

Sodium and T2W MRI Fusion Representation of Arbitrary Resolution via Implicit Neural Representation

Keyu Yang^a

^a*School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China*

Abstract

Sodium magnetic resonance imaging (MRI) provides rich metabolic information for early and accurate tumor diagnosis. However, due to hardware limitations, scan time constraints and low sodium concentration, sodium MR images often suffer from low resolution and signal-to-noise ratio (SNR). Although deep learning-based image super-resolution (SR) algorithms can recover high-resolution (HR) images from low-resolution (LR) inputs, these methods typically require paired HR-LR datasets, which are challenging to obtain for sodium MRI. Here, we propose ArRFR, a self-supervised Arbitrary Resolution Fusion Representation approach based on implicit neural representation (INR) that reconstructs HR fused images of arbitrary resolution from LR sodium and T2-weighted images. The ArRFR model first extracts modality-specific feature maps using encoder networks, then fuses these features via a convolutional block attention module (CBAM) to adaptively adjust channel-wise and spatial-wise contributions. Given the absence of HR references in self-supervised training, we adopt perceptual losses to optimize the model. A multi-layer perceptron (MLP) acts as a local implicit function, reconstructing HR images from the fused feature map through coordinate-wise prediction. To optimize the model without high-quality references, perceptual losses are computed. Experimental results demonstrate that ArRFR achieves superior image quality across various up-sampling scales. The fused images retain the metabolic information from sodium MRI while inheriting the high sharpness of T2-weighted modality, effectively balancing noise suppression and detail preservation. Analysis of lesion cases further shows that the proposed method enhances tumor boundary clarity, particularly for gliomas, highlighting its potential to improve multi-modal MRI-based diagnostic accuracy.

Keywords: Sodium MRI, Image fusion, Attention, Perceptual loss, Tumor diagnosis,

1. Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technology that has been widely used in the detection and diagnosis of various diseases. Currently, proton MRI is most widely used, since hydrogen is the most abundant element in the human body.

Apart from protons, other nuclei are also detectable with MRI, such as sodium. As the second most abundant element in human body detectable on MRI, sodium is crucial for the body's homeostasis (Boada et al., 2005) and plays a significant role in neural signal propagation within and between neurons. Hence, sodium MRI can reflect metabolic states of human tissues (Ouwerkerk, 2011). It shows remarkable potential for diagnosis, characterization and treatment monitoring in neurologic diseases. Clinical applications of sodium MRI emerge in diagnosis of various diseases, such as tumors, cerebral ischemic stroke and epilepsy (Madelin et al., 2014; Madelin and Regatte, 2013; Hagiwara et al., 2021). However, sodium MRI is confronted with challenges such as short T2 component, low gyromagnetic ratio and low concentration, which result in low signal-to-noise ratio (SNR) (Madelin et al., 2014; Hagiwara et al., 2021). Non-conventional sampling techniques are a typical solution for the challenge above. For example, non-Cartesian Sampling (Shah et al., 2016; Zaric et al., 2021) shortens the echo time by directly measuring free induction decay

(FID), which alleviates the problem of short T2 component. Although enhancing hardware facilities like using higher magnetic fields or multiple coils can improve the SNR of sodium MRI, implementing these measures is often costly. Furthermore, there is still a trade-off between scanning time and resolution, similar to proton MRI. To sum up, sodium MRI often suffer from low resolution and SNR, which greatly hinders its widespread applications.

Image super-resolution can be involved in sodium MRI reconstruction to address the low-resolution problem. A previous work on sodium MRI SR by Rodriguez et al. (2023) approached the problem statistically without employing deep-learning methods. With the development of convolutional neural network (CNN), many CNN-based MRI super-resolution models have been proposed (Pham et al., 2017, 2019; Sanchez and Vilaplana, 2018; Chen et al., 2018; de Leeuw den Bouter et al., 2022). Recent works involving MRI SR have employed models with more complex architectures. Man et al. (2023) enhanced its model by adding modified residual channel attention blocks for extracting high-level features of 3D brain structures. Wu et al. (2023) incorporated the principle of implicit neural representation (INR) to build a model that can handle SR tasks of arbitrary resolution. Sun et al. (2024) achieved MRI resolution and quality enhancement by a tissue-aware enhancement network following a tissue-classification network. To make full

use of information across modalities, multi-contrast MRI SR works emerge as another pioneering field in MRI SR, as shown in (Lyu et al., 2020; Mao et al., 2023; Lucas et al., 2023; Feng et al., 2021), (Li et al., 2022). Besides image super-resolution, image denoising can serve as a solution to improve the SNR of sodium MR images. For example, CNN-based image denoising models such as Zhang et al. (2017) are powerful tools to reduce random noise in images. Moreover, most SR works mentioned above also considered denoising as a part in their model design. However, conventional SR or denoising models are not suitable for our task, which involves sodium MRI suffering from limited availability of HR references. These conventional models often require HR references during training, a condition that our sodium MRI reconstruction task cannot meet. Moreover, unless INR is involved, the SR algorithms above cannot achieve output of arbitrary resolution.

Inspired by the model Local Implicit Image Function (LIIF) (Chen et al., 2021), an INR approach, we propose ArRFR, an Arbitrary Resolution Fusion Representation approach that recovers HR fused images of arbitrary resolution from LR sodium and T2-weighted images. The model utilizes two encoders to extract two modality-specific feature maps from a pair of spatially aligned sodium and T2-weighted images. The two feature maps are then fused via a channel and spatial attention module. A feature unfolding operation follows the attention module to enrich information in the fused feature map. Finally, a local implicit function combined with a local ensemble operation reconstructs a fused image from the enriched feature map. By adopting a self-supervised strategy, our model eliminates the need for HR references during training and enables fused outputs of arbitrary resolution through the use of INR. Experiments demonstrate that the fused images inherit high-quality features such as high sharpness and low SNR from T2-weighted images while retaining the unique metabolic information of sodium MRI. Lesion case analyses further highlight the model’s potential to improve multi-modal MRI-based disease diagnosis, particularly by enhancing tumor boundary clarity in clinical scenarios.

2. Related Work

2.1. Implicit Neural Representation

Visual signals in the natural world are inherently continuous, representable as functions of variables across multiple dimensions. For example, 2D medical images (e.g., MRI, CT, PET) can be mathematically modeled as continuous functions over 2D spatial coordinates. However, digital storage necessitates discretizing these continuous signals into 2D pixel arrays, introducing a fundamental trade-off between representation complexity and precision. Discretization of images can lead to many problems. First of all, representing images in a discretized way, rather than a continuous way, can lead to a great loss of details. Moreover, discretized images are typically represented with a fixed resolution, whereas engineering tasks often require images with a series of resolution. INR was proposed as a solution to continuous representation of images. It was initially proposed to improve the visual quality

of shapes and surfaces (Chen and Zhang, 2019), (Park et al., 2019), (Mescheder et al., 2019). INR models the continuous representation problem as learning a continuous function that maps an image’s spatial coordinates to signal intensities. Various applications of INR have been discovered since then. Chen et al. (2021) proposed a local implicit image function to reconstruct HR images of arbitrary resolution from LR inputs. Skorokhodov et al. (2021); Dupont et al. (2022); Anokhin et al. (2021) applied INR to image generation and Shaham et al. (2021) used INR for image translation. In contrast to conventional deep-learning models that require architecture adjustments for different image resolutions, INR uses a generalized architecture that can adapt to all resolutions. Therefore, INR provides a much more efficient solution for image processing compared with conventional methods.

In recent years, there has been a surge of interest in INR within the medical imaging field, as its applications reduce the reliance on large-scale labeled datasets by minimizing the need for explicit labels (Molaei et al., 2023). Specifically, INR is widely applied to tasks such as reconstruction, segmentation, neural rendering, compression and registration of medical images. In terms of reconstruction, Shen et al. (2024) and Feng et al. (2024) utilized INR for reconstructing sparsely-sampled images. Lao et al. (2025) employed INR for simultaneous reconstruction of multiparametric quantitative MRI and Zhang et al. (2024) proposed a INR-based imaging approach for quantitative susceptibility mapping with high accuracy. Reed et al. (2021) and Feng et al. (2025) employed it for dynamic computed tomography (CT) reconstruction and MRI reconstruction, respectively. Besides these works, there are works that integrated INR into the framework of other models in medical image reconstruction. For example, Chu et al. (2025) creatively integrated INR into the posterior sampling of diffusion models, achieving high quality reconstruction for highly-accelerated MRI. Medical image super-resolution is another pioneering application of INR in medical image processing. Wu et al. (2023, 2021); Mao and Kamata (2022) proposed MRI SR algorithms that can up-sample LR images to HR images with high quality. These SR algorithms can significantly enhance the visualization of anatomical structures and diagnostic information of MR images.

2.2. Image Fusion

Restricted by theoretical and technical limitations of hardware, images captured by only one kind of method cannot provide comprehensive description of the imaging scene. For example, in MRI, T1-weighted images provide rich anatomical details, but have low signal values in lesion regions. On the contrary, T2-weighted images provide better visual effect for lesions, but offering less detailed information about anatomical structures. Aiming to enhance analytical and visual image quality, image fusion integrates complementary information from multi-modal images. An effective image fusion framework should preserve critical details while avoid inconsistencies from source images (Kaur et al., 2021). Recent years, there has been a great progress in image fusion methods with the promotion

of deep-learning models. Deep-learning models uses specialized networks for feature extraction, and learn a more reasonable feature fusion strategy with well-designed loss functions (Zhang et al., 2021).

In the field of medical imaging, deep-learning fusion has gained traction through a variety of frameworks (Wang et al., 2021; Ma et al., 2022; Tang et al., 2022; Fu et al., 2021; Xu et al., 2023; Xu and Ma, 2021). These image fusion methods exhibit diverse model architectures. Ma et al. (2022) and Tang et al. (2022) adopted Transformer-based (Vaswani et al., 2023) multi-modal image fusion. Xu et al. (2023) designed a pipeline with coarse registration, fusion and fine registration modules, jointly optimizing all components during training. As image fusion typically fuses structural features from different sources, works above mostly adopted a weighted sum of structural similarity loss between the fused output and input modalities, balancing information preservation via weights.

Notably, INR has emerged as a novel fusion tool in non-medical fields: Deng et al. (2023) and Liang et al. (2024) proposed INR-based frameworks for multispectral/hyperspectral image fusion. Although these studies did not focus on medical imaging, they inspired our exploration of INR in medical image fusion. Critically, since our model employs a self-supervised strategy, conventional losses like L1 or SSIM may propagate sodium MRI noise to the fused output. To guide the training in a more proper manner, we adopt perceptual losses as our loss function.

3. Method

3.1. Model Overview

Figure 1 illustrates the overall architecture of our proposed ArRFR. The model structure is primarily inspired by the INR framework in LIIF (Chen et al., 2021). During training, given a pair of spatially aligned LR sodium and T2W images $I = \{I_{\text{Na}} \in \mathbb{R}^{1 \times w \times h}, I_{\text{T2}} \in \mathbb{R}^{1 \times w \times h}\}$, we first apply a degradation model to I_{Na} to simulate a low-quality version $I_{\text{Na}'}$. We then down-sample both $I_{\text{Na}'}$ and I_{T2} by scaling factors $k = \{1, 2, 3, 4\}$, generating down-sampled input pairs $I^{1/k} = \{I_{\text{Na}'}^{1/k} \in \mathbb{R}^{1 \times \frac{w}{k} \times \frac{h}{k}}, I_{\text{T2}}^{1/k} \in \mathbb{R}^{1 \times \frac{w}{k} \times \frac{h}{k}}\}$. Two encoders extract modality-specific feature maps from down-sampled sodium and T2-weighted images. The generated pair of feature maps, $\{M_{\text{Na}'}^{1/k} \in \mathbb{R}^{128 \times \frac{w}{k} \times \frac{h}{k}}, M_{\text{T2}}^{1/k} \in \mathbb{R}^{128 \times \frac{w}{k} \times \frac{h}{k}}\}$, is then concatenated along the channel dimension to form a single, combined feature map $M^{1/k} \in \mathbb{R}^{256 \times \frac{w}{k} \times \frac{h}{k}}$. A convolutional block attention module (CBAM) (Woo et al., 2018) then adaptively adjusts the values in $M^{1/k}$ based on channel and spatial attention. Following work Chen et al. (2021), a feature unfolding operation enriches the local contextual information in $M^{1/k}$ by incorporating 3×3 neighboring latent codes, expanding the feature map dimension to $(9 \cdot 256) \times \frac{w}{k} \times \frac{h}{k}$. Finally, a multi-layer perceptron (MLP) serves as the local implicit function, taking the unfolded feature map, the original-resolution pixel grid $x \in \mathbb{R}^{2 \times w \times h}$ and a cell input $c \in \mathbb{R}^{2 \times w \times h}$ as inputs to reconstruct the output image $\hat{I}_k \in \mathbb{R}^{1 \times w \times h}$ with the same resolution as I_{Na} and I_{T2} . The subscript k in \hat{I}_k denotes that the output is

reconstructed from inputs down-sampled by the factor k . The reconstruction is refined via a local ensemble operation, which aggregates predictions from sub-regions to enhance spatial consistency. To compute perceptual losses, \hat{I}_k , I_{Na} and I_{T2} are passed through a pretrained VGG-16 network (Simonyan and Zisserman, 2015) to extract feature and style statistics. By minimizing the perceptual losses using gradient descent via back-propagation (Rumelhart et al., 1986), the two encoders, CBAM and MLP are jointly optimized without requiring HR images.

During inference, the degraded sodium image input is replaced with the original LR sodium image. For a target up-sampling scale r , the MLP receives a pixel grid and cell size corresponding to the desired HR dimensions $2 \times (w \cdot r) \times (h \cdot r)$, enabling arbitrary-resolution output generation.

3.2. Networks and Operations in ArRFR

3.2.1. Degradation Model

In pioneering works of MRI super-resolution, reconstruction and enhancement, degradation models are commonly used to synthesize low-quality images from high-quality references. For example, Sun et al. (2024) introduced Gaussian and Rician noise (Gudbjartsson and Patz, 1995) respectively to get corrupted images. Man et al. (2023) and de Leeuw den Bouter et al. (2022) adopted k-space sampling as the first step: the former added Rician Noise to the magnitude image, whereas the latter added complex Gaussian noise in k-space. Shi et al. (2015) applied Gaussian blurring followed by downsampling to generate low-quality inputs. In our work, the sodium MR images we acquire inherently have low SNR and resolution, and down-sampling is performed during training. Therefore, we simply use Rician noise (Gudbjartsson and Patz, 1995) as our degradation model. Specifically, after preprocessing sodium images with normalization to $[0, 1]$, we then add Rician noise with the standard deviation of $\sigma = 0.4$, followed by normalization again. This procedure can change the mapping feature of our model, which will be demonstrated in more detail in the experimental results.

3.2.2. CBAM

The CBAM in our model follows the standard architecture in Woo et al. (2018), consisting of a channel attention branch followed by a spatial attention branch, as shown in Fig. 1.

In the channel attention part, the channel weight M_C is computed as:

$$M_C = \sigma(f(\text{AvgPool}(M)) + f(\text{MaxPool}(M))), \quad (1)$$

where σ denotes the sigmoid function and f is a shared 2-layer MLP with ReLU activation. For an input feature map $M \in \mathbb{R}^{256 \times w \times h}$, max/average pooling along the spatial dimension reduce its dimension to $256 \times 1 \times 1$. The feature map is updated via element-wise multiplication with $M_C \in \mathbb{R}^{256 \times 1 \times 1}$.

In the spatial attention part, the spatial weight is computed as:

$$M_P = \sigma(\text{Conv}^{7 \times 7}([\text{AvgPool}(M); \text{MaxPool}(M)])), \quad (2)$$

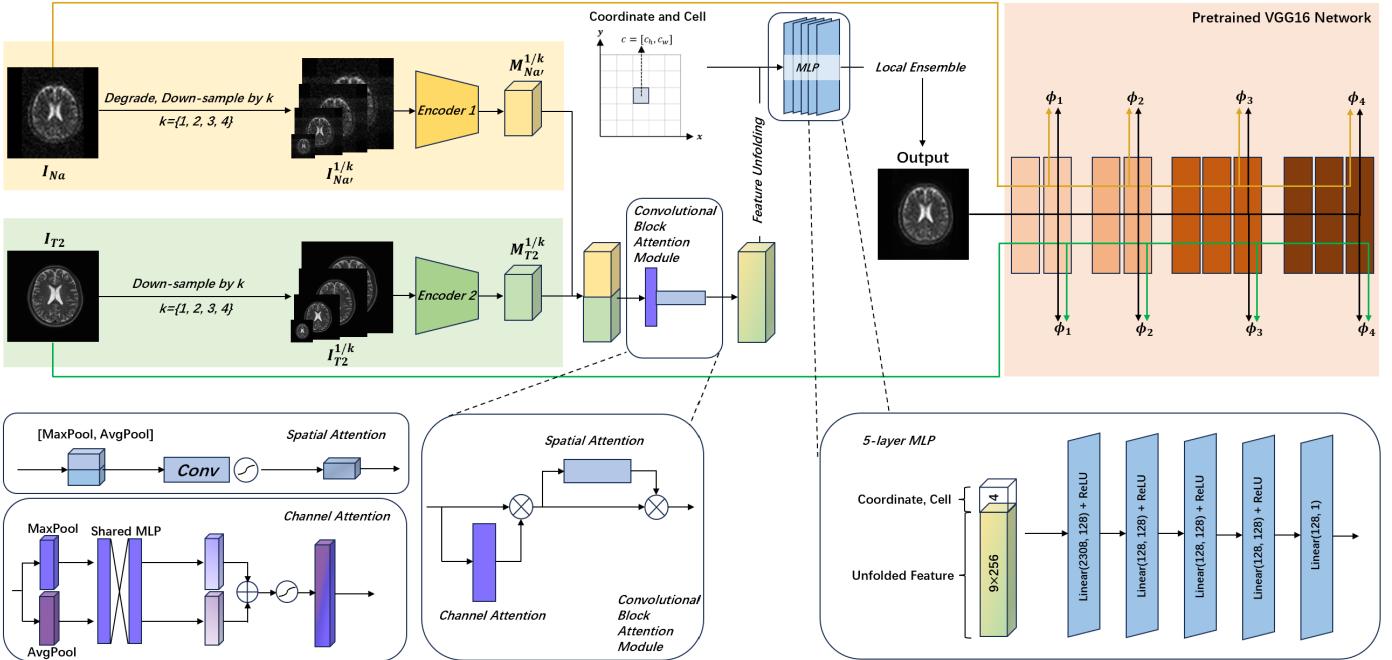


Figure 1: An overview of our proposed ArRFR model. \oplus denotes element-wise addition and \otimes denotes element-wise multiplication.

where σ denotes the sigmoid function and $Conv^{7 \times 7}$ denotes a 7×7 convolutional layer with two input channels and one output channel. Max/average pooling along the channel dimension reduce the shape of feature map from $256 \times w \times h$ to $1 \times w \times h$. The concatenated feature is processed by the convolutional layer to generate $M_P \in \mathbb{R}^{1 \times h \times w}$, which is multiplied with the feature map for updating.

3.2.3. Encoder Network

In our work, we used EDSR-baseline (Lim et al., 2017) (without up-sampling module) as the encoders for feature extraction of both sodium and T2W images. The encoders generate 128-channel feature maps with the same height and width as input images.

3.2.4. Feature Unfolding

We adopted the feature unfolding method from the LIIF framework (Chen et al., 2021) to enrich local contextual information. For a feature map $M \in \mathbb{R}^{256 \times w \times h}$, let $M[i, j]$ denote the latent code at spatial position (i, j) . Feature unfolding updates each latent code by concatenating its 3×3 neighboring codes:

$$M[i, j] = \text{Concat}(M[i + l, j + m]_{l, m \in \{-1, 0, 1\}}), \quad (3)$$

where Concat denotes channel-wise concatenation. This operation expands the feature map from $256 \times w \times h$ to $2304 \times w \times h$ (9 neighbors \times 256 channels), incorporating richer spatial context compared to the original single latent code.

3.2.5. MLP Network

For the MLP used in image reconstruction, we employ a 5-layer MLP with ReLU Activation and hidden dimension of 128, as shown in Fig. 1. The unfolded feature map, a coordinate and

a cell are the inputs to the MLP. During training, the coordinate grid $x \in \mathbb{R}^{2 \times w \times h}$ represents the (x, y) coordinates of pixel centers in I_{Na} and cell $c \in \mathbb{R}^{2 \times w \times h}$ represents the height and width of pixels. During inference, the coordinate grid and cell are scaled to $2 \times (w \cdot r) \times (h \cdot r)$, where r denotes the target up-sampling scale.

3.2.6. Local Ensemble

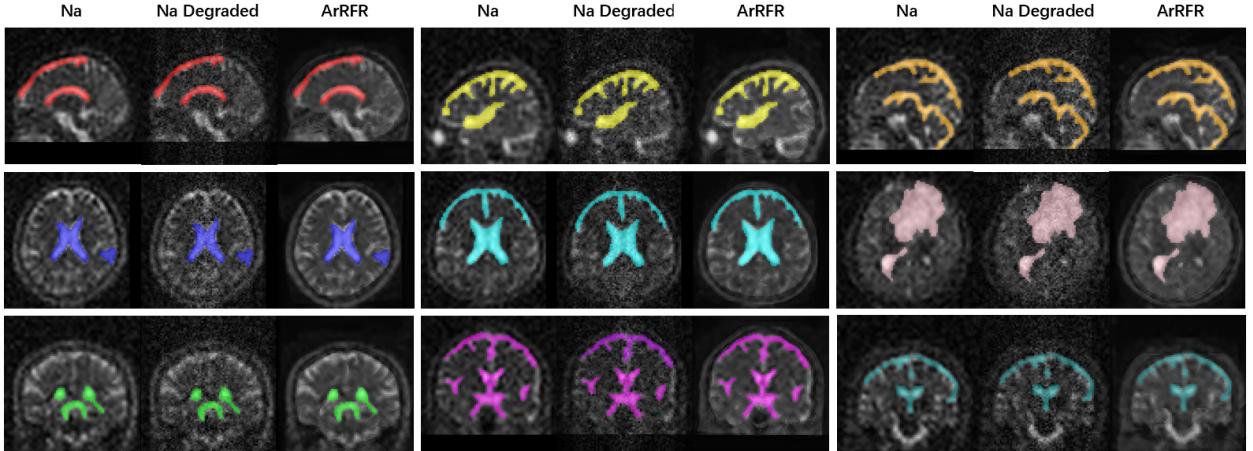
Following LIIF (Chen et al., 2021), we refine each pixel in MLP output via local ensemble, aggregating predictions of the four nearest latent codes from the pixel's coordinate. The final pixel intensity is a weighted average of these four predictions, where each weight is proportional to the area of the rectangle between the pixel point and its nearest latent code's diagonal counterpart.

3.3. Loss Function

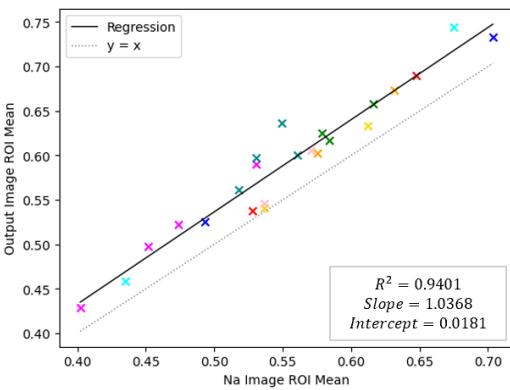
Considering the lack of paired references in our self-supervised task, we abandon conventional L1 or SSIM loss and adopt perceptual losses (Johnson et al., 2016). As shown in Fig. 1, a VGG-16 network (Simonyan and Zisserman, 2015) (denoted as ϕ) is pretrained on a shuffled dataset of sodium and T2W images. During the training of ArRFR, the VGG-16 is fixed. For a image, the VGG-16 extracts four hierarchical latent representations $\phi_i \in \mathbb{R}^{c_i \times w_i \times h_i}$ ($i \in \{1, 2, 3, 4\}$). For the input sodium image I_{Na} and output image \hat{I}_k , the feature loss is the normalized, squared Euclidean distance between latent representations:

$$L_{Na}^{\text{feat}} = \frac{1}{c_2 \times w_2 \times h_2} \|\phi(\hat{I}_k) - \phi(I_{Na})\|_2^2, \quad (4)$$

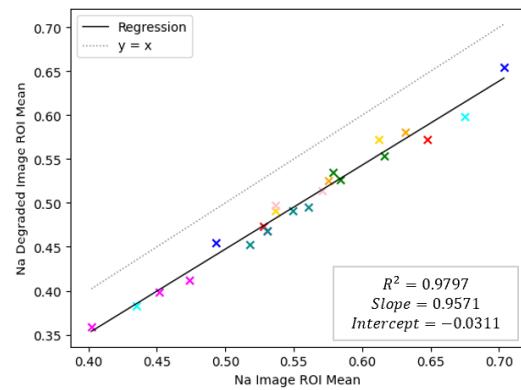
where c_2, w_2, h_2 correspond to the dimensions of ϕ_2 . For style loss, ϕ_i is reshaped to a matrix $\psi_i \in \mathbb{R}^{c_i \times (w_i \cdot h_i)}$. A Gram matrix is



(a) Segmentation results for ROIs with high pixel intensity (typically, CSF and tumors). Each slice is segmented with the same color.



(b) Relationship between means of input ROIs and degraded ROIs.



(c) Relationship between means of input ROIs and output ROIs.

Figure 2: Quantitative analysis of ArRFR model.

defined as:

$$G_i = \frac{\psi_i \psi_i^T}{c_i \times w_i \times h_i}, \quad (5)$$

where ψ_i^T denotes matrix transposition. The style loss is the squared Frobenius norm of the difference between the Gram matrices of \hat{I}_k and I_{Na} :

$$L_{\text{Na}}^{\text{style}} = \sum_{i=1}^4 \|G_i(\hat{I}_k) - G_i(I_{\text{Na}})\|_F^2. \quad (6)$$

Losses between \hat{I}_k and I_{T2} , i.e. $L_{\text{T2}}^{\text{feat}}$ and $L_{\text{T2}}^{\text{style}}$, are computed identically. For the output \hat{I}_k reconstructed from input $I^{1/k} = \{I_{\text{Na}}^{1/k}, I_{\text{T2}}^{1/k}\}$, the overall loss can be represented as a weighted sum of feature losses and style losses:

$$L_k = \lambda_{\text{Na}}^{\text{feat}} \cdot L_{\text{Na}}^{\text{feat}} + \lambda_{\text{Na}}^{\text{style}} \cdot L_{\text{Na}}^{\text{style}} + \lambda_{\text{T2}}^{\text{feat}} \cdot L_{\text{T2}}^{\text{feat}} + \lambda_{\text{T2}}^{\text{style}} \cdot L_{\text{T2}}^{\text{style}}, \quad (7)$$

where $\lambda_{\text{Na}}^{\text{feat}}, \lambda_{\text{Na}}^{\text{style}}, \lambda_{\text{T2}}^{\text{feat}}, \lambda_{\text{T2}}^{\text{style}}$ are hyper-parameters fixed during training. The final training loss aggregates across all down-sampling factors:

$$L = \frac{1}{4} \sum_{k=1}^4 L_k. \quad (8)$$

4. Experiments

4.1. Experimental Results

To comprehensively evaluate the proposed ArRFR approach, we conduct the following three experiments:

4.1.1. Quantitative Analysis

In order to analyze the effect of the degradation model and the mapping feature of the ArRFR model, we conducted a quantitative analysis regarding region of interest (ROI). In T2W and sodium MRI, cerebrospinal fluid (CSF) and tumor regions with high signal intensity were selected as ROIs in our task. We first randomly selected nine slices from the testing set. Outputs without up-sampling were generated by our trained ArRFR model. Then we segmented shared ROIs in two modalities. The same segmentation mask was applied to the original sodium image, the degraded sodium image and the output image. Figure 2a shows all segmentation results. A unique color is used for each slice. After segmentation, means of the pixels within the ROIs were calculated. To evaluate the quantitative relationship between the ROIs of three kinds of images, we performed two sets of linear regressions. Figure 2b shows that the ROI of degraded sodium image has lower mean value than the ROI

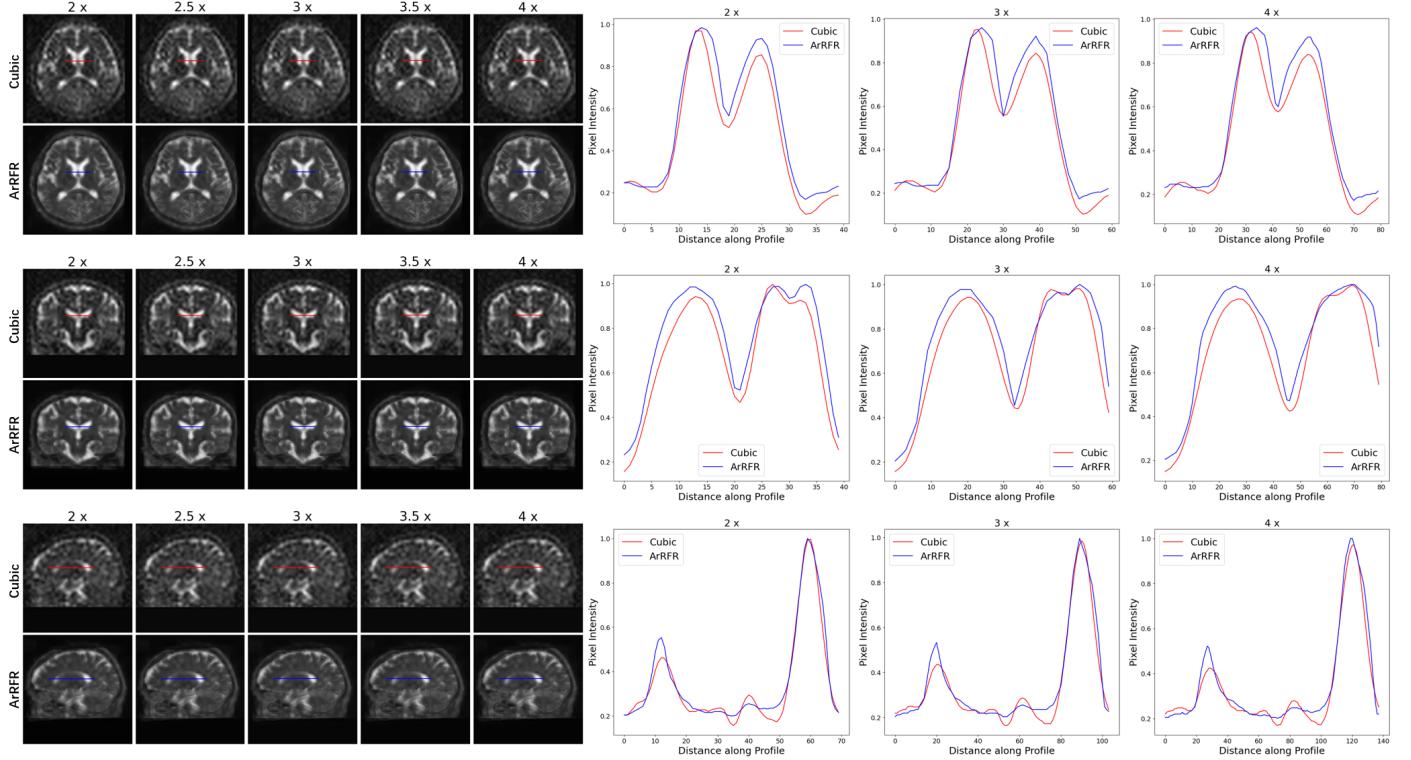


Figure 3: On the left: ArRFR outputs compared with direct cubic spline interpolation results with up-sampling scales $k = \{2, 2.5, 3, 3.5, 4\}$. On the right: corresponding profile plot with up-sampling scales $k = \{2, 3, 4\}$

of original sodium image. This result proves the effect of our degradation model. Figure 2c shows that the ROI of output image has higher mean value than the ROI of the sodium image. Results of two sets align with our model’s logic, since ArRFR model takes degraded sodium images as inputs when training and original sodium images when testing. When training, the model learns a mapping from degraded images to original images, which is a mapping from relatively low pixel intensities to high pixel intensities. When testing, the model therefore maps the high pixel intensities to even higher ones. The R^2 value showed in Figure 2c also indicates that the model performs considerably steady on different testing cases, which implies the robustness of the model.

4.1.2. Qualitative Analysis

Fig. 3 shows the outputs of our model with up-sampling scales of $k = \{1, 2, 2.5, 3, 3.5, 4\}$. To evaluate the model’s performance on high resolution outputs, we also applied direct cubic interpolation to the original image. Profile lines (colored for slice differentiation) crossing the area of CSF are overlaid on the results. Figure 3 also shows the pixel intensities along the profile line. As is shown, in high signal intensity regions such as the CSF, the model’s outputs exhibit sharper boundaries compared to direct interpolated baselines. Conversely, in low signal intensity regions, the model suppresses noise in the inputs, producing smoother intensity transitions.

Table 1: Values of Metrics Used for both Spline Interpolation and Our ArRFR Model with Up-sampling Scales $k = \{1, 2, 2.5, 3, 3.5, 4\}$

Scales	Models	PSI↑	LPC-SI↑	BRISQUE↓
2 x	Cubic	0.1825	0.9376	51.0612
	ArRFR	0.2123	0.9434	37.0899
2.5 x	Cubic	0.1460	0.9224	52.1949
	ArRFR	0.2266	0.9323	41.9754
3 x	Cubic	0.1275	0.9059	52.2954
	ArRFR	0.2318	0.9280	40.6400
3.5 x	Cubic	0.1148	0.8879	50.6304
	ArRFR	0.1941	0.9158	43.2371
4 x	Cubic	0.0985	0.8685	50.7053
	ArRFR	0.1718	0.9050	47.3980

To further evaluate the model’s qualitative performance, we adopted three non-referenced image quality metrics. The Perceptual Sharpness Index (PSI) (Feichtenhofer et al., 2013) and Local Phase Coherence-based Sharpness Index LPC-SI (Hassen et al., 2013) were calculated to compare the sharpness of model outputs and interpolated sodium images. Additionally, the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) (Mittal et al., 2012) was used to evaluate the quality in terms of “naturalness”. We selected 10 testing cases and evaluated the model’s performance across all up-sampling scales, as shown in 1. Results show that the model achieved higher sharpness and better quality across all up-sampling

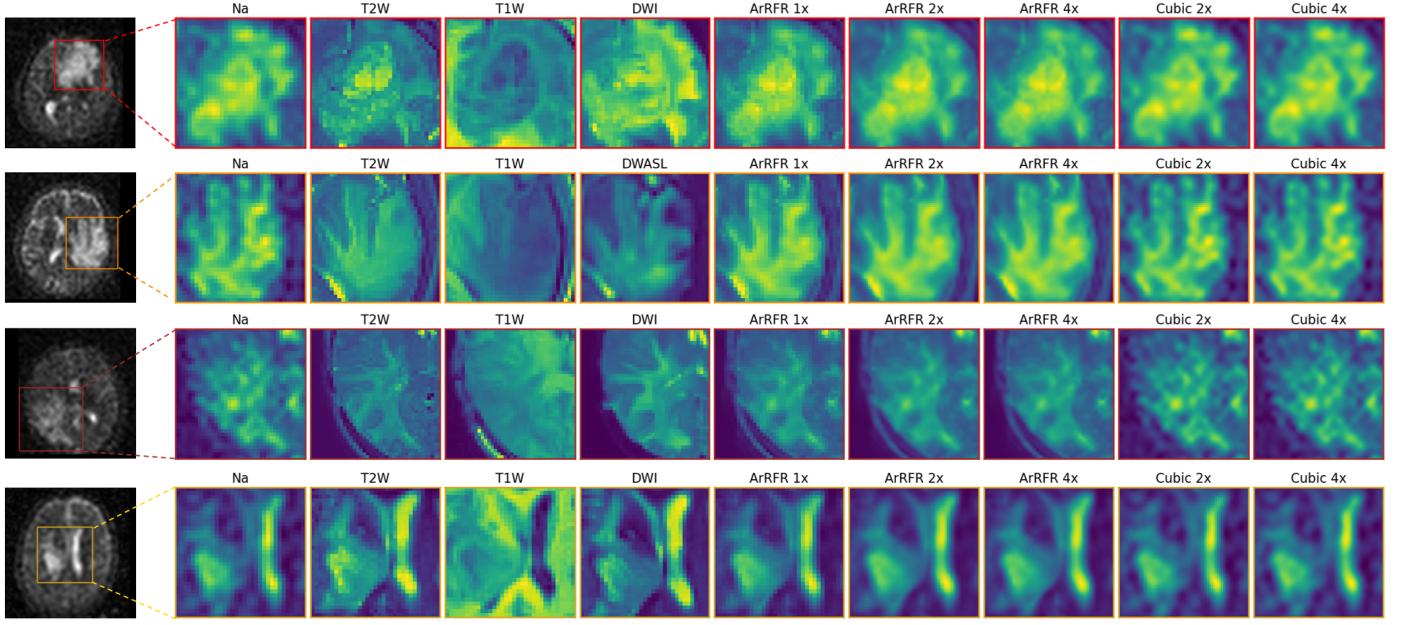


Figure 4: ArRFR results on glioma cases with up-sampling scales $k = \{1, 2, 4\}$. Direct interpolation results are also shown for comparison. T2W, T1, DWI (DWASL in the second case) images are also shown as a comparison between modalities.

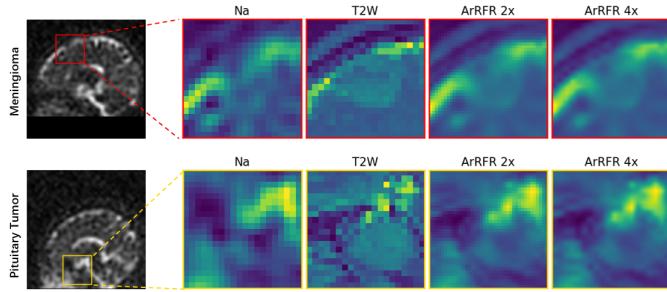


Figure 5: ArRFR results on a meningioma case and a pituitary tumor case.

scales compared to direct interpolation.

4.1.3. Analysis on Lesion Cases

To further discover our approach’s clinical utility, we tested our model on cases with different types of tumors. Figure 4 shows the model’s performance on glioma cases. The model’s outputs provides better details regarding the boudaries of the tumors. Cross-modal comparison shows that glioma exhibits high signal values in sodium images. In addition to glioma, we also tested our model on two other types of tumors, the meningioma and the pituitary tumor, as shown in Figure 5. These two types of tumors are not clearly visible on the sodium modality. Therefore, the model can not enhance the region of these two types of tumors. Results above demonstrates our model’s potential to improve glioma diagnosis via sodium MRI boundary refinement, while highlighting its limitations in tumors with inherently low sodium contrast.

4.2. Experimental Settings

4.2.1. Data Preprocessing

The data initially acquired include raw k-space data of sodium MRI and DICOM data of T2W MRI. For sodium MRI, we first employed the non-uniform discrete Fourier transform (NUFFT) to reconstruct sodium images with an isotropic resolution of $4.8572 \text{ mm} \times 4.8572 \text{ mm} \times 4.75 \text{ mm}$ from raw k-space data. In contrast, the T2W DICOM data have a higher resolution of $1 \text{ mm} \times 1 \text{ mm} \times 2 \text{ mm}$. Given the significant resolution gap between sodium and T2W MRI, direct registration was infeasible, since down-sampling T2W data by a factor of 4 would incur substantial information loss. Instead, we upsampled sodium data to $2.4286 \text{ mm} \times 2.4286 \text{ mm} \times 2.375 \text{ mm}$ using cubic spline interpolation, reducing the resolution mismatch. 3D linear registration was then applied to align T2W data to the upsampled sodium data. After registration, both modalities shared the common shape of $76 \times 112 \times 112$. To facilitate model training, we applied zero-padding to standardize the shape to $112 \times 112 \times 112$. Subsequently, we sliced 3D volumes into 2D images along the coronal, sagittal and axial planes, forming the dataset for both training and testing. To ensure modal consistency between 2-channel complex sodium data and 1-channel T2W magnitude data, we retained only the magnitude component of sodium images during training. Finally, min-max normalization was applied to both training and testing sets normalize pixel intensities to $[0, 1]$, optimizing model convergence.

4.2.2. Training Details

After preprocessing of all data, we excluded all cases with tumors for testing set. The remaining healthy cases were randomly split into a training set and a testing set at a ratio of 3:1. Training was conducted with a batch size of 16 on a single

NVIDIA GeForce RTX 3090 (24GB VRAM). An Adam optimizer Kingma and Ba (2017) was used to minimize the loss function in Eq. 8. The learning rate started from 1×10^{-4} , decaying by 0.1 every 10 epochs. The total number of training epochs was 50, taking about 10.5 hours. Checkpoints were saved every 10 epochs, with the best-performing model retained for testing.

4.3. Conclusion

In this study, we propose ArRFR, a self-supervised framework for reconstructing arbitrary-resolution fused images from LR sodium and T2-weighted MRI images. By leveraging INR, ArRFR addresses the challenges of sodium MRI—low resolution, low SNR, and scarce high-quality references—through a novel fusion strategy that integrates modality-specific features via attention and coordinated-wise prediction. This suggests potential to improve multi-modal diagnostic accuracy, especially for diseases where sodium MRI signal is indicative of tissue metabolic changes.

References

- Anokhin, I., Demochkin, K., Khakhulin, T., Sterkin, G., Lempitsky, V., Korzhenkov, D., 2021. Image generators with conditionally-independent pixel synthesis, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14273–14282.
- Boada, F., LaVerde, G., Jungreis, C., Nemoto, E., Tanase, C., Hancu, I., 2005. Loss of cell ion homeostasis and cell viability in the brain: what sodium MRI can tell us. *Current Topics in Development Biology* 70, 77–101.
- de Leeuw den Bouter, M.L., Ippolito, G., O'Reilly, T.P.A., Remis, R.F., van Gijzen, M.B., Webb, A.G., 2022. Deep learning-based single image super-resolution for low-field MR brain images. *Scientific Reports* 12, 6362.
- Chen, Y., Liu, S., Wang, X., 2021. Learning continuous image representation with local implicit image function, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8628–8638.
- Chen, Y., Xie, Y., Zhou, Z., Shi, F., Christodoulou, A.G., Li, D., 2018. Brain MRI super resolution using 3D deep densely connected neural networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging, p. 739–742.
- Chen, Z., Zhang, H., 2019. Learning implicit fields for generative shape modeling, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5932–5941.
- Chu, J., Du, C., Lin, X., Zhang, X., Wang, L., Zhang, Y., Wei, H., 2025. Highly accelerated MRI via implicit neural representation guided posterior sampling of diffusion models. *Medical Image Analysis* 100, 103398.
- Deng, S., Wu, R., Deng, L.J., Ran, R., Vivone, G., 2023. Implicit neural feature fusion function for multispectral and hyperspectral image fusion. [arXiv:2307.07288](https://arxiv.org/abs/2307.07288).
- Dupont, E., Teh, Y.W., Doucet, A., 2022. Generative models as distributions of functions. [arXiv:2102.04776](https://arxiv.org/abs/2102.04776).
- Feichtenhofer, C., Fassold, H., Schallauer, P., 2013. A perceptual image sharpness metric based on local edge gradient analysis. *IEEE Signal Processing Letters* 20, 379–382.
- Feng, C.M., Fu, H., Yuan, S., Xu, Y., 2021. Multi-contrast MRI super-resolution via a multi-stage integration network, in: Medical Image Computing and Computer Assisted Intervention 2021, pp. 140–149.
- Feng, J., Feng, R., Wu, Q., Shen, X., Chen, L., Li, X., Feng, L., Chen, J., Zhang, Z., Liu, C., Zhang, Y., Wei, H., 2025. Spatiotemporal implicit neural representation for unsupervised dynamic MRI reconstruction. *IEEE Transactions on Medical Imaging* , 1–1.
- Feng, R., Wu, Q., Feng, J., She, H., Liu, C., Zhang, Y., Wei, H., 2024. IMJENSE: Scan-specific implicit representation for joint coil sensitivity and image estimation in parallel MRI. *IEEE Transactions on Medical Imaging* 43, 1539–1553.
- Fu, J., Li, W., Du, J., Huang, Y., 2021. A multiscale residual pyramid attention network for medical image fusion. *Biomedical Signal Processing and Control* 66, 102488.
- Gudbjartsson, H., Patz, S., 1995. The rician distribution of noisy mri data. *Magnetic Resonance in Medicine* 34, 910–914.
- Hagiwara, A., Bydder, M., Oughourlian, T.C., Yao, J., Salamon, N., Jahan, R., Villablanca, J.P., Enzmann, D.R., Ellingson, B.M., 2021. Sodium MR neuroimaging. *American Journal of Neuroradiology* 42, 1920 – 1926.
- Hassen, R., Wang, Z., Salama, M.M.A., 2013. Image sharpness assessment based on local phase coherence. *IEEE Transactions on Image Processing* 22, 2798–2810.
- Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution, in: Computer Vision – ECCV 2016, pp. 694–711.
- Kaur, H., Koundal, D., Kadyan, V., 2021. Image fusion techniques: A survey. *Archives of Computational Methods in Engineering* 28, 4425–4447.
- Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- Lao, G., Feng, R., Qi, H., Lv, Z., Liu, Q., Liu, C., Zhang, Y., Wei, H., 2025. Coordinate-based neural representation enabling zero-shot learning for fast 3d multiparametric quantitative MRI. *Medical Image Analysis* 102, 103530.
- Li, G., Lv, J., Tian, Y., Dou, Q., Wang, C., Xu, C., Qin, J., 2022. Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast MRI super-resolution. [arXiv:2203.13963](https://arxiv.org/abs/2203.13963).

- Liang, Y.J., Cao, Z., Deng, L.J., Wu, X., 2024. Fourier-enhanced implicit neural fusion network for multispectral and hyperspectral image fusion. [arXiv:2404.15174](https://arxiv.org/abs/2404.15174).
- Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M., 2017. Enhanced Deep Residual Networks for Single Image Super-Resolution, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1132–1140.
- Lucas, A., Campbell Arnold, T., Okar, S.V., Vadali, C., Kawatra, K.D., Ren, Z., Cao, Q., Shinohara, R.T., Schindler, M.K., Davis, K.A., Litt, B., Reich, D.S., Stein, J.M., 2023. Multi-contrast high-field quality image synthesis for portable low-field MRI using generative adversarial networks and paired data. medRxiv .
- Lyu, Q., Shan, H., Steber, C., Helis, C., Whitlow, C., Chan, M., Wang, G., 2020. Multi-contrast super-resolution MRI through a progressive network. *IEEE Transactions on Medical Imaging* 39, 2738–2749.
- Ma, J., Tang, L., Fan, F., Huang, J., Mei, X., Ma, Y., 2022. SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer. *IEEE/CIA Journal of Automatica Sinica* 9, 1200–1217.
- Madelin, G., Lee, J.S., Regatte, R.R., Jerschow, A., 2014. Sodium MRI: methods and applications. *Progress in Nuclear Magnetic Resonance Spectroscopy* 79, 14 – 47.
- Madelin, G., Regatte, R.R., 2013. Biomedical applications of sodium MRI *in vivo*. *Journal of Magnetic Resonance Imaging* 38, 511 – 529.
- Man, C., Lau, V., Su, S., Zhao, Y., Xiao, L., Ding, Y., Leung, G.K., Leong, A.T., Wu, E.X., 2023. Deep learning enabled fast 3D brain MRI at 0.055 tesla. *Science Advances* 9, eadi9327.
- Mao, S., Kamata, S., 2022. Mri super-resolution using implicit neural representation with frequency domain enhancement, in: Proceedings of the 7th International Conference on Biomedical Signal and Image Processing, p. 30–36.
- Mao, Y., Jiang, L., Chen, X., Li, C., 2023. DisC-Diff: disentangled conditional diffusion model for multi-contrast MRI super-resolution, in: Medical Image Computing and Computer Assisted Intervention 2023, pp. 387–397.
- Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A., 2019. Occupancy networks: learning 3D reconstruction in function space, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4455–4465.
- Mittal, A., Moorthy, A.K., Bovik, A.C., 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing* 21, 4695–4708.
- Molaei, A., Aminimehr, A., Tavakoli, A., Kazerouni, A., Azad, B., Azad, R., Merhof, D., 2023. Implicit neural representation in medical imaging: a comparative survey. [arXiv:2307.16142](https://arxiv.org/abs/2307.16142).
- Ouwerkerk, R., 2011. Sodium MRI. *Methods in Molecular Biology* 711, 175 – 201.
- Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S., 2019. DeepSDF: learning continuous signed distance functions for shape representation, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 165–174.
- Pham, C.H., Ducournau, A., Fablet, R., Rousseau, F., 2017. Brain MRI super-resolution using deep 3d convolutional networks, in: IEEE 14th International Symposium on Biomedical Imaging, pp. 197–200.
- Pham, C.H., Tor-Díez, C., Meunier, H., Bednarek, N., Fablet, R., Passat, N., Rousseau, F., 2019. Multiscale brain MRI super-resolution using deep 3D convolutional networks. *Computerized Medical Imaging and Graphics* 77, 101647.
- Reed, A.W., Kim, H., Anirudh, R., Mohan, K.A., Champley, K., Kang, J., Jayasuriya, S., 2021. Dynamic CT reconstruction from limited views with implicit neural representations and parametric motion fields, in: 2021 IEEE/CVF International Conference on Computer Vision, pp. 2238–2248.
- Rodriguez, G.G., Yu, Z., Shaykevich, S., O'Donnell, L.F., Aguilera, L., Cloos, M.A., Madelin, G., 2023. Super-resolution of sodium images from simultaneous ¹H MRF/²³Na MRI acquisition. *NMR in Biomedicine* 36, e4959.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533–536.
- Sanchez, I., Vilaplana, V., 2018. Brain MRI super-resolution using 3D generative adversarial networks. [arXiv:1812.11440](https://arxiv.org/abs/1812.11440).
- Shah, N.J., Worthoff, W.A., Langen, K.J., 2016. Imaging of sodium in the brain: a brief review. *NMR in Biomedicine* 29, 162 – 174.
- Shaham, T.R., Gharbi, M., Zhang, R., Shechtman, E., Michaeli, T., 2021. Spatially-adaptive pixelwise networks for fast image translation, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14877–14886.
- Shen, L., Pauly, J., Xing, L., 2024. NeRP: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems* 35, 770–782.
- Shi, F., Cheng, J., Wang, L., Yap, P.T., Shen, D., 2015. LRTV: MR image super-resolution with low-rank and total variation regularizations. *IEEE Transactions on Medical Imaging* 34, 2459–2466.

- Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- Skorokhodov, I., Ignatyev, S., Elhoseiny, M., 2021. Adversarial generation of continuous images, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10748–10759.
- Sun, Y., Wang, L., Li, G., Lin, W., Wang, L., 2024. A foundation model for enhancing magnetic resonance images and downstream segmentation, registration and diagnostic tasks. *Nature Biomedical Engineering* .
- Tang, W., He, F., Liu, Y., Duan, Y., 2022. MATR: Multimodal medical image fusion via multiscale adaptive transformer. *IEEE Transactions on Image Processing* 31, 5134–5149.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2023. Attention Is All You Need. [arXiv:1706.03762](https://arxiv.org/abs/1706.03762).
- Wang, Z., Li, X., Duan, H., Su, Y., Zhang, X., Guan, X., 2021. Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform. *Expert Systems with Applications* 171, 114574.
- Woo, S., Park, J., Lee, J.Y., Kweon, I.S., 2018. CBAM: Convolutional block attention module, in: Computer Vision – ECCV 2018, pp. 3–19.
- Wu, Q., Li, Y., Sun, Y., Zhou, Y., Wei, H., Yu, J., Zhang, Y., 2023. An arbitrary scale super-resolution approach for 3D MR images via implicit neural representation. *IEEE Journal of Biomedical and Health Informatics* 27, 1004–1015.
- Wu, Q., Li, Y., Xu, L., Feng, R., Wei, H., Yang, Q., Yu, B., Liu, X., Yu, J., Zhang, Y., 2021. IREM: High-resolution magnetic resonance image reconstruction via implicit neural representation, in: Medical Image Computing and Computer Assisted Intervention 2021, pp. 65–74.
- Xu, H., Ma, J., 2021. EMFusion: An unsupervised enhanced medical image fusion network. *Information Fusion* 76, 177–186.
- Xu, H., Yuan, J., Ma, J., 2023. MURF: Mutually reinforcing multi-modal image registration and fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 12148–12166.
- Zaric, O., Juras, V., Szomolanyi, P., Schreiner, M., Raudner, M., Giraudo, C., Trattnig, S., 2021. Frontiers of sodium MRI revisited: from cartilage to brain imaging. *Journal of Magnetic Resonance Imaging* 54, 58 – 75.
- Zhang, H., Xu, H., Tian, X., Jiang, J., Ma, J., 2021. Image fusion meets deep learning: A survey and perspective. *Information Fusion* 76, 323–336.
- Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L., 2017. Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing* 26, 3142–3155.
- Zhang, M., Feng, R., Li, Z., Feng, J., Wu, Q., Zhang, Z., Ma, C., Wu, J., Yan, F., Liu, C., Zhang, Y., Wei, H., 2024. A subject-specific unsupervised deep learning method for quantitative susceptibility mapping using implicit neural representation. *Medical Image Analysis* 95, 103173.