Rethinking Multi-Contrast MRI Super-Resolution: Rectangle-Window Cross-Attention Transformer and Arbitrary-Scale Upsampling

Guangyuan Li[1], Lei Zhao[1,*], Jiakai Sun[1], Zehua Lan[1], Zhanjie Zhang[1],
Jiafu Chen[1], Zhijie Lin[2], Huaizhong Lin[1,*], Wei Xing[1,*]
[1]College of Computer Science and Technology, Zhejiang University, Hangzhou, China
[2]Zhejiang University of Science and Technology, Hangzhou, China
{cslgy, cszhl, csjk, zjucslzh, cszzj, chenjiafu, linhz, wxing}@zju.edu.cn
linzhijie@zust.edu.cn

Abstract

Recently, several methods have explored the potential of multi-contrast magnetic resonance imaging (MRI) super-resolution (SR) and obtain results superior to single-contrast SR methods. However, existing approaches still have two shortcomings: (1) They can only address fixed integer upsampling scales, such as 2×, 3×, and 4×, which require training and storing the corresponding model separately for each upsampling scale in clinic. (2) They lack direct interaction among different windows as they adopt the square window (e.g., *8×8) transformer network architecture, which results in inadequate modelling of longer-range dependencies. Moreover, the relationship between reference images and target images is not fully mined. To address these issues, we develop a novel network for multi-contrast MRI arbitrary-scale SR, dubbed as McASSR. Specifically, we design a rectangle-window cross-attention transformer to establish longer-range dependencies in MR images without increasing computational complexity and fully use reference information. Besides, we propose the reference-aware implicit attention as an upsampling module, achieving arbitrary-scale super-resolution via implicit neural representation, further fusing supplementary information of the reference image. Extensive and comprehensive experiments on both public and clinical datasets show that our McASSR yields superior performance over SOTA methods, demonstrating its great potential to be applied in clinical practice. Code will be available at https://github.com/GuangYuanKK/McASSR.*

1. Introduction

Magnetic resonance imaging (MRI) can provide clear information on tissue structure and function by acquiring high-resolution (HR) magnetic resonance (MR) images, which is a non-invasive, radiation-free clinical imaging technique. However, acquiring these HR MR images is time-consuming, and the long acquisition time will cause discomfort to patients and introduce motion artifacts [15, 5]. Image super-resolution (SR) technology is utilized to alleviate this problem by generating the corresponding HR version from the low-resolution (LR) images obtained by MRI.

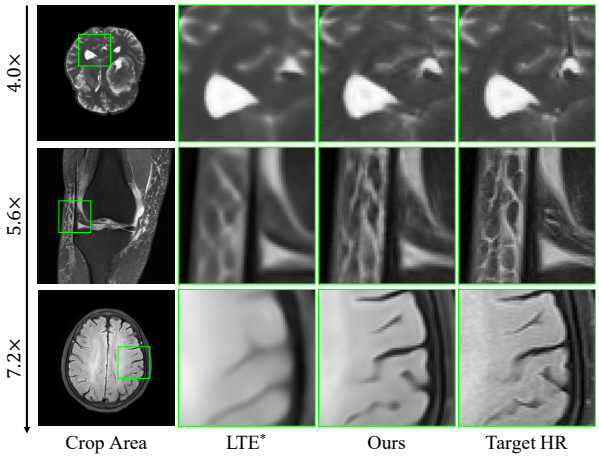

Figure 1: Visual comparison with SOTA arbitrary-scale SR method LTE [13], the reconstructed SR images by our network contain sharper edges, more visual details, and fewer blurring artefacts. ∗ means our implementation that uses SwinIR [17] as the backbone and concatenates the reference and target image as input.

At the early stage, the model-based [27, 23] and learning-based [37, 40] traditional methods are utilized to generate SR images. However, these methods have insufficient reconstruction performance under high upsampling scale, such as 4×. Recently, some deep learning

*Corresponding author.

methods based on single-contrast for MRI SR reconstruction [22, 29, 31, 2, 24, 26, 32, 41] have been proposed to cope with the limitations of traditional methods and acquire higher-quality MR images. However, the aforementioned single-contrast methods only focus on feature extraction and restoration using single-contrast MR images, ignoring the complementary information in the corresponding other contrast images, which can be utilized to improve the reconstruction quality of the images.

MRI can offer multi-contrast images with the same tissue and anatomical structure by setting different scanning parameters. Therefore, using reference HR images of one contrast with a shorter acquisition time (*e.g.*, T1 and PD) to complement target LR images of another contrast with a longer scan time (*e.g.*, T2 and FS-PD) is a promising method. Several multi-contrast approaches [21, 4, 14, 16, 39, 44] have been proposed to utilize the reference images for multi-contrast MRI SR and obtain better results than single-contrast methods. Nevertheless, these methods still have the following shortcomings that restrict their application in clinic. **Limitation 1**: They can only address fixed integer upsampling scales, such as $2\times$, $3\times$, and $4\times$, and cannot directly perform arbitrary-scale upsampling. Moreover, each fixed integer scale requires training and storing the corresponding deep neural network model separately, which seriously hinders the application in the medical field. **Limitation 2**: They lack direct interaction among different windows since they employ the square window (*e.g.*, 8×8) transformer network architecture, which results in inadequate modelling of longer-range dependencies. Moreover, the relationship between reference images and target images is not fully mined. Intuitively, for knee and brain MR images, the vertical/horizontal rectangular window (*e.g.*, $4\times16/16\times4$) can effectively establish longer-range dependencies and capture more similar features to accelerate the increase of receptive fields.

To cope with the above shortcomings, in this paper, we propose a novel and effective network for multi-contrast MRI arbitrary-scale SR. we call it McASSR. First, we design a rectangle-window cross-attention transformer as the feature extraction backbone. Specifically, it employs horizontal and vertical rectangle window cross-attention in different heads parallelly to expand the attention area, aggregate the features among different windows, and cross-fuse the complementary information of the reference and target images. Second, inspired by implicit neural representation (INR) [1], we propose the reference-aware implicit attention as an upsampling module, which achieves arbitrary-scale super-resolution and further fuses complementary information of the reference and target images. Our contributions can be summarized as follows:

(1) We propose a novel network for multi-contrast MRI arbitrary-scale SR, named as McASSR. To the best of our knowledge, our study is the first one to achieve arbitrary-scale upsampling in multi-contrast MRI tasks.

(2) The rectangle-window cross-attention transformer is designed to increase the receptive field, which can effectively establish the longer-range dependencies to fully use reference information without increasing computational complexity.

(3) The reference-aware implicit attention is proposed to realize multi-contrast MRI arbitrary-scale SR via INR and improve the quality of the reconstruction images by utilizing supplementary information from the reference images.

(4) Our McASSR outperforms SOTA approaches on four benchmark datasets (FastMRI, BraTs, healthy Brain, tumor Brain), demonstrating its effectiveness and tremendous potential to be used in clinical practice.

2. Related Works

2.1. Multi-Contrast MRI SR

In clinic, T1 or PD images usually have shorter repetition time and echo time than T2 or FS-PD images, and they have the same anatomical structure. Therefore, T1/PD can be used as a reference image to provide high-frequency texture and detail to T2/FS-PD (target image). Recently, some studies have explored the potential of multi-contrast MRI SR. For instance, Lyu *et al*. [21] designed a novel two-level progressive network for multi-contrast MRI SR. Feng *et al*. [4] proposed a novel network with the multi-stage feature fusion mechanism for multi-contrast MRI SR. Li *et al*. [14] first explored the use of the transformer in multi-contrast MRI SR and proposed a transformer-empowered multi-scale contextual matching and aggregation network. Li *et al*. [16] introduced a novel model to synergize wavelet transforms with a new cross-attention transformer for multi-contrast MRI SR. Although the above methods achieved impressive results, they only apply to a fixed upsampling factor and cannot perform arbitrary-scale SR, which is not well suited to the requirements of clinicians.

2.2. Arbitrary-Scale SR

Recently, researchers have proposed some methods for arbitrary-scale SR of natural images using implicit neural representation (INR) [1]. For example, Chen *et al*. [1] first used INR in the SR algorithm and proposed Local Implicit Image Function (LIIF) for continuous image representation. Yang *et al*. [36] designed a novel Implicit Transformer Super-Resolution Network (ITSRN) for screen content images arbitrary-scale SR. Wu *et al*. [35] proposed a novel network using INR for 3D MR images arbitrary-scale SR. Lee *et al*. [13] proposed a Local Texture Estimator (LTE) for natural images arbitrary-scale SR. Inspired by the above studies, we consider using INR in multi-contrast MRI SR to achieve arbitrary-scale upsampling using a single network.

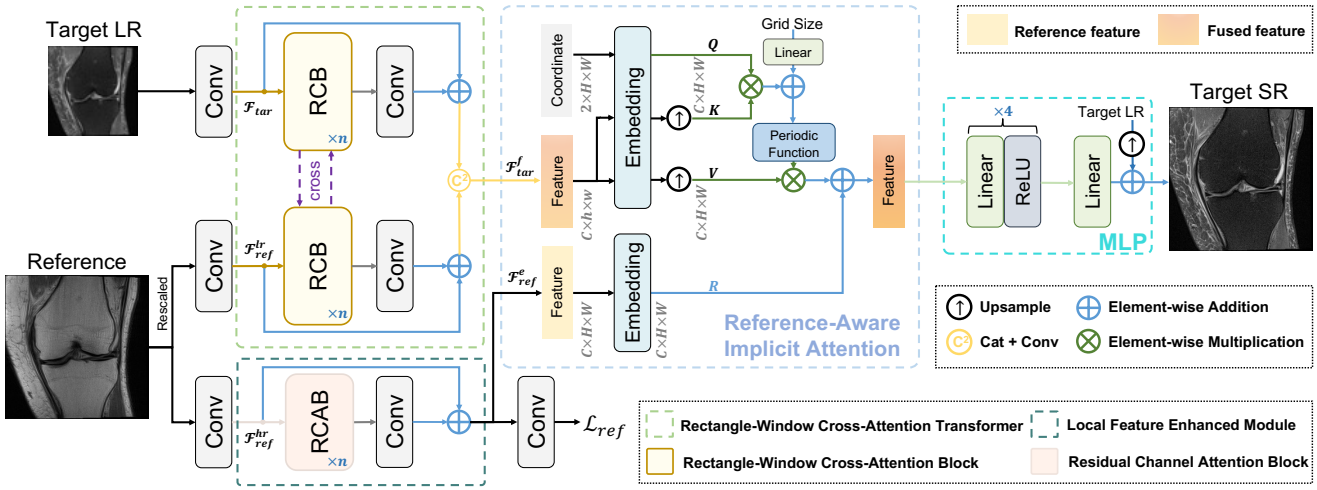

Figure 2: The overall architecture of our proposed McASSR. It divides into two components: (1) feature extraction & cross-fusion network, including rectangle-window cross-attention transformer and local feature enhanced module, and (2) implicit upsampling network, including reference-aware implicit attention and MLP.

2.3. MRI Transformer

Transformer [34] is usually utilized to solve the issue that the convolution kernel cannot capture long-range/non-local dependencies and has been widely employed in MRI reconstruction [7, 6, 14, 16, 20, 18, 19]. For instance, Feng *et al*. [7] proposed a task transformer for simultaneous MRI reconstruction and SR reconstruction and designed a multi-modal transformer [6] for multi-contrast MRI reconstruction. Lyu *et al*. [20] proposed a novel dual-domain cross-attention fusion transformer for multi-contrast MRI reconstruction. Nevertheless, the above methods employ the square window transformer to extract features of the reference image, which lacks the ability to model longer-range dependencies [43], resulting in insufficient utilization of the reference information. In contrast, the vertical/horizontal rectangle window transformer is more suitable for MR images as they usually have global similarities, such as knee and brain, which can effectively establish longer-range dependencies and capture more similar features.

3. Methodology

The overall architecture of our proposed McASSR is shown in Figure 2, which is divided into two components: (1) feature extraction & cross-fusion network (Sec. 3.1), including **R**ectangle-Window **C**ross-Attention **T**ransformer (RCT) and **L**ocal **F**eature **E**nhanced **M**odule (LFEM); (2) implicit upsampling network (Sec. 3.2), including **R**eference-Aware **I**mplicit **A**ttention (RIA) and MLP.

First, we utilize one convolution layer to perform shallow feature extraction on the target LR image $I_{tar} \in \mathbb{R}^{h \times w}$, the rescaled reference image $I^{lr}_{ref} \in \mathbb{R}^{h \times w}$, and the reference image $I^{hr}_{ref} \in \mathbb{R}^{H \times W}$ to obtain low-level features $\mathcal{F}_{tar} \in \mathbb{R}^{C \times h \times w}$, $\mathcal{F}^{lr}_{ref} \in \mathbb{R}^{C \times h \times w}$, and $\mathcal{F}^{hr}_{ref} \in \mathbb{R}^{C \times H \times W}$, respectively. Here, the rescaled means dynamically adjusting the reference image size to obtain the same size as I_{tar}, and $\mathcal{F}^{lr}_{ref}$ is utilized to supplement the low-frequency component in $\mathcal{F}_{tar}$. Then, RCT is employed to fuse $\mathcal{F}_{tar}$ and $\mathcal{F}^{lr}_{ref}$ to obtain the fused high-level feature $\mathcal{F}^{f}_{tar} \in \mathbb{R}^{C \times h \times w}$. Meanwhile, LFEM is used to perform local feature enhancement on $\mathcal{F}^{hr}_{ref}$ to obtain local enhanced high-level feature $\mathcal{F}^{e}_{ref} \in \mathbb{R}^{C \times H \times W}$. Next, $\mathcal{F}^{f}_{tar}$ and $\mathcal{F}^{e}_{ref}$ are fed into the RIA for arbitrary-scale upsampling according to the coordinate grid. Finally, the MLP is used to predict the final pixel values.

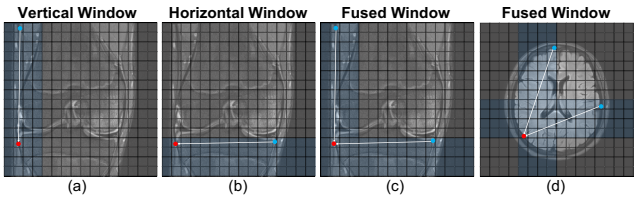

Figure 3: Illustration of rectangle window, including vertical window, horizontal window, and fused window. The blue point represents similar contents to the red point.

3.1. Feature Extraction & Cross-Fusion Network

Rectangle-Window Cross-Attention Transformer. MR images of certain tissues have global similarities, such as knee and brain, which means that the vertical/horizontal rectangular window can effectively establish longer-range dependencies and capture more similar features. As shown

in Figure 3 (a)(b), the blue patches indicate the rectangle window with a size of 4×16 or 16×4. As can be seen, there is a high similarity in the longer-range regions (red and blue). However, for the square window, such as 8×8, it is impossible to establish the longer-range dependencies of the red and blue regions without expanding the window size. Hence, we consider using the transformer with the rectangle window in the multi-contrast MRI SR (MCSR) task to enhance further the network's ability to establish longer-range dependencies.

Inspired by recent studies [43, 3, 16], we design a new rectangle-window cross-attention transformer, effectively establishing longer-range dependencies in MR images while fusing complementary information from reference images. As shown in Figure 2, the RCT is composed of n rectangle-window cross-attention blocks (RCBs) and one convolutional layer (we set n=6). The shallow features $\mathcal{F}_{tar}$ and $\mathcal{F}_{ref}^{lr}$ are fed into the RCB, respectively, and feature extraction and cross-fusion are performed to obtain the fused deep features $\mathcal{F}_{tar}^{f}$. In addition, a residual connection is involved in the RCT to stabilize the training. The process can be expressed as:

$$\mathcal{F}_{tar}^{f} = [\hat{\mathcal{F}}_{tar}+\mathcal{F}_{tar}, \hat{\mathcal{F}}_{ref}^{lr}+\mathcal{F}_{ref}^{lr}], \quad (1)$$

where $\hat{\mathcal{F}}_{tar}$= $\Theta(\Psi(\mathcal{F}_{tar}, \mathcal{F}_{ref}^{lr}))$, $\hat{\mathcal{F}}_{ref}^{lr}$= $\Theta(\Psi(\mathcal{F}_{ref}^{lr}, \mathcal{F}_{tar}))$, Ψ is RCB, Θ means Conv, and [,] denotes concatenate.

Rectangle-Window Cross-Attention Block. As shown in Figure 4, the RCB consists of two rectangle-window cross-attentions (Rwin-CA) and two MLPs. Similar to [43], Rwin-CA is divided into two types, vertical window (S^h>S^w, as shown in Figure 3 (a), named as Vwin-CA), and horizontal window (S^h<S^w, as shown in Figure 3 (b), named as Hwin-CA), and employ them in parallel for different attention heads. We set the number of attention heads H is even, where $H/2$ heads perform Vwin-CA, and the remaining heads perform Hwin-CA. Then, their outputs are concatenated along the channel dimension. As shown in Figure 3 (c)(d), by aggregating Vwin-CA and Hwin-CA, the attention area can be expanded to establish longer-range dependencies and capture more similar features without extending the window size.

For Vwin-CA, given the input $\mathcal{F}_{tar}$ and $\mathcal{F}_{ref}^{lr}$, we split them into non-overlapping rectangle windows with the number $\frac{h\times w}{S^h\times S^w}$. Specifically, for a rectangle window feature $\mathcal{F}^i$, $i = [1, \cdots, \frac{h\times w}{s^h\times S^w}]$, the query, key, and value are denoted as:

$$\left(\mathcal{F}^{Q,i}, \mathcal{F}^{K,i}, \mathcal{F}^{V,i}\right) = \left(\mathcal{F}^i E_Q, \mathcal{F}^i E_K, \mathcal{F}^i E_V\right), \quad (2)$$

where E_Q, E_K, and E_V denote the projection matrix. Then the cross-attention between $\mathcal{F}_{tar}$ and $\mathcal{F}_{ref}^{lr}$ can be computed as:

$$A(\mathcal{F}_{tar}^{Q,i}, \mathcal{F}_{ref}^{K,i}, \mathcal{F}_{ref}^{V,i}) = \mathcal{S}(\frac{\mathcal{F}_{tar}^{Q,i}(\mathcal{F}_{ref}^{K,i})^T}{\sqrt{d}} + B)\mathcal{F}_{ref}^{V,i}, \quad (3)$$

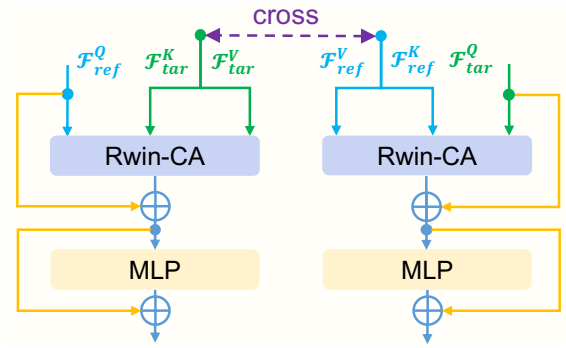

Figure 4: Illustration of rectangle-window cross-attention block, including two Rwin-CAs and two MLPs.

$$A(\mathcal{F}_{ref}^{Q,i}, \mathcal{F}_{tar}^{K,i}, \mathcal{F}_{tar}^{V,i}) = \mathcal{S}(\frac{\mathcal{F}_{ref}^{Q,i}(\mathcal{F}_{tar}^{K,i})^T}{\sqrt{d}} + B)\mathcal{F}_{tar}^{V,i}, \quad (4)$$

where $\mathcal{S}$ represents SoftMax, d means the channel dimension, and B denotes the dynamic relative position encoding. Performing the attention operation on all rectangle window features and obtaining the final output through MLP. Note that the above operation is the same for Vwin-CA and Hwin-CA.

Local Feature Enhanced Module. Current MCSR methods only consider local feature extraction [21, 4] or global feature extraction [14, 16] of MR images and fail to perform local and global feature extraction on MR images simultaneously. Therefore, we introduce a Local Feature Enhanced Module to address this issue comprehensively. Specifically, as shown in Figure 2, we use n RCABs [42] as a local feature extractor to obtain the high-level local feature $\mathcal{F}_{ref}^{e}$, which is utilized to supplement the high-frequency component information of $\mathcal{F}_{tar}^{f}$ in the implicit upsampling.

3.2. Implicit Upsampling Network

Current MCSR methods [14, 16] can only conduct fixed-scale upsampling, which is unsuitable for clinical applications as they require training and storing a separate model for each upsampling factor. Therefore, we propose implicit upsampling to solve this problem comprehensively, which consists of a reference-aware implicit attention (RIA) and an MLP.

Reference-Aware Implicit Attention. Some methods [1, 36, 13] currently achieve arbitrary-scale SR of natural images via implicit neural representation (INR). INR refers to a continuously differentiable function that utilizes the pixel coordinates to generate pixel values, such as NeRF [25]. Inspired by recent studies [1, 36, 13], we design RIA to achieve arbitrary-scale upsampling while fusing high-frequency component features from reference images, as shown in Figure 5. Unlike explicit attention, which takes

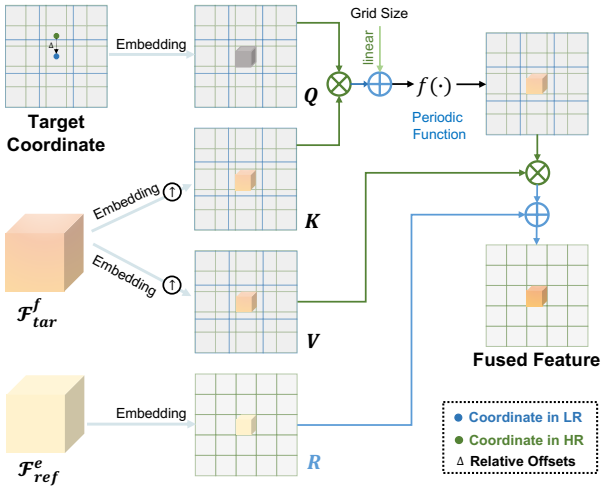

Figure 5: The detailed architecture of reference-aware implicit attention.

pixel values as the query, RIA takes pixels' coordinates as the query. In other words, RIA is not learning the pixel2pixel mapping but the coordinate2pixel mapping.

Specifically, we have the following definitions:

$$\begin{aligned} \Delta = \mathcal{C}_{HR} - \mathcal{C}^{\uparrow}_{LR}, \quad R = \mathbb{L}(\mathcal{F}^e_{ref}), \\ (Q, K, V) = (\mathbb{L}(\Delta), \uparrow \mathbb{C}(\mathcal{F}^f_{tar}), \uparrow \mathbb{C}(\mathcal{F}^f_{tar})), \end{aligned} \tag{5}$$

where $\mathbb{L}$ denotes Linear transform, $\mathbb{C}$ means Conv2D, $\uparrow$ represents nearest neighbor interpolation, $\Delta \in \mathbb{R}^{2\times H\times W}$ means the relative offsets between the query points (HR space, as shown in Figure 5 blue point) and their corresponding nearest neighboring points (target LR space, as shown in Figure 5 green point). We embed Δ to $Q \in \mathbb{R}^{C\times H\times W}$, $\mathcal{F}^f_{tar}$ to $K \in \mathbb{R}^{C\times H\times W}$ and $V \in \mathbb{R}^{C\times H\times W}$, and $\mathcal{F}^e_{ref}$ to $R \in \mathbb{R}^{C\times H\times W}$. For a query point P, the query feature is Q_P, its corresponding key feature and value feature are $K_{P\uparrow}$ and $V_{P\uparrow}$ (in the target LR space), respectively, and the reference feature is R_P (in the reference HR space), where $P\uparrow$ denotes the nearest neighboring point of P. Therefore, the RIA can be expressed as:

$$\mathcal{F}_P = (Q_P \otimes K_{P\uparrow}) \otimes V_{P\uparrow} + R_P, \tag{6}$$

where $\mathcal{F}_P$ is the fused feature that needs to be predicted, $\otimes$ means element-wise multiplication. Note that coordinates are normalized into the $[-1, 1]$. Inspired by [1, 13], we add grid size in RIA to address the issue that the location of the edge changes within a small area in its HR space when the scale factor changes. Hence, the RIA can be optimized as follows:

$$\mathcal{F}_P = (Q_P \otimes K_{P\uparrow} + \mathbb{L}(\mathcal{G})) \otimes V_{P\uparrow} + R_P, \tag{7}$$

where $\mathcal{G}$ is the grid size.

In addition, some studies [30, 33] have demonstrated that employing periodic activation functions in INR, such as *sin* and *cos*, can drive the network to learn high-frequency components. For MR images, high-frequency information is usually represented as complicated anatomical structures. Therefore, we use *sin* as a non-linear function to reweight the weights in RIA to preserve these structures:

$$\mathcal{F}_P = \sigma(Q_P \otimes K_{P\uparrow} + \mathbb{L}(\mathcal{G})) \otimes V_{P\uparrow} + R_P, \tag{8}$$

where σ denotes *sin* non-linear mapping function.

MLP. After acquiring the predicted feature $\mathcal{F}_P$, using MLP to obtain the predicted pixel value I_P of point P. Moreover, we utilize a long skip connection to stabilize the training and increase learning accuracy [11]. Thus, the predicted pixel value I_P can be expressed as:

$$I_P = \Phi(\mathcal{F}_P) + \uparrow I^{tar}_P, \tag{9}$$

where Φ denotes MLP, as shown in Figure 2.

3.3. Loss Function

The L_1 pixel loss is utilized to evaluate the reconstruction results of the target image and reference image:

$$\mathcal{L}_{tar} = \|I^{SR}_{tar} - I^{HR}_{tar}\|_1, \mathcal{L}_{ref} = \|I^{SR}_{ref} - I^{HR}_{ref}\|_1, \tag{10}$$

where I^{SR}_{tar} is the reconstructed target SR image, I^{HR}_{tar} and I^{HR}_{ref} are the original target image and the reference image, respectively, and I^{SR}_{ref} is the reconstructed reference image obtained by LFEM and one convolution layer. Therefore, the final loss function is:

$$\mathcal{L}_{full} = \lambda_{tar}\mathcal{L}_{tar} + \lambda_{ref}\mathcal{L}_{ref}, \tag{11}$$

where hyperparameters λ_{tar}=0.7 and λ_{ref}=0.3 are utilized to control the weight between $\mathcal{L}_{tar}$ and $\mathcal{L}_{ref}$ [4].

4. Experiments

4.1. Datasets and Baselines

Datasets. We employ four datasets to evaluate the performance of our proposed network, including two public multi-contrast MRI datasets: FastMRI [38] and BraTs [9], and two in-house datasets: Healthy brain and Tumor brain, as shown in Table 1. The tumor brain dataset is only used to test the model trained by the healthy brain dataset. For the training and validation datasets, we downsample the ground truth (GT) images (256×256) at random scale $s \in (1, 4]$ by bicubic interpolation during training to generate LR images [35]. For the test dataset, we utilize Fourier transform to convert the images into k-space, crop the k-space, and then obtain the LR test set by inverse Fourier transforms [14].

Baselines. We compare our method with existing fixed integer scale MCSR methods, including MINet [4] (MICCAI2021), WavTrans [16] (MICCAI2022), McMRSR [14]

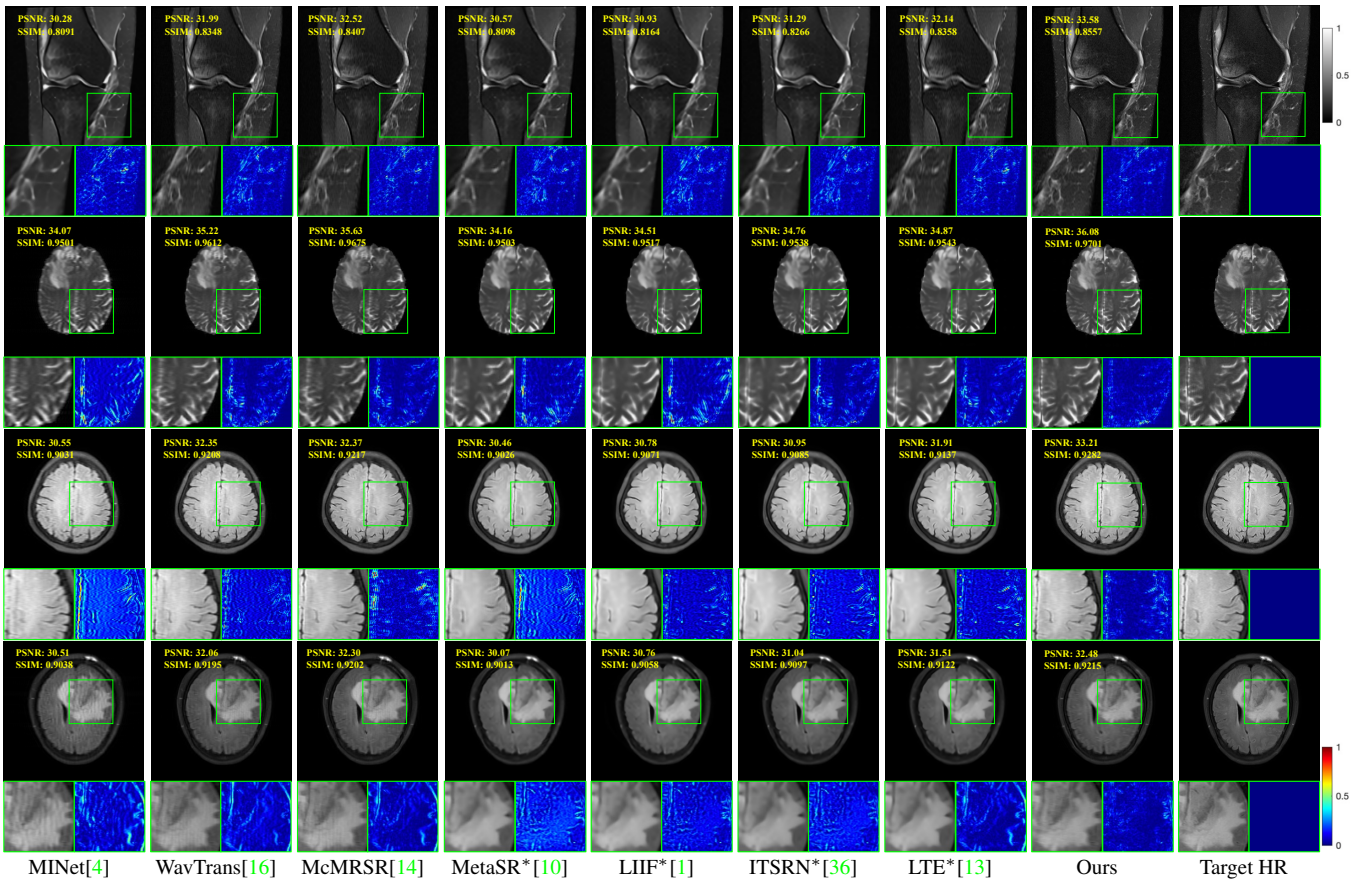

Figure 6: Qualitative comparison of different SR reconstruction methods on four datasets with an in-scale of 4×. ∗ means our implementation that uses SwinIR [17] as the backbone and concatenates the reference and target image as input.

Table 1: Four datasets used for the experiments.

Datasets	FastMRI	BraTs	Health	Tumor
Reference	PD	T1	T1	T1
Target	FS-PD	T2	T2-FLAIR	T2-FLAIR
Train	640	640	512	-
Validation	160	160	125	-
Test	160	160	125	305

(CVPR2022), and arbitrary-scale SR methods, including MetaSR [10] (CVPR2019), LIIF [1] (CVPR2021), ITSRN [36] (NeurIPS2021), LTE [13] (CVPR2022). For a fair comparison, we use SwinIR [17] as the backbone and concatenate the reference and the target image as input for arbitrary-scale SR methods. In addition, we evaluate our McASSR for both in-scale $s \in$(1,4] and out-of-scale s=[6, 8] to verify the generalization capability of our proposed network and set rectangle window with a size of 4×16/16×4.

4.2. Implementation Details

We implement our proposed approach in PyTorch [28] with a single NVIDIA RTX A6000 GPU (48GB). The Adam [12] optimizer is adopted for network training with epochs of 1000. We set the batch size as 8 and the learning rate as 2e-4 and decayed by factor 0.5 at [500, 800, 900, 950]. In addition, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity index (SSIM) are employed as metrics to measure the performance of the SR reconstruction.

4.3. Qualitative results

Figure 6 provides the qualitative comparison of three fixed integer scale MCSR methods (1^c_{st} to 3^c_{rd}) and four arbitrary-scale SR methods (4^c_{th} to 7^c_{th}) on four datasets with an in-scale of 4× ([64 × 64] → [256 × 256]). The top, second, third, and bottom rows are the SR results under the FastMRI, BraTs, healthy brain, and tumor brain datasets, respectively. The green box is the local zoom-in region and the corresponding error map, and the more textures in the error map, the poorer quality of the reconstructed SR results. As can be seen, our method can restore the compli-

Table 2: Quantitative comparison with SOTA methods on four datasets (PSNR (dB) ↑ and SSIM ↑). The best results are highlighted in red (best) and blue (2nd best). $*$ means our implementation that uses SwinIR [17] as the backbone and concatenates the reference and target image as input.

Dataset	Method	in-scale						out-of-scale			
		2×		3×		4×		6×		8×	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
FastMRI Knee	MINet [4]	35.02	0.8933	32.85	0.8504	30.59	0.8101	-	-	-	-
	WavTrans [16]	36.15	09054	34.23	0.8576	31.95	0.8343	-	-	-	-
	McMRSR [14]	36.17	0.9063	34.35	0.8588	31.96	0.8345	-	-	-	-
	MetaSR* [10]	34.26	0.8902	32.71	0.8443	30.66	0.8102	28.57	0.7914	27.42	0.7258
	LIIF* [1]	34.79	0.8931	32.90	0.8508	30.87	0.8155	28.86	0.7945	27.74	0.7263
	ITSRN* [36]	35.13	0.8948	33.02	0.8513	31.02	0.8258	29.05	0.7995	27.92	0.7306
	LTE* [13]	35.16	0.8955	33.14	0.8574	31.19	0.8263	29.12	0.8007	27.85	0.7316
	Ours	36.31	0.9071	34.54	0.8743	32.52	0.8416	29.92	0.8096	28.76	0.7457
BraTs	MINet [4]	40.90	0.9906	36.85	0.9724	34.27	0.9514	-	-	-	-
	WavTrans [16]	41.65	0.9911	38.35	0.9785	35.36	0.9621	-	-	-	-
	McMRSR [14]	41.73	0.9912	38.41	0.9793	35.49	0.9663	-	-	-	-
	MetaSR* [10]	40.82	0.9904	35.90	0.9700	34.15	0.9503	30.25	0.9107	28.50	0.9003
	LIIF* [1]	40.96	0.9907	36.73	0.9713	34.53	0.9519	30.73	0.9156	28.98	0.9037
	ITSRN* [36]	41.18	0.9909	36.92	0.9728	34.62	0.9523	30.77	0.9157	29.35	0.9094
	LTE* [13]	41.40	0.9910	37.24	0.9750	34.77	0.9535	31.06	0.9192	29.97	0.9135
	Ours	41.98	0.9917	38.47	0.9801	35.62	0.9659	32.05	0.9237	30.68	0.9197
Healthy Brain	MINet [4]	37.59	0.9836	32.91	0.9453	30.75	0.9049	-	-	-	-
	WavTrans [16]	38.52	0.9904	33.85	0.9521	32.42	0.9215	-	-	-	-
	McMRSR [14]	38.59	0.9911	33.89	0.9537	32.48	0.9217	-	-	-	-
	MetaSR* [10]	36.37	0.9751	31.83	0.9372	30.72	0.9042	26.61	0.8394	26.27	0.7996
	LIIF* [1]	37.22	0.9826	32.80	0.9451	30.82	0.9073	27.95	0.8559	26.35	0.8122
	ITSRN* [36]	37.66	0.9838	32.97	0.9455	31.24	0.9117	28.23	0.8596	26.68	0.8163
	LTE* [13]	37.73	0.9856	33.31	0.9486	31.96	0.9140	28.39	0.8612	26.92	0.8176
	Ours	38.68	0.9914	34.12	0.9579	32.65	0.9227	29.38	0.8697	27.56	0.8219
Tumor Brain	MINet [4]	37.18	0.9812	32.31	0.9406	30.37	0.9015	-	-	-	-
	WavTrans [16]	38.19	0.9882	33.52	0.9495	31.99	0.9177	-	-	-	-
	McMRSR [14]	38.22	0.9891	33.59	0.9502	32.07	0.9191	-	-	-	-
	MetaSR* [10]	36.34	0.9751	31.70	0.9353	30.43	0.9032	26.47	0.8384	26.12	0.7989
	LIIF* [1]	37.14	0.9815	32.53	0.9414	30.72	0.9057	27.87	0.8545	26.29	0.8107
	ITSRN* [36]	37.55	0.9834	32.65	0.9428	31.10	0.9102	28.19	0.8581	26.68	0.8162
	LTE* [13]	37.62	0.9851	33.16	0.9456	31.74	0.9165	28.26	0.8606	26.90	0.8173
	Ours	38.43	0.9901	33.97	0.9529	32.39	0.9206	29.14	0.8685	27.57	0.8219

Table 3: Ablation study on various variant models under FastMRI dataset with an in-scale of 4×. The best quantitative result (PSNR (dB) ↑ and SSIM ↑) is marked in **bold**.

Variants	Square	w/o reference	w/o $\mathcal{L}_{ref}$	w/o RIA	w/o GS	w/o PAF	w/o RCB	w/o LFEM	Ours Full
PSNR	31.63	31.11	32.26	31.69	31.42	31.85	31.52	32.01	**32.52**
SSIM	0.8289	0.8265	0.8357	0.8293	0.8273	0.8302	0.8281	0.8348	**0.8416**

cated anatomical structure in MR images and preserve the original information in the HR images, which demonstrates that our McASSR can establishes longer-range dependencies and effectively fuse high-frequency information in the reference images. Furthermore, Figure 7 illustrates the SR reconstruction results of our method with other arbitrary-scale SR methods using a single network under the out-of-scale 8× ([32 × 32] → [256 × 256]). Even at an extremely large magnification factor of 8×, our method can still reconstruct the closest image to the original, indicating that our method has significant clinical application prospects.

4.4. Quantitative results

Table 2 offers the quantitative comparison between our proposed method and other SR methods, including fixed integer scale MCSR and arbitrary-scale SR methods. We

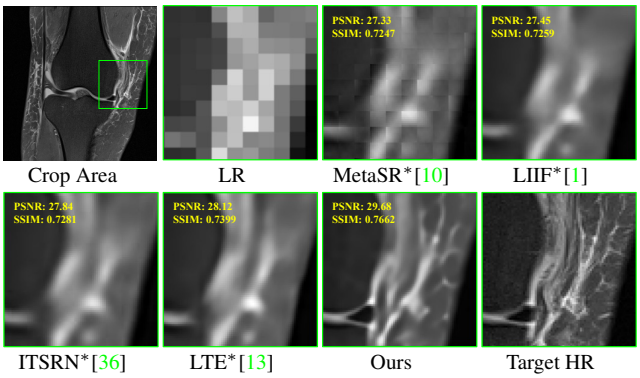

Figure 7: Qualitative results of different arbitrary-scale SR methods on the FastMRI with an out-of-scale of 8×.

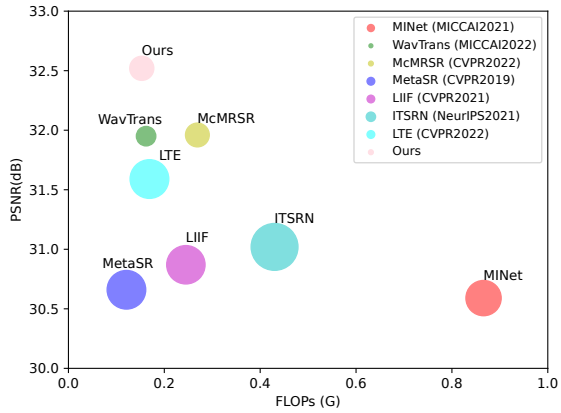

Figure 8: Complexity comparison. The PSNR are evaluated on the FastMRI dataset for 4× upsampling, and the FLOPs are calculated with a 64×64 input. FLOPs (G): $\times 10^3$.

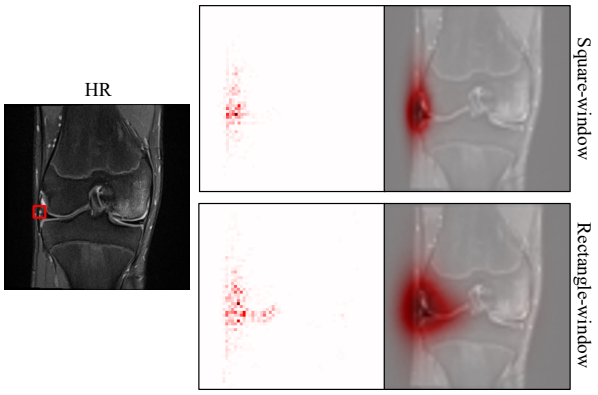

Figure 9: Local attribution maps comparison between square-window and rectangle-window.

notice that, WavTrans and McMRSR have significant benefits for in-scale as they train corresponding models for each fixed integer scale. Nevertheless, the best results in terms of all metrics are still acquired by our method regardless of in-scale or out-of-scale, demonstrating that our McASSR achieves arbitrary-scale upsampling while maximizing the restoration of detailed information in the MR image.

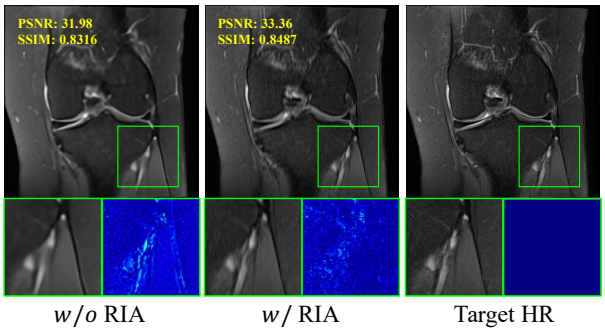

Figure 10: Qualitative comparison of w/o RIA and $w/$ RIA.

Moreover, we calculated the Parameters (M) and FLOPs (G $\times 10^3$) for each network model, as shown in Figure 8. The Parameters are represented as circles; a larger circle diameter means a larger parameter. As can be seen, our method can get the most satisfactory SR results under smaller Parameters and FLOPs. Furthermore, the inference time per image is as follows: MINet: 128ms, WavTrans: 193ms, McMRSR: 187ms, MetaSR: 56ms, LIIF: 40ms, ITSRN: 80ms, LTE: 24ms, and Ours: 35ms.

4.5. Ablation Study

In this section, we explore the significance of each key component of our proposed method. All these variant models are retrained the same way as before on the FastMRI dataset and tested with an in-scale of 4×. The quantitative metrics results of these variants are shown in Table 3.

4.5.1 Effect of rectangle-window.

To investigate the effectiveness of the rectangle window, we design a variant by replacing the rectangle window with the square window (8×8), as shown in Table 3. As can be seen, the PSNR value of the square-window variant decreases by about 0.89dB, demonstrating that using the rectangle-window can effectively establish longer-range dependencies to accelerate the increase of receptive fields.

Moreover, we utilize local attribution maps (LAM) [8] to visualize the receptive fields of rectangle-window and square-window, as shown in Figure 9. Note that LAM means the contribution of pixels in other regions when reconstructing the red region. That is, the more pixels that can be utilized, the larger the long-range dependencies of the network. We notice that, the square window variant utilizes only a limited range of pixels. In contrast, the rectangular window can employ a longer range of pixels, which means that it effectively establishes longer-range dependencies.

4.5.2 Effect of rectangle-window size

To explore the effectiveness of rectangle-window size, we design two variants by adjusting the rectangle-window size to 1×64 / 64×1, named as R-window-1, and adjusting the rectangle-window size to 2×32 / 32×2, named as R-window-2, as shown in Table 4. As can be seen, the PSNR values of R-window-1 and R-window-2 decrease by about 1.5dB and 1.09dB, respectively, demonstrating that the rectangle-window width/height compression will lead to the reduction of the reconstruction performance of the network.

4.5.3 Effect of reference image.

To evaluate the contribution of the reference image, we conduct three ablation studies by removing the reference image, referred as w/o reference, removing the $\mathcal{L}_{ref}$, referred as w/o $\mathcal{L}_{ref}$, removing the reference-aware in RIA, referred as w/o RIA. As can be seen, using supplementary information in the reference image can enhance the quality of target SR reconstruction. In addition, RIA can effectively supplement high-frequency information for the target image, as shown in Figure 10. As can be seen, the reconstructed image of variant w/o RIA loses some high-frequency detail information, resulting in a large reconstruction error.

4.5.4 Effect of GS and PAF

To verify the contribution of Grid Size (GS) and Periodic Activation Functions (PAF), we design two variants by removing grid size and removing periodic activation functions, named w/o GS and w/o PAF, respectively. As shown in Table 3, the introduction of grid size and periodic activation functions can enhance the reconstruction performance of the network.

4.5.5 Effect of RCB and LFEM

To explore the effect of the rectangle-window cross-attention block (RCB) and local feature enhanced module (LFEM), we design two variant models by removing the cross in RCB, named as w/o RCB, and replacing LFEM with one convolution layer, named as w/o LFEM. As shown in Table 3, by introducing the RCB and LFEM, the reconstruction performance of our proposed network is optimized. This confirms the strong ability of RCB and LFEM modules to ensure that the target LR features make maximum use of the reference information.

4.5.6 Effect of IUN

To validate the contribution of the proposed implicit upsampling network (IUN), we design a variant model named *w/o* IUN, which replaces IUN with the simplest implicit neural representation [1]. Utilizing the public dataset FastMRI to conduct experiments at in-scale $4\times$ and out-of-scale $6\times$ and $8\times$, the quantitative results (PSNR/SSIM) are shown in Table 5. The results indicate that without IUN, the performance of the network significantly decreases.

Table 4: Ablation study on various window-size under FastMRI dataset with an in-scale of $4\times$. The best quantitative result (PSNR (dB) ↑ and SSIM ↑) is marked in **bold**.

Variant	Window-Size	PSNR	SSIM
R-window-1	1×64 / 64×1	31.02	0.8257
R-window-2	2×32 / 32×2	31.43	0.8273
Ours Full	4×16 / 16×4	**32.52**	**0.8416**
S-window	8×8 / 8×8	31.63	0.8289

Table 5: Effect of implicit upsampling network. The best quantitative result (PSNR (dB) ↑ and SSIM ↑) is marked in **bold**.

Scale	Metrics	*w/o* IUN	*w/* IUN
$4\times$	PSNR	31.21	**32.52**
	SSIM	0.8266	**0.8416**
$6\times$	PSNR	29.21	**29.92**
	SSIM	0.8017	**0.8096**
$8\times$	PSNR	28.03	**28.76**
	SSIM	0.7339	**0.7457**

5. Conclusion

This study proposes a novel multi-contrast MRI SR method to establish longer-range dependencies in MR images and provide sufficient complementary information for the target LR image via a rectangle-window cross-attention transformer and achieve arbitrary-scale super-resolution by harnessing reference-aware implicit attention. Experimental results demonstrate that our McASSR outperforms existing SOTA methods, showing the potential to be applied in clinical practice.

Future Work. The multi-contrast image pairs need to be co-registered in advance, which is tedious and time-consuming. In the future, we shall work on designing a multi-task framework to perform registration and SR reconstruction simultaneously.

Acknowledgments

This work was supported in part by the National Program of China (2020YFC1523201, 62172365,19ZDA197) and Zhejiang Elite Program (2022C01222).

References

[1] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8628–8638, 2021. 2, 4, 5, 6, 7, 8, 9

[2] Yuhua Chen, Yibin Xie, Zhengwei Zhou, Feng Shi, Anthony G Christodoulou, and Debiao Li. Brain mri super resolution using 3d deep densely connected neural networks. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 739–742. IEEE, 2018. 2

[3] Xiaoyi Dong, Jianmin Bao, Dongdong Chen, Weiming Zhang, Nenghai Yu, Lu Yuan, Dong Chen, and Baining Guo. Cswin transformer: A general vision transformer backbone with cross-shaped windows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12124–12134, 2022. 4

[4] Chun-Mei Feng, Huazhu Fu, Shuhao Yuan, and Yong Xu. 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*, pages 140–149. Springer, 2021. 2, 4, 5, 6, 7

[5] Chun-Mei Feng, Kai Wang, Shijian Lu, Yong Xu, and Xuelong Li. Brain mri super-resolution using coupled-projection residual network. *Neurocomputing*, 456:190–199, 2021. 1

[6] Chun-Mei Feng, Yunlu Yan, Geng Chen, Yong Xu, Ying Hu, Ling Shao, and Huazhu Fu. Multi-modal transformer for accelerated mr imaging. *IEEE Transactions on Medical Imaging*, 2022. 3

[7] Chun-Mei Feng, Yunlu Yan, Huazhu Fu, Li Chen, and Yong Xu. 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*, pages 307–317. Springer, 2021. 3

[8] Jinjin Gu and Chao Dong. Interpreting super-resolution networks with local attribution maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9199–9208, 2021. 8

[9] Ali Hatamizadeh, Demetri Terzopoulos, and Andriy Myronenko. End-to-end boundary aware networks for medical image segmentation. In *Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10*, pages 187–194. Springer, 2019. 5

[10] Xuecai Hu, Haoyuan Mu, Xiangyu Zhang, Zilei Wang, Tieniu Tan, and Jian Sun. Meta-sr: A magnification-arbitrary network for super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1575–1584, 2019. 6, 7, 8

[11] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 5

[12] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. 6

[13] Jaewon Lee and Kyong Hwan Jin. Local texture estimator for implicit representation function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1929–1938, 2022. 1, 2, 4, 5, 6, 7, 8

[14] Guangyuan Li, Jun Lv, Yapeng Tian, Qingyu Dou, Cheng-Yan Wang, Chenliang Xu, and Jing Qin. Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast mri super-resolution. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20604–20613, 2022. 2, 3, 4, 5, 6, 7

[15] Guangyuan Li, Jun Lv, Xiangrong Tong, Chengyan Wang, and Guang Yang. High-resolution pelvic mri reconstruction using a generative adversarial network with attention and cyclic loss. *IEEE Access*, 9:105951–105964, 2021. 1

[16] Guangyuan Li, Jun Lyu, Chengyan Wang, Qi Dou, and Jing Qin. Wavtrans: Synergizing wavelet and cross-attention transformer for multi-contrast mri super-resolution. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, pages 463–473. Springer, 2022. 2, 3, 4, 5, 6, 7

[17] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 1, 6, 7

[18] Jun Lyu, Guangyuan Li, Chengyan Wang, Qing Cai, Qi Dou, David Zhang, and Jing Qin. Multicontrast mri super-resolution via transformer-empowered multiscale contextual matching and aggregation. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 3

[19] Jun Lyu, Guangyuan Li, Chengyan Wang, Chen Qin, Shuo Wang, Qi Dou, and Jing Qin. Region-focused multi-view transformer-based generative adversarial network for cardiac cine mri reconstruction. *Medical Image Analysis*, 85:102760, 2023. 3

[20] Jun Lyu, Bin Sui, Chengyan Wang, Yapeng Tian, Qi Dou, and Jing Qin. Dudocaf: Dual-domain cross-attention fusion with recurrent transformer for fast multi-contrast mr imaging. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, pages 474–484. Springer, 2022. 3

[21] Qing Lyu, Hongming Shan, Cole R. Steber, Corbin A. Helis, Chris Whitlow, Michael Chan, and Ge Wang. Multi-contrast super-resolution mri through a progressive network. *IEEE Transactions on Medical Imaging*, 39:2738–2749, 2019. 2, 4

[22] Qing Lyu, Hongming Shan, and Ge Wang. Mri super-resolution with ensemble learning and complementary priors. *IEEE Transactions on Computational Imaging*, 6:615–624, 2019. 2

[23] José V Manjón, Pierrick Coupé, Antonio Buades, Vladimir Fonov, D Louis Collins, and Montserrat Robles. Non-local

mri upsampling. *Medical image analysis*, 14(6):784–792, 2010. 1

[24] Steven McDonagh, Benjamin Hou, Amir Alansary, Ozan Oktay, Konstantinos Kamnitsas, Mary Rutherford, Jo V Hajnal, and Bernhard Kainz. Context-sensitive super-resolution for fast fetal magnetic resonance imaging. In *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment: Fifth International Workshop, CMMI 2017, Second International Workshop, RAMBO 2017, and First International Workshop, SWITCH 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017, Proceedings 5*, pages 116–126. Springer, 2017. 2

[25] B Mildenhall, PP Srinivasan, M Tancik, JT Barron, R Ramamoorthi, and R Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, 2020. 4

[26] Ozan Oktay, Wenjia Bai, Matthew Lee, Ricardo Guerrero, Konstantinos Kamnitsas, Jose Caballero, Antonio de Marvao, Stuart Cook, Declan O'Regan, and Daniel Rueckert. Multi-input cardiac image super-resolution using convolutional neural networks. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part III 19*, pages 246–254. Springer, 2016. 2

[27] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003. 1

[28] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 6

[29] Defu Qiu, Shengxiang Zhang, Ying Liu, Jianqing Zhu, and Lixin Zheng. Super-resolution reconstruction of knee magnetic resonance imaging based on deep learning. *Computer methods and programs in biomedicine*, 187:105059, 2020. 2

[30] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020. 5

[31] Jennifer A Steeden, Michael Quail, Alexander Gotschy, Kristian H Mortensen, Andreas Hauptmann, Simon Arridge, Rodney Jones, and Vivek Muthurangu. Rapid whole-heart cmr with single volume super-resolution. *Journal of Cardiovascular Magnetic Resonance*, 22(1):1–13, 2020. 2

[32] Kun Sun, Liangqiong Qu, Chunfeng Lian, Yongsheng Pan, Dan Hu, Bingqing Xia, Xinyue Li, Weimin Chai, Fuhua Yan, and Dinggang Shen. High-resolution breast mri reconstruction using a deep convolutional generative adversarial network. *Journal of Magnetic Resonance Imaging*, 52(6):1852–1858, 2020. 2

[33] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547, 2020. 5

[34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 3

[35] Qing Wu, Yuwei Li, Yawen Sun, Yan Zhou, Hongjiang Wei, Jingyi Yu, and Yuyao Zhang. An arbitrary scale super-resolution approach for 3d mr images via implicit neural representation. *IEEE Journal of Biomedical and Health Informatics*, 2022. 2, 5

[36] Jingyu Yang, Sheng Shen, Huanjing Yue, and Kun Li. Implicit transformer network for screen content image continuous super-resolution. *Advances in Neural Information Processing Systems*, 34:13304–13315, 2021. 2, 4, 6, 7, 8

[37] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010. 1

[38] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastmri: An open dataset and benchmarks for accelerated mri. *arXiv preprint arXiv:1811.08839*, 2018. 5

[39] Kun Zeng, Hong Zheng, Congbo Cai, Yu Yang, Kaihua Zhang, and Zhong Chen. 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, 2018. 2

[40] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, pages 711–730. Springer, 2012. 1

[41] Yulun Zhang, Kai Li, Kunpeng Li, and Yun Fu. Mr image super-resolution with squeeze and excitation reasoning attention network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13425–13434, 2021. 2

[42] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 4

[43] Chen Zheng, Yulun Zhang, Jinjin Gu, Yongbing Zhang, Linghe Kong, and Xin Yuan. Cross aggregation transformer for image restoration. In *Advances in Neural Information Processing Systems*. 3, 4

[44] Bo Zhou and S Kevin Zhou. 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*, pages 4273–4282, 2020. 2