MRI Super-Resolution using Implicit Neural Representation with Frequency Domain Enhancement

SHUANGMING MAO
SEIICHIRO KAMATA
shuangming@suou.waseda.jp
kamata@kamata-lab.org
Waseda University
Fukuoka, Japan

ABSTRACT

High resolution (HR) Magnetic Resonance Imaging (MRI) is a popular diagnostic tool, which provides detail structural information and rich textures, benefiting accurate diagnosis and disease detection. However, obtaining HR MRI remains a challenge due to longer scan time and lower peak signal-to-noise ratio (PSNR). Recently, Single Image Super-Resolution (SISR) has generated interest, which shows promising ability for recovering an HR image only relies on a Low Resolution (LR) image. MR images have some characteristics different with natural images: *derived from frequency domain, simpler textures and structural information.* However, Most of previous methods treat MR images as same as natural images, they only apply SR methods on natural images to MR images and fail to preserve low-frequency information and capture high-frequency details. In this paper, we mimic the process of an MRI machine produces an MRI in practice and propose an Implicit Neural Representation based module, which enable reconstruct high frequency contents effectively while preserving low frequency contents unchanged. Moreover, vanilla L1 loss cannot reflect the differences for each frequency, to address this problem, we design a *frequency loss* to disentangle each frequency and calculate the differences respectively. Finally, to further capture high frequency contents, we propose *High-Frequency Pixel Loss*, which can decouples the HF contents from pixel domain and emphasize the HF differences between SR and HR images. Extensive experiments show the effectiveness of our proposed method in terms of visual quality and PSNR score, which produces sharper edges and clearer details compared to previous works.

CCS CONCEPTS

• **Computing methodologies** → **Biometrics**.

KEYWORDS

MRI, Super-Resolution, Implicit Neural Representation

ICBIP 2022, August 19–21, 2022, Suzhou, China

ACM ISBN 978-1-4503-9669-1/22/08…$15.00
https://doi.org/10.1145/3563737.3563759

ACM Reference Format:
SHUANGMING MAO and SEIICHIRO KAMATA. 2022. MRI Super-Resolution using Implicit Neural Representation with Frequency Domain Enhancement. In *2022 7th International Conference on Biomedical Signal and Image Processing (ICBIP) (ICBIP 2022), August 19–21, 2022, Suzhou, China.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3563737.3563759

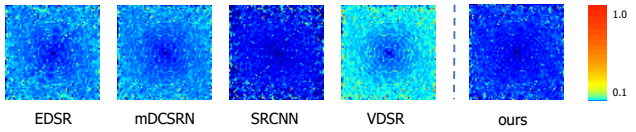


Figure 1: Visual comparison with error maps of low-frequency domain for different methods on dataset IXI-T2

1 INTRODUCTION

High Resolution (HR) Magnetic Resonance Imaging (MRI) provides accurate structural information and textures, which is useful to clinical diagnosis and quantitative image analyses. However, obtaining HR MR images remains a difficult challenge since it takes a lot of scan time, lower peak signal-to-noise ratio (PSNR) and smaller spatial converge [28]. Single Image Super-Resolution (SISR) becomes a promising technique to address this issue. SISR can produce HR image from a single Low-Resolution (LR) image, which can significantly reduce the scanning time. Therefore, improve the ability of SISR is urgently needed for decades.

Deep convolutional neural networks (CNNs) show its outstanding performance for Image Super-Resolution [4, 14, 17, 18, 40, 45]. There are 3 major directions to improve the performance for CNN-based SR: 1. Design a more complicated and deeper architectures. 2. Provide additional information with help of extra inputs. 3. Convert SR tasks to another tasks. For instance, VDSR [14] increase the network depth with Residual Learning [8]. SRNTT [45] provides an extra HR image as input and design a *Neural Texture Transformer* module to integrate texture information to CNN. [21] utilizes a pre-trained StyleGAN [13] to generate HR images, which converts SR problem to generative problem and produces promising results in Face SR.

However, previous CNN-based works ignore the characteristics of MR images and treat MR images as same as natural images. First, MRI Super-Resolution should keep the original information unchanged. Fig 1 shows the visual comparison for error maps of k-space between our method and another methods. As shown in

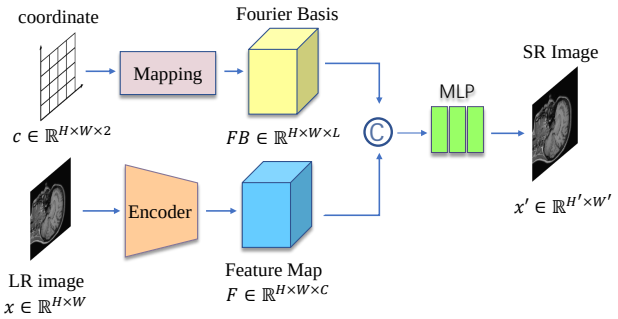


Figure 2: Single-scale Fourier Basis Module. The 2-D coordinates are mapped to a set of pre-defined Fourier Bases, a CNN Encoder extracts intermediate features from LR image. Then the Fourier bases and features are concatenated together and used as input to an MLP.

Fig 1, the low-frequency contents should be equal for SR and LR images. But those CNN-based methods cannot guarantee the consistency and show higher errors. Second, MRI is derived from k-space. However, most previous works do not take into account frequency domain and only make modification on pixel domain. Third, neural networks tend to prioritize learning the low frequency modes, this phenomenon called *spectral bias* [29], which leads to blurry results. Although GAN based methods [7] can produce clear results, they result in very low PSNR.

To tackle these problems, we mimic the process of MRI machine produces MR images in practice. Specifically, we propose an *Implicit Neural Representation* (INR) based architecture to convert SR task to learn a continuous function between a set of orthogonal bases and HR images, which is similar with using Inverse Fourier Transform (IFFT) to transform a set of Fourier Bases to image in practice. As shown in Fig 2, unlike previous works that rely only on LR image to get HR image, we aim to reconstruct HR images by the combination of Fourier bases. Specifically, we first define a series of orthogonal bases according to Fourier Series, then an MLP is used to approximate the coefficients of each basis given an LR image and combine them to an HR image. However, the behavior of neural network may not meet our expectation. Actually, it will omit the Fourier bases and convert to depend on LR image during optimization. To address the problem, we propose *Frequency Loss* to reflect the differences for each component, which explicitly forces neural networks focus on all frequencies. Moreover, to further capture high-frequency contents, we propose *High-Frequency Pixel Loss*, which is more efficient than L1 loss to capture high-frequency contents. Extensive experiments demonstrate the effectiveness of our compared to another methods.

In summary, the main contributions of our work are:

- We propose an INR based *Fourier Basis Module* to convert SR task to learn a continuous function between Fourier bases and HR image, which is similar with using Inverse Fourier Transform in practice.
- We propose a *Frequency Loss* to disentangle each frequency and calculate the differences respectively, which explicitly forces neural networks focus on all frequencies.
- We propose a *High-Frequency Pixel Loss*, which is more efficient than L1 loss to capture high-frequency contents.

2 RELATED WORK

2.1 Image Super-Resolution

Image Super-Resolution (SR) has gained rapid development due to Deep Learning, specifically, deep convolutional neural networks (CNNs). SRCNN [4] first develops CNN architecture to SR, which produces promising results compared to traditional SR methods though it is just a 3-layers CNN. After that, many works explore more complicated and deeper architectures to improve representational ability for NNs. VDSR [14] deepens the depth of NNs with skip-connection, it further improves the ability by using a very deep convolutional neural networks. EDSR [18] adopts Residual Learning [8] and removes unnecessary layers to largely extends the model size. RDN [44] adopts the densely connected network [10] to integrate low-level and high level-features. However, those methods still suffer severely ill-posed problem. On the other hand, recent works pay attention to Generative adversarial networks [17, 36], reference-based SR [40, 45] and generative model [20, 21, 30]. For instance, SRGAN uses GAN and perceptual loss [11] to improve perceptual similarity and high-frequency details. Saharia *et at.* [30] applies *Denoising diffusion probabilistic models* (DDPM) [9] to convert SR tasks to denoising process. It learns reversible noise-adding process, and image or details can be recovered by inverse the process.

In MRI Super-Resolution, there are some works promote the SR methods for natural images into MRI SR [3, 5, 6, 20, 27, 35, 42]. mDCSRN [3] develops a lightweight densely connected network, which makes a trade-off between time complexity and model capacity. It can achieve satisfying SR performance within an acceptable complexity. MRDG [35] enhances the discriminator by Unet-GAN [31] to improve the ability to repair local details. T2Net [6] makes a multi-task learning for SR and Reconstruction. Since multi-task models can express both shared and task-specific characteristics, both SR and reconstruction task can be improved. SERAN [42] overcomes the problem of restricted receptive field in CNN by GCN [16]. It aggregates the global descriptors to enhance the network's ability to focus on more informative regions and structures in MR images.

2.2 Implicit Neural Representation

For convenience, a signal is usually represented discretely in computer. In order to represent a signal continuously, an implicit function need to be established. For example, an object can be represented by Signed Distance Function (SDF). Continuously represent a scene in computer requires establish rendering equation. However, calculating such a rendering equation requires a lot of computational resources, making it impossible to perform in real time. *Implicit Neural Representation* (INR) is a novel way to represent a signal as continuous function. In the INR, a multi-layer perception (MLP) learns such a function between spatial coordinates and corresponding signal. It provides a general solution for various

applications of object reconstruction [1, 22–26]. For example, NeRF [23] approximates rendering equation for a scene, after the training of INR, it can synthesize the scene from different views by sampling different positions. DeepSDF [26] proposes to learn an SDF for each 3D object. Combine an auto-decoder model, it can represent any object by an INR. Mescheder *et at.* [22] propose to estimate an occupancy function by a Resent performing 3D reconstruction. StyleGAN3 [12] models images as infinitely continuous signals, which achieves translation invariance and rotation invariance for generated images. Recent works also demonstrate the potential of INR in SR. LIIF [2] aims to learn a contiguous function for images of any scales, which performs arbitrary up-sampling scales for LR images even if trained at a limited resolution. ArSSR [38] promotes LIIF to 3D MRI SR, which combines coordinates and LR MR images to predict the voxel intensity for HR MR images.

On the other hand, directly operate on low dimension coordinates performs poorly to represent high-frequency variation. There are some works attempt to address the problem to fit various tasks. NeRF [23] designs a Positional Encoding [34] $\gamma(p) = [\sin 2^0\pi p, \cos 2^0\pi p, \cdots, \sin 2^{L-1}\pi p, \cos 2^{L-1}\pi p]$ that projects coordinates to a higher dimensional space. Tancik *et at.* [33] propose a not axis-aligned encoding Fourier Features, which performs superior results than basic Positional Encoding. In Fourier Features, it uses a set of random bases sample from Gaussian distribution $w \sim \mathcal{N}(0, \sigma^2)$, then the Fourier features define as $\gamma(v) = [\cos(2\pi w_0 v), \cdots, \cos(2\pi w_i v)]$. Moreover, for different tasks, SIREN [32] argues that ReLU based network architectures are poor at learning complex periodic signals and proposes using sinusoidal as activation function. Xu *et at.* [39] propose Sinusoidal Positional Encoding (SPE) to improve spatial inductive bias for image generation. However, none of previous works focus on MR images. To process MR images, in our work, we follow the characteristic of MRI derived from well-defined k-space. We pre-define a set of Fourier bases as the representation of coordinates to replace the Positional Encoding. This is more effective to approximate the desired features for MR images because all defined frequencies are relevant to composite the target signal in practice.

3 METHODOLOGY

In this section, we will elaborate our proposals. We first briefly describe how to derive the Fourier basis, then detail introduce the components of our model and present its advantages over other methods.

3.1 Fourier Basis

Fourier Series is a weighted summation of sines and cosines functions, where each function is orthogonal so that all functions form a set of orthonormal basis, which are able to reconstruct any periodic function. For a 2-D image, we can see it as bounded periodic function and decompose it by Fourier Series. Here, we consider a 1-D discrete function $f(x)$, if we assume the period is 1, the Fourier expansion define for the function as:

$$f(x) = a_0 + \sum_{n \in N} a_n \sin(2\pi n x) + b_n \cos(2\pi n x) \tag{1}$$

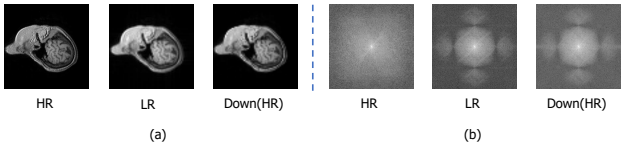


Figure 3: (a). Examples of HR MRI, LR MRI and LR MRI downsampled from HR MRI. (b). Corresponding k-space for the three images.

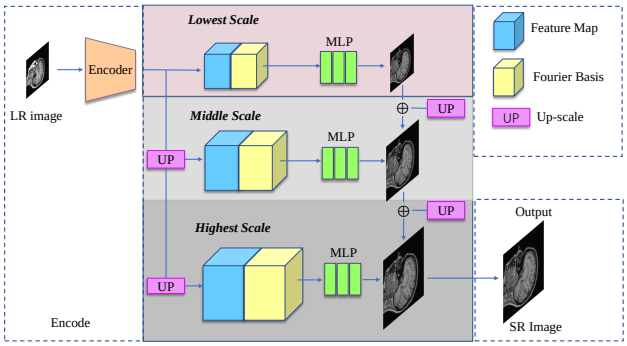


Figure 4: Multi-scale Fourier Basis Module. We adopt Single-scale module to the lowest scale, middle scale and highest scale respectively. An encoder first extracts the lowest features from LR image, then the middle feature and the highest can be got by one up-scale and twice up-scale. The three MLPs take features and Fourier basis as input and output an image corresponding to current resolution. Finally, the SR image is a sum of the three scale images.

where, N is the number of sampling points for $f(x)$, if write equation 1 into vector form, we can get:

$$f(x) = a_0 + WB^\top \tag{2}$$

where $W = [a_1, b_1, \cdots, a_n, b_n]$ and $B = [\sin(2\pi x), \cos(2\pi x), \cdots, \sin(2\pi n x), \cos(2\pi n x)]$. The vector B we call it Fourier Basis. We can see that the form of Fourier Basis is similar with Positional Encoding proposed in previous works. Indeed, the Positional Encoding can be seen as a sample in Fourier Basis. Since natural images have various contents and scales, we have no idea which frequency is relevant for all images. However, the MRI has single modal and contents so that we can easily select relevant frequencies as representation to obtain more efficient reconstruction.

3.2 Multi-scale Fourier Basis Module

The Fig 2 shows the single scale architecture of our model. However, it is not enough to reconstruct all high-frequency details, and suffers unstable training. Different with previous works [2, 42], the LR images are downsampled from HR images. The original LR MR images are derived from low frequency domain. Fig 3 (a) shows the comparison of the two types of LR images, we can see the downsampled LR image still contains some details. But the original LR image seldom contains high-frequency information. As shown

in Fig 3 (b), the k-space of downsampled LR image contains some high frequency components, but the high frequency components almost lost in original LR image.

Therefore, to capture more of this lost information. We extend our model to multi-scale. As shown in Fig 4, given an LR image $I_{LR} \in \mathbb{R}^{64\times64}$, A encoder E extracts the lowest scale features $F_l = E(I_{LR}), F_l \in \mathbb{R}^{64\times64\times128}$, then the middle scale features and highest scale features can be got: $F_m = UP(F_l), F_m \in \mathbb{R}^{128\times128\times128}$ and $F_h = UP(F_m), F_h \in \mathbb{R}^{256\times256\times128}$, where UP denotes bilinear interpolation. Next we adopt three Fourier basis modules for these three scales, the images of the three scales are obtained by:

$$\begin{aligned} I_l &= MLP(concat[F_l, B_l]) \\ I_m &= UP(I_l) + MLP(concat[F_m, B_m]) \\ I_{SR} &= UP(I_m) + MLP(concat[F_h, B_h]) \end{aligned} \tag{3}$$

Where, B denotes Fourier basis, *concat* denotes concatenation operation. We can see that we do not obtain the full resolution image directly, instead we recursively obtain the SR image from low resolution. There are two merits compared to single scale model. First, all scales contribute to the final output, this provides stable gradients for each module during backpropagation phase. Therefore, it greatly stabilizes the training. Second, learning the all frequency contents in full resolution is difficult during optimization, to alleviate the problem, our model separates this problem to three sub-tasks. Specifically, the next module can supplement the missing information of previous module. Each module does not need to learn all frequency information thus reducing the difficulty of training.

3.3 Frequency Loss

However, we observe that our model would ignore the Fourier basis and converts to rely only on LR image during training. Specifically, we find the improvement is limited compared to baseline We argue that this is because there is no guidance forces the model to learn high frequency contents with Fourier basis and the training will be deviated without guidance. To tackle this, we explicitly decompose an image into individual frequency components by Fourier Transform.

The Fourier Transform is widely used to analyze the frequency for a piece of signal. For an image $I \in \mathbb{R}^{H\times W}$, it can decompose the image depending on space into sets of functions depending on spatial frequency. Specifically, it calculates the coefficients in Fourier Series by inner product between the image and functions. Mathematically, the discrete Fourier transform (DFT) can be written as:

$$c_{u,v} = \frac{1}{\sqrt{HW}} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} I(x,y) e^{-2\pi i \frac{ux}{H}} e^{-2\pi i \frac{vy}{W}} \tag{4}$$

We can see that the Fourier Transform can decompose an image to sets of 2D waves $e^{-2\pi i(\frac{ux}{H}+\frac{vy}{W})}$ and c_{uv} denotes the corresponding coefficient. With the help of Fourier Transform, we define our Frequency Loss as:

$$\mathcal{L}_{freq} = ||FFT(I_{SR}) - FFT(I_{HR})||_1 \tag{5}$$

Unlike the vanilla L1 Loss that cannot reflect the differences for each frequency component, the Frequency Loss can be further decomposed as:

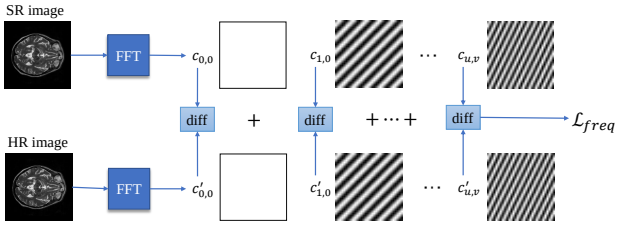


Figure 5: The concept of Frequency Loss

$$\begin{aligned} \mathcal{L}_{freq} &= ||FFT(I_{SR}) - FFT(I_{HR})||_1 \\ &= |c_{0,0} - c'_{0,0}| + |c_{1,0} - c'_{1,0}| + \cdots + |c_{u,v} - c'_{u,v}| \end{aligned} \tag{6}$$

As show in Fig 5, we can see that each term represents the difference for each 2D wave $e^{-2\pi i(\frac{ux}{H}+\frac{vy}{W})}$ between SR and HR image. And it plays two key roles in our model. First, it makes a supervision at low frequency domain in original LR image and SR image so that it can remain the original information unchanged, and this is crucial for MRI SR. Second, it directly emphasizes on the lost high frequency contents, which forces the model focuses on learning such high frequency contents and makes the model utilizes the information in Fourier basis.

3.4 High-Frequency Pixel Loss

The Frequency loss provides a supervision at frequency domain, we also want to supervise spatial domain to provide complementary information. The most straightforward way is to adopt L1 loss as supervision for spatial domain. However, we find the L1 loss does not contribute to the final performance (degrade 0.01 dB). We argue that L1 loss prefers to optimize the low-frequency contents thus converging to a local minimum. To tackle this problem, we decompose an image to a low frequency image and a high frequency image, then we only adopt L1 loss on the high frequency image. Specifically, given an image $I \in \mathbb{R}^{H\times W}$, we first adopt Gaussian filter as low pass filter to extract low frequency image:

$$\kappa[i,j] = \frac{1}{2\pi\sigma^2} e^{\frac{-(i^2+j^2)}{2\sigma^2}} \tag{7}$$

Then using convolution on the image I with the Gaussian filter:

$$I_L[x,y] = \sum_i \sum_j \kappa[i,j] I[x+i, y+j] \tag{8}$$

where I_L denotes low frequency image. Next we get the high frequency image by:

$$I_H = I - I_L \tag{9}$$

Note that more detail can be preserved in this way than using high pass filter directly. After obtaining the high frequency image, we define our High-Frequency Pixel Loss as:

$$\mathcal{L}_{hfp} = ||I_H^{SR} - I_H^{HR}||_1 \tag{10}$$

3.5 Overall Loss

The total loss function is as follows:

$$\mathcal{L}_{total} = \lambda_{freq}\mathcal{L}_{freq} + \lambda_{hfp}\mathcal{L}_{hfp} \tag{11}$$

Table 1: PSNR/SSIM values of different SR approaches. The best and second-best results are marked in red and blue, respectively.

Datasets	T1	T2
Metrics	PSNR/SSIM	PSNR/SSIM
SRCNN [4]	26.49/0.8466	26.68/0.8657
VDSR [14]	27.97/0.8822	28.56/0.9079
mDCSRN [3]	28.18/0.8832	28.56/0.9073
EDSR [18]	28.68/0.8955	28.96/0.9155
Ours	28.92/0.9009	29.20/0.9180

Where λ_{freq} and λ_{hfp} are balance weights that control the importance for each loss.

4 EXPERIMENTS

4.1 Datasets

We evaluate our method on IXI datasets. Three types of modality include in the datasets (PD, T1 and T2). Here we select T1 and T2 to evaluation. For each dataset, we use 450, 20 and 100 volumes for training, validation, and testing respectively. We exclude the first and last slices foe each volume since such slices are much noisier, so the size of each 3D volume is cut to $256 \times 256 \times 90$, where 256×256 indicates image size and 90 indicates the number of slices. Eventually, we get 40,500 samples for training.

4.2 Data preprocessing

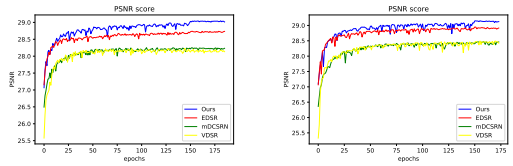
Unlike natural images that obtain LR images through bicubic interpolation from HR images. We obtain the paired LR and HR MR images following the steps as in [3]: 1) converting HR image into k-space using FFT: 2) truncating the outer part of the k-space with a factor 2: 3) converting back to image space by IFFT. This mimics the actual acquisition of LR and HR images in practice.

4.3 Implementation details

We use EDSR [18] backbone (16 residual blocks) as our encoder and set all convolutional layers as 3×3. In Fourier basis module, we set the dimension of all MLP layers to 128. In addition, the number of layers in MLP we set to 6, 5 and 4 for lowest, middle and highest scales respectively. Our network is trained using ADAM optimizer [15] with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$ for 200 epochs. The batch size is set to 8. The learning rate is initialized as 10^{-4} and decreases to half after 150 epochs.

4.4 Results and evaluation

Compared approaches. To verify the effectiveness of our proposed method, we compare with several typical SR techniques, include SRCNN [4], VDSR [14], EDSR [18] and mDCSRN [3], where SRCNN, EDSR and VDSR are for natural images, mDCSRN is for MRI. In addition, in order to fair comparison, we enhance the mDCSRN to similar number of layers and complexity with our network. Specifically, we enhance the original mDCSRN from its original configuration *b4u4* to *b8u4*, 8 dense blocks with 4 dense layers within each block.



(a) T1 (b) T2

Figure 6: PSNR score curve on validation set

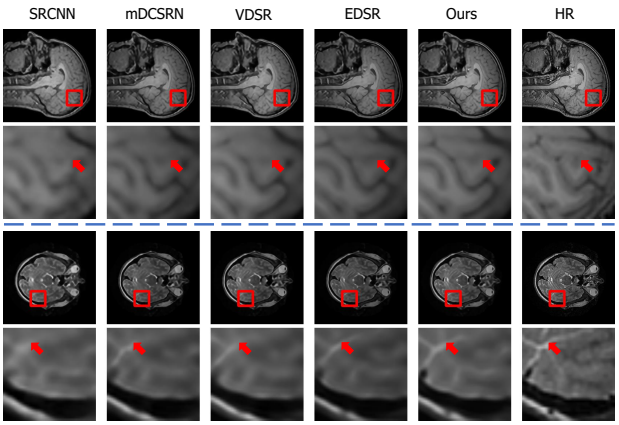


Figure 7: The visual comparison between different SR approaches on two datasets: T1 (top), T2 (bottom)

Quantitative results. We adopt PSNR and structural similarity index metric (SSIM) [37] to evaluate the quality for SR. Table 1 shows the quantitative results of the compared methods. As we can see, the performance of our proposed method obviously better than other methods on both datasets. Compared to the mDCSRN that used to MRI SR, our method can perform better by a wide margin, this demonstrates the effectiveness of the proposed model in MRI SR. Compared to EDSR, our method can steadily boost PSNR by 0.24 dB for both datasets. And most important, since the encoder of our model is based on EDSR, the performance is still limited by the capability of the EDSR. Therefore, our model can further improve with stronger backbone such as RCAN [43] and RDN [44]. Fig 6 shows the curve of PSNR score during training on validation set. Since SRCNN performs poorly, we exclude it from the figures. We can obviously observe that our model is steadily outperforms other methods. This excludes the situation of the improvement comes from randomness.

Qualitative Evaluation. Fig 7 shows the visual comparison between several SR methods. As we can see that our model produces clearer edges and perseveres better structure. For instance, in the T2 dataset, the area indicated by red arrow, other methods almost blurred except EDSR, but it only retains a little structure. This demonstrates the effectiveness of our model for recovering high frequency details. Fig 8 provides the visual comparison of error maps between several SR methods. In error maps, we can clearly observe that our model has lower error in detail and edge

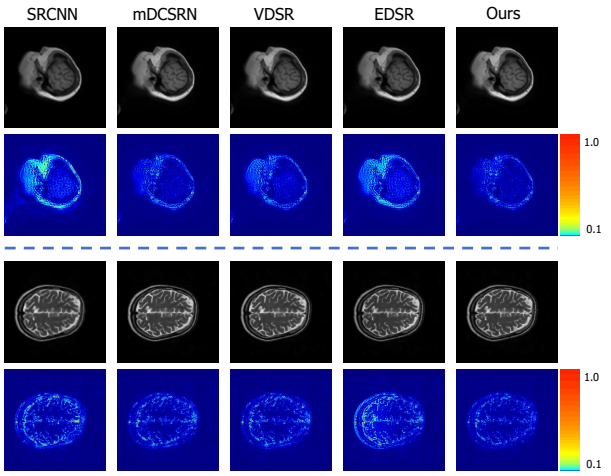


Figure 8: The visual comparison of error maps between different SR approaches on two datasets: T1 (top), T2 (bottom)

areas. Notably, the EDSR shows very high errors in these areas, this indicates the architecture of EDSR is poor at capturing the high frequency contents even if it shows high PSNR score.

4.5 Ablation Study

In this section, we conduct ablation study to verify the effectiveness of the components proposed in our method. Specifically, we conduct several experiments to evaluate Multi-scale Fourier Basis Module (MSFB), Frequency loss and High-Frequency Loss respectively, As shown in Table 2.

Multi-scale Fourier Basis Module. Compare config. A and B. As can be seen that without MSFB module, the performance degrades a lot, this demonstrates the importance of the Fourier basis, which can provide the complementary high-frequency information in network.

Frequency Loss. Compare config. C, D and E. Note that config D does not contain any components, thus it is actually EDSR. As can be seen that without Frequency Loss, the performance dropped significantly. And MSFB module cannot work well and performs to similar with baseline. This situation meets our expectation, Frequency loss can reflect the differences for each frequency component and directly forces network focuses on all frequencies. Therefore, the model would ignore the Fourier basis if without the guidance of Frequency loss.

High-Frequency Pixel Loss. Compare config. A and E. As can be seen that High-Frequency Pixel loss also contribute to the final performance. This demonstrates that the High-Frequency Pixel Loss can capture much high-frequency contents than vanilla L1 loss.

5 CONCLUSION

In this paper, we rethink the difference between MR images and natural images. From the characteristics of MRI, we mimic the process of an MRI machine produces an MRI in practice and propose an Implicit Neural Representation based Fourier Basis module. Unlike previous SR methods which rely only on LR images, the Fourier Basis module enables make use of information in Fourier basis to achieve more efficient reconstruction. Then to force network focuses on all frequency component, we propose a Frequency Loss to directly emphasizes on frequency domain. Final, we propose a High-Frequency Pixel Loss for spatial domain, which is more efficient than L1 loss. Experiments on two datasets have shown that our model is useful for MRI SR and outperforms the existing MRI SR methods.

Table 2: Ablation study on dataset T2, The best and second-best results are marked in red and blue, respectively.

config.	MSFB	Freq	HFP	L1	PSNR	SSIM
A	✓	✓	✓	✗	29.20	0.9180
B	✗	✓	✓	✗	29.10	0.9159
C	✓	✗	✗	✓	29.02	0.9165
D	✗	✗	✗	✓	28.96	0.9155
E	✓	✓	✗	✗	29.16	0.9173

REFERENCES

[1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5855–5864.

[2] Yinbo Chen, Sifei Liu, and Xiaolong Wang. 2021. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8628–8638.

[3] Yuhua Chen, Feng Shi, Anthony G Christodoulou, Yibin Xie, Zhengwei Zhou, and Debiao Li. 2018. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 91–99.

[4] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2015. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence* 38, 2 (2015), 295–307.

[5] Chun-Mei Feng, Huazhu Fu, Shuhao Yuan, and Yong Xu. 2021. Multi-contrast mri super-resolution via a multi-stage integration network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 140–149.

[6] Chun-Mei Feng, Yunlu Yan, Huazhu Fu, Li Chen, and Yong Xu. 2021. Task Transformer Network for Joint MRI Reconstruction and Super-Resolution. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[9] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* 33 (2020), 6840–6851.

[10] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

[11] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*. Springer, 694–711.

[12] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2021. Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems* 34 (2021).

[13] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and Improving the Image Quality of StyleGAN. In *Proc. CVPR*.

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

[15] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[16] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[17] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4681–4690.

[18] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 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.

[19] Zhisheng Lu, Hong Liu, Juncheng Li, and Linlin Zhang. 2021. Efficient transformer for single image super-resolution. *arXiv preprint arXiv:2108.11084* (2021).

[20] Razvan V Marinescu, Daniel Moyer, and Polina Golland. 2020. Bayesian image reconstruction using deep generative models. *arXiv preprint arXiv:2012.04567* (2020).

[21] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. 2020. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*. 2437–2445.

[22] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. 2018. Occupancy Networks: Learning 3D Reconstruction in Function Space. https://doi.org/10.48550/ARXIV.1812.03828

[23] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2020. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*. Springer, 405–421.

[24] Michael Niemeyer and Andreas Geiger. 2021. Giraffe: Representing scenes as compositional generative neural feature fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11453–11464.

[25] Michael Oechsle, Lars Mescheder, Michael Niemeyer, Thilo Strauss, and Andreas Geiger. 2019. Texture fields: Learning texture representations in function space. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4531–4540.

[26] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. 2019. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 165–174.

[27] Chi-Hieu Pham, Aurélien Ducournau, Ronan Fablet, and François Rousseau. 2017. Brain MRI super-resolution using deep 3D convolutional networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 197–200.

[28] Esben Plenge, Dirk Poot, Monique Bernsen, Gyula Kotek, Gavin Houston, Piotr Wielopolski, Louise van der Weerd, W.J. Niessen, and Erik Meijering. 2012. Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? *Magnetic Resonance in Medicine* 68 (12 2012). https://doi.org/10.1002/mrm.24187

[29] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. 2019. On the spectral bias of neural networks. In *International Conference on Machine Learning*. PMLR, 5301–5310.

[30] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. 2021. Image super-resolution via iterative refinement. *arXiv preprint arXiv:2104.07636* (2021).

[31] Edgar Schonfeld, Bernt Schiele, and Anna Khoreva. 2020. A u-net based discriminator for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8207–8216.

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

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

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

[35] Jiancong Wang, Yuhua Chen, Yifan Wu, Jianbo Shi, and James Gee. 2020. Enhanced generative adversarial network for 3D brain MRI super-resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3627–3636.

[36] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*. 0–0.

[37] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

[38] Qing Wu, Yuwei Li, Yawen Sun, Yan Zhou, Hongjiang Wei, Jingyi Yu, and Yuyao Zhang. 2021. An Arbitrary Scale Super-Resolution Approach for 3-Dimensional Magnetic Resonance Image using Implicit Neural Representation. *arXiv preprint arXiv:2110.14476* (2021).

[39] Rui Xu, Xintao Wang, Kai Chen, Bolei Zhou, and Chen Change Loy. 2021. Positional encoding as spatial inductive bias in gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13569–13578.

[40] Fuzhi Yang, Huan Yang, Jianlong Fu, Hongtao Lu, and Baining Guo. 2020. Learning texture transformer network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5791–5800.

[41] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. 2019. RankSRGAN: Generative Adversarial Networks with Ranker for Image Super-Resolution. *CoRR* abs/1908.06382 (2019). arXiv:1908.06382 http://arxiv.org/abs/1908.06382

[42] Yulun Zhang, Kai Li, Kunpeng Li, and Yun Fu. 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.

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

[44] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. 2018. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2472–2481.

[45] Zhifei Zhang, Zhaowen Wang, Zhe Lin, and Hairong Qi. 2019. Image super-resolution by neural texture transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7982–7991.