FDF-VQVAE: A Frequency Disentanglement and Fusion Learning Framework for Multi-Sequence MRI Enhancement

Xinghe Xie[1], Luyi Han[2,3], Yue Sun[1], Chi Kin Lam[1], Jian Zheng[4], Tong Tong[5], Wei Ke[1], Chan-Tong Lam[1], and Tao Tan[1⋆(✉)]

[1] Faculty of Applied Sciences, Macao Polytechnic University, Macao, China
[2] Radboud University Medical Centre, Nijmegen, Netherlands
[3] Department of Radiology, The Netherlands Cancer Institute, Amsterdam, Netherlands
[4] Department of Medical Imaging, Suzhou Institute of Biomedical Engineering, Chinese Academy of Sciences, Suzhou, China
[5] Fuzhou University, Fuzhou, China
taotan@mpu.edu.mo

Abstract. Multi-sequence magnetic resonance imaging (MRI) faces critical challenges in balancing accelerated acquisition and image quality: Rapid scanning typically induces degradation, including resolution reduction, increased noise, motion artifacts, and image blurring. While existing image enhancement models partially mitigate these issues, they often exhibit insufficient exploitation of complementary information across multi-sequence data. To address this issue, we propose an interpretable deep learning framework, FDF-VQVAE, for MRI image enhancement through frequency-domain feature disentanglement and fusion. Our framework constructs a dual-branch frequency-domain disentanglement module (DBFD) that precisely decouples high-frequency and low-frequency features of different sequences through parallel high-frequency feature and low-frequency feature extraction pathways. The multi-frequency-domian feature weighting mechanism (MFDFW) adaptively fuses the high and low frequency features of different sequences. Finally, feature recombination and decoding achieve MRI enhancement through joint optimization. We conducted denoising, super-resolution, and deblurring experiments on the IXI dataset (546 subjects) with external validation on the BraTS2021 dataset (357 subjects). Experimental results demonstrate that our method significantly outperforms the state-of-the-art approaches in denoising, motion artifact removal, and super-resolution tasks. Our code is available at https://github.com/kkllxh/FDF-VQVAE.

Keywords: Multi-Sequence Brain MRI · Image Enhancement · Frequency-Disentanglement and Fusion · Interpretability.

⋆ Corresponding author

1 Introduction

Multi-sequence magnetic resonance imaging (MRI) has become an indispensable tool in clinical diagnosis, providing doctors with rich detailed tissue information [3, 11–13], and is widely used in disease diagnosis [12], lesion segmentation [13], and treatment prognosis evaluation [2]. However, to improve scanning efficiency and save time, it is often necessary to reduce image resolution or increase the acceleration factor, which can lead to issues such as noise, motion artifacts, and image blurring [6]. To address these problems, deep learning-based MRI image enhancement techniques have gradually become an effective solution, attracting widespread attention.

In recent years, numerous deep learning-based methods have been applied to medical image enhancement [15, 6, 19, 4]. However, these methods mainly focus on unimodal processing and fail to fully utilize the complementary information between multi-sequence MRI. Although some methods have attempted to leverage the complementary information across multi-sequence MRI, current approaches still have fundamental limitations. For instance, Dalmaz et al. [5] proposed a shared encoder architecture for multi-sequence feature extraction, but their framework inadequately decouples sequence-specific features, resulting in suboptimal feature discriminability. Furthermore, other methods [9, 20] employing independent encoders for individual sequence feature extraction often struggle to effectively capture inter-sequence shared representations. The inability to concurrently learn sequence-specific distinctiveness and cross-sequence commonality significantly constrains their practical efficacy. To improve MRI image enhancement performance, it is crucial to effectively decouple and fuse this complementary information.

MRI images can be decoupled into high-frequency and low-frequency features [14, 18]. The low-frequency features contain shared information about anatomical structures (such as organ morphology and tissue distribution), while the high-frequency features capture complementary information (such as texture details and edge features). Therefore, we propose an interpretable deep learning framework that can adaptively decouple and fuse high-frequency and low-frequency features from different MRI sequences, further synthesizing enhanced images. Our contributions can be summarized as follows: (1) We propose an interpretable network architecture capable of decoupling and multi-frequency-domian feature weighting mechanism (MFDFW) of high-frequency and low-frequency features from different sequences while quantifying the contribution of high-frequency and low-frequency features in each sequence. (2) We design a wavelet transform decoupling loss that guides the model to effectively decouple high-frequency and low-frequency features, thereby enhancing the feature extraction process. (3) We propose a dual-branch frequency-domain disentanglement module (DBFD) that better decouples high-frequency and low-frequency features from different sequences. (4) We validate the model through tasks such as super-resolution, denoising, and motion artifact removal, demonstrating its generalization ability on external datasets.

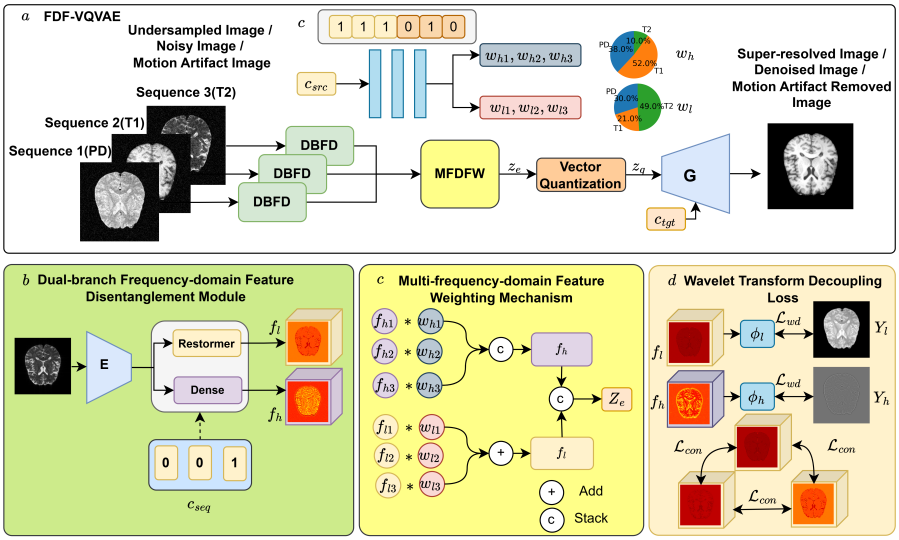

Fig. 1. The overview of the FDF-VQVAE framework.

2 Method

2.1 Frequency Disentanglement and Fusion Learning Framework

Figure 1 illustrates the overall architecture of the proposed FDF-VQVAE network. The network mainly consists of an encoder E, a dual-branch frequency-domain disentanglement module, multi-frequency-domian feature weighting mechanism, and a decoder G. The decoder G is composed of two hyper ConvNeXt blocks, a hyper convolution layer, and LeakyReLU. First, the input MRI sequence is processed by the encoder E to extract shallow features. Then, the dual-branch frequency-domain disentanglement module decouples the high-frequency and low-frequency features of all input sequences based on c_{seq}. Next, the weights for the high and low frequency features are generated through c_{src} to enable the effective fusion of multi-sequence features. Finally, the fused features Z_e are transformed into quantized features Z_q using vector quantization [16], which are then passed to the decoder G along with c_{tgt} to generate the target sequence. In addition, in this network, c is composed of c_{src} and c_{tgt}.

Dual-Branch Frequency-Domain Disentanglement Module Define a set of N sequences MRI images with its corresponding simulated degraded image $\mathcal{I} = \{Y_i, X_i \mid i = 1, \ldots, N\}$, where $X_i = \mathcal{F}_d(Y_i)$. $\mathcal{F}_d$ represents the degradation function. We propose a dual-branch frequency-domain feature disentanglement module to extract high and low-frequency features $f_{\text{hi}}, f_{\text{li}} = \text{DBFD}(X_i, c_{\text{seq}})$, as

shown in Figure 1(b). Here, c_{seq} represents the one-hot encoding of the sequence. The module consists of an Encoder, a hyper Restormer block, and a hyper dense block. The hyper Restormer module [17], where hyper convolutions [7] replace standard convolutions, is used to extract global features from MRI images. The hyper dense module [21], similarly using hyper convolutions instead of standard convolutions, is designed to extract local features from MRI images. This enables targeted disentangling for different MRI sequences. The encoder consists of two ConvNeXt blocks, a convolutional layer, LayerNorm, and LeakyReLU, aimed at extracting more complex high-dimensional features while ensuring computational efficiency, thereby enhancing feature representation ability.

Multi-frequency-domian Feature Weighting Mechanism The low-frequency features represent shared information, while the high-frequency features represent complementary information. Therefore, we integrate the low-frequency features using a weighted sum and the high-frequency features using weighted concatenation, followed by further fusion of the integrated high and low-frequency information. To enable the model to learn the optimal weights, we define a learnable parameter to predict the best weights $\omega_i \in \mathbb{R}^N$ from c_{src}.

$$\omega_i = \text{sigmoid}(c_{\text{src}} \cdot \omega + b) + \epsilon, \quad i \in \{low, high\} \tag{1}$$

where ω and b are the weights and biases of the fully connected layer. $\epsilon = 10^-5$ to avoid dividing 0. To simulate modality missing, during training, we randomly set c_{src} to 0 in the MFDFW mechanism and ensure that the output ω_i sums to 1.

$$\omega_i = \frac{\omega_i \cdot c_{\text{src}}}{\sum_{j=1}^{N} \omega_{ij} + \epsilon}, \quad i \in \{low, high\} \tag{2}$$

where $\cdot$ refers to the element-wise product, j represents the j^{th} weight in ω_i, N is the number of ω_i, $\epsilon = 10^{-5}$ prevent the denominator from being zero.

By using the Dual-Branch Frequency-Domain Disentanglement Module and the high-frequency and low-frequency weights ω_h and ω_l, the weighted fusion of high-frequency and low-frequency features from multiple sequences is achieved.

$$\begin{aligned} f_h &= [\omega_{h1} \cdot f_{h1}, \omega_{h2} \cdot f_{h2}, \omega_{h3} \cdot f_{h3}] \\ f_l &= \sum_{i=1}^{N} \omega_{li} \cdot f_{li} \\ f &= [f_h, f_l] \end{aligned} \tag{3}$$

where f_{hi} and f_{li} represent the high-frequency and low-frequency features of the i-th input sequence, respectively. $[\cdot, \cdot]$ presents channel-wise concatenation.

2.2 Loss Function

Wavelet Transform Decoupling Loss To better decouple high-frequency and low-frequency information, we propose a wavelet transform decoupling loss.

First, we introduce a low-frequency consistency loss $\mathcal{L}_{\text{con}}$ to enforce the consistency of the low-frequency features extracted from different sequences. using wavelet constraint loss $\mathcal{L}_{\text{wd}}$ to guide model to futher decouple the high and low-frequency information, we use the wavelet-transformed high and low-frequency images Y_{hi} and Y_{li} of the original image to supervise the extracted features f_{hi} and f_{li}. We remap the decoupled features back to the image space and apply pixel-level loss for constraint. Wavelet transform decoupling loss consists of $\mathcal{L}_{\text{con}}$ and $\mathcal{L}_{\text{wd}}$.

$$\mathcal{L}_{\text{con}}(i,j) = \|f_{li} - f_{lj}\|_2^2 \tag{4}$$

where f_{li} and f_{lj} represent the low-frequency features extracted from the different input sequences.

$$\mathcal{L}_{\text{wd}}(f_{li}, Y_{li}, f_{hi}, Y_{hi}) = \mathcal{L}_{rec}(\phi_l(f_{li}), Y_{li}) + \mathcal{L}_{rec}(\phi_{hi}(f_{hi}), Y_{hi}) \tag{5}$$

where ϕ_l denotes the shared weight projection for low-frequency features, and ϕ_{hi} denotes the private weight projection for high-frequency features. The reconstruction loss, $\mathcal{L}_{rec}$, consists of the mean squared error (MSE) and structure similarity index measure (SSIM) loss.

Image Reconstruction Loss To improve the enhancement effect, we impose constraints on the generated images at the image level, structural level, and semantic level. $\hat{Y}$ represents the model's output image.

$$\begin{aligned}\mathcal{L}_{\text{pixel}}(Y,\hat{Y}) =& \mathcal{L}_{\text{MSE}}(Y,\hat{Y}) + \mathcal{L}_{\text{SSIM}}(Y,\hat{Y}) + \mathcal{L}_{\text{Lap}}(Y,\hat{Y}) \\ &+ \mathcal{L}_{\text{per}}(Y,\hat{Y}) + L_{vq}(z_e, z_q)\end{aligned} \tag{6}$$

Here, $\mathcal{L}_{\text{MSE}}$ denotes the MSE loss, $\mathcal{L}_{\text{SSIM}}$ denotes the SSIM loss, $\mathcal{L}_{\text{Lap}}$ denotes the Laplacian loss, $\mathcal{L}_{\text{per}}$ denotes the perceptual loss based on a pre-trained VGG19 network, and $\mathcal{L}_{\text{vq}}$ denotes the commitment loss and vector quantization loss [16].

Total Loss The loss function combines image reconstruction loss, wavelet transform decoupling loss, and low-frequency consistency loss to constrain the network. The implementation is as follows:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{pixel}}(Y,\hat{Y}) + \sum_{i=1}^{N} \mathcal{L}_{\text{wd}}(f_{li}, Y_{li}, f_{hi}, Y_{hi}) + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j\neq i}}^{N} \mathcal{L}_{\text{con}}(i,j) \tag{7}$$

3 Experiments

3.1 Experimental Settings

Dataset and Evaluation Metrics We conducted a study using brain MRI images from 568 subjects in the IXI dataset [10], which included three aligned sequences: T1, T2, and proton density(PD). We selected 340 subjects for training,

57 for validation, and 171 for testing. Additionally, we used brain MRI images from 357 subjects in the BraTS2021 dataset [1] for external validation, which included T1 and T2 sequences. All MRI images underwent intensity normalization, with the intensity range standardized to [0, 1]. We simulated different low-quality images in the following ways: for the super-resolution task, we reduced the MRI image quality by applying 8x Fast Fourier Transform(FFT) undersampling to simulate low-resolution images; for the denoising task, gaussian noise with a mean of 0 and a standard deviation of 1 was added to the MRI images to simulate real-world noise interference; for motion artifact generation, we applied 6x cartesian acceleration undersampling to simulate image distortions caused by motion. To evaluate the synthesis performance, we used three common performance metrics: peak signal-to-noise ratio (PSNR), structure similarity index measure (SSIM), and learned perceptual image patch similarity (LPIPS). Using these three metrics, we conducted a comprehensive evaluation of the synthesized images from the perspectives of intensity, structure, and perception.

3.2 Implementation Details

The models are implemented with PyTorch and trained on the NVIDIA A40 GPU. The E and G are trained using the AdamW optimizer with an initial learning rate of 0.0001, betas = (0.9, 0.95), weight decay = 0.05, a batch size of 1, for 500 epochs.

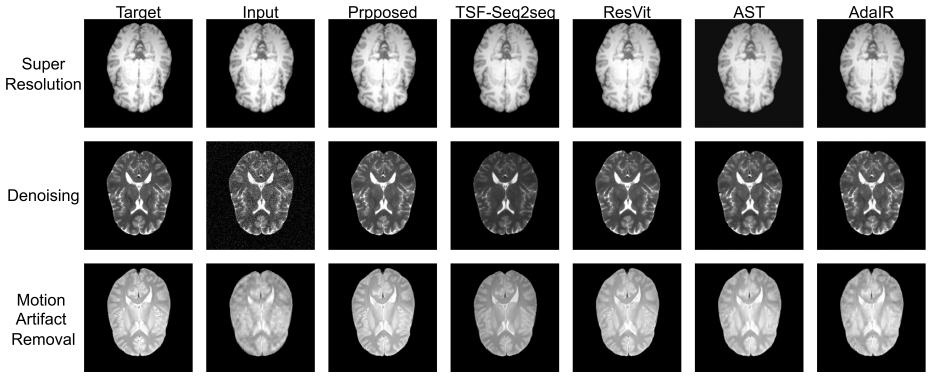

Fig. 2. The synthesized results of PD, T1, and T2 for our model and the comparison model in the tasks of super-resolution, denoising, and motion artifact removal.

3.3 Experimetal Results

Comparative Experiment We compared the proposed method with multi-sequence models (ResVit[5], TSF-seq2seq[8]) and single-sequence models (AST[19],

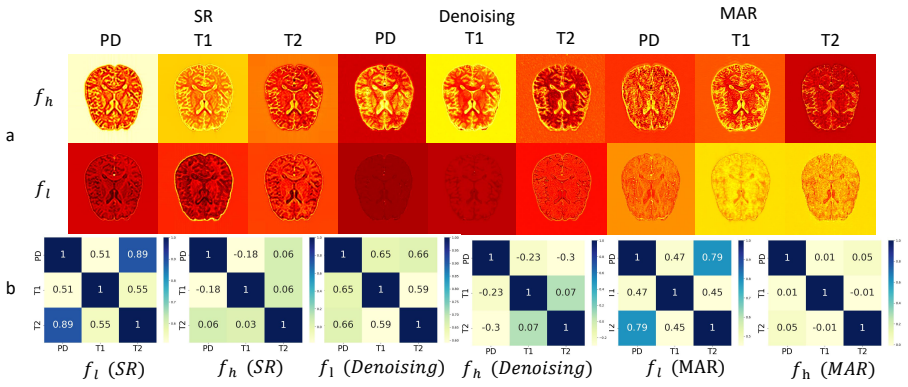

Fig. 3. (a) The visualization of high and low frequency features. (b) The CC between high-frequency features and the CC between low-frequency features in the three tasks.

Table 1. The results of super-resolution, denoising, and motion artifact removal on the IXI dataset and BraTS2021 dataset.

Dataset		IXI Dataset			BraTS2021 dataset(External Validation)				
Task name	Methods	SSIM ↑	PSNR (dB) ↑	LPIPS ↓	SSIM ↑	PSNR (dB) ↑	LPIPS ↓	Parameters (M)	GFLOPs (G)
SR	Our Model	0.972 ± 0.01	32.5 ± 1.8	0.039 ± 0.01	0.963 ± 0.01	31.2 ± 1.4	0.060 ± 0.02	0.3	157.2
	TSF-Seq2Seq	0.967 ± 0.01	31.6 ± 1.7	0.046 ± 0.04	0.948 ± 0.02	25.3 ± 2.4	0.090 ± 0.03	7.3	67.1
	AST	0.954 ± 0.01	31.3 ± 1.7	0.112 ± 0.03	0.949 ± 0.01	30.8 ± 1.7	0.145 ± 0.04	47.7	189.0
	AdaIR	0.957 ± 0.01	31.1 ± 2.1	0.117 ± 0.05	0.933 ± 0.02	30.8 ± 1.2	0.093 ± 0.02	28.7	344.9
	ResVit	0.969 ± 0.01	31.6 ± 2.0	0.076 ± 0.03	0.947 ± 0.01	26.6 ± 1.6	0.106 ± 0.03	11.3	167.6
Denoising	Our Model	0.975 ± 0.01	33.9 ± 1.1	0.028 ± 0.01	0.972 ± 0.01	32.4 ± 1.1	0.049 ± 0.02	0.3	157.2
	TSF-Seq2Seq	0.971 ± 0.01	33.4 ± 1.3	0.042 ± 0.01	0.953 ± 0.01	27.0 ± 2.7	0.085 ± 0.03	7.3	67.1
	AST	0.942 ± 0.01	32.6 ± 0.8	0.052 ± 0.01	0.910 ± 0.01	32.1 ± 0.7	0.067 ± 0.01	47.7	189.0
	AdaIR	0.951 ± 0.01	32.1 ± 1.0	0.050 ± 0.01	0.903 ± 0.01	31.4 ± 0.6	0.067 ± 0.02	28.7	344.9
	ResVit	0.972 ± 0.01	33.4 ± 1.3	0.042 ± 0.01	0.963 ± 0.01	30.8 ± 1.9	0.083 ± 0.02	11.3	167.6
Motion artifact removal	Our Model	0.973 ± 0.01	33.5 ± 1.4	0.026 ± 0.01	0.970 ± 0.01	32.1 ± 1.2	0.041 ± 0.01	0.3	157.2
	TSF-Seq2Seq	0.964 ± 0.01	33.3 ± 1.6	0.046 ± 0.03	0.940 ± 0.01	25.4 ± 2.4	0.100 ± 0.03	7.3	67.1
	AST	0.954 ± 0.01	31.2 ± 1.2	0.067 ± 0.01	0.928 ± 0.01	30.7 ± 1.2	0.075 ± 0.01	47.7	189.0
	AdaIR	0.946 ± 0.01	30.9 ± 1.2	0.068 ± 0.01	0.926 ± 0.02	30.9 ± 1.1	0.082 ± 0.02	28.7	344.9
	ResVit	0.966 ± 0.01	32.4 ± 1.9	0.047 ± 0.02	0.962 ± 0.01	31.2 ± 1.2	0.060 ± 0.02	11.3	167.6

AdaIR[4]). Figure 2 shows the effect maps for super-resolution (SR), denoising, and motion artifact removal (MAR). Table 1 summarizes the performance of these models in these tasks. The proposed method outperforms all others, with multi-sequence models showing better results than single-sequence models, highlighting the benefit of using complementary information from multiple sequences. Our method uses only about 0.3 M parameters and around 157.2 GFLOPS of computation, achieving a superior parameter–compute balance while maintaining high performance.

Ablation Study To validate the rationality of each module in the model, we conducted ablation experiments on wavelet transform decoupling loss, DBFD, and MFDFW, with the results shown in Table 2. The wavelet transform decoupling loss effectively constrains the model's performance in high and low frequency feature extraction; the DBFD module effectively decouples high and

Table 2. The results of the ablation experiments on the model conducted on the IXI dataset.

Dataset					IXI Dataset		
Task name		L_{wd}, L_{con}	MFDFW	DBFD	SSIM ↑	PSNR (dB) ↑	LPIPS ↓
SR	Our Model	✓	✓	✓	$\underline{0.972 \pm 0.01}$	$\underline{32.5 \pm 1.8}$	$\underline{0.039 \pm 0.01}$
	Our Model		✓	✓	0.958 ± 0.02	29.3 ± 1.6	0.067 ± 0.02
	Our Model	✓		✓	0.957 ± 0.01	29.8 ± 1.5	0.063 ± 0.01
	Our Model	✓	✓		0.936 ± 0.03	30.0 ± 1.7	0.079 ± 0.02
Denoising	Our Model	✓	✓	✓	$\underline{0.975 \pm 0.01}$	$\underline{33.9 \pm 1.1}$	$\underline{0.028 \pm 0.01}$
	Our Model		✓	✓	0.962 ± 0.01	30.4 ± 1.5	0.049 ± 0.01
	Our Model	✓		✓	0.963 ± 0.01	31.0 ± 1.9	0.043 ± 0.01
	Our Model	✓	✓		0.962 ± 0.01	30.8 ± 1.0	0.051 ± 0.01
Motion artifact removal	Our Model	✓	✓	✓	$\underline{0.973 \pm 0.01}$	$\underline{33.5 \pm 1.4}$	$\underline{0.026 \pm 0.01}$
	Our Model		✓	✓	0.955 ± 0.01	30.2 ± 1.5	0.048 ± 0.01
	Our Model	✓		✓	0.959 ± 0.01	30.6 ± 1.5	0.046 ± 0.01
	Our Model	✓	✓		0.956 ± 0.01	30.0 ± 1.3	0.052 ± 0.01

low frequency features, enhancing the model's performance; and the MFDFW mechanism effectively fuses high and low frequency features across multiple sequences.

Visualization of Feature Extraction As shown in Figure 3, the visualization of high and low frequency features and their correlation coefficients (CC) across three tasks is presented. As shown in Figure 3, the model effectively extracts both high and low frequency features in the MRI image.

Contribution of Low and High-Frequency Features In the SR task, the low-frequency weight of the PD sequence (0.98) helps maintain the stability of anatomical structures, while the low-frequency weight of the T1/T2 sequences (0.01) mitigates information conflicts caused by inter-modality contrast differences. The high-frequency weight of the T2 sequence (0.58) enhances the texture of lesion boundaries, the high-frequency weight of the T1 sequence (0.39) improves the sharpness of the gray-white matter interface, and the high-frequency weight of the PD sequence (0.03) effectively suppresses interference caused by low signal-to-noise ratio. In the denoising task, the low-frequency weight of the T2 sequence (0.49) stabilizes pathological signals, the weight of the PD sequence (0.3) performs brightness correction, and the low-frequency weight of the T1 sequence (0.21) prevents the amplification of contrast noise. The high-frequency weight of the T1 sequence (0.52) preserves the details of the gray-white matter structure, the PD sequence (0.38) supplements anatomical texture, and the high-frequency weight of the T2 sequence (0.1) supplements other details. In the MAR task, regarding low-frequency weights, the PD sequence (0.53) is primarily used to correct motion artifacts, the T1 sequence (0.26) assists in locating low-frequency structures at the gray-white matter boundary, and the T2 sequence (0.21) supplements the stability of fluid regions. In terms of high-frequency weights, the

T2 sequence (0.49) helps to enhance true edges, the T1 sequence (0.46) aids in identifying the direction of artifacts, and the PD sequence (0.05) suppresses motion-related high-frequency noise.

External Validation We conducted an external validation of our model and comparison models using the BraTS2021 dataset, which includes 357 subjects, to explore the model's generalization capability. Table 1 shows that the proposed method outperforms all other models in every task, demonstrating superior performance and excellent generalization ability.

3.4 Conclusion

In this study, we propose an interpretable network architecture designed to effectively extract high-frequency and low-frequency features from different sequences and adaptively adjust the fusion weights of these features according to specific task requirements. Additionally, we introduce a wavelet transform decoupling loss function to guide the model in extracting high-frequency and low-frequency features from MRI images. This method is suitable for multi-sequence synthesis tasks such as super-resolution reconstruction, denoising, and motion artifact removal. Experimental results on the IXI and BraTS2021 datasets demonstrate that our method effectively utilizes complementary information between sequences, significantly outperforming existing state-of-the-art techniques.

Acknowledgments This work was supported in part by Macao Polytechnic University Grant (RP/FCA-08/2024), the grant from Science and Technology Development Fund of Macao (0004/2024/E1B1).

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314 (2021)
2. Chefer, H., Gur, S., Wolf, L.: Transformer interpretability beyond attention visualization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 782–791 (2021)
3. Chen, J.H., Su, M.Y.: Clinical application of magnetic resonance imaging in management of breast cancer patients receiving neoadjuvant chemotherapy. BioMed research international **2013**(1), 348167 (2013)
4. Cui, Y., Zamir, S.W., Khan, S., Knoll, A., Shah, M., Khan, F.S.: Adair: Adaptive all-in-one image restoration via frequency mining and modulation. arXiv preprint arXiv:2403.14614 (2024)

5. Dalmaz, O., Yurt, M., Çukur, T.: Resvit: residual vision transformers for multimodal medical image synthesis. IEEE Transactions on Medical Imaging **41**(10), 2598–2614 (2022)
6. Feng, C.M., Yan, Y., Fu, H., Chen, L., Xu, Y.: Task transformer network for joint mri reconstruction and super-resolution. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. pp. 307–317. Springer (2021)
7. Han, L., Tan, T., Zhang, T., Huang, Y., Wang, X., Gao, Y., Teuwen, J., Mann, R.: Synthesis-based imaging-differentiation representation learning for multi-sequence 3d/4d mri. Medical Image Analysis **92**, 103044 (2024)
8. Han, L., Tan, T., Zhang, T., Wang, X., Gao, Y., Lu, C., Liang, X., Dou, H., Huang, Y., Mann, R.: Non-adversarial learning: Vector-quantized common latent space for multi-sequence mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 481–491. Springer (2024)
9. Han, L., Zhang, T., Huang, Y., Dou, H., Wang, X., Gao, Y., Lu, C., Tan, T., Mann, R.: An explainable deep framework: Towards task-specific fusion for multi-to-one mri synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 45–55. Springer (2023)
10. Jiang, D., Dou, W., Vosters, L., Xu, X., Sun, Y., Tan, T.: Denoising of 3d magnetic resonance images with multi-channel residual learning of convolutional neural network. Japanese journal of radiology **36**, 566–574 (2018)
11. Li, F., Wang, C., Xu, Z., Li, M., Deng, L., Wei, M., Zhang, H., Wu, K., Ning, R., Li, D., et al.: Biomed research international. Heart **24**(69), 8–74 (2015)
12. Mann, R.M., Cho, N., Moy, L.: Breast mri: state of the art. Radiology **292**(3), 520–536 (2019)
13. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
14. Qu, L., Zhang, Y., Wang, S., Yap, P.T., Shen, D.: Synthesized 7t mri from 3t mri via deep learning in spatial and wavelet domains. Medical image analysis **62**, 101663 (2020)
15. Sadikov, A., Wren-Jarvis, J., Pan, X., Cai, L.T., Mukherjee, P.: Generalized diffusion mri denoising and super-resolution using swin transformers. arXiv preprint arXiv:2303.05686 (2023)
16. Van Den Oord, A., Vinyals, O., et al.: Neural discrete representation learning. Advances in neural information processing systems **30** (2017)
17. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022)
18. Zhao, Z., Bai, H., Zhang, J., Zhang, Y., Xu, S., Lin, Z., Timofte, R., Van Gool, L.: Cddfuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5906–5916 (2023)
19. Zhou, S., Chen, D., Pan, J., Shi, J., Yang, J.: Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2952–2963 (2024)

20. Zhou, T., Fu, H., Chen, G., Shen, J., Shao, L.: Hi-net: hybrid-fusion network for multi-modal mr image synthesis. IEEE transactions on medical imaging **39**(9), 2772–2781 (2020)
21. Zhu, Y., Newsam, S.: Densenet for dense flow. In: 2017 IEEE international conference on image processing (ICIP). pp. 790–794. IEEE (2017)