CD-DPE: Dual-Prompt Expert Network Based on Convolutional Dictionary Feature Decoupling for Multi-Contrast MRI Super-Resolution (Supplementary Materials)

Xianming Gu[1], Lihui Wang[1*], Ying Cao[1], Zeyu Deng[1], Yingfeng Ou[1], Guodong Hu[1], Yi Chen[1,2]

[1]Key Laboratory of Advanced Medical Imaging and Intelligent Computing of Guizhou Province,
Engineering Research Center of Text Computing & Cognitive Intelligence, Ministry of Education,
College of Computer Science and Technology, Guizhou University, Guiyang, China
[2]The D-Lab, Department of Precision Medicine, GROW-School for Oncology and Reproduction,
Maastricht University, 6200 MD Maastricht, the Netherlands
xianming_gu@foxmail.com, lhwang2@gzu.edu.cn

Appendix A: Implementation description of different comparison methods

WavTrans

This method (Li et al. 2022) integrates wavelet transformation with Transformer architecture to enhance multi-contrast MRI super-resolution. By decomposing reference images into high-frequency (detail) and low-frequency (approximation) wavelet coefficients, the approach effectively preserves structural boundaries that are typically blurred during conventional feature extraction. The architecture employs a residual cross-attention dual Transformer to model long-range feature dependencies, augmented by a multi-residual fusion mechanism that intelligently combines high-frequency components from both target and reference images. During training, the model is optimized using a composite loss function incorporating both K-space data consistency and image reconstruction loss. Implementation details are available at: https://github.com/XAIMI-Lab/WavTrans.

SANet

This method (Feng et al. 2024) introduces a novel framework for multi-contrast MRI super-resolution by bidirectionally exploring high- and low-intensity regions from reference images. Leveraging complementary attention mechanisms, it computes high-intensity attention via sigmoid activation while deriving low-intensity attention through subtraction, followed by convolutional fusion of these weighted features. Furthermore, the method incorporates an adaptive multi-stage integration module that dynamically weights features across different stages based on their learned dependencies, enabling the network to selectively emphasize the most informative features for target image reconstruction. This approach effectively captures anatomical structures while preserving fine edge details for enhanced super-resolution performance. The specific implementation code link is: https://github.com/chunmeifeng/SANet.

DiffMSR

This study (Li et al. 2024) proposes an efficient diffusion-based MRI super-resolution framework that overcomes traditional limitations of computational complexity and image distortion. Key innovations include: (1) operating diffusion processes in a compact latent space to enable faster convergence, (2) a Prior-Guided Large Window Transformer (PLWformer) for distortion-free reconstruction, and (3) a two-phase training strategy separating prior extraction from diffusion modeling. The method leverages pixel uncoupling and conditional entropy compression to generate compact prior knowledge, which then guides PLWformer through iterative refinement. Implementation is available at: https://github.com/GuangYuanKK/DiffMSR.

DANCE

This work (Chen et al. 2025) addresses cross-modal misalignment challenges in multi-contrast MRI super-resolution through a deformable attention mechanism and neighborhood feature aggregation. The framework first employs transformer-based feature extraction, followed by a two-stage fusion process: initial deformable attention corrects spatial mismatches between low-resolution and reference images, while subsequent neighborhood aggregation transfers relevant features regardless of residual alignment errors. This dual strategy ensures robust feature utilization from reference images while maintaining computational efficiency. (Implementation available upon request from authors.)

A2-CDic

This work (Lei et al. 2025) presents an alignment-enhanced multi-contrast convolutional dictionary model that combines interpretable sparse coding with deep learning advantages. The framework incorporates: (1) a convolutional sparse coding model that decomposes images into shared and contrast-specific components, implemented through a multi-scale network for efficient optimization, and (2) a spatial alignment module to compensate for cross-contrast mismatches, enabling effective reference information utilization while

[*] Corresponding author.

Table C1: Quantitative comparison results of multi-contrast MRI super-resolution based on LPIPS and RMSE metric on both BraTS2018 and IXI datasets with 2× and 4× super-resolution.

Methods	BraTS2018 2×		BraTS2018 4×		IXI 2×		IXI 4×	
	LPIPS↓	RMSE↓	LPIPS↓	RMSE↓	LPIPS↓	RMSE↓	LPIPS↓	RMSE↓
WavTrans	0.0081	2.1295	0.0223	3.2932	0.0029	1.8684	0.0087	2.8221
SANet	0.0120	3.1754	0.0338	3.9502	0.0033	2.1584	0.0120	3.7744
DiffMSR	/	/	0.0227	7.1694	/	/	0.0096	3.5386
DANCE	0.0141	4.7836	0.0267	4.6641	0.0394	5.4005	0.0209	4.4493
A2-CDic	0.0074	2.0118	0.0200	3.0012	0.0039	2.0988	0.0097	2.9671
CD-DPE	**0.0067**	**1.9708**	**0.0177**	**2.9552**	**0.0025**	**1.7867**	**0.0083**	**2.7732**

suppressing misalignment artifacts. Implementation is available at: https://github.com/lpcccc-cv/A2-CDic.

Appendix B: Detailed Explanation of Quantitative Metrics

To quantitatively evaluate the performance of different methods on multi-contrast MRI super-resolution datasets, we used peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) as metrics.

PSNR is a measure of the fidelity of image pixel value differences, defined as:

$$\text{PSNR}(I_1, I_2) = 10\log_{10}\left(\frac{\text{MAX}_{I_1}^2}{\text{MSE}(I_1, I_2)}\right), \tag{B1}$$

where MAX is the maximum pixel value (e.g., 255), and MSE is the mean square error. The higher the PSNR value, the better the reconstruction quality.

SSIM is used to evaluate the brightness, contrast, and structural similarity between images, defined as:

$$\text{SSIM}(I_1, I_2) = \frac{\left(2\mu_{I_1}\mu_{I_2} + C_1\right)\left(2\sigma_{I_{12}} + C_2\right)}{\left(\mu_{I_1}^2 + \mu_{I_2}^2 + C_1\right)\left(\sigma_{I_1}^2 + \sigma_{I_2}^2 + C_2\right)}, \tag{B2}$$

where μ_{I_1} and μ_{I_2} are the mean intensities of I_1 and I_2, respectively. σ_{I_1} and σ_{I_2} are the variance of I_1 and I_2, respectively. $\sigma_{I_{12}}$ is the covariance between I_1 and I_2. C_2 and C_2 are two constants determined by the dynamic range of the image and scalar parameters. The closer the SSIM value is to 1, the better the image quality.

In addition, to further evaluate the performance of our model, we have adopted two additional evaluation metrics in the supplementary materials, including learned perceptual image patch similarity (LPIPS) (Zhang et al. 2018) and root mean squared error (RMSE). LPIPS is based on deep learning feature distance, simulating differences in human visual perception, defined as:

$$\text{LPIPS}(I_1, I_2) = \sum_l \frac{1}{H_l W_l} \|\phi_l(I_1) - \phi_l(I_2)\|_2^2, \tag{B3}$$

where ϕ_l is the feature map of the l-th layer of the pre-trained network. H and W are the height and width of the image, respectively. Here, we use VGG16 (Simonyan and Zisserman 2014) as the pre-trained network. The smaller the LPIPS value, the better the image quality.

RMSE directly reflects pixel-level error, defined as:

$$\text{RMSE}(I_1, I_2) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_1(i) - I_2(i))^2}, \tag{B4}$$

where N is the total number of pixels in the image. The smaller the RMSE value, the better the image quality.

Appendix C: Supplementary Results

Quantitative Evaluation Results with Supplementary metrics

Table C1 presents additional quantitative comparisons using LPIPS and RMSE metrics for both 2× and 4× super-resolution tasks on BraTS2018 and IXI datasets. Our method demonstrates superior performance across all evaluation metrics, achieving consistently lower values than state-of-the-art approaches. On BraTS2018, our LPIPS improvements of 10.45% (2×) and 12.99% (4×) demonstrate superior perceptual quality, while RMSE reductions of 2.08% (2×) and 1.56% (4×) indicate enhanced reconstruction accuracy. Similar advantages are observed on IXI, with LPIPS improvements of 16.00% (2×) and 4.82% (4×), coupled with RMSE reductions of 4.57% (2×) and 1.76% (4×). These comprehensive results confirm our method's consistent ability to generate both perceptually realistic and quantitatively accurate super-resolution outputs compared to existing approaches.

Qualitative Analysis of 2× Super-Resolution Performance

Visual comparison of 2× SR reconstruction between different methods is presented in Figures C1 and C2 for the BraTS2018 and IXI datasets. The pseudo-color difference maps reveal distinct performance characteristics among the evaluated methods. Conventional approaches like SANet and DANCE exhibit substantial intensity variations that compromise image fidelity, resulting in noticeable detail loss and structural distortion. While WavTrans achieves competent high-frequency reconstruction, its tendency toward excessive smoothing diminishes important textural features

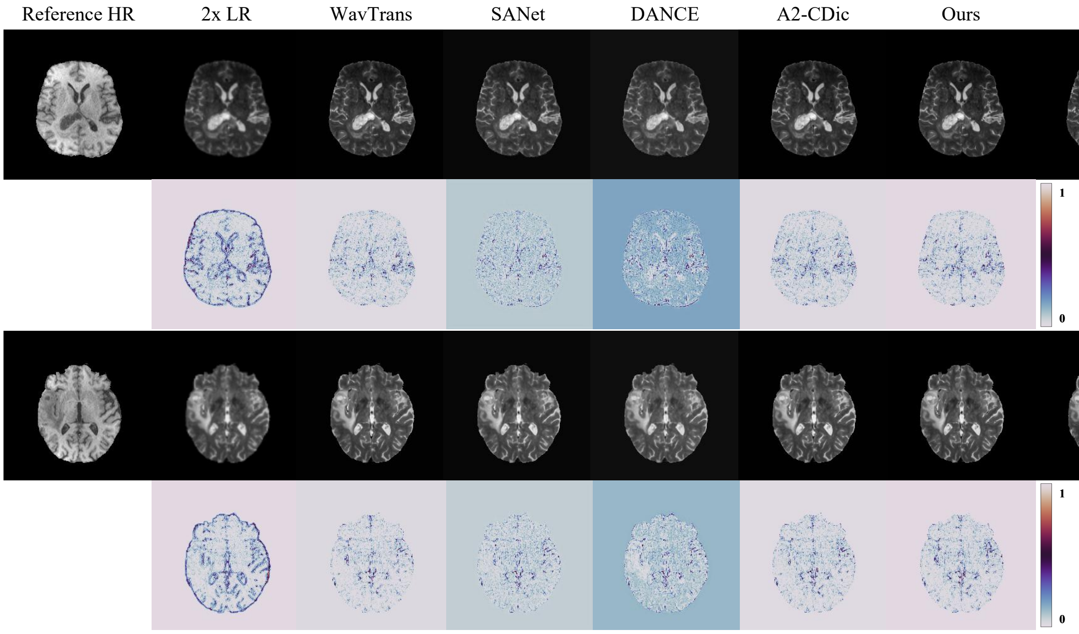


Figure C1: Qualitative comparison of various methods on the BraTS2018 dataset with 2× SR. The residual plot of the results and GT is shown below.

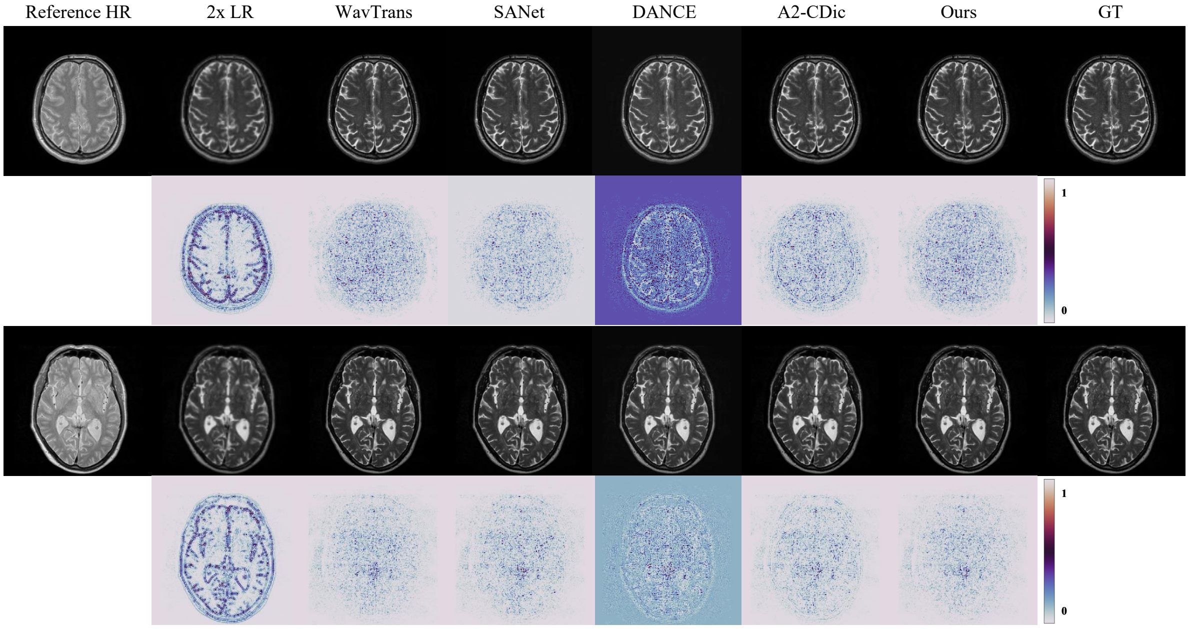
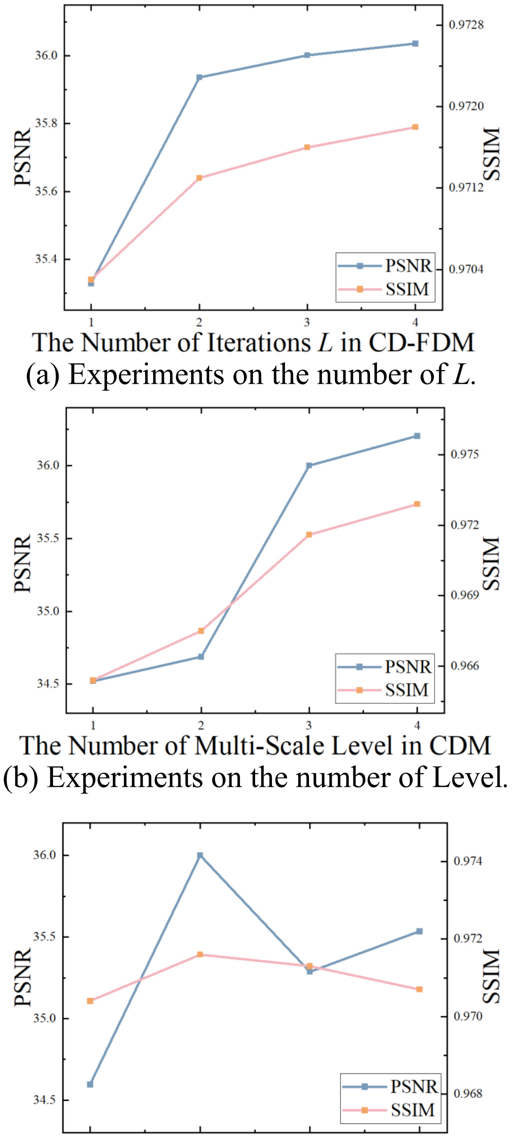


Figure C2: Qualitative comparison of various methods on the IXI dataset with 2× SR. The residual plot of the results and GT is shown below.

critical for medical interpretation. The A2-CDic framework shows artifacts in cerebrospinal fluid regions and demonstrates inadequate utilization of reference image data.

Our proposed method consistently outperforms these alternatives, as evidenced by the minimal residual errors in residual maps. The reconstructed images maintain anatomical authenticity through precise preservation of textural patterns and structural details. This visual superiority is particularly evident in the faithful incorporation of reference image information without introducing artificial textures or



(a) Experiments on the number of L.

(b) Experiments on the number of Level.

(c) Experiments on the number of E.

Figure C3: Quantitative results under different hyperparameter settings, including (a) the number of iterations L of CD-FDM, (b) the level of multi-scale feature of CDM, (c) the number of experts E of DP-FFEM.

compromising spatial resolution. The results demonstrate our method's ability to balance edge sharpness with natural texture representation, achieving superior reconstruction quality suitable for clinical applications. These qualitative findings align with and reinforce the quantitative advantages documented in previous sections.

Hyperparameters Ablation Analysis

We conducted systematic experiments to analyze key hyperparameter impacts on model performance, as demonstrated in Figure C3. For the CD-FDM module, we evaluated iteration counts $L = \{1, 2, 3, 4\}$ and observed that both PSNR and SSIM metrics progressively improved with increasing iterations. However, beyond L=2, the performance gains became marginal while computational complexity grew exponentially, leading us to select $L = 3$ as the optimal trade-off.

The convolutional dictionary scale analysis examined four configurations $\{1, 2, 3, 4\}$ with progressively increasing channel dimensions $\{64, 96, 128, 160\}$. Our results demonstrate that scaling from level 2 to 3 yielded substantial performance improvements, while further expansion to level 4 provided limited benefits at significantly increased parameter costs. Based on this nonlinear scaling behavior, we adopted level 3 as our standard configuration.

Regarding DP-FFEM architecture optimization, we explored expert counts $E = \{2, 4, 6, 8\}$. The analysis revealed that $E = 2$ proved insufficient for effective adaptive routing, while $E = 4$ achieved optimal performance. Notably, expanding beyond four experts ($E = 6/8$) actually degraded results, suggesting the existence of a saturation point for expert specialization benefits. All experiments consistently balanced performance metrics against computational complexity, with selected hyperparameters demonstrating superior performance compared to baseline methods across all configurations tested.

References

Chen, W.; Wu, S.; Wang, S.; Li, Z.; Yang, J.; Yao, H.; Tian, Q.; and Song, X. 2025. Multi-contrast image super-resolution with deformable attention and neighborhood-based feature aggregation (DANCE): Applications in anatomic and metabolic MRI. *Medical Image Analysis*, 99: 103359.

Feng, C.-M.; Yan, Y.; Yu, K.; Xu, Y.; Fu, H.; Yang, J.; and Shao, L. 2024. Exploring separable attention for multi-contrast MR image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*.

Lei, P.; Zhang, M.; Fang, F.; and Zhang, G. 2025. Robust Deep Convolutional Dictionary Model with Alignment Assistance for Multi-Contrast MRI Super-resolution. *IEEE Transactions on Medical Imaging*.

Li, G.; Lyu, J.; Wang, C.; Dou, Q.; and Qin, J. 2022. Wavtrans: Synergizing wavelet and cross-attention transformer for multi-contrast mri super-resolution. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 463–473. Springer.

Li, G.; Rao, C.; Mo, J.; Zhang, Z.; Xing, W.; and Zhao, L. 2024. Rethinking diffusion model for multi-contrast mri super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11365–11374.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.