Article

A Novel Implicit Neural Representation for Volume Data

Armin Sheibanifard *,† 🅾 **and Hongchuan Yu** †

National Center for Computer Animation, Faculty of Media and Communication, Bournemouth University, Bournemouth BH12 5BB, UK
* Correspondence: asheibanifard@bournemouth.ac.uk
† These authors contributed equally to this work.

Abstract: The storage of medical images is one of the challenges in the medical imaging field. There are variable works that use implicit neural representation (INR) to compress volumetric medical images. However, there is room to improve the compression rate for volumetric medical images. Most of the INR techniques need a huge amount of GPU memory and a long training time for high-quality medical volume rendering. In this paper, we present a novel implicit neural representation to compress volume data using our proposed architecture, that is, the Lanczos downsampling scheme, SIREN deep network, and SRDenseNet high-resolution scheme. Our architecture can effectively reduce training time, and gain a high compression rate while retaining the final rendering quality. Moreover, it can save GPU memory in comparison with the existing works. The experiments show that the quality of reconstructed images and training speed using our architecture is higher than current works which use the SIREN only. Besides, the GPU memory cost is evidently decreased.

1. Introduction

Contemporary microscopy technology is widely applied to biology, neuroscience, and medical imaging fields, which bring about large-scale multidimensional datasets and require terabytes or petabytes of data storage. Besides, recent developments in radiological hardware have increased the capabilities of medical imaging to produce high-resolution and 3D images that require high storage capacity. It is desired to manipulate multidimensional datasets for visualization and predictions from multi-channel image intensity values in a 2D or 3D scan. Such datasets pose significant challenges related to storage, manipulation, and rendering. In this paper, we suggest a novel implicit neural representation to compress high-resolution medical volume data using the implicit neural representation (INR) network with high speed and quality in comparison with current works.

There are some works using INR with periodic activation functions (SIREN) [1] to compress the medical images [2,3]. In order to compress the images, these techniques train a neural network using the voxel coordinates. The resulting trained networks represent the individual image stacks. A distinct advantage is that the size of the trained deep network is always less than that of the image stack. However, they also suffer some considerable issues. They usually depend on a huge GPU memory, specifically for high-resolution volume data. Thus, it is impossible for clinicians to access medical volume data based on usual workstations and laptops. On the other hand, in order to deal with high-resolution (HR) image stacks, it is natural to use large-scale neural networks with multiple layers, which results in a long training time and a low rate of compression. Obviously, the main challenge is to both keep a high rate of compression to decrease deep network size and keep the quality of reconstructed images.

To tackle this challenge, we present an architecture combining the SIREN to compress the volume data effectively and a super-resolution module to keep the quality of the reconstruction. Our architecture includes a downsampling module to decrease the resolution

of high-resolution (HR) and low-resolution (LR) images, which can fit the downsampled volume data in usual GPUs with low graphical memory. As a result, the SIREN's layers can be reduced. Training can be accelerated accordingly. The following super-resolution (SR) reconstruction network can recover images up to the quality of the original HR.

In summary, our contributions include:

- Our architecture consists of the Lanczos downsampling scheme, SIREN deep network, and SRDenseNet upsampling scheme, which increase the speed of training and decrease the demand for GPU memory in comparison with existing INR-based compression techniques;
- Our architecture can reach both a high compression rate and high quality of the final volume data rendering.

The following part of this paper is organized as follows. Section 2 presents the related work. Section 3 illustrates the methodology. Section 4 shows the experiments and results from the analysis. Section 5 draws our conclusions and presents future works.

2. Related Work

Recently, implicit neural representation (INR) is considered by variable works to represent medical imaging. Three-dimensional medical imaging is usually considered as a discrete grid of voxels. Although voxel-based techniques are simple and regular to use, they are used for small voxel grids. Some works try to overcome this limitation and increase the size of voxel grids, using shallow networks with a small batch size [4] which increases the time of the training.

2.1. Implicit Neural Representation

As the memory consumption of discrete voxel grids increases cubically, the Ref. [5] suggests an implicit organ segmentation network, using continuous implicit neural representations. As the IOSNET is a continuous function, it is completely independent of spatial resolution. Because of its continuity, high-resolution medical images can be processed quickly using IOSNET. The Ref. [6] suggests DeepSDF as a fully continuous and implicit representation for generative 3D modelling. DeepSDF is based on Sign Distance Function, but it uses a generative model to produce 3D continuous interfaces. DeepSDF reduces memory consumption in comparison with its counterparts. The suggested auto-encoder in DeepSDF needs explicit optimization which consumes more time during inference. In the Ref. [7] they introduced an encoder–decoder neural architecture to losslessly compress truncated signs in distance fields (TSDF) in a 3D voxel grid. Their deep network architecture is block-based that is trained end-to-end. The main limitation of their model architecture is that the blocks are independent and identically distributed. Ignoring this limitation may help to increase the rate of compression. A lossless compression technique (MedZip) using Long Short-Term Memory (LSTM) was introduced by the Ref. [8]. Their work predicts the next intensity value using LSTM in a set of the voxels' neighbourhood concept. MedZip is the first lossless compression technique that uses LSTM for volumetric MRI and CT.

In the Ref. [9] the authors suggested a technique to synthesize new views of a volumetric scene using implicit neural representation as a continuous function. Their technique represents a volume using a deep fully connected network with five inputs and four outputs. The network encodes the spatial location and direction to RGB and opacity. NeRF is time-consuming and its capability in complex images is not considerable. Following this work, the Ref. [1] demonstrates that using periodic activation functions outperforms ReLU-MLPs. They also suggest sinusoidal representation networks (SIRENs) that use periodic activation functions in implicit neural representation. This technique fits complex signals and natural images. In the Ref. [10] a 3D representation technique is presented to reduce the memory footprint. In their design, they use a neural network to predict the occupancy function to obtain a continuous volume. The Ref. [11] presents a compression technique using an implicit neural network (COIN) to compress natural images. It uses multi-layer perceptron (MLP) to encode geometric inputs. COIN's results are weaker than

state-of-the-art compression techniques. The Ref. [12] suggests an adversarial generation of continuous images using INRs. Their model is constituted of two techniques, multi-scale INRs and factorized multiplicative modulation (INR-GAN). With this architecture, it is possible to represent high-resolution images. Because high sensitivity of INR-based techniques to high-frequency features, it leads to cause artefacts. The Ref. [13] introduces an implicit neural representation technique with prior embedding to reconstruct sampled medical images [14] suggests a generative network based on implicit neural network representation to reconstruct new viewpoints by learning from different viewpoints from input images taken. The Ref. [15] suggests using Fourier features in MLP. As a standard MLP is slow in convergence, they suggest passing the inputs of the network through a simple Fourier feature mapping. By this technique, the network can learn high-frequency functions and improve the MLPs' performance.

2.2. Deep Neural Network in Medical Image Restoration

In the Ref. [16] the authors suggest a technique to use transfer learning to reconstruct the accelerated MR images. This work shows the capability of transfer learning for sparse training data. To accelerate dynamic MRI, the authors in the Ref. [17] used a convolutional recurrent neural network (CRNN-MRI). Through this method they reconstructed high-resolution MR image sequences. To accelerate MR image acquisition, the Ref. [18] introduced a self-attention CNN architecture to reconstruct MRI. Image quality can be easily degraded by noises and artefacts in low-dose computed tomography (LDCT). To address this issue, the Ref. [19] proposes a deep iterative reconstruction estimation (DIRE) with a 3D residual convolutional network (ResNet) architecture to improve the quality of images. Researchers in the Ref. [20] proposed a simultaneous algebraic reconstruction technique (SART) to reconstruct images for translational CT. Then they used a pre-trained CNN to remove artefacts and noise. To reduce the radiation dose, the Ref. [21] suggests a deep encoder–decoder adversarial reconstruction (DEAR) network to reconstruct 3D CT images directly from real clinical cone beam image data. To extract 3D details from generated slices with an adversarial network, they use DEAR-3D which is based on 3D convolutional layers. In low-dose CT imaging, noise and artefacts are inevitable. To address that issue, the Ref. [22] proposes an improved version of GoogLeNet to remove artefacts that are caused by missing projection during image reconstruction. To reconstruct an X-ray super-resolution image, the Ref. [23] suggests a GANs-based approach. It proposes spectral normalization super-resolution medical images (SNSR-GAN) to reconstruct high-resolution X-ray images. To reconstruct an ultra-fast CT image, the Ref. [24] suggests a multi-receptive field densely connected CNN (MRDC-CNN). They also use dense skip modules instead of simple skip modules to flow the information between the encoder and decoder. To decrease the memory footprint they ignore batch normalization. The Ref. [25] presents a deep regularization method to overcome the reconstruction of photo-acoustic computed tomography (PACT) with sparse view measurements. A non-local deep image before reconstructing the Positron Emission Tomography (PET) images is proposed by the Ref. [26]. They use a prior image of the patients as the input of the network. They suggest the 3D U-Net [27] as the backbone of their network structure. DeepPET is introduced by the Ref. [28] to reconstruct PET images from sinograms. They use a deep encoder–decoder network to reconstruct high-quality images. DeepPET is 108 times faster than its counterparts. DUG-RECON [29] is an Unet-based deep learning pipeline to reconstruct PET and CT images directly. It uses a convolutional generative network with three stages, denoising, image reconstruction and super-resolution segments. To reconstruct dynamic PET images, the Ref. [30] suggests non-negative matrix factorization (NMF) with a deep image prior. They also show the capabilities of DIP for PET image reconstruction. The Ref. [31] suggests a Deep Residual Error Iterative Minimization Network to reconstruct sparse-view CT. They optimize a hand-crafted function to reconstruct high-quality images. In CNN, the size of the convolutional kernel is smaller than the image size. In this case, they cannot understand the whole of the image. The Ref. [32] suggests using vision transformers

to remove restrictions of CNNs in image reconstruction. The slice-by-slice transformer network (SSTrans-3D) is a transformer-based technique that reconstructs 3D single-photon emission computed tomography (SPECT) images. In image reconstruction, incomplete projection data can cause considerable artefacts. Researchers in the Ref. [33] suggest Deep Iterative Optimization-based Residual-learning (DIOR) to reconstruct limited-angle CT. To improve generalization ability, DIOR combines deep learning and iterative optimization.

2.3. Deep Learning and Super-Resolution (SR) Techniques

Generally, super-resolution (SR) techniques can be clustered into traditional and deep learning methods. In SRCNN [34], an SR model is trained based on CNN to reconstruct HR from a given LR image. However, in terms of image quality, there are some limitations in SRCNN models. A huge amount of stacked layers causes gradient-vanishing issues. To solve those problems, researchers in the Ref. [35] suggested using a very deep convolutional network base on VGG-net. For some of the deep learning techniques in SR, overfitting is highly likely and models can be so big to be stored. To address these issues, a deeply recursive convolutional network (DRCN) [36] is suggested. In DRCN, a convolutional layer is repeated many times and the number of parameters depends on the number of applied recursions. As DRCN uses stochastic gradient descent widely, it cannot converge easily. To overcome the difficulties of the training, they use recursive supervision and skip connection. In the Ref. [37], authors suggest an end-to-end deep neural network that is constituted of an encoder, a fusion module and a decoder. The encoder takes the features of the LR image. Then they apply a Gated Recurrent Unit (GRU)-based module to combine the features. The super-resolution image is reconstructed by the decoder. The Ref. [38] suggests coupled-discriminate GAN (CDGAN). In CDGAN, HR and SR images are taken by a discriminator. The network can learn to discriminate low-frequency images. An unsupervised Image Super-Resolution is suggested in the Ref. [39]. They propose a framework that is based on two parts, unsupervised translation from the original LR image to the reconstructed LR image and supervised SR between reconstructed LR and HR images. In the Ref. [40] the authors present Meta-Transfer Learning for Zero-Shot Super-Resolution (MZSR). As CNNs are mostly limited to the trained supervised data, they are applicable to specific images. On the other hand, with CNNs it is not applicable to extract internal details of the images. To solve the problems they suggest using zero-shot super-resolution to train internal information. To overcome several gradient updates, they apply meta-learning for zero-shot super-resolution. In remote sensing, There are some issues in super-resolution reconstruction using deep learning, such as model training difficulties and blurred image edges. The Ref. [41] proposes a technique with the combination of residual channel attention (CA) to extract deep features. They combine shallow and deep futures, using skip connections to improve the model training. After training, the super-resolution images have sharper edges. In the Ref. [42] authors suggest using a Residual Back-Projection network (RBPNet) to reconstruct SR from extremely low-resolution face images. To generate the low-resolution feature map, RBPNet projects the high-resolution feature map to the low-resolution feature space. Then to make a residual feature map for the low-resolution feature map, they subtract the low-resolution feature map from the original feature map. Finally, to generate high-resolution feature space, the low-resolution feature map is projected into high-resolution feature space. RBPNet produces more precise high-resolution images using residual learning. During the application of SR techniques in low-resolution face images, face structure details cannot be recovered well. Thus, a novel SPatial Attention Residual Network (SPARNet) is suggested by the Ref. [43]. SPARNET uses a spatial attention mechanism to enable convolutional layers to focus on key face structures and pay less attention to less important details. The results show that their technique can detect key face structures well for extremely low-resolution (16×16) face images. They also extend SPARNet to SPARNetHD which can generate super-resolution face images (512×512). In the Ref. [44], the authors address the heavy computation and the high number of parameters of current SR techniques. They suggest the feedback

ghost residual dense network (FGRDN). Instead of the residual dense blocks (RDB) they use ghost modules (GM) which can prevent the rise of the parameters by increasing the network depth. FGRDN can converge faster than other corresponding algorithms. The Ref. [45] suggests a texture enhancement and generative adversarial network (TE-SAGAN), to generate super-resolution remote sensing images. They suggest an improved generator which is based on a residual network with self-attention and weight normalization. By improving training model stability, the generator can generate images with higher quality. The Ref. [46] proposes a fused recurrent network via channel attention (CA-FRN). Their main concerns are addressed regarding overfitting and the number of parameters. In SR techniques, by increasing the number of layers, the risk of overfitting increases too. CA-FRN uses a recursive channel attention block to pay attention to high-frequency information. In their model, high- and low-resolution information is fused to generate better results.

2.4. Volume Data Compression

Dealing with volumetric datasets may cause slow representation and huge file sizes. The Ref. [47] suggests that a quadtree encoding-based model compresses the volumetric medical images. Their technique is constituted of three stages, initialization, processing, and variable length encoding. In comparison to the octree technique, it shows better results in the image compression rate. In the Ref. [48], authors suggest a 3D hierarchical listless block (3D-HLCK). They utilize a 3D listless technique to compress the volumetric medical images. Their results illustrate that 3D-HLCK outperforms the 3D-SPIHT method. Researchers in the Ref. [49] present a GPU-based compression technique for volumetric medical images. In their work, they propose a caching strategy to render high-resolution volumetric medical images. In the Ref. [50] authors suggest a compression technique based on the wavelet transform domain. Their results show that the reconstruction quality of the volumetric medical images is not considered high. The Ref. [51] presents a volume data compression technique using the regression function. Their technique is based on a multi-layer perceptron (MLP) neural network to compress the volumetric medical images. In the Ref. [52] they used a stacked autoencoder to compress the malaria-infected blood cells. Authors in the Ref. [2] used neural networks with periodic activation functions (SIREN) to compress multidimensional medical images. Their results show that SIREN outperforms other INR-based compression techniques with ReLU or tanh activation functions. The Ref. [53] suggests CNN-based image compression to compress malaria cell images using a compressor–decompressor framework. According to their technique, there are two autoencoders, where one learns low-frequency components and another learns high-frequency components. As they use Huffman coding and decoding, it is necessary to have a large-sized training dataset to obtain better results.

3. Methodology

In the following section, we present our architecture and explain its stages in detail.

3.1. Our Architecture

We aim at the medical volume data compression technique to work on usual workstations and laptops instead of the overdependency on high-end GPUs. In order to increase the compression rate and the quality of the reconstruction, we present an architecture that can be applied in many scenarios, as shown in Figure 1. Our architecture consists of three modules, where the first module can downsample the original high-resolution images using Lanczos to decrease the volume data size. The second module can work on the LR volume data for the implicit neural representation. The third module can take the LR volume data back to the original size using SRDenseNet. As a result, the original volume data can be represented by the neural network to achieve compression purposes. The first module may result in information loss. The third module can recover it. There is no guarantee that it is lossless. However, the experiments show our architecture can reach high PSNR.
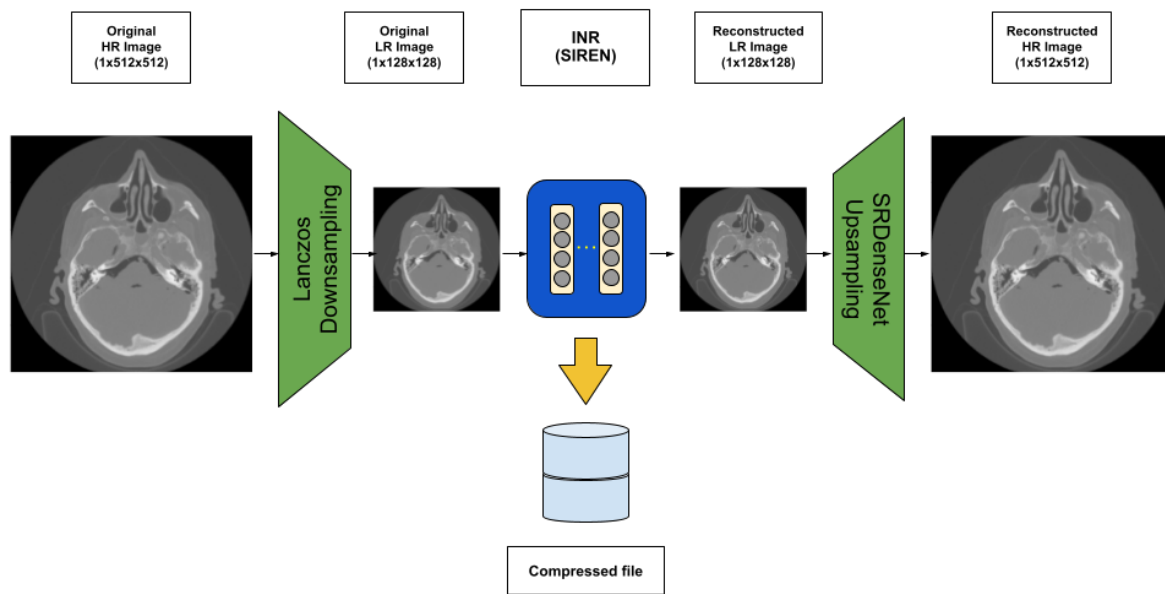
Figure 1. Our suggested architecture using INR (in this work, SIREN) to compress high-resolution (HR) medical images.

3.2. Lanczos Resampling

The Lanczos resampling technique [54] is an interpolation technique which is used to resample (downsample/upsample) the volumetric medical images. Here, we use Lanczos resampling to downsample our HR images. In the Lanczos resampling technique, each selected sample is replaced with a translated copy of the Lanczos kernel, which is a limited Sinc function. Equation (1) shows the Lanczos kernel formula.

$$L(x) = \begin{cases} sinc(\pi x)sinc(\pi x/a) & if |x| < a \\ 0 & otherwise. \end{cases} \tag{1}$$

3.3. Sinusoidal Representation Networks (SIREN)

We apply the SIREN deep network to medical image stacks. It may be regarded as a function that accepts the 3D coordinate of a voxel $\mathbf{x} = (x, y, z)$ and outputs the corresponding intensity value σ. In this case, the neural network becomes a continuous implicit function $F_\Theta : \mathbf{x} \rightarrow \sigma$ and can represent a continuous scene of medical image stacks. The training is to optimize its weights Θ for accurate representation. Recently, INRs have been one of the considerable areas of research focus. Most of the methods adopt a MLP structure using ReLU [9] as the activation function. ReLU-based MLPs are suffering from the lack of capacity to represent fine details. ReLU networks are linear and their second derivative is zero. As a result, it is impossible to model high-order signals. In the various versions of SIREN, the Sine function is used as the activation function to model fine details [1]. We applied a small version to training, and the resulting deep network was so small. As a new representation of the original volume data, it effectively compresses the volume data and saves training time.

3.4. SRDenseNet

In order to obtain the original resolution image, we use SRDenseNet [55] as a super-resolution module, which is a deep network. We selected SRDenseNet because of its generalization. We note that it can work well on a large class of medical images. For example, we use a small training image set from a large class and then perform it on unseen images of the same class. There are three types of SRDenseNets, SRDenseNet-H, SRDenseNet-HL, and SRDenseNet-All. SRDenseNet-H uses high-level features to

reconstruct HR images. SRDenseNet-HL uses a combination of low-level and high-level features to reconstruct HR images. SRDenseNet-All, which is used in our implementation, uses a combination of all levels of features with densely skipped connections to reconstruct HR images. SRDenseNet-All is constituted of four stages, low-level features, high-level features, deconvolution layers, and a reconstruction layer.

In the low-level feature stage, the network receives a low-resolution (LR) image as input. A convolution layer learns low-level features. There are eight dense blocks that learn high-level features. Each dense block is constituted of eight convolution layers and each layer produces 16 feature maps. Thus, each block produces 128 feature maps. In the bottleneck layer, the number of feature maps is decreased to a compact model. Then, image resolution is up-scaled from 128×128 to 512×512 using a deconvolution layer. Finally, with another convolutional layer with a 3×3 kernel, the output channels are reduced to a single channel. Figure 2 shows the SRDenseNet-All structure which is used in this paper. In this work, we use the pre-trained model of SRDenseNet in [55]. We trained the SRDenseNet network with LR and HR pairs of slices of our volumetric medical images dataset. Then a pre-trained model of our training is used in our architecture. According to this, SRDenseNet in our architecture can reconstruct HR images of downsampled images of our dataset.
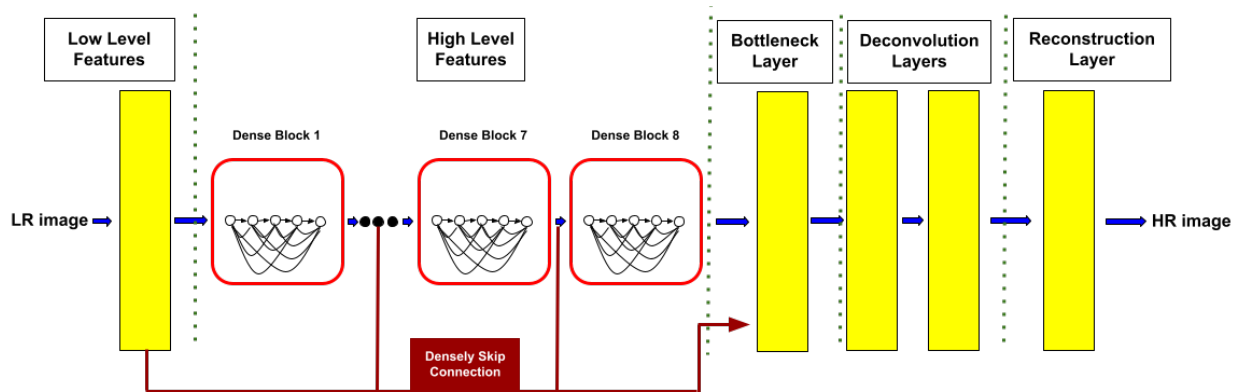

Figure 2. SRDensenet—all architecture with eight blocks.

3.5. Peak Signal-to-Noise Ratio

To compare the quality of the reconstructed image with the original one, we use the Peak Signal-to-Noise Ratio (PSNR) measurement technique. It is given for grey-scale images by the equation:

$$PSNR = 10.log_{10}\left[\frac{I^2}{MSE}\right] \tag{2}$$

where MSE is the mean square error which is given by the equation below:

$$MSE = \frac{1}{N.M}\sum_{i=1}^{N}\sum_{j=1}^{M}(y_{ij} - \bar{y}_{ij}) \tag{3}$$

I is the maximum intensity value of the image, which is 255 in this work. N is the number of rows and M is the number of columns. y is the original image and $\bar{y}$ is the reconstructed image.

4. Results and Discussion

In this section, we compare the numerical results of the applications of INR-based compression techniques with and without our architecture in terms of reconstruction quality, training speed, and GPU adaptivity. In this work, we examine an INR-based method, SIREN with two, three, and four MLP layers. Tables 1–3 show the size of the MLPs with two, three, and four layers, respectively, in detail.

Table 1. The estimation of SIREN network size with two layers.

Layer (Type)	Output Shape	Number of Parameters
Linear-1	$[-1, 1, 128]$	512
Linear-2	$[-1, 1, 128]$	16,512
Linear-3	$[-1, 1, 1]$	129
Total params: 17,153		
Trainable params: 17,153		
Non-trainable params: 0		
Input size (MB): 0.00		
Forward/backward pass size (MB): 0.00		
Params size (MB): 0.07		
Estimated total size (MB): 0.07		

Table 2. The estimation of SIREN network size with three layers.

Layer (Type)	Output Shape	Number of Parameters
Linear-1	$[-1, 1, 128]$	512
Linear-2	$[-1, 1, 128]$	16,512
Linear-3	$[-1, 1, 128]$	16,512
Linear-4	$[-1, 1, 1]$	129
Total params: 33,665		
Trainable params: 33,665		
Non-trainable params: 0		
Input size (MB): 0.00		
Forward/backward pass size (MB): 0.00		
Params size (MB): 0.13		
Estimated total size (MB): 0.13		

Table 3. The estimation of SIREN network size with four layers.

Layer (Type)	Output Shape	Number of Parameters
Linear-1	$[-1, 1, 128]$	512
Linear-2	$[-1, 1, 128]$	16,512
Linear-3	$[-1, 1, 128]$	16,512
Linear-4	$[-1, 1, 128]$	16,512
Linear-5	$[-1, 1, 1]$	129
Total params: 50,177		
Trainable params: 50,177		
Non-trainable params: 0		
Input size (MB): 0.00		
Forward/backward pass size (MB): 0.00		
Params size (MB): 0.19		
Estimated total size (MB): 0.20		

4.1. Dataset

In this paper, we are concerned with both 2D and 3D images. We use a volumetric dataset, containing Human CT scan slices obtained from the Visible Human project dataset (https://www.nlm.nih.gov, (accessed on 1 April 2022)). The database constitutes 463 DICOM axial CT scan slices of a human head with a resolution of 512 × 512 pixels and each pixel is made up of 12 bits of grey tone. The thickness of each slice is 0.5 mm.

4.2. Using SIREN with Our Architecture

Based on our architecture, we use the LANCZOS resampling technique to downsample our high-resolution volumetric medical images (512 × 512 × 463) to (128 × 128 × 115). This technique decreases the number of voxels from 121,372,672 to just 1,884,160. For simplicity, we show the results on a selected 2D slice, but it should be considered that it can be expanded to 3D data. Figure 3 shows a selected slice that is downsampled from 512 × 512 × 1 to 128 × 128 × 1 using LANCZOS, which decreases the number of voxels from 262,144 to just 16,384.
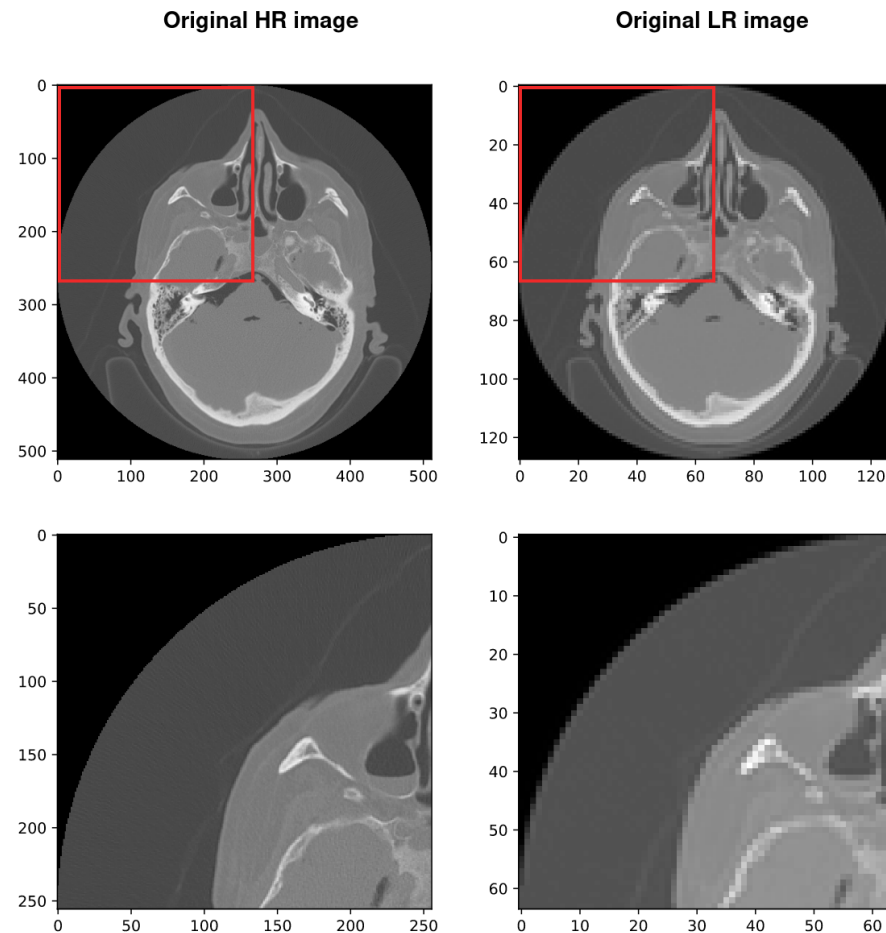
Figure 3. Left column shows the high-resolution slices with a size of 512 × 512 and the right is the low-resolution slices with a size of 128 × 128. Low-resolution slices were obtained by applying Lanczos resampling on the high-resolution slices.

After the downsampling stage, our architecture uses SIREN with 2, 3, and 4 layers and 128 neurons for each layer. Figures 4–6 show the results, respectively. The Adam was selected as the optimizer with a learning rate of 0.0015 and batch size of 2^{16}. MSE was used as the loss function of the MLP. In order to reconstruct the HR slice, we used SRDenseNet and applied it to reconstructed LR images. We also trained it on this kind of CT data in advance so that the trained SRDenseNe is suitable for the targeted class. The final results of using SIREN with our architecture are shown in Figures 7–9, respectively. Table 4 compares the results of the SIREN application with 2, 3 and 4 layers, using our architecture.
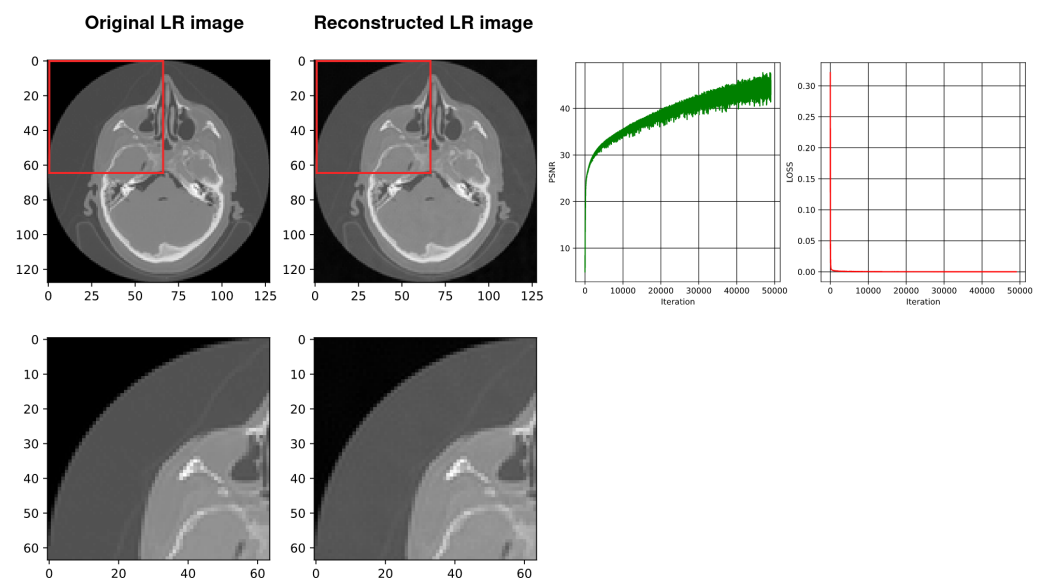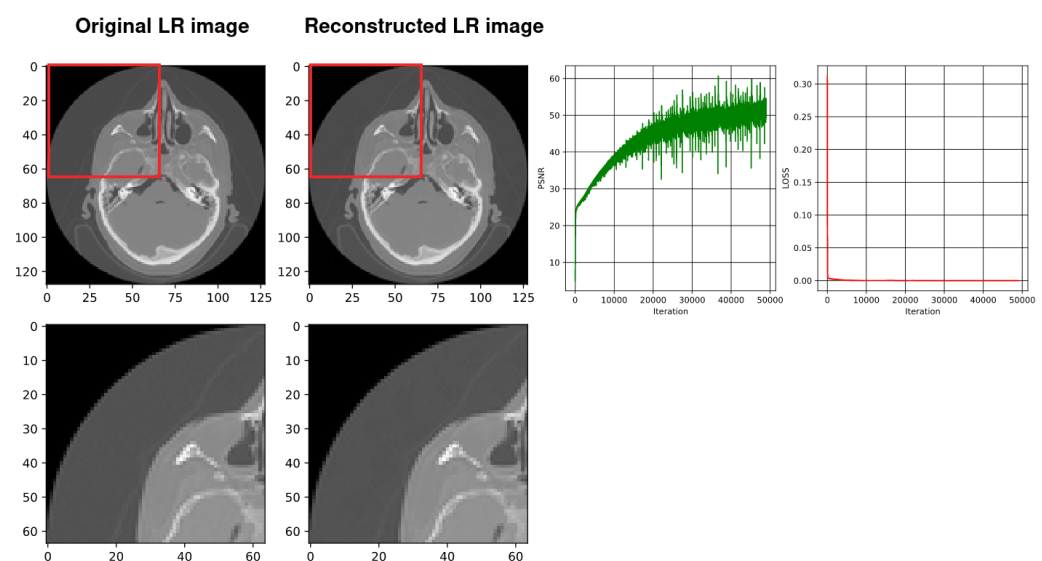
Figure 4. From left, the first image shows the original down-sampled image. The second image illustrates the reconstructed counterpart of the down-sampled image using SIREN. The next image shows the plot of the PSNR while training the network (best PSNR: 45.38) and the last image shows the plot of the loss values during the SIREN training (loss: 1.4805471×10^{-5}).
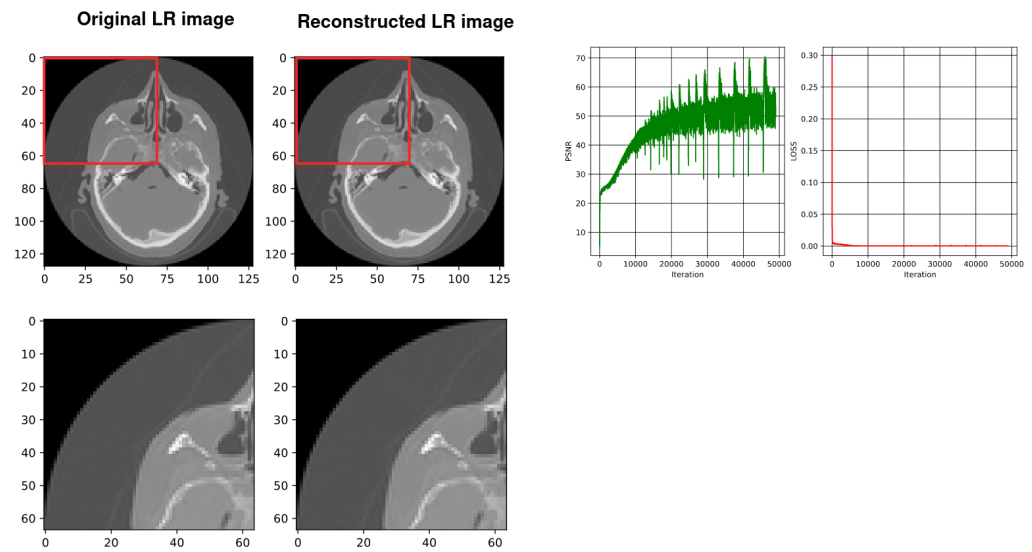

Figure 5. From the left, the first image shows the original down-sampled image. The second image illustrates the reconstructed counterpart of the down-sampled image using SIREN. The next image shows the plot of the PSNR while training the network (best PSNR: 48.29) and the last image shows the plot of the loss values during the SIREN training (loss: 1.0805471×10^{-5}).

Table 4. Shows comparison results of the implementation of SIREN with our architecture in terms of quality, speed, GPU memory allocation, and compression rate with 2, 3, and 4 layers of SIREN.

Number of Layers	Best PSNR	Training Time(s)/50,000 Iters	Compression Rate	GPU Memory (KB)
Two Layer	34.670	55.76	3.65	1038
Three Layer	34.865	74.80	1.96	1296
Four Layer	35.140	99.61	1.28	1554

Figure 6. From the left, the first image shows the original down-sampled image. The second image illustrates the reconstructed counterpart of the down-sampled image using a four-layer SIREN. The next image shows the plot of the PSNR during training of the network (best PSNR: 70.63) and the last image shows the plot of the loss values during the four-layer SIREN training (loss: 8.6481705×10^{-8}).
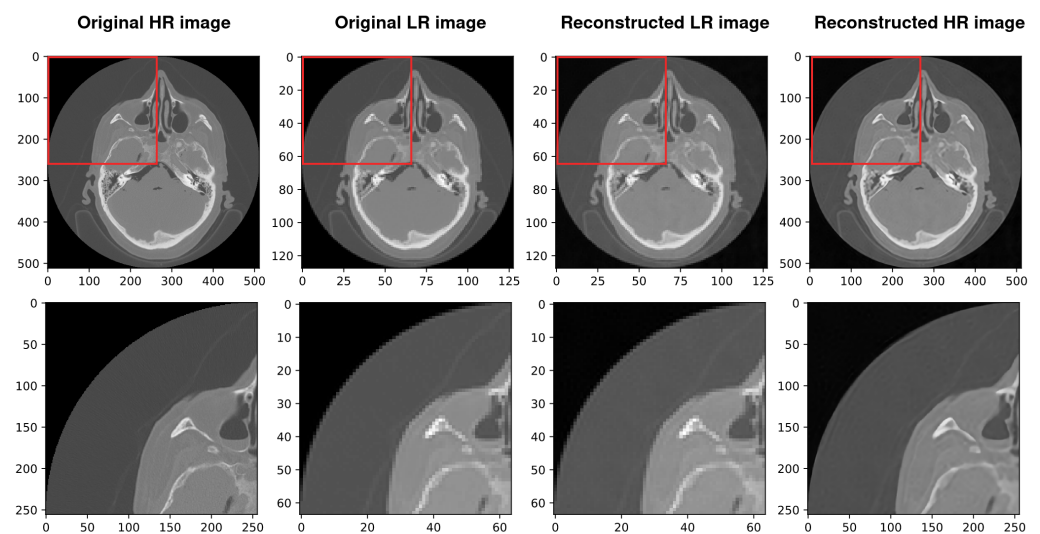

Figure 7. Results of the whole procedure of our architecture. From the left, the first image shows the original high-resolution (HR) image. The second image shows the original low resolution (LR). The third image shows the reconstructed low resolution (LR) and the last shows the reconstructed high resolution (HR). The best PSNR of reconstructed HR in comparison with the original HR is 34.670.

4.3. Using SIREN without Our Architecture [2]

In this section, we show the results of SIREN implementation without our architecture which was implemented in the Ref. [2]. Figures 10–12 show the results of high-resolution (HR) reconstruction of 2, 3 and 4 layers of SIREN without our architecture, respectively. Table 5 illustrates the results of using SIREN without our architecture with 2, 3, and 4 layers.
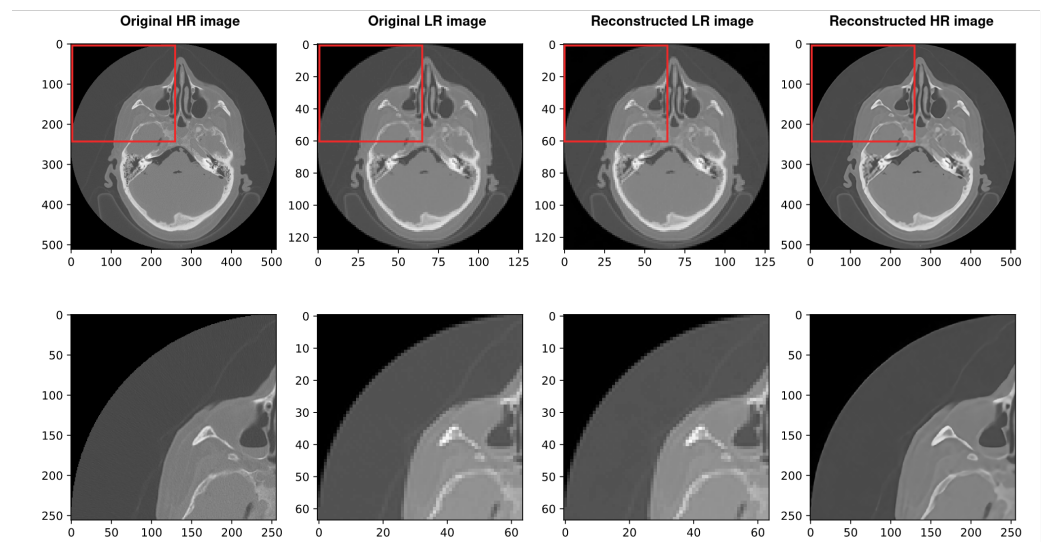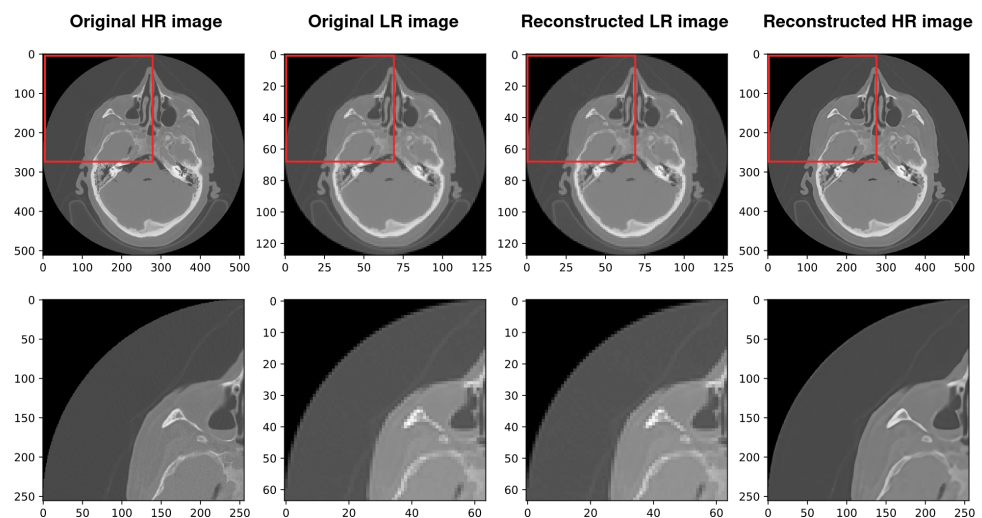
Figure 8. Results of the whole procedure of our architecture. From the left, the first image shows the original high-resolution (HR) image. The second shows the original low resolution (LR). The third image shows the reconstructed low resolution (LR) and the last shows the reconstructed high resolution (HR). The best PSNR of reconstructed HR in comparison with original HR is 34.865.

Figure 9. Results of the whole procedure of our architecture. From the left, the first image shows the original high-resolution (HR) image. The second shows the original low resolution (LR). The third image shows the reconstructed low resolution (LR) and the last shows the reconstructed high resolution (HR). The best PSNR of reconstructed HR in comparison with the original HR is 35.140.

Table 5. Shows comparison results of the implementation of SIREN without our architecture in terms of quality, speed, GPU memory allocation, and compression rate with 2, 3, and 4 layers with SIREN.

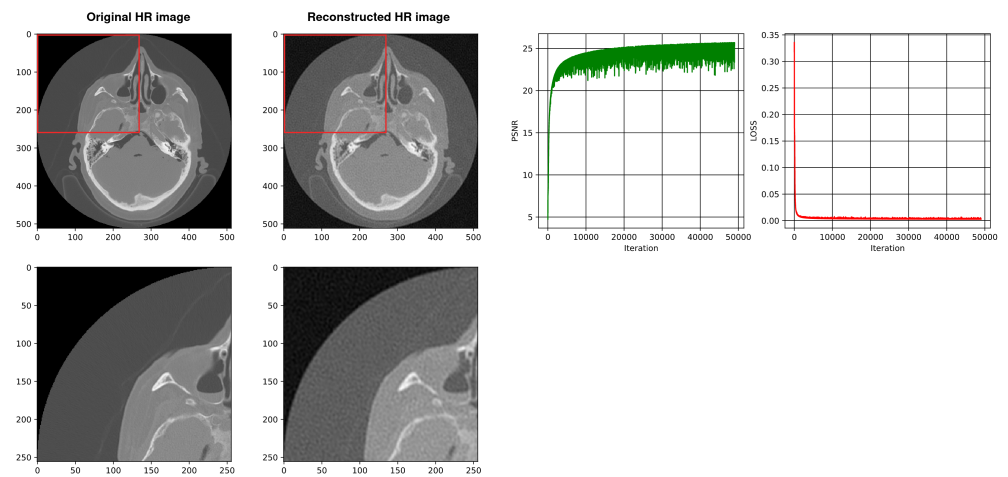| Number of Layers	Best PSNR	Training Time(s)/ 50,000 Iters	Compression Rate	GPU Memory (KB)
Two Layer	25.269	664.92	3.65	10,254
Three Layer	28.800	1004.21	1.96	10,512
Four Layer	30.689	1336.53	1.28	10,770

Figure 10. Shows the reconstruction of HR images using a two-layer SIREN without our architecture. The best PSNR is 25.26.
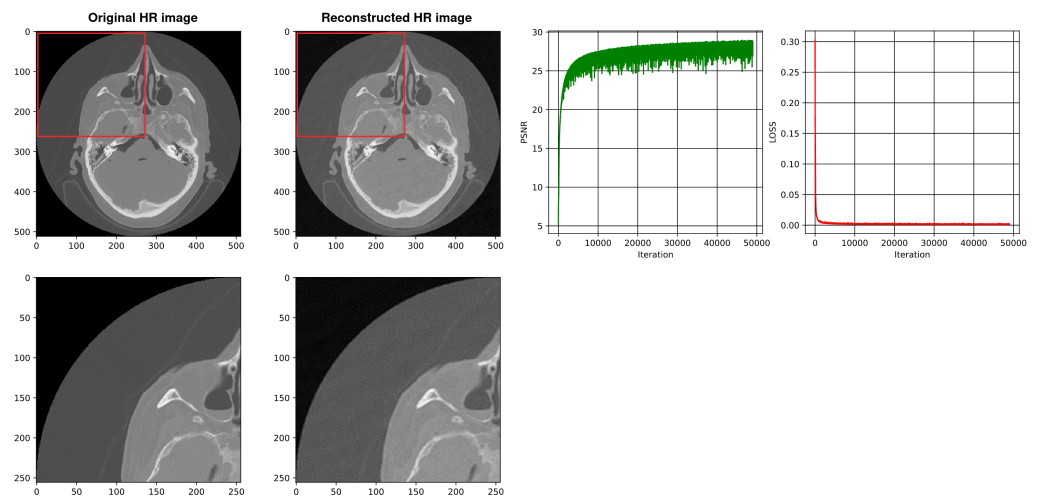

Figure 11. Shows the reconstruction of HR images using a three-layer SIREN without our architecture. The best PSNR is 28.80.
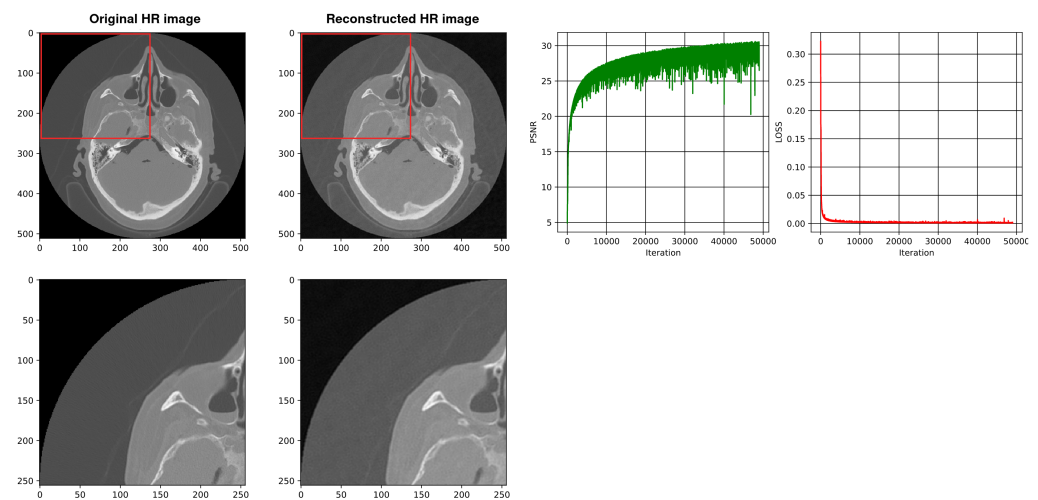

Figure 12. Shows the reconstruction of HR images using a four-layer SIREN without our architecture. The best PSNR is 30.68.

4.4. Comparison with Existing Methods

In this section, we compare the performance of SIREN with our architecture and SIREN without our architecture which is used in the Ref. [2] with 2, 3 and 4 layers and 128 neurons for each layer. From the general perspective, both methods use SIREN to compress volumetric medical images. In the Ref. [2] they applied SIREN without any downsampling and upsampling techniques which led to high GPU memory consumption and low training speed. In Figure 13 there are three bar charts, showing that SIREN with our architecture outperforms SIREN without our architecture in the Ref. [2] in terms of quality, speed, and GPU memory consumption with 2, 3, and 4 MLP layers. As can be seen, there are considerable gaps between our technique and SIREN in the Ref. [2] in terms of training speed and GPU memory consumption. The PSNR gap is decreased by increasing the number of layers, but our architecture still outperforms the SIREN in the Ref. [2]. Table 6 compares the results of using SIREN to compress volumetric medical images in the Ref. [2] with using SIREN with our architecture in three factors of PSNR, training speed and GPU memory consumption. Figure 14 compares the final results of using SIREN in the Ref. [2] and using SIREN with our architecture with 2, 3 and 4 layers. Table 7 shows the comparison of the results of Shen's [52], Mishra's [53], SIREN [2] and ours in terms of PSNR.

Table 6. Display and comparison of the final results that were obtained using SIREN in the Ref. [2] and using SIREN with our architecture. As it can be seen, using SIREN with our architecture outperforms using SIREN in the Ref. [2] in three terms of quality, speed, and GPU memory allocation with 2, 3 and 4 MLP layers.

	Number of Layers	**SIREN without Our Architecture [2]**	**SIREN with Our Architecture**
Best PSNR (dB)	2 layers	25.269	**34.670**
	3 layers	28.800	**34.865**
	4 layers	30.689	**35.140**
Training time(s) (for 15000 iters)	2 layers	664.92	**55.76**
	3 layers	1004.21	**74.80**
	4 layers	1336.53	**99.61**
GPU memory consumption (KB)	2 layers	10,254	**1,038**
	3 layers	10,512	**1,296**
	4 layers	10,770	**1,554**

Table 7. Comparison of results of our technique with SIREN [2], Mishra's [53] and Shen's [52] in terms of PSNR.

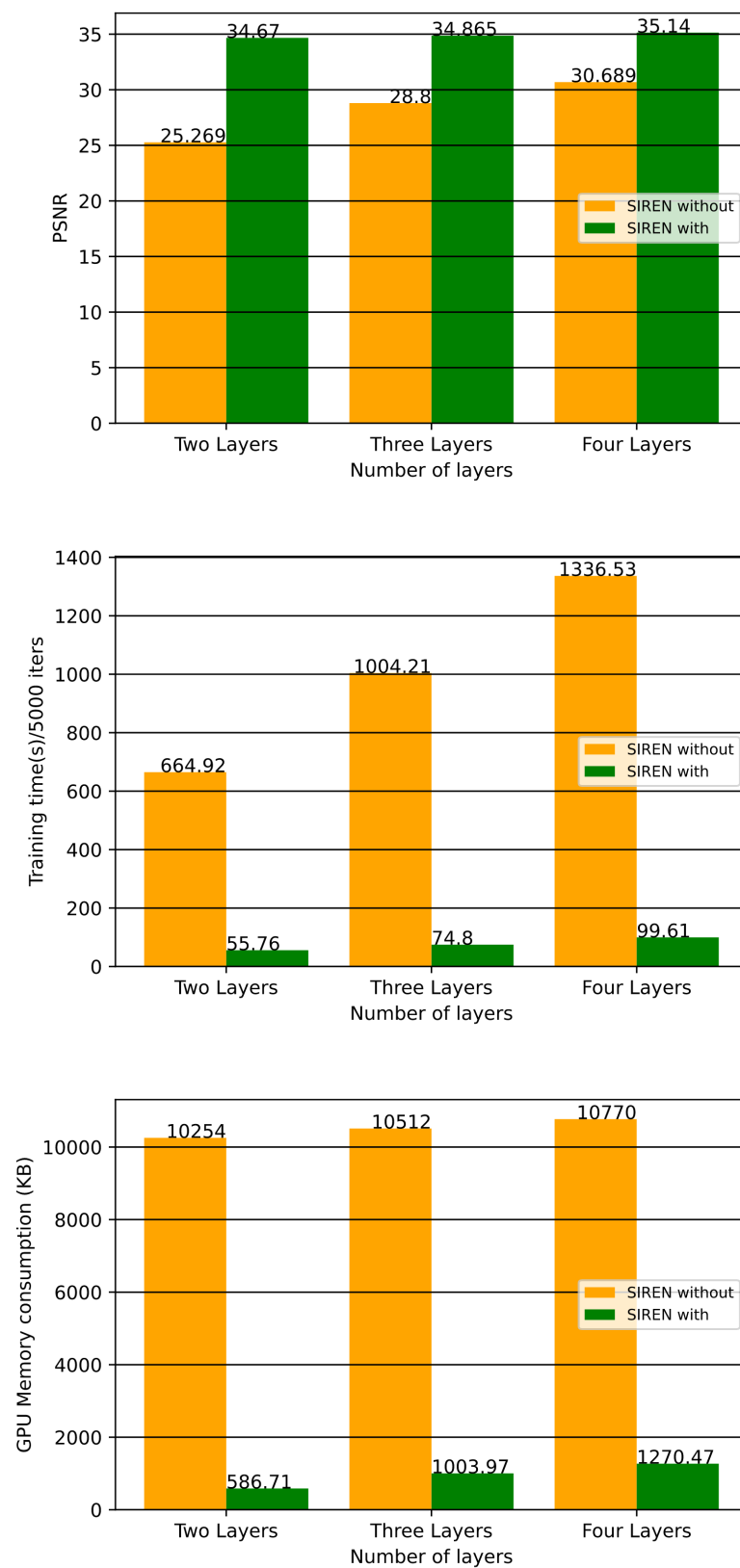| **Methods**	**Best PSNR**
Shen's [52]	27.50
Mishra's [53]	29.33
SIREN [2] (2 layers)	25.269
SIREN [2] (3 layers)	28.800
SIREN [2] (4 layers)	30.689
Ours (2 layers)	**34.670**
Ours (3 layers)	**34.865**
Ours (4 layers)	**35.140**

Figure 13. Bar charts compare the results of using SIREN without our architecture [2] and with our architecture. As can be seen, SIREN implementation with our architecture outperforms using SIREN without our architecture [2] in terms of quality, speed, and GPU memory allocation with 2, 3, and 4 MLP layers.
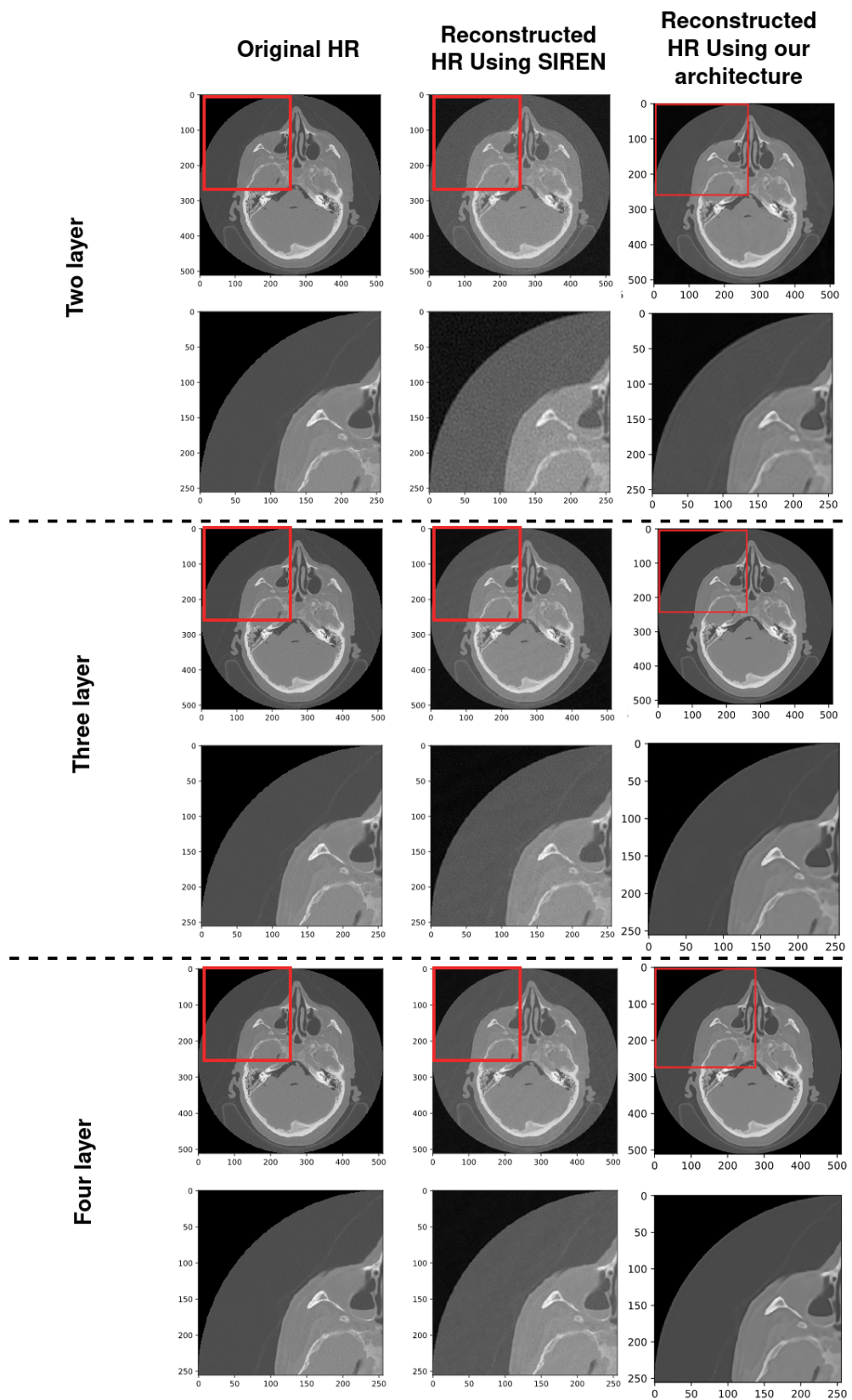
Figure 14. Comparison of the quality of reconstructed HR using SIREN in the Ref. [2] and using SIREN with our architecture with 2, 3 and 4 MLP layers.

5. Conclusions

Our architecture significantly outperforms other INR-based techniques without our architecture, which are regular in medical image compression. The experiments show that our proposed architecture is a novel implicit neural representation of medical volume data, reaching both a high compression rate and high quality of reconstruction. The architecture is simple and re-configurable. It may replace volume data and be regarded as a new representation form. Moreover, the three modules can be replaced by others according to different applications. For example, manual annotation may work on low-resolution volume to save time, while rendering may work on high-resolution volume for accuracy.

The quality of our reconstructed high-resolution images with a small version of SIREN is considerably higher than direct SIREN with the same size. We also note that the compression rate depends on the deep network structure. To reach a high compression rate and low loss for volume data, the SR module is crucial. Moreover, the generalization of the SR module is another concern. These are worth our effort in the future.

Abbreviations

The following abbreviations are used in this manuscript:

INR	Implicit Neural Representation
MLP	Multi-Layer Perceptron
CT	Computed Tomography
SIREN	Sinusoidal representation network

References

1. Sitzmann, V.; Martel, J.N.P.; Bergman, A.W.; Lindell, D.B.; Wetzstein, G. Implicit Neural Representations with Periodic Activation Functions. *arXiv*, **2020**, arXiv:2006.09661. [CrossRef]
2. Mancini, M.; Jones, D.K.; Palombo, M. Lossy compression of multidimensional medical images using sinusoidal activation networks: An evaluation study. *arXiv*, **2022**, arXiv:2208.01602. [CrossRef]
3. Strümpler, Y.; Postels, J.; Yang, R.; van Gool, L.; Tombari, F. Implicit Neural Representations for Image Compression. *arXiv*, **2021**, arXiv:2112.04267. [CrossRef]
4. Wu, J.; Zhang, C.; Zhang, X.; Zhang, Z.; Freeman, W.T.; Tenenbaum, J.B. Learning Shape Priors for Single-View 3D Completion and Reconstruction. *arXiv*, **2018**, arXiv:1809.05068. [CrossRef]
5. Khan, M.O.; Fang, Y. Implicit Neural Representations for Medical Imaging Segmentation. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2022*; Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S., Eds.; Springer Nature Switzerland: Cham, Switzerland, 2022; pp. 433–443.
6. Park, J.J.; Florence, P.; Straub, J.; Newcombe, R.; Lovegrove, S. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. *arXiv*, **2019**, arXiv:1901.05103. [CrossRef]
7. Tang, D.; Singh, S.; Chou, P.A.; Haene, C.; Dou, M.; Fanello, S.; Taylor, J.; Davidson, P.; Guleryuz, O.G.; Zhang, Y.; et al. Deep Implicit Volume Compression. *arXiv*, **2020**, arXiv:2005.08877. [CrossRef]
8. Nagoor, O.H.; Whittle, J.; Deng, J.; Mora, B.; Jones, M.W. MedZip: 3D Medical Images Lossless Compressor Using Recurrent Neural Network (LSTM). In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021 ; pp. 2874–2881. [CrossRef]

9.   Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *arXiv*, **2020**, arXiv:2003.08934. [CrossRef]

10.   Mescheder, L.; Oechsle, M.; Niemeyer, M.; Nowozin, S.; Geiger, A. Occupancy Networks: Learning 3D Reconstruction in Function Space. *arXiv*, **2018**, arXiv:1812.03828. [CrossRef]

11.   Dupont, E.; Goliński, A.; Alizadeh, M.; Teh, Y.W.; Doucet, A. COIN: COmpression with Implicit Neural representations. *arXiv* **2021**, arXiv:2103.03123. [CrossRef]

12.   Skorokhodov, I.; Ignatyev, S.; Elhoseiny, M. Adversarial Generation of Continuous Images. *arXiv* **2020**, arXiv:2011.12026. [CrossRef]

13.   Shen, L.; Pauly, J.; Xing, L. NeRP: Implicit Neural Representation Learning with Prior Embedding for Sparsely Sampled Image Reconstruction. *arXiv* **2021**, arXiv:2108.10991. [CrossRef]

14.   Eslami, S.M.A.; Rezende, D.J.; Besse, F.; Viola, F.; Morcos, A.S.; Garnelo, M.; Ruderman, A.; Rusu, A.A.; Danihelka, I.; Gregor, K.; et al. Neural scene representation and rendering. *Science* **2018**, *360*, 1204–1210. [CrossRef] [PubMed]

15.   Tancik, M.; Srinivasan, P.P.; Mildenhall, B.; Fridovich-Keil, S.; Raghavan, N.; Singhal, U.; Ramamoorthi, R.; Barron, J.T.; Ng, R. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. In Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–12 December 2020; NIPS'20; Curran Associates Inc.: Red Hook, NY, USA, 2022.

16.   Dar, S.U.H.; Özbey, M.; Çatlı, A.B.; Çukur, T. A Transfer-Learning Approach for Accelerated MRI Using Deep Neural Networks. *Magn. Reson. Med.* **2020**, *84*, 663–685. [CrossRef]

17.   Qin, C.; Schlemper, J.; Caballero, J.; Price, A.N.; Hajnal, J.V.; Rueckert, D. Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction. *IEEE Trans. Med Imaging* **2019**, *38*, 280–290. [CrossRef]

18.   Wu, Y.; Ma, Y.; Liu, J.; Du, J.; Xing, L. Self-attention convolutional neural network for improved MR image reconstruction. *Inf. Sci.* **2019**, *490*, 317–328. [CrossRef]

19.   Liu, J.; Zhang, Y.; Zhao, Q.; Lv, T.; Wu, W.; Cai, N.; Quan, G.; Yang, W.; Chen, Y.; Luo, L.; et al. Deep iterative reconstruction estimation (DIRE): Approximate iterative reconstruction estimation for low dose CT imaging. *Phys. Med. Biol.* **2019**, *64*, 135007. [CrossRef] [PubMed]

20.   Wang, J.; Liang, J.; Cheng, J.; Guo, Y.; Zeng, L. Deep learning based image reconstruction algorithm for limited-angle translational computed tomography. *PLoS ONE* **2020**, *15*, e0226963. [CrossRef]

21.   Xie, H.; Shan, H.; Wang, G. Deep Encoder-Decoder Adversarial Reconstruction (DEAR) Network for 3D CT from Few-View Data. *Bioengineering* **2019**, *6*, 111. [CrossRef]

22.   Xie, S.; Zheng, X.; Chen, Y.; Xie, L.; Liu, J.; Zhang, Y.; Yan, J.; Zhu, H.; Hu, Y. Artifact removal using improved GoogLeNet for sparse-view CT reconstruction. *Sci. Rep.* **2018**, *8*, 6700. [CrossRef]

23.   Xu, L.; Zeng, X.; Huang, Z.; Li, W.; Zhang, H. Low-dose chest X-ray image super-resolution using generative adversarial nets with spectral normalization. *Biomed. Signal Process. Control* **2020**, *55*, 101600. [CrossRef]

24.   Khodajou-Chokami, H.; Hosseini, S.A.; Ay, M.R. A deep learning method for high-quality ultra-fast CT image reconstruction from sparsely sampled projections. *Nucl. Instruments Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.* **2022**, *1029*, 166428. [CrossRef]

25.   Wang, T.; He, M.; Shen, K.; Liu, W.; Tian, C. Learned regularization for image reconstruction in sparse-view photoacoustic tomography. *Biomed. Opt. Express* **2022**, *13*, 5721–5737. [CrossRef] [PubMed]

26.   Gong, K.; Catana, C.; Qi, J.; Li, Q. Direct Reconstruction of Linear Parametric Images From Dynamic PET Using Nonlocal Deep Image Prior. *IEEE Trans. Med Imaging* **2022**, *41*, 680–689. [CrossRef]

27.   Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 424–432.

28.   Häggström, I.; Schmidtlein, C.R.; Campanella, G.; Fuchs, T.J. DeepPET: A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem. *Med Image Anal.* **2019**, *54*, 253–262. [CrossRef] [PubMed]

29.   Kandarpa, V.S.S.; Bousse, A.; Benoit, D.; Visvikis, D. DUG-RECON: A Framework for Direct Image Reconstruction Using Convolutional Generative Networks. *IEEE Trans. Radiat. Plasma Med Sci.* **2021**, *5*, 44–53. [CrossRef]

30.   Yokota, T.; Kawai, K.; Sakata, M.; Kimura, Y.; Hontani, H. Dynamic PET Image Reconstruction Using Nonnegative Matrix Factorization Incorporated With Deep Image Prior. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019.

31.   Zhang, Y.; Hu, D.; Hao, S.; Liu, J.; Quan, G.; Zhang, Y.; Ji, X.; Chen, Y. DREAM-Net: Deep Residual Error Iterative Minimization Network for Sparse-View CT Reconstruction. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 480–491. [CrossRef]

32.   Xie, H.; Thorn, S.; Liu, Y.H.; Lee, S.; Liu, Z.; Wang, G.; Sinusas, A.J.; Liu, C. Deep-Learning-Based Few-Angle Cardiac SPECT Reconstruction Using Transformer. *IEEE Trans. Radiat. Plasma Med Sci.* **2023**, *7*, 33–40. [CrossRef]

33.   Hu, D.; Zhang, Y.; Liu, J.; Luo, S.; Chen, Y. DIOR: Deep Iterative Optimization-Based Residual-Learning for Limited-Angle CT Reconstruction. *IEEE Trans. Med Imaging* **2022**, *41*, 1778–1790. [CrossRef]

34.   Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [CrossRef]

35. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1646–1654.

36. Wang, X.; Yi, J.; Guo, J.; Song, Y.; Lyu, J.; Xu, J.; Yan, W.; Zhao, J.; Cai, Q.; Min, H. A Review of Image Super-Resolution Approaches Based on Deep Learning and Applications in Remote Sensing. *Remote Sens.* **2022**, *14*. [CrossRef]

37. Arefin, M.R.; Michalski, V.; St-Charles, P.-L.; Kalaitzis, A.; Kim, S.; Kahou, S.E.; Bengio, Y. Multi-image super-resolution for remote sensing using deep recurrent networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 206–207.

38. Yu, Y.; Li, X.; Liu, F. E-DBPN: Enhanced deep back-projection networks for remote sensing scene image superresolution. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 5503–5515. [CrossRef]

39. Chen, S.; Han, Z.; Dai, E.; Jia, X.; Liu, Z.; Xing, L.; Zou, X.; Xu, C.; Liu, J.; Tian, Q. Unsupervised image super-resolution with an indirect supervised path. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 468–469.

40. Soh, J.W.; Cho, S.; Cho, N.I. Meta-Transfer Learning for Zero-Shot Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.

41. Gao, L.; Sun, H.-M.; Cui, Z.; Du, Y.; Sun, H.; Jia, R. Super-resolution reconstruction of single remote sensing images based on residual channel attention. *J. Appl. Remote Sens.* **2021**, *15*, 016513. [CrossRef]

42. Chen, X.; Wang, X.; Lu, Y.; Li, W.; Wang, Z.; Huang, Z. RBPNET: An asymptotic Residual Back-Projection Network for super-resolution of very low-resolution face image. *Neurocomputing* **2020**, *376*, 119–127. [CrossRef]

43. Chen, C.; Gong, D.; Wang, H.; Li, Z.; Wong, K.Y.K. Learning Spatial Attention for Face Super-Resolution. *IEEE Trans. Image Process.* **2021**, *30*, 1219–1231. [CrossRef]

44. Wang, J.; Wu, Y.; Wang, L.; Wang, L.; Alfarraj, O.; Tolba, A. Lightweight Feedback Convolution Neural Network for Remote Sensing Images Super-Resolution. *IEEE Access* **2021**, *9*, 15992–16003. [CrossRef]

45. Xu, Y.; Luo, W.; Hu, A.; Xie, Z.; Xie, X.; Tao, L. TE-SAGAN: An Improved Generative Adversarial Network for Remote Sensing Super-Resolution Images. *Remote Sens.* **2022**, *14*, 2425. [CrossRef]

46. Li, X.; Zhang, D.; Liang, Z.; Ouyang, D.; Shao, J. Fused Recurrent Network Via Channel Attention For Remote Sensing Satellite Image Super-Resolution. In Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME), London, UK, 6–10 June 2020; pp. 1–6. [CrossRef]

47. Klajnšek, G.; Žalik, B. Progressive lossless compression of volumetric data using small memory load. *Comput. Med. Imaging Graph.* **2005**, *29*, 305–312. . [CrossRef]

48. Senapati, R.K.; Prasad, P.M.K.; Swain, G.; Shankar, T.N. Volumetric medical image compression using 3D listless embedded block partitioning. *SpringerPlus* **2016**, *5*, 2100. [CrossRef]

49. Guthe, S.; Goesele, M. GPU-based lossless volume data compression. In Proceedings of the 2016 3DTV-Conference: The True Vision—Capture, Transmission and Display of 3D Video (3DTV-CON), Hamburg, Germany, 4–6 July 2016; pp. 1–4. [CrossRef]

50. Nguyen, K.G.; Saupe, D. Rapid High Quality Compression of Volume Data for Visualization. *Comput. Graph. Forum* **2001**, *20*, 49–57. [CrossRef]

51. Dai, Q.; Song, Y.; Xin, Y. Random-Accessible Volume Data Compression with Regression Function. In Proceedings of the 2015 14th International Conference on Computer-Aided Design and Computer Graphics (CAD/Graphics), Xi'an, China, 26–28 August 2015; pp. 137–142. [CrossRef]

52. Shen, H.; David Pan, W.; Dong, Y.; Alim, M. Lossless compression of curated erythrocyte images using deep autoencoders for malaria infection diagnosis. In Proceedings of the 2016 Picture Coding Symposium (PCS), Nuremberg, Germany, 4–7 December 2016; pp. 1–5. [CrossRef]

53. Mishra, D.; Singh, S.K.; Singh, R.K. Lossy Medical Image Compression using Residual Learning-based Dual Autoencoder Model. *arXiv*, **2021**, arXiv:2108.10579. [CrossRef]

54. Moraes, T.; Amorim, P.; Silva, J.V.D.; Pedrini, H. Medical image interpolation based on 3D Lanczos filtering. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2020**, *8*, 294–300. [CrossRef]

55. Tong, T.; Li, G.; Liu, X.; Gao, Q. Image Super-Resolution Using Dense Skip Connections. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4809–4817. [CrossRef]