# Supplementary Material for "Joint Deep-Unfolding Optimization Learning for Depth Map Arbitrary-Scale Super-Resolution"

Jialong Zhang, Lijun Zhao, Jinjing Zhang, Anhong Wang, Huihui Bai

## I. Appendix for experimental results and analysis

### A. More Performance Comparison Between the Proposed Method and Many State-Of-the-Art Methods

**Performance comparison about MAD indicator with fixed-scale DSR methods.** Six images (Art, Books, Dolls, Laundry, Moebius and Reindeer) are selected from the Middlebury RGB-D dataset for the Depth map Super-Resolution (DSR) performance comparison in term of the Mean Absolute Deviation (MAD), as shown in Table I. We retrain SUFT [15] and SGNet [16] in the Graphics Processing Unit (GPU) of NVIDIA GeForce RTX 3090 using the source codes provided by the authors. FDASU-Net$^+$ is obtained by increasing the number of channels on the basis of FDASU-Net. FDASU-Net$^+$ and GeoDSR [17] have similar performance, and their performance far exceeds other methods. To further compare FDASU-Net$^+$ and GeoDSR [17], we list the average MAD of six images, parameters and inference speed (runtime) on Middlebury RGB-D dataset, as shown in Table II. Obviously, under similar total parameter numbers condition, the performance of FDASU-Net$^+$ is better than GeoDSR [17], and the runtime of FDASU-Net$^+$ at $4\times$ and $8\times$ is faster than that of GeoDSR [17]. However, the performance of FDASU-Net$^+$ at $16\times$ is slightly slower than that of GeoDSR [17]. Note that GeoDSR [17], FDASU-Net, FDASU-Net$^+$ and DASU-Net belong to the class of arbitrary-scale DSR method.

**Performance comparison about RMSE indicator with arbitrary-scale DSR methods.** To demonstrate the continuous representation ability during arbitrary-scale up-sampling of DASU-Net, we randomly select three non-integer factors ($3.75\times, 14.60\times$ and $17.05\times$) for evaluation, where $17.05\times$ is the factor that exceeds the scope of our training. As shown in Table III, we present RMSE values that are tested on the NYU-v2, Middlebury, and Lu RGB-D datasets for performance comparison. GeoDSR [17] uses the "AND" operation, which is implemented through element-wise multiplication to modulate color and depth features, but this single fusion strategy is difficult to effectively aggregate dual-modality information. The proposed DASU-Net adopts three fusion strategies, including element-wise maximization, element-wise addition and element-wise maximization, and they are effectively combined, as shown in Section III-D in our paper. In addition, the proposed joint DSR and High-Low Frequency (H-LF) decomposition optimization model effectively enhances the continuous representation ability of the network. Therefore, our method obtains lower RMSE values in three datasets in most cases. In the last three columns of Table III, we also compare the RMSE values of three datasets on average. DASU-Net achieves the most outstanding performance, with improvements of 4.7%, 4.9% and 2.8% compare to GeoDSR [17] on three up-sampling factors respectively.

### B. More Ablation Study

In this section, NYU-v2 and Lu RGB-D datasets are respectively used as training dataset and validation dataset.

**The Discussion about how the proposed method alleviates three challenges.** DSR task still faces three challenges. Firstly, existing CNN-based DSR methods are designed as black-box network architectures. Secondly, few approaches consider utilizing single model to handle arbitrary-scale DSR. Thirdly, due to structural inconsistency between dual-modality, the reconstructed depth map guided by color images always faces texture-copying issues.

1) Analysis of the first challenge. Designing additional H-LF auxiliary branches can contribute to the image reconstruction performance [18]. However, in the field of unfolding networks, there is no method to explore H-LF branch to assist depth map reconstruction yet. Therefore, we formulate the DSR issue as a novel joint optimization model of DSR and H-LF decomposition. Unlike CDCN [18] relying on professional experience to design the interaction between H-LF features, our network embeds the prior structure of the optimization model, which makes the H-LF feature flow in the network more explicit. In summary, compared with previous methods, the proposed joint optimization model can make H-LF feature flows interpretable easier in depth reconstruction networks.

This work was supported by National Natural Science Foundation of China (62202323, 62331003, 62072325), Fundamental Research Program of Shanxi Province (202103021223284, 2310700017MZ), Taiyuan University of Science and Technology Scientific Research Initial Funding (20192023, 20192055), Shanxi Province Science Foundation for Youth (202203021222047), The Shanxi Province Third Batch of Outstanding Doctoral Research Initial Funding in 2022 (98001836), The First Batch of Doctoral Research Initial Funding in 2023 (110136051), Beijing Natural Science Foundation (L223022) and the Special Fund for Science and Technology Innovation Teams of Shanxi Province (202304051001035). (*Corresponding author: Lijun Zhao (zlj_ty@163.com).*)

Jialong Zhang, Lijun Zhao and Anhong Wang are with the Institute of Digital Media and Communication, Taiyuan University of Science and Technology, Taiyuan 030024, China.

Jinjing Zhang is with the Data Science and Technology, North University of China, Jiancaoping District, Taiyuan 030051, China.

Huihui Bai is with Beijing Key Laboratory of Advanced Information Science and Network Technology and also with Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China.

TABLE I
OBJECTIVE PERFORMANCE COMPARISON OF DIFFERENT DSR APPROACHES FOR FIXED-SCALE UPSAMPLING ON SIX IMAGES FROM MIDDLEBURY RGB-D DATASET IN TERM OF THE MAD (THE LOWER,THE BETTER).

| Image Name | Art | | | Books | | | Dolls | | | laundry | | | Moebius | | | Reindeer | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Methods | 4× | 8× | 16× | 4× | 8× | 16× | 4× | 8× | 16× | 4× | 8× | 16× | 4× | 8× | 16× | 4× | 8× | 16× |
| **CLMF** [1] | 0.76 | 1.44 | 2.87 | 0.28 | 0.51 | 1.02 | 0.34 | 0.60 | 1.01 | 0.50 | 0.80 | 1.67 | 0.29 | 0.51 | 0.97 | 0.51 | 0.84 | 1.55 |
| **TGV** [2] | 0.65 | 1.17 | 2.30 | 0.27 | 0.42 | 0.82 | 0.33 | 0.70 | 2.20 | 0.55 | 1.22 | 3.37 | 0.29 | 0.49 | 0.90 | 0.49 | 1.03 | 3.05 |
| **JGU** [3] | 0.47 | 0.78 | 1.54 | 0.24 | 0.43 | 0.81 | 0.33 | 0.59 | 1.06 | 0.36 | 0.64 | 1.20 | 0.25 | 0.46 | 0.80 | 0.38 | 0.64 | 1.09 |
| **CDLLC** [4] | 0.53 | 0.76 | 1.41 | 0.19 | 0.46 | 0.75 | 0.31 | 0.53 | 0.79 | 0.30 | 0.48 | 0.96 | 0.27 | 0.46 | 0.79 | 0.43 | 0.55 | 0.98 |
| **EG** [5] | 0.48 | 0.71 | 1.35 | 0.15 | 0.36 | 0.70 | 0.27 | 0.49 | 0.74 | 0.28 | 0.45 | 0.92 | 0.23 | 0.42 | 0.75 | 0.36 | 0.51 | 0.95 |
| **DEIN** [6] | 0.40 | 0.64 | 1.34 | 0.22 | 0.37 | 0.78 | 0.22 | 0.38 | 0.73 | 0.23 | 0.36 | 0.81 | 0.20 | 0.35 | 0.73 | 0.26 | 0.40 | 0.80 |
| **DSR** [7] | 0.25 | 0.53 | 1.44 | 0.11 | 0.26 | 0.67 | 0.16 | 0.36 | 0.65 | 0.16 | 0.36 | 0.76 | 0.13 | 0.27 | 0.69 | 0.17 | 0.35 | 0.77 |
| **PMBA-Net** [8] | 0.26 | 0.51 | 1.22 | 0.15 | 0.26 | 0.59 | 0.19 | 0.32 | 0.59 | 0.17 | 0.34 | 0.71 | 0.16 | 0.26 | 0.67 | 0.17 | 0.34 | 0.74 |
| **Bridge-Net** [9] | 0.30 | 0.58 | 1.49 | 0.14 | 0.24 | 0.51 | 0.19 | 0.34 | 0.64 | 0.17 | 0.34 | 0.71 | 0.15 | 0.26 | 0.54 | 0.19 | 0.31 | 0.70 |
| **CGN** [10] | 0.23 | 0.47 | 1.05 | 0.14 | 0.26 | 0.56 | 0.21 | 0.35 | 0.74 | 0.19 | 0.37 | 0.77 | 0.15 | 0.27 | 0.58 | 0.24 | 0.36 | 0.83 |
| **MFR** [11] | 0.29 | 0.57 | 1.19 | 0.18 | 0.31 | 0.53 | 0.25 | 0.38 | 0.64 | 0.30 | 0.50 | 0.92 | 0.20 | 0.32 | 0.55 | 0.29 | 0.40 | 0.84 |
| **MIG** [12] | **0.21** | 0.46 | 1.08 | 0.14 | 0.24 | 0.53 | 0.21 | 0.35 | 0.70 | 0.18 | 0.35 | 0.80 | 0.15 | 0.26 | 0.55 | 0.22 | 0.36 | 0.82 |
| **RDN** [13] | 0.23 | 0.51 | 1.07 | 0.16 | 0.28 | 0.53 | 0.23 | 0.39 | 0.66 | 0.20 | 0.41 | 0.76 | 0.16 | 0.29 | 0.55 | 0.22 | 0.38 | 0.80 |
| **RMIG** [14] | 0.23 | 0.46 | 1.17 | 0.16 | 0.24 | 0.63 | 0.23 | 0.36 | 0.76 | 0.20 | 0.36 | 0.97 | 0.17 | 0.26 | 0.57 | 0.23 | 0.35 | 0.88 |
| **SUFT** [15] | 0.33 | 0.53 | 1.01 | 0.20 | 0.25 | 0.42 | 0.28 | 0.37 | 0.60 | 0.23 | 0.35 | 0.61 | 0.19 | 0.27 | 0.48 | 0.27 | 0.36 | 0.62 |
| **SGNet** [16] | 0.33 | 0.49 | 0.88 | 0.20 | 0.26 | 0.38 | 0.30 | 0.37 | 0.57 | 0.24 | 0.34 | 0.55 | 0.20 | 0.26 | 0.43 | 0.28 | 0.35 | 0.55 |
| **GeoDSR** [17] | 0.27 | 0.46 | 0.87 | **0.08** | **0.15** | 0.30 | **0.13** | 0.23 | **0.46** | 0.14 | **0.25** | **0.52** | **0.10** | **0.18** | **0.33** | 0.16 | 0.27 | 0.48 |
| **FDASU-Net** | 0.31 | 0.46 | 0.88 | 0.09 | 0.16 | 0.32 | 0.14 | 0.24 | 0.50 | 0.16 | 0.28 | 0.60 | 0.11 | 0.19 | 0.41 | 0.17 | 0.26 | 0.49 |
| **FDASU-Net$^+$** | 0.28 | **0.42** | **0.81** | **0.08** | 0.16 | **0.29** | **0.13** | **0.22** | 0.49 | **0.13** | 0.27 | 0.57 | **0.10** | **0.17** | 0.39 | **0.15** | **0.25** | **0.45** |
| **DASU-Net** | 0.27 | **0.42** | 0.87 | 0.09 | 0.16 | 0.33 | 0.14 | 0.23 | 0.51 | 0.15 | 0.27 | 0.87 | 0.11 | 0.19 | **0.33** | **0.15** | **0.25** | 0.51 |

The comparison between DASU-Net and Baseline+CDCN [18] is shown in Table IV and Fig. 1. Specifically, the Baseline achieves the worst results. Adding H-LF decomposition (CDCN [18]) to the Baseline can significantly improve the performance of DSR tasks. However, compared to LNet, HNet, and DNet that inherit the prior structure of H-LF decomposition optimization model, it is hard for CDCN [18] to fully utilize the advantages of H-LF decomposition in DSR tasks. The design method of Baseline and Baseline+CDCN [18] can be found in Section IV-C of our paper.

2) Analysis of the second challenge. Arbitrary-scale DSR is more advantageous for real-world scenarios, but currently most DSR methods still focus on fixed and integer scales. In addition, achieving effective arbitrary-scale DSR is not a simple task. We use the fusion part and interpretable part as the cores to enhance the features obtained by the grid sampler for achieving effective arbitrary-scale DSR. The experiments results demonstrate the effectiveness of the proposed method.

3) Analysis of the third challenge. To avoid introducing unnecessary texture details in color images, we design the proposed method from two aspects. Firstly, regarding the overall network structure, to avoid structural inconsistency caused by dual-modality fusion, the structural information of color images is indirectly introduced into three reconstruction sub-networks through the proposed AUF branch, as shown in Fig. 1 (d) in our paper. Secondly, we design an AUF module to effectively reduce the dual-modality gap. Some methods [19], [20] directly feed multi-modal signals into multiple fusion strategies, which makes it difficult for networks to fully leverage the strengths of each strategy. Our method uses DAM [21] to reduce the dual-modality gap and adaptively assigns weights to each fusion strategy to capture more effective fusion features. To confirm the effectiveness of these two methods, we obtain two variants based on DASU-Net. For the first method, we directly add the color features into the interpretable part to obtain the variant DASU-Net$^2$. Specifically, we remove the fusion function of AUF and only retain the up-sampling part. Then, we fuse color features with $L_{k-1}$, $H_{k-1}$ and $D_{k-1}$ by element-wise addition operation, and feed them to LNet, HNet and DNet so as to output $L_k$, $H_k$ and $D_k$. For the second method, we replace it with grid up-sampling, convolution layer and several ResBlocks, to obtain the variant DASU-Net$^3$. The high-frequency information of DASU-Net$^2$ and DASU-Net$^3$ both have significant errors, and DASU-Net$^2$ has significant artifacts in the amplification area, as shown in Fig. 2.

TABLE II
OBJECTIVE COMPARISON OF GEODSR AND FDASU-NET$^+$ ON MIDDLEBURY RGB-D DATASET IN TERMS OF AVERAGE MAD, PARAMETERS (**Paras**, $1M = 10^6$) AND RUNTIME(MS).

| Methods | Paras(M) | Runtime(ms) | MAD | | |
|---|---|---|---|---|---|
| | | | 4× | 8× | 16× |
| **GeoDSR [17]** | 5.52 | 917 | 0.146 | 0.256 | **0.493** |
| **FDASU-Net$^+$** | 5.38 | 310 | **0.145** | **0.248** | 0.500 |

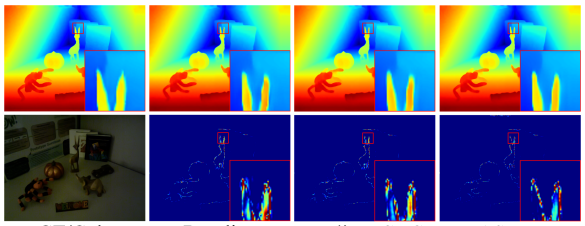


Fig. 1. Comparison of 8× visualization results and error maps of the 1-st image from the Lu RGB-D dataset for Baseline, Baseline+CDCN [59] and DASU-Net (The brighter the area in the error map, the greater the pixel error).

**The Discussion about how one sub-network is influenced by another sub-network in the proposed method.** In our method, the reconstruction performance between DNet, HNet and LNet can be balanced by adjusting the hyper-parameter of the loss function. The loss function of DASU-Net is represented as $Loss_{sum}$=$\tau_0 Loss_D$+$\tau_1 Loss_H$+$\tau_2 Loss_L$, where $Loss_D$,

TABLE III
OBJECTIVE PERFORMANCE COMPARISON OF DIFFERENT DSR APPROACHES ON NON-INTEGER FACTOR ON NYU-V2, MIDDLEBURY AND LU RGB-D DATASET IN TERMS OF RMSE (THE LOWER, THE BETTER).

| Methods | NYU-v2 | | | Middlebury | | | Lu | | | Average | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 3.75× | 14.60× | 17.05× | 3.75× | 14.60× | 17.05× | 3.75× | 14.60× | 17.05× | 3.75× | 14.60× | 17.05× |
| **Bicubic** | 4.07 | 10.90 | 11.93 | 2.13 | 6.00 | 6.52 | 2.35 | 7.20 | 7.96 | 2.85 | 8.03 | 8.80 |
| **GeoDSR** [17] | **1.35** | 4.56 | 5.18 | 1.01 | 2.77 | **3.17** | 0.80 | 3.54 | 4.24 | 1.05 | 3.62 | 4.19 |
| **FDASU-Net** | 1.46 | 4.78 | 5.51 | 0.98 | 2.95 | 3.42 | 0.75 | 3.56 | 4.07 | 1.06 | 3.76 | 4.33 |
| **DASU-Net** | 1.37 | **4.49** | **5.17** | **0.96** | **2.72** | 3.18 | **0.67** | **3.12** | **3.81** | **1.00** | **3.44** | **4.05** |

TABLE IV
OBJECTIVE PERFORMANCE COMPARISON OF MULTIPLE VARIANTS OF DASU-NET ON LU RGB-D DATASET IN TERMS OF THE RMSE, PSNR, SSIM AND TOTAL PARAMETER NUMBERS(**Paras**, $1M = 10^6$).

| Methods | Paras(M) | RMSE | | | PSNR | | | SSIM | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 4× | 8× | 16× | 4× | 8× | 16× | 4× | 8× | 16× |
| **Baseline** | 1.71 | 0.87 | 1.88 | 4.18 | 49.73 | 44.19 | 36.33 | 0.9974 | 0.9925 | 0.9778 |
| **Baseline + CDCN [18]** | 1.72 | 0.86 | 1.64 | 3.92 | 49.78 | 44.16 | 36.96 | 0.9974 | 0.9923 | 0.9797 |
| **DASU-Net** | 1.69 | **0.73** | **1.39** | **3.49** | **51.13** | **45.31** | **37.66** | **0.9977** | **0.9935** | **0.9820** |

TABLE V
THE ABLATION STUDY ON THE HYPER-PARAMETER $\tau_0$, $\tau_1$ AND $\tau_2$ ABOUT THE LOSS FUNCTION.

| $Loss_{sum}$=$\tau_0 Loss_D$+$\tau_1 Loss_H$+$\tau_2 Loss_L$ | RMSE | | | PSNR | | | SSIM | | |
|---|---|---|---|---|---|---|---|---|---|
| | $D_{SR}$ | $H_{SR}$ | $L_{SR}$ | $D_{SR}$ | $H_{SR}$ | $L_{SR}$ | $D_{SR}$ | $H_{SR}$ | $L_{SR}$ |
| $D_1 : \tau_0 = 1, \tau_1 = 0.01, \tau_2 = 0.01$ | **1.39** | 13.46 | 13.21 | **45.31** | 25.58 | 25.74 | **0.9935** | 0.9617 | 0.9868 |
| $D_2 : \tau_0 = 0.01, \tau_1 = 1, \tau_2 = 0.01$ | 1.62 | **3.26** | 6.37 | 43.96 | **37.90** | 32.09 | 0.9917 | **0.9633** | 0.9876 |
| $D_3 : \tau_0 = 0.01, \tau_1 = 0.01, \tau_2 = 1$ | 1.44 | 3.63 | **2.92** | 44.97 | 37.00 | **38.91** | 0.9931 | 0.9631 | **0.9913** |
| $D_4 : \tau_0 = 1, \tau_1 = 0.001, \tau_2 = 0.001$ | **1.39** | 14.95 | 13.80 | **45.01** | 24.62 | 25.69 | **0.9934** | 0.8816 | 0.9869 |
| $D_5 : \tau_0 = 0.001, \tau_1 = 1, \tau_2 = 0.001$ | 1.66 | **3.57** | 7.57 | 43.70 | **37.19** | 30.62 | 0.9912 | **0.9625** | 0.9812 |
| $D_6 : \tau_0 = 0.001, \tau_1 = 0.001, \tau_2 = 1$ | 1.48 | 3.99 | **3.24** | 44.94 | 36.81 | **38.21** | 0.9930 | 0.9919 | **0.9913** |

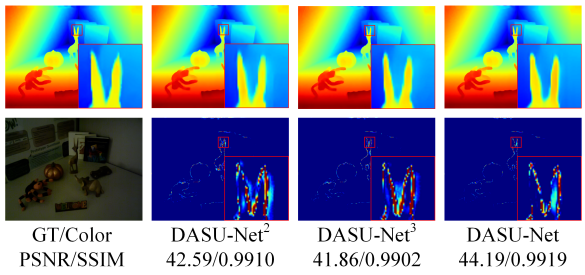


Fig. 2. Comparison of 8× visualization results and error maps of the 1-st image from the Lu RGB-D dataset for DASU-Net$^2$, DASU-Net$^3$ and DASU-Net (The brighter the area in the error map, the greater the pixel error).

$Loss_H$, and $Loss_L$ are respectively used for the regularization of Super-Resolution (SR) depth map $D_{SR}$, SR high-frequency map $H_{SR}$ and SR low-frequency map $L_{SR}$. $\tau_0$, $\tau_1$ and $\tau_2$ are three trade-off factors. In our paper, $\tau_0$ is set to 1. We conduct ablation study on hyper-parameters with six cases: $D_1$, $D_2$ , $D_3$, $D_4$, $D_5$ and $D_6$, as shown in Table V. The experimental record is the result of 8× up-sampling. When the hyper-parameters of the loss function for reconstructing depth maps are much larger than the other two ($\tau_0 > \tau_1$ and $\tau_0 > \tau_2$), the reconstruction quality of depth maps is better than the one ($\tau_0$ is less than $\tau_1$ and $\tau_2$). Specifically, the reconstruction performance of depth map $D_1$ is better than that of $D_2$ and $D_3$, where $\tau_0$ is much greater than $\tau_1$ and $\tau_2$. The performance of the HF map restored by $D_2$ is better than that of $D_1$ and $D_3$, when $\tau_1$ is much greater than $\tau_0$ and $\tau_2$. As for $D_3$, $D_4$, $D_5$ and $D_6$, we can get similar conclusions.

## REFERENCES

[1] J. Lu, K. Shi, D. Min, L. Lin, and M. N. Do, "Cross-based local multipoint filtering," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 430–437.
[2] D. Ferstl, C. Reinbacher, R. Ranftl, M. Ruether, and H. Bischof, "Image guided depth upsampling using anisotropic total generalized variation," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 993–1000.
[3] M.-Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 169–176.
[4] J. Xie, C.-C. Chou, R. Feris, and M.-T. Sun, "Single depth image super resolution and denoising via coupled dictionary learning with local constraints and shock filtering," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, 2014, pp. 1–6.
[5] J. Xie, R. S. Feris, and M.-T. Sun, "Edge-guided single depth image super resolution," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 428–438, 2016.
[6] X. Ye, X. Duan, and H. Li, "Depth super-resolution with deep edge-inference network and edge-guided depth filling," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 1398–1402.
[7] B. Sun, X. Ye, B. Li, H. Li, Z. Wang, and R. Xu, "Learning scene structure guidance via cross-task knowledge transfer for single depth super-resolution," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 7788–7797.
[8] 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.

[9] 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 the 29th ACM International Conference on Multimedia*, 2021, pp. 2148–2157.

[10] Y. Zuo, Y. Fang, P. An, X. Shang, and J. Yang, "Frequency-dependent depth map enhancement via iterative depth-guided affine transformation and intensity-guided refinement," *IEEE Transactions on Multimedia*, vol. 23, pp. 772–783, 2021.

[11] Y. Zuo, Q. Wu, Y. Fang, P. An, L. Huang, and 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*, vol. 30, no. 2, pp. 297–306, 2020.

[12] Y. Zuo, H. Wang, Y. Fang, X. Huang, X. Shang, and Q. Wu, "MIG-Net: Multi-scale network alternatively guided by intensity and gradient features for depth map super-resolution," *IEEE Transactions on Multimedia*, vol. 24, pp. 3506–3519, 2022.

[13] Y. Zuo, Y. Fang, Y. Yang, X. Shang, and B. Wang, "Residual dense network for intensity-guided depth map enhancement," *Information Sciences*, vol. 495, pp. 52–64, 2019.

[14] 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.

[15] 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*, ser. MM '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 3867C3876. [Online]. Available: https://doi.org/10.1145/3503161.3547873

[16] Z. Wang, Z. Yan, and J. Yang, "Sgnet: Structure guided network via gradient-frequency awareness for depth map super-resolution," *arXiv preprint arXiv:2312.05799*, 2023.

[17] 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, no. 3, 2023, pp. 2698–2706.

[18] Y. Wu, F. Li, H. Bai, W. Lin, R. Cong, and Y. Zhao, "Bridging component learning with degradation modelling for blind image super-resolution," *IEEE Transactions on Multimedia*, pp. 1–16, 2022.

[19] T. Zhou, H. Fu, G. Chen, J. Shen, and L. Shao, "Hi-net: Hybrid-fusion network for multi-modal mr image synthesis," *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2772–2781, 2020.

[20] F. Fang, Y. Yao, T. Zhou, G. Xie, and J. Lu, "Self-supervised multi-modal hybrid fusion network for brain tumor segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 11, pp. 5310–5320, 2022.

[21] Z. Zhang, H. Zheng, R. Hong, M. Xu, S. Yan, and M. Wang, "Deep color consistent network for low-light image enhancement," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 1889–1898.