



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Accelerated Multi-Contrast MRI Reconstruction via Frequency and Spatial Mutual Learning

Qi Chen¹, Xiaohan Xing^{2,*}, Zhen Chen³, and Zhiwei Xiong¹

¹ University of Science and Technology of China

² Stanford University

³ Centre for Artificial Intelligence and Robotics (CAIR), HKISI-CAS

Abstract. To accelerate Magnetic Resonance (MR) imaging procedures, Multi-Contrast MR Reconstruction (MCMR) has become a prevalent trend that utilizes an easily obtainable modality as an auxiliary to support high-quality reconstruction of the target modality with under-sampled k-space measurements. The exploration of global dependency and complementary information across different modalities is essential for MCMR. However, existing methods either struggle to capture global dependency due to the limited receptive field or suffer from quadratic computational complexity. To tackle this dilemma, we propose a novel Frequency and Spatial Mutual Learning Network (FSMNet), which efficiently explores global dependencies across different modalities. Specifically, the features for each modality are extracted by the *Frequency-Spatial Feature Extraction* (FSFE) module, featuring a frequency branch and a spatial branch. Benefiting from the global property of the Fourier transform, the frequency branch can efficiently capture global dependency with an image-size receptive field, while the spatial branch can extract local features. To exploit complementary information from the auxiliary modality, we propose a *Cross-Modal Selective fusion* (CMS-fusion) module that selectively incorporate the frequency and spatial features from the auxiliary modality to enhance the corresponding branch of the target modality. To further integrate the enhanced global features from the frequency branch and the enhanced local features from the spatial branch, we develop a *Frequency-Spatial fusion* (FS-fusion) module, resulting in a comprehensive feature representation for the target modality. Extensive experiments on the BraTS and fastMRI datasets demonstrate that the proposed FSMNet achieves state-of-the-art performance for the MCMR task with different acceleration factors.

Keywords: Magnetic Resonance Imaging · Multi-Contrast · Fourier Transform · Mutual Learning.

1 Introduction

Accelerated Magnetic Resonance (MR) imaging, which reconstructs MR images from under-sampled k-space measurements, can significantly reduce the cost of

* Corresponding author: Xiaohan Xing (xhxing@stanford.edu)

the MR imaging procedure and improve the patient experience [17,7,3]. However, the aliasing artifacts resulting from insufficient sampling often degrade the image quality and compromise clinical diagnoses [25]. In clinical practice, multi-contrast MR images, such as T1 and T2 weighted images (T1WIs and T2WIs), are often simultaneously acquired to provide complementary structural information for diagnosis and treatment planning. Recently, multi-contrast MR reconstruction (MCMR) [23,12] has become a prevalent trend that utilizes an easily obtainable modality (e.g., T1WIs) as an auxiliary to support high-quality reconstruction of under-sampled target modalities (e.g., T2WIs).

For multi-contrast MR images, similar structural features are distributed across different regions within each modality, and complementary information exists across different modalities. Therefore, the inherent problem of the MCMR task is how to comprehensively explore long-range dependencies within each modality and sufficiently leverage complementary information from the auxiliary modality. Toward this goal, a series of methods [5,10,21,19,20] integrate multi-modal features based on convolutional neural networks (CNNs). However, the actual receptive field of CNNs is restricted due to the vanishing gradient issue, thus failing to capture global dependencies. Recent advances in vision transformers [2] have led to self-attention-based feature integration methods [6,9], enabling the modeling of long-range dependencies and enhancing the performance of the MCMR task. However, these methods [6,9] face significant challenges in clinical deployment due to their quadratic computation complexity ($\mathcal{O}(N^2)$ for N tokens). Therefore, there is highly demanded to develop an efficient method that can capture global dependencies and extract complementary information from both modalities without introducing a heavy computational burden.

Inspired by [13,24], we resort to frequency information in the Fourier domain as an efficient global feature extractor. Since each pixel in the Fourier domain interacts with all pixels in the spatial domain, the frequency features naturally encompass global properties and can achieve an image-size receptive field. Thus, the global features extracted in the Fourier domain complement the local features extracted in the spatial domain, providing a more efficient way for global feature interaction across different modalities. In light of this, reconstruction accuracy for the target modality can be improved by addressing the following issues: 1) How to enhance the feature representation of each modality with the assistance of the global view from the Fourier domain. 2) How to effectively integrate the global frequency and local spatial features across different modalities.

In this work, we propose a novel Frequency and Spatial Mutual learning Network (FSMNet) for MCMR, offering an efficient method to extract and integrate global and local features across different modalities. FSMNet extracts features for each modality using a stack of *Frequency-Spatial Feature Extraction* (FSFE) modules, which include a frequency branch and a spatial branch. The frequency branch is designed to capture global features with an image-size receptive field, while the spatial branch specializes in extracting local structural features. To fully incorporate complementary information from the auxiliary modality, we introduce a *Cross-Modal Selective fusion* (CMS-fusion) module. This module

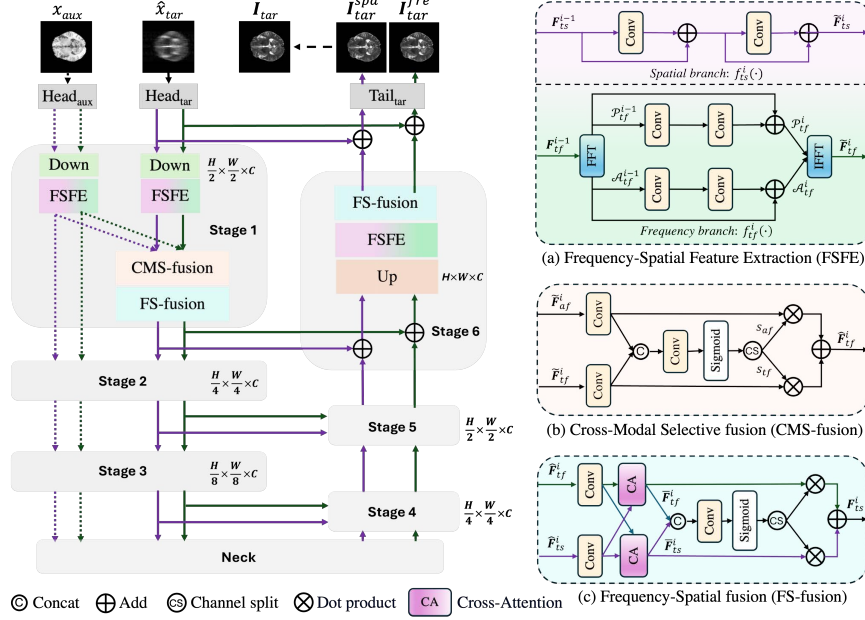


Fig. 1. Overview of FSMNet: In each stage, the FSFE module extracts global and local features from the frequency and spatial branches, respectively. The CMS-fusion module integrates the multi-modal features for each branch, and subsequently, the FS-fusion module combines the features across the frequency and spatial branches.

selectively integrates frequency and spatial features from the auxiliary modality to enhance the respective branch of the target modality. To further integrate enhanced global features from the frequency branch and enhanced local features from the spatial branch, we present a *Frequency-Spatial fusion* (FS-fusion) module that exchanges beneficial information across these two branches based on an adaptive attention map. The main contributions are summarized as:

- Our FSMNet provides an efficient method for capturing global dependency via Fourier transform. The proposed FSFE module enhances feature representations for each modality by extracting global features from the frequency branch and local features from the spatial branch. (§2.1)
- We propose a CMS-fusion module to enhance global and local features by selectively incorporating complementary information from the auxiliary modality, along with an FS-fusion module to further integrate the enhanced global and local features for the target modality. (§2.2)
- Experimental results on two MRI datasets (BraTS [14] and fastMRI [22]) demonstrate that FSMNet outperforms existing MCMR methods across various acceleration factors. (§3.2)

2 Methodology

An overview of FSMNet is depicted in Fig. 1. Given the under-sampled target image \hat{x}_{tar} and the fully-sampled image x_{aux} from the auxiliary modality, FSMNet aims to reconstruct a high-quality image \mathbf{I}_{tar} for the target modality. Each modality comprises two branches, the frequency branch (green lines) and the spatial branch (purple lines). In each encoder stage (Stage 1-3), the FSFE module extracts global features from the frequency branch and local features from the spatial branch. A total of 4 sets of features are encoded for the two modalities. The CMS-fusion module selectively integrates features from both modalities to generate enhanced global and local features for the target modality, which are mutually enhanced by the subsequent FS-fusion module. The neck stage mirrors the encoder stages but omits the downsample operation. As only target modality is decoded, the decoder stage (Stage 4-6) consist of the upsample operation, FSFE module, and FS-fusion module. The image \mathbf{I}_{tar}^{spa} recovered by the spatial branch is considered the high-quality image \mathbf{I}_{tar} for the target modality.

2.1 Frequency-Spatial Feature Extraction

Since similar structural features are distributed across different regions in an MR image, improving MRI reconstruction performance entails capturing global dependency and integrating similar features from various image locations. To address the limitation of CNNs in global modeling, we propose the *Frequency-Spatial Feature Extraction* (FSFE) module. Benefiting from the global property of Fourier transform, the FSFE module efficiently extracts local and global features through the spatial and frequency branches for both modalities.

Taking the i -th stage of the target modality as an example, the input spatial and frequency features (i.e., \mathbf{F}_{ts}^{i-1} and \mathbf{F}_{tf}^{i-1} obtained from the previous stage) are fed into the spatial and frequency branches, respectively,

$$\tilde{\mathbf{F}}_{ts}^i = f_{ts}^i(\mathbf{F}_{ts}^{i-1}), \tilde{\mathbf{F}}_{tf}^i = f_{tf}^i(\mathbf{F}_{tf}^{i-1}), \quad (1)$$

where $f_{ts}^i(\cdot)$ and $f_{tf}^i(\cdot)$ represent the networks for the spatial and frequency branches at the i -th stage for the target modality. As illustrated in Fig. 1 (a), the spatial branch consists of a cascade of residual convolutional blocks [8], extracting local features within the limited receptive field. In the frequency branch, the input feature \mathbf{F}_{tf}^{i-1} is first converted to the frequency domain via Fast Fourier Transform (FFT) [4]. Subsequently, the frequency feature is decomposed into amplitude and phase components (i.e., \mathcal{A}_{tf}^{i-1} , \mathcal{P}_{tf}^{i-1}), defined as:

$$\begin{aligned} \mathcal{A}_{tf}^{i-1} &= \sqrt{\mathcal{R}^2(\text{FFT}(\mathbf{F}_{tf}^{i-1})) + \mathcal{I}^2(\text{FFT}(\mathbf{F}_{tf}^{i-1}))}, \\ \mathcal{P}_{tf}^{i-1} &= \arctan \left[\mathcal{I}(\text{FFT}(\mathbf{F}_{tf}^{i-1})) / \mathcal{R}(\text{FFT}(\mathbf{F}_{tf}^{i-1})) \right], \end{aligned} \quad (2)$$

where $\mathcal{R}(\cdot)$ and $\mathcal{I}(\cdot)$ represent the real and imaginary parts of the frequency feature. According to Fourier transform theory [15], each pixel in the Fourier domain interacts with all pixels in the spatial domain, enabling image-size receptive

field and global feature extraction by extracting features from the amplitude and phase components in the Fourier domain. As depicted in Fig. 1 (a), we employ two convolutional layers with a residual connection to extract features \mathcal{A}_{tf}^i and \mathcal{P}_{tf}^i from the two components, respectively. The output $\tilde{\mathbf{F}}_{tf}^i$ of the frequency branch can be obtained via inverse Fourier transform (IFFT). Due to the global property of the Fourier transform, $\tilde{\mathbf{F}}_{tf}^i$ contains global information from the target modality, which is complementary to the local feature $\tilde{\mathbf{F}}_{ts}^i$.

Similarly, given the input features \mathbf{F}_{as}^{i-1} and \mathbf{F}_{af}^{i-1} for the i -th stage of the auxiliary modality, the FSFE module extracts local and global features $\tilde{\mathbf{F}}_{as}^i$ and $\tilde{\mathbf{F}}_{af}^i$ from the spatial and frequency branches, respectively.

2.2 Cross-Modal Frequency-Spatial Feature Fusion

For multi-contrast MR images, local features (i.e., $\tilde{\mathbf{F}}_{as}^i$ and $\tilde{\mathbf{F}}_{ts}^i$) and global features (i.e., $\tilde{\mathbf{F}}_{af}^i$ and $\tilde{\mathbf{F}}_{tf}^i$) are extracted by the spatial and frequency branches, respectively. Given the existence of complementary information across different modalities, as well as the spatial and frequency branches, there is a pressing need to design an effective fusion strategy to explore this complementary information. Toward this goal, we first fuse the multi-modal features for the different branches separately, and then integrate the spatial and frequency branches to generate a comprehensive feature representation for the target modality.

Cross-Modal Selective fusion (CMS-fusion): The CMS-fusion module generates enhanced local and global features by selectively integrating features from the respective branch in the auxiliary modality. Taking the frequency branch as an example, the features $\tilde{\mathbf{F}}_{tf}^i$ and $\tilde{\mathbf{F}}_{af}^i$ are processed by two convolutional layers, then combined to learn the pixel-wise modality complementary score s_{tf} and s_{af} for each modality:

$$[s_{tf}, s_{af}] = \text{sigmoid}(\text{conv}[\tilde{\mathbf{F}}_{tf}^i, \tilde{\mathbf{F}}_{af}^i]). \quad (3)$$

The complementary features from different modalities are fused to generate an enhanced global feature $\hat{\mathbf{F}}_{tf}^i$ for the target modality:

$$\hat{\mathbf{F}}_{tf}^i = s_{tf} * \tilde{\mathbf{F}}_{tf}^i + s_{af} * \tilde{\mathbf{F}}_{af}^i. \quad (4)$$

Similarly, the local features $\tilde{\mathbf{F}}_{as}^i$ and $\tilde{\mathbf{F}}_{ts}^i$ from the spatial branch are selectively integrated via the CMS-fusion module, producing the enhanced local feature $\hat{\mathbf{F}}_{ts}^i$ for the target modality. CMS-fusion in the spatial branch enables the exploration of detailed structural information from the auxiliary modality.

Frequency-Spatial fusion (FS-fusion):

By incorporating complementary information from the auxiliary modality, the enhanced features $\hat{\mathbf{F}}_{tf}^i$ and $\hat{\mathbf{F}}_{ts}^i$ offer a better representation of global information (corresponding to the basic structure and overall appearance) and local information (such as detailed structures), respectively. Given that global and local information are complementary and both are crucial for MRI reconstruction,

we introduce the FS-fusion module to adaptively merge the global frequency information and local spatial information.

As shown in Fig. 1 (c), the FS-fusion module employs the cross-attention mechanism to achieve mutual enhancement of the spatial and frequency branches. Aiming at enhancing the spatial information by exploring complementary frequency information, $\widehat{\mathbf{F}}_{ts}^i$ is used as the query (Q), and $\widehat{\mathbf{F}}_{tf}^i$ serves as both the key (K) and value (V). The integration process can be formalized as follows:

$$\overline{\mathbf{F}}_{ts}^i = f(\text{softmax}(\widehat{\mathbf{F}}_{tf}^i(\widehat{\mathbf{F}}_{ts}^i)^T/\sqrt{d})\widehat{\mathbf{F}}_{ts}^i) + \widehat{\mathbf{F}}_{ts}^i, \quad (5)$$

where f is the convolution function and $\overline{\mathbf{F}}_{ts}^i$ is the spatial feature enhanced by the global frequency information. Similarly, we obtain the enhanced frequency feature $\overline{\mathbf{F}}_{tf}^i$. The enhanced features $\overline{\mathbf{F}}_{ts}^i$ and $\overline{\mathbf{F}}_{tf}^i$ are integrated into \mathbf{F}_{ts}^i using pixel-wise selective fusion, as described in Eq. (3) and Eq. (4).

2.3 Loss Function

FSMNet reconstructs target modality image \mathbf{I}_{tar}^{spa} and \mathbf{I}_{tar}^{fre} from the spatial and frequency branch, respectively. The model is supervised by pixel-level loss $\mathcal{L}_{\text{pixel}}$ and frequency-level loss \mathcal{L}_{fre} , defined as:

$$\mathcal{L}_{\text{pixel}} = \|\mathbf{I}_{tar}^{spa} - \mathbf{I}_{tar}^{full}\|_1 + \|\mathbf{I}_{tar}^{fre} - \mathbf{I}_{tar}^{full}\|_1, \quad (6)$$

$$\mathcal{L}_{\text{fre}} = \|\mathcal{A}(\text{FFT}(\mathbf{I}_{tar}^{fre})) - \mathcal{A}(\text{FFT}(\mathbf{I}_{tar}^{full}))\|_1 + \|\mathcal{P}(\text{FFT}(\mathbf{I}_{tar}^{fre})) - \mathcal{P}(\text{FFT}(\mathbf{I}_{tar}^{full}))\|_1, \quad (7)$$

where \mathbf{I}_{tar}^{full} corresponds to the fully-sampled target modality image, serving as the ground-truth. $\mathcal{A}(\cdot)$ and $\mathcal{P}(\cdot)$ denote the amplitude and phase components in the Fourier domain. The total loss is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{pixel}} + \lambda * \mathcal{L}_{\text{fre}}, \quad (8)$$

where λ is the trade-off coefficient and is empirically set as 0.01.

3 Experiments

3.1 Datasets and Implementation Details

Dataset description: We evaluate our method on the BraTS [14] and fastMRI [22] datasets. For BraTS dataset, we use T1-weighted (T1WI) and T2-weighted (T2WI) brain MRI volumes, with T1WI supporting T2WI acceleration. We randomly select 100 patients, sample slices from the 3D volumes, and split the dataset 3:1 for training and testing. The fastMRI dataset consists of single-coil proton density-weighted (PDWI) and fat-suppressed proton density-weighted (FS-PDWI) knee MRI volumes. Following [5], we filter out 227 PDWI and FS-PDWI pairs for training and 24 pairs for testing, using PDWI as the auxiliary and FS-PDWI as the target modality. The reconstruction performance is evaluated under acceleration factors (AF) of $4\times$ and $8\times$ using the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

Implementation details: We implement FSMNet and all comparison methods on RTX 3090 GPUs using the PyTorch library [16]. The image resolution is

Table 1. Quantitative results on the BraTS and fastMRI datasets with different acceleration factors. We report mean \pm std for the PSNR and SSIM metrics.

BraTS		$4\times$		$8\times$	
Method	Year	PSNR	SSIM	PSNR	SSIM
Zero-filling [1]	JMRI'01	30.04 \pm 1.51	0.748 \pm 0.034	26.58 \pm 1.49	0.673 \pm 0.036
MDUNet [18]	TBME'18	37.94 \pm 1.66	0.975 \pm 0.006	35.19 \pm 1.64	0.960 \pm 0.009
MINet [5]	MICCAI'21	38.26 \pm 1.74	0.976 \pm 0.006	35.23 \pm 1.72	0.961 \pm 0.009
MCCA [10]	JBHI'23	38.03 \pm 1.68	0.975 \pm 0.006	35.37 \pm 1.66	0.962 \pm 0.009
SwinIR [11]	CVPR'21	37.87 \pm 1.73	0.974 \pm 0.006	34.95 \pm 1.72	0.960 \pm 0.009
MTrans [6]	TMI'22	36.02 \pm 1.67	0.962 \pm 0.007	34.81 \pm 1.57	0.957 \pm 0.009
DCAMSR [9]	MICCAI'23	38.60 \pm 1.75	0.978 \pm 0.006	35.99 \pm 1.74	0.965 \pm 0.009
Ours (FSMNet)	-	41.76\pm1.88	0.986\pm0.004	38.60\pm1.85	0.977\pm0.006
fastMRI		$4\times$		$8\times$	
Method	Year	PSNR	SSIM	PSNR	SSIM
Zero-filling [1]	JMRI'01	24.5 \pm 1.37	0.442 \pm 0.10	22.9 \pm 1.25	0.369 \pm 0.12
MDUNet [18]	TBME'18	28.6 \pm 1.00	0.600 \pm 0.05	27.9 \pm 0.86	0.544 \pm 0.05
MINet [5]	MICCAI'21	29.4 \pm 1.88	0.639 \pm 0.06	28.1 \pm 1.74	0.563 \pm 0.08
MCCA [10]	JBHI'23	29.4 \pm 1.87	0.637 \pm 0.06	28.2 \pm 1.75	0.562 \pm 0.08
SwinIR [11]	CVPR'21	29.4 \pm 1.87	0.636 \pm 0.06	28.1 \pm 1.74	0.560 \pm 0.08
MTrans [6]	TMI'22	29.0 \pm 1.79	0.619 \pm 0.06	27.3 \pm 1.68	0.526 \pm 0.08
DCAMSR [9]	MICCAI'23	29.4 \pm 1.87	0.637 \pm 0.06	28.4 \pm 1.79	0.569 \pm 0.08
Ours (FSMNet)	-	29.7\pm1.93	0.646\pm0.07	28.5\pm1.80	0.572\pm0.08

240 \times 240 and 320 \times 320 for the BraTS and fastMRI datasets, respectively. FSMNet is trained using AdamW optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The training process for FSMNet consists of 100k iterations with a batch size of 4. The learning rate is initially set to 1×10^{-4} and decays by a factor of 0.1 every 20k iterations. The comparison methods are implemented following the original configurations described in their respective papers [18,5,10,11,6,9].

3.2 Experimental Results

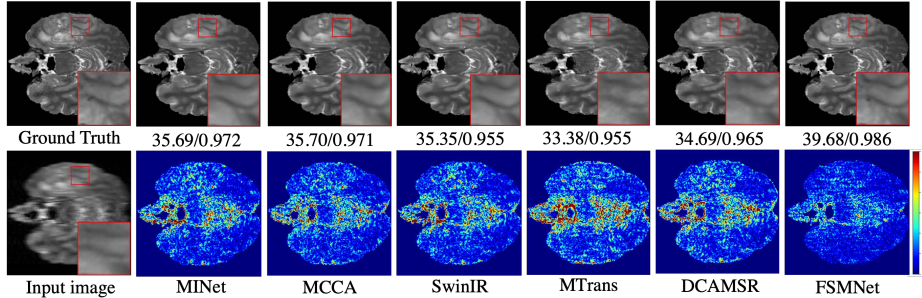
Comparison with state-of-the-art methods: We conduct a comprehensive comparison between FSMNet and state-of-the-art MCMR methods on both datasets. As shown in Table 1, FSMNet consistently outperforms existing MCMR methods [18,5,10,11,6,9] across various acceleration factors on both the BraTS and fastMRI datasets.

Specifically, when compared with the current state-of-the-art method (i.e., DCAMSR [9]), our method improves the PSNR by 3.16 dB and 2.61 dB under $4\times$ and $8\times$ AF on the BraTS dataset, respectively. We further illustrate the reconstructed images and error maps of different methods in Fig. 2. For the example with $4\times$ AF on the BraTS dataset, FSMNet recovers more image details and leads to a smaller error map. More examples are provided in the supplementary material. The visualization results align with the improvement in the PSNR and SSIM metrics shown in Table 1.

Ablation Study: For the T2WI MRI reconstruction on the BraTS dataset with different AF, we conduct an ablation study to evaluate the contributions of our proposed FSFE, CMS-fusion, and FS-fusion modules. As shown in Table 2, compared with the vanilla variant (multi-modal UNet with concat fusion, 1st

Table 2. Ablation study on the BraTS dataset with different AF.

FSFE	CMS-fusion	FS-fusion	4×		8×	
			PSNR	SSIM	PSNR	SSIM
			37.94±1.66	0.975±0.006	35.19±1.64	0.960±0.009
✓			41.05±1.80	0.985±0.004	37.97±1.78	0.974±0.007
✓	✓		41.39±1.83	0.986±0.004	38.27±1.78	0.975±0.007
✓	✓	✓	41.76±1.88	0.986±0.004	38.60±1.85	0.977±0.006

**Fig. 2. Qualitative visualization** of the reconstructed images (1st row) and error maps (2nd row) for different MCMR methods with 4× AF on the BraTS dataset. Additional qualitative results are provided in the supplementary material.

row), the reconstruction performance significantly improves with the introduction of the FSFE module (2nd row). Specifically, the PSNR improves by 3.11 dB and 2.78 dB under 4× and 8× AF, respectively. These results indicate that the global frequency and local spatial features extracted by the FSFE module can effectively enhance the feature representation and benefit MRI reconstruction. As demonstrated in the 3-4 rows of Table 2, we observe performance degradation when replacing the CMS-fusion and FS-fusion with the simple summation fusion method. These results suggest that the proposed CMS-fusion and FS-fusion modules can effectively explore complementary information across different modalities and branches. Moreover, a feature analysis for the FS-fusion module is provided in the supplementary material.

4 Conclusion

For the accelerated MCMR task, we propose a novel FSMNet to effectively explore and integrate the global frequency feature and local spatial feature across different modalities. In our proposed FSFE module, the spatial branch captures local dependency, while the frequency branch efficiently captures global dependency and achieves image-size receptive field via Fourier transform. The CMS-fusion is devised to generate enhanced local and global features by adaptively integrating complementary information from the auxiliary modality. The enhanced local and global features are further integrated via the FS-fusion module to produce a comprehensive feature representation for the target modality.

Extensive experiments on the BraTS and fastMRI datasets demonstrate the superiority of FSMNet for the MCMR task under various acceleration factors.

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

References

1. Bernstein, M.A., Fain, S.B., Riederer, S.J.: Effect of windowing and zero-filled reconstruction of mri data on spatial resolution and acquisition strategy. *Journal of magnetic resonance imaging: an official journal of the international society for magnetic resonance in medicine* **14**(3), 270–280 (2001) [7](#)
2. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020) [2](#)
3. Du, T., Zhang, H., Li, Y., Pickup, S., Rosen, M., Zhou, R., Song, H.K., Fan, Y.: Adaptive convolutional neural networks for accelerating magnetic resonance imaging via k-space data interpolation. *Medical image analysis* **72**, 102098 (2021) [2](#)
4. Elliott, D., Rao, K.: Fast fourier transform and convolution algorithms (1982) [4](#)
5. Feng, C.M., Fu, H., Yuan, S., Xu, Y.: Multi-contrast mri super-resolution via a multi-stage integration network. 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. 140–149. Springer (2021) [2](#), [6](#), [7](#)
6. Feng, C.M., Yan, Y., Chen, G., Xu, Y., Hu, Y., Shao, L., Fu, H.: Multi-modal transformer for accelerated mr imaging. *IEEE Transactions on Medical Imaging* (2022) [2](#), [7](#)
7. Han, Y., Sunwoo, L., Ye, J.C.: k-space deep learning for accelerated mri. *IEEE transactions on medical imaging* **39**(2), 377–386 (2019) [2](#)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016) [4](#)
9. Huang, S., Li, J., Mei, L., Zhang, T., Chen, Z., Dong, Y., Dong, L., Liu, S., Lyu, M.: Accurate multi-contrast mri super-resolution via a dual cross-attention transformer network. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 313–322. Springer (2023) [2](#), [7](#)
10. Li, B., Hu, W., Feng, C.M., Li, Y., Liu, Z., Xu, Y.: Multi-contrast complementary learning for accelerated mr imaging. *IEEE Journal of Biomedical and Health Informatics* (2023) [2](#), [7](#)
11. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 1833–1844 (2021) [7](#)
12. Lyu, J., Sui, B., Wang, C., Tian, Y., Dou, Q., Qin, J.: Dudocaf: Dual-domain cross-attention fusion with recurrent transformer for fast multi-contrast mr imaging. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 474–484. Springer (2022) [2](#)

13. Mao, X., Liu, Y., Shen, W., Li, Q., Wang, Y.: Deep residual fourier transformation for single image deblurring. arXiv preprint arXiv:2111.11745 **2**(3), 5 (2021) [2](#)
14. 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) [3](#), [6](#)
15. Nussbaumer, H.J., Nussbaumer, H.J.: The fast Fourier transform. Springer (1982) [4](#)
16. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* **32** (2019) [6](#)
17. Tsao, J., Kozerke, S.: Mri temporal acceleration techniques. *Journal of Magnetic Resonance Imaging* **36**(3), 543–560 (2012) [2](#)
18. Xiang, L., Chen, Y., Chang, W., Zhan, Y., Lin, W., Wang, Q., Shen, D.: Deep-learning-based multi-modal fusion for fast mr reconstruction. *IEEE Transactions on Biomedical Engineering* **66**(7), 2105–2114 (2018) [7](#)
19. Xing, X., Chen, Z., Hou, Y., Yuan, Y.: Gradient modulated contrastive distillation of low-rank multi-modal knowledge for disease diagnosis. *Medical Image Analysis* **88**, 102874 (2023) [2](#)
20. Xing, X., Chen, Z., Zhu, M., Hou, Y., Gao, Z., Yuan, Y.: Discrepancy and gradient-guided multi-modal knowledge distillation for pathological glioma grading. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 636–646. Springer (2022) [2](#)
21. Xing, X., Zhu, M., Chen, Z., Yuan, Y.: Comprehensive learning and adaptive teaching: Distilling multi-modal knowledge for pathological glioma grading. *Medical Image Analysis* **91**, 102990 (2024) [2](#)
22. Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M.J., Defazio, A., Stern, R., Johnson, P., Bruno, M., et al.: fastmri: An open dataset and benchmarks for accelerated mri. arXiv preprint arXiv:1811.08839 (2018) [3](#), [6](#)
23. Zhou, B., Zhou, S.K.: Dudornet: learning a dual-domain recurrent network for fast mri reconstruction with deep t1 prior. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 4273–4282 (2020) [2](#)
24. Zhou, M., Huang, J., Guo, C.L., Li, C.: Fourmer: An efficient global modeling paradigm for image restoration. In: *International Conference on Machine Learning*. pp. 42589–42601. PMLR (2023) [2](#)
25. Zhuo, J., Gullapalli, R.P.: Mr artifacts, safety, and quality control. *Radiographics* **26**(1), 275–297 (2006) [2](#)