

# CuNeRF: Cube-Based Neural Radiance Field for Zero-Shot Medical Image Arbitrary-Scale Super Resolution

Zixuan Chen<sup>1</sup>, Jianhuang Lai<sup>1,2,3</sup>, Lingxiao Yang<sup>1</sup>, Xiaohua Xie<sup>1,2,3\*</sup>

<sup>1</sup>School of Computer Science and Engineering, Sun Yat-sen University, China

<sup>2</sup>Guangdong Province Key Laboratory of Information Security Technology, China

<sup>3</sup>Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China

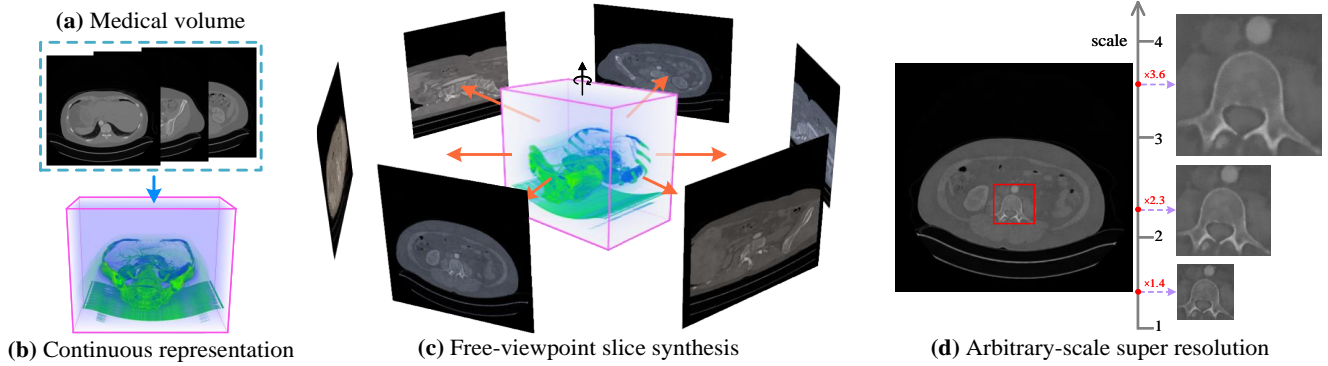


Figure 1. *CuNeRF* is a zero-shot Medical Image Arbitrary-Scale Super Resolution framework. Our model is trained on medical volumes (a) to generate (blue arrow) coordinate-intensity continuous representations (b). A single model can achieve free-viewpoint slice synthesis (c), e.g., simulating a rotating camera 360 degrees around the  $y$ -axis (black arrow) to yield (orange arrow) the corresponding medical images. Moreover, *CuNeRF* can handle arbitrary upsampling scales in a continuous domain (d).

## Abstract

Medical image arbitrary-scale super-resolution (MI-ASSR) has recently gained widespread attention, aiming to super sample medical volumes at arbitrary scales via a single model. However, existing MIASSR methods face two major limitations: (i) reliance on high-resolution (HR) volumes and (ii) limited generalization ability, which restricts their application in various scenarios. To overcome these limitations, we propose Cube-based Neural Radiance Field (*CuNeRF*), a zero-shot MIASSR framework that can yield medical images at arbitrary scales and viewpoints in a continuous domain. Unlike existing MIASSR methods that fit the mapping between low-resolution (LR) and HR volumes, *CuNeRF* focuses on building a coordinate-intensity continuous representation from LR volumes without the need for HR references. This is achieved by the proposed differentiable modules: including cube-based sampling, isotropic volume rendering, and cube-based hierarchical rendering. Through extensive experiments on magnetic resource imaging (MRI) and computed tomography (CT) modalities, we demonstrate that *CuNeRF* outperforms state-of-the-art MI-ASSR methods. *CuNeRF* yields better visual verisimilitude

and reduces aliasing artifacts at various upsampling factors. Moreover, our *CuNeRF* does not need any LR-HR training pairs, which is more flexible and easier to be used than others. Our code will be publicly available soon.

## 1. Introduction

Medical imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) are critical tools in assisting clinical diagnosis by providing visual information. However, obtaining medical slices at desired views and scales is a resource-intensive process that demands patients to expose themselves to considerable ionizing radiation while remaining immobile for a long time [19]. To reduce the burden on patients, a feasible approach is to reconstruct high-quality medical volumes from low-resolution (LR) medical ones.

Mainstream medical image processing methods aim to reconstruct high-resolution (HR) medical images from LR ones via medical image super-resolution (MISR) techniques. Early studies employed optimization methods [11, 35] and interpolation methods [16]. With the advent of deep learning, a series of methods [7, 39, 42, 5, 31] have

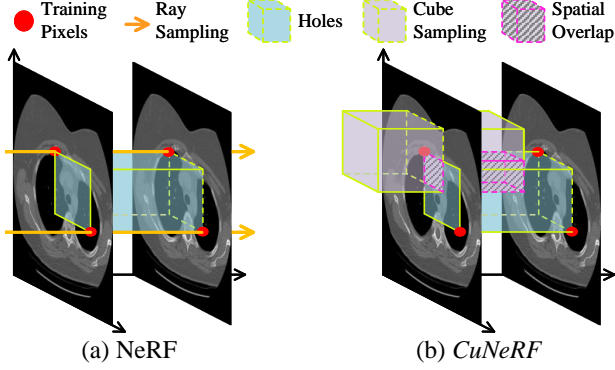


Figure 2. Visualization of the modeling forms between NeRF [21] (a) and *CuNeRF* (b). Visually, NeRF *only* models the ray corresponding to each training pixel (red circle), which cannot cover the whole representation ranges, leaving some “holes” (i.e., unmodeled space within (blue cube) between adjacent training pixels. To circumvent this issue, *CuNeRF* samples a cube (purple cube) centered by each training pixel, and therefore the “holes” are well-filled by the spatial overlaps (grey cube).

adopted convolutional neural networks to learn the mapping between LR and HR domains.

Recently, medical image arbitrary-scale super-resolution (MIASSR) methods [25, 43, 32, 37] have received widespread attention in the field of MISR. These methods aim to employ a single model to upsample medical volumes at arbitrary scales. Although these MIASSR methods achieve reasonable HR results, they still have two major issues: (i) Existing MIASSR methods rely on the supervision from HR volumes, yet high-quality HR volumes are not always available; (ii) These methods may be susceptible to the distribution gap between training and test data, leading to the produce non-existent details.

To address the above-mentioned limitations, we present a zero-shot MIASSR framework, which is trained on LR medical volumes without any supervision from HR ones. Specifically, we draw inspiration from the neural radiance field (NeRF) [21] to estimate the continuous representation from discrete samples (LR volumes) instead of fitting the mapping between LR and HR volumes. However, NeRF’s modeling strategy may not be suitable for direct application on medical volumes. As depicted in Figure 2 (a), due to the difference between medical imaging systems and conventional cameras, NeRF may remain some “holes” (i.e., unmodeled space) in the continuous fields, leading to the synthesis of grid-like artifacts (see Figure 3). Detailed explanation is provided in Section 4.1.

In this paper, we present Cube-based NeRF (*CuNeRF*), a zero-shot solution for MIASSR challenges. Specifically, we tackle NeRF’s hole-forming issues via the proposed modules: cube-based sampling, isotropic volume rendering, and cube-based hierarchical rendering. Since the modules are

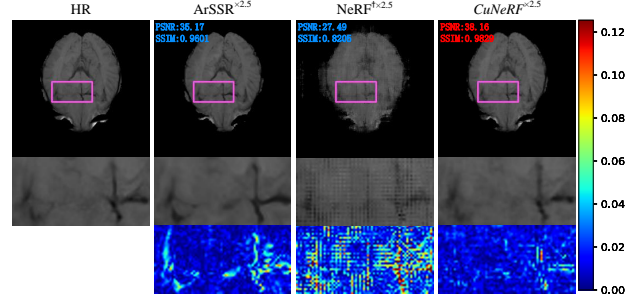


Figure 3. Visual examples of 3D MISR results at  $\times 2.5$  scale between ArSSR [37], SAINT [25], NeRF<sup>†</sup> [21] and the proposed *CuNeRF* on T1-weighted brain tumor MRI volumes selected from MSD [30] dataset. The heatmaps at the bottom are provided to clearly visualize the difference related to the HR image. PSNR and SSIM for each image are inserted, red text indicates the best one. Visually, NeRF<sup>†</sup> renders blurry results with severe grid-like artifacts, while ArSSR produces non-existent details (zooming in purple box). By contrast, *CuNeRF* renders high-quality medical images, achieving better visual verisimilitude to the HR images and reducing objectionable artifacts.

differentiable, *CuNeRF* can build a continuous mapping between the coordinate and the corresponding intensity value in the training data, which is capable of generating medical slices at arbitrary scales and viewpoints in a continuous domain, as shown in Figure 1. Comprehensive experiments on the MSD Brain Tumour (MRI) [30] and KiTS19 (CT) [12] datasets show that *CuNeRF* surpasses state-of-the-art MIASSR methods in 3D MISR and achieves impressive performance in anisotropic MISR challenges. Figure 3 shows the proposed *CuNeRF* outperforms other methods in visual verisimilitude. The main contributions are summarized:

- To the best of our knowledge, *CuNeRF* is the first zero-shot MIASSR framework that can continuously upsample medical volumes at arbitrary scales.
- We address the hole-forming issues via the proposed techniques: cube-based sampling, isotropic volume rendering, and cube-based hierarchical rendering.
- Extensive experiments on CT and MRI modalities for 3D MISR and anisotropic MISR show *CuNeRF* favorably surpasses state-of-the-art MIASSR methods.

## 2. Related Works

In this section, we first review implicit neural representation and subsequently introduce some impressive progress in medical image super-resolution. Recent surveys [34, 17] provide a comprehensive literature review of deep learning-based super-resolution methods.

## 2.1. Implicit Neural Representation

Learning implicit neural representations (INRs) from discrete samples to form a continuous function has been a long-standing research problem in computer vision for numerous tasks. A recent trend in this field is to map discrete representations to coordinate-based continuous neural representations through implicit functions formed by neural networks, such as multi-layer perceptron (MLP). Chen *et al.* [4] proposed a method to learn the INR of 2D images using the local implicit image function. Subsequent work [6] extended this [4] to apply in the video domain. Currently, most 3D view synthesis methods are based on the neural radiance fields (NeRF) [21] framework. NeRF can model a volumetric radiance field to render novel views with impressive visual quality using standard volumetric rendering [13] and alpha compositing techniques [26]. However, NeRF has a drawback of requiring massive training views and lengthy optimization iterations to learn the correct 3D geometry. Several follow-up works have attempted to optimize NeRF's training procedures, such as reducing the required training views [40, 9], accelerating convergence and rendering speed [22], and adapting NeRF to various domains, such as generative modeling [28, 23], anti-aliasing [2], unbounded representation [3], and RGB-D scene synthesis [1]. Recently, some researchers have attempted to employ INR-based methods to reconstruct high-quality medical images in both  $k$ -space [29, 41, 36] and image space [8]. [29, 41, 36] aim to synthesize high-quality sinograms to reconstruct sparse sampled CT from sparse views to dense views. [8] employs [28] to synthesize novel-view X-rays with training on multi-view X-rays.

## 2.2. Medical Image Super Resolution

Medical image super resolution (MISR) is an important task in medical image processing, which aims to reconstruct high-resolution (HR) medical slices from corresponding low-resolution (LR) ones. Initially, some conventional methods like [11, 35] and widely-used interpolation methods like bicubic and tricubic interpolations [16] were employed in early researches. Inspired by [10], recent studies have shifted their focus towards using deep learning-based super-resolution networks in the medical domain. Lim *et al.* [18] employ deep learning-based super-resolution networks to upsample medical images. Some studies upsample each 2D LR medical slice to acquire the corresponding HR one, such as [7, 39, 42]. On the other hand, Chen *et al.* [5] and Wang *et al.* [31] use 3D DenseNet-based networks to generate HR volumetric patches from LR ones. Recent researches have been focusing on medical image arbitrary-scale super-resolution (MIASSR), which aims to upsample medical slices at arbitrary scales by a single model. Peng *et al.* [25] introduce SAINT, a method that deals with anisotropic super resolution on the  $z$ -axis at integer scales.

Wu *et al.* [37] propose ArSSR, an INR-based method that can upsample MRI volumes at arbitrary scales in a continuous domain. Additionally, Wang *et al.* [32] propose a weakly-supervised framework that uses unpaired LR and HR medical volumes for supervision. As discussed in Section 1, MIASSR methods may be susceptible to the distribution gap between training and test data. Moreover, the availability of high-quality HR medical volumes can limit the applicability of these models in certain scenarios.

## 3. Preliminary: NeRF

Neural radiance field (NeRF) [21] aims to build the continuous mapping from  $(\mathbf{x}, \mathbf{d})$  to  $(\mathbf{c}, \sigma)$ , where  $\mathbf{x} = (x, y, z)$  and  $\mathbf{d}$  denote location and viewing direction, while  $\mathbf{c}$  and  $\sigma$  represent content color and volume density, respectively. NeRF's techniques can be summarized as follow:

**Ray sampling.** NeRF first constructs the ray  $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$  that emits from the center of projection  $\mathbf{o}$  and passes through the pixel along the viewing direction  $\mathbf{d}$ . Subsequently, NeRF samples  $N$  points along the ray from near plane  $t_n$  to far plane  $t_f$  predefined. For each sampling point  $\mathbf{r}(t_k)$ , NeRF employs a positional encoding function  $\gamma(\cdot)$  to map the location  $\mathbf{x}_k$  and view direction  $\mathbf{d}$  into higher dimensional space as:

$$\gamma(\rho) = \rho \bigcup_{i=0}^{L-1} (\sin(2^i \rho), \cos(2^i \rho)), \text{ where } L \in \mathbb{N}. \quad (1)$$

where  $\rho$  denotes an arbitrary vector and  $L$  is a hyperparameter set to 10 as default.

**Volume rendering.** The pixel color  $\mathbf{C}(\mathbf{r})$  can be modeled as the integral of the corresponding ray  $\mathbf{r}$  as:

$$\mathbf{C}(\mathbf{r}) = \int_{t_n}^{t_f} \frac{\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt}{\exp(\int_{t_n}^t \sigma(\mathbf{r}(s))ds)}, \quad (2)$$

where  $\mathbf{c}(\cdot)$  and  $\sigma(\cdot)$  denote the color and volume density functions. In practice, NeRF employs a multi-layer perceptron (MLP)  $F_\Theta$  to estimate these two functions. For each sampling point  $\mathbf{r}(t_k)$ , MLP  $F_\Theta$  predicts the corresponding color  $\mathbf{c}_k$  and volume density  $\sigma_k$  by:

$$(\mathbf{c}_k, \sigma_k) = F_\Theta(\gamma(\mathbf{x}_k), \gamma(\mathbf{d})). \quad (3)$$

Given the estimated results of the  $N$  sampling points from  $t_n$  to  $t_f$ , we can approximate the volume rendering integral using numerical quadrature as introduced by [20]:

$$\hat{\mathbf{C}}(\mathbf{r}) = \sum_{i=1}^N \frac{1 - \exp(-\sigma_i(t_{i+1} - t_i))}{\exp(\sum_{j=1}^i \sigma_j(t_{j+1} - t_j))} \mathbf{c}_i, \quad (4)$$

where  $\hat{\mathbf{C}}(\mathbf{r})$  is the predicted color of the pixel.

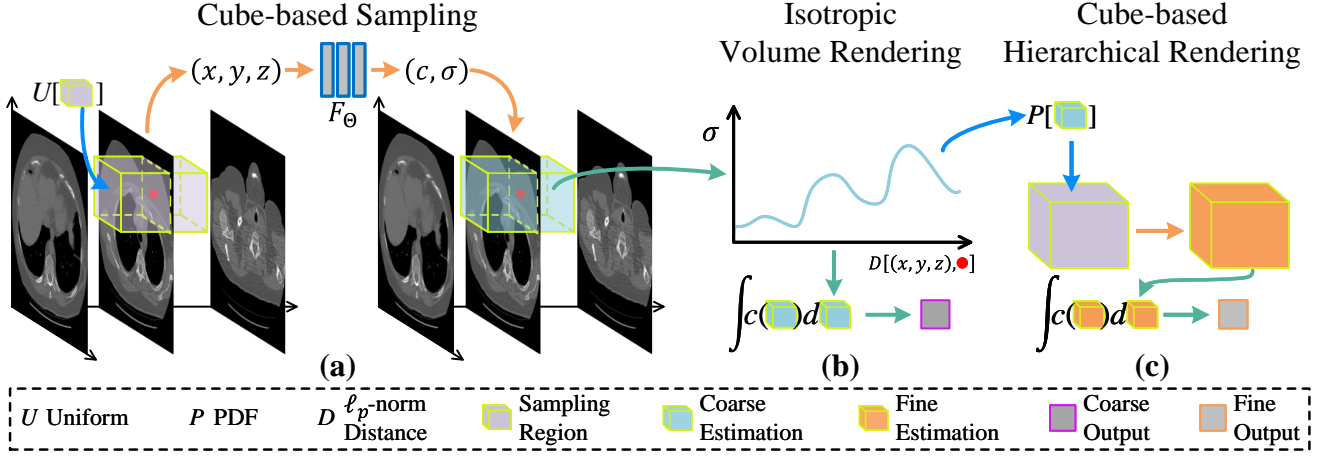


Figure 4. The overall framework of the proposed *CuNeRF*. To synthesize a pixel (red circle), (a) *CuNeRF* first uniformly samples the points  $(x, y, z)$  within the cube space (purple cube). Then, *CuNeRF* obtains the coarse estimation (blue cube) by feeding the sampling points into an MLP  $F_\theta$  to produce the corresponding pixel intensity  $c$  and volume density  $\sigma$ . (b) Subsequently, assuming  $\sigma$  of each sampling point is only related to the distance with the target pixel, *CuNeRF* computes the coarse output of the target pixel via volume integral. (c) Finally, *CuNeRF* resamples the points under the PDF of coarse estimation to acquire the fine estimation (orange cube) of the cube. The fine output is generated by the same procedures of (b). These two rendering functions are differentiable, and thus *CuNeRF* can be optimized by minimizing the rendering loss in Eq 13. We employ the fine output as the result of the target pixel.

**Hierarchical volume rendering.** NeRF also renders the scenes to refine the result by allocating samples proportionally to their expected volume distribution. NeRF simultaneously optimizes two MLPs, *i.e.*, the coarse one  $F_\theta^c$  and the fine one  $F_\theta^f$ . NeRF first samples  $N_c$  points and obtain the coarse output  $\hat{C}_c(\mathbf{r})$  by Eq 4, which can be rewritten as  $\hat{C}_c(\mathbf{r}) = \sum_{i=1}^{N_c} w_i c_i$ . A piecewise-constant PDF related to the sampling points along the ray can be produced by  $\hat{w} = w_i / \sum_{j=1}^{N_c} w_j$ . NeRF then samples  $N_f$  points from this distribution by inverse transform sampling (ITS) and computes the fine outputs  $\hat{C}_f(\mathbf{r})$  using all  $N_c + N_f$  sorted sampling points. Let  $\mathcal{R}$  represent the batch, and these two MLPs can be optimized by the following rendering loss:

$$\mathcal{L} = \sum_{\mathbf{r} \in \mathcal{R}} \left[ \|g.t. - \hat{C}_c(\mathbf{r})\|_2^2 + \|g.t. - \hat{C}_f(\mathbf{r})\|_2^2 \right], \quad (5)$$

where *g.t.* denotes the ground truth of the rendering pixels.

## 4. Method

In this section, we first explain our motivation, analyzing the limitations of NeRF for medical volumes. Subsequently, we propose cube-based NeRF (*CuNeRF*), a novel yet efficient method for “zero-shot” medical image arbitrary-scale super resolution (MIASSR), extending NeRF’s application scenarios in the medical domain. Specifically, we first normalizes the medical volumes by volumetric normalization, and then train the model via proposed differentiable modules: cube-based sampling, isotropic volume rendering, and cube-based hierarchical rendering. During training, *CuNeRF* is building a coordinate-intensity continuous

function whose input is a 3D location  $\mathbf{x} = (x, y, z)$  and output is a pixel value  $c$ . After optimization, *CuNeRF* can predict pixels at any position within the range. As a result, *CuNeRF* achieves free-viewpoint and arbitrary-scale rendering, which can render desired medical slices (desired viewpoints and scales) by feeding the corresponding plane equations. Figure 4 depicts the framework of *CuNeRF*, and the subsequent techniques are described in the following.

### 4.1. Motivation

NeRF’s modeling strategy may not be suitable for direct application on medical volumes. To explain this limitation, we provide an example of NeRF’s modeling strategies applied to medical volumes in Figure 2 (a). Visually, NeRF is trained to model the volumetric space along the ray emitted by each training pixel. Medical volumes only contain three orthogonal slices, which differs from multi-view photos. Hence, NeRF’s modeling techniques cannot cover the entire fields, leaving some “holes” (*i.e.*, unmodeled space) between adjacent training pixels, as depicted in Figure 2 (a). Consequently, NeRF may produce sub-optimal results while rendering the contents within the holes. As shown in Figure 3, NeRF<sup>†1</sup> struggles to deal with 3D MISR issues, producing grid-like artifacts. Consequently, NeRF may struggle to represent high-quality HR medical volumes at arbitrary upsampled scales.

To address the hole-forming issues in NeRF, we introduce cube-based sampling, which samples cubes (3D volumetric space) instead of rays (1D space) to fill the holes

<sup>1</sup>We use the MLP depicted in Figure 5 to adapt volume data.



between adjacent training pixels by the spatial overlaps, as demonstrated in Figure 2 (b). To adapt cube-based sampling, we further propose isotropic volume rendering and cube-based hierarchical rendering modules. These modules will be introduced in the following subsections.

## 4.2. Volumetric Normalization

To build the continuous representations for the given medical volumes, we first normalize the whole volumetric space  $H \times W \times L$  into an  $\ell_\infty$ -norm open ball as:

$$\mathcal{B}(\hat{\mathbf{x}}_o, \pi) = \{\hat{\mathbf{x}} : \|\hat{\mathbf{x}} - \hat{\mathbf{x}}_o\|_\infty < \pi\}, \quad (6)$$

where  $\hat{\mathbf{x}}_o$  is set to  $(0, 0, 0)$  related to the center  $\mathbf{x}_o = (\frac{H}{2}, \frac{W}{2}, \frac{L}{2})$  of the medical volume. To adapt the positional encoding  $\gamma(\cdot)$  introduced in Eq 1, each positional coordinate  $\mathbf{x}_t = (x_t, y_t, z_t)$  within the medical volume is transformed into the field coordinate  $\hat{\mathbf{x}}_t = (\hat{x}_t, \hat{y}_t, \hat{z}_t)$ . The normalization function  $\mathcal{N}(\cdot)$  is formulated as:

$$\hat{\mathbf{x}}_t = \left( \frac{2\pi(x_t - \frac{H}{2})}{H + 2P}, \frac{2\pi(y_t - \frac{W}{2})}{W + 2P}, \frac{2\pi(z_t - \frac{L}{2})}{L + 2P} \right), \quad (7)$$

where  $P$  is a hyperparameter as the padding size.

## 4.3. Cube-based Sampling

Neural representation methods aim to build the continuous representation of medical volumes. However, NeRF suffers from hole-forming issues, which may leave some holes and thus synthesizes grid-like artifacts. To circumvent the hole forming in representation fields, we propose a novel sampling strategy: cube-based sampling, which samples the points within a cube (3D volumetric space). Specifically, for the pixel location  $\hat{\mathbf{x}}_t$ , *CuNeRF* uniformly samples a set of points within the cube space  $\mathcal{B}(\hat{\mathbf{x}}_t, \frac{l}{2})$ . Each sampling point  $\hat{\mathbf{x}}_i$  is chosen under the uniform distribution  $\mathcal{U}$  by:

$$\hat{\mathbf{x}}_i \sim \mathcal{U} \left[ \mathcal{B}(\hat{\mathbf{x}}_t, \frac{l}{2}) \right], \quad (8)$$

where  $l$  denotes the edge length of the cube. We employ the group of these  $N$  sampling points to approximate the cube space. Due to the spatial overlaps between adjacent cubes, the representation fields can be well-covered by employing the proposed cube-based sampling in optimization. As a result, the representation field can be densely modeled with the same sampling time as NeRF.

## 4.4. Isotropic Volume Rendering

As introduced in 3, the pixel color related to the ray  $\mathbf{r}$  is computed by an integral in Eq 2. Intuitively, the pixel color  $\mathbf{C}(\hat{\mathbf{x}}_t, l)$  related to the cube space  $\mathcal{B}(\hat{\mathbf{x}}_t, \frac{l}{2})$  can be computed by the following triple integral as:

$$\mathbf{C}(\hat{\mathbf{x}}_t, l) = \iiint_{\mathcal{B}(\hat{\mathbf{x}}_t, \frac{l}{2})} \frac{\sigma(\hat{x}, \hat{y}, \hat{z}) \mathbf{c}(\hat{x}, \hat{y}, \hat{z}) d\hat{x} d\hat{y} d\hat{z}}{\exp(\int_{\hat{x}_n}^{\hat{x}} \int_{\hat{y}_n}^{\hat{y}} \int_{\hat{z}_n}^{\hat{z}} \sigma(x, y, z) dx dy dz)}, \quad (9)$$

where  $(\hat{x}_n, \hat{y}_n, \hat{z}_n) = (\hat{x}_t - \frac{l}{2}, \hat{y}_t - \frac{l}{2}, \hat{z}_t - \frac{l}{2})$  denotes the initial location of the triple integral while  $\mathbf{c}(\cdot)$  and  $\sigma(\cdot)$  represent the color and volume density functions. However, since NeRF samples  $N$  points to approximate the volume rendering integral of the ray using numerical quadrature in Eq 4, it is required to sample  $N^3$  points to model the cube with the same density, leading to massive computational costs.

Inspired by CRF [15] that assigns the nearby pixels with similar potentials, we assume the volume density  $\sigma$  of each point  $\hat{\mathbf{x}}$  within the cube  $\mathcal{B}(\hat{\mathbf{x}}_t, \frac{l}{2})$  is only related to the  $\ell_p$ -norm distance  $r = \|\hat{\mathbf{x}} - \hat{\mathbf{x}}_t\|_p$  between the centroid and itself. Hence, the volumetric distribution of the cube is isotropic towards the value of  $r$ . The above triple integral can be converted into the spherical coordinates system and simplified by:

$$\mathbf{C}(\hat{\mathbf{x}}_t, l) = 4\pi \int_0^{\hat{r}} \frac{r^2 \sigma(\hat{\mathbf{x}}_t, r) \mathbf{c}(\hat{\mathbf{x}}_t, r) dr}{\exp(4\pi \int_0^r s^2 \sigma(\hat{\mathbf{x}}_t, s) ds)}, \quad (10)$$

where  $\hat{r} = \|(\frac{l}{2}, \frac{l}{2}, \frac{l}{2})\|_p$  denotes the max distance of  $r$  within the cube. The derivation detail of Eq 10 is shown in the supplementary materials. Given  $N$  sampling points by the proposed cube-based sampling, *CuNeRF* first sorts these points by the distance  $r$ . Subsequently, the integral of the cube is approximated via numerical quadrature rules:

$$\hat{\mathbf{C}}(\hat{\mathbf{x}}_t, l) = 4\pi \sum_{i=1}^N \frac{r_i^2 (1 - \exp(-\sigma_i(r_{i+1} - r_i)))}{\exp(4\pi \sum_{j=1}^i r_j^2 \sigma_j(r_{j+1} - r_j))} \mathbf{c}_i, \quad (11)$$

where  $\hat{\mathbf{C}}(\hat{\mathbf{x}}_t, l)$  denotes the predicted color of  $\hat{\mathbf{x}}_t$ .

## 4.5. Cube-based Hierarchical Rendering

To refine the results, *CuNeRF* allocates sampling points proportionally to their expected volume distribution within the cube. Similar to NeRF, *CuNeRF* also simultaneously optimizes the coarse and fine MLPs. As obtaining the coarse output  $\hat{\mathbf{C}}_c(\hat{\mathbf{x}}_t, l)$ , *CuNeRF* first samples  $N_f$  numbers of  $r$  using ITS. Subsequently, for each  $r$ , we use the hierarchical sampling function  $\zeta_p(\cdot)$  to sample the corresponding sampling point:

$$\hat{\mathbf{x}}_f = \zeta_p(r, \varphi, \theta), \quad (12)$$

where  $\varphi$  and  $\theta$  are the randomly sampled spherical coordinates, and  $\zeta_p(\cdot)$  converts the  $\ell_p$ -norm spherical coordinates  $(r, \varphi, \theta)$  to the Cartesian coordinates  $\hat{\mathbf{x}}$ . If  $p \neq \infty$ , we allow  $\hat{\mathbf{x}}_f$  can beyond the cube space  $\mathcal{B}(\hat{\mathbf{x}}_t, \frac{l}{2})$ . After obtaining fine outputs  $\hat{\mathbf{C}}_f(\hat{\mathbf{x}}_t, l)$  at Eq 11 using the sorted union of  $N_c + N_f$  sampling points, *CuNeRF* can be optimized in the following proposed adaptive rendering loss:

$$\mathcal{L}_A = \sum_{\hat{\mathbf{x}}_t \in \mathcal{R}} \left[ \lambda \|g.t. - \hat{\mathbf{C}}_c(\hat{\mathbf{x}}_t, l)\|_2^2 + \|g.t. - \hat{\mathbf{C}}_f(\hat{\mathbf{x}}_t, l)\|_2^2 \right], \quad (13)$$

where  $\lambda = \|g.t. - \hat{\mathbf{C}}_f(\hat{\mathbf{x}}_t, l)\|^{\frac{1}{2}}$  is an adaptive regularization term to alleviate the overfitting brought by the ‘‘coarse’’ loss while  $\mathcal{R}$  denotes the batch.

Table 1. Quantitative comparisons of start-of-the-arts on brain tumor MRI volumes selected from MSD [30] dataset in terms of PSNR/SSIM metrics for 3D MISR in this table. **Bold** and underline texts indicate the best and second best performance.

	$\times 2$	$\times 2.5$	$\times 3$	$\times 3.5$	$\times 4$	$\times 5$	$\times 6$	$\times 8$
Bicubic	33.75 / 0.9469	30.84 / 0.9271	30.74 / 0.9161	29.19 / 0.8873	28.67 / 0.8721	28.14 / 0.8687	26.83 / 0.8521	26.54 / 0.8376
ArSSR [37]	36.98 / 0.9690	35.24 / 0.9398	34.69 / 0.9199	34.01 / 0.8974	33.21 / 0.8910	30.50 / 0.8624	29.96 / 0.8543	28.43 / 0.8353
NeRF <sup>†</sup> [21]	29.33 / 0.8472	27.03 / 0.8392	25.98 / 0.8220	25.47 / 0.8157	25.12 / 0.8088	24.50 / 0.7767	23.45 / 0.7549	22.63 / 0.7275
<i>CuNeRF</i>	<b>39.62 / 0.9786</b>	<b>37.56 / 0.9441</b>	<b>36.24 / 0.9267</b>	<b>35.81 / 0.9189</b>	<b>35.01 / 0.9031</b>	<b>34.73 / 0.8952</b>	<b>33.69 / 0.8800</b>	<b>31.19 / 0.8675</b>

Table 2. Quantitative comparisons of start-of-the-arts on kidney tumor CT volumes selected from KiTS19 [12] dataset in terms of PSNR/SSIM metrics for anisotropic MISR in this table. **Bold** and underline texts indicate the best and second best performance.

	$\times 2$	$\times 4$	$\times 6$
Bicubic	37.75 / 0.9498	33.76 / 0.9149	31.24 / 0.8967
SAINT [25]	<b>39.47 / 0.9782</b>	<u>36.61 / 0.9574</u>	<b>33.63 / 0.9301</b>
NeRF <sup>†</sup> [21]	36.50 / 0.9383	34.14 / 0.9181	31.18 / 0.9020
<i>CuNeRF</i>	38.33 / 0.9663	<b>36.64 / 0.9480</b>	32.93 / 0.9229

#### 4.6. Free-Viewpoint & Arbitrary-Scale Rendering

After optimization, *CuNeRF* can predict the pixels corresponding to arbitrary 3D coordinates. Therefore, *CuNeRF* can represent medical slices with free viewpoints and arbitrary scales by feeding the corresponding plane coordinates. Specifically, to render a medical slice with the given position  $\hat{\mathbf{x}}$ , viewpoint  $\mathbf{d}$  and scale  $\delta$ , we first construct a base plane at  $\hat{\mathbf{x}}_o$  with the sampling scale  $\delta$ . Subsequently, we employ the translation matrix  $\mathcal{M}_T$  to move slices from  $\hat{\mathbf{x}}_o$  to  $\hat{\mathbf{x}}$ . Finally, since the viewpoint  $\mathbf{d}$  can be formulated as rotating  $\phi$  degrees around a certain axis  $\vec{n}$ , we can obtain the corresponding rotation matrix  $\mathcal{M}_R$  via Rodrigues' rotation formula [27].

We provide some visual examples in the anonymous link. As shown, *CuNeRF* is capable to yield high-quality medical slices at continuous-valued scales from 1.0 to 8.0. Moreover, *CuNeRF* can synthesize medical images with a viewpoint rotating 360 degrees around an arbitrary coordinate axis  $\vec{n}$ . Consequently, compared to existing methods, *CuNeRF* is capable to provide richer visual information.

### 5. Experiments

In this section, we conduct extensive experiments and in-depth analysis to demonstrate the superiority of our *CuNeRF* in representing high-quality slices at arbitrary scales. For fair comparisons, the hyperparameters and model settings are consistent in all experiments.

#### 5.1. Experimental Details

**Datasets.** We comprehensively compare the proposed *CuNeRF* and the existing advances on 80 medical volumes. These medical volume data consist of 2 different modalities: CT and MRI. More specifically, we select 40 T1-weighted brain tumor MRI volumes from the publicly available Medical Segmentation Decathlon (MSD) [30] while we also take

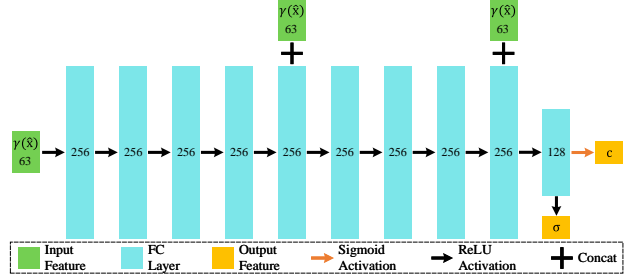


Figure 5. The architecture of our multi-layer perceptron (MLP). For a given field coordinate  $\hat{\mathbf{x}}$ , it is first encoded by the positional encoding  $\gamma(\cdot)$  in Eq 1 as the input feature (green blocks). Subsequently, the input features pass through 9 fully-connected layers (blue blocks), each with 256 channels and the ReLU activation (black arrows), where we concatenate (+) the input features to the 5th and 9th intermediate hidden layers as the skip connections. Finally, we downscale the feature channels from 256 to 128 and obtain the outputs (yellow blocks), which contain volume density  $\sigma$ , and pixel value  $\mathbf{c}$  with sigmoid activations (orange arrow).

40 CT volumes from the 2019 Kidney Tumor Segmentation Challenge (KiTS19) [12] datasets, respectively. All MRI volumes have the same dimension of  $240 \times 240 \times 155$ . The number of CT scan slices along the  $z$ -axis is different, while the dimension of each slice is  $512 \times 512$ . In experiments, all the CT volumes are resized into  $512 \times 512 \times 512$ .

**Multi-Layer Perceptron Architecture.** Due to the vast difference between the medical imaging system and conventional cameras, we further reform NeRF's MLP to adapt to medical volume data. Figure 5 depicts MLP's architecture, where the input is a 3D location  $\hat{\mathbf{x}} = (\hat{x}, \hat{y}, \hat{z})$  and the output is a 2D union of color  $\mathbf{c}$  and volume density  $\sigma$ . Similar to NeRF, we first employ  $\gamma(\cdot)$  to encode the input  $\mathbf{x}$ , and then feed it to the fully-connected network to obtain the output  $(\mathbf{c}, \sigma)$ . Thus MLP is optimized by minimizing  $\mathcal{L}_A$ .

**Implementation Details.** *CuNeRF* is implemented on top of [38], a Pytorch [24] re-implementation of NeRF. Meanwhile, our experiments are also based on Pytorch [24] framework, and run on a single GeForce RTX 3090 GPU with 24G memory. Similar to NeRF, *CuNeRF* first samples 64 points for the coarse model  $F_\Theta^c$  and feeds 192 sampling points (the sorted union of 64 coarse and 128 fine samples) into the fine model  $F_\Theta^f$ . We employ the  $\ell_2$ -norm distance for isotropic volume rendering and hierarchical cubic rendering functions as default while the edge length  $l$  of the cube is set to 1 pixel distance as default. Hence, the hierarchical sam-

Table 3. Quantitative comparisons of *CuNeRF* and its ablation variants on brain tumor MRI volumes selected from MSD [30] dataset in terms of PSNR/SSIM metrics for 3D MISR in this table. **Bold** text indicates the best performance.

<i>CuS</i>	<i>IVR</i>	$\mathcal{L}_A$	$\times 2$	$\times 4$	$\times 8$
			29.33 / 0.8472	25.12 / 0.8088	22.63 / 0.7275
✓			35.82 / 0.9244	33.29 / 0.8703	27.77 / 0.8398
✓	✓		38.15 / 0.9524	34.27 / 0.8887	29.34 / 0.8455
✓	✓	✓	<b>39.62 / 0.9786</b>	<b>35.01 / 0.9031</b>	<b>31.19 / 0.8675</b>

pling function  $\zeta_2(\cdot)$  converts  $(r, \varphi, \theta)$  to  $\hat{\mathbf{x}} = (\hat{x}, \hat{y}, \hat{z})$  by:

$$\hat{\mathbf{x}} = (r \sin \varphi \cos \theta, r \sin \varphi \sin \theta, r \cos \varphi), \quad (14)$$

where  $\varphi \sim \mathcal{U}[0, \pi]$ ,  $\theta \sim \mathcal{U}[0, 2\pi]$ .

For training, we employ Adam [14] as the optimizer with a weight decay of  $10^{-6}$  and a batch size of 2048. The maximum iteration is set to 250,000, and the learning rate is annealed logarithmically from  $2 \times 10^{-3}$  to  $2 \times 10^{-5}$  during the whole training time.

For testing, the results are obtained by feeding all the coordinates of the given plane equations into our model (seeing details in Section 4.6). Note that we do not use any pre-processing and post-processing techniques to improve our results in the following experiments.

**Evaluation Metrics.** We use two quantitative metrics: Peak Signal-to-Noise Ratio (PSNR) and Structured Similarity Index (SSIM) [33] to measure the image quality of different medical image super-resolution methods.

## 5.2. Comparison With State-of-the-art Methods

We compare the proposed *CuNeRF* with 6 state-of-the-art methods, including 2 recent MIASSR methods: ArSSR [37] and SAINT [25], 1 conventional method: bicubic interpolation, and NeRF<sup>†</sup> [21]. As introduced in Section 2.2, SAINT [25] can only upsample LR CT volumes on a single axis. For fair comparisons, as given a upsampling scale  $\delta$ , we evaluate these methods under the following two settings: (i) **3D MISR**. Upsampling the donwsampled volume from  $\frac{H}{\delta} \times \frac{W}{\delta} \times \frac{L}{\delta}$  to  $H \times W \times L$ ; (ii) **Anisotropic MISR**. Upsampling the donwsampled volume from  $H \times W \times \frac{L}{\delta}$  to  $H \times W \times L$ ; Note that NeRF<sup>†</sup> and *CuNeRF* are trained with the same experimental settings, including the same architecture of the MLP  $F_\Theta$  introduced in Section 5.1.

**Quantitative Comparison.** We report 3D MISR and anisotropic MISR based on the increasing upsampled scales in Table 1 and Table 2, respectively. As demonstrated, for the 3D MISR challenge on MRI volumes, *CuNeRF* surpasses all the competitors with consistent preferable performance at various upsampling scales. For anisotropic MISR challenge on CT volumes, *CuNeRF* achieves comparable performance to SAINT [25]. Note that *CuNeRF* is a zero-shot framework that can represent medical volumes in a continuous domain, while SAINT [25] only deals with inte-

Table 4. Quantitative comparisons of *CuNeRF* under different settings on brain tumor MRI volumes selected from MSD [30] dataset in terms of PSNR/SSIM metrics for 3D MISR in this table. **Bold** and underline texts indicate the best and second best performance.

	$\times 2$	$\times 4$	$\times 8$
<i>CuNeRF</i> default.	39.62 / <b>0.9786</b>	<b>35.01 / 0.9031</b>	31.19 / 0.8675
<i>CuNeRF</i> $p = \infty$	35.78 / 0.9348	32.85 / 0.8853	29.25 / 0.8433
<i>CuNeRF</i> $l = 0.5$	38.17 / 0.9621	<u>35.00 / 0.9018</u>	<b>31.53 / 0.8704</b>
<i>CuNeRF</i> $l = 2$	<b>39.65 / 0.9723</b>	<u>34.57 / 0.9011</u>	30.85 / 0.8608

ger factors. Compared to fully-supervised MIASSR methods: ArSSR [37] and SAINT [25], the *CuNeRF* is more robust at presenting large-scale medical slices and capable to deal with different modalities (CT and MRI), suggesting *CuNeRF* owns broader application scenarios. It is also worth noting that NeRF<sup>†</sup> achieves comparable performance for anisotropic MISR but fails in 3D MISR. Because the rendering content of anisotropic MISR is along the ray ( $z$ -axis) while 3D MISR must query the contents within the holes, the experimental results of NeRF<sup>†</sup> confirm the correctness of our motivation.

**Visual Comparison.** We visualize the rendering results of *CuNeRF* and other competitors on MRI (rows 1 and 2) and CT (rows 3 and 4) modalities in Figure 6. It can be observed that *CuNeRF* well represents the medical slices at various scales. Compared to the exhibited methods, *CuNeRF* is most similar to the ground truths, achieving better visual verisimilitude and reducing aliasing artifacts, especially in representing large-scale medical slices. Since NeRF<sup>†</sup> exhibits grid-like artifacts in rendering high-quality medical slices at larger-valued scales, the visualization results prove the effectiveness of *CuNeRF*, which extends NeRF’s capability to build high-quality continuous representation for medical volumes.

## 5.3. Ablation Study

In this subsection, we conduct comprehensive experiments to prove the correctness of *CuNeRF*’s design. We first carry out ablation studies to investigate the effectiveness of the proposed modules. Subsequently, we evaluate the *CuNeRF*’s performance under different settings.

***CuNeRF*’s ablation variants.** We evaluate against several ablations of the proposed *CuNeRF* with each module: *CuS*, *IVR* and  $\mathcal{L}_A$  represent cube-based sampling, isotropic volume rendering, and adaptive rendering loss, respectively. The baseline model here is NeRF<sup>†</sup>. As reported in Table 3, the baseline model struggles to deal with 3D MISR issues (row 1), while adopting *CuS* instead of ray sampling can significantly improve the performance (row 2). Compared to NeRF’s volume rendering function, employing *IVR* (row 3) can further improve the slice synthesis quality, suggesting *IVR* can better estimate the volumetric distribution, reducing aliasing artifacts raised by undersampling. Since the coarse term of NeRF’s rendering loss may affect the opti-

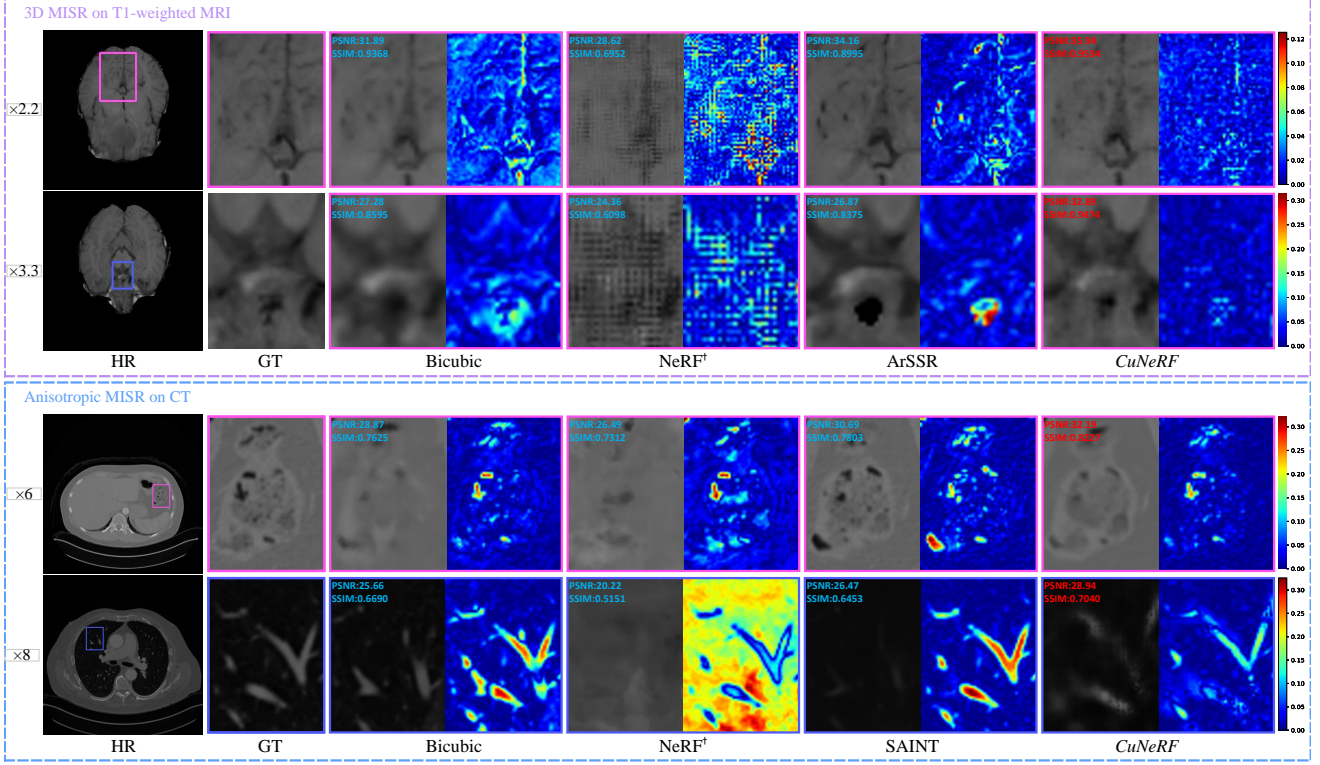


Figure 6. Visual comparisons between *CuNeRF* and state-of-the-art methods for 3D MISR and anisotropic MISR. The heatmaps on the right are provided to clearly visualize the difference related to GT images. For better visualization, the difference maps are provided to the right of the results. PSNR and SSIM for each image are inserted, red text indicates the best one.

mization,  $\mathcal{L}_A$  (row 4) is able to alleviate this distraction for achieving better results.

***CuNeRF* under different settings.** We evaluate the performance of *CuNeRF* under different settings: “ $p = \infty$ ” employs  $\ell_\infty$ -norm distance of  $r$ , “ $l = 0.5$ ” and “ $l = 2$ ” represent setting the edge length of cube to 0.5 and 2 pixel distance, respectively. The default setting of *CuNeRF* is introduced in Section 5.1, where  $p = 2$  and  $l = 1$ .

As reported in Table 4, the default setting of *CuNeRF* achieves consistent outperformance at various scales. In contrast, employing the  $\ell_\infty$ -norm is significantly inferior to the default one, which means  $\ell_2$ -norm distance is more suitable to model the continuous representation for medical volumes. Meanwhile, different values of cube edge  $l$  also acquire comparable performance to the default one, suggesting the proposed *CuNeRF* is a parameter-insensitive method with good robustness under different experimental settings.

## 6. Conclusion

Medical image arbitrary-scale super resolution (MI-ASSR) aims to employ a single well-trained model to up-sample medical slices at arbitrary scales. However, there are two major limitations of existing MIASSR methods: (i) relying on HR volumes; (ii) limited generalization abil-

ity, limiting the model application scenarios. In this paper, we present *CuNeRF*, a zero-shot MIASSR framework that can yield medical images at arbitrary viewpoints and scales in a continuous domain. Instead of learning the mapping between low-resolution (LR) and high-resolution (HR) volumes, *CuNeRF* models the continuous representation from LR volumes via the proposed differentiable modules: cube-based sampling, isotropic volume rendering, and cube-based hierarchical rendering. Extensive experiments on magnetic resource imaging (MRI) and computed tomography (CT) modalities demonstrate that *CuNeRF* outperforms state-of-the-art MIASSR methods, yielding better visual synthesis effects and reducing aliasing artifacts at various upsampling factors.

## Acknowledgement

This project is supported by the Natural Science Foundation of China (No. 62072482).

## References

- [1] Dejan Azinović, Ricardo Martin-Brualla, Dan B Goldman, Matthias Nießner, and Justus Thies. Neural rgb-d surface reconstruction. In *Proceedings of the IEEE/CVF Conference*



- on *Computer Vision and Pattern Recognition*, pages 6290–6301, 2022. 3
- [2] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. *ICCV*, 2021. 3
  - [3] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 3
  - [4] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8628–8638, 2021. 3
  - [5] Yuhua Chen, Feng Shi, Anthony G Christodoulou, Yibin Xie, Zhengwei Zhou, and Debiao Li. Efficient and accurate mri super-resolution using a generative adversarial network and 3d multi-level densely connected network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 91–99. Springer, 2018. 1, 3
  - [6] Zeyuan Chen, Yinbo Chen, Jingwen Liu, Xingqian Xu, Vidit Goel, Zhangyang Wang, Humphrey Shi, and Xiaolong Wang. Videoinr: Learning video implicit neural representation for continuous space-time super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2047–2057, 2022. 3
  - [7] Venkateswararao Cherukuri, Tiantong Guo, Steven J Schiff, and Vishal Monga. Deep mr brain image super-resolution using spatio-structural priors. *IEEE Transactions on Image Processing*, 29:1368–1383, 2019. 1, 3
  - [8] Abril Corona-Figueroa, Jonathan Frawley, Sam Bond-Taylor, Sarath Bethapudi, Hubert P. H. Shum, and Chris G. Willcocks. Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray, 2022. 3
  - [9] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised nerf: Fewer views and faster training for free. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12882–12891, 2022. 3
  - [10] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014. 3
  - [11] Ali Gholipour, Judy A Estroff, and Simon K Warfield. Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain mri. *IEEE transactions on medical imaging*, 29(10):1739–1758, 2010. 1, 3
  - [12] Nicholas Heller, Niranjana Sathianathan, Arveen Kalapara, Edward Walczak, Keenan Moore, Heather Kaluzniak, Joel Rosenberg, Paul Blake, Zachary Rengel, Makinna Oestreich, et al. The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. *arXiv preprint arXiv:1904.00445*, 2019. 2, 6
  - [13] James T Kajiya and Brian P Von Herzen. Ray tracing volume densities. *ACM SIGGRAPH computer graphics*, 18(3):165–174, 1984. 3
  - [14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7
  - [15] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems*, 24, 2011. 5
  - [16] Francois Lekien and J Marsden. Tricubic interpolation in three dimensions. *International Journal for Numerical Methods in Engineering*, 63(3):455–471, 2005. 1, 3
  - [17] Y Li, Bruno Sixou, and F Peyrin. A review of the deep learning methods for medical images super resolution problems. *Irbm*, 42(2):120–133, 2021. 2
  - [18] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 3
  - [19] Pechin Lo, Bram Van Ginneken, Joseph M Reinhardt, Tarunashree Yavarna, Pim A De Jong, Benjamin Irving, Catalin Fetita, Margarete Ortner, Rômulo Pinho, Jan Sijbers, et al. Extraction of airways from ct (exact’09). *IEEE Transactions on Medical Imaging*, 31(11):2093–2107, 2012. 1
  - [20] Nelson Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, 1995. 3
  - [21] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, pages 405–421. Springer, 2020. 2, 3, 6, 7
  - [22] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *arXiv preprint arXiv:2201.05989*, 2022. 3
  - [23] Michael Niemeyer and Andreas Geiger. Giraffe: Representing scenes as compositional generative neural feature fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11453–11464, 2021. 3
  - [24] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *Proceedings of Neural Information Processing Systems (NIPS)*, 2017. 6
  - [25] Cheng Peng, Wei-An Lin, HaoFu Liao, Rama Chellappa, and S. Kevin Zhou. Saint: Spatially aware interpolation network for medical slice synthesis. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2, 3, 6, 7
  - [26] Thomas Porter and Tom Duff. Compositing digital images. In *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, pages 253–259, 1984. 3
  - [27] Olinde Rodrigues. Des lois géométriques qui régissent les déplacements d’un système solide dans l’espace, et de la

variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. *J. Math. Pures Appl*, 5(380-400):5, 1840. 6

- [28] Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger. Graf: Generative radiance fields for 3d-aware image synthesis. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 20154–20166. Curran Associates, Inc., 2020. 3
- [29] Liyue Shen, John Pauly, and Lei Xing. Nerp: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. 3
- [30] Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019. 2, 6, 7
- [31] Jiancong Wang, Yuhua Chen, Yifan Wu, Jianbo Shi, and James Gee. Enhanced generative adversarial network for 3d brain mri super-resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3627–3636, 2020. 1, 3
- [32] Jiale Wang, Runze Wang, Rong Tao, and Guoyan Zheng. Uassr: Unsupervised arbitrary scale super-resolution reconstruction of single anisotropic 3d images via disentangled representation learning. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, pages 453–462. Springer, 2022. 2, 3
- [33] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 7
- [34] Zhihao Wang, Jian Chen, and Steven CH Hoi. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3365–3387, 2020. 2
- [35] Stefan Wesarg et al. Combining short-axis and long-axis cardiac mr images by applying a super-resolution reconstruction algorithm. In *Medical Imaging 2010: Image Processing*, volume 7623, page 76230I. International Society for Optics and Photonics, 2010. 1, 3
- [36] Qing Wu, Ruimin Feng, Hongjiang Wei, Jingyi Yu, and Yuyao Zhang. Self-supervised coordinate projection network for sparse-view computed tomography. *arXiv preprint arXiv:2209.05483*, 2022. 3
- [37] Qing Wu, Yuwei Li, Yawen Sun, Yan Zhou, Hongjiang Wei, Jingyi Yu, and Yuyao Zhang. An arbitrary scale super-resolution approach for 3d mr images via implicit neural representation. *IEEE Journal of Biomedical and Health Informatics*, 27(2):1004–1015, 2023. 2, 3, 6, 7
- [38] Lin Yen-Chen. Nerf-pytorch. <https://github.com/yenchenlin/nerf-pytorch/>, 2020. 6
- [39] Chenyu You, Guang Li, Yi Zhang, Xiaoliu Zhang, Hongming Shan, Mengzhou Li, Shenghong Ju, Zhen Zhao, Zhuiyang Zhang, Wenxiang Cong, et al. Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *IEEE transactions on medical imaging*, 39(1):188–203, 2019. 1, 3
- [40] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4578–4587, 2021. 3
- [41] Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1960–1970, 2021. 3
- [42] Xiaole Zhao, Yulun Zhang, Tao Zhang, and Xueming Zou. Channel splitting network for single mr image super-resolution. *IEEE Transactions on Image Processing*, 28(11):5649–5662, 2019. 1, 3
- [43] Jin Zhu, Chuan Tan, Junwei Yang, Guang Yang, and Pietro Lio'. Arbitrary scale super-resolution for medical images. *International Journal of Neural Systems*, 31(10):2150037, 2021. 2