experiments.

V. CONCLUSION

In this letter, we proposed AMD-NLOS, an adaptive zero-shot approach for NLOS imaging. This method dynamically detects the distinct characteristics of residuals across different locations and assigns appropriate weights by leveraging a mixture distribution, achieving promising performance in public dataset.

REFERENCES

[1] M. Liu, H. Liu, X. He, S. Jin, P. Chen, and M. Xu, "Research advances on non-line-of-sight imaging technology," *J. Shanghai Jiaotong Univ. Sci.*, pp. 1–22, Jan. 2024. [Online]. Available: https://doi.org/10.1007/s12204-023-2686-8

[2] J. Liu, Y. Zhou, X. Huang, Z.-P. Li, and F. Xu, "Photon-efficient non-line-of-sight imaging," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 639–650, 2022.

[3] B. Ahn, A. Dave, A. Veeraraghavan, I. Gkioulekas, and A. C. Sankaranarayanan, "Convolutional approximations to the general non-line-of-sight imaging operator," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 7889 7899.
[4] J. H. Nam et al., "Low-latency time-of-flight non-line-of-sight imaging at 5 frames per second," *Nature Commun.*, vol. 12, no. 1, 2021, Art. no. 6526.
[5] A. Young et al., "Enhancing autonomous navigation by imaging hidden objects using single-photon LiDAR," 2024, *arXiv:2410.03555*.
[6] Y. Wen et al., "Non-line-of-sight sparse aperture ISAR imaging via a novel detail-aware regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5220218.
[7] M. Isogawa, Y. Yuan, M. O'Toole, and K. M. Kitani, "Optical non-line-of-sight physics-based 3D human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7013–7022.
[8] M. O'Toole, D. B. Lindell, and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," *Nature*, vol. 555, no. 7696, pp. 338–341, 2018.
[9] S. I. Young, D. B. Lindell, B. Girod, D. Taubman, and G. Wetzstein, "Non-line-of-sight surface reconstruction using the directional light-cone transform," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2020, pp. 1407–1416.
[10] X. Liu et al., "Non-line-of-sight imaging using phasor-field virtual wave optics," *Nature*, vol. 572, no. 7771, pp. 620–623, 2019.
[11] D. B. Lindell, G. Wetzstein, and M. O'Toole, "Wave-based non-line-of-sight imaging using fast *f-k* migration," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 1–13, 2019.
[12] D. Royo et al., "Virtual mirrors: Non-line-of-sight imaging beyond the third bounce," *ACM Trans. Graph.*, vol. 42, no. 4, pp. 1–15, 2023.
[13] S. Xin, S. Nousias, K. N. Kutulakos, A. C. Sankaranarayanan, S. G. Narasimhan, and I. Gkioulekas, "A theory of fermat paths for non-line-of-sight shape reconstruction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6800–6809.
[14] S. Shen et al., "Non-line-of-sight imaging via neural transient fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2257–2268, Jul. 2021.
[15] D. Royo et al., "Non-line-of-sight transient rendering," in *Proc. ACM SIGGRAPH Posters*, 2022, pp. 1–2.
[16] C. Wu et al., "Non–line-of-sight imaging over 1.43 km," *Proc. Natl. Acad. Sci.*, vol. 118, no. 10, 2021, Art. no. e2024468118.
[17] X. Liu, J. Wang, Z. Li, Z. Shi, X. Fu, and L. Qiu, "Non-line-of-sight reconstruction with signal–object collaborative regularization," *Light Sci. Appl.*, vol. 10, no. 1, 2021, Art. no. 198.
[18] R. Ding, J. Ye, Q. Gao, F. Xu, and Y. Duan, "Curvature regularization for non-line-of-sight imaging from under-sampled data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 8474–8485, Dec. 2024.
[19] J. Wang, X. Liu, L. Xiao, Z. Shi, L. Qiu, and X. Fu, "Non-line-of-sight imaging with signal superresolution network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 17420–17429.
[20] Y. Li, J. Peng, J. Ye, Y. Zhang, F. Xu, and Z. Xiong, "NLOST: Non-line-of-sight imaging with transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 13313–13322.
[21] X. Su, Y. Hong, J.-t. Ye, F. Xu, and X. Yuan, "Model-guided iterative diffusion sampling for NLOS reconstruction," *IEEE J. Sel. Top. Quantum Electron.*, vol. 30, no. 1: Single-Photon Technologies and Applications, Jan./Feb. 2024, Art. no. 7700111.
[22] Y. Li, Y. Zhang, J. Ye, F. Xu, and Z. Xiong, "Deep non-line-of-sight imaging from under-scanning measurements," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023, vol. 36, pp. 59095–59106.
[23] S. Zhang, S. Jin, H. Liu, Y. Li, X. Jiang, and M. Xu, "CMFormer: Non-line-of-sight imaging with a memory-efficient MetaFormer network," *Opt. Laser Eng.*, vol. 187, 2025, Art. no. 108875.
[24] G. Spoorthi, S. Gorthi, and R. K. S. S. Gorthi, "PhaseNet: A deep convolutional neural network for two-dimensional phase unwrapping," *IEEE Signal Process. Lett.*, vol. 26, no. 1, pp. 54–58, Jan. 2019.
[25] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Commun. ACM*, vol. 64, no. 3, pp. 107–115, Feb. 2021.
[26] W. Xu et al., "Feature enhanced non-line-of-sight imaging using graph model in latent space," *Opt. Laser Technol.*, vol. 181, 2025, Art. no. 111538.
[27] A. Assa and K. N. Plataniotis, "Wasserstein-distance-based gaussian mixture reduction," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1465–1469, Oct. 2018.
[28] X. Liu, J. Wang, L. Xiao, Z. Shi, X. Fu, and L. Qiu, "Non-line-of-sight imaging with arbitrary illumination and detection pattern," *Nature Commun.*, vol. 14, no. 1, 2023, Art. no. 3230.
[29] C. Zhao, J. Liu, and J. Zhang, "A dual model for restoring images corrupted by mixture of additive and multiplicative noise," *IEEE Access*, vol. 9, pp. 168869–168888, 2021.
[30] D. Bertsekas, A. Nedic, and A. Ozdaglar, "Convex analysis and optimization," *Athena Sci.*, vol. 1, 2003.
[31] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.
[32] U. S. Kamilov, H. Mansour, and B. Wohlberg, "A plug-and-play priors approach for solving nonlinear imaging inverse problems," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1872–1876, Dec. 2017.
[33] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, vol. 503, pp. 92–108, 2022.
[34] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," *Nature Commun.*, vol. 3, no. 1, 2012, Art. no. 745.
[35] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, pp. 89–97, 2004.
[36] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.
[37] U. Sara, M. Akter, and M. S. Uddin, "Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study," *J. Comput. Commun.*, vol. 7, no. 3, pp. 8–18, 2019.
[38] H. Wu, S. Liu, X. Meng, X. Yang, and Y. Yin, "Non-line-of-sight imaging based on an untrained deep decoder network," *Opt. Lett.*, vol. 47, no. 19, pp. 5056–5059, 2022.