

Supplementary Materials for "DeepGSR: Deep group-based sparse representation network for solving image inverse problems"

Ke Jiang, Xinya Ji, Baoshun Shi, *Senior Member, IEEE*

This document is organized as follows. We first report the experimental results in Sec. I, and then provide the detailed proofs in Sec. II.

I. EXPERIMENTAL RESULTS

A. Image denoising

In the proposed GSR-Net denoising network, the core component DeepGSR is seamlessly integrated into the SwinIR [1] framework. SwinIR is adopted as the backbone due to its proven effectiveness in image restoration tasks, offering a strong balance between local representation and non-local contextual modeling. This makes it a suitable foundation for embedding a deep group-based sparse representation mechanism. The network begins with a series of convolutional layers for initial feature extraction, which are followed by residual connections to enhance training stability. The core of the architecture consists of N feature extraction modules, each containing our proposed deep group-based sparse representation layer, termed DeepGSR-L. Structurally analogous to a Swin Transformer block, DeepGSR-L replaces the original self-attention mechanism with a novel sparse representation-based design to improve both interpretability and representation ability. Specifically, each DeepGSR-L module adopts a two-stage processing strategy. The input features are first processed by a lightweight variant, DeepGSR-B, which omits the shifting operation and is tailored for efficient local modeling and sparse representation of texture details. The output is then fed into the full DeepGSR module, which introduces a wavelet-domain shifting-based patch partitioning scheme, along with adaptive patch matching, learned low-rank shrinkage, adaptive aggregation, and shifting wavelet-domain recomposition. This enables the model to effectively capture global structures and exploit non-local self-similarities across the image. Both DeepGSR-B and DeepGSR are equipped with layer normalization and feed-forward networks to refine the feature distribution. The outputs are integrated via residual connections to ensure smooth information flow and network stability. Finally, the aggregated features are further refined through multiple convolutional layers to generate high-quality denoised outputs.

This work was supported by the National Natural Science Foundation of China under Grants 62371414, by the Hebei Natural Science Foundation under Grant F2023203043, by the Hebei Key Laboratory Project under Grants 202250701010046, and by the Postgraduate Innovation Fund Project of Hebei Province under Grant No. CXZZBS2025064. (Corresponding author: Baoshun Shi.)

Ke Jiang, Xinya Ji and Baoshun Shi are with the School of Information Science and Engineering, Yanshan University, Qinhuang Dao 066004, China, and also with the Hebei Key Laboratory of Information Transmission and Signal Processing, Qinhuang Dao 066004, China (e-mail: shibaoshun@ysu.edu.cn).

B. Single image deraining

TABLE I
THE AVERAGE RESULTS OF PSNR (dB) AND SSIM VALUES BY DIFFERENT ALGORITHMS ON RAIN100L AND RAIN200L DATASETS. BEST RESULTS ARE HIGHLIGHTED IN RED, AND THE SECOND-BEST RESULTS ARE MARKED IN BLUE.

Method	Rain100L		Rain200L	
	PSNR	SSIM	PSNR	SSIM
JCAS [2]	28.54	0.8523	29.87	0.8771
RESCAN [3]	38.06	0.9764	36.62	0.9764
PReNet [4]	37.42	0.9774	36.59	0.9749
MSPFN [5]	37.16	0.9783	37.58	0.9789
RCDNet [6]	38.60	0.9520	38.21	0.9848
DGUNet [7]	37.42	0.9694	38.18	0.9832
ECNet [8]	38.19	0.9821	37.88	0.9809
MSResNet-TRNR [9]	38.16	0.9839	38.02	0.9823
HCT-FFN [10]	38.89	0.9769	38.91	0.9777
OptNet [11]	39.45	0.9851	39.51	0.9859
DeepGSR-Derain	40.08	0.9820	40.57	0.9845

To evaluate the effectiveness of our proposed framework in the task of single image deraining, we adopt the experimental protocol established in OptNet [11]. Specifically, the proposed DeepGSR is integrated into the multi-channel channel attention Transformer of each iteration in OptNet, resulting in a new model referred to as DeepGSR-Derain throughout the following comparisons. We compare DeepGSR-Derain against a diverse set of state-of-the-art deraining methods, including a prior-based approach known as JCAS [2], several CNN-based models such as RESCAN [3], PReNet [4], MSPFN [5], RCDNet [6], DGUNet [7], ECNet [8], and MSResNet-TRNR [9], as well as a recently developed Transformer-based method named HCT-FFN [10]. To ensure a fair comparison, all models are retrained when official pre-trained weights are unavailable, using publicly released code and recommended training settings. We conduct evaluations under the matched rain type setting, where the types of rain in the testing images are consistent with those in the training data. Specifically, we adopt two widely used benchmark datasets, Rain100L and Rain200L, both of which contain synthetic images with single rain streak degradations of varying intensity and scale. As summarized in Table I, DeepGSR-Derain consistently achieves the best quantitative results across both datasets, outperforming all existing methods in terms of peak signal-to-noise ratio and structural similarity index. Notably, DeepGSR-Derain yields superior performance compared to strong baselines such as RCDNet and HCT-FFN, thereby demonstrating the efficacy of integrating deep group-based sparse representation into the deraining framework.

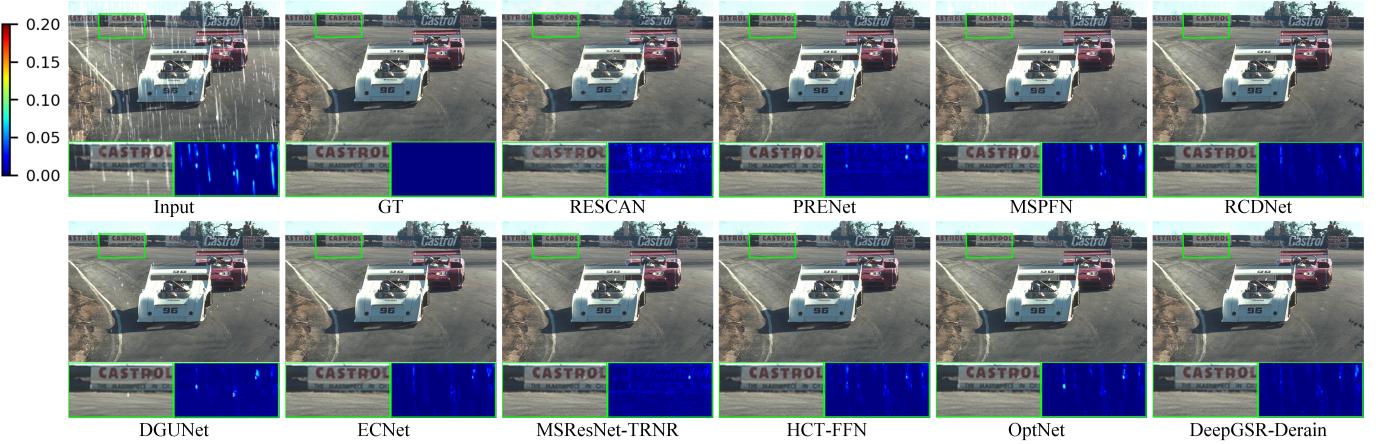


Fig. 1. Visual comparison of various SID methods under a rain streak intensity of Rain100L dataset. Local zoom-in views and the corresponding difference maps are provided to highlight the detail restoration abilities.

To further evaluate the visual quality of restored images, we provide qualitative comparisons in Fig. 1. As illustrated, prior-based approaches such as JCAS often fail to remove noticeable rain streaks and tend to retain residual degradation artifacts. Although convolutional neural network based methods show improved deraining capability, they frequently oversmooth image content and suppress fine-grained textures, particularly in scenes containing high-frequency components or detailed structures. In contrast, DeepGSR-Derain exhibits superior visual performance by effectively eliminating multi-scale rain streaks while maintaining structural integrity and preserving intricate texture information. This improvement can be attributed to the integration of local sparsity and non-local self-similarity modeling, together with the adaptively learned low-rank regularization within our network.

C. Metal artifact reduction

TABLE II

QUANTITATIVE EVALUATIONS FOR DIFFERENT MAR METHODS ON THE DEEPLESION DATASET UNDER THE FULL-VIEW CONDITION. WE REPORT THE AVERAGE PSNR (DB) AND SSIM VALUES ON THE TEST SET FOR EACH METHOD. THE BEST RESULTS ARE HIGHLIGHTED IN RED, AND THE SECOND-BEST RESULTS ARE MARKED IN BLUE.

Method	Average	
	PSNR	SSIM
Input	26.98	0.4311
LI [12]	35.40	0.8377
NMAR [13]	36.36	0.8758
CNNMAR [14]	36.90	0.8901
DSCMAR [15]	37.53	0.9596
DuDoNet [16]	39.68	0.9776
DICDNet [17]	44.61	0.9879
ACDNet [18]	44.75	0.9884
InDuDoNet [19]	44.87	0.9867
DeepGSR-MAR	45.03	0.9881

To further validate the generality and effectiveness of the proposed DeepGSR module, we integrate it into the well-established InDuDoNet [19] framework, resulting in a novel

MAR solution referred to as DeepGSR-MAR. InDuDoNet is a dual-domain architecture originally developed for MAR, leveraging both image-domain and sinogram-domain priors for artifact suppression. We embed the DeepGSR module into the image-domain proximal sub-network of InDuDoNet, replacing its fixed regularization components with a learnable and interpretable sparse modeling mechanism. As shown in Table II, the resulting DeepGSR-MAR consistently surpasses not only its baseline InDuDoNet but also other recent state-of-the-art MAR methods under the full-view setting on the DeepLesion dataset. This performance gain is attributed to the synergy between the wavelet-domain patch processing and the adaptive low-rank modeling capabilities of DeepGSR, which together provide a more expressive and context-aware image representation. The superior results further highlight the robustness and flexibility of the proposed DeepGSR framework as a plug-and-play module for complex clinical inverse problems, offering a promising direction for medical image reconstruction.

To comprehensively evaluate the effectiveness of different MAR methods on real clinical images, we perform a visual comparison on the DeepLesion dataset. As shown in Fig. 2, we present the outputs of several representative MAR approaches on the same CT slice, along with corresponding local zoom-in views and difference maps to highlight their capabilities in detail restoration. It can be observed that traditional interpolation-based method LI and early model-based algorithm NMAR are able to partially suppress metal artifacts, yet often at the cost of blurred edges and structural distortions. In contrast, learning-based methods such as CNNMAR, DuDoNet, and its enhanced variant InDuDoNet+ achieve more effective artifact removal while better preserving anatomical structures. Notably, DeepGSR-MAR demonstrates superior performance in both artifact suppression and detail recovery. This is evidenced by its smoother difference maps with lower residuals in the artifact-affected regions, indicating stronger fidelity and generalization capabilities.

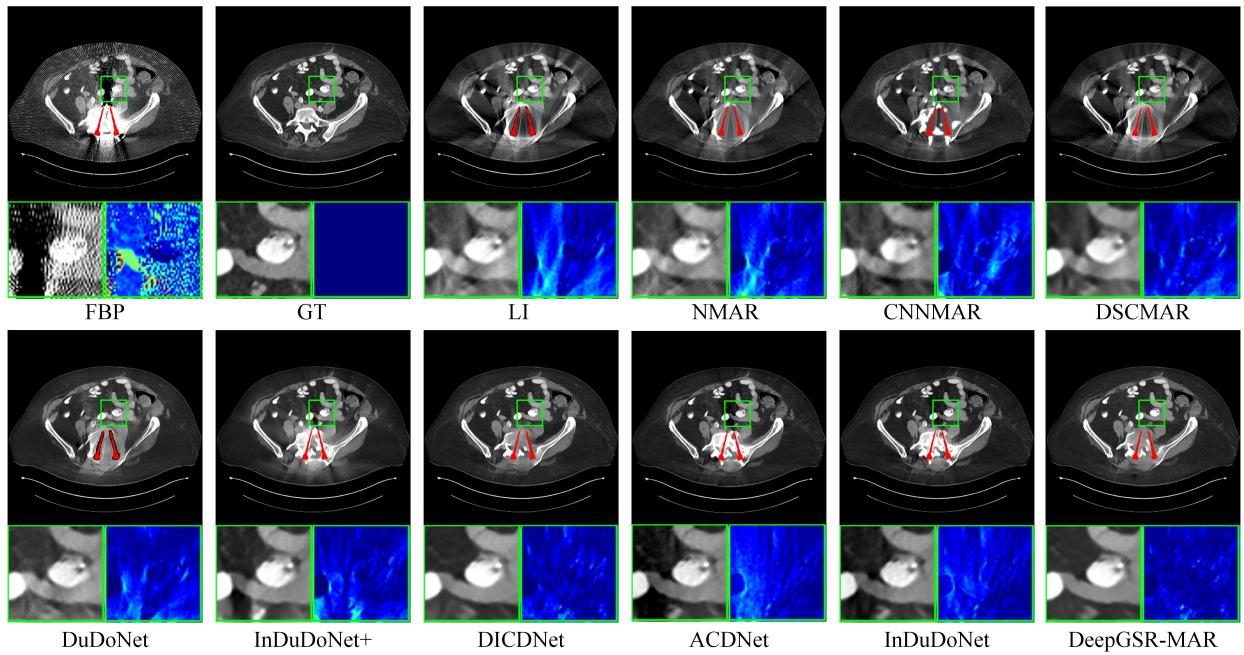


Fig. 2. Visual comparison of various MAR methods on the DeepLesion dataset. Local zoom-in views and the corresponding difference maps are provided to highlight the detail restoration abilities.

II. THE PROOFS

The traditional GSR model aims to adaptively obtain the sparse representation of natural images at the group level \mathbf{x}_{G_k} . The group sparsity model enforces that similar patches within \mathbf{x}_{G_k} share the same dictionary atoms in their sparse representations. In contrast, the low-rank model seeks a low-rank approximation of \mathbf{x}_{G_k} to obtain a robust estimate. Although these three models appear distinct due to their differing theoretical motivations, they exhibit underlying connections [20], [21]. Specifically, by employing a particular dictionary learning strategy based on SVD, a variant of the group sparsity model and a form of the low-rank model can both be derived from the traditional GSR formulation. This reveals that the three models are, to some extent, equivalent, offering a unified perspective that facilitates a deeper understanding of their interrelationships. Based on this observation, we summarize the following theorem:

Theorem 1 (Equivalence and unification of low-rank, group sparsity, and SVD models): For each patch group \mathbf{x}_{G_k} , given its noisy estimate \mathbf{r}_{G_k} , the associated dictionary \mathbf{D}_{G_k} , and its sparse coefficient matrix $\boldsymbol{\alpha}_{G_k}$, the classical GSR model seeks to recover the clean group via the following optimization:

$$\hat{\boldsymbol{\alpha}}_{G_k} = \arg \min_{\boldsymbol{\alpha}_{G_k}} \frac{1}{2} \|\mathbf{r}_{G_k} - \mathbf{D}_{G_k} \boldsymbol{\alpha}_{G_k}\|_F^2 + \lambda \|\boldsymbol{\alpha}_{G_k}\|_0 \quad (1)$$

with the reconstructed group $\hat{\mathbf{x}}_{G_k} = \mathbf{D}_{G_k} \hat{\boldsymbol{\alpha}}_{G_k}$. Note that, ℓ_0 -norm is exploited to measure the real sparsity of $\boldsymbol{\alpha}_{G_k}$ in the group domain in order to enhance the image restoration quality. Assume that each image group can be expressed globally as $\mathbf{x}_{G_k} = \mathbf{D}\mathbf{A}$, where \mathbf{D} is the dictionary to sparsely represent all the patches in \mathbf{x}_{G_k} and \mathbf{A} denotes the coefficient

matrix shared across all patches. Then, the original GSR problem equivalently transforms into:

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \frac{1}{2} \|\mathbf{r}_{G_k} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_{l_0,\infty} \quad (2)$$

where $\|\cdot\|_{l_0,\infty}$ denotes the number of the nonzero rows of a matrix and is a pseudo norm, reflecting structured group-level sparsity. The reconstructed group $\hat{\mathbf{x}}_{G_k}$ is then given by $\mathbf{D}\hat{\mathbf{A}}$. According to the definition of the group sparsity model, Eq. 2 can be regarded as a specific instance that employs the $\ell_{0,\infty}$ norm constraint, in contrast to earlier formulations that typically imposed constraints on the entire matrix norm.

Similarly, let $\gamma_{\mathbf{x}_{G_k}}$ denote the vector consisting of all singular values of \mathbf{x}_{G_k} , i.e., $\gamma_{\mathbf{x}_{G_k}} = [\gamma_{\mathbf{x}_{G_k} \otimes 1}; \gamma_{\mathbf{x}_{G_k} \otimes 2}; \dots; \gamma_{\mathbf{x}_{G_k} \otimes m}]$. Given that $\mathbf{x}_{G_k} = \mathbf{D}_{G_k} \boldsymbol{\alpha}_{G_k}$ and based on the definition of \mathbf{D}_{G_k} , the following equivalence holds:

$$\|\gamma_{\mathbf{x}_{G_k}}\|_0 = \text{rank}(\mathbf{x}_{G_k}) = \|\boldsymbol{\alpha}_{G_k}\|_{l_0,\infty} \quad (3)$$

where $\text{rank}(\cdot)$ represents the rank of a matrix. Consequently, the unified optimization form, explicitly linking group sparsity, low rank, and SVD, is expressed by:

$$\hat{\boldsymbol{\alpha}}_{G_k} = \arg \min_{\boldsymbol{\alpha}_{G_k}} \frac{1}{2} \|\mathbf{r}_{G_k} - \mathbf{D}_{G_k} \boldsymbol{\alpha}_{G_k}\|_F^2 + \lambda \|\gamma_{\mathbf{x}_{G_k}}\|_0. \quad (4)$$

This theorem establishes the fundamental equivalence among the traditional GSR model, the group sparsity model, and the low-rank model, providing a comprehensive unified perspective for robust sparse image representation.

REFERENCES

- [1] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte, SwinIR: Image restoration using swin transformer, in: IEEE/CVF International Conference on Computer Vision Workshops, 2021, pp. 1833–1844.

- [2] S. Gu, D. Meng, W. Zuo, L. Zhang, Joint convolutional analysis and synthesis sparse representation for single image layer separation, in: IEEE International Conference on Computer Vision, 2017, pp. 1717–1725.
- [3] X. Li, J. Wu, Z. Lin, H. Liu, H. Zha, Recurrent squeeze-and-excitation context aggregation net for single image deraining, in: the European Conference on Computer Vision, 2018.
- [4] D. Ren, W. Zuo, Q. Hu, P. Zhu, D. Meng, Progressive image deraining networks: A better and simpler baseline, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3932–3941.
- [5] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, J. Jiang, Multi-scale progressive fusion network for single image deraining, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 8343–8352.
- [6] H. Wang, Q. Xie, Q. Zhao, Y. Li, Y. Liang, Y. Zheng, D. Meng, RCDNet: An interpretable rain convolutional dictionary network for single image deraining, *IEEE Transactions on Neural Networks and Learning Systems* 35 (6) (2024) 8668–8682.
- [7] C. Mou, Q. Wang, J. Zhang, Deep generalized unfolding networks for image restoration, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 17378–17389.
- [8] Y. Li, Y. Monno, M. Okutomi, Single image deraining network with rain embedding consistency and layered lstm, in: IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 3957–3966.
- [9] W. Ran, B. Yang, P. Ma, H. Lu, TRNR: Task-driven image rain and noise removal with a few images based on patch analysis, *IEEE Transactions on Image Processing* 32 (2023) 721–736.
- [10] X. Chen, J. Pan, J. Lu, Z. Fan, H. Li, Hybrid CNN-transformer feature fusion for single image deraining, *Proceedings of the AAAI Conference on Artificial Intelligence* 37 (2023) 378–386.
- [11] K. Jiang, Y. Zhao, B. Shi, OptNet: Optimization-inspired network beyond deep unfolding for structural artifact reduction, *Knowledge-Based Systems* 315 (2025) 113235.
- [12] E. Klotz, W. A. Kalender, R. Sokiransky, D. Felsenberg, Reduction of CT artifacts caused by metallic implants., *Radiology* 164 (1987) 576–577.
- [13] E. Meyer, F. Bergner, R. Raupach, T. Flohr, M. Kachelrieß, Normalized metal artifact reduction (NMAR) in computed tomography, in: IEEE Nuclear Science Symposium Conference Record, 2009, pp. 3251–3255.
- [14] Y. Zhang, H. Yu, Convolutional neural network based metal artifact reduction in x-ray computed tomography, *IEEE Transactions on Medical Imaging* 37 (6) (2018) 1370–1381.
- [15] L. Yu, Z. Zhang, X. Li, L. Xing, Deep sinogram completion with image prior for metal artifact reduction in CT images, *IEEE Transactions on Medical Imaging* 40 (1) (2021) 228–238.
- [16] H. Wang, Y. Li, H. Zhang, D. Meng, Y. Zheng, Indudonet+: A deep unfolding dual domain network for metal artifact reduction in CT images, *Medical Image Analysis* 85 (2023) 102729.
- [17] H. Wang, Y. Li, N. He, K. Ma, D. Meng, Y. Zheng, DICDNet: Deep interpretable convolutional dictionary network for metal artifact reduction in CT images, *IEEE Transactions on Medical Imaging* 41 (4) (2021) 869–880.
- [18] H. Wang, Y. Li, D. Meng, Y. Zheng, Adaptive convolutional dictionary network for CT metal artifact reduction, in: the Thirty-First International Joint Conference on Artificial Intelligence, 2022, pp. 1401–1407.
- [19] H. Wang, Y. Li, H. Zhang, J. Chen, K. Ma, D. Meng, Y. Zheng, InDuDoNet: An interpretable dual domain network for CT metal artifact reduction, in: Medical Image Computing and Computer-Assisted Intervention, Vol. 12906, 2021, pp. 107–118.
- [20] W. Dong, G. Shi, X. Li, Nonlocal image restoration with bilateral variance estimation: A low-rank approach, *IEEE Transactions on Image Processing* 22 (2) (2013) 700–711.
- [21] J. Tropp, A. Gilbert, M. Strauss, Simultaneous sparse approximation via greedy pursuit, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 5, 2005, pp. 721–724.