Supplementary Materials for
Degradation-Resistant Unfolding Network for Heterogeneous Image Fusion

Chunming He[1,3,†] , Kai Li[2*], Guoxia Xu[3,4,*] , Yulun Zhang[5] , Runze Hu[6] , Zhenhua Guo[7] , Xiu Li[1,*]
[1]Shenzhen International Graduate School, Tsinghua University, [2]NEC Laboratories America,
[3]Smart Vision, [4]Nanjing University of Posts and Telecommunications,
[5]ETH Zürich, [6]Beijing Institute of Technology, [7]Tianyi Traffic Technology
{chunminghe19990224, li.gml.kai, gxxu.re, yulun100, hrzlpk2015, cszguo}@gmail.com
li.xiu@sz.tsinghua.edu.cn

Contents

A. Methodology

A.1. Derivations of the HIFM solution

We first introduce two auxiliary variables $\mathbf{E}_l$ and $\mathbf{H}_k$, and rewrite HIFM as follows:

$$\min_{\mathbf{Y},\mathbf{E}_l,\mathbf{H}_k} \frac{1}{2}\|\mathbf{Y}-\mathbf{M}\|_2^2+\sum_{l=1}^{L}\lambda_l\psi\left(\mathbf{E}_l\right)+\sum_{k=1}^{K}\mu_k\phi\left(\mathbf{H}_k\right), \quad \text{s.t. } \mathbf{E}_l=\nabla_l\mathbf{Y}, \mathbf{H}_k=\nabla_k(\mathbf{Y}-\mathbf{M}). \tag{1}$$

We then solve Eq. (1) with the alternative direction method of multipliers (ADMM) [17] to eliminate the corresponding constraints and acquire the final solutions. Eq. (1) can be converted into its augmented Lagrangian form by introducing two dual variables $\mathbf{F}_l$ and $\mathbf{G}_k$:

$$\begin{aligned} L(\mathbf{Y},\mathbf{E}_l,\mathbf{H}_k,\mathbf{F}_l,\mathbf{G}_k)=&\frac{1}{2}\|\mathbf{Y}-\mathbf{M}\|_2^2+\sum_{l=1}^{L}\lambda_l\psi\left(\mathbf{E}_l\right)+\sum_{k=1}^{K}\mu_k\phi\left(\mathbf{H}_k\right)\\ &+\sum_{l=1}^{L}\frac{\rho_l}{2}\|\nabla_l\mathbf{Y}-\mathbf{E}_l+\mathbf{F}_l\|_2^2+\sum_{k=1}^{K}\frac{\tau_k}{2}\|\nabla_k\left(\mathbf{Y}-\mathbf{M}\right)-\mathbf{H}_k+\mathbf{G}_k\|_2^2, \end{aligned} \tag{2}$$

where ρ_l and τ_k are penalty weights. Through the variable splitting strategy, the solution of the proposed HIFM is equivalent to solving the following decoupled sub-problems $\{\mathbf{Y},\mathbf{E}_l,\mathbf{H}_k\}$ with the Lagrangian multipliers omitted:

$$\begin{cases} \mathbf{Y}^{(n)}=\arg\min\limits_{\mathbf{Y}} \frac{1}{2}\|\mathbf{Y}-\mathbf{M}\|_2^2+\sum_{l=1}^{L}\frac{\rho_l}{2}\left\|\nabla_l\mathbf{Y}-\mathbf{E}_l^{(n-1)}+\mathbf{F}_l^{(n-1)}\right\|_2^2\\ \qquad +\sum_{k=1}^{K}\frac{\tau_k}{2}\left\|\nabla_k\left(\mathbf{Y}-\mathbf{M}\right)-\mathbf{H}_k^{(n-1)}+\mathbf{G}_k^{(n-1)}\right\|_2^2,\\ \mathbf{E}_l^{(n)}=\arg\min\limits_{\mathbf{E}_l}\lambda_l\psi\left(\mathbf{E}_l\right)+\frac{\rho_l}{2}\left\|\nabla_l\mathbf{Y}^{(n)}-\mathbf{E}_l+\mathbf{F}_l^{(n-1)}\right\|_2^2,\\ \mathbf{H}_k^{(n)}=\arg\min\limits_{\mathbf{H}_k}\mu_k\phi\left(\mathbf{H}_k\right)+\frac{\tau_k}{2}\left\|\nabla_k\left(\mathbf{Y}^{(n)}-\mathbf{M}\right)-\mathbf{H}_k+\mathbf{G}_k^{(n-1)}\right\|_2^2. \end{cases} \tag{3}$$

Then we further alternatively solve the sub-problems $\mathbf{Y},\mathbf{E}_l,\mathbf{H}_k$ in Eq. (3). Note that the dual variables, i.e., $\mathbf{F}_l$ and $\mathbf{G}_k$, will also be updated in their corresponding sub-problems.

Solving sub-problem $\mathbf{Y}$. We use $\mathcal{Y}(\cdot)$ to represent sub-problem $\mathbf{Y}$, whose formulation at the n^{th} stage can be written as follows:

$$\begin{aligned} &\mathcal{Y}\left(\mathbf{E}_l^{(n-1)},\mathbf{F}_l^{(n-1)},\mathbf{H}_k^{(n-1)},\mathbf{G}_k^{(n-1)}\right)\\ &=\frac{1}{2}\|\mathbf{Y}-\mathbf{M}\|_2^2+\sum_{l=1}^{L}\frac{\rho_l}{2}\left\|\nabla_l\mathbf{Y}-\mathbf{E}_l^{(n-1)}+\mathbf{F}_l^{(n-1)}\right\|_2^2\\ &+\sum_{k=1}^{K}\frac{\tau_k}{2}\left\|\nabla_k\left(\mathbf{Y}-\mathbf{M}\right)-\mathbf{H}_k^{(n-1)}+\mathbf{G}_k^{(n-1)}\right\|_2^2. \end{aligned} \tag{4}$$

Then we obtain the partial derivative of sub-problem $\mathbf{Y}$:

$$\begin{aligned} &\partial_{\mathbf{Y}}\mathcal{Y}\left(\mathbf{E}_l^{(n-1)},\mathbf{F}_l^{(n-1)},\mathbf{H}_k^{(n-1)},\mathbf{G}_k^{(n-1)}\right)\\ &=\left(\mathbf{I}+\sum_{l=1}^{L}\rho_l\nabla_l^T\nabla_l+\sum_{k=1}^{K}\tau_k\nabla_k^T\nabla_k\right)\mathbf{Y}-\mathbf{Ens}_{\mathbf{Y}}, \end{aligned} \tag{5}$$

where $\mathbf{I}$ denotes the identity map. $\mathbf{Ens}_{\mathbf{Y}}=\mathbf{M}+\sum_{k=1}^{K}\tau_k\nabla_k^T\left(\nabla_k\mathbf{M}+\mathbf{H}_k^{(n-1)}-\mathbf{G}_k^{(n-1)}\right)+\sum_{l=1}^{L}\rho_l\nabla_l^T\left(\mathbf{E}_l^{(n-1)}-\mathbf{F}_l^{(n-1)}\right)$. Let the partial derivative be equal to zero, we achieve the closed-form solution for $\mathbf{Y}$ at the n^{th} stage:

$$\mathbf{Y}^{(n)}=(\mathbf{I}+\sum_{l=1}^{L}\rho_l\nabla_l^T\nabla_l+\sum_{k=1}^{K}\tau_k\nabla_k^T\nabla_k)^{-1}\mathbf{Ens}_{\mathbf{Y}}. \tag{6}$$

*Corresponding author, † Work done during the internship at Smart Vision.

Metrics	N=5	N=7	N=9 (Ours)	N=11
SSIM	0.80	**0.81**	**0.81**	**0.81**
PSNR	18.06	18.47	**18.73**	18.72
AG	7.10	7.23	**7.47**	7.39
FMI	0.92	0.92	0.93	**0.94**

(a) Different stage numbers of DeRUN.

Metrics	B=3,7	B=3,5 (Ours)	B=5,7	B=3,5,7
SSIM	**0.81**	**0.81**	0.80	**0.81**
PSNR	18.63	**18.73**	18.56	**18.73**
AG	7.37	7.47	7.29	**7.49**
FMI	**0.93**	**0.93**	0.92	**0.93**

(b) Different combinations of LGDE loss.

	I=4	I=8 (Ours)	I=16	I=32
SSIM	**0.81**	**0.81**	**0.81**	0.80
PSNR	18.43	**18.73**	18.62	18.57
AG	7.20	**7.47**	7.31	7.16
FMI	0.92	**0.93**	**0.93**	0.92

(c) Different numbers of quantified gradient directions.

Table 1: Ablation study in the IVF task on the M^3FD dataset. The best results are marked in **bold**.

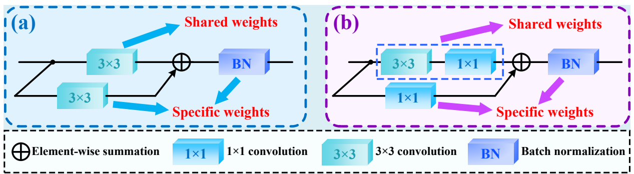

Figure 1: Different feature extractors with the same channel number, where (a) is a common symmetrical feature extractor for domain interaction and (b) is the proposed PWSM.

Solving sub-problem $\mathbf{E}_l$. Following [16], there exist closed-form solution for Sub-problem $\mathbf{E}_l$ at the n^{th} stage:

$$\mathbf{E}_l^{(n)} = S_{\lambda_l/\rho_l}\left(\nabla_l \mathbf{Y}^{(n)} + \mathbf{F}_l^{(n-1)}; \{\theta_{l,i}\}_{i=1}^{I_t}\right), \quad (7)$$

where $S_{\lambda_l/\rho_l}(\cdot)$ is a nonlinear map following [17], which is constrained by the predefined value map $\{\theta_{l,i}\}_{i=1}^{I_t}$ with $I_t = 101$ with the following definition (the value map is learnable in our DeRUN):

$$S_{\frac{\lambda_l}{\rho_l}}(a; \{\theta_{l,i}\}_{i=1}^{I_t}) = \begin{cases} a + \theta_{l,1}, & a < -\lambda_l/\rho_l \\ a - \theta_{l,I_t}, & a > \lambda_l/\rho_l \\ \frac{\Delta\mathcal{V}\theta_{l,(i+1)} + (\mathcal{V}-\Delta\mathcal{V})\theta_{l,i}}{\mathcal{V}}, & |a| \leq \lambda_l/\rho_l \end{cases} \quad (8)$$

where $\mathcal{V}$ is for step adjustment and is set as 0.02 following [17]. $i = \lfloor (a + \lambda_l/\rho_l)/\mathcal{V} \rfloor$. $\{\theta_{l,i}\}_{i=1}^{I_t}$ is the predefined value map to the corresponding points within $[-\lambda_l/\rho_l, \lambda_l/\rho_l]$ at a interval of $\mathcal{V}$ for noise removal.

Given the optimized $\mathbf{E}_l^{(n)}$, the dual variable $\mathbf{F}_l^{(n)}$ can be updated by dual ascent strategy:

$$\mathbf{F}_l^{(n)} = \mathbf{F}_l^{(n-1)} + \varphi_l\left(\nabla_l \mathbf{Y}^{(n)} - \mathbf{E}_l^{(n)}\right), \quad (9)$$

where φ_l is a fixed parameter for the step size.

Solving sub-problem $\mathbf{H}_k$. Owing to the non-convex property of sub-problem $\mathbf{H}_k$, we apply the gradient descent strategy to update $\mathbf{H}_k$ [2]. To begin with, we formulate sub-problem $\mathbf{H}_k$ at the n^{th} stage with function $\mathcal{H}(\cdot)$:

$$\mathcal{H}\left(\mathbf{G}_k^{(n-1)}, \mathbf{Y}^{(n)}\right) = \frac{\tau_k}{2}\left\|\nabla_k\left(\mathbf{Y}^{(n)} - \mathbf{M}\right) - \mathbf{H}_k + \mathbf{G}_k^{(n-1)}\right\|_2^2 + \mu_k \phi(\mathbf{H}_k), \quad (10)$$

where $\phi(\mathbf{a}) = \sum_i \log\left(1 + \theta \mathbf{a}_i^2\right)$, $\mathbf{a}_i$ denotes the i^{th} element of $\mathbf{a}$ and θ is the sparsity controlled parameter for the salient texture information [12]. Then we update $\mathbf{H}_k^{(n)}$ with gradient descent strategy [12]:

$$\mathbf{H}_k^{(n)} = \mathbf{H}_k^{(n-1)} - \sigma_k\left(\partial_{\mathbf{H}_k}\mathcal{H}\left(\mathbf{G}_k^{(n-1)}, \mathbf{Y}^{(n)}\right)\right), \quad (11)$$

where σ_k is the parameter that controls step size.

Having achieved $\mathbf{H}_k^{(n)}$, we further optimize $\mathbf{G}_k^{(n)}$ from the perspective of dual ascent:

$$\begin{aligned} \mathbf{G}_k^{(n)} &= \mathcal{G}\left(\mathbf{G}_k^{(n-1)}, \mathbf{Y}^{(n)}, \mathbf{H}_k^{(n)}\right), \\ &= \mathbf{G}_k^{(n-1)} + \omega_k\left(\nabla_k\left(\mathbf{Y}^{(n)} - \mathbf{M}\right) - \mathbf{H}_k^{(n)}\right), \end{aligned} \quad (12)$$

where ω_k is a controlled parameter for the step size.

B. Experiment

B.1. Ablation Study and Analysis

We evaluate the effect of DeRUN with four metrics, *i.e.*, SSIM, PSNR, AG, and FMI, on M^3FD, including the effect of PWSM and parameter analyses for the stage number N, and the hyper-parameters I and B in the LGDE loss.

Parameter analysis of stage number N. To ensure the appropriate stage number, we set the stage number N to be 5, 7, 9, and 11, and test the fusion performance. As shown in Tab. 1, to keep the trade-off between fusion performance and inference time, we assign N to be 9. Note that even $N = 7$ can outperform the SOTA techniques. **Parameter analysis of L_{lgde}.** In this section, we will analyze two significant hyper-parameters of our proposed local gradient directional entropy (LGDE) loss L_{lgde}, i.e., the number of the quantified gradient directions I and the size of the local block B. As for I, we set I as 8 because most pixels are surrounded by 8 adjacent pixels and we want to characterize the effect of the surrounding 8 pixels of the middle pixel on the gradient direction. As shown in Tab. 1, the performance of $I = 8$ is higher than those of $I = 4$ and $I = 16$, which arises from the fact that a too-small I leads to the weak representational capacity of the gradient direction, while a too large I results in the too strict requirement of the group property, i.e., tougher judgment on gradient direction consistency. Furthermore, Tab. 1 verifies our claim mentioned in the manuscript that LGDE is sensitive to the block size B, where a larger size can ignore some detailed

Figure 2: Failure cases of DeRUN. We simulate misty fog and heavy fog on the visible image (b) based on its depth map (a), where DeRUN fails to enhance part of the salient texture information under the scenario with heavy fog.

texture and a smaller size can suppress the diversity of entropy. To accommodate performance and efficiency, we select the LGDE with the block size of $B = 3, 5$ to jointly extract the texture information in multi-scale.

Effect of PWSM. To demonstrate the advancement of partial weight sharing module (PWSM) in feature extraction, we compare PWSM with a fixed extractor, i.e., Canny operator [1], a Siamesed extractor (from [17]), and a symmetrical extractor, i.e., (a) in Fig. 1, with the same channel number. Note that the symmetrical extractor shares the same number of parameters as our PWSM. As presented in Tab. 1, the best performance illustrates the superiority of the proposed PWSM as a feature extractor.

B.2. Failure Cases and Future Works

In Fig. 2, we simulate misty fog and heavy fog on the visible image (b) based on its depth map (a) following [18, 8]. As shown in Fig. 2, the proposed DeRUN can preserve the detailed component and enhance the salient texture information when the visible image has a clean background or is even degraded with misty fog. However, DeRUN fails to enhance part of the salient texture information under the scenario with uneven and heavy fog, which is mainly due to the fact that the existing components of DeRUN do not accommodate severe and uneven degradation. Therefore, we will consider proposing targeted solutions for realistic degradation scenarios, e.g., bi-level optimization [9], to generate fused results that cater to more downstream tasks, such as semantic segmentation [19, 15, 13].

Additionally, we will consider using other self-excavation techniques to mine the valuable information from a grouping perspective [20, 5] or incorporating more powerful architectures, e.g., dynamic networks [4, 3], transformer [14], and diffusion model [10], with more strategic pretrain networks, such as SimVTP [11]. Furthermore, it would be desirable to employ image quality assessment techniques [6, 7] to generate visual-friendly fusion results.

References

[1] John Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-8(6):679–698, 1986. 3

[2] Xueyang Fu, Xi Wang, Aiping Liu, Junwei Han, and Zheng-Jun Zha. Learning dual priors for jpeg compression artifacts removal. In *ICCV*, pages 4086–4095, 2021. 2

[3] Yizeng Han, Yifan Pu, Zihang Lai, Chaofei Wang, Shiji Song, Junfeng Cao, Wenhui Huang, Chao Deng, and Gao Huang. Learning to weight samples for dynamic early-exiting networks. In *European Conference on Computer Vision*, 2022. 3

[4] Yizeng Han, Zhihang Yuan, Yifan Pu, Chenhao Xue, Shiji Song, Guangyu Sun, and Gao Huang. Latency-aware spatial-wise dynamic networks. In *Advances in Neural Information Processing Systems*, 2022. 3

[5] Chunming He, Kai Li, Yachao Zhang, Guoxia Xu, Longxiang Tang, Yulun Zhang, Zhenhua Guo, and Xiu Li. Weakly-supervised concealed object segmentation with sam-based pseudo labeling and multi-scale feature grouping. *arXiv preprint arXiv:2305.11003*, 2023. 3

[6] Runze Hu, Yutao Liu, Ke Gu, Xiongkuo Min, and Guangtao Zhai. Toward a no-reference quality metric for camera-captured images. *IEEE T. Cybernetics*, 2021. 3

[7] Runze Hu, Yutao Liu, Zhanyu Wang, and Xiu Li. Blind quality assessment of night-time image. *Displays*, 69:102045, 2021. 3

[8] Mingye Ju, Can Ding, Charles A Guo, Wenqi Ren, and Dacheng Tao. Idrlp: image dehazing using region line prior. *IEEE Trans. Image Process.*, 30:9043–9057, 2021. 3

[9] Risheng Liu, Jiaxin Gao, Jin Zhang, Deyu Meng, and Zhouchen Lin. Investigating bi-level optimization for learning and vision from a unified perspective: A survey and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021. 3

[10] Yue Ma, Yingqing He, Xiaodong Cun, Xintao Wang, Ying Shan, Xiu Li, and Qifeng Chen. Follow your pose: Pose-guided text-to-video generation using pose-free videos. *arXiv preprint arXiv:2304.01186*, 2023. 3

[11] Yue Ma, Tianyu Yang, Yin Shan, and Xiu Li. Simvtp: Simple video text pre-training with masked autoencoders. *arXiv preprint arXiv:2212.03490*, 2022. 3

[12] Stefan Roth and Michael J Black. Fields of experts. *Int. J. Comput. Vis.*, 82(2):205–229, 2009. 2

[13] Longxiang Tang, Kai Li, Chunming He, Yulun Zhang, and Xiu Li. Source-free domain adaptive fundus image segmentation with class-balanced mean teacher. *arXiv preprint arXiv:2307.09973*, 2023. 3

[14] Rui Yang, Hailong Ma, Jie Wu, Yansong Tang, Xuefeng Xiao, Min Zheng, and Xiu Li. Scalablevit: Rethinking the context-oriented generalization of vision transformer. In *European Conference on Computer Vision*, pages 480–496. Springer, 2022. 3

[15] Rui Yang, Lin Song, Yixiao Ge, and Xiu Li. Boxsnake: Polygonal instance segmentation with box supervision. *arXiv preprint arXiv:2303.11630*, 2023. 3

[16] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Admm-csnet: A deep learning approach for image compressive sensing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(3):521–538, 2018. 2

[17] Yan Yang, Jian Sun, Huibin Li, Zongben Xu, et al. Deep admm-net for compressive sensing mri. *NIPS*, 29, 2016. 1, 2, 3

[18] Tian Ye, Yunchen Zhang, Mingchao Jiang, Liang Chen, Yun Liu, Sixiang Chen, and Erkang Chen. Perceiving and modeling density for image dehazing. In *European Conference on Computer Vision*, pages 130–145. Springer, 2022. 3

[19] Yachao Zhang, Yanyun Qu, Yuan Xie, Zonghao Li, Shanshan Zheng, and Cuihua Li. Perturbed self-distillation: Weakly supervised large-scale point cloud semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15520–15528, 2021. 3

[20] Yachao Zhang, Yuan Xie, Cuihua Li, Zongze Wu, and Yanyun Qu. Learning all-in collaborative multiview binary representation for clustering. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–14, 2022. 3