

FNIN: A Fourier Neural Operator-based Numerical Integration Network for Surface-from-gradients

Jiaqi Leng^{1,2}, Yakun Ju^{3*}, Yuanxu Duan⁴, Jiangnan Zhang⁵, Qingxuan Lv⁵,
Zuxuan Wu², Hao Fan^{5*}

¹Haide College, Ocean University of China

²Shanghai Key Lab of Intell. Info. Processing, School of CS, Fudan University

³School of Computing and Mathematic Sciences, University of Leicester

⁴Department of Computer Engineering, Qingdao City University

⁵College of Computer Science and Technology, Ocean University of China
jq leng24@m.fudan.edu.cn, kelvin.yakun.ju@gmail.com, fanhao@ouc.edu.cn

Abstract

Surface-from-gradients (SfG) aims to recover a three-dimensional (3D) surface from its gradients. Traditional methods encounter significant challenges in achieving high accuracy and handling high-resolution inputs, particularly facing the complex nature of discontinuities and the inefficiencies associated with large-scale linear solvers. Although recent advances in deep learning, such as photometric stereo, have enhanced normal estimation accuracy, they do not fully address the intricacies of gradient-based surface reconstruction. To overcome these limitations, we propose a Fourier neural operator-based Numerical Integration Network (FNIN) within a two-stage optimization framework. In the first stage, our approach employs an iterative architecture for numerical integration, harnessing an advanced Fourier neural operator to approximate the solution operator in Fourier space. Additionally, a self-learning attention mechanism is incorporated to effectively detect and handle discontinuities. In the second stage, we refine the surface reconstruction by formulating a weighted least squares problem, addressing the identified discontinuities rationally. Extensive experiments demonstrate that our method achieves significant improvements in both accuracy and efficiency compared to current state-of-the-art solvers. This is particularly evident in handling high-resolution images with complex data, achieving errors of fewer than 0.1 mm on tested objects.

Introduction

Numerical integration is foundational in computational mathematics, supporting various tasks in computer vision. Among these, reconstructing a 3D surface from its gradients, known as normal integration (Quéau, Durou, and Aujol 2017a), is crucial for accurate surface reconstruction. This process has become increasingly important as a follow-up step to photometric stereo (PS) (Shi et al. 2019a; Ju et al. 2023, 2024), a method that estimates 3D surface normals from varying light reflectance. Recent advancements in PS have greatly improved normal estimation by overcoming the limitations of the Lambertian reflectance model (Chen, Han, and Wong 2018; Ikehata 2018). To achieve precise 3D mesh

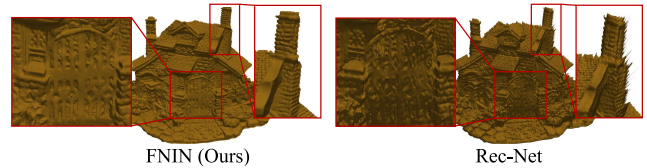


Figure 1: Reconstruction of the “HOUSE” in the LUCES (Mecca et al. 2021) dataset. Previous Rec-Net (Lichy, Sengupta, and Jacobs 2022) introduces sharp features and distortion at discontinuities, while our FNIN preserves the desired details.

reconstructions, Surface-from-gradients (SfG), as shown in Fig. 1, emerges as a natural progression of these PS advancements. Furthermore, methods like normal integration are also essential in other vision tasks, such as computed tomography (Bradshaw et al. 2013) and image restoration (Chen, Ng, and Zhao 2015; Yang et al. 2022).

Although various strategies have been employed to tackle the SfG problem, existing methods often struggle with noise and integrability issues (Badri, Yahia, and Aboutajdine 2014; Quéau, Durou, and Aujol 2017a; Cao et al. 2021). Most advanced approaches are optimization-based and rely on hand-crafted features (Cao et al. 2022; Kim, Jung, and Lee 2024). However, these features often oversimplify real-world complexities. Poorly designed objectives can also lead to high computational costs or severe numerical instability, hindering practical application and complicating integration with other algorithms, like the PS normal estimation network in (Lichy, Sengupta, and Jacobs 2022).

Inspired by advancements in other vision tasks (Peng et al. 2021; Ouyang, Wang, and Chen 2022; Jin et al. 2017; Fan et al. 2018), replacing optimization-based methods with end-to-end deep CNNs offers a promising improvement. However, straightforward applications (Lichy, Sengupta, and Jacobs 2022) have not achieved satisfactory results due to CNNs’ reliance on discretization and inherent bias toward locality, which is problematic for the Surface-from-gradients (SfG) problem that depends on global information. Fortunately, a recently proposed neural network architecture called the neural operator (NO), originally developed for

*Corresponding author.

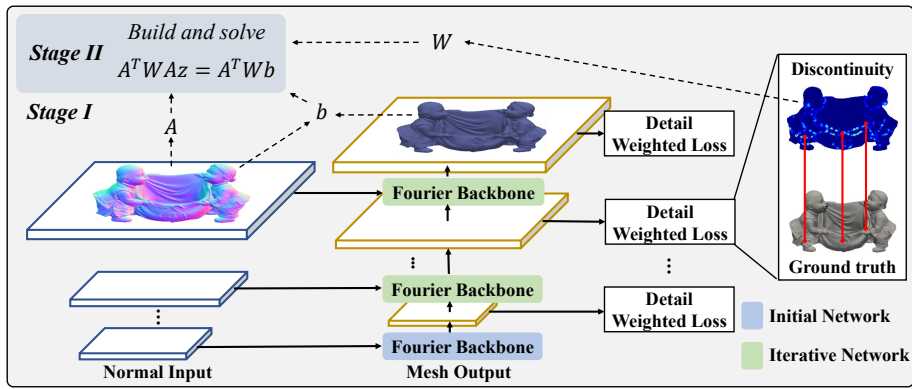


Figure 2: The Fourier neural operator-based numerical integration network operates within a two-stage framework. In stage I, the input normals are downsampled to different resolutions. In this iterative structure, the output depth is initially set by the initial network (blue) and subsequently refined by the iterative network (green). At each resolution, the Fourier backbone receives the upsampled output from the previous resolution and performs integration under the supervision of a detail weighted loss to preserve discontinuities. In stage II, the result is refined through a one-step weighted least squares optimization.

solving partial differential equations (PDEs) in computational physics (Lu et al. 2021; Li et al. 2020b,a), presents a potential solution. Neural operators leverage operator theory to learn mappings between infinite-dimensional function spaces without relying on PDEs as a physics prior. This capability enables us to introduce a learning-based approach to tackle the SfG problem effectively.

In this article, we propose a novel Fourier neural operator-based Numerical Integration Network (FNIN) to address the SfG problem using a learning-based approach. A key feature that distinguishes the SfG problem from typical NO equations is that, the discontinuities not satisfying the PDEs significantly influence the solution and tend to be extremely sparse, which will lead to confusion if learned pixel wise. To address this, we decouple the integration process of smooth area from optimization of discontinuity and introduce a two-stage scheme (Fig. 2). In the first stage, FNIN performs integration based on the formulated PDEs, approximating the solution through a Fourier neural operator (FNO) (Li et al. 2020a) in Fourier space, maintaining spatial globality. FNIN is constructed in a Rec-Net-like structure (Lichy et al. 2021) to ensure solution uniqueness. To detect discontinuities and mitigate disturbances, we incorporate an attention network, inspired by (Ju et al. 2020, 2022), to generate an adaptive attention map. Consequently, we creatively obtain the solution at smooth areas without optimization and transferring the discontinuities into a relative weight.

In the second stage, we globally optimize the sparse discontinuities to avoid determining huge jumps pixel wise. We formulate a weighted least squares problem following (Cao et al. 2022), but perform the optimization only once, as discontinuities are detected in the first stage, making the process more time-efficient. Our approach achieves state-of-the-art results on several objects from the DiLiGenT (Shi et al. 2019b) and LUCES (Mecca et al. 2021) datasets, demonstrating the novelty of our method. Additionally, we conduct experiments to validate our approach further. Our method excels in handling high-resolution images with com-

plex data, achieving errors of fewer than 0.1 mm on tested objects.

Our contributions are summarized as follows:

- We propose a Fourier neural operator-based integration network that performs integration on smooth areas and identifies discontinuities using a detail weighted loss.
- We develop an optimization process that ensures optimal solutions at discontinuities through a weighted least squares equation.
- Our experiments show that the proposed approach achieves exceptional accuracy and efficiency, making it a promising tool for future research.

Code — <https://github.com/nailwatts/FNIN>

Related Work

Our work is inspired by many existing studies but offers substantial novelty. In this section, we review several related works and highlight the key differences between them and our FNIN.

SfG Methods Based on Optimization

In the SfG problem, the ultimate goal is to solve the following unified partial differential equation

$$\nabla z = \mathbf{g}, \quad (1)$$

where z represents the depth from the camera, and $\mathbf{g} = [p, q]^T$ denotes the gradient map (Quéau, Durou, and Aujol 2017a). Ideally, when the surface is smooth, the solution should be independent of the integration path. However, in practice, due to natural discontinuities in real-world objects, perfect integrability is rarely achievable.

Recent works (Cao et al. 2022; Kim, Jung, and Lee 2024) struggled with a coupled depth and discontinuity weight, requiring repeated optimization to account for discontinuities, which consumes significant time. In contrast, the proposed FNIN leverages a neural network that directly approximates

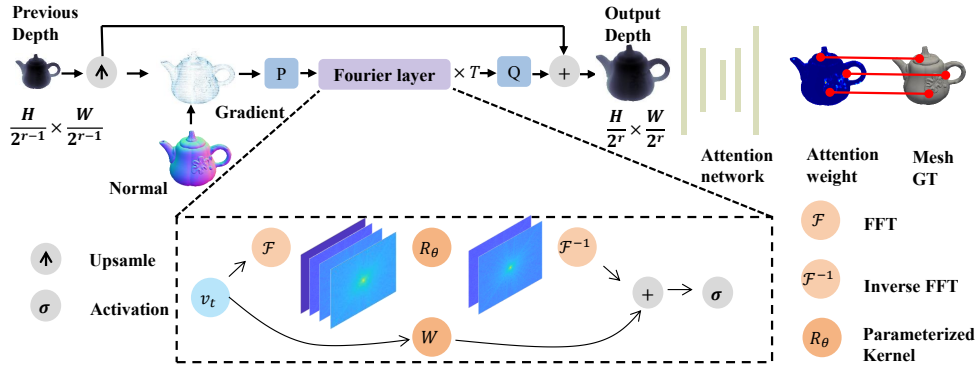


Figure 3: Integration is performed at an example resolution. Gradient \mathbf{g} for the input is computed using the normal and the upsampled output from the previous resolution with Eq. 5. The integration is then carried out within NO framework, with the solution approximated in Fourier space. Finally, the relative weight for discontinuities is estimated by the attention network.

the solution without relying on the PDEs, allowing us to decouple discontinuity detection from optimization. This data-driven detection method makes our approach highly innovative.

PDE Solvers Based on Neural Operator

The Neural Operator (NO) is a series of architectures designed to solve PDEs (Morton and Mayers 2005) in a purely data-driven manner. The core idea is to convert PDEs into a mapping between the parameter space and the solution space. Consider a general form of PDEs defined on $D \subseteq \mathbb{R}^d$

$$\begin{aligned} (\mathcal{L}_\alpha z)(\mathbf{u}) &= f(\mathbf{u}) \quad \mathbf{u} \in D, \\ z(\mathbf{u}) &= 0 \quad \mathbf{u} \in \partial D, \end{aligned} \quad (2)$$

where $z : D \rightarrow \mathbb{R}$ represents the solution in the Banach space $\mathcal{Z} \subseteq \mathbb{R}^{d_z}$, and $a : D \rightarrow \mathbb{R}$ is the parameter in the Banach space $\mathcal{A} \subseteq \mathbb{R}^{d_a}$ that determines the operator \mathcal{L}_α . The goal is to approximate the solution map $\mathcal{F}^\dagger : \mathcal{A} \rightarrow \mathcal{Z}$ by training a neural network with finite-dimensional parameters $\theta^\dagger \in \Theta$, such that $\mathcal{F} : \mathcal{A} \times \Theta \rightarrow \mathcal{Z}$ satisfies $\mathcal{F}(\cdot; \theta^\dagger) \approx \mathcal{F}^\dagger$.

Li et al. (2020b) observed that the operator learning task described above is equivalent to an empirical risk minimization problem, which can be solved through an iterative update. For $v_t \in \mathbb{R}^{d_v}$, $t = 0, 1, \dots, T$, the update is given by

$$v_{t+1}(\mathbf{u}) = \sigma(Wv_t(\mathbf{u}) + \mathcal{K}_\theta(a; \theta)v_t(\mathbf{u})), \quad (3)$$

where $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ is a non-linear activation function, $W : \mathbb{R}^{d_v} \rightarrow \mathbb{R}^{d_v}$ is a linear transformation, and the kernel integration is defined as

$$\mathcal{K}_\theta(a)(\mathbf{u}) := \int_D \kappa_\theta(\mathbf{u}, \mathbf{y}, a(\mathbf{u}), a(\mathbf{y}); \theta)v_t(\mathbf{y})d\mathbf{y}, \quad (4)$$

where $\kappa_\theta : \mathbb{R}^{2(d+d_a)} \rightarrow \mathbb{R}^{d_v \times d_v}$ is a neural network parameterized by $\theta \in \Theta$. Typically, a NO first lifts the input $a \in \mathcal{A}$ into a higher-dimensional representation by $P : \mathbb{R}^{d_a} \rightarrow \mathbb{R}^{d_v}$, such that $v_0(\mathbf{u}) = P(a(\mathbf{u}))$ and projects v_T back by $z(\mathbf{u}) = Q(v_T(\mathbf{u}))$, where $Q : \mathbb{R}^{d_v} \rightarrow \mathbb{R}^{d_z}$.

In this work, the choice of NO architectures is flexible. We apply the popular FNO (Li et al. 2020a) based on our assumptions, but it should be noted that these components can be replaced with other NO architectures, such as (Cao

2021; Raonic et al. 2023), if necessary. However, the most significant difference is that, unlike most physical equations, our SfG problem is characterized by severe non-uniqueness and discontinuities, which heavily influence the solution. To address this, we construct a Rec-Net (Lichy et al. 2021) structure to maintain solution uniqueness and incorporate an attention network with a detail weighted loss, which helps suppress discontinuities during training.

The Proposed FNIN

In this section, we introduce our Fourier neural operator-based numerical integration network (FNIN) with two stages. Fig. 3 illustrates the evaluation performed at one of the resolutions.

Stage I of FNIN: Approximating in Fourier Space

Variants of SfG equation in a Neural Operator settings

As Eqs. 2-4 apply universally to the class of equations represented by Eq. 2, a straightforward approach to solving Eq. 1 is to use a mature NO architecture. However, previous research (Quéau, Durou, and Aujol 2017a; Simchony, Chelappa, and Shao 1990) has shown that enforcing the Dirichlet boundary condition (setting the boundary value to 0) as implied by Eq. 2, may lead to problematic solutions, despite ensuring uniqueness. To avoid this conflict with the natural boundary condition in Eq. 1, a proper transformation is needed.

Inspired by (Lichy, Sengupta, and Jacobs 2022), we introduce an additional variable \hat{z} , which serves as an approximate solution to Eq. 1, such that

$$\begin{aligned} \nabla(z - \hat{z}) &= \mathbf{g} - \nabla\hat{z} \quad \mathbf{u} \in D, \\ z - \hat{z} &= 0 \quad \mathbf{u} \in \partial D, \end{aligned} \quad (5)$$

where $\mathbf{g} = [p, q]^T$, and \mathbf{u} represents the image coordinates. Ideally, if z and \hat{z} are smooth and fully satisfy Eq. 1, Eq. 5 would become an identity with a unique solution, $z \equiv \hat{z}$. In practice, we cannot provide such a solution, so we opt for a learning-based approach. Following (Lichy, Sengupta, and Jacobs 2022), and considering the lack of detail in the synthetic dataset (Lichy et al. 2021), we approximate \hat{z} using

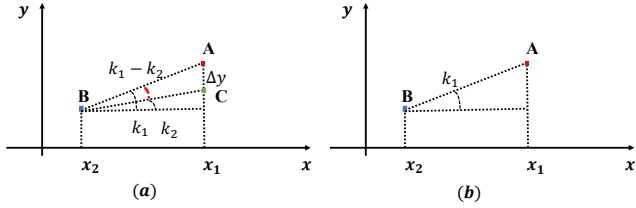


Figure 4: A simple example of (a) our assumption and (b) original natural boundary condition. Our assumption on given approximation green point C ensures the uniqueness otherwise the solution is ambiguous because value at other coordinates (blue point B) is also unknown.

the upsampled output from a lower resolution, and $\nabla \hat{z}$ is approximated by its central difference.

Integration operator in spatial domain After transforming our SfG problem into an NO setting, the assumption of kernel integration in Eq. 4 determines the specific architecture of the NO. It’s necessary to guarantee that our assumption will not break the uniqueness of the solution.

As shown in Fig. 4, in a 1D scenario, the green point C represents our known approximated solution, and our goal is to derive the value at the red point A. Our assumption is that Δy depends only on the distance $x_1 - x_2$ between the red point A and the blue point B, and the difference in slope $k_1 - k_2$. Here, $\mathbf{g} - \nabla \hat{z}$ corresponds to $k_1 - k_2$, and \hat{z} corresponds to the green point C. Once we derive Δy , we can determine the desired z . This example also illustrates the non-uniqueness issue caused by the original boundary condition, as the value at A cannot be determined solely by $x_1 - x_2$ and its slope k_1 as the value at B is also unknown.

Therefore, by applying our assumption to Eq. 4, we obtain

$$\mathcal{K}_\theta(\mathbf{u}) := \int_D \kappa_\theta(\mathbf{u} - \mathbf{y}; \theta) v_t(\mathbf{y}) d\mathbf{y}, \quad (6)$$

where $v_t(\mathbf{y})$ corresponds to the lifted $\mathbf{g} - \nabla \hat{z}$, and the dependence on the distance between coordinates is incorporated into the kernel. Clearly, this forms a perfect convolution with κ_θ as the kernel, enabling operator approximation in Fourier space. In fact, Eq. 6 can be interpreted as adaptively determining the integration path. Compared to earlier works (Wu and Li 1988; Karaçalı and Snyder 2003), our parameterized kernel offers significantly greater robustness.

Operator approximation in Fourier space As a convolution operator, a straightforward approach is to train a CNN to approximate Eq. 6. While increasing the receptive field could cover the entire domain D , the theory of effective receptive field (Luo et al. 2016) suggests that the network may struggle to focus globally, especially when locality is not included in Eq. 6. Moreover, since Eq. 6 is meant to operate on a continuous domain rather than a discretized one, we choose to approximate it in Fourier space.

Consider the Fourier transform \mathcal{F} of a function $f : D \rightarrow \mathcal{R}^{d_v}$ and its inverse \mathcal{F}^{-1} . By applying the convolution theorem, Eq. 6 becomes

$$\mathcal{K}_\theta(\mathbf{u}) = \mathcal{F}^{-1}(R_\theta(\mathcal{F}v_t))(\mathbf{u}), \quad (7)$$

where R_θ is parameterization tensor. As we cannot approximate infinite terms of the Fourier series, we truncate it with a maximum node k_{max} . Further detailed discussion of FNO including the case under discrete Fourier transform can be referred to in (Li et al. 2020a).

Adaptive discontinuity detection So far, we have considered integration in major smooth areas. In theory, our kernel in Eq. 6 can handle jumps at discontinuities. However, since discontinuities are much sparser than smooth areas, the network may make incorrect jumps, disrupting the approximation in other regions. To address this issue inspired by other vision tasks (Fang et al. 2024), we transfer them into a relative weight depending on their extension to be suppressed.

During the training process, we adopt a self-learning attention map. Unlike hand-crafted features in previous methods, we employ an extractor and regressor for attention, which can be formulated as follows

$$\begin{aligned} \Phi &= f_{\theta_{ae}}(\Delta z; \theta_{ae}), \\ \omega &= \text{normalize}(f_{\theta_{ar}}(\Phi; \theta_{ar})), \end{aligned} \quad (8)$$

where $f_{\theta_{ae}}$ and $f_{\theta_{ar}}$ are both three-layer CNNs, z represents the current depth estimation, and Δ represents the discontinuity modeled by the difference between the one-sided limit and its value measured in point-to-plane distance (Cao et al. 2021), following (Cao et al. 2022; Kim, Jung, and Lee 2024). The variable ω is the relative discontinuity weight, which is normalized to maintain values within $[0, 1]$.

To enable the attention network to detect discontinuities, we designed a detail weighted loss inspired by the ideas in (Ju et al. 2020, 2022), which has been used for detecting discontinuities in normals. For each resolution r , the normalized attention map ω balances the trade-off between the normal calculated from adjacent pixels and the absolute depth value. The loss function is expressed as

$$\mathcal{L}_r = \frac{1}{N} \sum_{i=1}^N ((1 - \omega_i) \|z_i - z_i^{gt}\|_1 + \gamma \omega_i \| \mathbf{n}_i(\mathbf{x}) - \mathbf{n}_i(\mathbf{x}^{gt}) \|_1), \quad (9)$$

where $N = |D|$ represents the number of all valid points, ω_i is the weight at pixel i in the attention map ω , and γ is a hyperparameter set to 0.25. The terms z_i and z_i^{gt} are the estimated and ground truth depths at pixel i , respectively, while \mathbf{x} and \mathbf{x}^{gt} represent the point clouds calculated via camera projection. The normals $\mathbf{n}_i(\mathbf{x})$ and $\mathbf{n}_i(\mathbf{x}^{gt})$ are computed using point clouds with central differences. Although many previous works have used similar losses to enhance high-frequency areas, we use the detail weighted loss to suppress discontinuities, which is a completely different approach. Combining all resolutions, the network is supervised by

$$\mathcal{L} = \sum_{r=1}^R \mathcal{L}_r \quad (10)$$

where R is the number of different resolutions.

Stage II of FNIN: Optimizing Discontinuity

To this end, our integration network has performed integration on most smooth areas and provided a discontinuity map. The remaining issue is to determine the jump at discontinuities through optimization.

Assumption	Method	Ball	Bell	Bowl	Buddha	Bunny	Cup	Die	Hippo	House	Jar	Owl	Queen	Squirrel	Tool	Runtime (s)
Smooth	HB	2.12	3.77	4.09	✘	5.80	1.25	3.30	5.95	8.61	7.46	5.43	6.68	2.22	5.72	> 6000
	FFT	0.50	4.29	4.66	23.78	4.85	0.80	2.28	5.59	9.57	3.57	4.77	5.79	1.64	4.96	0.23
	DCT	0.30	1.45	2.13	22.56	4.07	0.81	1.27	4.67	9.08	2.01	4.18	6.01	1.54	4.71	0.49
	DF	0.46	2.54	✘	✘	✘	0.37	✘	✘	✘	✘	4.15	✘	✘	2.71	136.75
	DP	0.17	0.33	0.38	✘	2.66	0.09	0.66	1.77	9.30	1.12	3.62	4.09	1.11	1.88	6.64
	ZS								OOM							
	IPF	✘	0.98	1.15	6.93	3.09	✘	✘	1.68	✘	3.06	3.86	1.86	1.84	1.11	45.22
Discontinue	L^1	0.29	1.47	2.12	6.88	3.81	0.21	1.21	4.61	9.03	1.96	2.89	4.42	1.51	4.42	110.77
	AD	0.18	0.36	0.46	3.40	2.62	0.11	0.68	2.06	9.26	1.41	3.22	3.71	1.15	0.80	59.70
	BiNI*	0.54	0.96	1.16	3.59	2.92	0.70	2.03	1.32	11.13	3.20	1.62	0.90	1.43	1.11	7.31
Learning	Rec-Net*	0.93	1.38	1.28	18.44	2.25	1.63	1.07	1.79	15.60	2.05	1.72	2.52	2.48	4.16	0.09
	FNIN*	0.15	0.12	0.08	3.35	2.45	< 0.01	0.70	1.26	9.96	0.28	3.77	3.38	1.18	0.28	1.81
	FNIN-S*	0.15	0.15	0.05	3.63	2.37	0.03	0.59	1.21	10.50	0.27	3.39	2.63	1.11	0.25	1.28

Table 1: Results on the LUCES (Mecca et al. 2021) Dataset. The metric for each object is Mean Absolute Error (MAE) in millimeters (mm). ✘ indicates a numerical failure, and OOM indicates an out-of-memory error. Methods marked with * are GPU-compatible and were tested on the same GPU. Best result for each object is highlighted in bold.

Method	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot 1	Pot 2	Reading
FFT	3.43	5.51	7.23	2.89	15.69	3.37	2.54	1.72	6.13
DST	3.41	5.71	7.26	2.89	15.66	3.39	2.56	1.73	6.16
DCT	2.99	5.72	7.29	2.80	15.26	3.34	2.48	1.69	5.75
ZS	1.82	3.80	1.72	1.16	12.72	11.17	2.00	0.78	7.09
IPF	0.89	3.33	1.48	0.86	11.23	11.11	1.73	0.64	6.56
WLS	2.52	3.50	6.12	2.10	14.75	2.60	8.00	1.15	4.54
TV	4.30	4.72	3.71	2.00	9.89	14.47	3.38	1.35	3.30
MS					✘				
Rec-Net					✘				
FNIN	1.35	3.74	1.69	1.18	12.61	11.17	1.96	0.82	7.07
FNIN-S	0.11	3.52	1.41	0.66	11.53	12.42	1.53	0.49	6.38

Table 2: Results on the DiLiGenT (Shi et al. 2019b) Dataset. Best result for each object is highlighted in bold.

The key idea is to suppress the effect of discontinuities in the objective function using weights, where the attention mechanism plays a crucial role. Since our attention-based discontinuity map is point-wise, meaning it assigns a value to each pixel, we can formulate a convex weighted least squares problem with four directions in the discrete case by

$$\begin{aligned}
\min_{z(u,v)} \sum_D \frac{w_r}{2} (n_3 \partial_u^+ z + n_1)^2 + \frac{w_l}{2} (n_3 \partial_u^- z + n_1)^2 \\
+ \frac{w_t}{2} (n_3 \partial_v^+ z + n_2)^2 + \frac{w_b}{2} (n_3 \partial_v^- z + n_2)^2 \\
+ \lambda (z(u,v) - z_R(u,v))^2,
\end{aligned} \quad (11)$$

where w_r, w_l, w_t, w_b are the weights at adjacent pixels in ω from the right, left, top, and bottom directions, respectively. z_R is the log depth output from FNIN, and $\lambda = 1e^{-3}$ is a fixed constant. We directly solve the normal equation of Eq. 11, similar to the approach in (Cao et al. 2022). Consider the flattened vectors $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$, and $\mathbf{z}_R \in \mathbb{R}^N$, where the depth map $\mathbf{z} \in \mathbb{R}^N$, flattened in the same order, is given by

$$\mathbf{A}^T \mathbf{W} \mathbf{A} \mathbf{z} = \mathbf{A}^T \mathbf{W} \mathbf{b}, \quad (12)$$

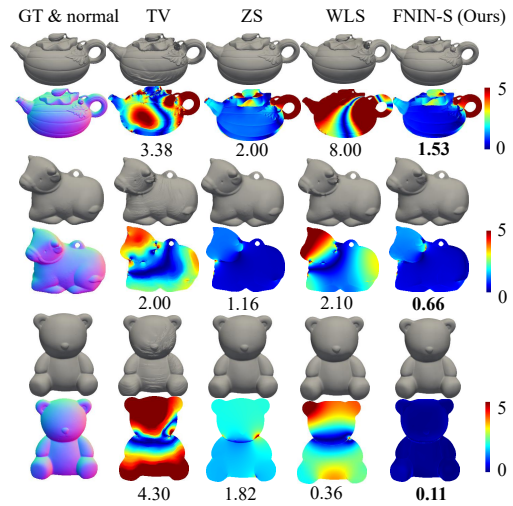


Figure 5: Qualitative results on “BRAR”, “COW” and “POT 1” objects from DiLiGenT (Shi et al. 2019b) Dataset. The black numbers under the error maps indicate the MAE (mm). The best result for each object is highlighted in bold.

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{N}_z \mathbf{D}_u^+ + \lambda \mathbf{1} \\ \mathbf{N}_z \mathbf{D}_u^- + \lambda \mathbf{1} \\ \mathbf{N}_z \mathbf{D}_v^+ + \lambda \mathbf{1} \\ \mathbf{N}_z \mathbf{D}_v^- + \lambda \mathbf{1} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -\mathbf{n}_1 + \lambda \mathbf{z}_R \\ -\mathbf{n}_1 + \lambda \mathbf{z}_R \\ -\mathbf{n}_2 + \lambda \mathbf{z}_R \\ -\mathbf{n}_2 + \lambda \mathbf{z}_R \end{bmatrix}, \quad \mathbf{W} = \text{diag} \left(\begin{bmatrix} w_r/2 \\ w_l/2 \\ w_t/2 \\ w_b/2 \end{bmatrix} \right), \quad (13)$$

and $\mathbf{N}_z = \text{diag}(\mathbf{n}_3)$, $\mathbf{D}_u^+, \mathbf{D}_u^-, \mathbf{D}_v^+, \mathbf{D}_v^- \in \mathbb{R}^{N \times N}$ are four discrete partial derivative matrices, $\mathbf{1} \in \mathbb{R}^N$ is all-one vector. We ultimately obtain the desired depth map after solving Eq. 12.

Experiments

Our FNIN is implemented in PyTorch (Paszke et al. 2017), with Eq.12 solved using the conjugate gradient method (Hestenes and Stiefel 1952). We used a synthetic dataset

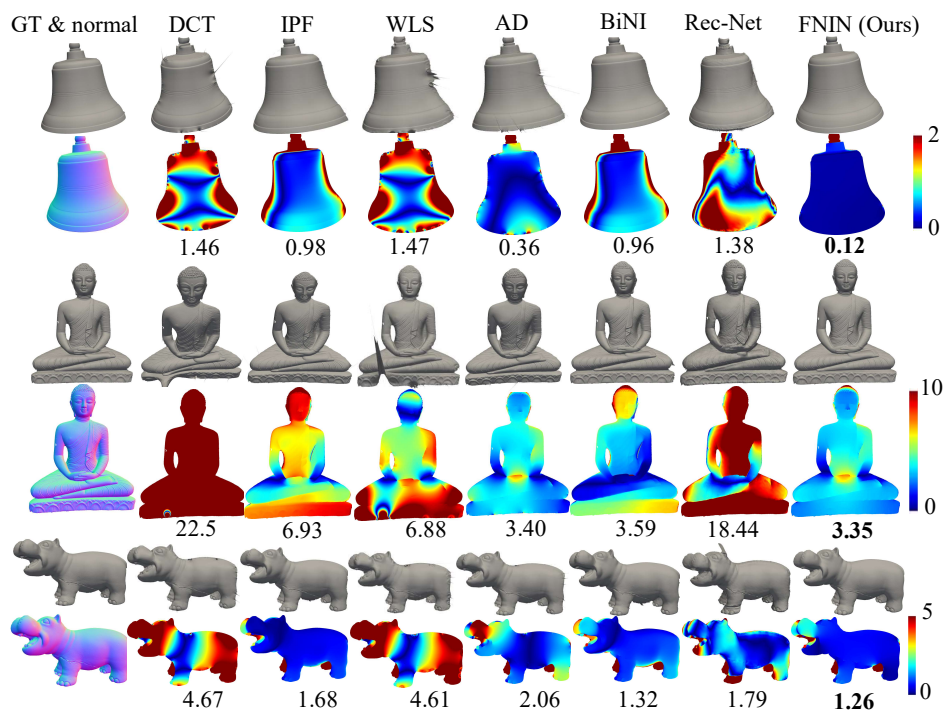


Figure 6: Qualitative results on “BELL”, “BUDDHA” and “HIPPO” objects from LUCES (Mecca et al. 2021) Dataset. Ground truth, input normals, estimated meshes, and absolute difference maps are shown in each pair of rows. The black numbers under the error maps indicate the MAE (mm). The best result for each object is highlighted in bold.

rendered by (Lichy, Sengupta, and Jacobs 2022), consisting of 14 objects from the statue dataset (Zisserman and Wiles 2017), with 1.75k training samples and 175 validation samples, each sized at 512×512 . To reduce memory consumption from large kernels, we applied Tucker factorization as in (Kossaihi et al. 2023). For discontinuity detection, we additionally provide a sigmoid function (FNIN-S) (Cao et al. 2022), handling discontinuities with varying intensities.

Comparison with Previous Methods

Previous optimization-based methods excel in generalization. Classic neural networks, based on the universal approximation theorem, often struggle with generalization errors (Lu et al. 2021), making them less competitive with traditional numerical PDE solvers, especially for purely geometric problems like SfG. In contrast, our FNIN, derived from minimizing cost functionals in infinite dimensions, achieves exceptional generalization. We conducted a comparison where our learning-based method, without fine-tuning, was directly compared with state-of-the-art numerical methods. We used the DiLiGenT (Shi et al. 2019b) and LUCES (Mecca et al. 2021) datasets. The DiLiGenT dataset contains 10 real-world objects scanned at 612×512 resolution, while the LUCES dataset includes 14 objects at 2048×1536 resolution, closer to common camera resolutions. This higher resolution challenges efficiency and detail preservation. We report the Mean Absolute Error (MAE) for each object, with results presented in Tables 1 and 2 and vi-

sualized in Figs. 5 and 6.

Comparison with methods based on smooth surface assumption For methods based on the smooth surface assumption, we selected several approaches: the symbolic Horn and Brooks (HB) method (Horn and Brooks 1986), the fast Fourier transform (FFT) integrability enforcing method (Frankot and Chellappa 1988), the discrete cosine transform (DCT), discrete sine transform (DST) integrability enforcing method (Simchony, Chellappa, and Shao 1990), the discretized functional (DF) variational method (Harker and O’Leary 2014), the 2D Savitzky-Golay filters variational method (ZS) (Zhu and Smith 2020), and the four-point inverse plane fitting (IPF) method (Cao et al. 2021).

Our approach outperforms others in both accuracy and stability. The attention network effectively prevents FNIN from being affected by large discontinuities (e.g., “BUDDHA” in Fig.6), and the discontinuity optimization step preserves these details, significantly improving mesh accuracy. Additionally, our iterative structure maintains numerical stability, minimizing spikes caused by exponential functions.

Comparison with methods focusing on discontinuity preservation For methods focusing on discontinuity preservation, we selected the variational method using Split-Bregman iterations to solve the weighted least squares problem (WLS) (Quéau and Durou 2015), the variational method using anisotropic diffusion (AD), isotropic total variation (TV), Mumford and Shah Functional (MS) (Quéau, Durou,

Experiment	Method	Parameters	DiLiGenT (612×512)		LUCES (2048×1536)		
			MAE (mm)	RAM+VRAM (GB)	MAE (mm)	Runtime (s)	RAM+VRAM (GB)
FNIN	w/o FNIN (BiNI, k=0)	–	4.59	1.6 + 1.0	2.79	0.91	2.0 + 2.0
	FNIN-S	1.53M	4.23	4.0 + 4.8	1.88	1.28	4.5 + 13.9
Structure*	w/o attention	1.05M	5.97	3.7 + 4.5	4.02	0.21	3.9 + 12.5
	w/ mask concatenation	1.53M	15.15	3.7 + 4.49	3.96	0.50	4.1 + 13.5
Discontinuity	w/o discontinuity optimization	1.53M	5.91	3.7 + 4.8	3.00	0.39	4.1 + 13.5
	FNIN	1.53M	4.55	4.1 + 4.8	1.92	1.66	4.6 + 13.5

Table 3: Ablation study of our approach. Discontinuity optimization is omitted in “Structure*”. “Runtime” shows the average evaluation time across all objects, and “RAM+VRAM” records the peak memory usage.

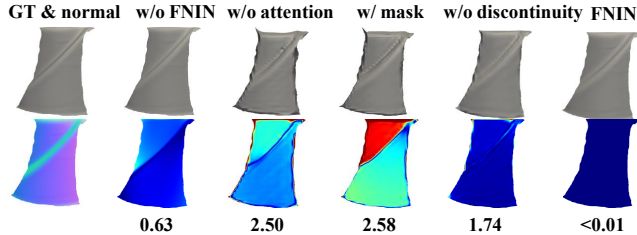


Figure 7: Ablation study on the “CUP” object from the LUCES (Mecca et al. 2021) Dataset. The black numbers under the error maps indicate the MAE (mm).

and Aujol 2017b), and the variational method using bilaterally weighted functional (Cao et al. 2022).

Our approach outperforms these methods in capturing fine details. For objects like “BELL” in Fig.6 and “BEAR” in Fig.5, where discontinuities are small and difficult to detect, our method performs significantly better. Notably, our accuracy on the “CUP” object in Table 1 is impressive, with an MAE below 0.01 mm. Additionally, FNIN integrates easily into other deep learning methods due to its efficiency on high-resolution images and ability to parallelize operations.

Comparison with Rec-Net To our knowledge, Rec-Net (Lichy, Sengupta, and Jacobs 2022) is the only deep learning method designed for the SfG problem. Key differences between our approach and Rec-Net are: First, while both use iterative structures, Rec-Net cannot be formulated into Eq. 6 due to CNNs’ inherent locality, which prevents it from guaranteeing the uniqueness of Eq. 5. Second, Rec-Net fails to account for discontinuities, causing irregular jumps and hindering convergence. Third, simply concatenating the input and mask in Rec-Net can cause confusion during transfer, leading to numerical failures (Table 2). Our FNIN addresses these issues, often outperforming traditional solvers, as shown in Table 1.

Ablation Study

In this section, we focus on the most crucial characteristics of our approach, including the effects of FNIN, the iterative structure, the attention network, and discontinuity optimization. Additionally, we analyze the computational cost and the handling of non-rectangular domains.

Ablation on the effect of FNIN Without FNIN, our approach degenerates into a smoother version of BiNI (Cao et al. 2022), resulting in blurred discontinuities and reduced accuracy (Fig. 7). Although optimization is performed only once, most runtime is spent solving the WLS problem, ensuring efficient training. The factorization we apply also facilitates integration into other deep learning algorithms.

Ablation on the effect of attention The attention mechanism in FNIN and the weights in discontinuity optimization suppress discontinuities at different levels. Attention in the detail-weighted loss prevents incorrect pixel-level normals from adjacent pixels. Without it, accuracy drops significantly at edges and discontinuities (Fig. 7), as the PDEs in Eq. 1 and Eq. 5 do not account for discontinuities, necessitating the decoupling of these processes. The WLS weights balance the impact of discontinuities and determine their jumps. Both components are essential.

Ablation on handling non-rectangular domain Handling arbitrary domains is crucial, as shown by Rec-Net’s instability on DiLiGenT. While the domain should ideally be embedded in the input (Liu, Jafarzadeh, and Yu 2023), our attention network allows us to directly multiply the gradient by the mask, ensuring edge discontinuities. Even if FNIN is affected by boundaries, the attention mechanism and discontinuity optimization resolve the issues seamlessly, as demonstrated by FNIN and FNIN-S results.

Ablation on discontinuity optimization Finally, our network, designed to avoid being misled by jumps at sparse discontinuities, cannot recover them individually. Due to this conflict, pixel-level accuracy in regions of discontinuity is not advisable. Removing discontinuity optimization, as shown in Table 3, results in a significant drop in accuracy.

Conclusion

In this work, we developed a numerical integration neural network within a two-stage framework for the Surface-from-gradients (SfG) problem. In the first stage, the solution is approximated by a neural operator in Fourier space, using a detail-weighted loss to address discontinuities. In the second stage, the identified discontinuities are optimized using a weighted least squares equation to determine their jumps rationally. Our experiments demonstrate that our approach achieves state-of-the-art results with exceptional efficiency.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 42106193 and Grant 41927805.

References

- Badri, H.; Yahia, H.; and Aboutajdine, D. 2014. Robust Surface Reconstruction via Triple Sparsity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2291–2298.
- Bradshaw, A. P.; Taubman, D. S.; Todd, M. J.; Magnussen, J. S.; and Halmagyi, G. M. 2013. Augmented Active Surface Model for the Recovery of Small Structures in CT. *IEEE Transactions on Image Processing*, 22(11): 4394–4406.
- Cao, S. 2021. Choose a Transformer: Fourier or Galerkin. In *Proceedings of the Advances in Neural Information Processing Systems*, volume 34, 24924–24940. Curran Associates, Inc.
- Cao, X.; Santo, H.; Shi, B.; Okura, F.; and Matsushita, Y. 2022. Bilateral Normal Integration. In *Proceedings of the European Conference on Computer Vision*.
- Cao, X.; Shi, B.; Okura, F.; and Matsushita, Y. 2021. Normal Integration via Inverse Plane Fitting with Minimum Point-to-Plane Distance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2382–2391.
- Chen, C.; Ng, M. K.; and Zhao, X.-L. 2015. Alternating Direction Method of Multipliers for Nonlinear Image Restoration Problems. *IEEE Transactions on Image Processing*, 24(1): 33–43.
- Chen, G.; Han, K.; and Wong, K.-Y. K. 2018. PS-FCN: A Flexible Learning Framework for Photometric Stereo. In *Proceedings of the European Conference on Computer Vision*, 3–19.
- Fan, L.; Huang, W.; Gan, C.; Ermon, S.; Gong, B.; and Huang, J. 2018. End-to-End Learning of Motion Representation for Video Understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6016–6025.
- Fang, X.; Easwaran, A.; Genest, B.; and Suganthan, P. N. 2024. Your Data Is Not Perfect: Towards Cross-Domain Out-of-Distribution Detection in Class-Imbalanced Data. *ESWA*.
- Frankot, R.; and Chellappa, R. 1988. A Method for Enforcing Integrability in Shape from Shading Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4): 439–451.
- Harker, M.; and O’Leary, P. 2014. Regularized Reconstruction of a Surface from its Measured Gradient Field. *Journal of Mathematical Imaging and Vision*, 51(1): 46–70.
- Hestenes, M. R.; and Stiefel, E. 1952. Methods of Conjugate Gradients for Solving Linear Systems. *Journal of research of the National Bureau of Standards*, 49: 409–435.
- Horn, B. K.; and Brooks, M. J. 1986. The Variational Approach to Shape from Shading. *Computer Vision, Graphics, and Image Processing*, 33(2): 174–208.
- Ikehata, S. 2018. CNN-PS: CNN-Based Photometric Stereo for General Non-convex Surfaces. In *Proceedings of the European Conference on Computer Vision*, 3–19.
- Jin, K. H.; McCann, M. T.; Froustey, E.; and Unser, M. 2017. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing*, 26(9): 4509–4522.
- Ju, Y.; Lam, K.-M.; Chen, Y.; Qi, L.; and Dong, J. 2020. Pay Attention to Devils: A Photometric Stereo Network for Better Details. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 694–700.
- Ju, Y.; Lam, K.-M.; Xiao, J.; Zhang, C.; Yang, C.; and Dong, J. 2023. Efficient feature fusion for learning-based photometric stereo. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.
- Ju, Y.; Lam, K.-M.; Xie, W.; Zhou, H.; Dong, J.; and Shi, B. 2024. Deep Learning Methods for Calibrated Photometric Stereo and Beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Ju, Y.; Shi, B.; Jian, M.; Qi, L.; Dong, J.; and Lam, K.-M. 2022. Normattention-psn: A High-frequency Region Enhanced Photometric Stereo Network with Normalized Attention. *International Journal of Computer Vision*, 130(12): 3014–3034.
- Karaçalı, B.; and Snyder, W. 2003. Reconstructing discontinuous surfaces from a given gradient field using partial integrability. *Computer Vision and Image Understanding*, 92(1): 78–111.
- Kim, H.; Jung, Y.; and Lee, S. 2024. Discontinuity-preserving Normal Integration with Auxiliary Edges. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11915–11923.
- Kossaifi, J.; Kovachki, N.; Azizzadenesheli, K.; and Anandkumar, A. 2023. Multi-grid Tensorized Fourier Neural Operator for High-resolution PDEs. *arXiv:2310.00120*.
- Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Bhat-tacharya, K.; Stuart, A.; and Anandkumar, A. 2020a. Fourier Neural Operator for Parametric Partial Differential Equations. *arXiv:2010.08895*.
- Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Bhat-tacharya, K.; Stuart, A.; and Anandkumar, A. 2020b. Neural Operator: Graph Kernel Network for Partial Differential Equations. *arXiv:2003.03485*.
- Lichy, D.; Sengupta, S.; and Jacobs, D. W. 2022. Fast Light-Weight Near-Field Photometric Stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12602–12611.
- Lichy, D.; Wu, J.; Sengupta, S.; and Jacobs, D. W. 2021. Shape and Material Capture at Home. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6119–6129.
- Liu, N.; Jafarzadeh, S.; and Yu, Y. 2023. Domain Agnostic Fourier Neural Operators. In *Proceedings of the Advances in Neural Information Processing Systems*, volume 36, 47438–47450. Curran Associates, Inc.

- Lu, L.; Jin, P.; Pang, G.; Zhang, Z.; and Karniadakis, G. E. 2021. Learning Nonlinear Operators via DeepONet Based on the Universal Approximation Theorem of Operators. *Nature Machine Intelligence*, 3(3): 218–229.
- Luo, W.; Li, Y.; Urtasun, R.; and Zemel, R. 2016. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. In *Proceedings of the Advances in Neural Information Processing Systems*, 4905–4913.
- Mecca, R.; Logothetis, F.; Budvytis, I.; and Cipolla, R. 2021. Lucas: A dataset for Near-field Point Light Source Photometric Stereo. arXiv:2104.13135.
- Morton, K. W.; and Mayers, D. F. 2005. *Numerical Solution of Partial Differential Equations: An Introduction*. Cambridge university press.
- Ouyang, H.; Wang, T.; and Chen, Q. 2022. Restorable Image Operators with Quasi-Invertible Networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(2): 2008–2016.
- Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; and Lerer, A. 2017. Automatic differentiation in pytorch.
- Peng, C.; Liao, H.; Wong, G.; Luo, J.; Zhou, S. K.; and Chellappa, R. 2021. XraySyn: Realistic View Synthesis From a Single Radiograph Through CT Priors. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1): 436–444.
- Quéau, Y.; and Durou, J.-D. 2015. Edge-Preserving Integration of a Normal Field: Weighted Least-Squares, TV and L^1 Approaches. In *Proceedings of Scale Space and Variational Methods in Computer Vision*, 576–588. Springer.
- Quéau, Y.; Durou, J.-D.; and Aujol, J.-F. 2017a. Normal Integration: A survey. *Journal of Mathematical Imaging and Vision*, 60(4): 576–593.
- Quéau, Y.; Durou, J.-D.; and Aujol, J.-F. 2017b. Variational Methods for Normal Integration. *Journal of Mathematical Imaging and Vision*, 60(4): 609–632.
- Raonic, B.; Molinaro, R.; De Ryck, T.; Rohner, T.; Bartolucci, F.; Alaifari, R.; Mishra, S.; and de Bézenac, E. 2023. Convolutional Neural Operators for Robust and Accurate Learning of PDEs. In *Proceedings of the Advances in Neural Information Processing Systems*, volume 36, 77187–77200. Curran Associates, Inc.
- Shi, B.; Mo, Z.; Wu, Z.; Duan, D.; Yeung, S.-K.; and Tan, P. 2019a. A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(02): 271–284.
- Shi, B.; Mo, Z.; Wu, Z.; Duan, D.; Yeung, S.-K.; and Tan, P. 2019b. A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2): 271–284.
- Simchony, T.; Chellappa, R.; and Shao, M. 1990. Direct Analytical Methods for Solving Poisson Equations in Computer Vision Problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5): 435–446.
- Wu, Z.; and Li, L. 1988. A Line Integration Based Method for Depth Recovery from Surface Normals. In *Proceedings of International Conference on Pattern Recognition*, 591–595.
- Yang, Z.; Yao, M.; Huang, J.; Zhou, M.; and Zhao, F. 2022. SIR-Former: Stereo Image Restoration Using Transformer. In *Proceedings of the 30th ACM International Conference on Multimedia*, 6377–6385.
- Zhu, D.; and Smith, W. A. P. 2020. Least Squares Surface Reconstruction on Arbitrary Domains. In *Proceedings of the European Conference on Computer Vision*, 530–545. Springer.
- Zisserman, A.; and Wiles, O. 2017. SilNet: Single-and Multi-View Reconstruction by Learning from Silhouettes. In *Proceedings of the British Machine Vision Conference*. British Machine Vision Association and Society for Pattern Recognition.