This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Learnable Fractional Reaction-Diffusion Dynamics for Under-Display ToF Imaging and Beyond

Xin Qiao[1] Matteo Poggi[2] Xing Wei[3]
Pengchao Deng[1] Yanhui Zhou[1,*] Stefano Mattoccia[2]
[1]Xi'an Jiaotong University [2]University of Bologna [3]Anyang Institute of Technology

Abstract

*Under-display ToF imaging aims to achieve accurate depth sensing through a ToF camera placed beneath a screen panel. However, transparent OLED (TOLED) layers introduce severe degradations—such as signal attenuation, multi-path interference (MPI), and temporal noise—that significantly compromise depth quality. To alleviate this drawback, we propose **L**earnable **F**ractional **R**eaction-**D**iffusion **D**ynamics (LFRD^2), a hybrid framework that combines the expressive power of neural networks with the interpretability of physical modeling. Specifically, we implement a time-fractional reaction-diffusion module that enables iterative depth refinement with dynamically generated differential orders, capturing long-term dependencies. In addition, we introduce an efficient continuous convolution operator via coefficient prediction and repeated differentiation to further improve restoration quality. Experiments on four benchmark datasets demonstrate the effectiveness of our approach. The code is publicly available at https://github.com/wudiqx106/LFRD2.*

1. Introduction

The ascendancy of full-screen featuring a high screen-to-body ratio within the realm of intelligent terminals (e.g. smartphones) design has marked a significant leap forward in enhancing both user experience and aesthetic appeal. This trend underscores the importance of integrating cameras to facilitate an immersive and visually captivating interactive environment. Recently, extensive research [10, 50] on under-display RGB image restoration has been conducted by both academia and industry, leading to its successful deployment in the front cameras of mass-produced smartphones. Furthermore, as the pursuit of a truly seamless display experience advances, under-display depth cameras, such as under-display ToF (UD-ToF) sensors [32], have drawn considerable interest. This technology, capable of capturing three-dimensional spatial information through the screen panel, represents a crucial step forward in the evolution of intelligent terminal design. Indeed, compared to under-display RGB imaging, Transparent Organic Light-Emitting Diode (TOLED) panels pose more severe challenges for ToF cameras, such as reduced ranging accuracy and the loss of depth details, among others.

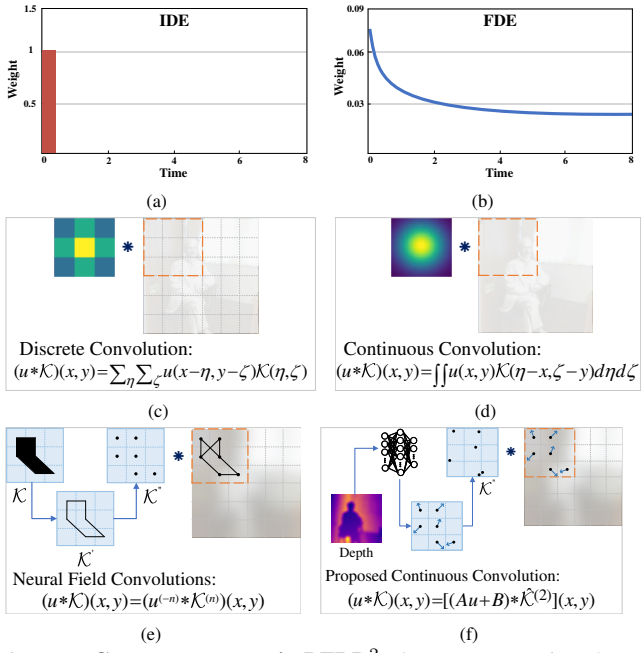


Figure 1. **Core components in LFRD^2.** On top: comparison between (a) Integer Differential Equation (IDE), (b) Fractional Differential Equation (FDE); At bottom: comparison between (c) Discrete Convolution and (d) Continuous Convolution, followed by some implementations of the latter, i.e., (e) Neural Field Convolution, and (f) ours.

In addressing these issues, classical diffusion processes that leverage domain-specific physical priors, such as Perona-Malik (P-M) diffusion [31] and its variants [38], are recognized as robust tools for improving depth accuracy and preserving details. Mathematically, given the spatial infor-

*Corresponding author

mation $\mathbf{z}$ on a bounded domain $\Omega \in \mathbb{R}^2$ and time t, the latent clear image $u(\mathbf{z})$ can be obtained by solving the equation

$$\begin{cases} \dfrac{\partial u}{\partial t} = \mathrm{div}(g|\nabla u|\nabla u), & (\mathbf{z}, t) \in \Omega \times (0, T] \\ u(\mathbf{z}, 0) = u_0(\mathbf{z}), & \mathbf{z} \in \Omega \end{cases} \tag{1}$$

with $u_0(\mathbf{z})$ being the initial condition, $g(\cdot)$ the diffusivity function and $u = u(\mathbf{z}, t)$ the solution at time t. These methods exhibit robust adaptability and generalization, yet the requirement for modeling numerous and complex parameters, extensive computational demands, and the neglect or misassumptions of non-primary factors in imaging degradation burden them. In this context, deep learning methods, with their strengths in high-level image understanding and contextual reasoning, have garnered considerable attention as potential solutions. However, their reliance on the meticulous design of network architectures, as well as the abundance and quality of data, remains a significant factor.

Notably, efforts [6, 27] to establish specific and systematic constraints between commonly used iterative algorithms in diffusion processes and deep neural networks – also known as algorithm unrolling – have shown promising advances in depth restoration. Nonetheless, these iterative methods typically employ integer differential equations (IDE), where the predicted state u_{n+1} depends solely on the current state u_n, as illustrated in Fig. 1(a). In real systems, the predicted state often depends not only on the current state but also on previous ones. Fig. 1(b) further illustrates the memory properties inherent to fractional-order dynamic systems, which capture this historical dependency. Some traditional methods [13, 22] have leveraged this property to advance image processing, whereas challenges in analytical solutions and parameter selection persist in fractional-order systems. This difficulty motivates us to harness the fitting capabilities of neural networks to approximate the solutions, thereby improving UD-ToF imaging quality.

In a diffusion step for image processing, the central pixel is updated based on a weighted combination of its neighboring pixels, effectively performing a discrete convolution to propagate information. However, scenes in the natural world are continuous rather than discrete, sparking significant interest in neural fields [39, 44, 46]. Also known as implicit neural representations, neural fields are typically implemented by Multi-Layer Perceptrons (MLPs) which learn a continuous function mapping spatial coordinates into signals or kernels [39, 44], these latter to replace the widespread discrete convolutions with continuous ones – both illustrated in Fig. 1(c) and (d). Despite demonstrating promising prospects in tasks such as 3D reconstruction and image super-resolution, they are still hindered by drawbacks like high computational costs and intricate hyperparameters tuning. Differently, using repeated differentiation [29], shown in Fig. 1(e), offers a pragmatic strategy for efficiently implementing continuous convolution, but its flexibility is limited by placing control points on fixed grids and predefining the convolution kernel.

In this paper, we develop a hybrid approach termed **L**earnable **F**ractional **R**eaction-**D**iffusion **D**ynamics (LFRD^2), which combines neural networks with physical modeling in an end-to-end training framework, enabling depth optimization in a coarse-to-fine manner. Here, the neural networks embedded within the time-fractional reaction-diffusion equation learn to optimize the iterative errors generated at each step based on previous states, rather than functioning as an end-to-end regressor. Notably, the differentiation orders are no longer fixed at predetermined values but are dynamically generated by a neural network. During the diffusion process, we propose a novel method for continuous convolution based on the properties of signal convolution, illustrated in Fig. 1(f), which can efficiently improve depth quality. This approach, simply leveraging several flat convolution layers instead of coordinate-based MLPs, achieves continuous convolution via parameter prediction while offering robust interpretability. This design enables efficient, interpretable continuous convolution, and offers potential extensibility to other depth-related tasks. Fig. 1 highlights the main properties differentiating our proposal concerning existing methods. Accordingly, our main contributions can be summarized as follows:

- We present a hybrid framework that integrates neural networks into a learnable fractional reaction-diffusion equation, leveraging prior physics knowledge to iteratively refine depth and enable effective learning with variable fractional order.
- We introduce an efficient continuous convolution operator that leverages coefficient prediction and repeated differentiation, boasting robust interpretability alongside parameter efficiency.
- The proposed framework was evaluated on two UD-ToF and two depth restoration benchmark datasets, confirming its theoretical and experimental consistency, and validating its effectiveness in UD-ToF imaging and beyond.

2. Related Work

We briefly review the literature relevant to our proposal.

Under-Display Sensor Imaging. Existing under-display sensor imaging is mainly divided into two types: Under-Display RGB and Under-Display ToF. Among them, the former was developed earlier. The optical system of an under-display RGB camera is analyzed for the first time by Zhou et al. [50], who also present a dataset for this analysis. The Point Spreading Function (PSF) of the under-display device is directly measured by utilizing a point light source [10], with this measurement being integrated as a pivotal component within their data synthesis process. A novel degradation model for under-display imaging is pro-

posed by Koh et al. [19], taking into account the color shift and signal attenuation that vary across different positions on the Transparent OLED screen. To address the challenge of low-contrast image enhancement, statistical properties of the H and S channels in HSV space of under-display images are analyzed, and a pixel-level estimation network is proposed by Luo et al. [25]. Feng et al. [11] design an innovative Transformer-based architecture to alleviate the non-negligible domain discrepancy and spatial misalignment, resulting in superior-quality target data. Recently, Liu et al. [24] proposed a network architecture that incorporates interactive learning between frequency and spatial domains to mitigate the effects of various scattering phenomena. Li et al. [21] design a lightweight network to estimate distortion-free images by leveraging wavelet transformation and multi-scale feature fusion. Since these methods fail to consider the physical correlations in ToF raw data, they cannot be directly applied to UD-ToF imaging.

A few frameworks were also designed to solve depth restoration for under-display ToF. The pioneering work [32] focuses on a depth restoration framework and synthetic data algorithm, tailored to overcome complex degradation in ToF imaging through Transparent OLED displays. A similar work [42] utilizes an optimized Restormer [48] to replace the second stage lightweight network by Qiao et al. [32]. Although these methods have achieved remarkable progress in under-display ToF depth restoration, they do not involve research on interpretability.

Nonlinear Diffusion for Image Enhancement. Nonlinear diffusion has been widely applied to address image enhancement. Traditional works usually leverage mathematical models to generate result images. To remove the noise of the image, a new model [23] based on the time-fractional diffusion equation is proposed, which is stable and the numerical solution converges. A denoising model [20] is built based on fractional-order and integer-order diffusions, taking advantage of texture-preserving and edge-preserving properties. For image denoising and restoration, a novel partial differential equation, utilizing a time-fractional order derivative, is proposed by Ben-Loghfyry and Charkaoui [2]. Recently, a coupled nonlinear diffusion system [9] is designed to both restore and binarize a degraded document image.

Since deep learning has become widely popular in various fields of image processing, researchers concentrate on incorporating diffusion models with deep learning methods. The trainable dynamic nonlinear reaction-diffusion model, featuring time-dependent filter parameters and influence functions learned from data, is introduced by Chen and Pock [4]. A denoising network [17] based on the discretization of a fractional-order differential equation is developed to consider long-term memory in both forward and backward passes. Moreover, Metzger et al. [27] leverage a novel approach utilizing guided anisotropic diffusion with a deep convolutional network for guided depth super-resolution. However, these methods either fail to account for the influence of previous states during the iteration, or expose unclear statistical or physical explainability.

Continuous Convolution. Continuous convolution has become popular in computer vision due to its advantages in handling irregular data and preserving original information. In 3D vision, increasing approaches adopt the continuous kernel to improve the quality of irregular 3D point clouds [26, 43, 45]. A small neural network can represent a convolutional kernel as a continuous function, enabling the parallel processing of arbitrarily long sequences within a single operation [36] A new approach [18] dynamically adjusts parameters, enabling continuous function construction via interpolation, which achieves a lightweight structure with enhanced performance. Recently, Nsampi et al. [29] propose to train a repeated integral field, requiring only a small number of point samples from the neural integral field to perform an exact continuous convolution. However, these methods bring high computational costs and intricate hyperparameter tuning.

3. Methodology

UD-ToF imaging strives to yield high-quality depth maps from corrupted raw measurements. To achieve this, we introduce our learnable fractional reaction-diffusion framework, comprised of two novel components. First, the overall paradigm of the proposed framework in Fig. 2 is presented. Then we elaborate on the fractional reaction-diffusion dynamics with the derivation of underlying physics. Finally, the efficient continuous convolution operator is illustrated.

3.1. Proposed Framework

As shown in Fig. 2, our framework primarily encompasses two processes: deep initial state builder (DISB) and deep fractional reaction-diffusion. DISB is introduced to generate the initial state u_0 for subsequent iterative depth refinement. For denoising, it serves as a denoising network or an identity mapping, while for image super-resolution, it functions as an upsampling network. In the deep fractional reaction-diffusion process, we combine the time-fractional reaction-diffusion equation with neural networks to iteratively refine the depth map, where u_t denotes the intermediate depth at iteration t. Since each current state depends on all previous ones, the recurrent refinement captures long-term memory. Mathematically, the evolution of this process is expressed as:

$$\begin{aligned} u_{n+1} &= \Phi(\mathbf{z}, t, u, \nabla_{\mathbf{z}} u, \cdots) \\ &= \sum_{t=0}^{n} w_t u_t + \mathcal{D}_t(u_n) + \mathcal{R}_t(u_n, u_0) \end{aligned} \tag{2}$$

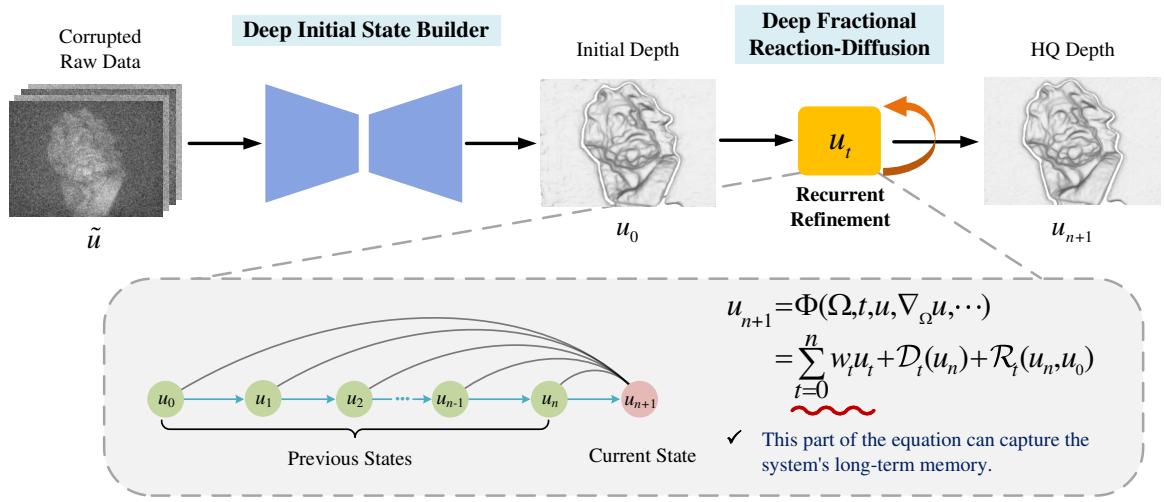


Figure 2. **Overview of LFRD^2.** Our framework deploys a Deep Initial State Builder, which can be any among the existing networks for UD-ToF imaging, to obtain an initial depth map. Then, the Deep Fractional Reaction-Diffusion module iteratively optimizes it to obtain the final, high-quality (HQ) depth map.

where Φ represents a nonlinear operator that characterizes the right-hand side of the PDE, w_t indicates the memory weight of the previous state in defining the present stage at time t, and $\nabla_{\mathbf{z}}$ is the gradient operator in spatial information. The $\mathcal{D}_t(\cdot)$ and $\mathcal{R}_t(\cdot)$ denote the diffusion term and reaction term, respectively. The detailed description of Eq. (2) will be provided in the subsequent section.

3.2. Fractional Reaction-Diffusion Dynamics

IDEs, such as the P-M model, rely solely on the current state for prediction during iteration, often leading to blur and artifacts [22] – for further details, see the **supplementary material**. In contrast, FDEs benefit from long-term dependence, accumulating historical information over iterations to better mitigate these drawbacks. Furthermore, the nonlocal properties of fractional FDEs offer a suitable framework for explaining dynamic processes that exhibit memory effects, thereby enhancing the description of real-world physical phenomena. In practice, three commonly used fractional-order derivatives are the Riemann-Liouville, Caputo, and Grünwald-Letnikov formulations, each exhibiting distinct characteristics in numerical computations. Among them, the Caputo derivative stands out for its clear physical interpretation and straightforward initial condition handling, making it suitable for physical and engineering modeling. Therefore, we adopt the Caputo derivative. For an order α $(0 < \alpha < 1)$ and a state $u(t)$, the Caputo derivative $D_t^\alpha u(t)$ can be expressed as:

$$ {}_0^C D_t^\alpha u(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-\tau)^{-\alpha} u'(\tau)\, d\tau \qquad (3) $$

where $\Gamma(\cdot)$ represents the Gamma function, while $u'(t)$ denotes the first-order derivative of the state $u(t)$. The integral is evaluated over the interval from 0 to t.

To solve fractional-order differentials, we discretize the Caputo derivative with $L1$ approximation. The formula at $t = t_{n+1}$ can be approximated as:

$$ {}_0^C D_t^\alpha u_{n+1} \approx \frac{(\Delta t)^{-\alpha}}{\Gamma(2-\alpha)} \sum_{k=0}^{n} a_k^{(\alpha)} [u_{n+1-k} - u_{n-k}] \qquad (4) $$

where $u_i = u(t_i, \mathbf{z})$, $i = 0, 1, 2, ..., n$, $a_k^{(\alpha)} = (k+1)^{1-\alpha} - k^{1-\alpha}, l \geq 0$. $(\cdot)^{(\alpha)}$ represents the derivative, distinguishing it from the power index.

Here, we choose the $L1$ approximation over other ones for two main reasons [28]: firstly, the estimation accuracy from the $L1$ approximation is adequate for our purposes, and secondly, the $L1$ approximation involves a lower computational complexity. From the physical perspective, the nonlinear fractional reaction-diffusion process can be formulated in an explicit numerical scheme as:

$$ {}_0^C D_t^\alpha u_{n+1} = \text{div}(g(|\nabla u_n|)\nabla u_n) + \lambda(u_0 - u_n) \qquad (5) $$

where $\text{div}(g(|\nabla u_n|)\nabla u_n$ is the Perona-Malik diffusion process $\mathcal{D}_t(\cdot)$ [31], and $\lambda(u_0 - u_n)$ is the reaction term $\mathcal{R}_t(\cdot)$, also known as the additional bias term, which serves to drive the depth evolution toward a target state while preserving essential features of the source. Here, $g(\cdot)$ is generated by a neural network with flat convolution layers, rather than derived from a conductance function [31]. Following TNRD [4], we set $\lambda = 0.01$.

When $\Delta t = 1$, we can derive the physical model-driven refinement by combining Eq. (4) and Eq. (5):

$$ \begin{aligned} u_{n+1} = u_n + S\,[&\text{div}(g|\nabla u_n|\nabla u_n) + \lambda(u_0 - u_n))] \\ &- \left[\sum_{k=1}^{n} a_k^{(\alpha)} (u_{n+1-k} - u_{n-k})\right] \end{aligned} \qquad (6) $$

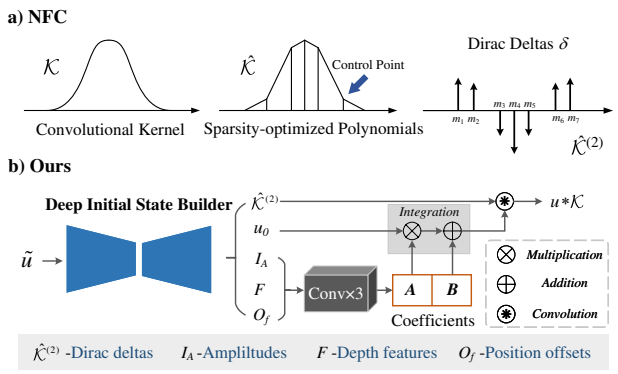


Figure 3. **Overview of our Continuous Convolution module.** Comparison between NFC and our proposal.

with $S = \frac{\Gamma(2-\alpha)}{a_0^\alpha}$. Building on the work [1] of Ashurov et al., the fractional order in the subdiffusion equation is guaranteed to exist and be unique when the initial condition is given (i.e., the depth map output by DISB) and appropriate boundary conditions (i.e., Neumann boundary condition [35]) and constraints are imposed. While noise may affect the numerical stability of the solution, it does not compromise its existence or uniqueness. This suggests that neural networks, as a flexible and efficient alternative to traditional numerical methods, can be leveraged to estimate the fractional order, which can be viewed as a form of physics-informed neural networks (PINNs).

The iterative refinement integrates the intrinsic physical information within the learnable diffusion dynamics [35], thereby endowing the entire process with interpretability and aligning it closely with the physical process.

3.3. Continuous Convolution

In iterative optimization, each step of image diffusion can be interpreted as a discrete convolution, which overlooks the inherent continuity of natural scenes. To address this, we implement $div(g(|\nabla u|)\nabla u)$ using a learnable continuous convolution operator, enabling flexible and data-adaptive spatial propagation. This motivation aligns with recent advances in neural fields, where continuous convolution is often implemented via MLPs to approximate coordinate-based functions, though such designs suffer from high computational cost and complex tuning. Employing repeated differentiation and integration [15], as shown in Fig. 3(a), can be an efficient pathway towards achieving continuous convolution [29]:

$$u * \mathcal{K} = \underbrace{\left(\int^n \cdots \int^n u \mathrm{d}\mathbf{z}_1^n \ldots \mathrm{d}\mathbf{z}_d^n \right)}_{u^{(-n)}} * \underbrace{\left(\frac{\partial^{dn}}{\partial \mathbf{z}_1^n \ldots \partial \mathbf{z}_d^n} \mathcal{K} \right)}_{\mathcal{K}^{(n)}} \quad (7)$$

where u and $\mathcal{K}$ represent signals and kernels, while $(\cdot)^{(-n)}$ and $(\cdot)^{(n)}$ denote, respectively, multidimensional repeated antiderivatives and derivatives, and n is the number of repeated operations.

In NFC Nsampi et al. [29], the authors introduce a predefined Gaussian kernel with a continuous second derivative and set the control points to approximate the kernel using piecewise linear functions. When n is set to 2, the estimated kernel $\hat{\mathcal{K}}^{(2)}$ reduces to a sparse set of Dirac deltas δ, which facilitates convolution calculations. Different from it, which predefines the Gaussian kernel and control points, and then estimates Dirac deltas for convolution, our method directly generates estimated Dirac deltas $\hat{\mathcal{K}}^{(2)}$ through DISB. This design enhances the flexibility of kernel selection and reduces the complexity of the estimated procedure. For the antiderivative computation $u^{(-2)}$ of the signal, a common approach is to use a trained coordinate-based neural network, typically structured as MLPs. However, the MLPs exhibit significant computational costs and are constrained to limited scenarios for training, inadequately addressing the demands of UD-ToF imaging. In contrast, we propose a simple yet efficient formulation of repeated antiderivatives in the continuous convolution based on Eq.(7). Our repeated antiderivatives are estimated as:

$$u^{(-2)} \approx Au(x_0, y_0) + B \quad (8)$$

where A and B are coefficients, and $(x_0, y_0) \in \mathbf{z}$ is the pixel where the continuous convolution is to be performed. The detailed proof of this process is given as:

$$\begin{aligned} &\int\int u \mathrm{d}\mathbf{z} \approx \sum_{i=0}^{m}\sum_{j=0}^{n} u(x_i, y_j) \cdot \Delta\mathbf{z} \\ &= u(x_0, y_0) \cdot \Delta\mathbf{z} + u(x_1, y_0) \cdot \Delta\mathbf{z} + \cdots + u(x_m, y_n) \cdot \Delta\mathbf{z} \\ &= [u(x_0, y_0) + C_{0,0}] \cdot \Delta\mathbf{z} + [u(x_0, y_0) + C_{1,0}] \cdot \Delta\mathbf{z} + \cdots \\ &\quad + [u(x_0, y_0) + C_{m,n}] \cdot \Delta\mathbf{z} \\ &= \underbrace{(m+1)(n+1)\Delta\mathbf{z}}_{A} \cdot u(x_0, y_0) + \underbrace{\sum_{i=0}^{m}\sum_{j=0}^{n} C_{i,j} \cdot \Delta\mathbf{z}}_{B} \end{aligned} \quad (9)$$

where $C_{0,0} = 0$. As shown in Fig. 3, the deep initial state builder additionally outputs amplitudes I_A, features F, and offsets O_f before iteration. During iteration, these are concatenated and processed through three convolution layers, with the middle layer having only 32 channels, to yield the coefficients A and B.

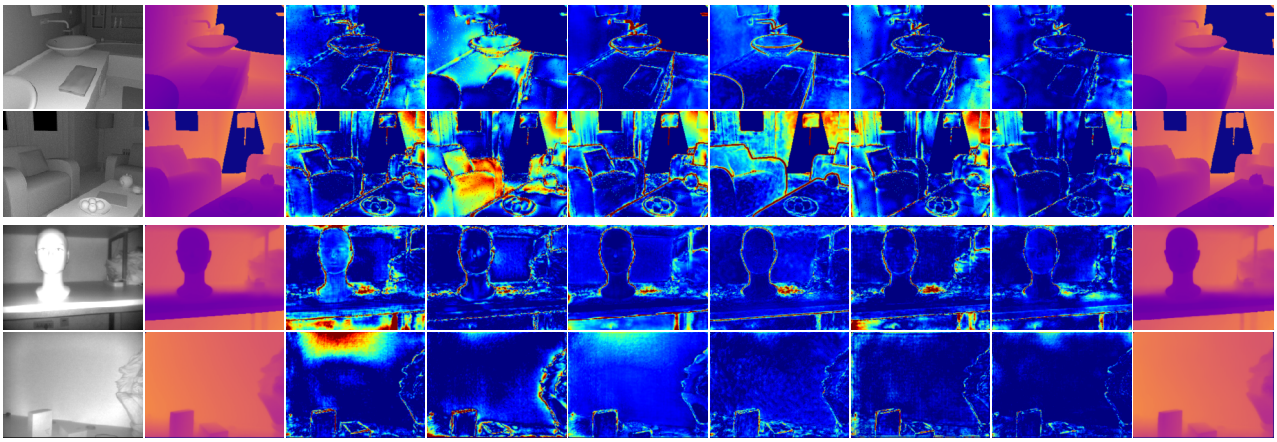
Although both our proposed continuous convolution and NFC [29] are inspired by repeated differentiation, there are fundamental differences in the process, from the generation of Dirac deltas to the computation of signal antiderivatives.

4. Experiments

We now report the outcome of our evaluation. We first introduce the experimental settings, then we compare LFRD[2]

Dataset	Metrics	CDNLM	JGDR	ToFnet	ToF-KPN	SHARPnet	PE-ToF	NAFNet	Restomer	UD-ToFnet	LFRD^2
	Input	Raw	Depth	Raw	Depth	Depth	Raw	Depth	Depth	Raw	Raw
SUD-ToF	MAE↓	33.23	9.14	10.34	13.39	14.84	9.77	11.08	9.75	<u>8.88</u>	**8.41**
	RMSE↓	48.43	34.43	28.28	21.05	23.00	15.92	18.24	14.76	<u>11.50</u>	**10.99**
	$\rho_{1.02}$ ↑	51.57	95.40	92.57	86.79	80.41	95.23	91.96	96.11	<u>97.09</u>	**97.19**
	$\rho_{1.05}$ ↑	87.64	97.82	97.01	98.77	94.22	98.76	98.20	99.10	<u>99.70</u>	**99.72**
	$\rho_{1.10}$ ↑	97.01	98.66	98.29	99.57	96.82	99.53	99.31	99.63	**99.94**	**99.94**
RUD-ToF	MAE↓	42.38	37.05	25.13	27.60	24.63	21.22	20.41	18.94	<u>17.29</u>	**16.73**
	RMSE↓	121.61	71.36	61.50	49.94	43.68	48.76	33.83	31.78	<u>31.11</u>	**30.94**
	$\rho_{1.02}$ ↑	63.99	48.04	66.23	61.65	56.04	62.03	67.30	68.41	**70.13**	<u>69.97</u>
	$\rho_{1.05}$ ↑	84.71	80.40	87.33	81.57	79.25	87.04	<u>90.08</u>	84.51	90.01	**90.66**
	$\rho_{1.10}$ ↑	92.09	91.79	95.17	89.90	90.01	95.39	<u>96.91</u>	94.92	96.74	**96.97**

Table 1. **Comparison with the state-of-the-art on the SUD-ToF and RUD-ToF datasets.** The best and second-best results are marked in **bold** and <u>underline</u>, respectively. The direction of arrows in metrics represents their trends (the lower/higher, the better).



(a) IR **(b)** GT **(c)** PE-ToF **(d)** SHARPnet **(e)** Restormer **(f)** NAFNet **(g)** UD-ToFnet **(h)** Ours **(i)** Our Pred

Figure 4. **Qualitative results on the SUD-ToF (top two rows) and RUD-ToF (bottom two rows) dataset.** From left to right: (a) IR image and (b) ground-truth depth, followed by (c-g) error maps achieved by SoTA solutions and (h) LFRD^2, (i) depth maps by LFRD^2.

with state-of-the-art solutions in UD-ToF imaging, conduct ablation studies, and conclude by discussing limitations.

4.1. Experimental Settings

We use two public UD-ToF datasets to evaluate the effectiveness of our proposal: SUD-ToF and RUD-ToF [32]. The former is a synthetic dataset crafted via transient rendering, encompassing 100K images, of which 10% are randomly allocated for testing. The latter is a real UD-ToF dataset featuring diverse indoor scenes, divided into 1171 scenes for training and 105 for evaluation. MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error), both measured in millimeters (mm), are used as evaluation metrics. Additionally, we measure the proportion (denoted as ρ_{th}) of pixels that fall within a specified relative error range compared to the total number of pixels. Following Qiao et al. [32], we set $th \in \{1.02, 1.05, 1.10\}$. To facilitate the network processing, the original image of 180×240 is cropped to a patch of 176×240. We use the Adam optimizer and a batch size of 16. The initial learning rate is set to 1×10^{-4}. The total number of epochs is 250 for SUD-ToF and 1000 for RUD-ToF. The proposed method is implemented using the Pytorch framework and experiments are conducted on a Nvidia RTX 3090 GPU. Moreover, the deep initial state builder is based on UD-ToFnet [32], keeping the original settings.

To prove the effectiveness of our framework beyond UD-ToF imaging, we conduct experiments on two additional datasets. FLAT [14] is a synthetic dataset of 2000 ToF measurements, capturing several nonidealities affecting real ToF sensors; we use it to perform ToF depth map denoising. NYUv2 [40] comprises video sequences from various indoor scenes, collected by both the RGB and Depth cameras from the Microsoft Kinect for a total of 1449 RGBD frames; we deploy this dataset to perform depth super-resolution, following standard settings from the literature [33].

4.2. Comparisons With State-of-the-Art Methods

To assess the effectiveness of LFRD^2, we compare it with state-of-art methods, including traditional algorithms like

Dataset	Metrics	CDNLM	JGDR	ToFnet	ToF-KPN	SHARPnet	Cardioid	PE-ToF	UD-ToFnet	LFRD^2
FLAT	MAE↓	13.86	8.86	54.33	4.65	4.62	6.74	7.93	<u>4.41</u>	**4.13**
	RMSE↓	21.00	45.78	74.99	12.83	10.26	19.94	32.60	<u>8.23</u>	**7.35**

Table 2. **Comparison with the state-of-the-art on the FLAT dataset.** The best and second best results are marked in **bold** and <u>underline</u>, respectively. The direction of arrows in metrics represents their trends (the lower/higher, the better).

Methods	4×	8×	16×
MSG	6.85 / 0.81	24.1 / 1.66	84.5 / 3.35
FDKN	9.07 / 0.85	29.9 / 1.80	113 / 3.95
PMBANet	10.8 / 0.93	17.2 / 1.38	84.9 / 3.26
FDSR	10.1 / 0.94	19.5 / 1.38	86.4 / 3.35
DCTNet	3.63 / 0.68	20.9 / 1.79	77.0 / 3.61
LGR	6.45 / 0.73	19.6 / 1.42	67.5 / 2.90
DADA	4.83 / 0.64	16.6 / 1.30	59.0 / 2.64
SGNet	3.22 / 0.54	14.9 / 1.26	58.8 / <u>2.63</u>
DSR-EI	<u>2.94</u> / <u>0.49</u>	<u>13.3</u> / <u>1.19</u>	<u>57.0</u> / 2.70
LFRD^2	**2.85 / 0.47**	**12.8 / 1.16**	**52.3 / 2.58**

Table 3. **Results on NYUv2 dataset.** We report MSE and MAE metrics, the lower the better.

Method	PE-ToF	NAFNet	Restomer	UD-ToFnet
Baseline	21.2 / 48.7	20.4 / 39.8	18.9 / 31.8	17.3 / 31.1
LFRD^2	**20.0 / 33.7**	**19.8 / 36.4**	**17.5 / 30.4**	**16.7 / 30.9**

Table 4. **Ablations on different baselines.** Our method adopts PE-ToF, NAFNet, Restomer, and UD-ToFnet as DISB to validate its effectiveness, with MAE / RMSE reported in the table.

CDNLM [12], JGDR [37] and learning-based frameworks, namely ToFnet [41], ToF-KPN [34],SHARPnet [8], PE-ToF [5], NAFNet [3], Restomer [48] and UD-ToFnet [32].

In Table 1, we present quantitative results on the SUD-ToF and RUD-ToF datasets. Unsurprisingly, traditional methods perform worse than learning-based approaches. Both LFRD^2 and UD-ToFnet [32] consistently outperform other methods, with LFRD^2 achieving better results than UD-ToFnet, except for the $\rho_{1.10}$ and $\rho_{1.02}$ scores on SUD-ToF and RUD-ToF, respectively.

We also report qualitative examples of depth maps restored by LFRD^2 compared to other learning-based methods. As shown by the error maps in Fig. 4, ToFnet [41], ToF-KPN [34], and PE-ToF [5] struggle to recover structural details, especially edges. Despite the improvement observed at edges for the result of SHARPnet [8], the irreversible loss of information inherent in the mapping from raw data to depth estimation methods poses challenges in achieving optimal depth estimation. Compared to other methods, our LFRD^2 demonstrates notable advantages in restoring depth quality and exhibits superior performance in edge-preserving, highlighting its effectiveness on the UD-ToF task.

4.3. Beyond UD-ToF imaging.

Furthermore, we demonstrate how LFRD^2 is also effective for enhancing the quality of the depth maps obtained through classical ToF sensors – e.g., not deployed under displays. Tab. 2 shows results concerning ToF denoising on the FLAT dataset [14]; it highlights that our framework significantly outperforms existing methods for this task. Additionally, Tab. 3 reports the results for the depth super-resolution task on the NYUv2 dataset [40], performed according to three different upsampling factors – 4×, 8× and 16×. Once again, LFRD^2 achieves the best results with any upsampling factor, outperforming state-of-the-art DSR-EI [33]. We refer the reader to the **supplementary material** for qualitative results.

4.4. Ablation Analysis

We now study in deeper detail the impact of the different modules composing our framework. For further details, see the **supplementary material**.

Ablations on different baselines. In Table 4, we show on the RUD-ToF dataset how different existing models can be used as internal state builder, and how any of them get improved by our approach.

Comparison with RNNs. In Table 5, we analyze the results by LFRD^2 and the use of three Recurrent Neural Networks (RNNs), i.e., dilated convolution [47], NLSPN [30], GRUs [7] and LSTM [16], all of which were integrated with the baseline model for iterative optimization of depth. "Dilated" refers to using dilated convolution with a fixed position rather than a flexible one to execute the diffusion process. NLSPN can be viewed as a diffusion process that employs deformable discrete convolution, exhibiting negligible performance gains compared to the baseline. Gated recurrent units (GRUs) and Long short-term memory (LSTM) are RNN variants, with the latter being used in particular to capture long-term dependency; both attain improvements in terms of MAE, with negligible improvements – or even drops – in RMSE. We ascribe this to the higher dependency of LSTM on large amounts of training data. Finally, although our goal is to privilege interpretability rather than outperform any alternative methods, we can appreciate that LRFD^2 consistently surpasses its counterparts, achieving state-of-the-art accuracy.

Furthermore, Table 5 includes ablations without Continuous Convolutions (CC) or Fractional Calculus (FC), where "w/o FC" corresponds to the integer-order variant. Results

Config.	Params / M	Flops / G	Speed / ms	MAE	RMSE
Baseline	2.17	8.65	15.20	17.29	31.11
Dilated	≈ 0	≈ 0	17.54	17.25	31.16
NLSPN	+0.01	+3.23	17.70	17.23	31.14
GRU	+0.18	+7.62	19.89	17.02	31.09
LSTM	+0.24	+10.4	22.15	16.96	31.22
Ours	+0.18	+7.69	22.75	**16.73**	**30.94**
w/o FC	+0.01	+0.41	20.67	17.00	30.99
w/o CC	+0.17	+7.28	22.11	16.88	31.03
NFC	+0.13	+20.5	28.42	16.97	31.00

Table 5. **Ablation study – LFRD^2 main components.** We compare LFRD^2 with RNN variants (top) and evaluate the impact of key modules – Fractional Calculus (FC), Continuous Convolution (CC), and NFC [29] – in the iterative process (bottom).

	Depth	Features	Amplitude	MAE	RMSE
I	✗	✗	✓	17.19	31.20
II	✗	✓	✗	17.16	31.22
III	✓	✓	✓	16.91	30.98
IV	✗	✓	✓	**16.73**	**30.94**

Table 6. **Ablation study – Continuous Convolution.** Comparison among four variants using different inputs.

show that both components are essential for optimal UD-ToF performance. We also compare with NFC [29], a continuous convolution baseline with similar accuracy but significantly higher FLOPs and runtime, highlighting the efficiency of our design.

Ablations on Continuous Convolution. This section presents how inputs of continuous convolution affect the performance of LFRD^2. We compare the outcomes of employing different concatenations of depth, intermediate features of depth (Simplified as features), amplitude, and offsets during the iteration. As shown in Table 6, we notice that concatenating amplitude and features results in optimal performance, whereas concatenating four elements fails to yield better results. Upon analysis, we ascribe this to the fact that during the iterative process, the depth tends to steer the attention of the network toward previous depth states, compromising final accuracy. Therefore, we select amplitude, features, and offsets as inputs to be forwarded to our continuous convolution modules.

Effect of Fractional Order Selection. Following FF [49], we compare our variable fractional order selection with different fixed orders ranging from 0 to 1. Tab. 7 shows that as the fractional order varies, the model performance undergoes notable changes in accuracy. Although our method achieves only marginal improvement over fixed orders of 0.1 and 0.2 in terms of $\rho_{1.02}$, it significantly outperforms them in the MAE metric. Overall, our variable order yields the optimal results.

Qualitative Results at Different Iterations To evaluate

Order	0.1	0.3	0.5	0.7	0.9	Ours
MAE/mm	18.12	18.38	18.86	19.01	18.29	**17.62**
$\sigma_{1.02}$ /%	66.79	66.80	66.16	65.73	66.04	**67.43**

Table 7. **Ablation study – fractional order.** Comparison between models with fixed fractional orders and ours.

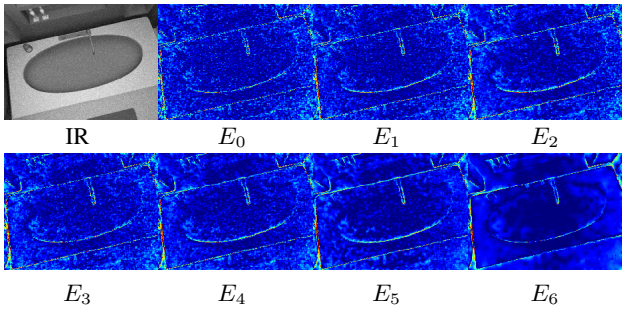


Figure 5. **Qualitative results of different iterations.** IR represents the IR image, and E_i denotes the error map after i-th iteration.

the effectiveness of our iterative process, we present the error maps at different iterations. In Fig. 5, as the iteration number increases, depth consistently and progressively approximates the ground truth depth, indicating a steady improvement in accuracy toward the desired outcome.

4.5. Limitations

Despite the higher interpretability, it is crucial to carefully set the training strategy to avoid instabilities – i.e., NaN values. In the future, we aim to develop a more robust method and validate its effectiveness across various tasks for depth restoration. Besides, we plan to devise a more efficient hybrid architecture based on the implicit numerical scheme.

5. Conclusion

In this paper, we proposed a physical model-driven deep framework for UD-ToF depth restoration, which integrates the underlying physical knowledge into a convolutional neural network, thereby iteratively facilitating the learning of spatio-temporal dynamics from the depth information. This approach encodes the time-fractional reaction-diffusion equation into the designed neural module, endowing the diffusion process with long-term memory properties. To further enhance depth quality, an efficient non-local continuous convolution operator is introduced. This operator enables the framework to achieve continuous convolution in a discrete form by predicting coefficients based on linear approximation and repeated differentiation. The experimental results demonstrate that our framework not only leverages the powerful representation learning capabilities of neural networks but also respects the underlying physics, resulting in more accurate and robust UD-ToF imaging.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62088102, No. U24A20265, No. 62376208), the Natural Science Basic Research Program of Shaanxi (Program No.2025SYS-SYSZD-021), the Natural Science Foundation of Sichuan Province (Grant No. 2025ZNSFSC1501), the China Postdoctoral Science Foundation (Certificate Number: 2025M771538), and the Postdoctoral Science Foundation of Shaanxi Province (Grant No. 2024BSHSDZZ100).

References

[1] Ravshan Ashurov and Sabir Umarov. Determination of the order of fractional derivative for subdiffusion equations. *Fractional Calculus and Applied Analysis*, 23(6): 1647–1662, 2020. 5

[2] Anouar Ben-Loghfyry and Abderrahim Charkaoui. Regularized perona & malik model involving caputo time-fractional derivative with application to image denoising. *Chaos, Solitons & Fractals*, 175:113925, 2023. 3

[3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Proceedings of the European Conference on Computer Vision*, pages 17–33. Springer, 2022. 7

[4] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1256–1272, 2016. 3, 4

[5] Yan Chen, Jimmy Ren, Xuanye Cheng, Keyuan Qian, Luyang Wang, and Jinwei Gu. Very power efficient neural time-of-flight. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2257–2266, 2020. 7

[6] Xinjing Cheng, Peng Wang, and Ruigang Yang. Learning depth with convolutional spatial propagation network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10):2361–2379, 2019. 2

[7] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1724–1734, 2014. 7

[8] Guanting Dong, Yueyi Zhang, and Zhiwei Xiong. Spatial hierarchy aware residual pyramid network for time-of-flight depth denoising. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*, pages 35–50. Springer, 2020. 7

[9] Zhongjie Du and Chuanjiang He. Nonlinear diffusion system for simultaneous restoration and binarization of degraded document images. *Computers & Mathematics with Applications*, 153:237–248, 2024. 3

[10] Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Chen Change Loy, and Jinwei Gu. Removing diffraction image artifacts in under-display camera via dynamic skip connection network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 662–671, 2021. 1, 2

[11] Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Jinwei Gu, and Chen Change Loy. Generating aligned pseudo-supervision from non-aligned data for image restoration in under-display camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5013–5022, 2023. 3

[12] Mihail Georgiev, Robert Bregović, and Atanas Gotchev. Time-of-flight range measurement in low-sensing environment: Noise analysis and complex-domain non-local denoising. *IEEE Transactions on Image Processing*, 27(6):2911–2926, 2018. 7

[13] Lin Guo, Xi-Le Zhao, Xian-Ming Gu, Yong-Liang Zhao, Yu-Bang Zheng, and Ting-Zhu Huang. Three-dimensional fractional total variation regularized tensor optimized model for image deblurring. *Applied Mathematics and Computation*, 404:126224, 2021. 2

[14] Qi Guo, Iuri Frosio, Orazio Gallo, Todd Zickler, and Jan Kautz. Tackling 3d tof artifacts through learning and the flat dataset. In *Proceedings of the European Conference on Computer Vision*, pages 368–383, 2018. 6, 7

[15] Paul S Heckbert. Filtering by repeated integration. *ACM SIGGRAPH Computer Graphics*, 20(4):315–321, 1986. 5

[16] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. 7

[17] Xixi Jia, Sanyang Liu, Xiangchu Feng, and Lei Zhang. Focnet: A fractional optimal control network for image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6054–6063, 2019. 3

[18] Sanghyeon Kim and Eunbyung Park. Smpconv: Self-moving point representations for continuous convolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10289–10299, 2023. 3

[19] Jaihyun Koh, Jangho Lee, and Sungroh Yoon. Bnudc: A two-branched deep neural network for restoring images from under-display cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1950–1959, 2022. 3

[20] Chengxue Li and Chuanjiang He. Fractional-order diffusion coupled with integer-order diffusion for multiplicative noise removal. *Computers & Mathematics with Applications*, 136: 34–43, 2023. 3

[21] Yuenan Li, Jin Wu, and Zetao Shi. Lightweight neural network for enhancing imaging performance of under-display camera. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(1):71–84, 2023. 3

[22] Wenhui Lian and Xinwu Liu. Non-convex fractional-order tv model for impulse noise removal. *Journal of Computational and Applied Mathematics*, 417:114615, 2023. 2, 4

[23] Xingran Liao and Minfu Feng. Time-fractional diffusion equation-based image denoising model. *Nonlinear Dynamics*, 103:1999–2017, 2021. 3

[24] Chengxu Liu, Xuan Wang, Shuai Li, Yuzhi Wang, and Xueming Qian. Fsi: Frequency and spatial interactive learning for image restoration in under-display cameras. In

Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 12537–12546, 2023. 3

[25] Jun Luo, Wenqi Ren, Tao Wang, Chongyi Li, and Xiaochun Cao. Under-display camera image enhancement via cascaded curve estimation. *IEEE Transactions on Image Processing*, 31:4856–4868, 2022. 3

[26] Jiageng Mao, Xiaogang Wang, and Hongsheng Li. Interpolated convolutional networks for 3d point cloud understanding. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1578–1587, 2019. 3

[27] Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Guided depth super-resolution by deep anisotropic diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18237–18246, 2023. 2, 3

[28] Kassem Mustapha. An l1 approximation for a fractional reaction-diffusion equation, a second-order error analysis over time-graded meshes. *SIAM Journal on Numerical Analysis*, 58(2):1319–1338, 2020. 4

[29] Ntumba Elie Nsampi, Adarsh Djeacoumar, Hans-Peter Seidel, Tobias Ritschel, and Thomas Leimkühler. Neural field convolutions by repeated differentiation. *ACM Transactions on Graphics*, 42(6):1–11, 2023. 2, 3, 5, 8

[30] Jinsun Park, Kyungdon Joo, Zhe Hu, Chi-Kuei Liu, and In So Kweon. Non-local spatial propagation network for depth completion. In *Proceedings of the European Conference on Computer Vision*, pages 120–136. Springer, 2020. 7

[31] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990. 1, 4

[32] Xin Qiao, Chenyang Ge, Pengchao Deng, Hao Wei, Matteo Poggi, and Stefano Mattoccia. Depth restoration in under-display time-of-flight imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5668–5683, 2023. 1, 3, 6, 7

[33] Xin Qiao, Chenyang Ge, Youmin Zhang, Yanhui Zhou, Fabio Tosi, Matteo Poggi, and Stefano Mattoccia. Depth super-resolution from explicit and implicit high-frequency features. *Computer Vision and Image Understanding*, 237: 103841, 2023. 6, 7

[34] Di Qiu, Jiahao Pang, Wenxiu Sun, and Chengxi Yang. Deep end-to-end alignment and refinement for time-of-flight rgb-d module. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9994–10003, 2019. 7

[35] Chengping Rao, Pu Ren, Qi Wang, Oral Buyukozturk, Hao Sun, and Yang Liu. Encoding physics to learn reaction–diffusion processes. *Nature Machine Intelligence*, 5(7):765–779, 2023. 5

[36] David W Romero, Anna Kuzina, Erik J Bekkers, Jakub M Tomczak, and Mark Hoogendoorn. Ckconv: Continuous kernel convolution for sequential data. *arXiv preprint arXiv:2102.02611*, 2021. 3

[37] Mattia Rossi, Mireille El Gheche, Andreas Kuhn, and Pascal Frossard. Joint graph-based depth refinement and normal estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12154–12163, 2020. 7

[38] Daniel Scharstein and Richard Szeliski. Stereo matching with nonlinear diffusion. *International Journal of Computer Vision*, 28:155–174, 1998. 1

[39] J Ryan Shue, Eric Ryan Chan, Ryan Po, Zachary Ankner, Jiajun Wu, and Gordon Wetzstein. 3d neural field generation using triplane diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20875–20886, 2023. 2

[40] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *Proceedings of the European Conference on Computer Vision*, pages 746–760. Springer, 2012. 6, 7

[41] Shuochen Su, Felix Heide, Gordon Wetzstein, and Wolfgang Heidrich. Deep end-to-end time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6383–6392, 2018. 7

[42] Yuan Sun, Hao Wei, Xin Qiao, Pengchao Deng, and Chenyang Ge. Under-display tof imaging with efficient transformer. In *2023 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 1–5. IEEE, 2023. 3

[43] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6411–6420, 2019. 3

[44] Cristina N Vasconcelos, Cengiz Oztireli, Mark Matthews, Milad Hashemi, Kevin Swersky, and Andrea Tagliasacchi. Cuf: Continuous upsampling filters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9999–10008, 2023. 2

[45] Shenlong Wang, Simon Suo, Wei-Chiu Ma, Andrei Pokrovsky, and Raquel Urtasun. Deep parametric continuous convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2589–2597, 2018. 3

[46] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. In *Computer Graphics Forum*, pages 641–676. Wiley Online Library, 2022. 2

[47] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *International Conference on Learning Representations*, 2016. 7

[48] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 3, 7

[49] Julio Zamora, Jesus A Cruz Vargas, Anthony Rhodes, Lama Nachman, and Narayan Sundararajan. Convolutional filter approximation using fractional calculus. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 383–392, 2021. 8

[50] Yuqian Zhou, David Ren, Neil Emerton, Sehoon Lim, and Timothy Large. Image restoration for under-display camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9179–9188, 2021. 1, 2