Supplementary Material for "An Interpretable Depth Map Super-Resolution Method via Unrolling Dual-Boundary Consistency Constrained Optimization"

Lijun Zhao[a,*], Hao Ren[a], Jinjing Zhang[b], Huihui Bai[c], Anhong Wang[a]

[a]*Taiyuan University of Science and Technology, Taiyuan 030051, China*
[b]*North University of China, Jiancaoping District, Taiyuan 030051, China*
[c]*Beijing Jiaotong University, No.3 Shangyuancun Haidian District, Beijing 100044, China*

1. Appendix for Experimental Results and Analysis

1.1. Performance Comparison

To extensively evaluate the generalization capability of the proposed EDC-Net beyond the Middlebury and Lu datasets, we conduct comprehensive experiments on the NYU-v2 dataset. This dataset contains a diverse range of real-world indoor scenes, posing greater challenges for Depth Map Super-Resolution (DSR) due to complex structures and varying lighting conditions. In this evaluation, we perform super-resolution tasks under upsampling factors of $4\times$, $8\times$, and $16\times$. To ensure a rigorous and holistic comparison, we select a wide range of state-of-the-art (SOTA) competitors, covering both non-interpretable deep learning methods (e.g., DCTNet[1], BridgeNet[2], AHMF[3], GeoDSR[4], PMBANet[5]) and interpretable model-based networks (e.g., MADUNet[6], EC-DSRNet[7], DASUNet[8]). Table 1 shows the quantitative comparison results in terms of Root Mean Square Error (RMSE). As observed, our EDC-Net consistently achieves the lowest RMSE across all upsampling factors.

*Corresponding author.
Email address: zlj_ty@163.com (Lijun Zhao)

Table 1: Objective performance comparison of different DSR approaches on NYU-v2 RGB-D dataset in terms of average RMSE ("Inter" indicates whether the networks is interpretable).

Methods	**Inter**	**4×**	**8×**	**16×**
Bicubic	✓	8.16	14.22	22.32
DCTNet[1]	✗	1.59	3.16	5.84
AHMF[3]	✗	1.40	2.89	5.64
BridgeNet[2]	✗	1.54	2.63	4.98
GeoDSR [4]	✗	1.42	2.62	4.86
PMBA-Net[5]	✗	1.06	2.28	4.98
MADUNet[6]	✓	1.51	3.02	6.23
EC-DSRNet[7]	✓	1.05	1.96	3.49
DASUNet [8]	✓	1.44	2.57	4.79
EDC-Net(Ours)	✓	**0.97**	**1.88**	**3.18**

Table 2: FLOPS/PARAS COMPARISON OF DIFFERENT DSR METHODS ON NYU-V2 RGB-D DATASET ($1G = 10^9$, $1M = 10^6$).

Methods	**DKN**[9]	**DMSG**[10]	**DJFR**[11]	**FDSR**[12]	**AHMF**[3]	**HCGNet**[13]	**EC-DSRNet**[7]	**EDC-Net(Ours)**
4×	688.04G/1.16M	92.91G/0.38M	22.91G/0.53M	**18.16G/0.60M**	101.95G/2.54M	3364.19G/22.26M	291.07G/1.20M	279.43G/1.65M
8×	688.04G/1.16M	95.54G/0.52M	22.91G/0.53M	**18.16G/0.64M**	54.94G/3.36M	6455.76G/41.42M	298.43G/1.79M	331.42G/1.31M
16×	688.04G/1.16M	96.29G/0.65M	22.91G/0.53M	**18.16G/0.68M**	90.45G/5.75M	16723.09G/102.38M	549.66G/1.78M	395.68G/2.43M

1.2. The Comparison of Computational Complexity

To evaluate the model complexity, we present a quantitative comparison of FLoating-point OPerations (FLOPs) and the total parameter number for various DSR methods in Table 2. The evaluation is standardized by using an input resolution of 640×480 across $4\times$, $8\times$, and $16\times$ upsampling scales. As compared to interpretable EC-DSRNet[7], our method achieves a lower computational cost at $16\times$ (e.g., 395.68GFLOPs vs. 549.66GFLOPs) and improves reconstruction accuracy, as shown in Table 1, albeit with a slight increase in total parameter number (e.g., 2.43M vs. 1.78M). Compared to unexplainable models such as HCGNet [13] requiring 16723.09GFLOPs and 102.38M Parameters at $16\times$), our method is substantially more lightweight in both aspects. Conversely, when compared to ultra-lightweight methods specifically designed for efficiency, such as FDSR [12] with 18.16GFLOPs and AHMF [3] with 90.45GFLOPs, our method requires higher FLOPs of 395.68G,

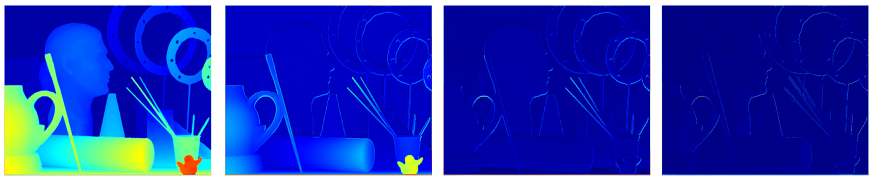

Figure 1: Feature visualizations of the intermediate outputs across unfolding stages ($n = 1, 2, 3, 4$) on the "Art" sample from the Middlebury dataset.

although it maintains a comparable yet smaller total parameter number (e.g., 2.43M vs. AHMF's 5.75M). This increases the computational cost of FLOPs, which is inherently caused by the iterative unfolding architecture and the explicit step-by-step computation of the triple-task branches of color edge, depth edge, and depth map based on the ADMM algorithm. Although our approach may not be the best option for strictly resource-constrained or extreme real-time applications because of its higher FLOPs, it ensures mathematical transparency. The moderate sacrifice in inference speed yields a transparent, mathematically interpretable mechanism and state-of-the-art reconstruction accuracy, as shown in Table 1. The moderate sacrifice in FLOPs results in a transparent, mathematically interpretable mechanism and state-of-the-art reconstruction accuracy, as shown in Table 1. This is a justifiable and highly advantageous trade-off for high-fidelity depth sensing applications.

1.3. Visualization of intermediate features

To further interpret the internal optimization mechanism and the stage-wise behavior of the proposed EDC-Net for DSR, we visualize the intermediate feature outputs across the unfolding stages. Figure 1 displays this feature evolution using the "Art" sample from the Middlebury dataset. Since our network is intrinsically formulated by unrolling the ADMM algorithm, each stage ($n = 1, 2, 3, 4$) mathematically corresponds to a physical iteration step. The visualizations reveal a highly interpretable coarse-to-fine reconstruction process. In the early stage, the network initially focuses on recovering the coarse global structure and the low-frequency depth layout from the low-resolution input. High feature activation are

Table 3: Quantitative ablation study on different combinations of L_1 and L_2 norms for the depth and edge loss functions on the Middleburry RGB-D dataset.

Methods	$L_1 + L_1$	$L_2 + L_2$	$L_2 + L_1$	$L_1 + L_2$ **(EDC-Net(Ours))**
PSNR	48.61	48.36	47.38	**50.48**
SSIM	0.9971	0.9969	0.9966	**0.9973**
RMSE	0.96	1.02	1.11	**0.79**

Table 4: Quantitative evaluation of Boundary Alignment Rate (BAR) on the Middlebury RGB-D dataset.

Methods	**4**×	**8**×	**16**×
CGN[14]	60.39%	59.95%	51.49%
MFR[15]	65.53%	64.85%	57.38%
MIG[16]	62.20%	61.08%	59.66%
RDN [17]	61.38%	60.62%	57.30%
EC-DSRNet [7]	75.13%	71.87%	61.12%
EDC-Net(Ours)	**79.89%**	**74.81%**	**65.84%**

broadly distributed across continuous, flat regions to establish the base high-resolution structure. As the iterative unfolding progresses, the network gradually refines the depth features in homogeneous areas. The activation in these flat background regions are gradually suppressed, indicating that the smooth depth structures have been well reconstructed. Consequently, the network begins to shift its attention towards structural transitions. In the deep unfolding stage, the feature responses become highly sparse and are intensely concentrated solely on the high-frequency details and sharp object boundaries. Driven by the explicit dual-boundary consistency constraint, the unrolled network progressively reconstructs high-fidelity yet high-resolution depth maps with sharp boundaries.

1.4. Quantitative Evaluation of Boundary Alignment

To explicitly validate the effectiveness of our dual-boundary consistency, we introduce a quantitative geometric metric: the Boundary Alignment Rate (BAR). Standard metrics like RMSE evaluate global pixel intensity but fail to strictly measure the geometric coincidence of physical edges. Therefore, we extract single-pixel-width binary edges from the reconstructed depth maps and calculate their exact spatial

overlap with the Ground Truth (GT) boundaries by applying a standard 1-pixel tolerance to account for negligible sub-pixel shifts. As presented in Table 4, we compared our EDC-Net with state-of-the-art methods, including CGN[14], MFR[15], MIG[16], RDN[17], and EC-DSRNet[7]. Our method achieves the highest BAR score across all upsampling scales. This geometric evidence confirms that our method reconstructs depth boundaries that align most accurately with the physical GT.

1.5. Ablation Study on Loss Function Combinations

To validate the rationale behind our loss function design, we conduct an ablation study on the combinations of L_1 and L_2 norms for the depth reconstruction loss and the edge consistency loss. As presented in Table 3, different loss combinations significantly impact the final super-resolution performance. Using the L_2 norm for depth reconstruction (e.g., the L_2+L_1 and L_2+L_2 combinations) inevitably leads to lower PSNR and higher RMSE. This performance drop occurs because the L_2 norm tends to heavily penalize large errors at depth discontinuities, resulting in over-smoothed edges. On the other hand, applying the L_1 norm to the highly sparse edge map, e.g., the L_1+L_1 combination, degrades the structural continuity of the guidance features, yielding sub-optimal results compared to our proposed method. The quantitative results clearly verify that our specific configuration, employing the robust L_1 loss to preserve sharp depth boundaries and the L_2 loss to maintain continuous structural edge guidance, yields the optimal DSR performance across all metrics of PSNR, SSIM, and RMSE.

Acknowledgments

This work was supported by Shanxi Scholarship Council of China (2024-130), Fundamental Research Program of Shanxi Province (202503021211182), National Natural Science Foundation of China (62202323, 62331003), Fundamental Research Program of Shanxi Province (202103021223284), The Shanxi Province Third Batch of Outstanding Doctoral Research Initial Funding in 2022 (98001836) and The First Batch of Doctoral Research Initial Funding in 2023 (110136051).

[1] Z. Zhao, J. Zhang, S. Xu, C. Zhang, J. Liu, Discrete cosine transform network for guided depth map super-resolution, IEEE Conference on Computer Vision and Pattern Recognition (2021) 5687–5697.

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

[3] Z. Zhong, X. Liu, J. Jiang, D. Zhao, Z. Chen, X. Ji, High-resolution depth maps imaging via attention-based hierarchical multi-modal fusion, IEEE Transactions on Image Processing 31 (2021) 648–663.

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

[5] X. Ye, B. Sun, Z. Wang, J. Yang, R. Xu, H. Li, B. Li, PMBANet: Progressive multi-branch aggregation network for scene depth super-resolution, IEEE Transactions on Image Processing 29 (2020) 7427–7442.

[6] M. Zhou, K. Yan, J. Pan, W. Ren, Q. Xie, X. Cao, Memory-augmented deep unfolding network for guided image super-resolution, International Journal of Computer Vision 131 (1) (2023) 215–242.

[7] L. Zhao, J. Zhang, J. Zhang, H. Bai, A. Wang, Joint discontinuity-aware depth map super-resolution via dual-tasks driven unfolding network, IEEE Transactions on Instrumentation and Measurement 73 (2024) 1–14.

[8] J. Zhang, L. Zhao, J. Zhang, A. Wang, H. Bai, Joint deep-unfolding optimization learning for depth map arbitrary-scale super-resolution, IEEE Transactions on Multimedia.

[9] B. Kim, J. Ponce, B. Ham, Deformable kernel networks for joint image filtering, International Journal of Computer Vision 129 (2) (2021) 579–600.

[10] T.-W. Hui, C. C. Loy, X. Tang, Depth map super-resolution by deep multi-scale guidance, in: European conference on computer vision, Springer, 2016, pp. 353–369.

[11] Y. Li, J.-B. Huang, N. Ahuja, M.-H. Yang, Joint image filtering with deep convolutional networks, IEEE transactions on pattern analysis and machine intelligence 41 (8) (2019) 1909–1923.

[12] L. He, H. Zhu, F. Li, H. Bai, R. Cong, C. Zhang, C. Lin, M. Liu, 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, 2021, pp. 9229–9238.

[13] R. Cong, R. Sheng, H. Wu, Y. Guo, Y. Wei, W. Zuo, Y. Zhao, S. Kwong, Learning hierarchical color guidance for depth map super-resolution, IEEE Transactions on Instrumentation and Measurement 73 (2024) 1–13.

[14] Y. Zuo, Y. Fang, P. An, X. Shang, J. Yang, Frequency-dependent depth map enhancement via iterative depth-guided affine transformation and intensity-guided refinement, IEEE Transactions on Multimedia 23 (2020) 772–783.

[15] Y. Zuo, Q. Wu, Y. Fang, P. An, L. Huang, Z. Chen, Multi-scale frequency reconstruction for guided depth map super-resolution via deep residual network, IEEE Transactions on Circuits and Systems for Video Technology 30 (2) (2019) 297–306.

[16] Y. Zuo, H. Wang, Y. Fang, X. Huang, X. Shang, Q. Wu, Mig-net: Multi-scale network alternatively guided by intensity and gradient features for depth map super-resolution, IEEE Transactions on Multimedia 24 (2021) 3506–3519.

[17] Y. Zuo, Y. Fang, Y. Yang, X. Shang, B. Wang, Residual dense network for intensity-guided depth map enhancement, Information Sciences 495 (2019) 52–64.