A Transform-Domain Approach with Symmetric and Edge Constraints for MRI Super-Resolution

Han Zhang[1&], Yu Lu[1&], Ran Wang[1], Dian Ding[1*], Mengying Zhu[3],
Shengyun He[3], Ling Ma[4], Yi-Chao Chen[1], Ruokun Li[2], Shikui Tu[1],
Guangyu Wu[3], Guangtao Xue[1]
[1] Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China
[2] Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, China
[3] Department of Radiology, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[4] Library, Shanghai Jiao Tong University, Shanghai, China
Email:{han_zhang, yulu01, wang_r, dingdian94}@sjtu.edu.cn, zhu_mengying93@163.com, 810218553@qq.com,
{maling920827, yichao}@sjtu.edu.cn, lrk12113@rjh.com.cn,
tushikui@sjtu.edu.cn, danielrau@163.com, gt_xue@sjtu.edu.cn

***Abstract*—Magnetic resonance imaging (MRI) provides high-quality soft tissue contrast images and is crucial in medical diagnosis. However, systems face trade-offs between image resolution and scan time. Low-resolution MRI scans reduce scan time and patient burden but lose critical details needed for accurate diagnosis. To address this problem, super-resolution techniques have been developed to improve the clarity of low-resolution input images. Single-image super-resolution (SISR), which minimizes patient scanning time, has gradually become a research focus, but existing methods often struggle to balance the reconstruction of low-frequency structural information and high-frequency details. In this paper, we propose a novel super-resolution up-sampling pipeline that enhances both the high-frequency and low-frequency components of magnetic resonance imaging. In addition, we introduce an enhanced loss function that includes symmetry and edge constraints to preserve critical structural details for improved diagnostic accuracy. The extensive experiments across multiple datasets validate the effectiveness of our SISR model. Source code will be made publicly available.**

***Index Terms*—Magnetic Resonance Imaging, Single-image Super-resolution, Deep Learning**

I. Introduction

Magnetic Resonance Imaging (MRI) is of great importance in medicine, helping doctors to accurately diagnose and treat various conditions by providing high-quality contrast images of soft tissue [1]. However, high-resolution images require longer scanning times [2], [3], increasing patient discomfort and the risk of motion artifacts. Moreover, high-resolution scans usually require more expensive equipment and complex operational procedures, which increase healthcare costs and are not conducive to their rollout in resource-limited healthcare organizations. However, low-resolution images lose details, making it difficult to detect small but critical lesions, thus affecting the accuracy and timeliness of diagnosis. Image blurring and artifacts can also interfere with a doctor's judgment, increasing the risk of misdiagnosis and missed diagnosis. To address these issues, researchers have developed various super-resolution techniques [4], [5] that improve image clarity based on low-resolution images and up-sampling techniques.

Based on the number of input low-resolution MRI images, existing super-resolution systems can be classified into multi-image super-resolution (MISR) and single-image super-resolution (SISR). Although the acquisition time of a single low-resolution MRI image is significantly shorter, the scanning time associated with multi-frame image input is non-negligible. To minimize patient burden, SISR systems offer significant advantages.

To recover high-frequency information from low-resolution MRI images, recent researches [4], [6]–[9] have proposed SISR networks based on deep learning. EDSR [6] proposed a SISR network with optimized residual structure and achieves multi-scale image super-resolution. LIIF [7] introduced a novel method for continuous image representation that predicts RGB values at arbitrary resolutions using image coordinates and local 2D deep features. LTE [8] enhanced implicit neural functions in single image super-resolution (SISR), enabling high-frequency texture estimation in the Fourier domain for accurate and continuous image reconstruction at arbitrary scales. For MRI image super-resolution, ArSSR [9] proposed an Arbitrary Scale Super-Resolution approach for 3D MRI that uses implicit neural representations to reconstruct high-resolution images from low-resolution inputs, enabling continuous and arbitrary up-sampling rates with a single model. DRFSA [4] proposed a multiple perceptron (MLP) based on depth residual Fourier to extract image high-frequency features.

Conventional super-resolution methods often struggle to balance low- and high-frequency feature processing, leading to suboptimal reconstruction of structural details and fine textures. Most existing approaches handle these components separately, lacking a unified optimization framework. Moreover, they typically rely on pixel-wise losses such as mean squared error (MSE), which emphasize intensity accuracy but fail to preserve critical edge structures—resulting in blurred or distorted details that are vital for medical image analysis.

To address these challenges, first, we propose a novel super-

[&] Both authors contributed equally to the research.
[*] Dian Ding is Corresponding author.

resolution up-sampling structure that employs a single up-sampling module to enhance both the high-frequency and low-frequency components of the MRI image in the transform domain by a factor of 2, thereby achieving 2^N-fold super-resolution reconstruction with N up-sampling modules ($N = 1, 2, 3, \dots$). For the low-frequency component, we utilize a convolutional block based on a soft-thresholding operator to filter and extract high-energy low-frequency features. In contrast, for the high-frequency component, we implement an LTE-based structure to estimate texture details. Then, we employ an enhanced loss function to train the model. In addition to the traditional MSE loss, we incorporate a constraint loss to enforce symmetric constraints and an edge loss to preserve and enhance edge details, thereby optimizing the model's performance.

In this paper, we evaluate the SISR model on the publicly available IXI and SIMON datasets, selected for their diversity, multi-center acquisition, and clinical relevance. Both datasets feature high-quality preprocessing and minimal noise, ensuring reliable and reproducible evaluation. IXI includes multiple MRI modalities, enabling multimodal analysis, while SIMON focuses on clinical scenarios, validating the model's practical applicability. Together, they provide a comprehensive benchmark for assessing our method.

The main contributions of this work are as follows:

- We propose a novel super-resolution up-sampling structure that simultaneously extracts the low and high-frequency features of an image over the MRI transform domain, taking into account both the structural low-frequency information and the high-frequency detail features.
- We propose a hybrid loss function that combines L1 loss, symmetric constraint loss, and edge loss to enhance the recovery of image edge information and improve the medical diagnostic value of the SR image output.
- We conduct extensive experiments to validate the effectiveness of our SISR model. The results demonstrate that, compared to previous MRI SISR models, our model delivers superior image quality across multiple datasets.
- We contribute SISRMRI, a public MRI dataset focused on ccRCC with pseudocapsule, featuring multi-resolution T2-weighted imaging and comprehensive acquisition parameters.

II. Preliminary

MRI images are generated through a complex interplay of magnetic fields, radiofrequency pulses, and advanced computational techniques [10]. The process involves aligning protons in a strong magnetic field, perturbing them with RF pulses, detecting the emitted signals during relaxation, spatially encoding these signals using gradient fields, and reconstructing the Fourier space data into high-resolution images using Fourier transform and image processing algorithms. In the medical imaging community, this Fourier space is referred to as *k-space* [11]. The reconstructed MRI image I_r can be obtained

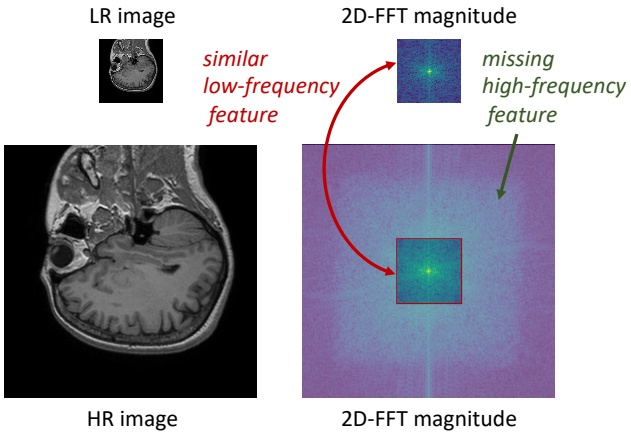

Fig. 1. HR images and LR images possess similar low-frequency features that correspond to image structural information; however, LR images lack detailed information corresponding to high-frequency features.

from the full *k-space* sampling y by performing an inverse multidimensional Fourier transform:

$$I_r = \mathcal{F}^{-1}(y) \tag{1}$$

However, the full *k-space* sampling is constrained by several limitations, such as prolonged sampling durations, necessitating patient immobility, attenuation of magnetic resonance signals, and physical constraints inherent to hardware facilities. In clinical practice, MRI images are typically reconstructed through *k-space* sub-sampling techniques, which include both fast and slow sampling methods. These approaches result in medical images of lower and higher spatial resolution, respectively:

$$y_f = \mathcal{P}_f(\mathcal{F}(I_r)) + \mathcal{N}_f \tag{2}$$

$$I_{lr} = \underset{I_{lr}}{argmin}\, \mathcal{R}(I_{lr}) \quad s.t. \|\mathcal{F}(I_{lr}) - y_f\|_2^2 \leq \epsilon \tag{3}$$

$$y_s = \mathcal{P}_s(\mathcal{F}(I_r)) + \mathcal{N}_s \tag{4}$$

$$I_{hr} = \underset{I_{hr}}{argmin}\, \mathcal{R}(I_{hr}) \quad s.t. \|\mathcal{F}(I_{hr}) - y_s\|_2^2 \leq \epsilon \tag{5}$$

Here, $\mathcal{P}$ denotes the *k-space* sampling function, and $\mathcal{N}$ corresponds to the system noise. y represents the *k-space* subsampled signal acquired at all measured spatial frequencies while simultaneously minimizing a sparsity-inducing objective $\mathcal{R}(\cdot)$ under certain domains, which penalizes unnatural reconstructions [12]. I_{hr} denotes the generated high-resolution medical image, I_{lr} represents the generated low-resolution medical image, and ϵ is the specified small threshold value.

Consequently, considering that both y_s and y_f are subsamples of the fully sampled *k-space* signal y, then I_{lr} and I_{hr} are also approximated in the frequency domain as subsamples y_f, y_s of the *k-space* signal y. Due to the different sampling rates employed, y_f and y_s share essentially the same low-frequency components, meanwhile the subsampled y_s contains richer high-frequency details. As illustrated in Fig. 1, the 2D Fourier transforms $\mathcal{F}(I_{hr})$ and $\mathcal{F}(I_{lr})$ of the high-resolution image I_{hr} and the low-resolution image I_{lr}, respectively, exhibit similarity in the central portion of

their magnitude Spectrum. Specifically, $\mathcal{F}(I_{hr})$ and $\mathcal{F}(I_{lr})$ share nearly identical low-frequency components. However, $\mathcal{F}(I_{lr})$ lacks the edge details present in $\mathcal{F}(I_{hr})$, indicating that $\mathcal{F}(I_{hr})$ contains a greater number of high-frequency components. The SISR technique for medical imaging focuses on learning a transformation function that maps input low-resolution images I_{lr} to their corresponding high-resolution images I_{hr}. Simultaneously, their frequency-domain representations are transformed from an approximation $\widehat{y}_f$ of y_f to an approximation $\widehat{y}_s$ of y_s. We use the following formula to convert $\widehat{y}_f$ to $\widehat{y}_s$:

$$\dot{y}_s = soft(\widehat{y}_f; \theta_s) + \mathcal{G}(I_{lr}; \theta_g) \tag{6}$$

where *soft* denotes a soft threshold function [13] with θ_s as the threshold, which removes the high-frequency details with lower energy in $\widehat{y}_f$ while retaining the low-frequency components where energy is concentrated. Meanwhile, $\mathcal{G}$ represents a priori model with conditional inputs parameterized by θ_g, generating the high-frequency components of $\dot{y}_s$ based on the input low-resolution image. In this case, the objective of the SISR system is to find the optimal parameter θ_s, θ_g such that the difference between $\dot{y}_s$ and $\widehat{y}_s$ is minimized:

$$\min_{\theta_s, \theta_g} \|\dot{y}_s - \widehat{y}_s\|_2^2 \tag{7}$$

III. Method

In this section, we introduce our neural network-based SISR model. As shown in Fig. 2, our network model takes a low-resolution MRI image $I_{lr} \in \mathcal{R}^{1\times H\times W}$ as input and obtains a high-resolution output of size $I_{hr} \in \mathcal{R}^{1\times 2^N H\times 2^N W}$ through N up-sampling modules. Next, we provide a detailed description of the various modules within the network.

A. Initialization Module

The initialization module serves as the foundational stage of the network architecture, responsible for transforming the input low-resolution MRI images I_{lr} into rich, high-level feature representations e. Unlike standard preprocessing pipelines that may rely on interpolation or handcrafted filters, this module is learnable and designed to capture both local anatomical structures and global intensity patterns directly from the raw image data. By applying a series of learnable convolutional layers, the model maps the pixel intensities of I_{lr} into a structured latent space, effectively encoding spatial and contextual information that is crucial for subsequent processing stages. This deep feature embedding not only preserves essential diagnostic content but also enhances robustness to noise and artifacts commonly present in clinical MRI acquisitions.

To enable high-dimensional feature learning and improve the model's capacity for discriminative representation, we employ a dedicated convolutional block that expands the channel dimensionality of the initial features from a low-dimensional input space to a higher-dimensional latent space. Specifically, the feature dimension is expanded from 1 channel (grayscale intensity) to n_{emb} dimensions, where n_{emb} is set to 96 by default. This dimensional expansion allows the network to learn a diverse set of feature maps that capture complementary aspects of the underlying anatomy, such as edges, textures, and regional contrasts. The convolutional block typically consists of a 3×3 convolution followed by normalization and a non-linear activation function (e.g., GELU), ensuring effective feature separation and expressive power for downstream tasks such as super-resolution or segmentation.

B. Up-sampling Module

We employ a novel up-sampling module to achieve $2\times$ up-sampling of image features. According to Eq. 6, the up-sampling module separates feature up-sampling into two components, high frequency and low frequency, to achieve the operations described as $soft$ and $\mathcal{G}$.

a) Low-Frequency Feature Module: We first up-sample the feature $e \in \mathcal{R}^{n_{emb}\times H\times W}$ using a transpose convolution block $\mathcal{U}_l$ to obtain $\mathcal{U}_l(e) \in \mathcal{R}^{n_{fea}\times 2H\times 2W}$, where n_{fea} is the output feature dimensions (default is 64). Given that e is a deep feature extracted from the image space, to preserve its high-energy components, we need to first transform e into an alternative domain before applying soft threshold filtering. Motivated by the invertible properties of domain transformation methods, such as the Fourier transform [14], wavelet transform [15], and discrete cosine transform [16], we design two domain transformation operators ϕ and $\tilde{\phi}$, where $\tilde{\phi}$ serves as the inverse operator of ϕ, i.e., $\tilde{\phi} \times \phi = \mathcal{I}$, with $\mathcal{I}$ being the identity operator. Specifically, ϕ is designed to exhibit a structure symmetric to that of $\tilde{\phi}$, and is therefore modeled as two linear convolutional operators separated by a GeLU operator, as illustrated in Fig. 2. Since both ϕ and $\tilde{\phi}$ are learnable, we incorporate symmetric constraints of the form $\|\tilde{\phi} \times \phi - \mathcal{I}\|_2^2 = 0$ into the loss function during network training. Therefore, Low frequency up-sampling features $e_l \in \mathcal{R}^{n_{fea}\times 2H\times 2W}$ can be efficiently computed in closed-form as:

$$e_l = \tilde{\phi}(soft(\phi(\mathcal{U}_l(e)); \theta_s)) \tag{8}$$

b) High-Frequency Feature Module: We use an LTE-based [8] network structure to generate high-frequency features e_h for images. First, we use the standard RSTB (residual swin transformer blocks) [17] as the main component of the encoder $\mathbb{E}$. The RSTB is a building block within the Swin Transformer that combines the strengths of hierarchical attention mechanisms and residual learning. The encoder $\mathbb{E}$ consists of 4 RSTB layers, each with a depth of 6 and 6 attention heads per Swin Transformer layer. For an input $e \in \mathcal{R}^{n_{emb}\times H\times W}$, the encoder processes it through the RSTB, utilizing a convolutional block to generate the output $\mathbb{E}(e) \in \mathcal{R}^{n_{fea}\times H\times W}$. The encoder is capable of extracting natural image features within the receptive field (RF), thereby aiding the LTE in estimating crucial high-frequency information in the transform domain. Inspired by position encoding [18] and Fourier feature mapping [19], the Local Texture Estimator(LTE) is a dominant frequency estimator for images. LTE transforms input coordinates into the transform domain before passing them through an MLP to address

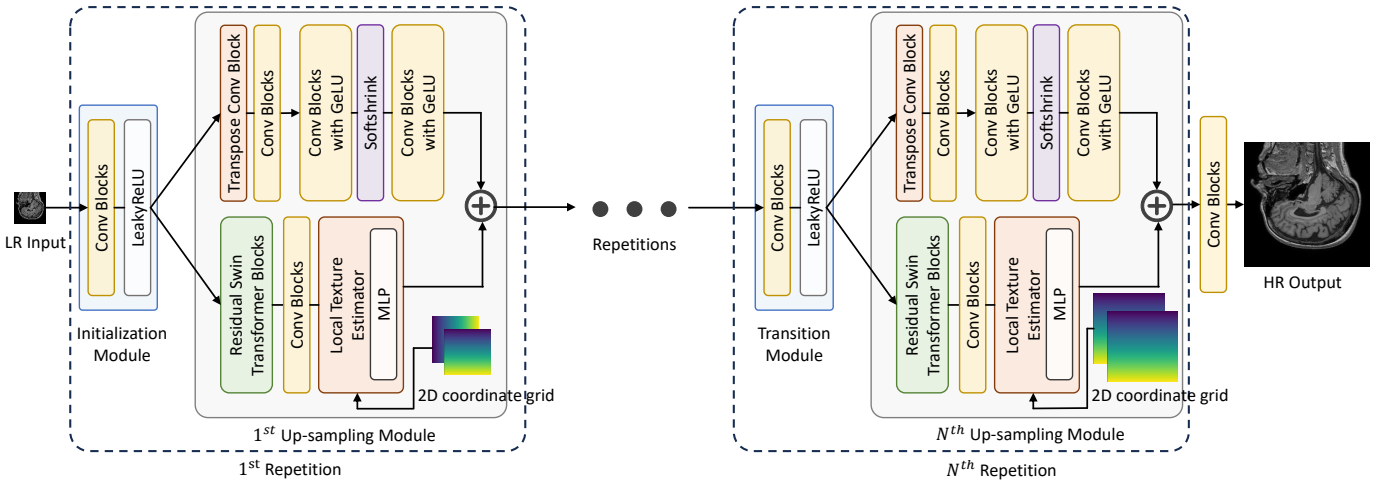

Fig. 2. Detailed framework of our SISR model for MRI. Each repetition comprises a transition module and an up-sampling module, designed to upscale the image features by a factor of 2. The network model contains N repetitions, allowing it to up-sample the image features by a factor of 2^N. Notably, the transition module in the first repetition is replaced by an initialization module.

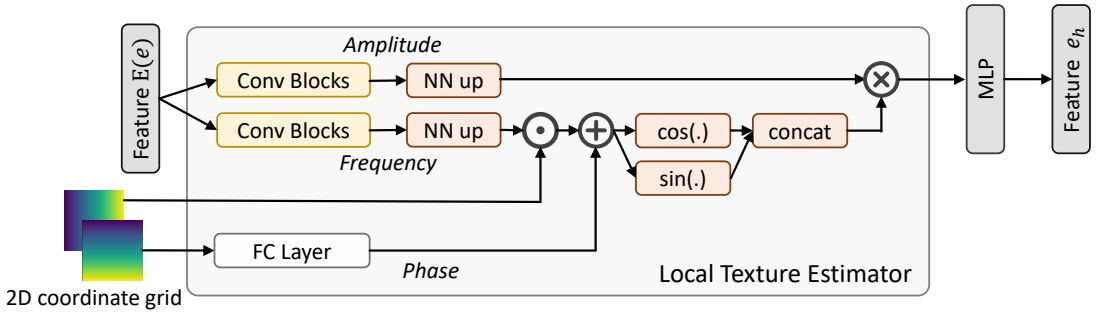

Fig. 3. The architecture of Local Texture Estimator(LTE). The LTE takes as inputs the feature map $\mathbb{E}(e)$ from the encoder, along with a local grid and cell. LTE then transforms these input coordinates into the transform domain by extracting amplitude, frequency, and phase information. The resulting output is passed through a decoder (MLP) to generate the high-frequency features e_h.

the spectral bias problem inherent in implicit neural functions. LTE's estimated transform information is data-driven and captures image textures within the 2D transform space. Immediately after that, the decoder $\mathbb{D}$ maps the latent tensor and local coordinates back to the image feature domain. The decoder $\mathbb{D}$ is a 4-layer MLP with ReLU activation, with default hidden dimensions 256 and default output dimensions n_{fea}. Ultimately, the local implicit neural representation as a high-frequency feature can be obtained by LTE using the following equation:

$$e_h(x,y) = \sum_{i\in\Psi} w_i \mathbb{D}(\Upsilon(\mathbb{E}(e), x - x_i, y - y_i)) \tag{9}$$

Here, Υ denotes the LTE, which is shift-invariant. As shown in Fig. 3, it consists of three elements: an amplitude estimator, a frequency estimator, and a phase estimator. Moreover, Ψ is a set of indices for the four nearest latent codes (in terms of Euclidean distance) around position (x, y), and w_i represents the bilinear interpolation weight corresponding to latent code i (referred to as the local ensemble weight [7]), with $\sum_{i\in\Psi} w_i = 1$). We fix LTE to perform only 2x scaling, resulting in $e_h \in \mathcal{R}^{n_{fea}\times 2H\times 2W}$.

Finally, the up-sampled image features can be obtained as follows:

$$e_{up} = e_l + e_h, e_{up} \in \mathcal{R}^{n_{fea}\times 2H\times 2W} \tag{10}$$

C. Transition Module

When feature $e \in \mathcal{R}^{n_{emb}\times H\times W}$ is input, the up-sampling module outputs an upsampled feature $e_{up} \in \mathcal{R}^{n_{fea}\times 2H\times 2W}$. Similarly, when the feature $e^N \in \mathcal{R}^{n_{emb}\times 2^{N-1}H\times 2^{N-1}W}$ is input, the up-sampling module outputs the image feature $e^N_{up} \in \mathcal{R}^{n_{fea}\times 2^N H\times 2^N W}$. To enable the inputs from the previous up-sampling module to be passed to the next up-sampling module, we introduce a transition module to adjust the feature shapes accordingly. The transition module uses a convolutional block to map the output $e^j_{up} \in \mathcal{R}^{n_{fea}\times 2^j H\times 2^j W}$ of the j-th up-sampling module to $e^{j+1} \in \mathcal{R}^{n_{emb}\times 2^j H\times 2^j W}$, which is then used as the input for the $(j+1)$-th up-sampling module ($j = 1, 2, 3, \ldots, N$).

TABLE I
PSNR, SSIM, AND LPIPS RESULTS FOR ALL COMPARED MODELS AT UPSCALING FACTORS 2X, 4X, AND 8X ON THE IXI T1, T2, PD, AND SIMON DATASETS.

Dataset	Model	2x			4x			8x		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
IXI T1	Bicubic	28.89	0.899	0.057	22.64	0.677	0.308	19.53	0.510	0.524
	EDSR [6]	35.39	0.965	0.025	27.90	0.849	0.118	24.24	0.714	0.233
	LIIF [7]	35.52	0.966	0.024	28.14	0.859	0.109	24.82	0.743	0.195
	LTE [8]	35.04	0.958	0.031	27.59	0.843	0.116	24.13	0.698	0.215
	DRFSA [4]	35.37	0.965	0.026	28.23	0.859	0.111	24.73	0.738	0.199
	ArSSR [9]	30.22	0.921	0.052	23.86	0.727	0.264	20.73	0.559	0.434
	ours	**35.71**	**0.968**	**0.023**	**28.56**	**0.870**	**0.101**	**24.92**	**0.750**	**0.188**
IXI T2	Bicubic	28.82	0.917	0.046	22.52	0.740	0.233	19.51	0.596	0.419
	EDSR [6]	35.51	0.969	0.020	26.83	0.864	0.101	22.59	0.738	0.214
	LIIF [7]	35.77	0.969	0.020	27.12	0.868	0.098	22.85	0.746	0.197
	LTE [8]	35.49	0.970	0.020	27.48	0.876	0.089	23.72	0.763	0.164
	DRFSA [4]	35.31	0.968	0.020	26.91	0.868	0.104	22.66	0.743	0.205
	ArSSR [9]	30.42	0.936	0.041	23.81	0.776	0.209	20.55	0.629	0.373
	ours	**37.41**	**0.976**	**0.013**	**28.48**	**0.899**	**0.076**	**24.37**	**0.803**	**0.143**
IXI PD	Bicubic	29.39	0.927	0.043	22.64	0.752	0.229	19.07	0.591	0.426
	EDSR [6]	36.98	0.974	0.018	28.77	0.884	0.087	23.86	0.765	0.179
	LIIF [7]	37.16	0.975	0.018	28.85	0.887	0.083	24.04	0.773	0.162
	LTE [8]	36.37	0.975	0.019	28.89	0.894	0.079	24.96	0.777	0.147
	DRFSA [4]	36.85	0.973	0.019	28.78	0.887	0.090	23.82	0.766	0.169
	ArSSR [9]	30.99	0.944	0.038	24.02	0.792	0.199	20.30	0.633	0.355
	ours	**38.62**	**0.980**	**0.012**	**30.46**	**0.915**	**0.069**	**26.25**	**0.829**	**0.128**
SIMON	Bicubic	31.84	0.924	0.050	25.58	0.745	0.238	22.55	0.600	0.450
	EDSR [6]	37.63	0.970	0.029	29.85	0.873	0.114	25.34	0.724	0.257
	LIIF [7]	**37.94**	**0.971**	0.028	30.21	**0.876**	0.119	25.61	0.734	0.239
	LTE [8]	36.03	0.965	0.038	29.75	0.858	0.127	24.73	0.678	0.310
	DRFSA [4]	36.09	0.963	0.039	29.93	0.874	0.115	25.53	0.727	0.242
	ArSSR [9]	33.29	0.940	0.047	27.02	0.783	0.214	23.68	0.635	0.414
	ours	36.52	**0.971**	**0.027**	**30.22**	0.873	**0.111**	**25.79**	**0.741**	**0.236**
SISRMRI	Bicubic	34.91	0.916	0.031	27.70	0.670	0.285	24.15	0.574	0.537
	EDSR [6]	43.62	0.981	0.014	31.78	0.797	0.181	27.85	0.672	0.394
	LIIF [7]	43.08	0.981	0.015	31.55	0.792	0.174	27.61	0.662	0.438
	LTE [8]	42.97	0.980	0.015	31.86	0.798	0.180	27.92	0.674	0.407
	DRFSA [4]	43.91	**0.983**	**0.012**	31.73	0.798	**0.170**	27.81	0.672	0.396
	ArSSR [9]	38.31	0.956	0.026	28.78	0.709	0.273	26.16	0.628	0.504
	ours	**44.57**	**0.983**	**0.012**	**31.96**	**0.801**	0.172	**27.99**	**0.676**	**0.388**

Finally, an additional convolutional block maps the output features e_{up}^{N} of the last up-sampling module back into the image space, producing the high-resolution MRI image $\widehat{I}_{hr} \in \mathcal{R}^{1\times 2^N H\times 2^N W}$.

D. Training Strategy

Given the training data pairs (I_{lr},I_{hr}), the model takes I_{lr} as input and generates the high-resolution MRI image $\widehat{I}_{hr}$. We seek to minimize the difference between I_{hr} and $\widehat{I}_{hr}$ while satisfying symmetry constraints: $\tilde{\phi}^j \times \phi^j = \mathcal{I}, \forall j \in [1,\ldots,N]$ [20], [21]. Considering that the integration of edge loss into the training process has been demonstrated to significantly enhance the quality of super-resolved images, especially in areas where fine details and sharp transitions are crucial for perceptual accuracy, we incorporate an edge loss function to penalize the discrepancies in edge features between I_{hr} and $\widehat{I}_{\text{hr}}$ [22]. Therefore, we design the end-to-end loss function for our model as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{difference} + \lambda\mathcal{L}_{constraint} + \mu\mathcal{L}_{edge} \tag{11}$$

with:

$$\begin{cases} \mathcal{L}_{difference} = |\widehat{I}_{hr} - I_{hr}| \\ \mathcal{L}_{constraint} = \sum_{j=1}^{N} \|\tilde{\phi}^j(\phi^j(\mathcal{U}_l(e^j))) - \mathcal{U}_l(e^j)\|_2^2 \\ \mathcal{L}_{edge} = \|Edge(I_{hr}) - Edge(\widehat{I}_{hr})\|_2^2 \end{cases} \tag{12}$$

where λ and μ is the regularization parameter, $j \in [1,\ldots,N]$ refers to the j-th up-sampling module, and $Edge(\cdot)$ represents the edge map of an image, typically obtained using a Sobel filter [23] by default. In our experiments, λ is set to 0.01 and μ is set to 0.01.

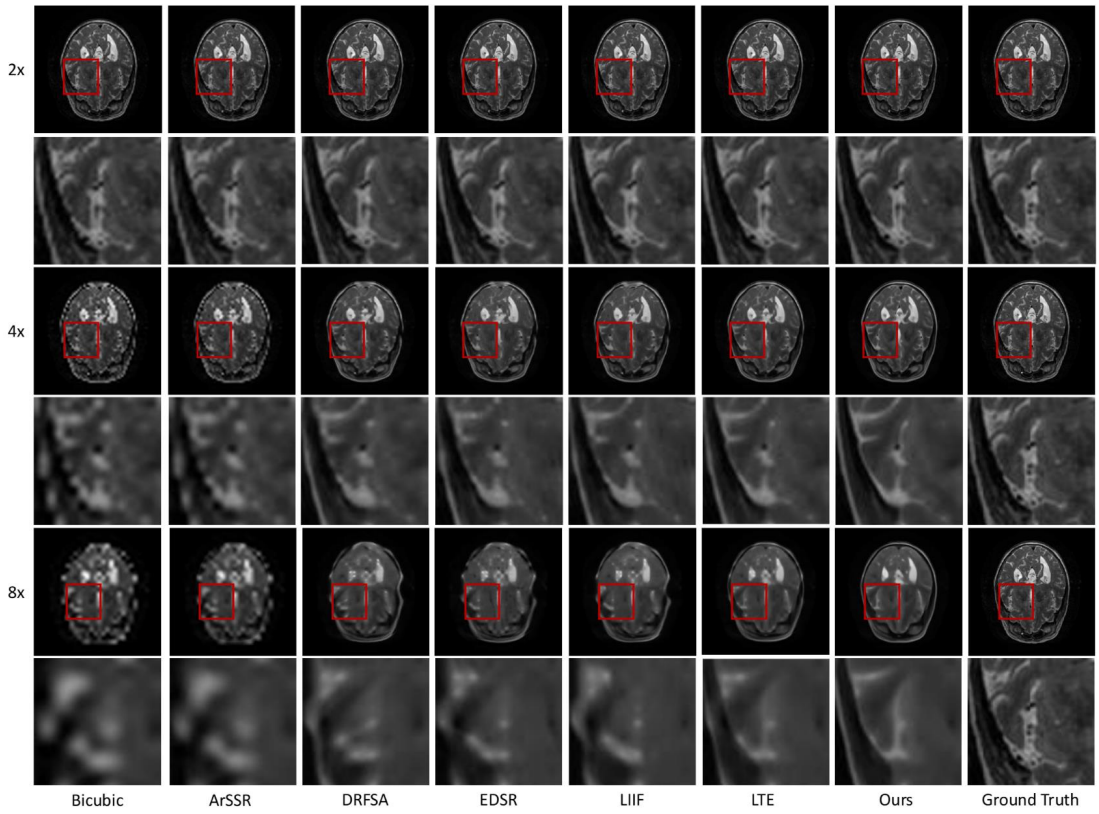

Fig. 4. Qualitative comparison of different methods at upscaling factors 2x, 4x, and 8x.

IV. Experiments

A. Experimental Setup

a) Datasets: To demonstrate the effectiveness of our method, we evaluate our model on three datasets.

Dateset1-IXI [24]. The IXI dataset, comprising over 600 diverse MRI scans from multiple London hospitals, provides high-resolution images ideal for evaluating SISR techniques. This dataset, particularly its T1-weighted images, is a standard benchmark in medical imaging research, offering variability that supports robust algorithm assessment.

Dateset2-SIMON [25]. The SIMON dataset includes various MRI modalities and captures a wide range of patient demographics, making it an excellent resource for developing and assessing image processing algorithms. Its extensive and varied data ensure robust evaluation of SISR methods, particularly in handling different imaging conditions and anatomical structures.

Dataset3-SISRMRI Our dataset comprises data collected from June 2014 to June 2017 at a famous hospital in China. During this period, a total of 1015 consecutive patients suspected of having renal cell carcinoma (RCC) underwent preoperative MRI examinations. Among these patients, those diagnosed with clear cell RCC (ccRCC) and who had pathologically confirmed pseudocapsule formation were included in this study.

All MRI scans were performed in the supine position using a 3.0T MR scanner (Ingenia; Philips Medical Systems, Best, the Netherlands). Two T2-weighted imaging sequences with different spatial resolutions were retrospectively analyzed. The imaging parameters were as follows: TR/TE = 1500/90 ms, flip angle = 90°, slice thickness = 4 mm. The high-resolution sequence had a voxel size of 0.3×0.3 mm with a scan time of 7 minutes and 30 seconds, while the low-resolution sequence had a voxel size of 0.9×0.9 mm and a scan time of 2 minutes and 5 seconds.

These MRI datasets are systematically partitioned into training, validation, and testing subsets with a ratio of 7:2:1, respectively. This division is carefully conducted based on distinct individuals to ensure that each subset adequately represents the diversity of the data.

Evaluation metrics. Peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [26], learned perceptual image patch similarity (LPIPS) [27] are employed to assess the similarity between the super-resolution (SR) results and the ground truth images.

b) Implementation Details: In our experiments, we utilized the Adam optimizer [28] to train the model for 100 epochs. The Adam optimizer is chosen due to its adaptive learning rate mechanism, which computes individual learning rates for different parameters. Specifically, the optimizer was initialized with a learning rate of $1e-4$, with $\beta_1 = 0.9, \beta_2 = 0.999$ and $\epsilon = 1e-8$. These values are standard for Adam and were selected based on their effectiveness in stabilizing training across a wide range of models. To enhance the conver-

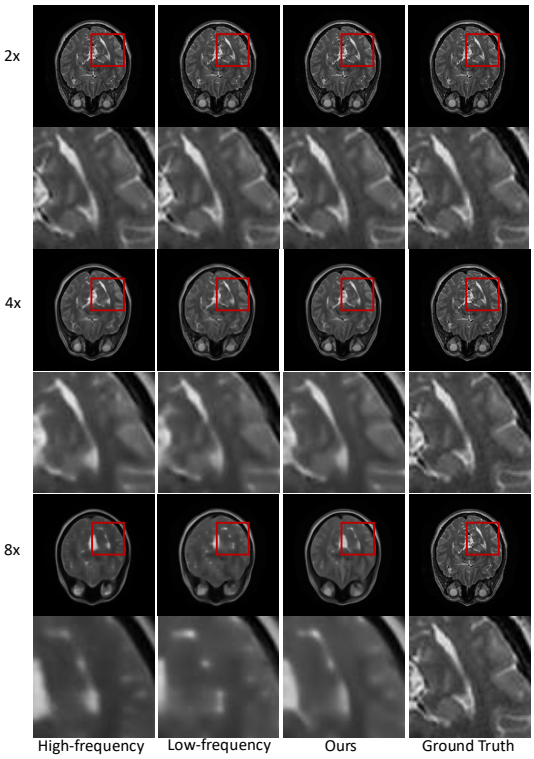

Fig. 5. The ablation experiments of our system at upscaling factors 2x, 4x, and 8x.

gence and performance of the model, we employed a learning rate scheduler, specifically the MultiStepLR scheduler [29]. The scheduler decreases the learning rate by a factor of 0.1 at predefined epochs, specifically at the 40th and 80th epochs. This step-wise reduction in the learning rate helps the model to fine-tune and reach a more optimal solution by allowing larger steps during the initial phase of training and smaller, more precise steps as the model converges. The batch size of the training phase is 8, and the size of each high-resolution MRI image is 256×256. We use PyTorch [29] to implement models on an NVIDIA Tesla V100 GPU.

B. Experimental Results

Qualitative Comparison. The experimental results summarized in Tab. I demonstrate the efficacy of our proposed neural network approach to MRI single-image super-resolution, as evidenced by consistent improvements across multiple datasets (IXI T1, T2, PD, and SIMON) and up-sampling factors (2x, 4x, and 8x). Compared to state-of-the-art methods such as EDSR [6], LIIF [7], LTE [8], DRFSA [4], and ArSSR [9], our model achieves superior performance in terms of peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS). Specifically, our model outperforms others with a notable margin, particularly on challenging up-sampling factors like $4\times$ and $8\times$. For instance, on the IXI T2 dataset at $4\times$ up-sampling, our model achieves a PSNR of 28.48 dB, an SSIM of 0.899 and an LPIPS of 0.076, which are significantly higher than the closest competitor, LTE, with a PSNR of 27.48 dB, an SSIM of 0.876 and an LPIPS of 0.089, respectively.

As shown in Fig. 4, the reconstructed high-resolution images using our method show significantly fewer artifacts and better preservation of fine details compared to the other methods. For example, while the Bicubic interpolation results are heavily blurred and fail to capture the intricate details of the MRI scans, and ArSSR introduces noticeable distortions, our approach maintains the structural integrity and sharpness of the images. These visual and quantitative results collectively demonstrate that our SISR model excels in the perceptual quality of the reconstructed high-resolution MRI images.

C. Ablation Study

To further validate the effectiveness of our proposed network architecture, we conducted an ablation study focusing on the contributions of the high-frequency and low-frequency feature modules in our model. The results, presented in Tab. II and Fig. 5, illustrate the performance of the network when using only the high-frequency feature module, only the low-frequency feature module, and the full model that integrates both. When the network is configured to utilize only the high-frequency feature module, a degradation in performance is observed across all up-sampling factors ($2\times$, $4\times$, and $8\times$). Specifically, the PSNR and SSIM values decrease, while the LPIPS score increases, indicating that although the model captures some fine details, it struggles with overall image quality and structural similarity. For instance, at $4\times$ upsampling, the PSNR is 27.44 dB and the SSIM is 0.878, both of which are lower than those achieved by the full model.

Conversely, when the network is constrained to utilize only the low-frequency feature module, its performance is also inferior compared to that of the full model. The network's ability to reconstruct fine details is constrained, as indicated by a relatively higher LPIPS score. For instance, at $8\times$ upscaling, the PSNR is 23.60 dB, which is lower than that of the full model, and the LPIPS score is 0.166, signifying a perceptual quality loss.

The full model, which integrates both high-frequency and low-frequency feature modules, yields the best results, underscoring the significance of combining both modules. This configuration attains the highest PSNR and SSIM values, as well as the lowest LPIPS scores, across all up-sampling factors. At $4\times$ up-sampling, the full model achieves a PSNR of 28.48 dB and an SSIM of 0.899, significantly outperforming the other configurations. These results confirm that both frequency feature modules are crucial for achieving superior image super-resolution, and their integration is essential for the model's optimal performance.

V. Conclusion

This study presents a novel Single-image Super-resolution system (SISR), the proposed up-sampling structure is designed to address the challenges of super-resolution magnetic resonance imaging by efficiently enhancing both high-frequency

TABLE II
QUANTITATIVE RESULTS OF ABLATION EXPERIMENTS ON THE IXI-T2 DATASET.

High-frequency	Low-frequency	2x			4x			8x		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
✓		36.32	0.972	0.015	27.44	0.878	0.090	22.78	0.728	0.163
	✓	35.11	0.965	0.019	27.34	0.875	0.090	23.60	0.776	0.166
✓	✓	**37.41**	**0.976**	**0.013**	**28.48**	**0.899**	**0.076**	**24.37**	**0.803**	**0.143**

and low-frequency image components. In addition, the enhanced loss function integrating symmetry and edge constraints greatly enhances the preservation of critical structural details essential for medical diagnosis. The results show that our method not only improves the quality of super-resolution images but also has great potential for widespread application in clinical settings, especially in resource-limited healthcare environments.

ACKNOWLEDGMENT

This work is supported in part by National Natural Science Foundation of China (No. 61936015), Natural Science Foundation of Shanghai (No. 24ZR1430600) and Shanghai Key Laboratory of Trusted Data Circulation and Governance, and Web3.

REFERENCES

[1] Z. Zhou, M. Qutaish, Z. Han, R. M. Schur, Y. Liu, D. L. Wilson, and Z.-R. Lu, "Mri detection of breast cancer micrometastases with a fibronectin-targeting contrast agent," *Nature communications*, vol. 6, no. 1, p. 7984, 2015.

[2] C.-M. Feng, H. Fu, S. Yuan, and Y. 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*. Springer, 2021, pp. 140–149.

[3] C.-M. Feng, Y. Yan, H. Fu, L. Chen, and Y. 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*. Springer, 2021, pp. 307–317.

[4] X. Wang, J. Xia, S. Gao, X. Hao, and Y. Zhou, "Deep residual fourier and self-attention for arbitrary scale mri super-resolution," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2023, pp. 666–671.

[5] Z. Liu, X. Wang, Z. Wu, Y.-C. Zhu, and A. F. Frangi, "Simultaneous super-resolution and denoising on mri via conditional stochastic normalizing flow," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2023, pp. 1313–1318.

[6] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[7] Y. Chen, S. Liu, and X. Wang, "Learning continuous image representation with local implicit image function," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 8628–8638.

[8] J. Lee and K. H. Jin, "Local texture estimator for implicit representation function," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1929–1938.

[9] Q. Wu, Y. Li, Y. Sun, Y. Zhou, H. Wei, J. Yu, and Y. Zhang, "An arbitrary scale super-resolution approach for 3d mr images via implicit neural representation," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 2, pp. 1004–1015, 2022.

[10] R. H. Hashemi, W. G. Bradley, and C. J. Lisanti, *MRI: the basics: The Basics*. Lippincott Williams & Wilkins, 2012.

[11] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno *et al.*, "fastmri: An open dataset and benchmarks for accelerated mri," *arXiv preprint arXiv:1811.08839*, 2018.

[12] Y. Yang, J. Sun, H. Li, and Z. Xu, "Admm-net: A deep learning approach for compressive sensing mri," *arXiv preprint arXiv:1705.06869*, 2017.

[13] D. L. Donoho, "De-noising by soft-thresholding," *IEEE transactions on information theory*, vol. 41, no. 3, pp. 613–627, 1995.

[14] H. J. Nussbaumer and H. J. Nussbaumer, *The fast Fourier transform*. Springer, 1982.

[15] D. Zhang and D. Zhang, "Wavelet transform," *Fundamentals of image data mining: Analysis, Features, Classification and Retrieval*, pp. 35–44, 2019.

[16] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE transactions on Computers*, vol. 100, no. 1, pp. 90–93, 1974.

[17] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.

[18] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[19] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in neural information processing systems*, vol. 33, pp. 7537–7547, 2020.

[20] J. Zhang and B. Ghanem, "Ista-net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1828–1837.

[21] D. You, J. Xie, and J. Zhang, "Ista-net++: Flexible deep unfolding network for compressive sensing," in *2021 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2021, pp. 1–6.

[22] G. Seif and D. Androutsos, "Edge-based loss function for single image super-resolution," in *2018 IEEE International conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2018, pp. 1468–1472.

[23] H. B. Fredj, M. Ltaif, A. Ammar, and C. Souani, "Parallel implementation of sobel filter using cuda," in *2017 International Conference on Control, Automation and Diagnosis (ICCAD)*. IEEE, 2017, pp. 209–212.

[24] I. C. London. [Online]. Available: https://brain-development.org/ixi-dataset/

[25] S. Duchesne, I. Chouinard, O. Potvin, V. S. Fonov, A. Khademi, R. Bartha, P. Bellec, D. L. Collins, M. Descoteaux, R. Hoge *et al.*, "The canadian dementia imaging protocol: harmonizing national cohorts," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 2, pp. 456–465, 2019.

[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

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

[28] D. Kingma, "Adam: a method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.