

MFTN: Multi-Level Feature Transfer Network Based on MRI-Transformer for MR Image Super-Resolution

Shuying Huang^{1*}, Ge Chen^{2*}, Yong Yang^{3†}, Xiaozheng Wang⁴, Chenbin Liang²

¹ School of Software, Tiangong University, Tianjin, China

² School of Electronics and Information Engineering, Tiangong University, Tianjin, China

³ School of Computer Science and Technology, Tiangong University, Tianjin, China

⁴ School of Control Science and Engineering, Tiangong University, Tianjin, China

huangshuying@tiangong.edu.cn, 2131070961@tiangong.edu.cn, greatyangy@126.com, xiaozhengwang95@gmail.com, 15802268381@163.com

Abstract

Due to the unique environment and inherent properties of magnetic resonance imaging (MRI) instruments, MR images typically have lower resolution. Therefore, improving the resolution of MR images is beneficial for assisting doctors in diagnosing the condition. Currently, the existing MR image super-resolution (SR) methods still have the problem of insufficient detail reconstruction. To overcome this issue, this paper proposes a multi-level feature transfer network (MFTN) based on MRI-Transformer to realize SR of low-resolution MRI data. MFTN consists of a multi-scale feature reconstruction network (MFRN) and a multi-level feature extraction branch (MFEB). MFRN is constructed as a pyramid structure to gradually reconstruct image features at different scales by integrating the features obtained from MFEB, and MFEB is constructed to provide detail information at different scales for low resolution MR image SR reconstruction by constructing multiple MRI-Transformer modules. Each MRI-Transformer module is designed to learn the transfer features from the reference image by establishing feature correlations between the reference image and low-resolution MR image. In addition, a contrast learning constraint item is added to the loss function to enhance the texture details of the SR image. A large number of experiments show that our network can effectively reconstruct high-quality MR Images and achieves better performance compared to some state-of-the-art methods. The source code of this work will be released on GitHub.

Introduction

Compared to CT and PET imaging, MRI is more effective in detecting early lesions that are not easily noticeable and has become the most efficient imaging diagnostic method for diseases such as the brain and spinal cord. Furthermore, MRI is nonionizing and harmless to the human body. However, MRI scans require longer acquisition times to obtain high-resolution MR images, and due to limitations in device hardware and post-processing, an increase in image resolution will also reduce the signal-to-noise ratio of the image. This will also increase the consumption of hardware facilities and shorten their lifespan, leading to an increase in the

cost of acquiring MR images. To obtain high-quality MR images, two solutions are usually adopted. One is to introduce more advanced equipment. However, state-of-the-art equipment is expensive and not widely accessible. The second approach is to utilize computer vision technology to achieve SR reconstruction of MR images, ensuring richer details are reconstructed and restored while improving the spatial resolution of the image.

Currently SR techniques for MR images can mainly be divided into two categories: single-contrast SR methods and multi-contrast SR methods. Single-contrast SR methods only utilizes the information of the single modality degraded image itself to improve its resolution, but the obtained feature information is limited and singular. The multi-contrast SR methods restore more information by utilizing high-resolution (HR) images from another modality as reference images, thereby reconstructing HR images with more texture details. Currently, most of the SR works for MR images use the multi-contrast reconstruction way, which achieve the reconstruction of one modality image by utilizing two different modalities of MR images. Figure 1 shows two sets of different modality MR images with the same anatomical structure, such as fluid-attenuated inversion recovery (FLAIR) and T2-weighted images, proton density-weighted (PDWI) and fat-suppressed proton density-weighted (FS-PDWI) images, which are generated by adjusting the parameter settings of the MRI machine. From the figure, it can be seen that the two images have the same structure, but emphasize different details. Therefore, one of the images can be used as a reference image for SR reconstruction of the other image.

In the recent study, MSPN (Lyu et al. 2020) introduced neural networks of different levels to perform multi-contrast image fusion. Although this method generates more information, there is also a problem of insufficient detail reconstruction due to a lack of attention to long-distance information. MINet (Feng et al. 2021a) adopted multi-stage integrated network to carry out multi-contrast MR Image SR, but it reconstructed features in HR space without reconstructing intermediate scale features, resulting in the problem of missing details. WavTrans (Li et al. 2022b) introduced wavelet transform into the network and designed

*These authors contributed equally.

†Corresponding author

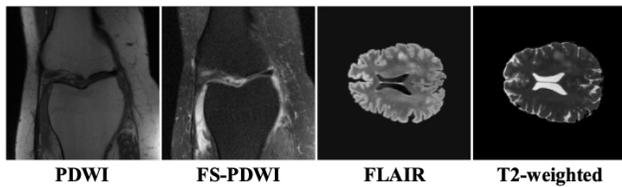


Figure 1: Examples of two modalities of MR images.

a new residual cross-attention swin-transformer network to realize multi-contrast MR Image SR. However, due to performing feature reconstruction in LR space, the reconstructed features in HR space are insufficient. McMRSR (Li et al. 2022a) indicated that feature information at different scales is different, and it is worth further research on how to better utilize features at different scales to supplement information for the reconstructed image. In addition, during the data preprocessing process, there is a subpixel displacement between the low-resolution (LR) MR image and the reference image, which makes it difficult to ensure that the features extracted from the reference image are fully matched with the same positional features in the LR MR image. This can lead to artifacts and edge blurring in the reconstructed image. Therefore, how to reduce the impact of subpixel displacement on the accuracy of feature extraction is also crucial for multi-contrast SR reconstruction.

To address the above issues, we design a multi-level feature transfer network (MFTN) based on MRI-Transformer to realize SR reconstruction of LR MR images. In MFTN, to reconstruct richer features, a multi-scale feature reconstruction network (MFRN) is constructed to learn multiply scale features and reconstruct a HR MR image, which solves the problem of insufficient feature reconstruction in a single scale network. To provide supplementary information for feature reconstruction in MFRN, a multi-level feature extraction branch (MFEB) is constructed. In MFEB, an MRI-Transformer module is designed to extract transfer features from the reference image, and it can reduce the impact of sub-pixel displacement on feature extraction by paying attention to long-distance features. The contributions of this work are as follows:

- An MRI-Transformer based MFTN, including two branches: MFRN and MFEB, is proposed to achieve SR reconstruction of LR MR images by learning supplementary information from reference images.
- A pyramid structured MFRN is constructed to gradually reconstruct image features at different scales and obtain HR MR images by integrating transfer features from the reference image.
- In view of the limitation of feature information in a single modality image, an MFEB using another modality image as a reference image is constructed to provide supplementary information for MFRN. In MFEB, an MRI-Transformer module is designed to learn transfer features by establishing feature correlations between the LR MR image and the reference image.
- To improve the visual effects of reconstructed MR im-

ages, a perceptual contrastive loss is defined and combined with commonly used loss functions to achieve network training.

Proposed Approach

In this section, a MFTN based on MRI-Transformer for MR image SR is proposed by learning transfer features from a reference image. As shown in Figure 2, MFTN consists of MFRN and MFEB, which are respectively used to reconstruct a HR MR image and learn transfer features from a reference image. The reference images and LR MR images are collected using the same sensor with different parameter settings on the same anatomical region, and images with higher resolution are usually chosen as reference images, which are used to assist in SR reconstruction of LR MR images. The specific details of each module in MFTN will be described in the following sections.

Multi-scale Feature Reconstruction Network (MFRN)

To reconstruct rich detail features, a pyramid structured network called MFRN is constructed to learn features at different scales and achieve SR reconstruction of LR MR images. The specific structure is described below.

Firstly, a simple convolutional layer is used to extract shallow features from the input LR MR image. Then, three feature integration blocks are designed to achieve the reconstruction of three scale features. In each feature integration block, the current scale features are first concatenated with the same level transfer features from MFEB, and then sent to a feature fusion (FF) block consisting of a 3×3 convolution and batch normalization (BN) layer for feature integration. In addition, a residual structure is adopted to achieve the reuse of shallow features at the current scale. Finally, n residual blocks (RBS, $n=4$) are used to reconstruct the current scale features. Between two adjacent feature integration blocks, a scale extension layer containing a convolutional layer and Pixelshuffle operation, called CP, is constructed to increase the size of the feature maps, and then these feature maps are fed into the next feature integration block.

Finally, the reconstructed feature maps at three scales are enlarged into HR space and concatenated together into n RBs to achieve the final HR MR image.

Multi-level Feature Extraction Branch (MFEB)

MFEB mainly consists of three multifold feature extractors and three MRI-transformer modules. The multifold feature extractors are constructed to extract different levels of shallow features from the input MR image. And subsequent MRI-transformer modules take features of different levels as inputs to extract transfer features, which are transferred as supplementary features to MFRN. Below, we will provide a detailed introduction to the structures of multifold feature extractor and MRI-transformer module.

1) Multifold Feature Extractor

The multifold feature extractor similar to VGG architecture is constructed to extract three levels of features from an input MR image, as shown in Figure 3. The extractor consists of three convolutional blocks, where the first two convolutional

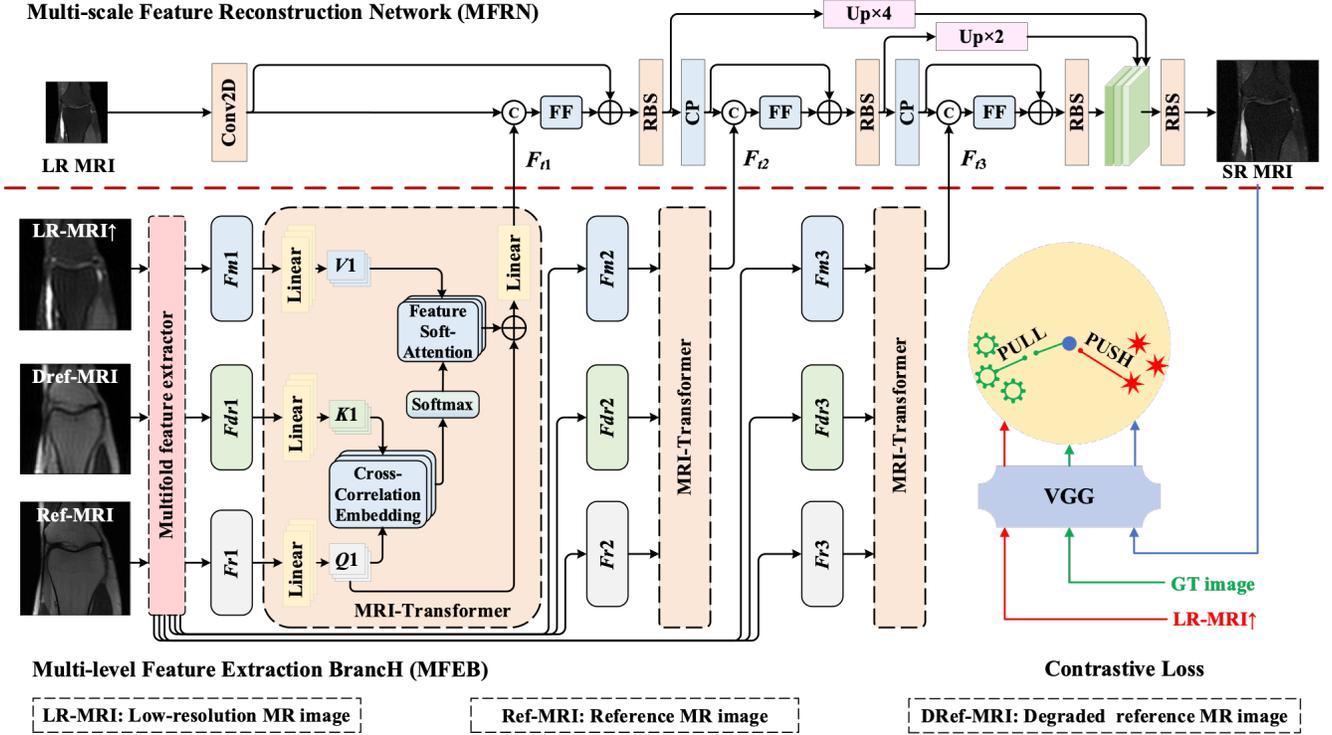


Figure 2: Architecture of MFTN.

blocks contain two convolutional layers composed of a convolutional operation, a BN layer, and a ReLU function, and the last convolutional block contains only one convolutional layer. In convolutional blocks, the Maxpooling operation is adopted to reduce the size of the feature maps.

Due to the fact that MFEB learns transfer features from the reference MR image, which has a higher resolution than the input LR MR image, it is necessary to establish a feature distribution relationship between the HR space and the LR space. Therefore, the reference image need be degraded through down-sampling and up-sampling operations to obtain a degraded reference image with a similar distribution to the input LR image. Thus, three multifold feature extractors are required to extract shallow features from three input images, namely the up-sampled LR MR image, reference MR image, and degraded reference MR image. And the same level shallow features from three extractors are used to define Query (Q), Key (K), and Value (V) for one of the subsequent MRI Transformer modules. Taking feature extraction for the up-sampling LR MR image I_{UM} as an example, the execution process of the multi-fold feature extractor is as follows.

$$CL(I_{UM}) = ReLU(BN(Conv_{3 \times 3}(I_{UM}))) \quad (1)$$

$$Fm1 = CB(I_{UM}) = CL(CL(I_{UM})) \quad (2)$$

$$Fm2 = CB(Maxpool(Fm1)) \quad (3)$$

$$Fm3 = CL(Maxpool(Fm2)) \quad (4)$$

where CL represents a convolutional layer containing a 3×3 convolutional operation $Conv_{3 \times 3}$, a BN layer, and a

$ReLU$ function. CB represents a convolutional block containing two convolutional layers. $Maxpool$ represents the max-pooling operation. $Fm1$, $Fm2$, and $Fm3$ represent three levels of feature maps extracted from the upsampling LR MR image.

Using the same operation, the three levels of feature maps extracted from the reference MR image can be represented as $Fr1$, $Fr2$, and $Fr3$, while the three levels of features extracted from the degraded reference MR image can be represented as $Fdr1$, $Fdr2$, and $Fdr3$.

MRI-Transformer Module

The proposed MRI-Transformer module can extract transfer features for feature reconstruction in MFRN by establishing feature correlations and long-distance dependencies between the input LR image and the reference image. The three MRI-Transformer modules are used to receive three levels of feature maps from three multifold feature extractors and learn corresponding levels of transfer features.

Taking the generation of first level transfer features as an example, the processing process of one MRI-Transformer module is described below. Because the first level of transfer features is transferred to the minimum scale feature reconstruction stage in MFRN, the first MRI-Transformer module takes the feature maps $Fr1$, $Fdr1$, and $Fm1$ as inputs. First, these feature maps are mapped into the Query1 ($Q1$), Key1 ($K1$), and Value1 ($V1$), respectively. Then, to establish the distribution relationship between the HR space and the LR space, the multi-head attention mechanism (Vaswani et al.

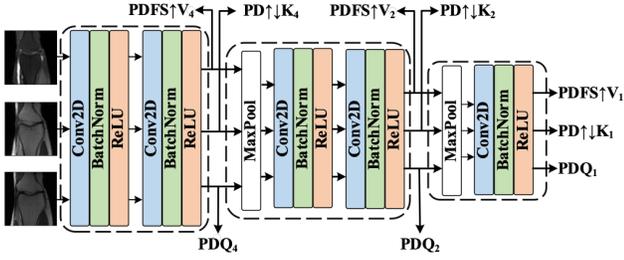


Figure 3: Architecture of Multifold feature extractor.

Dataset	FastMRI			
Scale	2×		4×	
Methods	Metric			
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
EDSR (Lim et al. 2017)	26.70	0.512	18.41	0.208
MCSR (Zeng et al. 2018)	28.50	0.661	21.92	0.463
NEU (Neu et al. 2020)	29.50	0.682	28.20	0.574
MINet (Feng et al. 2021)	31.80	0.709	29.80	0.601
MSDT (Zou et al. 2023)	31.98	0.713	30.38	0.615
Ours	32.17	0.736	31.04	0.622

Table 1: Average objective results of 2× and 4× SR of all comparison methods on the FastMRI dataset.

2017) is used to calculate a cross correlation matrix for the two feature tensors Q_1 and K_1 from the reference image. Next, a soft-attention mechanism is used for the feature tensors V_1 and cross correlation matrix to obtain the feature correlation matrix between the input LR image and the reference image. Finally, a residual link is established between the feature correlation matrix and feature tensors Q_1 to obtain the transfer features F_{t1} from the reference image.

As described above, the next second and third MRI-Transformer modules obtain transfer features F_{t2} and F_{t3} at the other two scales, respectively, which are sent to MFRN to provide supplementary information.

Loss Function

To better guide network training, a joint loss function containing four loss terms is defined: reconstruction loss, perceptual loss, perceptual contrastive loss, and SSIM loss.

Reconstruction Loss L1 loss is employed to evaluate the similarity between the reconstruction result and ground-truth (GT) image. It can be represented as follows:

$$L_{rec} = \frac{1}{CWH} \|I_{gt} - I_{sr}\|^1 \quad (5)$$

where I_{gt} denotes the GT image, and I_{sr} denotes the reconstruction result. W and H is the width and height of the images, and C is the number of channels.

Dataset	AXA			
Scale	2×		4×	
Methods	Metric			
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
EDSR (Lim et al. 2017)	29.71	0.803	24.32	0.628
MCSR (Zeng et al. 2018)	33.10	0.841	29.96	0.724
NEU (Neu et al. 2020)	34.77	0.867	30.24	0.746
MINet (Feng et al. 2021)	36.41	0.916	33.38	0.831
MSDT (Zou et al. 2023)	36.17	0.914	33.41	0.834
Ours	36.79	0.923	33.58	0.852

Table 2: Average objective results of 2× and 4× SR of all comparison methods on the AXA dataset.

Perceptual Loss Perception loss is the measurement of the distance between two images in the feature space, and is more in line with the human eye’s perception of image quality. In this paper, the pre-trained VGG network (Simonyan and Zisserman 2014) is trained on RGB images to obtain features of the reconstruction result and GT image. It can be represented as follows:

$$L_{per} = \frac{1}{C_i W_i H_i} \|y_i(I_{gt}) - y_i(I_{sr})\|^2 \quad (6)$$

where $y_i(\cdot)$ denotes the i th layer’s feature map of VGG-19.

Perceptual Contrastive Loss Inspired by the perceptual loss in AECR-Net (Wu et al. 2021), we define a perceptual contrastive loss to enhance the network’s discriminative capability by incorporating the idea of perceptual loss into the contrastive loss. In the latent feature space, the common intermediate features from the pre-trained VGG19 model are used to participate in the calculation of perceptual contrastive loss. The perceptual contrastive loss can be represented as follows:

$$L_{com} = \sum_{i=1}^n \sum_{j=1}^m \omega_j \cdot \frac{\|y_j(I_{sr}^i), y_j(I_{st}^i)\|^1}{\sum_{k=1}^K \|y_j(I_{sr}^i), y_j(I_{tr}^k)\|^1} \quad (7)$$

where m is the number of total hidden layers, and w_j is a weighted term to balance the shallow and deep features. n represents the total number of reconstructed images and K represents the total number of positive and negative samples for each image under each hidden layer.

SSIM Loss To achieve the reconstruction results with better visual effects in terms of brightness, contrast, and structure, in terms of brightness, contrast, and structure, we introduce the SSIM loss, which is represented as follows:

$$L_{ssim} = 1 - \frac{(2\mu_{I_{gt}}\mu_{I_{sr}} + c_1)(2\sigma_{I_{gt}I_{sr}} + c_2)}{(\mu_{I_{gt}}^2 + \mu_{I_{sr}}^2 + c_1)(\sigma_{I_{gt}}^2 + \sigma_{I_{sr}}^2 + c_2)} \quad (8)$$

Total Loss the joint loss function of our proposed network is defined as:

$$L_{all} = \alpha L_{rec} + \beta L_{per} + \delta L_{com} + \varepsilon L_{ssim} \quad (9)$$

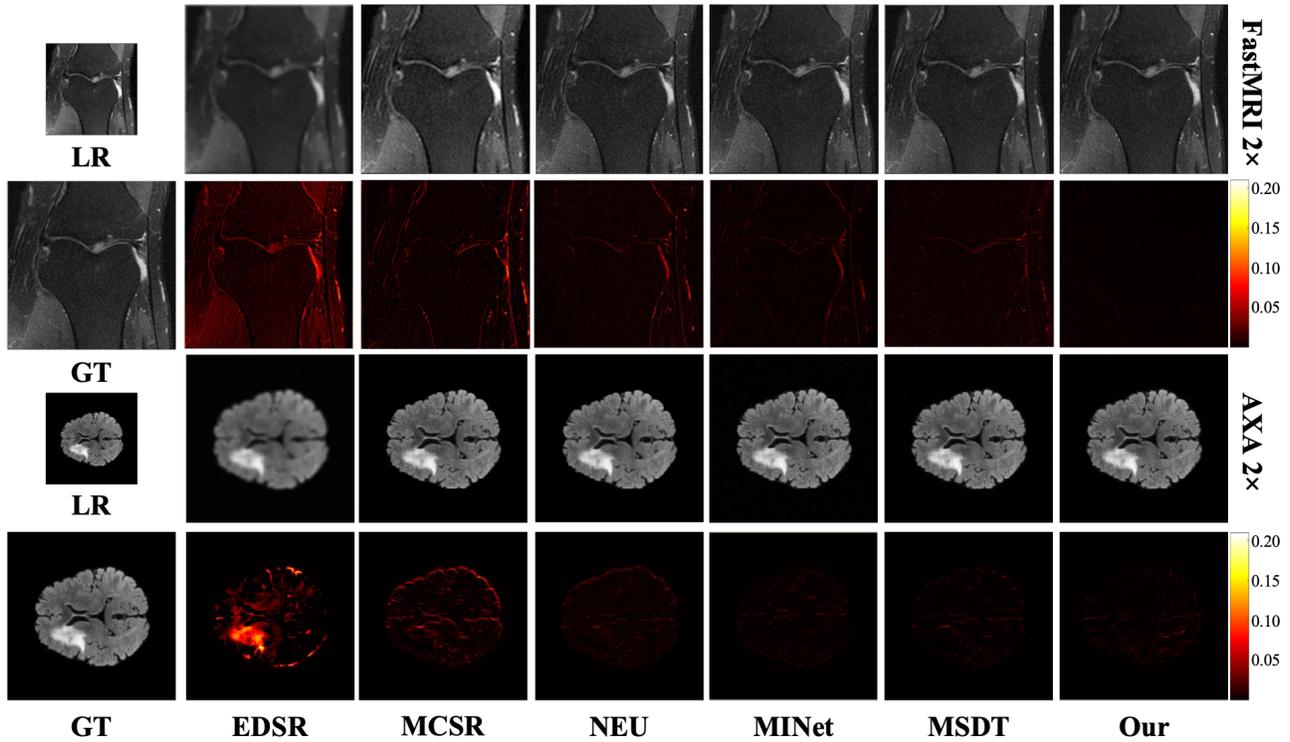


Figure 4: Qualitative results of different SR reconstruction methods on FastMRI and AXA dataset. The first/third row is the SR reconstructed images obtained by different methods, and the second/fourth row is the corresponding residual maps.

where α , β , δ , ε and are the hyperparameters, we set $\alpha = 1.0$, $\beta = 0.1$, $\delta = 0.01$, and $\varepsilon = 0.001$ based on experience.

Experimental Results and Analysis

Datasets and Baselines

To verify the effectiveness of the proposed method, we conduct extensive experiments on a public dataset and a self-built dataset. The public dataset (Zbontar et al. 2018) (the largest publicly available MRI dataset) is named FastMRI and the self-built dataset is named AXA. For the FastMRI dataset, we follow the approach in LMKSP (Xuan et al. 2020) and select 227 pairs and 24 pairs of PD and PDFS for training and validation, respectively. AXA dataset is constructed from data collected from 50 patients using a 3T Siemens Magnetom Skyra system. Each MRI scan collects data in T2 and Flair fully sampled k-spaces (TRFlair = 9000ms, TEFlair = 120ms, TRT2 = 5725ms, TET2 = 100ms). The collection of clinical datasets has been approved by the institutional review board. The AXA dataset is divided into training/validation/test sets in a ratio of 7:1:2. In the FastMRI experiments, we used PD images to guide SR reconstruction of PDFS images, while in the AXA experiments, we used T2 images to guide SR reconstruction of Flair images.

Experimental Setup

We implement our model using one NVIDIA RTX A6000 GPU with single-card 48GB memory. Our model is trained

using the Adam optimizer for 50 epochs, and a learning rate is set to $1e-5$. For the FastMRI dataset, the evaluation index values of all comparison methods are obtained from MSDT (Zou et al. 2023). For AXA dataset, the indicators of MSDT are obtained by direct testing it, and the other comparison methods are retrained and tested to obtain the corresponding indicators.

Objective and Subjective Comparison

In order to quantitatively evaluate the performance of comparison methods, two commonly used indexes, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), are adopted in the experiments. The higher the PSNR and SSIM values, the better the performance of the method. In order to demonstrate the validity of the proposed method, we objectively compare it with some state-of-the-art methods, including EDSR (Lim et al. 2017), MCSR (Zeng et al. 2018), NEU (Neubert et al. 2020), MINet (Feng et al. 2021a) and MSDT (Zou et al. 2023), and the results obtained by the comparison method are all from the official specifications. Table 1 and Table 2 show the average measure results of $2\times$ and $4\times$ SR reconstruction on the FastMRI and AXA datasets for all comparison methods. As can be seen from the tables, all quantitative indicators obtained by the proposed method are significantly higher than those obtained by other comparison methods. This indicates that the proposed method can recover images closer to GT images. Figure 4 and Figure 5 show the visualization results of $2\times$ and $4\times$ SR recon-

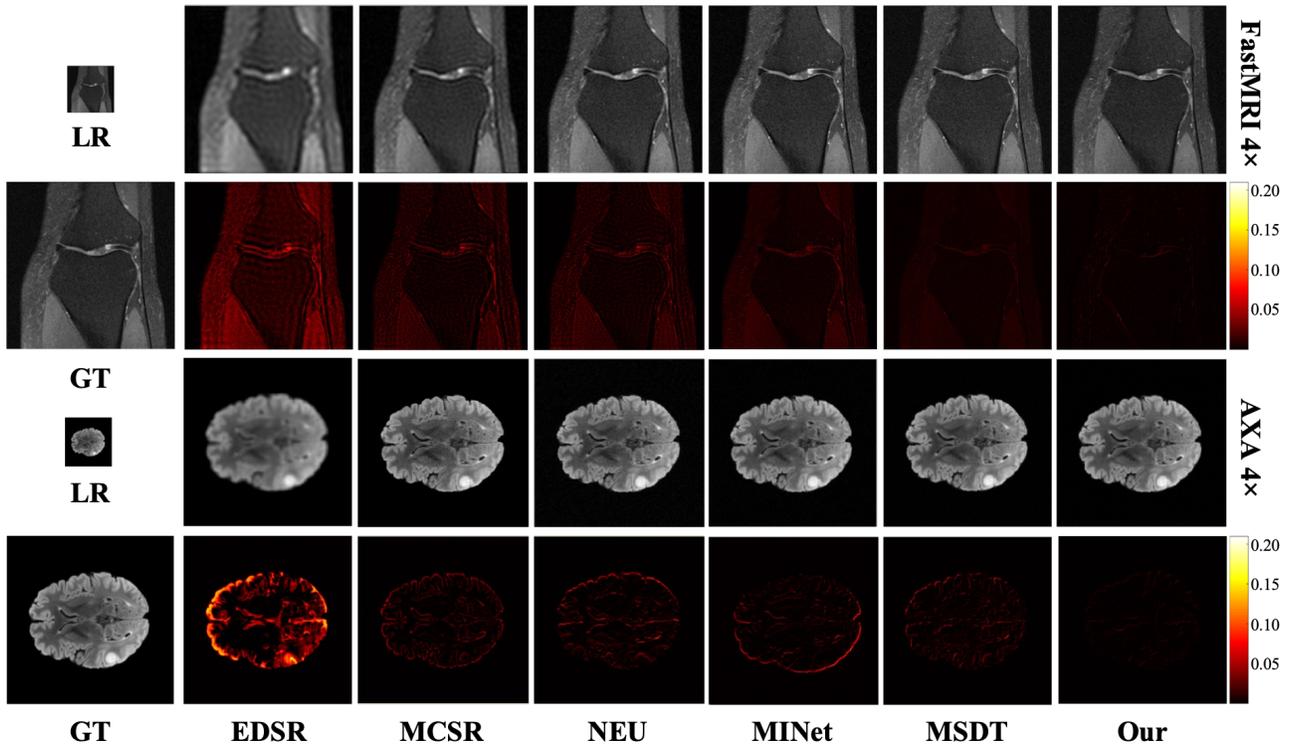


Figure 5: Qualitative results of different SR reconstruction methods on FastMRI and AXA dataset. The first/third row is the SR reconstructed images obtained by different methods, and the second/fourth row is the corresponding residual maps.

<i>MRIT/RT</i>	<i>w/o comp</i>	PSNR↑	SSIM↑
<i>MRIT</i>	<i>w comp</i>	32.17	0.736
	<i>o comp</i>	32.09	0.731
<i>RT</i>	<i>w comp</i>	30.34	0.613
	<i>o comp</i>	30.12	0.608

Table 3: Average objective results of $2\times$ SR of ablation experiments about MRI-Transformer modules and Comparative learning loss on the FastMRI dataset.

struction on two images from two datasets, respectively. To facilitate the observation of differences between images, the residual images between the reconstruction results and corresponding GT images are calculated and displayed. From the figures, it can be clearly seen that due to the fact that EDSR only uses one LR image to achieve image reconstruction, the reconstructed image contains fewer details and has blurred edges, and the residual images obtained by EDSR contain more residual information. Other comparison methods and proposed methods use reference images to achieve image reconstruction, thereby reconstructing more texture information. However, as the amplification factor increases, there is a significant ringing artifacts in the results of EDSR, MCSR, and Neu. From the residual images, it can be ob-

<i>Rec</i>	<i>Per</i>	<i>Com</i>	<i>SSIM</i>	PSNR↑	SSIM↑
✓	×	×	×	30.67	0.611
✓	✓	×	×	30.89	0.614
✓	✓	✓	×	30.96	0.617
✓	✓	✓	✓	31.04	0.622

Table 4: Average objective results of $4\times$ SR of ablation experiments about different combinations of loss items on the FastMRI dataset.

served that the residual images of other comparison methods contain more information than those of the proposed method, which means that our results are closest to the GT images.

Ablation Study

To verify the effectiveness of the proposed components, we design several groups of ablation experiments on the combination of the MRI-Transformer module and the perceptual contrastive loss, the joint loss function, and the multi-scale network structure, and the results are shown in Table 3, Table 4, and Table 5.

The experiment results of different combinations of MRI-Transformer module and the perceptual contrastive loss are shown in Table 3. *w comp* represents a model with the per-

Methods	PSNR \uparrow	SSIM \uparrow
<i>low-scale</i>	29.86	0.609
<i>high-scale</i>	30.41	0.616
<i>Ours(multi-scale)</i>	31.04	0.622

Table 5: Average objective results of $4\times$ SR of ablation experiments about multi-scale network structure on the FastMRI dataset.

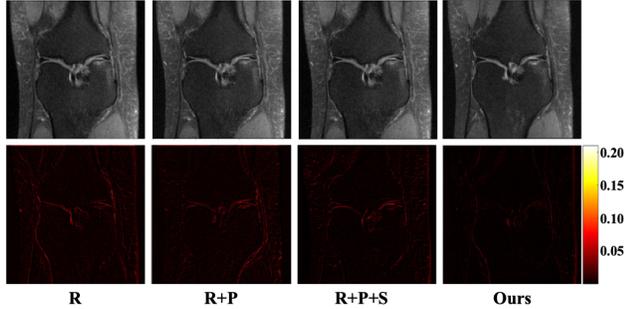


Figure 6: Visualization results of $2\times$ SR obtained by different loss terms on the FastMRI dataset.

ceptual contrastive loss, while o comp represents a model without the perceptual contrastive loss. MRIT represents two cases of a model with the MRI-Transformer module, which are trained with and without the perceptual contrastive loss. RT means two cases of a model with the Task Transformer module from T2Net (Feng et al. 2021b) but no MRI-Transformer module, which are trained with and without the perceptual contrastive loss. From the table, it can be seen that both cases included in MRIT achieved better results than the two cases in RT, indicating that the MRI-Transformer module is effective. And the model with the perceptual contrastive loss in MRIT achieved the best results in these four cases, indicating that the combination of both the MRI-Transformer module and the perceptual contrastive loss can enable the model to achieve better performance.

Table 4 shows the results of the network trained by different loss terms in the joint loss function. In the table, *Rec* refers to reconstruction loss, *Per* refers to perception loss, *Com* refers to perceptual contrastive loss, and *1-SSIM* refers to SSIM loss. From the table, we can see that the network trained by the joint loss function obtains the best results. Figure 6 shows the visual results and corresponding residual images of various cases in Table 4. From the figure, it can be seen that, the result of the joint loss function has richer information. Therefore, the joint loss function defined in this paper is effective.

Table 5 shows the results of the ablation experiment on the multi-scale network structure. In the Table 5, *low-scale* refers to the fact that all three feature reconstruction layers in MFRN perform feature reconstruction in LR space. *high-scale* refers to the fact that all three feature reconstruction layers in MFRN perform feature reconstruction in HR space, that is to say, the input LR image is magnified four

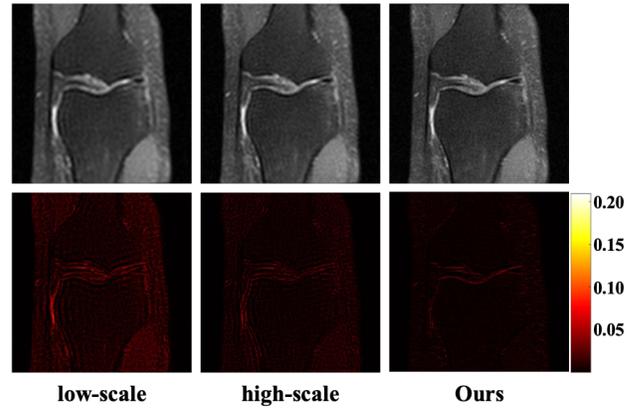


Figure 7: Visualization results of $4\times$ SR obtained by different-scale network structure on the FastMRI dataset.

times and then sent to the network for feature reconstruction. *Ours(multi-scale)* refers to the multi-scale network structure in this paper. Similarly, for the first two cases, the feature extractor in MFEB is also replaced by single-scale convolutional layers. From the table, it can be seen that the multi-scale network structure achieves the highest average metric on the FastMRI dataset. Figure 7 shows the visual results of the various cases in Table 5 and the corresponding residual images, and it can be seen that due to the lack of reconstruction of intermediate scale features, single-scale structures reconstruct less information, and the result of *low-scale* structure exhibit severe ringing phenomena. Although *high-scale* structure can reconstruct MR Images better, it is poor in detail reconstruction compared with multi-scale structure. The network structure in this paper achieves the result with more detail textures. Therefore, the constructed multi-scale SR network has better performance than the single-scale network.

Conclusion

In this paper, we propose a MFTN composed of two branches: MFRN and MFEB, to achieve SR reconstruction of LR MR image. In MFTN, a pyramid structured MFRN is constructed to gradually reconstruct features at different scales and generate a HR MR image by fusing transfer features from reference images. A MFEB based on MRI-Transformer modules is designed to learn the multi-level transfer features, which are sent to different scale layers in MFRN to supplement information for feature reconstruction. And each MRI-Transformer module learns one-level transfer features from the reference image by establishing feature correlations between input LR MR images and reference images. In addition, to improve the visual effect of reconstructed images, a perceptual contrastive loss term is defined to achieve network training. Extensive experiments on two datasets demonstrate that our method outperforms some state-of-the-art approaches in both quantitative and qualitative evaluations.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 62072218 and No. 61862030).

References

- Dong, X.; Pang, Y.; and Wen, J. 2010. Fast efficient algorithm for enhancement of low lighting video. In *ACM SIG-GRAPH 2010 Posters*, 1–1.
- Feng, C.-M.; Fu, H.; Yuan, S.; and Xu, Y. 2021a. 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*, 140–149. Springer.
- Feng, C.-M.; Yan, Y.; Fu, H.; Chen, L.; and Xu, Y. 2021b. 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*, 307–317. Springer.
- Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; and Wang, Z. 2021. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30: 2340–2349.
- Li, G.; Lv, J.; Tian, Y.; Dou, Q.; Wang, C.; Xu, C.; and Qin, J. 2022a. Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast MRI super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20636–20645.
- Li, G.; Lyu, J.; Wang, C.; Dou, Q.; and Qin, J. 2022b. Wav-Trans: 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.
- Lim, B.; Son, S.; Kim, H.; Nah, S.; and Mu Lee, K. 2017. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 136–144.
- Lyu, Q.; Shan, H.; Steber, C.; Helis, C.; Whitlow, C.; Chan, M.; and Wang, G. 2020. Multi-contrast super-resolution MRI through a progressive network. *IEEE transactions on medical imaging*, 39(9): 2738–2749.
- Lyu, Q.; Shan, H.; and Wang, G. 2020. MRI super-resolution with ensemble learning and complementary priors. *IEEE Transactions on Computational Imaging*, 6: 615–624.
- Neubert, A.; Bourgeat, P.; Wood, J.; Engstrom, C.; Chandra, S. S.; Crozier, S.; and Fripp, J. 2020. Simultaneous super-resolution and contrast synthesis of routine clinical magnetic resonance images of the knee for improving automatic segmentation of joint cartilage: data from the Osteoarthritis Initiative. *Medical Physics*, 47(10): 4939–4948.
- Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- Wang, S.; Cheng, H.; Ying, L.; Xiao, T.; Ke, Z.; Zheng, H.; and Liang, D. 2020. DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution. *Magnetic Resonance Imaging*, 68: 136–147.
- Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; and Ma, L. 2021. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10551–10560.
- Xiang, L.; Chen, Y.; Chang, W.; Zhan, Y.; Lin, W.; Wang, Q.; and Shen, D. 2018. Deep-learning-based multi-modal fusion for fast MR reconstruction. *IEEE Transactions on Biomedical Engineering*, 66(7): 2105–2114.
- Xuan, K.; Sun, S.; Xue, Z.; Wang, Q.; and Liao, S. 2020. Learning MRI k-space subsampling pattern using progressive weight pruning. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23*, 178–187. Springer.
- Zbontar, J.; Knoll, F.; Sriram, A.; Murrell, T.; Huang, Z.; Muckley, M. J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; et al. 2018. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*.
- Zeng, K.; Zheng, H.; Cai, C.; Yang, Y.; Zhang, K.; and Chen, Z. 2018. Simultaneous single- and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. *Computers in biology and medicine*, 99: 133–141.
- Zhang, Y.; Li, K.; Li, K.; and Fu, Y. 2021. MR image super-resolution with squeeze and excitation reasoning attention network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13425–13434.
- Zou, B.; Ji, Z.; Zhu, C.; Dai, Y.; Zhang, W.; and Kui, X. 2023. Multi-scale deformable transformer for multi-contrast knee MRI super-resolution. *Biomedical Signal Processing and Control*, 79: 104154.