Dual Graph Regularized Deep Unfolding Network for Guided Depth Map Super-resolution

Zhiwei Zhong[1], Peilin Chen[1*], Qiangqiang Shen[1], Bo Li[2], Shiqi Wang[1*]
[1]City University of Hong Kong, Hong Kong SAR, China
[2]vivo BlueImage Lab, vivo Mobile Communication Co., Ltd., China

Abstract

Depth map super-resolution with color guidance is a fundamental task in computer vision that aims to reconstruct high-resolution depth maps by leveraging structural correlations from corresponding guidance images. Recently, with the development of deep learning techniques, the performance of guided depth super-resolution (GDSR) models has been significantly improved. However, most existing approaches rely on black-box architectures that lack theoretical interpretability. Although graph optimization has been explored to integrate model-driven and data-driven frameworks, it remains computationally expensive and struggles to preserve the intrinsic structures of the depth maps. To overcome these limitations, we propose a novel GDSR framework based on a dual graph Laplacian prior, termed LapNet, which efficiently unfolds graph optimization into a deep neural network. Specifically, we first formulate a dual graph Laplacian prior that separately models structural dependencies along the row and column dimensions of the depth maps. This formulation explicitly enforces piecewise smoothness while reducing computational complexity from $\mathcal{O}(H^3W^3)$ to $\mathcal{O}(H^3+W^3)$ by avoiding the construction of global affinity graph. Furthermore, we develop a deep implicit prior to extract high-frequency structural cues from the guidance image, serving as a complementary component to the manually designed prior. Finally, we integrate these complementary priors into a unified variational optimization framework, which is efficiently solved through alternating minimization and subsequently unfolded into an interpretable multi-stage deep network. Extensive experiments on both synthetic and real-world datasets demonstrate that LapNet achieves state-of-the-art performance while maintaining low computational complexity.

1. Introduction

Scene depth, which records the per-pixel distance between objects in 3D space, is essential for various computer vision applications [1–4]. Over the past decades, we have seen a democratization of sensing technologies, and many consumer-grade RGB-D cameras have been introduced [5]. However, the depth maps captured by these depth sensors typically have a lower resolution than their associated color images. For example, the resolution of the depth maps acquired by the PMD CamCube camera is only about 200×200. Even for Microsoft Kinect v2, a widely used RGB-D camera, the resolution of the captured depth maps is only 512×424, which is still lower than its associated color images (1920×1080) [6]. These low-resolution (LR) depth maps restrict the applications based on them, especially for the tasks that require the resolution of the depth map to be the same as the color images [7–9]. Thus, it is crucial to develop effective super-resolution (SR) algorithms to bridge the resolution gap between depth maps and color images.

In practical scenarios, we can easily obtain aligned high-resolution (HR) color images when acquiring the depth maps. Thus, most existing depth SR approaches are designed to leverage the HR color image as a prior to guide the reconstruction of the depth map [11–19]. Conventional guided depth map super-resolution (GDSR) approaches for-

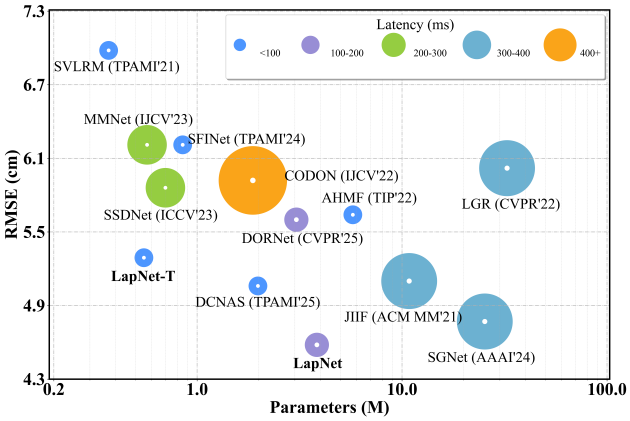


Figure 1. Quantitative comparison on the NYU v2 dataset [10] for 16× GDSR. "LapNet" denotes our full model with 24 feature maps, while "LapNet-T" represents a lightweight variant with 8 feature maps. The latency of each model is measured on the same NVIDIA RTX 5090 GPU using a 60×60 LR depth map as input.

*Co-corresponding authors.

mulate the task as an optimization problem in which multiple priors are imposed to limit the solution space and counteract the ill-posedness. Representative examples include the graph Laplacian prior [20–23] and Total Variation [24]. These approaches offer clear theoretical interpretability and allow explicit control over the optimization process. However, the limited representational capacity of handcrafted priors often leads to suboptimal performance, and the iterative optimization procedure is usually time-consuming.

Recently, the development of deep learning algorithms has substantially advanced the field of GDSR [25–33]. Compared with traditional model-based methods, learning-based approaches offer greater flexibility and more powerful feature representation capabilities. As a result, they typically achieve superior reconstruction accuracy and visual fidelity. However, due to their black-box nature, it remains difficult to interpret the roles of individual components. Moreover, performance improvements are often obtained by stacking additional modules, which increases computational complexity and further reduces model transparency. To bridge the gap between interpretability and performance, recent works [34, 35] have explored hybrid approaches that incorporate graph optimization into deep networks. Despite their promise, these methods still face notable limitations. On one hand, fully connected graphs capture long-range dependencies but incur prohibitive computational costs due to large-scale matrix multiplications and inversions. On the other hand, local neighborhood graphs offer improved efficiency but are limited to modeling short-range structures. In addition, many existing methods flatten the target image into a 1D vector when constructing the Laplacian, thereby disrupting the natural 2D spatial structure and potentially causing the loss of geometric details.

To overcome these challenges, we propose a novel guided depth map super-resolution framework, termed LapNet, which unfolds a model-based optimization formulation into a deep neural architecture. The core of LapNet is a dual graph Laplacian prior, which captures structural dependencies along the row and column dimensions through two compact affinity graphs. This design avoids constructing a dense global graph, and instead derives two low-dimensional Laplacian matrices, explicitly enforcing piecewise smoothness while reducing complexity from $\mathcal{O}(H^3W^3)$ to $\mathcal{O}(H^3 + W^3)$. In addition, we introduce a deep implicit prior, implemented via a lightweight neural network, to effectively extract structural details from the guidance image. By jointly integrating the data fidelity term, dual Laplacian regularization, and the deep implicit prior into a unified variational framework, we formulate an alternating minimization algorithm that is subsequently unfolded into an interpretable, multi-stage deep network. This hybrid design allows LapNet to benefit from both handcrafted priors and data-driven learning, achieving a favorable balance between accuracy, efficiency, and interpretability. The main contributions of this paper are as follows:

- We propose a dual graph Laplacian prior for guided depth map super-resolution, which constructs low-dimensional affinity graphs by independently capturing structural relationships in the row and column spaces of the input images. This formulation not only preserves piecewise smoothness, but also reduces computational complexity.
- We propose LapNet, a novel interpretable deep unfolding network for guided depth super-resolution (GDSR), which seamlessly integrates an explicit dual graph Laplacian prior and an implicit deep prior into a unified deep learning framework. This hybrid design effectively combines handcrafted structural regularization with data-driven feature learning, leading to improved interpretability and reconstruction accuracy.
- We design an efficient alternating optimization algorithm to iteratively solve the underlying variational problem, which jointly exploits the dual graph priors and deep implicit guidance within a unified framework. As shown in Fig. 1, our method achieves state-of-the-art performance with notably lower computational complexity.

2. Graph Laplacian Model

The most distinctive characteristic of depth maps is their piecewise smooth (PWS) structure, in which smoothly varying regions are separated by sharp discontinuities. Graph-based techniques are particularly effective for modeling this property, as the graph Laplacian operator enforces intra-region smoothness while preserving inter-region discontinuities [36]. Consequently, such methods have been widely adopted in depth super-resolution (SR). In the following, we first introduce the main forms of the graph Laplacian and then discuss their applications in depth SR.

In graph-based signal modeling, a depth map can be viewed as a discrete signal $\boldsymbol{x} \in \mathbb{R}^N$ defined on a weighted undirected graph $\boldsymbol{\mathcal{G}} = (\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}}, \boldsymbol{W})$, where the vertices $\boldsymbol{\mathcal{V}}$ correspond to image pixels and the edges $\boldsymbol{\mathcal{E}}$ represent pairwise pixel connections. The affinity matrix $\boldsymbol{W}$ encodes pairwise pixel similarities, with $w_{i,j} = w_{j,i} \geq 0$. The graph Laplacian $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$ (where $\boldsymbol{D}$ is the diagonal matrix with $\boldsymbol{D}_{ii} = \sum_j w_{ij}$) provides a smoothness measure for $\boldsymbol{x}$:

$$S(\boldsymbol{x}) = \frac{1}{2}\sum_{i,j} w_{i,j}(\boldsymbol{x}_i - \boldsymbol{x}_j)^2 = \boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}. \tag{1}$$

This formulation penalizes large variations between adjacent nodes and encourages piecewise smoothness. With this quantitative smoothness measure, we can incorporate the piecewise-smooth prior into a standard inverse problem:

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|_2^2 + \gamma \boldsymbol{x}^\top \boldsymbol{L}\boldsymbol{x}, \tag{2}$$

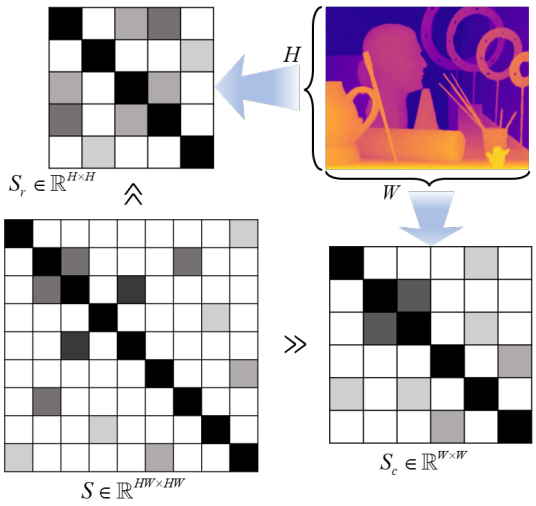


Figure 2. Affinity graphs $\boldsymbol{S}_r$ and $\boldsymbol{S}_c$ are constructed along the row and column dimensions of $\boldsymbol{Y}$, respectively. Unlike conventional full affinity matrix $\boldsymbol{S}$, which scales quadratically with image resolution, these graphs are much smaller, with sizes that grow linearly.

where $\boldsymbol{y}$ is the observed low-resolution (LR) depth, $\boldsymbol{H}$ is the degradation matrix, and γ is a regularization parameter.

Building upon the above formulation, researchers have explored various graph Laplacian designs and optimization strategies for depth SR. For example, Jiang *et al.* [37] introduce a dual-domain regularization method that leverages a Laplacian-based transform-domain prior and a spatial consistency constraint from color-depth pairs to recover sharp edges and structural details. Liu *et al.* [21] introduce a joint internal-external graph regularization framework, where the internal graph, constructed from the depth map based on spatial similarity, promotes prior propagation, while the external graph, derived from the guidance image, enhances cross-modal structural consistency. Wang *et al.* [22] present a dual regularization approach that enforces structural smoothness via graph Laplacian and maintains geometric fidelity through normal-depth consistency.

3. Method

3.1. Motivation

Graph-based regularization has been widely applied in guided depth super-resolution (GDSR) due to its ability to preserve the piecewise-smooth nature of depth maps [21, 34, 35, 37]. There are two main approaches for graph construction: fully connected graphs and sparse graphs.

In the fully connected graph method, pairwise similarities are computed across all pixel pairs, resulting in dense Laplacian matrices of size $HW \times HW$, where H and W denote the height and width of the image. While this method captures global dependencies, it leads to a high computational cost due to the large number of connections, with a complexity of $\mathcal{O}(H^3W^3)$. In contrast, sparse graphs connect each pixel to only a small neighborhood, reducing memory usage and computational complexity. However, this sparsity introduces two main drawbacks. First, the fixed local connections limit the receptive field, making it difficult to capture long-range dependencies and global structure. Second, the predefined adjacency pattern restricts the model to fixed input sizes, limiting its scalability to images with varying resolutions. Additionally, many of these methods flatten 2D images into 1D vectors when constructing the Laplacian, which discards the natural row-column spatial structure and may introduce potential artifacts.

To overcome these limitations, we propose a dual graph regularization framework that independently models structural dependencies along the horizontal and vertical directions (as shown in Fig. 2). **This design naturally preserves the two-dimensional topology of images, supports arbitrary input resolutions, and substantially reduces computational complexity to $\mathcal{O}(H^3 + W^3)$ by avoiding the explicit construction of a global pixel-wise graph.**

3.2. Problem Formulation

We define the low-resolution (LR) depth map and the high-resolution (HR) color image as $\boldsymbol{Y} \in \mathbb{R}^{h\times w}$ and $\boldsymbol{G} \in \mathbb{R}^{H\times W\times C}$, respectively, where C means the number of channels, and $H = h\cdot s, W = w\cdot s$ correspond to the height and width of the HR image with an upscaling factor s. To match the resolution of $\boldsymbol{Y}$ and $\boldsymbol{G}$, we first upsample $\boldsymbol{Y}$ via bicubic interpolation to obtain $\hat{\boldsymbol{Y}} \in \mathbb{R}^{H\times W}$. Our objective is to reconstruct the HR depth map $\boldsymbol{X} \in \mathbb{R}^{H\times W}$ by jointly optimizing three components within a unified framework.

1) Data fidelity term: We begin by formulating the data fidelity term to model the degradation relationship between the reconstruction result $\boldsymbol{X}$ and the observed LR input $\boldsymbol{Y}$:

$$\min_{\boldsymbol{X}} \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}\|_F^2, \tag{3}$$

where $\boldsymbol{D}$ is the degradation operator that simulates the downsampling process from HR to its LR counterpart.

2) Dual graph Laplacian regularization: Since the GDSR model in Eq. (3) lacks the exploration of the intrinsic prior of the target depth image, such as piece-wise smooth structures, we develop the dual graph Laplacian prior as shown in the following ***Definition 1*** and ***Lemma 1***.

Definition 1 (Dual graph Laplacian prior): *Unlike the traditional graph optimization strategies that focus on learning the similarity relationships of each pixel, the proposed dual graph Laplacian prior is designed to capture the row-wise and column-wise similarities simultaneously. Let $\boldsymbol{S}_r \in \mathbb{R}^{H\times H}$ and $\boldsymbol{S}_c \in \mathbb{R}^{W\times W}$ be the row-wise and column-wise affinity graphs, respectively. As shown in Fig. 2, these matrices are constructed based on the row and column subspaces of the input image, which are much smaller than the traditional affinity matrix. Consequently, the target*

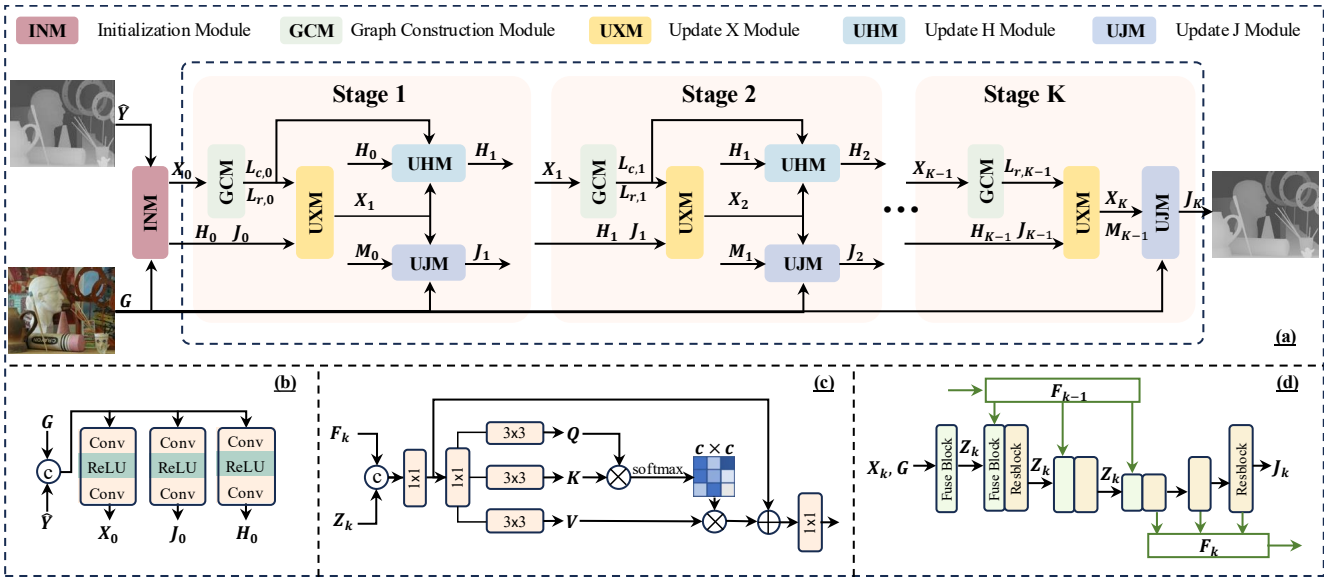


Figure 3. Detailed architecture of the proposed LapNet for guided depth map super-resolution. (a) Overall iterative unfolding architecture, which alternates between graph construction and three update modules across K stages; (b) Initialization Module (INM) that jointly encodes $\hat{\boldsymbol{Y}}$ and $\boldsymbol{G}$ to produce initial estimates of $\boldsymbol{X}_0$, $\boldsymbol{H}_0$ and $\boldsymbol{J}_0$. (c) Fuse block; (d) Proximal network.

image $\boldsymbol{X}$ can be optimized through the following model,

$$\min_{\boldsymbol{X}} \frac{1}{2}\sum_{i=1}^{H}\sum_{j=1}^{H}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_2^2(\boldsymbol{S}_r)_{ij}+ \frac{1}{2}\sum_{i=1}^{W}\sum_{j=1}^{W}\|(\boldsymbol{X}^\top)_i-(\boldsymbol{X}^\top)_j\|_2^2(\boldsymbol{S}_c)_{ij}. \tag{4}$$

Based on ***Definition 1***, we could derive the ***Lemma 1***.

Lemma 1: *By introducing the Laplacian matrices $\boldsymbol{L}_r \in \mathbb{R}^{H\times H}$ and $\boldsymbol{L}_c \in \mathbb{R}^{W\times W}$ in terms with $\boldsymbol{S}_r$ and $\boldsymbol{S}_c$, respectively, Eq. (4) is equivalent to the following expression,*

$$\min_{\boldsymbol{X}} tr(\boldsymbol{X}^\top\boldsymbol{L}_r\boldsymbol{X})+tr(\boldsymbol{X}\boldsymbol{L}_c\boldsymbol{X}^\top). \tag{5}$$

where the two Laplacian matrices are defined as:

$$\boldsymbol{L}_r=\boldsymbol{U}_r-\boldsymbol{S}_r, \tag{6}$$

$$\boldsymbol{L}_c=\boldsymbol{U}_c-\boldsymbol{S}_c, \tag{7}$$

with $\boldsymbol{U}_r$ and $\boldsymbol{U}_c$ denoting the corresponding degree matrices. The detailed proof of Lemma 1 is provided in the **Supplementary Material**. Based on ***Definition 1*** and ***Lemma 1***, our GDSR model in Eq. (3) can be improved as

$$\min_{\boldsymbol{X}}\|\boldsymbol{Y}-\boldsymbol{D}\boldsymbol{X}\|_F^2+\lambda tr(\boldsymbol{X}^\top\boldsymbol{L}_r\boldsymbol{X})+\alpha tr(\boldsymbol{X}\boldsymbol{L}_c\boldsymbol{X}^\top), \tag{8}$$

where λ and α denote two balance parameters.

3) Deep implicit prior: To compensate for the limitations of handcrafted priors, we introduce a deep prior term based on a learnable network $f(\cdot)$ that exploits high-frequency structural cues from the guidance image $\boldsymbol{G}$. This term helps refine the reconstruction result $\boldsymbol{X}$ by injecting complementary information from $\boldsymbol{G}$:

$$\min_{\boldsymbol{X}}\beta\cdot f(\boldsymbol{X},\boldsymbol{G}), \tag{9}$$

where β controls the influence of the learned prior. Combining all components, the overall objective becomes:

$$\min_{\boldsymbol{X}}\|\boldsymbol{Y}-\boldsymbol{D}\boldsymbol{X}\|_F^2+\lambda tr(\boldsymbol{X}^\top\boldsymbol{L}_r\boldsymbol{X})+ \alpha tr(\boldsymbol{X}\boldsymbol{L}_c\boldsymbol{X}^\top)+\beta f(\boldsymbol{X},\boldsymbol{G}). \tag{10}$$

This formulation integrates explicit graph-based structural regularization with data-driven priors, facilitating efficient and interpretable high-resolution depth reconstruction.

3.3. Optimization

To solve the optimization problem outlined in Eq. (10), we propose an efficient algorithm leveraging the Alternating Direction Method of Multipliers (ADMM) [38]. Specifically, we first introduce two additional variables $\boldsymbol{J}$ and $\boldsymbol{H}$, with constraints $\boldsymbol{X}=\boldsymbol{J}$ and $\boldsymbol{X}=\boldsymbol{H}$, and reformulate the problem using augmented Lagrangian terms. The augmented Lagrangian function is defined as:

$$\Phi_\mu=\|\boldsymbol{Y}-\boldsymbol{D}\boldsymbol{X}\|_F^2+\lambda tr(\boldsymbol{X}^\top\boldsymbol{L}_r\boldsymbol{X})+\alpha tr(\boldsymbol{H}\boldsymbol{L}_c\boldsymbol{H}^\top)+ \beta f(\boldsymbol{J},\boldsymbol{G})+tr(\boldsymbol{M}^\top(\boldsymbol{X}-\boldsymbol{J}))+\frac{\mu}{2}\|\boldsymbol{X}-\boldsymbol{J}\|_F^2+ tr(\boldsymbol{N}^\top(\boldsymbol{X}-\boldsymbol{H}))+\frac{\mu}{2}\|\boldsymbol{X}-\boldsymbol{H}\|_F^2, \tag{11}$$

where $\boldsymbol{M}$ and $\boldsymbol{N}$ are two Lagrange multipliers. μ denotes a penalty parameter and satisfies $\mu>0$. Then, by simple

Algorithm 1: ADMM Optimization for LapNet

Input: LR depth map $\boldsymbol{Y}$, HR guidance image $\boldsymbol{G}$;
Parameters $\lambda, \alpha, \beta, \mu$ Initial values:
$\boldsymbol{X}_0, \boldsymbol{J}_0, \boldsymbol{H}_0, \boldsymbol{M}_0, \boldsymbol{N}_0$
Output: Reconstructed HR depth map $\boldsymbol{X}^*$
for $t = 0$ **to** $K-1$ **do**
  Step 1: Update $\boldsymbol{X}_{t+1}$;
  Construct $\boldsymbol{S}_{r,t}, \boldsymbol{S}_{c,t}$ from $\boldsymbol{X}_t$ (Section 3.2);
  Compute $\boldsymbol{L}_{r,t}, \boldsymbol{L}_{c,t}$ via Eqs. 6 - 7;
  Solve $\boldsymbol{X}_{t+1}$ using closed-form Eq. 14;
  Step 2: Update $\boldsymbol{H}_{t+1}$;
  Solve column-wise regularization via Eq. 16;
  Step 3: Update $\boldsymbol{J}_{t+1}$;
  Apply proximal network: $\boldsymbol{J}_{t+1} = \text{Prox}(\cdot)$ (Eq. 18);
  Step 4: Update Multipliers;
  $\boldsymbol{M}_{t+1}, \boldsymbol{N}_{t+1}$ updated via Eq. 19;
end
return $\boldsymbol{X}^* = \boldsymbol{J}_K$;

algebraic operations, we could improve Eq. (11) as

$$\Phi_\mu = \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}\|_F^2 + \lambda tr(\boldsymbol{X}^\top \boldsymbol{L}_r \boldsymbol{X}) + \alpha tr(\boldsymbol{H}\boldsymbol{L}_c\boldsymbol{H}^\top) + \beta f(\boldsymbol{J}, \boldsymbol{G}) + \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{J} + \frac{\boldsymbol{M}}{\mu}\|_F^2 + \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{H} + \frac{\boldsymbol{N}}{\mu}\|_F^2, \tag{12}$$

The optimization proceeds by alternatingly updating variables $\boldsymbol{X}$, $\boldsymbol{J}$, and $\boldsymbol{H}$. For the $(t+1)$-th iteration, the detailed optimization process is listed as follows:

Step 1: The subproblem for $\boldsymbol{X}$ combines quadratic terms from data fidelity, graph regularization, and equality constraints:

$$\boldsymbol{X}_{t+1} = \arg\min_{\boldsymbol{X}} \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X}\|_F^2 + \lambda tr(\boldsymbol{X}^\top \boldsymbol{L}_{r,t} \boldsymbol{X}) + \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{J}_t + \frac{\boldsymbol{M}_t}{\mu}\|_F^2 + \frac{\mu}{2}\|\boldsymbol{X} - \boldsymbol{H}_t + \frac{\boldsymbol{N}_t}{\mu}\|_F^2. \tag{13}$$

By solving the quadratic problem, we could get the closed-form solution of $\boldsymbol{X}$ as

$$\boldsymbol{X}_{t+1} = (\boldsymbol{D}^\top \boldsymbol{D} + \lambda \boldsymbol{L}_{r,t} + \mu \boldsymbol{I}_r)^{-1} (\boldsymbol{D}^\top \boldsymbol{Y} + \frac{\mu_t}{2}(\boldsymbol{J}_t + \boldsymbol{H}_t - \frac{\boldsymbol{M}_t + \boldsymbol{N}_t}{\mu_t})), \tag{14}$$

where $\boldsymbol{I}_r \in \mathbb{R}^{H \times H}$ denotes an identity matrix.

Step 2: We could solve $\boldsymbol{H}$ by the following subproblem,

$$\boldsymbol{H}_{t+1} = \arg\min_{\boldsymbol{H}} \alpha tr(\boldsymbol{H}\boldsymbol{L}_{c,t+1}\boldsymbol{H}^\top) + \frac{\mu}{2}\|\boldsymbol{X}_{t+1} - \boldsymbol{H} + \frac{\boldsymbol{N}_t}{\mu}\|_F^2. \tag{15}$$

Then, we could obtain the closed-form solution of $\boldsymbol{H}$ as

$$\boldsymbol{H}_{t+1} = (\mu \boldsymbol{X}_{t+1} + \boldsymbol{N}_t)(\mu \boldsymbol{I}_c + 2\alpha \boldsymbol{L}_{c,t+1})^{-1}, \tag{16}$$

where $\boldsymbol{I}_c \in \mathbb{R}^{W \times W}$ denotes an identity matrix.

Step 3: We could solve $\boldsymbol{J}$ by the following subproblem,

$$\boldsymbol{J}_{t+1} = \arg\min_{\boldsymbol{J}} \frac{\mu}{2}\|\boldsymbol{X}_{t+1} - \boldsymbol{J} + \frac{\boldsymbol{M}_t}{\mu}\|_F^2 + \beta f(\boldsymbol{J}, \boldsymbol{G}). \tag{17}$$

Then, we could obtain the closed-form solution of $\boldsymbol{J}$ as

$$\boldsymbol{J}_{t+1} = Prox(\boldsymbol{X}_{t+1} + \frac{\boldsymbol{M}_t}{\mu}, \boldsymbol{G}). \tag{18}$$

Step 4: We could update the Lagrange multipliers $\boldsymbol{M}$ and $\boldsymbol{N}$ by the following expressions:

$$\begin{aligned} \boldsymbol{M}_{t+1} &= \boldsymbol{M}_t + \mu(\boldsymbol{X}_{t+1} - \boldsymbol{J}_{t+1}), \\ \boldsymbol{N}_{t+1} &= \boldsymbol{N}_t + \mu(\boldsymbol{X}_{t+1} - \boldsymbol{H}_{t+1}). \end{aligned} \tag{19}$$

3.4. Deep Unfolding Network

To combine the interpretability of iterative optimization with the efficiency of deep learning, we propose a deep unfolding network that explicitly unrolls the optimization steps into structured network stages. As shown in Fig. 3, the network consists of five key modules:

1) The Initialization Module (INM): The INM (Fig. 3 (b)) generates initial estimates of $\boldsymbol{X}_0$, $\boldsymbol{J}_0$ and $\boldsymbol{H}_0$ by fusing the upsampled LR depth map $\hat{\boldsymbol{Y}}$ and the guidance image $\boldsymbol{G}$.

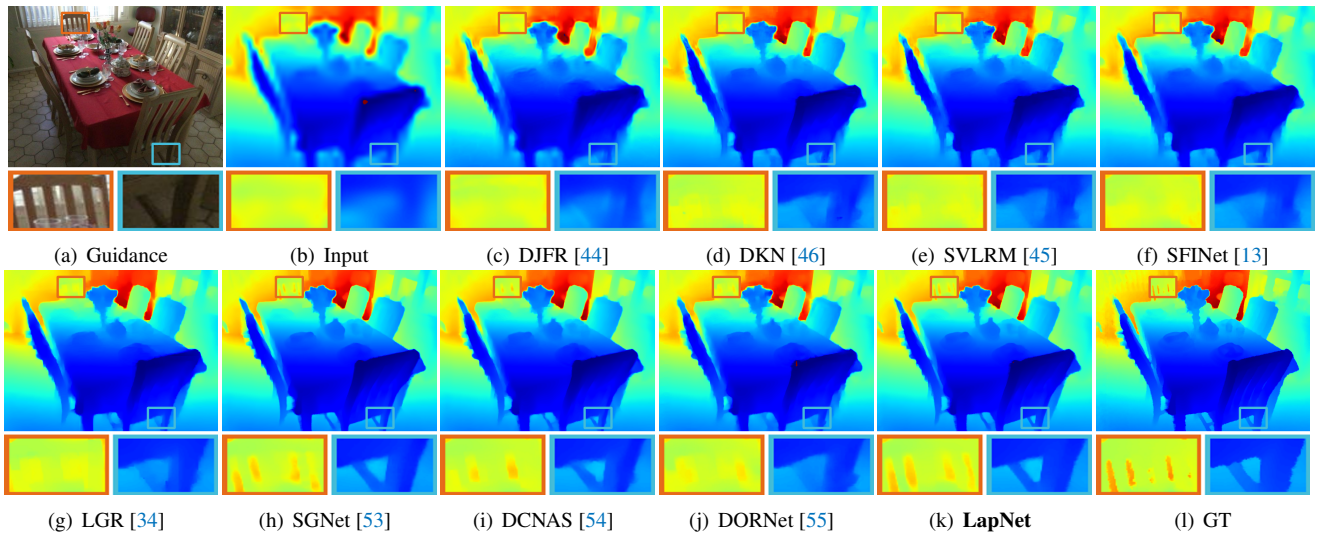
2) The Graph Construction Module (GCM): The GCM constructs the row and column Laplacian matrices $\boldsymbol{L}_r$ and $\boldsymbol{L}_c$ from the current estimate $\boldsymbol{X}_t$. It first computes row-wise and column-wise similarities to form the affinity matrices $\boldsymbol{S}_r$ and $\boldsymbol{S}_c$. Two similarity functions are considered: i) L_2-based similarity: $\boldsymbol{S}_{i,j} = \exp\left(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||_2^2}{\sigma^2}\right)$, where σ is adaptively learned during training; and ii) dot-product similarity: $\boldsymbol{S}_{i,j} = \frac{\boldsymbol{x}_i^\top \boldsymbol{x}_j}{||\boldsymbol{x}_i|| \cdot ||\boldsymbol{x}_j||}$. Here, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ denote the i-th and j-th rows (or columns) of $\boldsymbol{X}_t$, respectively. The Laplacian matrices are then computed from the affinity graphs according to Eq. (6) and Eq. (7).

3) The Update $\boldsymbol{X}$ Module (UXM) and the Update $\boldsymbol{H}$ Module (UHM): The UXM and UHM correspond to the optimization steps for updating the primary variable $\boldsymbol{X}$ and the auxiliary variable $\boldsymbol{H}$, respectively. The purpose of these modules is to refine the current depth estimate by incorporating data fidelity and structural regularization. They are implemented based on Eq. (14) and Eq. (16).

4) The Update J Module (UJM): This module corresponds to the optimization step for updating the auxiliary variable $\boldsymbol{J}$, which encodes the deep implicit prior extracted from input depth and guidance images. The purpose of UJM is to incorporate high-frequency structural information from the guidance image into the current depth reconstruction. As illustrated in Fig. 3 (d), UJM is implemented using a lightweight U-shaped network. Specifically, the input features are obtained by first fusing the reconstructed depth $\boldsymbol{X}_t$ and the guidance image $\boldsymbol{G}$ through a fusion block

Table 1. RMSE comparison of various guided depth map super-resolution (GDSR) methods on benchmark datasets. The top two results are highlighted in **first** and second, respectively.

Method	NYU v2 [10]			Sintel [39]			DIDOE [40]			SUN RGB [41]			RGB-D-D [42]			DIML [43]			Average		
	4×	8×	16×	4×	8×	16×	4×	8×	16×	4×	8×	16×	4×	8×	16×	4×	8×	16×	4×	8×	16×
DJFR [44]	2.38	4.94	9.18	4.90	7.39	10.33	5.63	8.24	9.89	0.81	1.54	2.80	1.50	2.72	5.05	1.27	2.34	4.13	2.75	4.53	6.90
SVLRM [45]	1.51	3.21	6.98	4.05	5.83	8.60	3.58	6.96	9.55	0.59	1.10	2.33	1.22	1.88	3.55	1.19	1.93	3.49	2.02	3.48	5.75
DKN [46]	1.62	3.26	6.51	4.38	5.89	8.40	3.49	6.96	9.31	0.63	1.10	2.16	1.31	1.87	3.26	1.27	1.86	3.22	2.12	3.49	5.48
JIIF [47]	1.37	2.76	5.27	3.82	5.50	7.46	2.94	6.17	8.58	0.54	0.95	1.79	1.15	1.77	2.79	1.17	1.79	2.86	1.83	3.16	4.79
FDSR [42]	1.61	3.18	5.86	4.14	5.67	7.86	3.62	6.54	8.85	0.64	1.05	1.97	1.16	1.82	3.06	1.10	1.71	2.87	2.05	3.33	5.08
CODON [48]	1.40	2.77	5.92	3.76	5.37	7.86	3.03	6.17	9.02	0.53	0.99	2.09	1.12	1.79	3.05	1.18	1.85	3.06	1.82	3.17	5.17
AHMF [49]	1.40	2.89	5.64	3.84	5.62	7.55	2.93	6.14	8.54	0.57	0.99	1.82	1.10	1.73	3.04	1.10	1.70	2.83	1.82	3.18	4.90
LGR [34]	1.79	3.04	6.02	4.29	5.56	7.93	3.93	6.37	9.06	0.67	1.05	2.03	1.30	1.83	3.12	1.25	1.79	3.03	2.20	3.27	5.20
DADA [50]	1.54	2.74	4.80	4.21	5.46	7.12	3.68	6.23	8.43	0.60	0.96	1.70	1.20	1.83	2.80	1.15	1.71	2.65	2.06	3.16	4.58
MMNet [51]	1.50	3.03	6.21	4.19	5.59	8.02	3.53	6.39	8.97	0.60	1.03	2.04	1.13	1.72	2.91	1.14	1.73	3.05	2.02	3.25	5.20
DCTNet [12]	1.59	3.08	5.80	4.18	6.31	9.16	3.84	7.20	9.69	0.65	1.28	2.43	1.08	1.74	3.05	1.07	1.71	2.99	2.07	3.55	5.52
SSDNet [52]	1.60	3.14	5.86	3.84	5.72	8.31	3.04	6.54	9.46	0.62	1.11	2.15	**1.04**	1.72	2.92	1.15	1.83	3.21	1.88	3.34	5.32
SFINet [13]	1.52	3.04	6.21	3.91	5.66	8.78	3.51	6.47	9.50	0.56	1.07	2.34	1.16	1.87	3.47	1.21	1.83	3.60	1.98	3.32	5.65
SGNet [53]	1.10	2.44	4.77	3.67	5.12	7.11	2.78	5.64	8.31	**0.51**	0.92	1.68	1.10	1.64	2.55	1.06	1.75	2.71	1.70	2.92	4.52
DCNAS [54]	1.21	2.55	5.06	3.68	5.24	7.19	2.82	5.88	8.36	**0.51**	**0.91**	1.70	1.06	1.70	2.61	1.04	1.64	2.65	1.72	2.99	4.60
DORNet [55]	1.19	2.70	5.60	3.88	5.60	8.26	3.21	6.43	9.18	0.54	1.04	2.15	1.19	1.91	3.35	1.20	1.88	3.21	1.87	3.26	5.29
LapNet	**1.05**	**2.33**	**4.55**	**3.57**	**5.05**	**7.02**	**2.63**	**5.51**	**8.24**	**0.51**	0.92	**1.58**	1.05	**1.56**	**2.45**	**1.02**	**1.52**	**2.52**	**1.64**	**2.82**	**4.39**



(a) Guidance (b) Input (c) DJFR [44] (d) DKN [46] (e) SVLRM [45] (f) SFINet [13]

(g) LGR [34] (h) SGNet [53] (i) DCNAS [54] (j) DORNet [55] (k) **LapNet** (l) GT

Figure 4. Visual comparison for 16× depth map super-resolution on NYU v2 dataset [10]. Best viewed by zooming.

(Fig. 3 (c)) followed by a proximal network that refines the fused representation and produces the updated $\boldsymbol{J}$. Moreover, to reduce information loss across iterations [51], we introduce a fine-grained feature propagation strategy, where the reconstructed image $\boldsymbol{X}_k$ and decoder features from the previous stage ($\boldsymbol{F}_{k-1}$) are jointly passed to the encoder of the current stage as input (Fig. 3 (d)).

4. Experiments

4.1. Dataset and Implementation Details

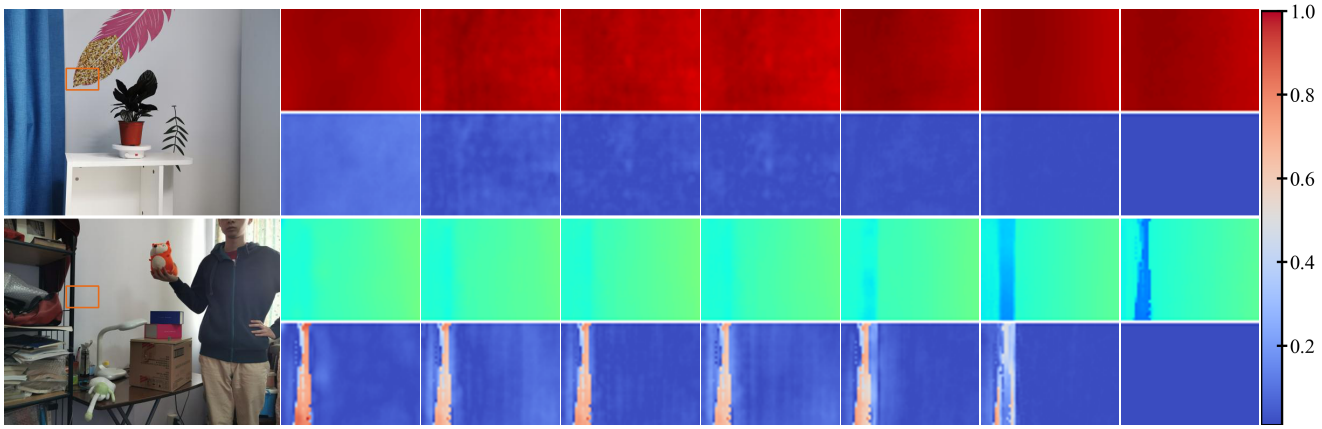
In this section, we conduct experiments to evaluate the performance of the proposed method. Following [11, 52, 54], two widely used benchmark datasets are used: NYU v2 [10] and RGB-D-D [42]. The RMSE is adopted as the evaluation metric. Our framework is implemented in PyTorch and trained using two NVIDIA RTX 5090 GPUs. We use the $\mathcal{L}_1$ loss for supervision. The number of unfolding iterations is set to three, and the proximal network contains 24 feature channels. More details about the datasets and training configurations are provided in the **Supplementary Material**.

4.2. Comparison with the State-of-the-arts

Experimental results on NYU dataset [10]. The quantitative results are summarized in Table 1. The proposed LapNet almost achieves the best performance across all datasets in terms of RMSE, demonstrating the effectiveness of the proposed method. Although the second-best method,

Table 2. RMSE comparison with the different methods on RGB-D-D [42] dataset for real-world depth map super-resolution.

Method	DKN [46]	FDSR [42]	DADA [50]	DCTNet [12]	SSDNet [52]	SFINet [13]	SGNet [53]	DCNAS [54]	SFG [56]	DORNet [55]	**LapNet**
RMSE↓	5.74	5.49	5.48	5.38	5.40	5.38	5.32	4.87	3.88	3.42	**3.25**



(a) RGB (b) Input (c) FDSR [42] (d) SGNet [53] (e) DCNAS [54] (f) DORNet [55] (g) **LapNet** (h) GT

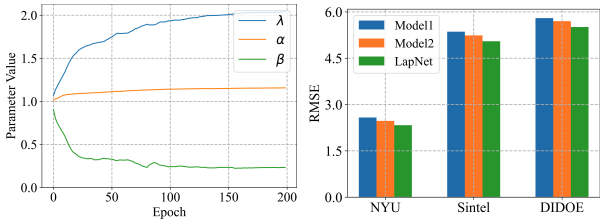
Figure 5. Visual comparisons on the RGB-D-D [42] dataset for real-world depth map super-resolution. For clearer illustration, a representative region is highlighted and enlarged, with its corresponding error map also provided. Best viewed with zoom.

SGNet [53], produces comparable results, LapNet is much more lightweight, with only 3.84M parameters compared to 25.33M in SGNet. This remarkable reduction in model size clearly shows the efficiency and compactness of our design. Furthermore, Fig. 4 presents the qualitative comparison of different methods for the 16× upscaling task. As observed, LapNet generates sharper edges and more distinct structural details, particularly in fine-grained regions such as the chair backrest (orange box) and the table leg (blue box). In contrast, competing approaches often yield over-smoothed outputs or structural distortions in these challenging areas.

Experimental results on RGB-D-D dataset [42]. In addition to the synthetic dataset, we also conduct experiments on the real-world RGB-D-D dataset [42]. The quantitative and qualitative comparisons are presented in Table 2 and Fig. 5, respectively. On this dataset, our LapNet achieves the best performance with the lowest RMSE. In the first row of Fig. 5, all competing methods exhibit noticeable texture transfer artifacts and fail to maintain consistent depth across smooth regions. In contrast, LapNet produces more uniform and coherent depth maps, effectively suppressing texture copying from the RGB guidance. In the second row, our method accurately reconstructs thin structures such as the bookshelf support, recovering sharp edges and fine geometric details, while other methods yield blurred or missing components. These results confirm that LapNet provides superior structural fidelity and visual consistency in both smooth and detailed regions.

4.3. Ablation Study

To assess the contribution of each component in LapNet, we perform four sets of ablation experiments on the NYU v2 dataset [10] for 8× guided depth map super-resolution.



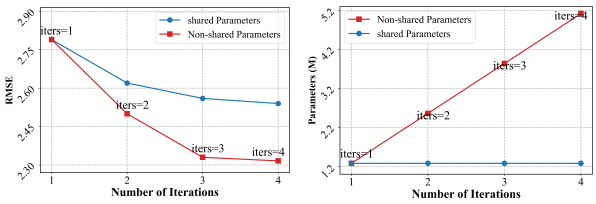
(a) (b)

Figure 6. **Ablation study**. (a) Evolution of the penalty parameters α, λ and β during training. (b) RMSE comparison of Model1, Model2, and LapNet on NYU v2 [10], Sintel [57], and DIDOE [40] datasets for 8× guided depth super-resolution.

1) **Effect of Initialization Strategy.** We first examine the impact of the learnable initialization strategy. The initial variables $\boldsymbol{X}_0, \boldsymbol{J}_0, \boldsymbol{H}_0$ are generated by an initialization module, while the penalty parameters α, λ, β are all initialized to 1 and updated through backpropagation. To assess their effectiveness, we design three variants:

- `Model1`: The initial variables $\boldsymbol{X}_0$, $\boldsymbol{J}_0$, and $\boldsymbol{H}_0$ are all set to the bicubic-upsampled depth map, and the penalty parameters α, λ, and β are fixed at 1 throughout training;
- `Model2`: The initial variables are learned by the initialization module, while the penalty parameters are fixed;
- `LapNet (Ours)`: Both the initialization variables and penalty parameters are learned jointly during training.

As shown in Fig. 6(b), LapNet consistently achieves the lowest RMSE across all datasets, confirming the effectiveness of jointly learning both initialization and penalty parameters. Compared with `Model1`, `Model2` shows clear improvements, indicating that a learnable initialization of the latent variables provides a more reliable starting point for iterative optimization. Moreover, Fig. 6(a) illustrates the evolution of the penalty parameters during training. All



(a) (b)

Figure 7. **Ablation study**. Effect of iteration number on performance and model complexity. (a) RMSE comparison on NYU v2 dataset [10] for 8× GDSR under shared and non-shared parameters. (b) Model size (M) for different iteration numbers.

Table 3. **Ablation Study.** Evaluation of various graph construction strategies for 8× GDSR in terms of RMSE.

Method	NYU v2 [10]	Sintel [57]	DIDOE [40]
`Model3`	2.52	5.26	5.78
`Model4`	2.60	5.24	5.73
LapNet	**2.33**	**5.05**	**5.51**

three parameters gradually converge to stable values: β decreases to control the influence of the proximal term, while λ and α increase to enhance graph regularization. This adaptive behavior allows LapNet to better balance fidelity and structure across different training stages.

2) **Effect of Iteration Number.** We analyze the impact of the number of unfolding iterations on model performance. The iteration count varies, and both RMSE and parameter size are recorded (see Fig. 7). The results show that performance steadily improves as the number of iterations increases from 1 to 3. When the number of iterations exceeds 3, the RMSE continues to decrease but only marginally. Considering the trade-off between performance gain and computational complexity, we set the number of iterations to 3 in our final model. In deep unfolding frameworks, two strategies are commonly used to handle parameters across iterations: parameter sharing, where the same parameters are reused at each step, and non-sharing, where each iteration has its own parameters. As shown in Fig. 7(a), the non-shared setting consistently yields lower RMSE values, particularly as the iteration number increases. This suggests that separate parameters enable each stage to better capture distinct transformations and refine predictions. However, as illustrated in Fig. 7(b), this design increases the total number of parameters. Despite this cost, LapNet adopts the non-shared strategy for its superior performance.

3) **Effect of Graph Construction Strategy.** We further compare different strategies for graph construction, focusing on two key aspects:

- **Affinity metric:** We evaluate the effect of computing row/column affinity matrices using either dot-product similarity (`Model3`) or L_2 distance (Ours);
- **Graph update frequency:** We compare a static graph (`Model4`), which is constructed only once from the initial estimate $\boldsymbol{X}_0$, with a dynamic graph, which is updated at each iteration using the current estimate $\boldsymbol{X}_t$.

Table 4. **Ablation Study.** Evaluation of different prior modeling strategies for 8× GDSR in terms of RMSE.

Method	NYU v2 [10]	Sintel [57]	DIDOE [40]
`Model5`	2.63	5.31	5.87
`Model6`	2.57	5.28	5.81
`Model7`	3.01	5.87	6.34
LapNet	**2.33**	**5.05**	**5.51**

As shown in Table 3, our method outperforms Model3, demonstrating that the L_2-based affinity metric provides more reliable and scale-invariant similarity estimation than dot-product similarity. Moreover, the dynamic graph strategy consistently achieves lower RMSE than the static one, as iterative updates allow the graph to adapt to progressively refined depth predictions. Therefore, we adopt the L_2-based dynamic graph construction as the default setting in LapNet.

3) Effect of Dual-Prior Modeling. To evaluate the dual-prior design, we conduct an ablation study by removing specific components from LapNet:

- `Model5`: Removes the row-directional graph while retaining the column graph and deep prior.
- `Model6`: Removes the column-directional graph while retaining the row graph and deep prior.

In addition, to examine the role of the deep implicit prior, we replace the learned proximal network with a fixed guided filter [58], forming `Model7`. As shown in Table 4, removing either directional graph results in a noticeable performance drop, confirming that the row-wise and column-wise priors capture complementary structural cues essential for accurate reconstruction. In addition, `Model7` exhibits a significant degradation, as replacing the learnable prior with a fixed filter reduces adaptability and representation capacity required to handle diverse scene structures. Overall, these results validate the effectiveness of our dual-prior design, which integrates both explicit graph regularization and deep implicit guidance in a unified framework.

5. Conclusion

This paper presents LapNet, a novel guided depth map super-resolution framework that incorporates dual graph Laplacian priors with deep learning. This design effectively preserves the two-dimensional structure of the depth maps, supports arbitrary input resolutions, and significantly reduces computational complexity to $\mathcal{O}(H^3+W^3)$ by eliminating the need for explicit global pixel-wise graph construction. Extensive experiments on both synthetic and real-world datasets demonstrate that LapNet achieves state-of-the-art performance while maintaining high computational efficiency. Future work will explore its scalability and potential for integrating other graph optimization techniques.

Acknowledgment

This work was partially supported in part by the Hong Kong Innovation and Technology Commission (ITC) under Grants GHP/044/21SZ and PRP/036/24FX, in part by the General Research Fund (GRF) of the Research Grants Council (RGC) of Hong Kong under Grant 11200323, and in part by the National Natural Science Foundation of China (NSFC)/Research Grants Council (RGC) Joint Research Scheme under Grant N_CityU198/24.

References

[1] R. Girshick, J. Shotton, P. Kohli, A. Criminisi, and A. Fitzgibbon, "Efficient regression of general-activity human poses from depth images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 415–422, IEEE, 2011. 1

[2] R. Liu, G. Zhang, J. Wang, and S. Zhao, "Cross-modal 360° depth completion and reconstruction for large-scale indoor environment," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[3] H. Adams, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer, "Depth perception in augmented reality: The effects of display, shadow, and position," in *VR*, pp. 792–801, IEEE, 2022.

[4] O. Natan and J. Miura, "End-to-end autonomous driving with semantic depth cloud mapping and multi-agent," *IEEE Transactions on Intelligent Vehicles*, 2022. 1

[5] P. L. Rosin, Y.-K. Lai, L. Shao, and Y. Liu, *RGB-D image analysis and processing*. Springer, 2019. 1

[6] G. Kurillo, E. Hemingway, M.-L. Cheng, and L. Cheng, "Evaluating the accuracy of the azure kinect and kinect v2," *Sensors*, vol. 22, no. 7, p. 2469, 2022. 1

[7] J. Choe, S. Im, F. Rameau, M. Kang, and I. S. Kweon, "Volumefusion: Deep depth fusion for 3d scene reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16086–16095, 2021. 1

[8] M. Schön, M. Buchholz, and K. Dietmayer, "Mgnet: Monocular geometric scene understanding for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 15804–15815, 2021.

[9] J. Xiang, X. Zhu, X. Wang, Y. Wang, H. Zhang, F. Guo, and X. Yang, "Depthor: Depth enhancement from a practical light-weight dtof sensor and rgb image," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6101–6111, October 2025. 1

[10] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *Proceedings of the European Conference on Computer Vision*, pp. 746–760, 2012. 1, 6, 7, 8

[11] Z. Zhong, X. Liu, J. Jiang, D. Zhao, and X. Ji, "Guided depth map super-resolution: A survey," *ACM Computing Surveys*, vol. 55, no. 14, pp. 1–36, 2023. 1, 6

[12] Z. Zhao, J. Zhang, S. Xu, Z. Lin, and H. Pfister, "Discrete cosine transform network for guided depth map super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5697–5707, 2022. 6, 7

[13] M. Zhou, J. Huang, K. Yan, D. Hong, X. Jia, J. Chanussot, and C. Li, "A general spatial-frequency learning framework for multimodal image fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 6, 7

[14] S. Gu, S. Guo, W. Zuo, Y. Chen, R. Timofte, L. V. Gool, and L. Zhang, "Learned dynamic guidance for depth image reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2437–2452, 2019.

[15] Y. Zuo, Y. Hu, Y. Xu, Z. Wang, Y. Fang, J. Yan, W. Jiang, Y. Peng, and Y. Huang, "Learning guided implicit depth function with scale-aware feature fusion," *IEEE Transactions on Image Processing*, vol. 34, pp. 3309–3322, 2025.

[16] Y. Wen, B. Sheng, P. Li, W. Lin, and D. D. Feng, "Deep color guided coarse-to-fine convolutional network cascade for depth image super-resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 994–1006, 2019.

[17] H. Wang, M. Yang, C. Zhu, and N. Zheng, "Rgb-guided depth map recovery by two-stage coarse-to-fine dense crf models," *IEEE Transactions on Image Processing*, vol. 32, pp. 1315–1328, 2023.

[18] X. Wang, X. Chen, B. Ni, Z. Tong, and H. Wang, "Learning continuous depth representation via geometric spatial aggregator," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 2698–2706, 2023.

[19] X. Deng and P. L. Dragotti, "Deep convolutional neural network for multi-modal image restoration and fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020. 1

[20] J. Yang, W. Xu, X. Ye, P. Frossard, and K. Li, "Graph based non-uniform sampling and reconstruction of depth maps," in *ICIP*, pp. 2324–2328, 2019. 2

[21] X. Liu, D. Zhai, R. Chen, X. Ji, D. Zhao, and W. Gao, "Depth super-resolution via joint color-guided internal and external regularizations," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1636–1645, 2019. 3

[22] J. Wang, L. Sun, R. Xiong, Y. Shi, Q. Zhu, and B. Yin, "Depth map super-resolution based on dual normal-depth regularization and graph laplacian prior," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 6, pp. 3304–3318, 2022. 3

[23] Y. Zhang, Y. Feng, X. Liu, D. Zhai, X. Ji, H. Wang, and Q. Dai, "Color-guided depth image recovery with adaptive data fidelity and transferred graph laplacian regularization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 320–333, 2020. 2

[24] Q. Wang, S. Li, H. Qin, and A. Hao, "Super-resolution of multi-observed rgb-d images based on nonlocal regression and total variation," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1425–1440, 2016. 2

[25] Y. Zuo, Y. Fang, Y. Yang, X. Shang, and Q. Wu, "Depth map enhancement by revisiting multi-scale intensity guidance within coarse-to-fine stages," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4676–4687, 2020. 2

[26] X. Ye, B. Sun, Z. Wang, J. Yang, R. Xu, H. Li, and B. Li, "Pmbanet: Progressive multi-branch aggregation network

for scene depth super-resolution," *IEEE Transactions on Image Processing*, vol. 29, pp. 7427–7442, 2020.

[27] X. Qiao, M. Poggi, P. Deng, H. Wei, C. Ge, and S. Mattoccia, "Rgb guided tof imaging system: A survey of deep learning-based methods," *International Journal of Computer Vision*, vol. 132, no. 11, pp. 4954–4991, 2024.

[28] Z. Yan, Z. Wang, H. Dong, , J. Li, J. Yang, and G. H. Lee, "Ducos: Duality constrained depth super-resolution via foundation model," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

[29] X. Deng, J. Xu, F. Gao, X. Sun, and M. Xu, "DeepM^2m2cdl: Deep multi-scale multi-modal convolutional dictionary learning network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 5, pp. 2770–2787, 2024.

[30] Y. Zuo, Y. Xu, Y. Zeng, Y. Fang, X. Huang, and J. Yan, "A2 gstran: Depth map super-resolution via asymmetric attention with guidance selection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 6, pp. 4668–4681, 2024.

[31] S. Du, Y. Zou, Z. Wang, X. Li, Y. Li, C. Shang, and Q. Shen, "Unsupervised hyperspectral image super-resolution via self-supervised modality decoupling," *International Journal of Computer Vision*, vol. 134, no. 4, p. 152, 2026.

[32] J. Yuan, H. Jiang, X. Li, J. Qian, J. Li, and J. Yang, "Recurrent structure attention guidance for depth super-resolution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 3331–3339, 2023.

[33] J. Wang, P. Liu, and F. Wen, "Self-supervised learning for rgb-guided depth enhancement by exploiting the dependency between rgb and depth," *IEEE Transactions on Image Processing*, vol. 32, pp. 159–174, 2023. 2

[34] R. de Lutio, A. Becker, S. D'Aronco, S. Russo, J. D. Wegner, and K. Schindler, "Learning graph regularisation for guided super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2, 3, 6

[35] G. Gartzonikas, E. Tsiligianni, N. Deligiannis, and L. P. Kondi, "A graph laplacian regularizer from deep features for depth map super-resolution," *Information*, vol. 16, no. 6, 2025. 2, 3

[36] A. Ortega, P. Frossard, J. Kovačević, J. M. F. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018. 2

[37] Z. Jiang, Y. Hou, H. Yue, J. Yang, and C. Hou, "Depth super-resolution from rgb-d pairs with transform and spatial domain regularization," *IEEE Transactions on Image Processing*, vol. 27, pp. 2587–2602, 2018. 3

[38] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Advances in Neural Information Processing Systems*, vol. 24, 2011. 4

[39] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Proceedings of European Conference on Computer Vision*, pp. 611–625, 2012. 6

[40] I. Vasiljevic, N. Kolkin, S. Zhang, R. Luo, H. Wang, F. Z. Dai, A. F. Daniele, M. Mostajabi, S. Basart, M. R. Walter, *et al.*, "Diode: A dense indoor and outdoor depth dataset," *arXiv preprint arXiv:1908.00463*, 2019. 6, 7, 8

[41] S. Song, S. P. Lichtenberg, and J. Xiao, "Sun rgb-d: A rgb-d scene understanding benchmark suite," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 567–576, 2015. 6

[42] L. He, H. Zhu, F. Li, H. Bai, R. Cong, C. Zhang, C. Lin, M. Liu, and Y. Zhao, "Towards fast and accurate real-world depth super-resolution: Benchmark dataset and baseline," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9229–9238, 2021. 6, 7

[43] J. Cho, D. Min, Y. Kim, and K. Sohn, "Deep monocular depth estimation leveraging a large-scale outdoor stereo dataset," *Expert Systems with Applications*, vol. 178, p. 114877, 2021. 6

[44] Y. Li, J. B. Huang, N. Ahuja, and M. H. Yang, "Joint image filtering with deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1909–1923, 2019. 6

[45] J. Dong, J. Pan, J. S. Ren, L. Lin, J. Tang, and M.-H. Yang, "Learning spatially variant linear representation models for joint filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8355–8370, 2022. 6

[46] B. Kim, J. Ponce, and B. Ham, "Deformable kernel networks for joint image filtering," *International Journal of Computer Vision*, pp. 1–22, 2021. 6, 7

[47] J. Tang, X. Chen, and G. Zeng, "Joint implicit image function for guided depth super-resolution," in *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 4390–4399, 2021. 6

[48] Y. Yang, Q. Cao, J. Zhang, and D. Tao, "Codon: on orchestrating cross-domain attentions for depth super-resolution," *International Journal of Computer Vision*, vol. 130, no. 2, pp. 267–284, 2022. 6

[49] Z. Zhong, X. Liu, J. Jiang, D. Zhao, Z. Chen, and X. Ji, "High-resolution depth maps imaging via attention-based hierarchical multi-modal fusion," *IEEE Trans. Image Process.*, vol. 31, pp. 648–663, 2022. 6

[50] N. Metzger, R. C. Daudt, and K. Schindler, "Guided depth super-resolution by deep anisotropic diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18237–18246, June 2023. 6, 7

[51] M. Zhou, K. Yan, J. Pan, W. Ren, Q. Xie, and X. Cao, "Memory-augmented deep unfolding network for guided image super-resolution," *International Journal of Computer Vision*, vol. 131, no. 1, pp. 215–242, 2023. 6

[52] Z. Zhao, J. Zhang, X. Gu, C. Tan, S. Xu, Y. Zhang, R. Timofte, and L. Van Gool, "Spherical space feature decomposition for guided depth map super-resolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12547–12558, 2023. 6, 7

[53] Z. Wang, Z. Yan, and J. Yang, "Sgnet: Structure guided network via gradient-frequency awareness for depth map super-resolution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 5823–5831, 2024. 6, 7

[54] Z. Zhong, X. Liu, J. Jiang, D. Zhao, and S. Wang, "Dual-level cross-modality neural architecture search for guided image super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 9, pp. 8249–8267, 2025. 6, 7

[55] Z. Wang, Z. Yan, J. Pan, G. Gao, K. Zhang, and J. Yang, "Dornet: A degradation oriented and regularized network for blind depth super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15813–15822, 2025. 6, 7

[56] J. Yuan, H. Jiang, X. Li, J. Qian, J. Li, and J. Yang, "Structure flow-guided network for real depth super-resolution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 3340–3348, 2023. 7

[57] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Proceedings of the European Conference on Computer Vision*, pp. 611–625, 2012. 7, 8

[58] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013. 8

Dual Graph Regularized Deep Unfolding Network for Guided Depth Map Super-resolution
–Supplementary Material–

This document provides additional details to complement the main paper:
- **Section A:** More related work.
- **Section B:** Proof for Lemma 1.
- **Section C:** Additional implementation details and experimental settings.
- **Section D:** Extended experimental results and visual comparisons.
- **Section E:** Convergence Discussion.

A. Related Work

A.1. Guided Depth Map Super-resolution

Guided depth map super-resolution (GDSR) refers to the process of restoring fine spatial details in a low-resolution (LR) depth map using a high-resolution (HR) color image as a reference. Depending on the algorithmic design paradigm, existing methods can be broadly grouped into filtering-based, optimization-based, and learning-based approaches.

Filtering-based methods enhance the resolution of LR images by using a weighted averaging technique, where the weights are derived from various strategies. For instance, Eisemann *et al.* [1] propose an edge-preserving filter that considers both spatial distance and intensity difference. Based on the local linear assumption, He *et al.* [2] introduce a translation variant image filter. These approaches are attractive for their simplicity and low computational cost. Nevertheless, their dependence on the local linear assumption restricts their capability to capture complex cross-modal correlations.

Optimization-based methods address the GDSR problem from a global optimization perspective. The objective function of these methods typically comprises two components: data fidelity term and regularization term. The former maintains data consistency, while the latter enhances structural alignment with the guidance image. Representative regularization functions include the Markov random field (MRF) [3], Auto-Regress model (AR) [4] and conditional random field (CRF) [5]. Although these methods overcome certain limitations of filtering-based approaches, they rely on manually designed objective functions, which may not fully capture complex image priors.

Recently, learning-based methods have emerged as the dominant paradigm in various computer vision tasks, outperforming traditional hand-designed methods with their ability to autonomously learn complex features. To this end, more and more advanced architectures are proposed and applied in the field of GDSR [6–12]. For example, Li *et al.* [6] present a data-driven filter capable of adaptively identifying and injecting critical structural cues from the guidance image into the target image. Kim *et al.* [13] develop deformable kernel networks for joint image filtering, where both sampling locations and kernel weights are learned in a content-adaptive manner. Tang *et al.* [14] introduce a dual-branch framework for jointly learning depth map SR and monocular depth estimation, aiming to bridge the information gap between the two tasks. Yuan *et al.* [10] address noise and distortion in real-world low-resolution depth maps by estimating structure flow from guidance RGB images. Tang *et al.* [15] develop a joint implicit image function framework that represents the high-resolution depth map as a continuous function of spatial coordinates guided by RGB input. This implicit formulation enables accurate depth prediction without explicit upsampling and effectively preserves structural details. Zhao *et al.* [7] propose a discrete cosine transform network with semi-coupled blocks to better extract informative features. Zuo *et al.* [16] develop a guided implicit function framework that models depth as a continuous function of spatial coordinates and introduces a scale-aware fusion module for effective multi-scale feature aggregation. Zhou *et al.* [17] design a spatial frequency information integration network that fuses cross-modality features in both the spatial and frequency domains. Wang *et al.* [11] propose a degradation-oriented blind depth super-resolution method that jointly learns degradation representation and reconstruction. Yan *et al.* [12] leverage a frozen foundation model to extract semantic-aligned global features and introduce a duality-constrained optimization

framework that enforces consistency between the degradation and reconstruction processes. Although these learning-based methods have achieved superior performance by leveraging powerful feature extraction capabilities, they are typically designed as black-box models with limited interpretability, and performance improvements often come at the cost of increased model complexity and parameters.

A.2. Deep Unfolding Networks

In recent years, Deep Unfolding Networks (DUNs) have gained wide attention for their ability to combine the interpretability of model-based optimization with the learning flexibility of deep neural networks. By unfolding iterative solvers into structured layers, DUNs offer transparent and controllable architectures for various image restoration tasks. A representative work, LISTA [18], unfolds sparse coding into a trainable network, inspiring a range of DUN-based models in denoising [19, 20], deblurring [21, 22], and super-resolution [23, 24]. In the context of GDSR, several studies have adopted DUN-based designs. Riegler *et al.* [25] propose a deep primal-dual network that mimics optimization dynamics, using dual gradients to enforce RGB-depth consistency. Marivani *et al.* [26] propose a multimodal unfolding framework that fuses RGB and depth priors at each stage to progressively refine structural details. Zhao *et al.* [27] propose a discontinuity-aware unfolding network that jointly refines depth and gradient for better structure preservation. Metzler *et al.* [28] introduce a deep anisotropic diffusion model that simulates diffusion iterations with learned directional filters. Dai et al. [29] proposed an indoor depth recovery framework based on deep unfolding with a non-local prior, where a non-local auto-regressive regularization term is introduced to exploit repetitive depth structures in the scene. More recently, De Lutio *et al.* propose LGR [30], a hybrid framework that incorporates graph-based regularization into convolutional neural networks for guided depth super-resolution. Specifically, LGR learns a data-driven similarity matrix to construct the Laplacian graph and embeds a differentiable regularization module into the end-to-end training pipeline. As the first work to explicitly combine deep learning with graph optimization in GDSR, LGR bridges an important research gap and achieves competitive performance. However, LGR still faces several notable limitations. Although sparse graph construction effectively reduces computational cost, it inherently restricts the receptive field to local neighborhoods, making it difficult to capture long-range structural dependencies. In addition, the fixed adjacency structure restricts the model to fixed input sizes, which reduces its ability to handle images in different resolutions. Finally, flattening the spatial features into one-dimensional vectors during graph processing breaks the natural two-dimensional layout of the depth map and may reduce geometric accuracy and spatial consistency.

B. Proof for Lemma 1

This proof is based on **Definition 1** in the main paper, which defines the dual-graph regularization term:

$$\min_{\boldsymbol{X}} \frac{1}{2}\sum_{i=1}^{H}\sum_{j=1}^{H}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_2^2(\boldsymbol{S}_r)_{ij}+\frac{1}{2}\sum_{i=1}^{W}\sum_{j=1}^{W}\|(\boldsymbol{X}^\top)_i-(\boldsymbol{X}^\top)_j\|_2^2(\boldsymbol{S}_c)_{ij}. \tag{1}$$

Lemma 1: *By introducing the Laplacian matrices $\boldsymbol{L}_r \in \mathbb{R}^{H\times H}$ and $\boldsymbol{L}_c \in \mathbb{R}^{W\times W}$ in terms with $\boldsymbol{S}_r$ and $\boldsymbol{S}_c$, respectively, Eq. (1) is equivalent to the following expression,*

$$\min_{\boldsymbol{X}} tr(\boldsymbol{X}^\top\boldsymbol{L}_r\boldsymbol{X})+tr(\boldsymbol{X}\boldsymbol{L}_c\boldsymbol{X}^\top). \tag{2}$$

Proof: *As the two Laplacian matrices $\boldsymbol{L}_r$ and $\boldsymbol{L}_c$ are constructed by the affinity graphs $\boldsymbol{S}_r \in \mathbb{R}^{H\times H}$ and $\boldsymbol{S}_c \in \mathbb{R}^{W\times W}$, respectively, we can formulate that*

$$\boldsymbol{L}_r=\boldsymbol{U}_r-\boldsymbol{S}_r, \tag{3}$$

$$\boldsymbol{L}_c=\boldsymbol{U}_c-\boldsymbol{S}_c, \tag{4}$$

where $\boldsymbol{U}_r \in \mathbb{R}^{H\times H}$ and $\boldsymbol{U}_c \in \mathbb{R}^{W\times W}$ are two degree matrices, i.e., $(\boldsymbol{U}_r)_{ii}=\sum_j(\boldsymbol{S}_r)_{ij}$ and $(\boldsymbol{U}_c)_{ii}=\sum_j(\boldsymbol{S}_c)_{ij}$. According to the definition of vector norm, we get

$$\frac{1}{2}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|_2^2=\frac{1}{2}(\|\boldsymbol{X}_i\|_2^2+\|\boldsymbol{X}_j\|_2^2-2\langle\boldsymbol{X}_i,\boldsymbol{X}_j\rangle), \tag{5}$$

By combining Eq. (1) and Eq. (5), we can rewrite Eq. (1) as

$$\min_{\boldsymbol{X}} \frac{1}{2}\sum_{i=1}^{H}\sum_{j=1}^{H}(\|\boldsymbol{X}_i\|_2^2+\|\boldsymbol{X}_j\|_2^2-2\langle\boldsymbol{X}_i,\boldsymbol{X}_j\rangle)(\boldsymbol{S}_r)_{ij}+\frac{1}{2}\sum_{i=1}^{W}\sum_{j=1}^{W}(\|(\boldsymbol{X}^\top)_i\|_2^2+\|(\boldsymbol{X}^\top)_j\|_2^2-2\langle(\boldsymbol{X}^\top)_i,(\boldsymbol{X}^\top)_j\rangle)(\boldsymbol{S}_c)_{ij}. \tag{6}$$

Then, Eq. (6) can be reformulated as

$$\begin{aligned}
&\min_{\boldsymbol{X}} \sum_{i=1}^{H}\sum_{j=1}^{H}(\|\boldsymbol{X}_i\|_2^2 - 2\langle \boldsymbol{X}_i, \boldsymbol{X}_j\rangle)(\boldsymbol{S}_r)_{ij} + \sum_{i=1}^{W}\sum_{j=1}^{W}(\|(\boldsymbol{X}^\top)_i\|_2^2 - 2\langle (\boldsymbol{X}^\top)_i (\boldsymbol{X}^\top)_j\rangle)(\boldsymbol{S}_c)_{ij} \\
=&\min_{\boldsymbol{X}} \sum_{i=1}^{H}((\boldsymbol{X}^\top \boldsymbol{U}_r \boldsymbol{X})_{ii} - (\boldsymbol{X}^\top \boldsymbol{S}_r \boldsymbol{X})_{ii}) + \sum_{i=1}^{W}((\boldsymbol{X}\boldsymbol{U}_c\boldsymbol{X}^\top)_{ii} - (\boldsymbol{X}\boldsymbol{S}_c\boldsymbol{X}^\top)_{ii}) \\
=&\min_{\boldsymbol{X}}\ tr(\boldsymbol{X}^\top \boldsymbol{L}_r \boldsymbol{X}) + tr(\boldsymbol{X}\boldsymbol{L}_c\boldsymbol{X}^\top).
\end{aligned} \tag{7}$$

The proof is completed.

C. Experimental Details

C.1. Dataset

Following the experimental protocols of existing guided depth super-resolution methods [17, 31–33], we evaluate our method on two benchmark datasets: NYU v2 [34] and RGB-D-D [35].

For the NYU v2 dataset [34], the first 1,000 RGB-D image pairs are used for training, while the remaining 449 pairs are used for testing. To further evaluate the generalization ability of our model, we also perform cross-dataset testing on five representative datasets: (1) 1,064 RGB-D images from Sintel [36], (2) the test set of DIDOE [37], (3) 500 image pairs from the SUN RGB-D test set [38], (4) the official test set of RGB-D-D [35], and (5) the test set of DIML Indoor [39]. Low-resolution (LR) depth maps are generated by downsampling the high-resolution (HR) ground-truth depth maps using bicubic interpolation with scaling factors of $4\times$, $8\times$, and $16\times$. The corresponding HR RGB images are used as guidance.

The RGB-D-D dataset [35] is a real-world benchmark dataset, where the HR depth maps are captured using a Helios ToF camera, and the LR depth maps are obtained with a Huawei P30 Pro. To ensure fair and consistent evaluation, we strictly follow the official training and testing splits provided by the dataset.

C.2. Implementation Details

Our framework is implemented in PyTorch and trained on a workstation equipped with two NVIDIA RTX 5090 GPUs. We use image patches of size 320×320 and a batch size of 16. Training is conducted for 200 epochs using the Adam optimizer [40] with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1 \times 10^{-8}$. The learning rate is initialized to 1×10^{-4} and reduced by half after the 100th epoch to ensure stable convergence.

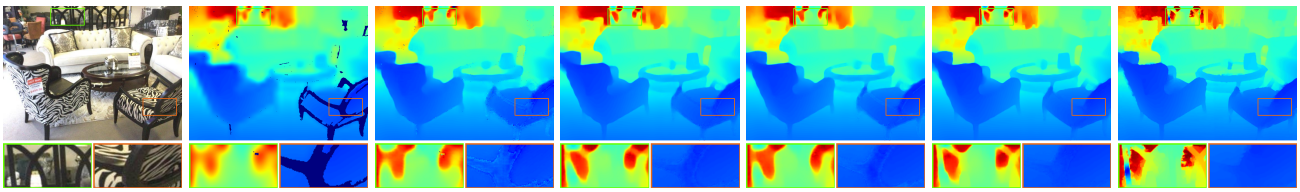
The proximal net is implemented as a lightweight U-Net with two downsampling layers and two upsampling layers. At each resolution level, a residual block is used as the basic building unit. The downsampling path aggregates multi-scale structural information, and the upsampling path restores fine details with skip connections. The iterative parameters α, μ, β, and λ are initialized to 1 and jointly optimized during training. The model performs three iterative refinement stages to progressively enhance reconstruction quality. The $\mathcal{L}_1$ loss is used for supervision. Following prior works [7, 17, 33], we apply standard data augmentation techniques, including random flipping and rotation, to improve the model's generalization ability. For fair comparison, all learning-based methods are trained and evaluated using the same datasets and experimental protocols. The root mean square error (RMSE) is used to measure the difference between the predicted depth values and the ground-truth depth values:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(D_i - \hat{D}_i)^2}, \tag{8}$$

where D_i and $\hat{D}_i$ denote the ground-truth and predicted depth values at pixel i, respectively, and N is the total number of pixels. A lower RMSE indicates better reconstruction performance.

D. Experiments

In this section, we present more visualization results as well as additional subjective and objective comparisons to further demonstrate the visual quality and quantitative performance of our method. In addition, we conduct a new experiment on RGB-guided joint depth map completion and super-resolution to further verify the effectiveness and generalization capability of the proposed method.



(a) Guidance (b) Input (c) DJFR [6] (d) LGR [30] (e) DORNet [11] (f) **LapNet** (g) GT

Figure 1. Qualitative comparison for joint depth super-resolution and completion on the SUN RGB-D dataset [38].

D.1. Joint Depth Map Completion and Super-resolution

Experimental Settings. To evaluate the robustness of our method under realistic conditions, we simulate Kinect-like degradation, which is commonly observed in structured light sensors. This degradation includes two types of artifacts: structured holes near object boundaries and random dropouts in flat or reflective regions. To replicate this, we first downsample the HR depth maps using bicubic interpolation. Then, we apply a binary degradation mask that combines: i) structured missing regions, generated by detecting and dilating depth edges to mimic boundary-related holes, and ii) random missing pixels, introduced by sampling a Bernoulli distribution with a predefined missing probability. The final degraded depth map is obtained by masking the downsampled depth map through element-wise multiplication. We conduct experiments on the SUN RGB-D [38], and allocate the first 1,500 image pairs for training, with the remaining 355 pairs reserved for evaluation.

Results. We compare our method with several representative baselines, as shown in Table 1. LapNet achieves the best performance at all three scaling factors ($4\times$, $8\times$, and $16\times$), highlighting its strong robustness in practical degradation settings, particularly in cases involving simultaneous depth value absence and spatial resolution loss. Visual comparisons are presented in Fig. 1. As illustrated, DJFR [6] fails to recover missing regions, while LGR [30] and DORNet [11] can inpaint the holes but often result in incomplete or distorted structures, as seen in the zoomed-in areas. In contrast, our method not only fills in the missing depth values but also reconstructs more complete and accurate structures. These results highlight the effectiveness of our dual-prior design in addressing complex degradations and preserving structural integrity under challenging conditions.

Table 1. RMSE comparison on SUN RGB [41] dataset for joint depth map completion and super-resolution.

Scale	Bicubic	DJFR [6]	DKN [13]	DCTNet [7]	MMNet [42]	AHMF [43]	LGR [30]	SFINet [17]	DORNet [11]	**LapNet**
$4\times$	22.69	3.59	2.73	2.57	3.17	<u>2.45</u>	2.78	2.73	2.46	**1.92**
$8\times$	22.83	3.39	3.06	2.66	2.69	2.52	2.87	3.04	<u>2.49</u>	**2.05**
$16\times$	23.15	3.87	3.42	3.05	3.59	2.82	3.22	3.40	<u>2.74</u>	**2.31**

D.2. Comparison with More Methods

We further compare the proposed method with two diffusion-based methods and one deep unfolding-based methods. The corresponding quantitative results are presented in Table 2. As can be seen, our method achieves the best performance among all the compared methods.

Table 2. RMSE/MAE comparison on NYU v2 [34] dataset for guided depth map super-resolution.

Scale	StableSR [44]	TVT [45]	DeepSN-Net [46]	**LapNet**
$4\times$	1.86/0.71	1.54/0.60	1.42/0.52	**1.05/0.38**
$8\times$	4.04/1.62	3.79/1.55	3.66/1.48	**2.33/0.91**
$16\times$	6.92/2.84	6.41/2.63	5.98/2.41	**4.55/1.86**

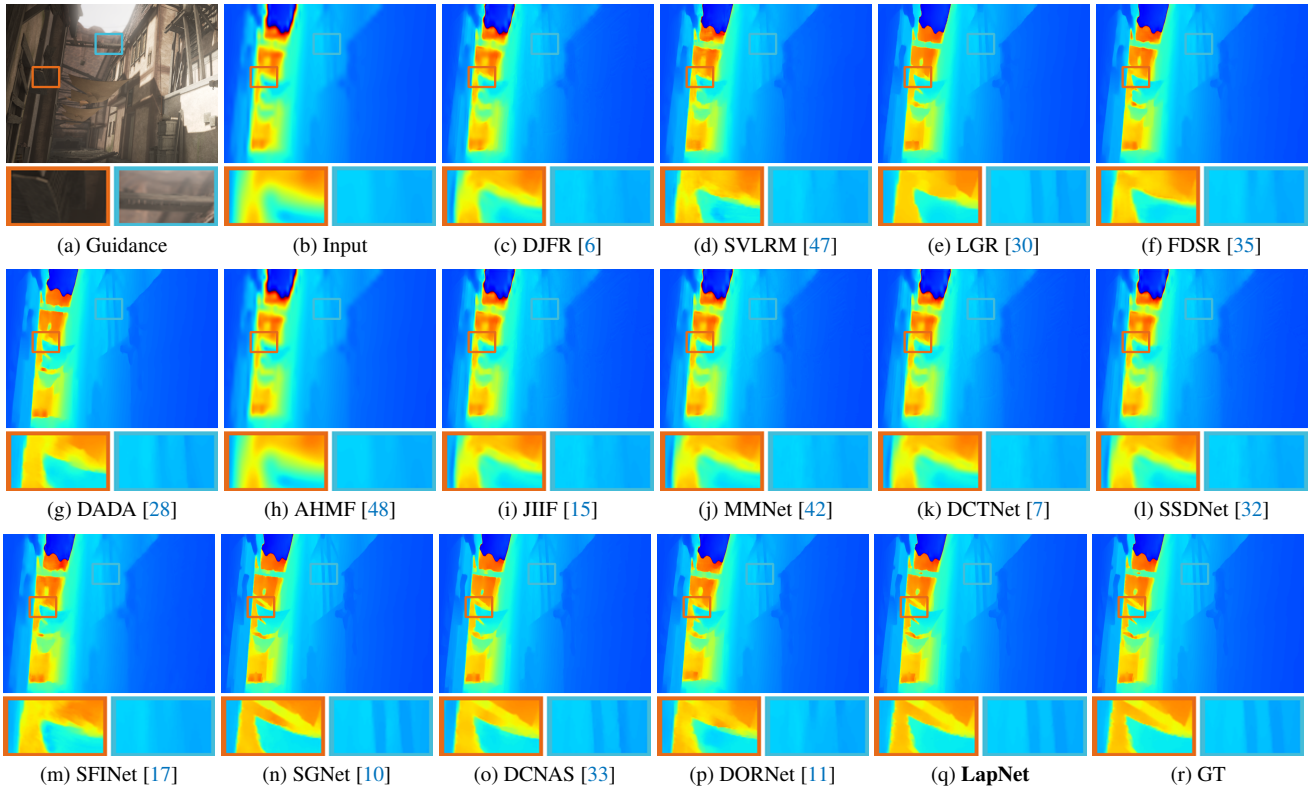
D.3. More ablation experiments

Effect of Skip Connections in the Proximal Network. To prevent the loss of important structural information during cross-stage propagation, we design a skip connection strategy in the proximal network, enabling the model to reuse features across iterations. Specifically, both the reconstructed depth $\boldsymbol{X}_k$ and the intermediate decoder features $\boldsymbol{F}_{k-1}$ from the previous stage are passed to the encoder of the current stage. To validate the effectiveness of this strategy, we construct a variant model, `Model8`, in which all skip connections are removed and only current-stage inputs are used. As presented in Table 3,

eliminating skip connections leads to consistent performance degradation on all three benchmark datasets, indicating that they are beneficial for reducing information loss and improving reconstruction quality.

Table 3. **Ablation Study.** RMSE comparison of proximal network design strategies for 8× GDSR.

Method	NYU v2 [34]	Sintel [36]	DIDOE [37]
`Model8`	2.52	5.26	5.84
LapNet	**2.33**	**5.05**	**5.51**



(a) Guidance (b) Input (c) DJFR [6] (d) SVLRM [47] (e) LGR [30] (f) FDSR [35]

(g) DADA [28] (h) AHMF [48] (i) JIIF [15] (j) MMNet [42] (k) DCTNet [7] (l) SSDNet [32]

(m) SFINet [17] (n) SGNet [10] (o) DCNAS [33] (p) DORNet [11] (q) **LapNet** (r) GT

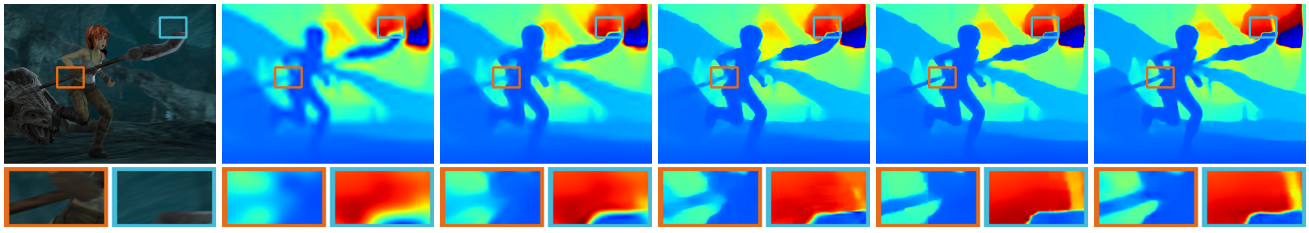
Figure 2. Qualitative comparison for 16× depth map super-resolution (*Sintel dataset [36] Image-85*).

D.4. More Visualization Results

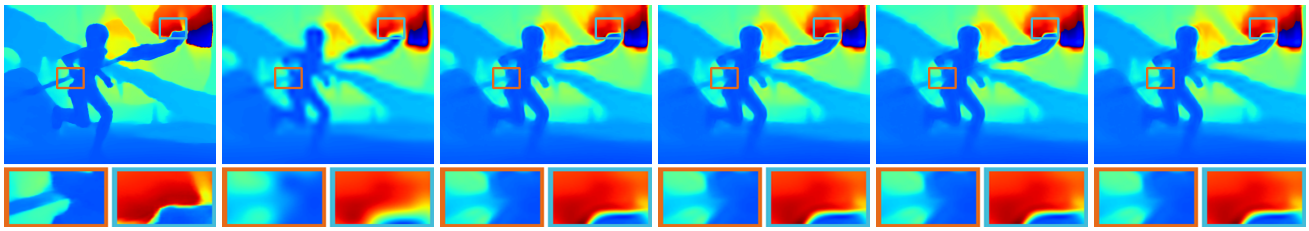
In this subsection, we present additional qualitative comparison results (see Fig. 2, Fig. 3, Fig. 4, Fig. 5). As shown in the figures, our method reconstructs more complete structural details and effectively suppresses artifacts caused by incorrect texture transfer.

E. Convergence Discussion

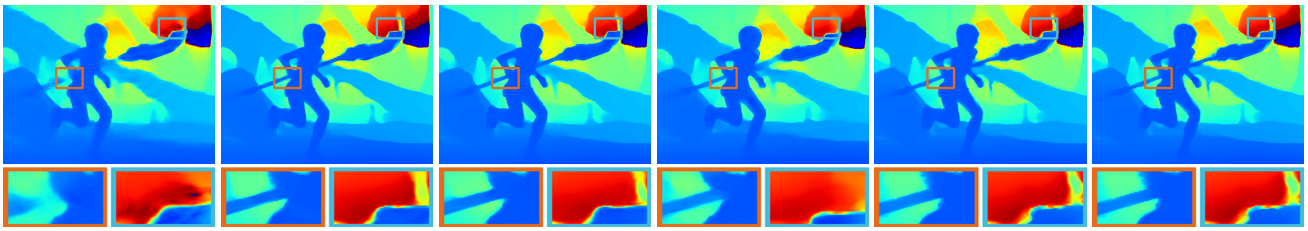
The proposed method is derived by unfolding a fixed number of ADMM iterations, and thus each stage of the network corresponds to one step of the underlying optimization algorithm. This gives the model a clear optimization interpretation and preserves the key structure of ADMM. While learnable modules are introduced to enhance representation power, the overall framework remains optimization-driven and retains strong interpretability. Following the common practice in deep unfolding literature, we do not pursue a strict theoretical proof for the convergence of the learned network, since the inclusion of neural parameterization generally makes such analysis difficult. Instead, we highlight that the proposed model demonstrates stable empirical convergence and consistent reconstruction improvement in practice across different experimental settings.



(a) Guidance (b) Input (c) DJFR [6] (d) SVLRM [47] (e) LGR [30] (f) FDSR [35]

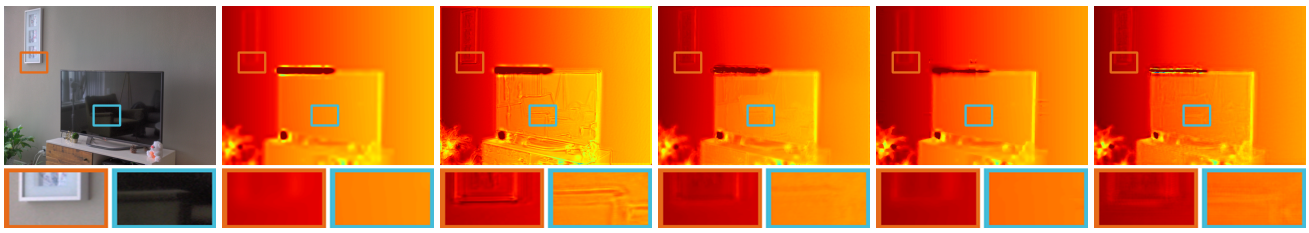


(g) DADA [28] (h) AHMF [48] (i) JIIF [15] (j) MMNet [42] (k) DCTNet [7] (l) SSDNet [32]

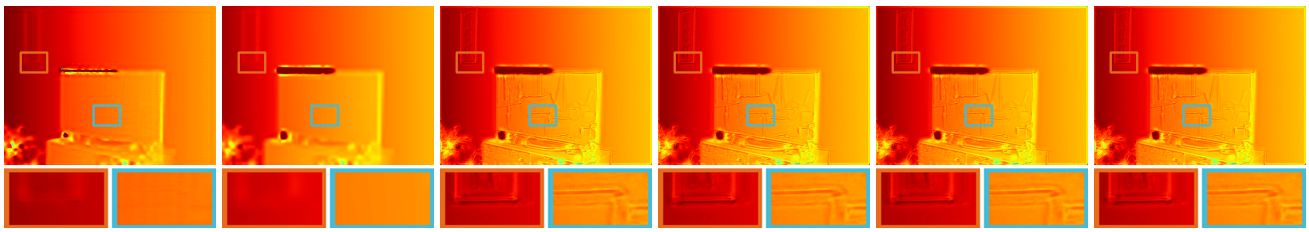


(m) SFINet [17] (n) SGNet [10] (o) DCNAS [33] (p) DORNet [11] (q) **LapNet** (r) GT

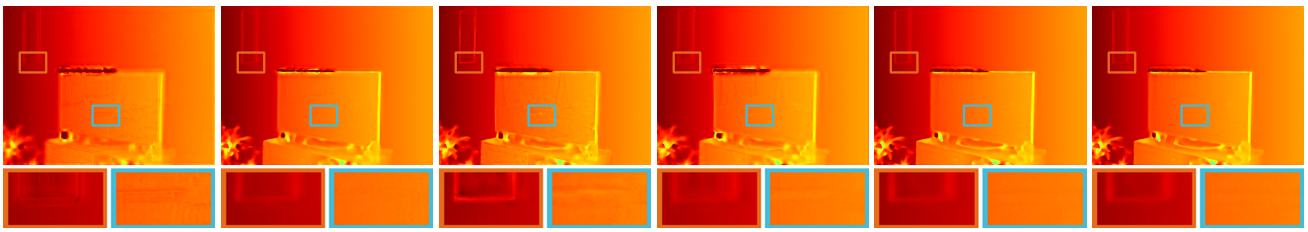
Figure 3. Qualitative comparison for 16× depth map super-resolution (*Sintel dataset [36] Image-483*).



(a) Guidance (b) Input (c) DJFR [6] (d) SVLRM [47] (e) LGR [30] (f) FDSR [35]

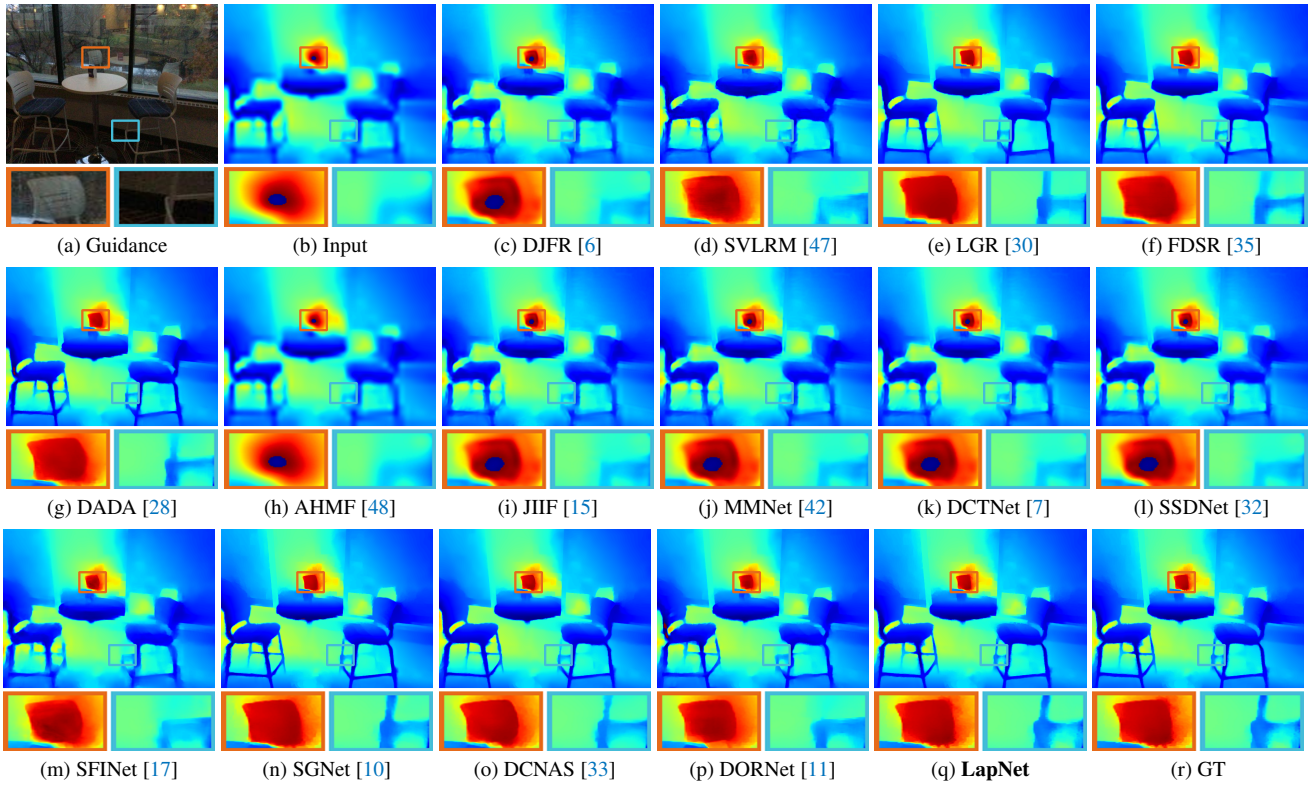


(g) DADA [28] (h) AHMF [48] (i) JIIF [15] (j) MMNet [42] (k) DCTNet [7] (l) SSDNet [32]



(m) SFINet [17] (n) SGNet [10] (o) DCNAS [33] (p) DORNet [11] (q) **LapNet** (r) GT

Figure 4. Qualitative comparison for 16× depth map super-resolution (*DIDOE dataset [37] Image-110*).



(a) Guidance (b) Input (c) DJFR [6] (d) SVLRM [47] (e) LGR [30] (f) FDSR [35]

(g) DADA [28] (h) AHMF [48] (i) JIIF [15] (j) MMNet [42] (k) DCTNet [7] (l) SSDNet [32]

(m) SFINet [17] (n) SGNet [10] (o) DCNAS [33] (p) DORNet [11] (q) **LapNet** (r) GT

Figure 5. Qualitative comparison for 16× depth map super-resolution (*SUN RGB-D dataset [38] Image-393*).

References

[1] E. Eisemann and F. Durand, "Flash photography enhancement via intrinsic relighting," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 673–678, 2004. 1

[2] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013. 1

[3] Y. Zuo, Q. Wu, J. Zhang, and P. An, "Minimum spanning forest with embedded edge inconsistency measurement model for guided depth map enhancement," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4145–4159, 2018. 1

[4] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from rgb-d data using an adaptive autoregressive model," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3443–3458, 2014. 1

[5] H. Wang, M. Yang, C. Zhu, and N. Zheng, "Rgb-guided depth map recovery by two-stage coarse-to-fine dense crf models," *IEEE Transactions on Image Processing*, vol. 32, pp. 1315–1328, 2023. 1

[6] Y. Li, J. B. Huang, N. Ahuja, and M. H. Yang, "Joint image filtering with deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1909–1923, 2019. 1, 4, 5, 6, 7

[7] Z. Zhao, J. Zhang, S. Xu, Z. Lin, and H. Pfister, "Discrete cosine transform network for guided depth map super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5697–5707, 2022. 1, 3, 4, 5, 6, 7

[8] W. Shi, M. Ye, and B. Du, "Symmetric uncertainty-aware feature transmission for depth super-resolution," in *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 3867–3876, 2022.

[9] X. Wang, X. Chen, B. Ni, Z. Tong, and H. Wang, "Learning continuous depth representation via geometric spatial aggregator," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, pp. 2698–2706, 2023.

[10] Z. Wang, Z. Yan, and J. Yang, "Sgnet: Structure guided network via gradient-frequency awareness for depth map super-resolution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 5823–5831, 2024. 1, 5, 6, 7

[11] Z. Wang, Z. Yan, J. Pan, G. Gao, K. Zhang, and J. Yang, "Dornet: A degradation oriented and regularized network for blind depth super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15813–15822, 2025. 1, 4, 5, 6, 7

[12] Z. Yan, Z. Wang, H. Dong, , J. Li, J. Yang, and G. H. Lee, "Ducos: Duality constrained depth super-resolution via foundation model," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. 1

[13] B. Kim, J. Ponce, and B. Ham, "Deformable kernel networks for joint image filtering," *International Journal of Computer Vision*, pp. 1–22, 2021. 1, 4

[14] Q. Tang, R. Cong, R. Sheng, L. He, D. Zhang, Y. Zhao, and S. Kwong, "Bridgenet: A joint learning network of depth map super-resolution and monocular depth estimation," in *Proceedings of ACM International Conference on Multimedia*, p. 2148–2157, 2021. 1

[15] J. Tang, X. Chen, and G. Zeng, "Joint implicit image function for guided depth super-resolution," in *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 4390–4399, 2021. 1, 5, 6, 7

[16] Y. Zuo, Y. Hu, Y. Xu, Z. Wang, Y. Fang, J. Yan, W. Jiang, Y. Peng, and Y. Huang, "Learning guided implicit depth function with scale-aware feature fusion," *IEEE Transactions on Image Processing*, vol. 34, pp. 3309–3322, 2025. 1

[17] M. Zhou, J. Huang, K. Yan, D. Hong, X. Jia, J. Chanussot, and C. Li, "A general spatial-frequency learning framework for multimodal image fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1, 3, 4, 5, 6, 7

[18] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pp. 399–406, 2010. 2

[19] S. Lefkimmiatis, "Non-local color image denoising with convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3587–3596, 2017. 2

[20] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1256–1272, 2017. 2

[21] H. Wang, T. Zhang, M. Yu, J. Sun, W. Ye, C. Wang, and S. Zhang, "Stacking networks dynamically for image restoration based on the plug-and-play framework," in *Proceedings of European Conference on Computer Vision*, pp. 446–462, Springer, 2020. 2

[22] C. Mou, Q. Wang, and J. Zhang, "Deep generalized unfolding networks for image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17378–17389, 2022. 2

[23] Q. Ning, W. Dong, G. Shi, L. Li, and X. Li, "Accurate and lightweight image super-resolution with model-guided deep unfolding network," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 2, pp. 240–252, 2021. 2

[24] K. Zhang, L. Van Gool, and R. Timofte, "Deep unfolding network for image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3214–3223, 2020. 2

[25] G. Riegler, D. Ferstl, M. Rüther, and B. Horst, "A deep primal-dual network for guided depth super-resolution," in *BMVC*, 2016. 2

[26] I. Marivani, E. Tsiligianni, B. Cornelis, and N. Deligiannis, "Multimodal deep unfolding for guided image super-resolution," *IEEE Transaction on Image Processing*, vol. 29, pp. 8443–8456, 2020. 2

[27] L. Zhao, J. Zhang, J. Zhang, H. Bai, and A. Wang, "Joint discontinuity-aware depth map super-resolution via dual-tasks driven unfolding network," *TIM*, vol. 73, pp. 1–14, 2024. 2

[28] N. Metzger, R. C. Daudt, and K. Schindler, "Guided depth super-resolution by deep anisotropic diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18237–18246, June 2023. 2, 5, 6, 7

[29] Y. Dai, J. Zhang, F. Fang, and G. Zhang, "Indoor depth recovery based on deep unfolding with non-local prior," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12355–12364, 2023. 2

[30] R. de Lutio, A. Becker, S. D'Aronco, S. Russo, J. D. Wegner, and K. Schindler, "Learning graph regularisation for guided super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2, 4, 5, 6, 7

[31] Z. Zhong, X. Liu, J. Jiang, D. Zhao, and X. Ji, "Guided depth map super-resolution: A survey," *ACM Computing Surveys*, vol. 55, no. 14, pp. 1–36, 2023. 3

[32] Z. Zhao, J. Zhang, X. Gu, C. Tan, S. Xu, Y. Zhang, R. Timofte, and L. Van Gool, "Spherical space feature decomposition for guided depth map super-resolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12547–12558, 2023. 5, 6, 7

[33] Z. Zhong, X. Liu, J. Jiang, D. Zhao, and S. Wang, "Dual-level cross-modality neural architecture search for guided image super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 9, pp. 8249–8267, 2025. 3, 5, 6, 7

[34] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from rgbd images," in *Proceedings of the European Conference on Computer Vision*, pp. 746–760, 2012. 3, 4, 5

[35] L. He, H. Zhu, F. Li, H. Bai, R. Cong, C. Zhang, C. Lin, M. Liu, and Y. Zhao, "Towards fast and accurate real-world depth super-resolution: Benchmark dataset and baseline," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9229–9238, 2021. 3, 5, 6, 7

[36] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Proceedings of the European Conference on Computer Vision*, pp. 611–625, 2012. 3, 5, 6

[37] I. Vasiljevic, N. Kolkin, S. Zhang, R. Luo, H. Wang, F. Z. Dai, A. F. Daniele, M. Mostajabi, S. Basart, M. R. Walter, *et al.*, "Diode: A dense indoor and outdoor depth dataset," *arXiv preprint arXiv:1908.00463*, 2019. 3, 5, 6

[38] S. Song, S. P. Lichtenberg, and J. Xiao, "Sun rgb-d: A rgb-d scene understanding benchmark suite," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 567–576, 2015. 3, 4, 7

[39] J. Cho, D. Min, Y. Kim, and K. Sohn, "Deep monocular depth estimation leveraging a large-scale outdoor stereo dataset," *Expert Systems with Applications*, vol. 178, p. 114877, 2021. 3

[40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv: Learning*, 2014. 3

[41] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Proceedings of European Conference on Computer Vision*, pp. 611–625, 2012. 4

[42] M. Zhou, K. Yan, J. Pan, W. Ren, Q. Xie, and X. Cao, "Memory-augmented deep unfolding network for guided image super-resolution," *International Journal of Computer Vision*, vol. 131, no. 1, pp. 215–242, 2023. 4, 5, 6, 7

[43] Z. Zhong, X. Liu, J. Jiang, D. Zhao, and X. Ji, "Deep attentional guided image filtering," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2023. 4

[44] J. Wang, Z. Yue, S. Zhou, K. C. Chan, and C. C. Loy, "Exploiting diffusion prior for real-world image super-resolution," *International Journal of Computer Vision*, vol. 132, no. 12, pp. 5929–5949, 2024. 4

[45] Q. Yi, S. Li, R. Wu, L. Sun, Y. Wu, and L. Zhang, "Fine-structure preserved real-world image super-resolution via transfer vae training," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 12415–12426, 2025. 4

[46] X. Deng, C. Zhang, L. Jiang, J. Xia, and M. Xu, "Deepsn-net: Deep semi-smooth newton driven network for blind image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 4, pp. 2632–2646, 2025. 4

[47] J. Dong, J. Pan, J. S. Ren, L. Lin, J. Tang, and M.-H. Yang, "Learning spatially variant linear representation models for joint filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8355–8370, 2022. 5, 6, 7

[48] Z. Zhong, X. Liu, J. Jiang, D. Zhao, Z. Chen, and X. Ji, "High-resolution depth maps imaging via attention-based hierarchical multi-modal fusion," *IEEE Trans. Image Process.*, vol. 31, pp. 648–663, 2022. 5, 6, 7