NExpR: Neural Explicit Representation for fast arbitrary-scale medical image super-resolution

Kaifeng Pang [a,b], Kai Zhao [b,*], Alex Ling Yu Hung [b,c], Haoxin Zheng [b,c], Ran Yan [b,d], Kyunghyun Sung [b]

[a] *Department of Electrical and Computer Engineering, University of California, Los Angeles, CA, 90095, United States*
[b] *Department of Radiological Sciences, University of California, Los Angeles, CA, 90095, United States*
[c] *Department of Computer Science, University of California, Los Angeles, CA, 90095, United States*
[d] *Department of Bioengineering, University of California, Los Angeles, CA, 90095, United States*

A R T I C L E   I N F O

A B S T R A C T

Medical images often require rescaling to various spatial resolutions to ensure interpretations at different levels. Conventional deep learning-based image super-resolution (SR) enhances the fixed-scale resolution. Implicit neural representation (INR) is a promising way of achieving arbitrary-scale image SR. However, existing INR-based methods require the repeated execution of the neural network (NN), which is slow and inefficient. In this paper, we present Neural Explicit Representation (NExpR) for fast arbitrary-scale medical image SR. Our algorithm represents an image with an explicit analytical function, whose input is the low-resolution image and output is the parameterization of the analytical function. After obtaining the analytical representation through a single NN inference, SR images of arbitrary scales can be derived by evaluating the explicit functions at desired coordinates. Because of the analytical explicit representation, NExpR is significantly faster than INR-based methods. In addition to speed, our method achieves on-par or better image quality than other strong competitors. Extensive experiments on Magnetic Resonance Imaging (MRI) datasets, including ProstateX, fastMRI, and our in-house clinical prostate dataset, as well as the Computerized Tomography (CT) dataset, specifically the Medical Segmentation Decathlon (MSD) liver dataset, demonstrate the superiority of our method. Our method reduces the rescaling time from the order of 1 ms to the order of 0.01 ms, achieving an over $100\times$ speedup without losing the image quality. Code is available at https://github.com/Calvin-Pang/NExpR.

1. Introduction

Medical images are essential in clinical diagnosis due to their ability to provide structural and functional information [1,2]. High-resolution medical images are desired as they can help improve the identification of clinically relevant information with enhanced delineation of anatomical structures and image analysis. However, acquiring high-resolution medical images is challenging due to the cost of data acquisition [3]. Besides, medical images are desired to be zoomed in and out to different scales to ensure interpretations at different levels [4,5]. Fast enhancing image resolution to arbitrary scales is a critical demand in medical imaging [6].

Medical image super-resolution (SR) is a promising way to improve the resolution. With the rapid development of computer vision and deep learning, numerous SR models have been proposed to obtain super-resolution images in both natural images [7,8] and medical images [9,10]. Deep learning-based image SR can be roughly divided into 3 categories: (i) feed-forward networks (FFNs), (ii) generative models, and (iii) implicit neural representation models. FFN-based methods [7] directly learn the mapping from low-resolution images to high-resolution images. However, they cannot reproduce fine details, especially with larger upsampling scales. Generative models, *e.g.* Generative Adversarial Networks (GANs) [11] and diffusion models [10], can generate highly realistic images, but they inadvertently generate synthetic details that do not represent any underlying anatomical structures. The inconsistency between these generated details and their lower-resolution inputs can be detrimental to clinical decision-making [12]. Moreover, both FFN-based and generative model-based methods can only perform SR at a prescribed fixed scale. While in
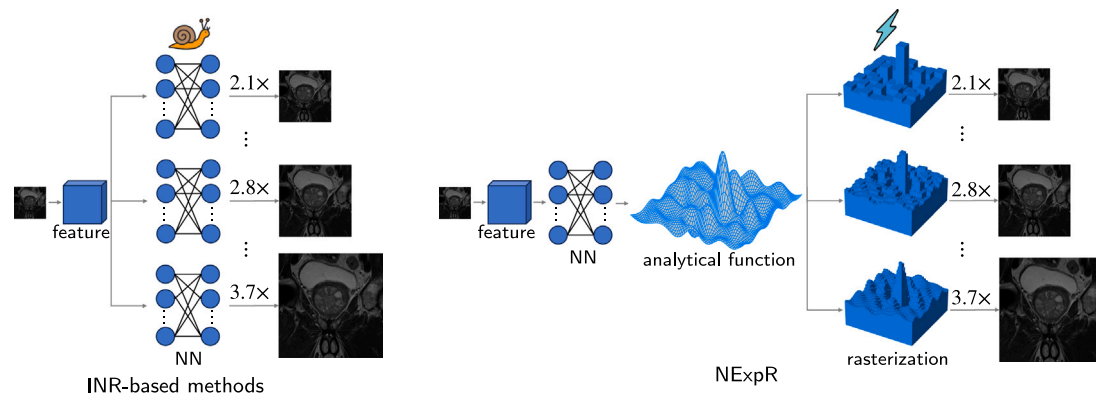

Fig. 1. Comparison of upscaling process between INR-based methods and NExpR. INR-based methods need to evaluate the implicit neural function to derive SR images, which is time-consuming. In contrast, our method evaluates (rasterizes) the explicit analytical function predicted by a neural network (NN) and is much faster than INR-based methods.

clinical applications, images are often zoomed in (or out) to visualize the tissue at different levels, and a flexible super-resolution is desired for different clinical purposes.

Implicit Neural Representations (INRs) are implicitly defined, continuous, differentiable functions parameterized by neural networks. Recent studies have demonstrated the potential of INRs for arbitrary-scale[1] image super-resolution [13,14]. INRs represent the image with an implicit neural function, and image super-resolution of arbitrary scales can be achieved by evaluating the neural implicit function at desired coordinates. However, generating SR images requires numerous repeated passes through the neural network, which is computationally inefficient [15]. As described in Fig. 1, in INR-based methods, each distinct SR scale necessitates a separate inference of the network, markedly slowing down the process in scenarios requiring continuous upscaling. Besides, INRs model the continuous signal directly in the spatial domain without maintaining a faithful representation of the original low-resolution image's fined details [16,17]. The SR results may deviate from the original LR input, leading to potential inconsistencies.

To the best of our knowledge, there has not been a study to represent medical images with a continuous and analytical function derived from a parameterized neural network. Here, we propose **N**eural **Exp**licit **R**epresentation (NExpR) for fast medical image SR. NExpR employs a Fourier transform-based *explicit* function for the continuous representation of medical images. Given a low-resolution medical image, NExpR predicts coefficients of a Fourier transform-based function that has a finite set of learnable frequencies and represents the image with this explicit analytical function. With the predicted coefficients, SR images at arbitrary scales can be achieved by evaluating the analytical function at desired locations. The explicit analytical function is more portable and compact than the implicit neural function, and can be fast evaluated on any device without specialized hardware or software. In contrast, the repeated execution of NN in INR-based methods is slower and requires deep learning infrastructure.

In summary, NExpR enjoys the following favorable properties for medical image super-resolution:

- NExpR is significantly faster than INR-based methods for arbitrary-scale SR. Our method is more lightweight, and the derived analytical function can rapidly generate SR images of arbitrary scale.
- NExpR achieves superior SR image quality compared to INR-based methods for arbitrary-scale SR and performs on par with or surpasses a wide range of methods for fixed-scale SR.

---

[1] We use the term "arbitrary" to refer to continuous scales within a specific range.

2. Related work

Image super-resolution. In literature, a variety of classical SR methods have been proposed to preserve specific characteristics of high-resolution images. These methods include edge-based methods [18,19], statistical methods [20,21], and sparse coding-based methods [22,23]. In the past decade, deep learning-based methods have dominated super-resolution and have achieved impressive performance in generating high-fidelity images. Deep learning-based SR methods can be broadly categorized into three classes: (i) feed-forward networks (FFNs), (ii) generative models, and (iii) Implicit Neural Representations (INRs).

FFN-based methods use neural networks, *e.g.* CNN [7,24], and Transformer [8,25] to directly estimate high-resolution images based on low-resolution inputs. Lim et al. used Convolutional Neural Networks (CNNs) [7], and Liang et al. [8] used the more powerful Transformers as the backbone networks. Han et al. proposed a multi-path recurrent networks [24] to recurrently refine the image textures. Li et al. built a pure transformer-style CNN, SRConvNet [25], for light-weight super-resolution. Various loss functions [26,27] have also been applied to or proposed for image super-resolution. Johnson et al. [26] proposed the perceptual loss to enhance the perceptual realism of super-resolution images. Sajjiadi et al. [27] extended the perceptual loss to texture synthesis. In the field of medical imaging, Li et al. [9] proposed a multi-level feature extraction and reconstruction model to recover the degraded high-resolution details of 3D MRI images. Fok et al. [28] incorporated the convolutional neural network (CNN) with multi-modality data training for computed tomography (CT) image super-resolution.

Generative model-based super-resolution methods employ state-of-the-art generative models, *e.g.* Generative Adversarial Networks (GANs) [29,30], Variational Autoencoders (VAEs) [31], and diffusion models [10,32,33] to generate high-resolution images conditioned on low-resolution inputs. Ledig et al. [29] proposed SRGAN, a GAN-based model using a deep residual network to recover photo-realistic textures for high-quality SR results. Saharia et al. [32] designed a diffusion-based SR model that progressively refines images from a low-resolution input to a high-resolution output. GANs [11] and diffusion models [10] have also been explored in medical image super-resolution. Wang et al. [34] introduced GAN to 3D brain MRI super-resolution. Zhao et al. [35] designed a multi-scale GAN-based model to achieve retinal fundus image super-resolution with a higher perceptual quality. Chung et al. [36] proposed a score-based reverse diffusion model for MRI denoising and super-resolution.

Although FFN-based methods have demonstrated impressive performance, they face challenges in recovering fine details, particularly with high upsampling factors. Generative models can produce realistic super-resolution results, but the synthetic details may not align well with the original low-resolution images. Besides, both FFN-based and
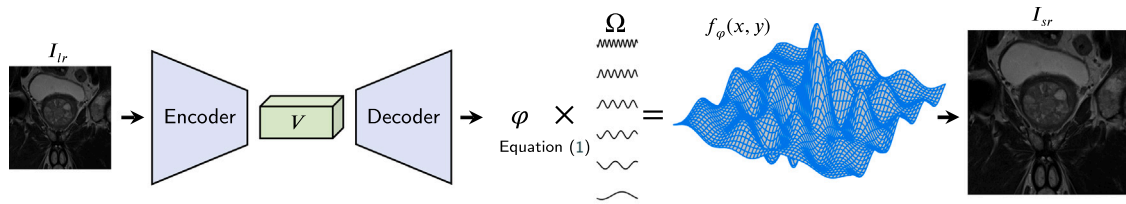
Fig. 2. The overall pipeline of NExpR. The encoder takes the discrete low-resolution image I_{lr} as input and extracts feature maps V. The decoder subsequently predicts coefficients φ of the Fourier transform function. The image is represented by the continuous Fourier transform function in Eq. (1). The discrete super-resolution image I_{sr} is obtained by rasterizing the function.

generative model-based methods are limited to upsampling images at a predefined scale. In contrast, by using a continuous explicit representation, NExpR not only achieves better consistency with low-resolution inputs, but also enables flexible upsampling without the need for multiple trainings at different SR scales.

Implicit Neural Representations. Recent studies have demonstrated the potential of neural network-parameterized implicit functions for object parts [37], 3D objects [38,39], object appearance [40,41], and 2D images [16,42]. Tancik et al. [16] and Sitzmann et al. [42] have demonstrated INRs tend to learn low-frequency content and propose Fourier mapping [16] or periodic activations [42] to overcome the spectral bias and enable INRs to capture higher-frequency details. Benbarka et al. [17] investigated the connection between INRs and Fourier series and proposed a new integer Fourier mapping strategy.

Chen et al. proposed the 'Local Implicit Image Function (LIIF)' [13] that employs INRs for arbitrary-scale image super-resolution. The implicit function in LIIF takes coordinates and local image features as input and predicts the corresponding pixel values. Due to its simplicity and impressive performance, a line of works has been proposed based on LIIF. Lee et al. [14] used a local texture estimator to enhance the model's ability to recover fine details. CiaoSR [43] and SRNO [44] incorporated attention mechanisms to exploit additional non-local information. Besides, flow-based models [45] and diffusion models [46] have also been integrated with LIIF for more realistic outputs.

Recently, INRs have been introduced to medical image super-resolution. Zhang et al. [47] applied spatial and temporal implicit neural representation model in cone-bean CT reconstruction. Xu et al. [48] proposed a slice-to-volume reconstruction method with INR to enhance robustness against the subject motions and image artifacts. Han et al. [49] used a diffusion-based model to achieve arbitrary-scale MRI super-resolution. Li et al. [50] and Zhang et al. [12] designed models based on LIIF for arbitrary-scale multi-contrast MRI super-resolution. Liu et al. [51] designed a neighborhood evaluator-based INR model for efficient medical image super-resolution at arbitrary scales. Duan et al. [52] utilized an INR-based model for texture enhancement in histopathology image super-resolution.

Although INR-based methods achieve descent performance on arbitrary-scale super-resolution, the repeated execution of neural networks brings a high computation cost. Our methods utilizes a compact explicit analytical representation to significantly reduces the computation cost and accelerates the rescaling process without losing the image quality.

3. Neural explicit representation

We introduce the details of our method. We first give an overview of the pipeline and then elaborate on how we convert a discrete image into a continuous function and rasterize the function to SR images at arbitrary scales.

3.1. Overview

Let $I_{lr} \in \mathbb{R}^{h \times w}$ be the low-resolution image to be processed, where $h \times w$ is the spatial dimension of the image. NExpR takes the I_{lr} as input and converts it into a continuous parametric function $f_{\varphi}(\cdot, \cdot)$, where φ denotes the parameters of the function. After obtaining the continuous function $f_{\varphi}(\cdot, \cdot)$, a super-resolution image of arbitrary resolution $I_{sr} \in \mathbb{R}^{sh \times sw}$ can be derived by rasterizing the continuous function, where s is the super-resolution scale.

Fig. 2 illustrates the overall pipeline of NExpR. We employ an encoder–decoder architecture, where the encoder takes the low-resolution image I_{lr} as input and extracts feature maps V. The feature maps are then flattened into a 1D vector before being input into the decoder. The decoder subsequently estimates parameters φ of the explicit analytic function: $f_{\varphi}(\cdot, \cdot)$. Following previous works [13,14], we use the EDSR [7] without the upsampling modules as the encoder because of its simplicity and efficiency. The decoder is a multi-layer perceptron (MLP) consisting of five linear layers, and a ReLU activation function follows each hidden linear layer. The final layer of the decoder predicts the parameters φ.

3.2. Parameterizing the explicit function

Fourier transform function approximation. Given the low-resolution image I_{lr} as input, we aim to learn an explicit function $f_{\varphi}(\cdot, \cdot)$ parameterized by φ to represent the input image. There are various ways of parameterizing an image signal, such as Fourier, Wavelets, b-spline, *etc.*. Without loss of generality and for easy implementation, we use a Fourier transform function to parameterize the explicit function $f_{\varphi}(\cdot, \cdot)$ where φ are the coefficients.

Let $f_{\varphi}(x, y)$ be the bivariate function to be determined, where $\{(x, y) \mid x, y \in [0, 1]\}$ are the normalized coordinates of an arbitrary point in the 2D plane:

$$f_{\varphi}(x, y) = \sum_{p=0}^{N} \sum_{q=0}^{N} \begin{bmatrix} \alpha_{p,q} & \beta_{p,q} \end{bmatrix} \begin{bmatrix} \sin(\omega_p x + \nu_q y) \\ \cos(\omega_p x + \nu_q y) \end{bmatrix}, \quad (1)$$

where $\{(\omega_p, \nu_q) \mid p, q \in \{0, 1, \dots, N\}\}$ are frequencies of Fourier components, and $\{(\alpha_{p,q}, \beta_{p,q}) \mid p, q \in \{0, 1, \dots, N\}\}$ are the coefficients. Our model predicts the coefficients $\alpha_{p,q}, \beta_{p,q}$ in Eq. (1) to obtain the continuous function. We use $\varphi \in \mathbb{R}^{(N+1) \times (N+1) \times 2}$ to denote all the coefficients to be estimated, and $\Omega \in \mathbb{R}^{(N+1) \times 2}$ are all the frequencies (ω_p, ν_q), as shown in Fig. 2.

The Fourier transform function in Eq. (1) requires an infinite number of frequencies ($N \to \infty$) for accurate approximation. This is not practical in numerical computation, and in practice, we use a finite number of frequencies. Experimental results in Section 4 demonstrate that a finite number of frequencies achieve a good level of approximation accuracy.

Learnable frequencies. Since we are using a limited number of frequencies in the Fourier transform function, the choice of frequencies Ω also affects the accuracy of the approximation, and carefully configured frequencies result in more accurate approximations.
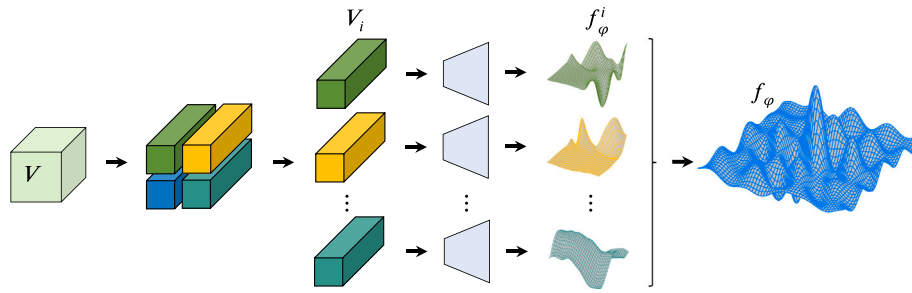
Fig. 3. Patch-wise continuous representation. The decoder processes each patch to obtain its respective continuous representation f_φ^i. Continuous representation of the entire image f_φ is the piece-wise function of the individual functions $f_\varphi^i, i = 1, 2, \ldots$.

Instead of manually tuning the frequencies, we treat the frequencies Ω as parameters of the neural network. In detail, the frequencies Ω for the two spatial directions, $\{\omega\}$ and $\{\nu\}$, are set independently and represented as two sets of learnable parameters in the model. The frequencies are uniformly initialized and then updated during training. Consequently, the model automatically finds the optimal frequencies that best fit the training data. Experiments in Section 4.5 demonstrate that learnable frequencies are superior to constant frequencies.

Patch-wise processing. We process the image in a patch-by-patch manner to capture fine-grained details effectively. In particular, after feature extraction, we partition the feature maps into non-overlapping patches in the spatial plane and then process the patches individually.

Let $V \in \mathbb{R}^{h \times w \times d}$ be the feature maps from the encoder, and $V_i \in \mathbb{R}^{a \times a \times d}$ be the feature of the ith patch, where a is the size of the patch, and d is the dimensionality of the feature. As shown in Fig. 3, each patch feature V_i corresponds to an independent patch-wise Fourier function f_φ^i of the same form as Eq. (1). The continuous representation of the entire image f_φ is a piece-wise function composed of patch-wise functions $\{f_\varphi^1, f_\varphi^2, \ldots\}$. Let $\mathbb{D}_i$ be the domain of f_φ^i, then f_φ is:

$$f_\varphi(x, y) = \begin{cases} f_\varphi^1 & \text{if } (x, y) \in \mathbb{D}_1 \\ f_\varphi^2 & \text{if } (x, y) \in \mathbb{D}_2 \\ \quad \ldots \end{cases} \tag{2}$$

3.3. Continuous function to discrete images: Rasterization

Given the function representation, the last step is to derive a super-resolution image of desired dimensions from the continuous function f_φ. This is achieved by rasterizing f_φ into pixel grids. The rasterization process basically queries (or samples) values from the function f_φ at desired locations. For example, to derive a super-resolution image with dimensions $sh \times sw$, we uniformly sample $sh \times sw$ intensity values within the $[0, 1] \times [0, 1]$ grid. The intensity value of the super-resolution image I_{sr} at location (i, j) can be determined by:

$$I_{sr}[i, j] = f_\varphi(\frac{i}{sh}, \frac{j}{sw}). \tag{3}$$

In this context, a higher upsampling factor s essentially means finer sampling strides $\frac{1}{sh}, \frac{1}{sw}$. The procedure of rasterizing I_{sr} with different upsampling factors is depicted in Fig. 1. The rasterization in Eq. (3) is significantly faster than evaluating a neural network. It is straightforward and efficient, allowing fast rescaling into arbitrary dimensions.

3.4. Model training

In each training iteration, we randomly sample a high-resolution image I_{hr} from the training set and a random downsampling factor s from a certain range. Then, we downsample the I_{hr} by a factor of s to derive the low-resolution image I_{lr}. Let $sh \times sw$ be the size of I_{hr} and $h \times w$ be the size of I_{lr}, and s is the downsampling scale. Note that s could be an arbitrary real number within a certain range, and

we round h and w into the nearest integers to ensure valid image sizes. For example, if $sh \times sw = 256 \times 256$ and $s = 3.141$, then $h = w = [256/3.141] = 81$.

The degraded image I_{lr} is subsequently fed into NExpR to derive the explicit function representation f_φ. Then, we rasterize the continuous function f_φ into a discrete high-resolution image: $I_{sr} \in \mathbb{R}^{sh \times sw}$. Training is performed by minimizing the L_1 disparity between I_{sr} and I_{hr}

$$\mathcal{L} = |I_{sr} - I_{hr}|_1. \tag{4}$$

We train the model, and update model parameters and learnable frequencies Ω with a back-propagation algorithm.

4. Experiments

In this section, we discuss our detailed implementation and report the performance of our proposed method.

4.1. Implementation details

Datasets. Our experiments are conducted on three MRI and one CT datasets. The three MRI datasets include ProstateX [53] fastMRI [54], and our in-house clinical prostate dataset. The CT data is from Medical Segmentation Decathlon (MSD) Liver datasets [55]. Table 1 summarizes the clinical applications (prostate, knee, liver), modalities, and the number of training/test cases of these datasets. We use magnitude images obtained from DICOM format for the MRI datasets and from NIfTI format data for the CT dataset. Our experimental study was performed in compliance with the United States Health Insurance Portability and Accountability Act (HIPAA) of 1996 and was approved by the institutional review board (IRB) with a waiver of the requirement for informed consent. The details of the four datasets are demonstrated below:

(1) ProstateX dataset: The ProstateX dataset [53] is an openly accessible dataset that consists of 348 prostate MR studies acquired on one of two types of Siemens 3T MR scanners, the Magnetom Trio and Skyra. Each study includes T2-weighted (T2 W), proton density-weighted (PD-W), dynamic-contrast-enhanced (DCE), and diffusion-weighted (DW) imaging. For our experiments, we utilize the coronal T2 W images, which were acquired from a turbo spin echo sequence (TSE) with an in-plane resolution of 0.5 mm and a slice thickness of 3.6 mm. The dataset is officially divided into 206 training scans and 142 testing scans, containing 3447 and 2321 slices, respectively.

(2) fastMRI dataset: The fastMRI dataset [54] is a public dataset that includes scans of the knee, brain, prostate, and breast. We select the coronal proton density-weighted images with fat suppression from the knee scans for our experiment. All data are acquired on one of three clinical 3T scanners (Siemens Magnetom Skyra, Prisma and Biograph mMR) or one 1.5T scanner (Siemens Magnetom Aera), with an in-plane resolution of 0.5 mm and a slice thickness of 3 mm. Due to the large scale of the dataset, we randomly selected 600 scans (20 015 slices) for training and 200 scans (6611 slices) for testing.
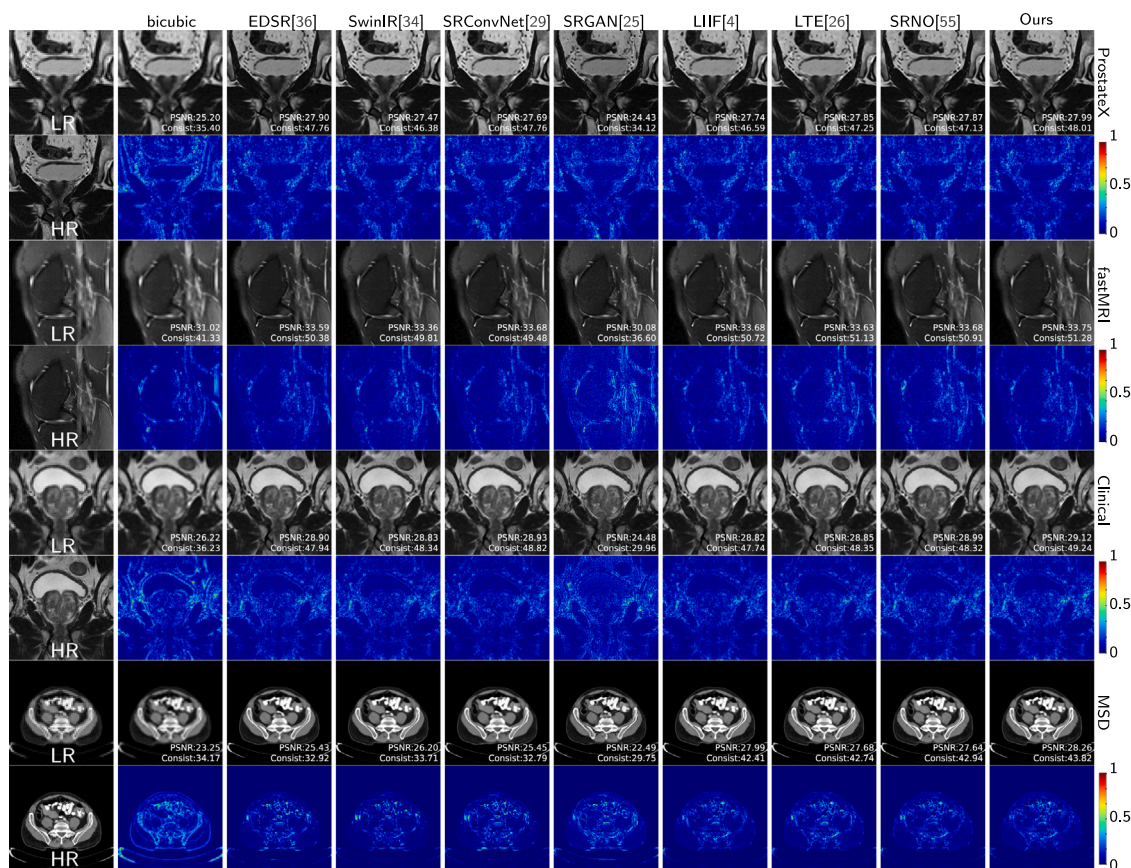
Fig. 4. Example super-resolution results of various methods on four different datasets along with the error maps depicting the differences between the super-resolution results and the high-resolution references.

Table 1

Summary of the four datasets used in our experiments. The numbers in each table cell denote the total number of cases and slices, respectively.

Datasets	ProstateX [53]	fastMRI [54]	Clinical prostate	MSD [55]
Region	prostate	knee	prostate	liver
Modality	MRI (T2-TSE)	MRI (T2-PD)	MRI (T2-TSE)	CT
Plane	coronal	sagittal	coronal	axial
# train	206 (3,447)	600 (20,015)	636 (13,475)	131 (11,679)
# test	142 (2,321)	200 (6,611)	200 (4,246)	70 (5,379)

(3) In-house clinical prostate dataset: This dataset is obtained from our in-house clinical data. We retrospectively reviewed clinical prostate MRI scans from March 2013 to December 2018 at a single academic institution and identified a total of 836 clinical subjects. These scans were acquired in the coronal plane using the multi-slice 2D T2-weighted turbo spin-echo (T2w-TSE) sequence, with an in-plane resolution of 0.625 mm and a slice thickness of 3.6 mm. All scans were performed on one of four Siemens 3T scanners (Prisma, Skyra, Vida, and Trio). We randomly split the dataset into 636 training scans (13 475 slices) and 200 testing scans (4246 slices).

(4) MSD liver dataset [55]: The Medical Segmentation Decathlon (MSD) dataset is an openly accessible medical imaging dataset that includes objects such as the liver, brain, and lung. For our experiment, we selected the liver subset acquired from portal venous phase CT imaging. This subset contains 201 3D CT volumes, which we reformat into axial 2D slices and clip the pixel values to $[-140, 260]$ for clear visualization. This dataset is officially split into 131 training scans (11 679 slices) and 70 testing scans (5379 slices).

Baselines. We compare our proposed method with various competitors including FFN-based models: EDSR [7], SwinIR [8], SRConvNet [25],

generative model: SRGAN [29], and INR-based methods: LIIF [13], LTE [14], SRNO [44]. FFN and generative methods require individual training for each upsampling factor, while INR-based methods and our approach need just one training for all factors.

Image Degradation. For MRI images, following previous works [12, 50,56–58], we adopt the k-space zero padding (KSZP) to generate low-resolution images from high-resolution images. Specifically, we first use the Fourier transform to convert the high-resolution image into k-space. Suppose a downsampling factor of s, we only retain the low-frequency components in the central $\frac{1}{s} \times \frac{1}{s}$ region of the k-space, and then convert the k-space data back to the image domain using inverse Fourier transform. This approach simulates the loss of high-frequency details, which aligns with practical MRI acquisition scenarios where spatial resolution is often compromised due to acquisition time constraints or physical hardware limitations. For CT images, while more sophisticated downsampling techniques, such as *e.g.* sparse-view downsampling [59], would be more representative of real-world scenarios, we follow the common practice of existing works [60–63] and employ bicubic interpolation to downsample the high-resolution images for fast experiments and evaluations.

Setup. We implement NExpR with the PyTorch [64] framework. The model was trained on four NVIDIA Quadro RTX 8000 GPUs and tested on a single GPU of the same model. Similar to the training protocol of LIIF [13], we use the Adam optimizer and the model is trained for 200 epochs. The learning rate is initialized to 10^{-4} and is halved every 50 epochs. Random crop and flipping are used for data augmentation. The size of output images is 256×256 and the batch size is 16. The downsampling factor s is randomly drawn from $(1, 4)$.

The number of frequencies N in Eq. (1) is set to 8 and the frequencies ω, ν are initialized as $0, \pi, 2\pi, \ldots, 8\pi$. ω_0 and ν_0 are fixed to 0 to
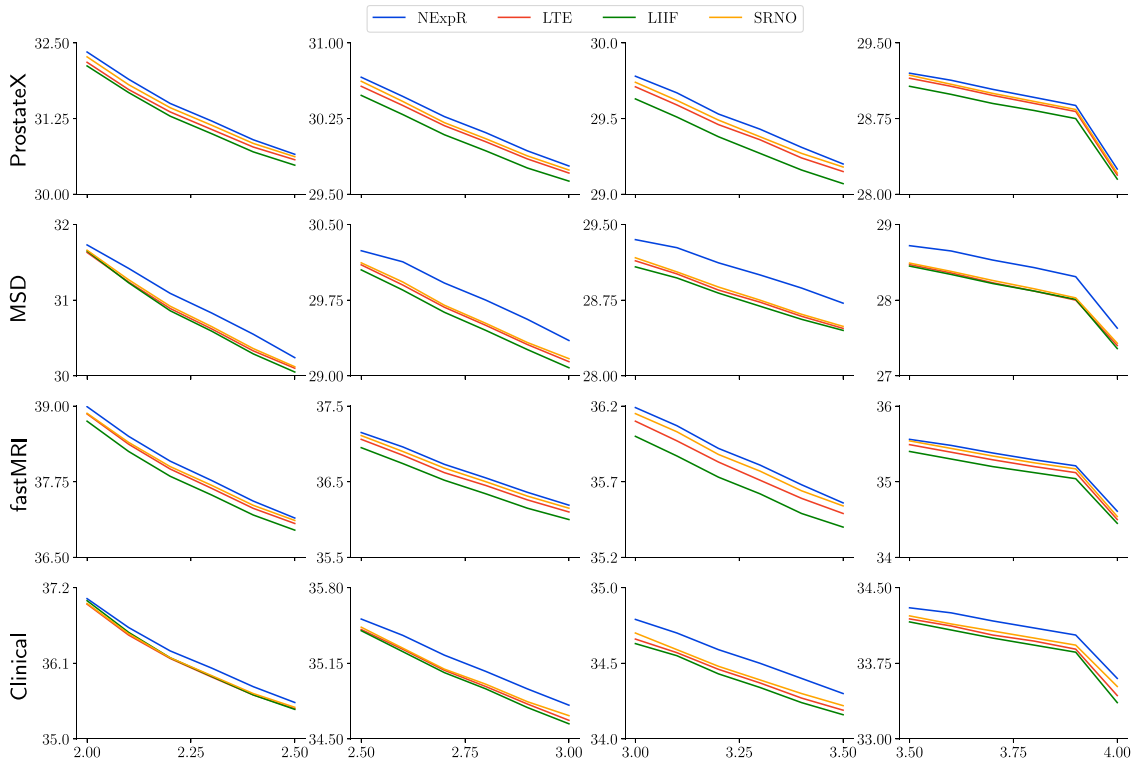
Fig. 5. PSNR under various super-resolution scales on respective test sets. The *x*-axis represents the super-resolution scale, ranging from 2.0 to 4.0, and the *y*-axis represents the PSNR (dB).

represent the direct current (DC) component. The patch size a is set to 2 for the trade-off between efficiency and accuracy. The choice of a, N, and the learned frequencies on different datasets will be discussed in the ablation study in Section 4.5. The settings, together with the model architecture used in this study, are carefully chosen to provide an optimal balance between image quality and efficiency.

4.2. Qualitative results

Fig. 4 presents qualitative comparisons at 4× super-resolution scale ($[64 \times 64] \rightarrow [256 \times 256]$) on four datasets. The error maps highlight the difference between the SR results and the corresponding high-resolution ground truth. In these maps, more textures and more vivid colors indicate larger errors and a lower quality of super-resolution outputs.

Among all the methods in comparison, SRGAN [29] shows decent perceptual quality with realistic fine-grained details, while the error maps reveal significant discrepancies between the SR output and HR ground truth. This is attributed to the training strategy of GANs, which prioritizes generating images that appeal to visual perception rather than faithfully adhering to the pixel-wise content of LR inputs, making their output unreliable for practical medical applications. Compared with other FFN-based and INR-based competitors, NExpR showcases the comparable ability to effectively restore blurred edges and minimize blocking artifacts without compromising consistency with low-resolution inputs. Additionally, NExpR achieves this image quality with the shortest inference time as discussed in Section 4.4.

4.3. Quantitative results

Evaluation metrics. We use Peak-Signal-to-Noise Ratio (**PSNR**) and Structural Similarity Index (**SSIM**) to quantitatively evaluate the super-resolution performance. In addition, we quantify the **consistency**

between low-resolution inputs and the super-resolution results because it is crucial in medical imaging to ensure that the SR images accurately reflect the original anatomical structures. In SR3 [32], the consistency metric was defined as the disparity between the original low-resolution image and the downsampled (DS) super-resolution image.

$$\text{Consist}(I_{sr}, I_{lr}) = \text{PSNR}\big(\text{DS}(I_{sr}), I_{lr}\big), \tag{5}$$

where 'DS' is the image degradation demonstrated in Section 4.1.

We also use the Learned Perceptual Image Patch Similarity (**LPIPS**) metric [65] to evaluate the perceptual quality of the super-resolution images. This is calculated by comparing the deep features extracted from a pre-trained neural network between the SR image and the HR ground truth.

For all results presented in tables, different background colors indicate the best and second best results.

Arbitrary-scale super-resolution. We first evaluate the arbitrary-scale super-resolution performance of NExpR and INR-based methods, *i.e.* LIIF [13], LTE [14], and SRNO [44]. We test scales within the range of $[2, 4]$ because extremely drastic upsampling scales are typically unnecessary for medical images. We first degrade the high-resolution images by a factor of 4 to simulate the low-resolution input I_{lr}. I_{lr} is then upscaled to various scales by different SR methods. We pick 20 upsampling scales uniformly, ranging from 2 to 4. When evaluating the performance, the ground-truth images are obtained via downsampling the original high-resolution image I_{hr} to desired dimensions using corresponding degradation methods. Figs. 5–7 compare the PSNR, SSIM, and consistency results on the four datasets.

As can be observed from Fig. 5, NExpR consistently achieves higher PSNR values. This indicates that NExpR is more effective at minimizing the errors during continuous rescaling, leading to more pixel-wise accurate SR results.

In terms of SSIM in Fig. 6, NExpR outperforms other INR-based methods on the ProstateX [53], MSD [55], and clinical prostate datasets. For the fastMRI dataset [54], NExpR and SRNO [44] are
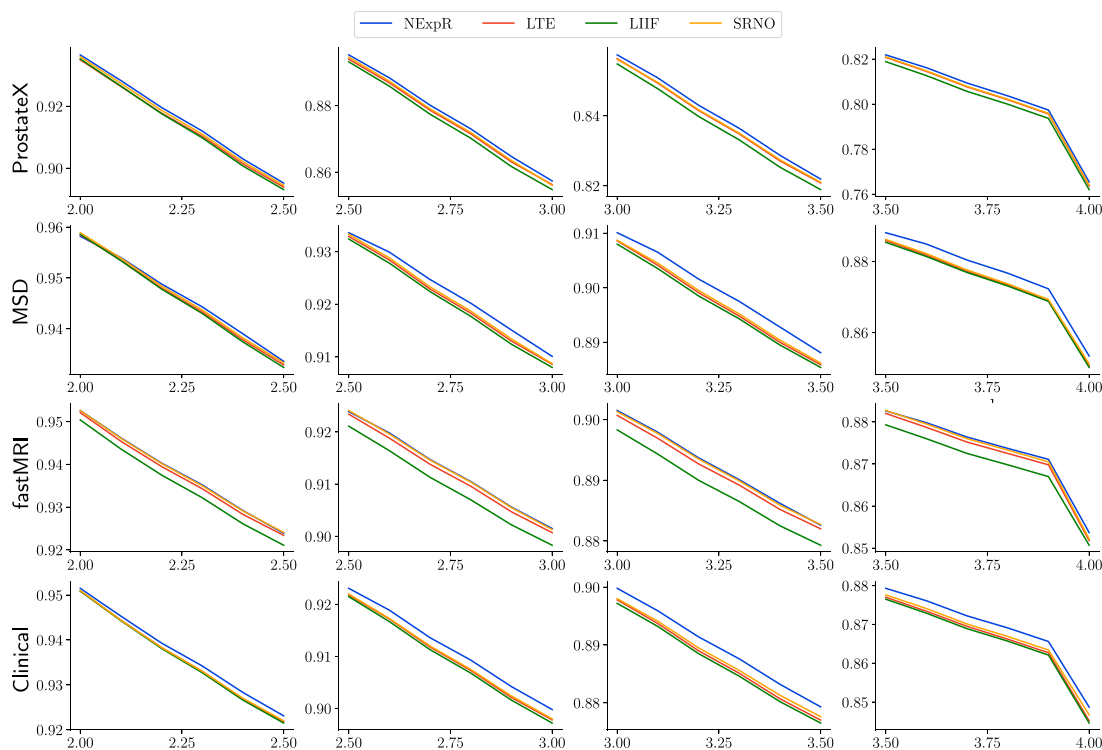
Fig. 6. SSIM under various super-resolution scales on respective test sets. The *x*-axis represents the super-resolution scale, ranging from 2.0 to 4.0, and the *y*-axis represents the SSIM value.
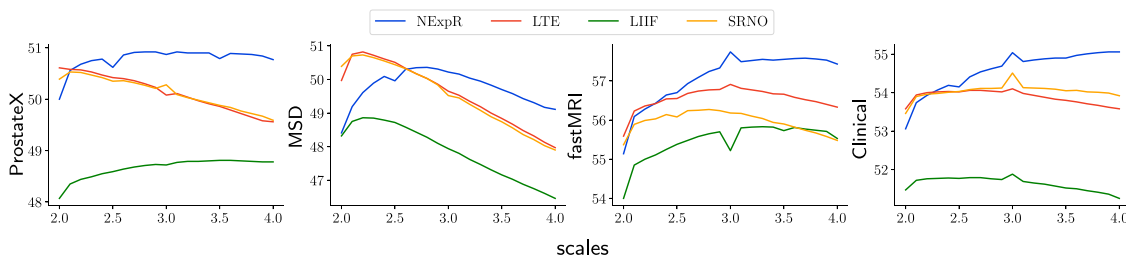

Fig. 7. Consistency under various super-resolution scales on respective test sets. The *x*-axis represents the super-resolution scale, ranging from 2.0 to 4.0, and the *y*-axis represents the Consistency (dB).

nearly tied for first place. The consistently higher SSIM values demonstrate that NExpR can not only maintain pixel-wise accuracy but also preserve important structural patterns such as edges and textures.

Fig. 7 presents the consistency results across different scales and datasets. NExpR does not achieve the highest consistency at small scales (closer to 2×), where methods such as LTE [14] and SRNO [44] perform slightly better. However, NExpR demonstrates the most strong and robust overall consistency across the test scales, particularly excelling at large super-resolution scales On the contrary, the consistency of other INR-based methods decreases more noticeably as the SR scale increases. This indicates NExpR's robustness in handling continuous SR, making it particularly suitable for medical imaging applications that require flexible and accurate rescaling.

Fixed-scale super-resolution. Table 2 summarizes the fixed-scale quantitative comparisons between NExpR and other baseline models on four datasets at fixed super-resolution scales, *i.e.* 2×, 3×, 4×. SwinIR is excluded from the 3× experiments as it only supports super-resolution scales that are powers of two.

At larger super-resolution scales (3× and 4×), NExpR demonstrates superior overall performance compared to other methods in terms of PSNR, SSIM, and consistency. At the 2× scale, NExpR does not achieve

the best performance across all metrics. In some cases, fixed-scale SR methods like EDSR [7], SwinIR [8] and SRConvNet [25] slightly outperform NExpR as they are trained to focus on this specific SR scale. The results suggest that NExpR is better at handling challenging large upscaling factors.

In addition to PSNR, SSIM, and consistency, we also present the LPIPS metric results for a perceptual quality evaluation. It is noticeable that SRGAN [29] exhibits the generally lowest LPIPS error due to the properties of generative models, which aligns with the qualitative results discussed in Section 4.2. Among all other non-generative methods, NExpR shows the lowest LPIPS errors across the majority of super-resolution scales and datasets, indicating the superior perceptual quality.

Statistical analysis. To evaluate the statistical significance of the performance differences between our method and other competitors, we conduct the two-sided paired t-test.

Table 3 presents the mean value of the metrics (PSNR, SSIM and Consistency) across all datasets and SR scales. In particular, they are derived from the entire combined test image cohort ($N = 18557$) of the four datasets and three SR scales (2×, 3×, 4×). Each result is marked

Table 2
PSNR, SSIM, and consistency of different methods under fixed super-resolution scales on respective test sets.

Dataset	Baselines	2×				3×				4×			
		PSNR↑	SSIM↑	Consist↑	LPIPS↓	PSNR↑	SSIM↑	Consist↑	LPIPS↓	PSNR↑	SSIM↑	Consist↑	LPIPS↓
ProstateX [53]	Bicubic	30.62	0.8568	38.56	0.2140	27.46	0.7452	37.20	0.3201	25.75	0.6589	36.27	0.3958
	EDSR [7]	33.88	0.8614	46.48	0.1458	30.15	0.8306	37.92	0.2541	28.24	0.7658	38.57	0.3204
	SwinIR [8]	33.82	0.9120	46.14	0.1472	–	–	–	–	27.89	0.7535	38.23	0.3267
	SRConvNet [25]	33.79	0.9123	46.23	0.1470	30.07	0.8295	37.86	0.2560	28.09	0.7602	38.28	0.3259
	SRGAN [29]	31.25	0.8579	40.55	0.0984	27.10	0.7224	34.24	0.1921	24.80	0.6756	32.45	0.2497
	LIIF [13]	33.68	0.9100	49.51	0.1506	30.09	0.8291	49.59	0.2566	28.15	0.7621	48.78	0.3261
	LTE [14]	33.77	0.9108	53.01	0.1484	30.14	0.8302	51.52	0.2567	28.19	0.7637	49.56	0.3259
	SRNO [44]	33.77	0.9107	52.79	0.1496	30.15	0.8301	51.35	0.2570	28.21	0.7641	49.59	0.3257
	Ours	33.81	0.9110	53.00	0.1469	30.20	0.8312	52.31	0.2505	28.25	0.7655	50.77	0.3140
fastMRI [54]	Bicubic	36.48	0.9059	44.21	0.1896	33.69	0.8454	43.45	0.2852	32.17	0.8454	42.86	0.3491
	EDSR [7]	39.52	0.9410	49.67	0.1235	36.14	0.8876	43.80	0.2312	34.49	0.8523	44.49	0.2884
	SwinIR [8]	39.53	0.9409	50.13	0.1243	–	–	–	–	34.22	0.8466	44.19	0.2957
	SRConvNet [25]	39.62	0.9423	49.90	0.1208	36.24	0.8898	43.96	0.2272	34.55	0.8535	44.52	0.2859
	SRGAN [29]	36.49	0.8994	42.23	0.0856	32.51	0.7997	38.00	0.1671	32.13	0.7536	36.36	0.2127
	LIIF [13]	39.38	0.9384	54.55	0.1350	36.12	0.8867	56.76	0.2310	34.45	0.8507	55.53	0.2906
	LTE [14]	39.44	0.9391	57.40	0.1290	36.17	0.8873	57.90	0.2275	34.50	0.8516	55.80	0.2885
	SRNO [44]	39.48	0.9395	56.98	0.1315	36.20	0.8877	57.20	0.2302	34.54	0.8522	55.48	0.2893
	Ours	39.53	0.9399	56.43	0.1246	36.27	0.8887	58.24	0.2201	34.61	0.8537	57.43	0.2777
Clinical Prostate	Bicubic	36.30	0.9163	42.98	0.1670	32.85	0.8403	41.66	0.2693	31.00	0.7783	40.81	0.3451
	EDSR [7]	40.16	0.9518	48.90	0.1090	35.67	0.8949	42.23	0.2046	33.48	0.8461	42.45	0.2748
	SwinIR [8]	38.88	0.9431	47.02	0.1213	–	–	–	–	33.45	0.8457	42.44	0.2745
	SRConvNet [25]	40.22	0.9519	48.79	0.1068	35.79	0.8969	42.31	0.2013	33.50	0.8467	42.77	0.2717
	SRGAN [29]	36.11	0.9128	39.98	0.0742	31.36	0.8292	34.39	0.1569	31.44	0.7911	31.82	0.2309
	LIIF [13]	39.87	0.9501	53.42	0.1158	35.60	0.8942	52.84	0.2071	33.36	0.8445	51.25	0.2780
	LTE [14]	39.99	0.9504	55.78	0.1151	35.66	0.8946	55.47	0.2076	33.43	0.8452	53.48	0.2792
	SRNO [44]	40.04	0.9506	55.34	0.1134	35.75	0.8953	55.33	0.2052	33.52	0.8467	53.92	0.2758
	Ours	40.06	0.9509	55.21	0.1082	35.82	0.8965	56.02	0.1965	33.60	0.8487	55.06	0.2631
MSD Liver [55]	Bicubic	27.34	0.8719	43.08	0.1394	24.57	0.7727	46.58	0.2175	23.11	0.6992	35.22	0.2700
	EDSR [7]	31.18	0.9164	39.77	0.0960	25.47	0.8203	25.75	0.1743	25.21	0.8109	28.56	0.1820
	SwinIR [8]	30.91	0.9155	38.80	0.0946	–	–	–	–	25.83	0.8243	28.33	0.1737
	SRConvNet [25]	31.30	0.9181	39.93	0.0924	25.59	0.8253	25.89	0.1621	25.38	0.8154	28.55	0.1749
	SRGAN [29]	28.55	0.8666	37.53	0.0606	22.83	0.7444	25.60	0.1318	22.51	0.7328	27.09	0.1405
	LIIF [13]	33.12	0.9382	51.43	0.0776	29.53	0.8896	48.80	0.1305	27.36	0.8502	46.46	0.1648
	LTE [14]	33.19	0.9385	52.94	0.0783	29.59	0.8899	50.68	0.1325	27.40	0.8506	47.98	0.1682
	SRNO [44]	33.19	0.9388	52.80	0.0772	29.61	0.8903	50.50	0.1322	27.43	0.8512	47.90	0.1679
	Ours	33.18	0.9390	52.78	0.0739	29.79	0.8915	51.31	0.1253	27.63	0.8534	49.11	0.1548

Table 3
Average performance across the test sets from all datasets and three SR scales.

Method	PSNR↑	SSIM↑	Consist↑
EDSR [7]	33.15*	0.8749*	45.62*
SwinIR [8]	33.53*	0.8816*	45.89*
SRConvNet [25]	33.18*	0.8747*	47.72*
SRGAN [29]	29.66*	0.7960*	36.59*
LIIF [13]	33.82*	0.8851*	48.62*
LTE [14]	33.89*	0.8857*	49.07*
SRNO [44]	33.92*	0.8861*	49.06*
Ours	34.00	0.8870	49.33

* Indicates p-value ≤ 0.001.

with an asterisk (∗) if the corresponding p-value is less than or equal to 0.001.

The analysis indicates that our method statistically outperforms other methods. Moreover, all p-values are less than 0.001, meaning that the observed differences are statistically significant.

Cross-dataset validation. We compare the generalizability of different methods via cross-dataset validation. In particular, we train the model on one dataset and test the performance on the other dataset. The results are in Table 4 where 'A → B' indicates the model is trained on dataset A and tested on dataset B. As shown in Table 4, our method achieves the best consistency and competitive PSNR and SSIM across all configurations. This demonstrates NExpR's strong generalization ability.

4.4. Runtime and model complexity

Walltime and model complexity. Table 5 reports the model parameters (Params), Giga Floating Point Operations (GFLOPS), and inference time.

Considering the intentional lightweight design of the model architecture of SRConvNet [25], it achieves the smallest model size and lowest GFLOPS. Similarly, NExpR, due to the compact explicit function representation, also demonstrates a significant reduction in computational complexity and model parameters, achieving the second smallest model size and the second lowest GFLOPS. Notably, NExpR delivers the best-accelerated inference speed in comparison to competing approaches, further highlighting its efficiency.

Rescaling speed. In this experiment, we compare the speed of rescaling images at arbitrary scales after obtaining the function representation (either neural implicit or our explicit).

We pass the low-resolution image into the models to get the function representation, and then measure the time of querying the function to get super-resolution images of desired dimensions. The measured time represents the speed of arbitrary-scale super-resolution after obtaining the function representation.

Fig. 8 compares runtime for executing these functions with various super-resolution scales on the ProstateX dataset. Notably, our method is over 100× faster than LIIF [13], LTE [14] and SRNO [44], reducing the rescaling time from the order of 1 ms to the order of 0.01 ms. This significant speed improvement suggests that our method achieves blazingly fast arbitrary-scale super-resolution and enables smooth zoom-in and out in clinical practice.

Table 4
Cross-dataset performance of different methods on respective test setsfor 4× super-resolution scale.

| Method	ProstateX → In-house			In-house → ProstateX			ProstateX → fastMRI			fastMRI → ProstateX		
	PSNR↑	SSIM↑	Consist↑	PSNR↑	SSIM↑	Consist↑	PSNR↑	SSIM↑	Consist↑	PSNR↑	SSIM↑	Consist↑
EDSR [7]	33.24	0.8435	42.45	28.17	0.7631	38.28	34.05	0.8442	44.30	27.79	0.7520	38.30
SwinIR [8]	32.44	0.8200	41.32	28.14	0.7621	38.29	33.28	0.7935	41.99	27.58	0.7433	37.90
SRConvNet [25]	33.09	0.8403	42.62	28.13	0.7620	38.35	33.79	0.8410	44.58	27.71	0.7488	38.17
SRGAN [29]	28.89	0.6916	34.46	24.26	0.6159	28.89	30.82	0.7061	36.75	24.45	0.6145	30.02
LIIF [13]	33.15	0.8413	49.93	28.12	0.7624	48.26	34.08	0.8430	50.95	27.92	0.7552	49.55
LTE [14]	33.22	0.8426	52.25	28.15	0.7627	49.70	34.09	0.8449	54.02	27.92	0.7552	50.08
SRNO [44]	33.26	0.8432	52.97	28.23	0.7641	49.96	34.07	0.8443	54.88	27.90	0.7546	49.52
Ours	33.21	0.8431	53.40	28.30	0.7674	51.28	34.08	0.8444	54.96	27.95	0.7567	50.33
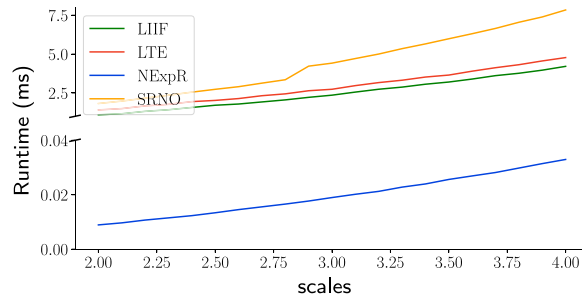
Fig. 8. Rescaling runtime of arbitrary-scale SR methods under various upscaling scales on the ProstateX test set.
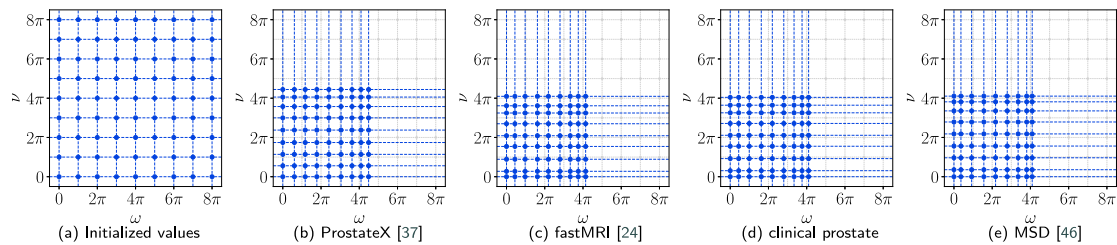

Fig. 9. Initialization and optimization of frequencies across four datasets. The connecting lines represent frequency values, while the points indicate different combinations of ω, ν.

Table 5
Model complexity and inference speed of different methods. The numbers are measured for 4× super-resolution($[64 \times 64] \rightarrow [256 \times 256]$) on the ProstateX test set.

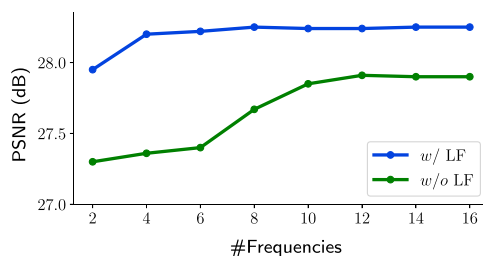| Method	Params (M)	GFLOPS	Time (ms)
EDSR [7]	5.8	8.04	4.2
SwinIR [8]	45.2	50.46	57.1
SRConvNet [25]	1.0	1.74	14.7
SRGAN [29]	5.8	10.31	3.2
LIIF [13]	1.6	95.52	4.6
LTE [14]	1.7	57.87	5.8
SRNO [44]	2.6	51729.14	8.9
Ours	1.6	5.36	3.0

Fig. 10. 4× SR with different numbers of frequencies and learnable frequencies on the ProstateX test set.

4.5. Ablation study

In this section, we investigate the impact of each component in our model design. Specifically, they are (i) the number of frequencies in Fourier transform function (N in Eq. (1)), (ii) the learnable frequencies (LF), and (iii) the patch size a in patch-wise processing. All ablation studies are conducted on the ProstateX [53] dataset with 4× super-resolution scale.

Number of Frequencies. We compare the performance of NExpR with different numbers of frequencies (N in Eq. (1)). The frequencies ω and ν are initialized uniformly from 0 to $N\pi$. As depicted by the blue polyline in Fig. 10, there is a noticeable decline in PSNR with smaller N values, while models with N larger than 8 have similar performance in terms of PSNR. This indicates that $N = 8$ is adequate for accurately approximating medical images.

Learnable Frequencies. To evaluate the contribution of learnable frequencies (LF), we conduct ablation studies by training the NExpR with *fixed frequencies* versus *learnable frequencies*. Quantitative and qualitative results are demonstrated in Fig. 10 and Fig. 11a, respectively. As shown in Fig. 10, NExpR with learnable frequencies significantly outperforms fixed frequencies. Moreover, as illustrated in Fig. 11a, the super-resolution output of NExpR w/o LF presents visible discontinuities across patches. In contrast, NExpR $w/$ LF gets a more continuous and smoother output. This underscores the benefits of enabling NExpR to adaptively learn optimal frequencies, thereby improving its representational capabilities for medical image super-resolution.
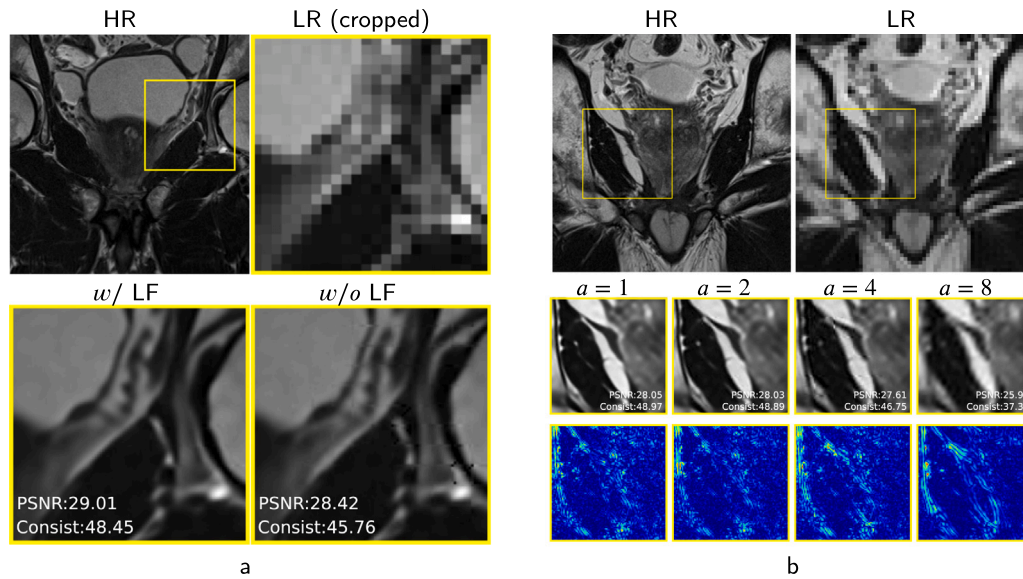
Fig. 11. (a) Qualitative comparison of *w* and *w/o* LF. The number of frequencies *N* is fixed to 8 for a fair comparison. (b) Example super-resolution results with different patch sizes *a*.

Table 6
PSNR and efficiency comparison of NExpR (×4) with different patch sizes *a* on the ProstateX test set.

a	PSNR (dB)	SSIM	Consist (dB)	GFLOPS	Time (ms)
1	28.27	0.7674	50.82	6.05	3.4
2	28.25	0.7655	50.77	5.36	3.0
4	27.82	0.7500	47.87	5.38	3.2
8	26.03	0.6735	37.34	6.14	3.3

We further illustrate the changes in the learnable frequencies ω, ν, including the initial values before training and the optimized values after convergence. Fig. 9a shows the uniformly initialized frequencies, and Fig. 9b–e present the optimized frequencies on four different datasets. We find that the learnable frequencies tend to converge to lower values.

Patch Size. To explore the effect of patch size *a* in the patch-wise processing, we implement ablation experiments with different patch sizes, *i.e.* 1, 2, 4, and 8. As shown in Table 6, PSNR, SSIM and consistency decrease when the patch size enlarges, indicating a drop in super-resolution quality with larger patches. Moreover, small patch sizes, *e.g.* $a = 1$, lead to an excessive segmentation of the image into numerous patches, while large patch sizes, like $a = 8$, increase the input feature dimensionality for the decoder. Both extreme selections of patch size result in increased computation for the decoder, negatively affecting the complexity and inference speed of the super-resolution process. Qualitative comparisons of super-resolution outputs from NExpR with varying patch sizes are presented in Fig. 11b. The super-resolution output with $a = 2$ is similar to that with $a = 1$, whereas larger patch sizes, *e.g.* 4 and 8, generate super-resolution images with blurred edges and discontinuous artifacts between patches. We use patch size $a = 2$ for an excellent trade-off between efficiency and image quality.

5. Discussion

We demonstrate that the proposed framework can significantly accelerate the medical image super-resolution at arbitrary scales while maintaining competitive image quality.

NExpR first extracts the feature map from a low-resolution image and then decodes it into a Fourier transform-based explicit function by predicting its coefficients. Using this explicit analytical function, NExpR can generate SR images at arbitrary scales via rasterization. This simple and efficient explicit analytical representation demonstrates its superiority in reducing computation costs and inference time. Experiments show that NExpR not only achieves low model complexity and inference time in fixed-scale super-resolution, but also significantly accelerates the continuous rescaling process. This is because, unlike conventional INR-based models that require repeated inference through a complex neural network for different SR scales, NExpR employs a simple explicit function representation for fast rasterization during rescaling.

NExpR also achieves competitive image quality. In the arbitrary-scale super-resolution comparison, NExpR robustly outperforms other INR-based methods, *i.e.* LIIF [13], LTE [14], and SRNO [44]. We believe that this is because the compact and concise explicit analytical function is easy for the model to learn and predict, enhancing its ability to continuously preserve image information across different scales. When compared to the fixed-scale super-resolution methods, such as EDSR [7] and SRConvNet [25], NExpR performs slightly below some of these methods at the 2× scale. This likely be because that 2× SR is relatively simple, and fixed-scale SR models are trained specifically for a single scale, allowing them to be finely tuned to that specific task. However, for larger scales like 3× and 4×, NExpR demonstrates superior performance, highlighting its ability to handle more challenging super-resolution tasks. Moreover, the cross-validation experiment indicates that NExpR performs well even when the model is trained and tested on different datasets, suggesting its robustness in generalizing to different image patterns.

Despite the advantages, our study has several limitations, which will guide our potential future research directions. Firstly, NExpR is currently restricted to representing 2D signals, limiting its generalizability in higher-dimensional domains such as volumetric medical imaging. Potential future work could explore analytical representations of higher-dimensional signals, such as 3D volumes (*e.g.*, 3D isotropic MRI scans) or 4D spatiotemporal data (*e.g.*, 4D Dynamic-Contrast-Enhancement (DCE) MRI scans).

Secondly, while NExpR excels in speed and flexibility, it still exhibits some degree of over-smoothing in the SR images, indicating that there is room for improvement in capturing fine-grained details before being fully implemented in real clinical scenarios. Future work may focus on exploring a more advanced framework with a better balance between efficiency and detail-resolving ability.

Lastly, our current evaluation pipeline does not include expert reader studies or clinical task-specific evaluations. This work primarily focuses on proposing a generally applicable method in medical imaging and utilizes standard super-resolution evaluation metrics. However, the clinical relevance of the results for real-world medical practice has not been fully explored. Future research could involve collaboration with clinical experts to ensure that the method meets the standard for clinical translation.

6. Conclusion

In this study, we propose the neural explicit representation for medical image super-resolution. Our method represents an image using an explicit analytical function based on the Fourier transform. Unlike previous implicit neural representation (INR) methods, which involve repeated inference through complex neural networks, our approach uses a concise analytical function that can be directly rasterized, enabling fast super-resolution at arbitrary scales.

We evaluated our method on three MRI datasets and one CT dataset. The experimental results demonstrate that NExpR achieves competitive or superior image quality with a much faster runtime than state-of-the-art super-resolution methods. Specifically, our method accelerates the rescaling process by over 100 times without sacrificing the image quality.

However, our approach also has certain limitations. NExpR is currently limited to 2D image representation, restricting its application in higher-dimensional imaging tasks. Also, our method shows some degree of over-smoothing, suggesting a need for better fine-detail preservation. Additionally, our experiments lack expert reader studies, limiting the assessment of its real-world performance in medical practice. Potential future work would explore enhanced explicit representations for higher-dimensional data while involving clinical task-specific evaluation.

CRediT authorship contribution statement

Kaifeng Pang: Writing – original draft, Validation, Software, Methodology, Data curation, Conceptualization. **Kai Zhao:** Writing – original draft, Methodology, Conceptualization. **Alex Ling Yu Hung:** Writing – original draft, Conceptualization. **Haoxin Zheng:** Writing – review & editing. **Ran Yan:** Writing – review & editing. **Kyunghyun Sung:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The author(s) declare no competing interests.

Acknowledgments

References

[1] R. Cao, A.M. Bajgiran, S.A. Mirak, S. Shakeri, X. Zhong, D. Enzmann, S. Raman, K. Sung, Joint prostate cancer detection and gleason score prediction in mp-MRI via FocalNet, IEEE TMI 38 (11) (2019) 2496–2506.

[2] V. Tsapaki, J. Damilakis, G. Paulo, A.A. Schegerer, J. Repussard, W. Jaschke, G. Frija, CT diagnostic reference levels based on clinical indications: results of a large-scale European survey, Eur. Radiol. 31 (2021) 4459–4469.

[3] Z.-P. Liang, P.C. Lauterbur, Principles of Magnetic Resonance Imaging, SPIE Optical Engineering Press Belllingham, WA, 2000.

[4] G.J. Grevera, J.K. Udupa, An objective comparison of 3-D image interpolation methods, IEEE Trans. Med. Imaging 17 (4) (1998) 642–652.

[5] T.M. Lehmann, C. Gonner, K. Spitzer, Survey: Interpolation methods in medical image processing, IEEE Trans. Med. Imaging 18 (11) (1999) 1049–1075.

[6] A. Li, K. Mueller, T. Ernst, Methods for efficient, high quality volume resampling in the frequency domain, in: IEEE Visualization 2004, IEEE, 2004, pp. 3–10.

[7] B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, Enhanced deep residual networks for single image super-resolution, in: CVPRW, 2017, pp. 136–144.

[8] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, R. Timofte, Swinir: Image restoration using swin transformer, in: ICCV, 2021, pp. 1833–1844.

[9] H. Li, Y. Jia, H. Zhu, B. Han, J. Du, Y. Liu, Multi-level feature extraction and reconstruction for 3D MRI image super-resolution, Comput. Biol. Med. (2024) 108151.

[10] A.L.Y. Hung, K. Zhao, H. Zheng, R. Yan, S.S. Raman, D. Terzopoulos, K. Sung, Med-cDiff: Conditional medical image generation with diffusion models, Bioengineering 10 (11) (2023) 1258.

[11] Q. Delannoy, C.-H. Pham, C. Cazorla, C. Tor-Díez, G. Dollé, H. Meunier, N. Bednarek, R. Fablet, N. Passat, F. Rousseau, SegSRGAN: Super-resolution and segmentation using generative adversarial networks—Application to neonatal brain MRI, Comput. Biol. Med. 120 (2020) 103755.

[12] J. Zhang, Y. Chi, J. Lyu, W. Yang, Y. Tian, Dual arbitrary scale super-resolution for multi-contrast MRI, in: MICCAI, Springer, 2023, pp. 282–292.

[13] Y. Chen, S. Liu, X. Wang, Learning continuous image representation with local implicit image function, in: CVPR, 2021, pp. 8628–8638.

[14] J. Lee, K.H. Jin, Local texture estimator for implicit representation function, in: CVPR, 2022, pp. 1929–1938.

[15] Q.H. Nguyen, W.J. Beksi, Single image super-resolution via a dual interactive implicit neural network, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 4936–4945.

[16] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, R. Ng, Fourier features let networks learn high frequency functions in low dimensional domains, NeurIPS 33 (2020) 7537–7547.

[17] N. Benbarka, T. Höfer, A. Zell, et al., Seeing implicit neural representations as fourier series, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 2041–2050.

[18] G. Freedman, R. Fattal, Image and video upscaling from local self-examples, ACM Trans. Graph. 30 (2) (2011) 1–11.

[19] J. Sun, Z. Xu, H.-Y. Shum, Image super-resolution using gradient profile prior, in: CVPR, IEEE, 2008, pp. 1–8.

[20] K.I. Kim, Y. Kwon, Single-image super-resolution using sparse regression and natural image prior, IEEE Trans. Pattern Anal. Mach. Intell. 32 (6) (2010) 1127–1133.

[21] Z. Xiong, X. Sun, F. Wu, Robust web image/video super-resolution, IEEE Trans. Image Process. 19 (8) (2010) 2017–2028.

[22] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution as sparse representation of raw image patches, in: CVPR, IEEE, 2008, pp. 1–8.

[23] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.

[24] W. Han, S. Chang, D. Liu, M. Yu, M. Witbrock, T.S. Huang, Image super-resolution via dual-state recurrent networks, in: CVPR, 2018, pp. 1654–1663.

[25] F. Li, R. Cong, J. Wu, H. Bai, M. Wang, Y. Zhao, SRConvNet: A transformer-style ConvNet for lightweight image super-resolution, Int. J. Comput. Vis. (2024) 1–17.

[26] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: ECCV, Springer, 2016, pp. 694–711.

[27] M.S. Sajjadi, B. Scholkopf, M. Hirsch, Enhancenet: Single image super-resolution through automated texture synthesis, in: ICCV, 2017, pp. 4491–4500.

[28] W.Y.R. Fok, A. Fieselmann, M. Herbst, L. Ritschl, S. Kappler, S. Saalfeld, Deep learning in computed tomography super resolution using multi-modality data training, Med. Phys. 51 (4) (2024) 2846–2860.

[29] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: CVPR, 2017, pp. 4681–4690.

[30] X. Wang, L. Xie, C. Dong, Y. Shan, Real-esrgan: Training real-world blind super-resolution with pure synthetic data, in: ICCV, 2021, pp. 1905–1914.

[31] D. Chira, I. Haralampiev, O. Winther, A. Dittadi, V. Liévin, Image super-resolution with deep variational autoencoders, in: ECCV, Springer, 2022, pp. 395–411.

[32] C. Saharia, J. Ho, W. Chan, T. Salimans, D.J. Fleet, M. Norouzi, Image super-resolution via iterative refinement, IEEE Trans. Pattern Anal. Mach. Intell. 45 (4) (2022) 4713–4726.

[33] K. Zhao, K. Pang, A.L.Y. Hung, H. Zheng, R. Yan, K. Sung, Mri super-resolution with partial diffusion models, IEEE Transactions on Medical Imaging (2024).

[34] J. Wang, Y. Chen, Y. Wu, J. Shi, J. Gee, Enhanced generative adversarial network for 3D brain MRI super-resolution, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 3627–3636.

[35] L. Zhao, H. Chi, T. Zhong, Y. Jia, Perception-oriented generative adversarial network for retinal fundus image super-resolution, Comput. Biol. Med. 168 (2024) 107708.

[36] H. Chung, E.S. Lee, J.C. Ye, MR image denoising and super-resolution using regularized reverse diffusion, IEEE TMI 42 (4) (2022) 922–934.

[37] K. Genova, F. Cole, D. Vlasic, A. Sarna, W.T. Freeman, T. Funkhouser, Learning shape templates with structured implicit functions, in: ICCV, 2019, pp. 7154–7164.

[38] K. Genova, F. Cole, A. Sud, A. Sarna, T. Funkhouser, Local deep implicit functions for 3d shape, in: CVPR, 2020, pp. 4857–4866.

[39] B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, R. Ng, Nerf: Representing scenes as neural radiance fields for view synthesis, Commun. ACM 65 (1) (2021) 99–106.

[40] S. Saito, Z. Huang, R. Natsume, S. Morishima, A. Kanazawa, H. Li, Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization, in: ICCV, 2019, pp. 2304–2314.

[41] V. Sitzmann, M. Zollhöfer, G. Wetzstein, Scene representation networks: Continuous 3d-structure-aware neural scene representations, NeurIPS 32 (2019).

[42] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, G. Wetzstein, Implicit neural representations with periodic activation functions, NeurIPS 33 (2020) 7462–7473.

[43] J. Cao, Q. Wang, Y. Xian, Y. Li, B. Ni, Z. Pi, K. Zhang, Y. Zhang, R. Timofte, L. Van Gool, Ciaosr: Continuous implicit attention-in-attention network for arbitrary-scale image super-resolution, in: CVPR, 2023, pp. 1796–1807.

[44] M. Wei, X. Zhang, Super-resolution neural operator, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 18247–18256.

[45] J.-E. Yao, L.-Y. Tsao, Y.-C. Lo, R. Tseng, C.-C. Chang, C.-Y. Lee, Local implicit normalizing flow for arbitrary-scale image super-resolution, in: CVPR, 2023, pp. 1776–1785.

[46] S. Gao, X. Liu, B. Zeng, S. Xu, Y. Li, X. Luo, J. Liu, X. Zhen, B. Zhang, Implicit diffusion models for continuous super-resolution, in: CVPR, 2023, pp. 10021–10030.

[47] Y. Zhang, H.-C. Shao, T. Pan, T. Mengke, Dynamic cone-beam CT reconstruction using spatial and temporal implicit neural representation learning (STINR), Phys. Med. Biol. 68 (4) (2023) 045005.

[48] J. Xu, D. Moyer, B. Gagoski, J.E. Iglesias, P.E. Grant, P. Golland, E. Adalsteinsson, NeSVoR: Implicit neural representation for slice-to-volume reconstruction in MRI, IEEE TMI (2023).

[49] Z. Han, W. Huang, Arbitrary scale super-resolution diffusion model for brain MRI images, Comput. Biol. Med. 170 (2024) 108003.

[50] G. Li, L. Zhao, J. Sun, Z. Lan, Z. Zhang, J. Chen, Z. Lin, H. Lin, W. Xing, Re-thinking multi-contrast MRI super-resolution: Rectangle-window cross-attention transformer and arbitrary-scale upsampling, in: ICCV, 2023, pp. 21230–21240.

[51] Z. Liu, J. Han, J. Liu, Z.-C. Li, G. Zhai, Neighborhood evaluator for efficient super-resolution reconstruction of 2D medical images, Comput. Biol. Med. 171 (2024) 108212.

[52] M. Duan, L. Qu, Z. Yang, M. Wang, C. Zhang, Z. Song, Towards arbitrary-scale histopathology image super-resolution: An efficient dual-branch framework via implicit self-texture enhancement, 2024, arXiv preprint arXiv:2401.15613.

[53] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, H. Huisman, Prostatex challenge data, Cancer Imaging Arch 10 (2017) K9TCIA.

[54] F. Knoll, J. Zbontar, A. Sriram, M.J. Muckley, M. Bruno, A. Defazio, M. Parente, K.J. Geras, J. Katsnelson, H. Chandarana, et al., FastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning, Radiol.: Artif. Intell. 2 (1) (2020) e190007.

[55] A.L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B.A. Landman, G. Litjens, B. Menze, et al., A large annotated medical image dataset for the development and evaluation of segmentation algorithms, 2019, arXiv preprint arXiv:1902.09063.

[56] G. Li, J. Lyu, C. Wang, Q. Dou, J. Qin, WavTrans: Synergizing wavelet and cross-attention transformer for multi-contrast MRI super-resolution, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2022, pp. 463–473.

[57] Q. Lyu, H. Shan, C. Steber, C. Helis, C. Whitlow, M. Chan, G. Wang, Multi-contrast super-resolution MRI through a progressive network, IEEE Trans. Med. Imaging 39 (9) (2020) 2738–2749.

[58] G. Li, J. Lv, Y. Tian, Q. Dou, C. Wang, C. Xu, J. Qin, Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast MRI super-resolution, in: CVPR, 2022, pp. 20636–20645.

[59] Y. Song, L. Shen, L. Xing, S. Ermon, Solving inverse problems in medical imaging with score-based generative models, 2021, arXiv preprint arXiv:2111.08005.

[60] J. Chi, Z. Sun, H. Wang, P. Lyu, X. Yu, C. Wu, CT image super-resolution reconstruction based on global hybrid attention, Comput. Biol. Med. 150 (2022) 106112.

[61] X. Jiang, M. Liu, F. Zhao, X. Liu, H. Zhou, A novel super-resolution CT image reconstruction via semi-supervised generative adversarial network, Neural Comput. Appl. 32 (2020) 14563–14578.

[62] C. Jiang, Q. Zhang, R. Fan, Z. Hu, Super-resolution CT image reconstruction based on dictionary learning and sparse representation, Sci. Rep. 8 (1) (2018) 8799.

[63] X. Jiang, Y. Xu, P. Wei, Z. Zhou, Ct image super resolution based on improved srgan, in: 2020 5th International Conference on Computer and Communication Systems, ICCCS, IEEE, 2020, pp. 363–367.

[64] 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, NeurIPS 32 (2019).

[65] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, 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.