MRI Super-Resolution via Realistic Downsampling with Adversarial Learning

Bangyan Huang[1], Haonan Xiao[2], Weiwei Liu[3], Yibao Zhang[3], Hao Wu[3], Weihu Wang[3], Yunhuan Yang[1], Yidong Yang[1], G. Wilson Miller[4], Tian Li[2], Jing Cai[2]

[1]Department of Engineering and Applied Physics, University of Science and Technology of China, Hefei, China

[2]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong SAR, China

[3]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiation Oncology, Beijing Cancer Hospital and Institute, Peking University Cancer Hospital and Institute, Beijing, China

[4]Department of Radiology and Medical Imaging, The University of Virginia, Charlottesville, VA, USA

Corresponding Author:

Jing Cai, Ph.D.

Department of Health Technology and Informatics

The Hong Kong Polytechnic University

11 Yuk Choi Rd, Hung Hom

Hong Kong, China

Email: jing.cai@polyu.edu.hk

Tian Li, Ph.D.

Department of Health Technology and Informatics

The Hong Kong Polytechnic University

11 Yuk Choi Rd, Hung Hom

Hong Kong, China

Email: litian.li@polyu.edu.hk

Short Title: MRI SR via Real-World Downsampling

Abstract

Many deep learning (DL) frameworks have demonstrated state-of-the-art performance in the super-resolution (SR) task of magnetic resonance imaging (MRI), but most performances have been achieved with simulated low-resolution (LR) images rather than LR images from real acquisition. Due to the limited generalizability of the SR network, enhancement is not guaranteed for real LR images because of the unreality of the training LR images. In this study, we proposed a DL-based SR framework with an emphasis on data construction to achieve better performance on real LR MR images. The framework comprised two steps: (a) downsampling training using a generative adversarial network (GAN) to construct more realistic and perfectly matched LR/high-resolution (HR) pairs. The downsampling GAN input was real LR and HR images. The generator translated the HR images to LR images and the discriminator distinguished the patch-level difference between the synthetic and real LR images. (b) Super-resolution training was performed using an enhanced deep super-resolution network (EDSR). In the controlled experiments, three EDSRs were trained using our proposed method, Gaussian blur, and k-space zero-filling. As for the data, liver MR images were obtained from 24 patients using breath-hold serial LR and HR scans (only HR images were used in the conventional methods). The k-space zero-filling group delivered almost zero enhancement on the real LR images and the Gaussian group produced a considerable number of artifacts. The proposed method exhibited significantly better resolution enhancement and fewer artifacts compared with the other two networks. Our method outperformed the Gaussian method by an improvement of 0.111 ± 0.016 in the structural similarity index (SSIM) and 2.76 ± 0.98 dB in the peak signal-to-noise ratio (PSNR). The blind/reference-less image spatial quality evaluator (BRISQUE) metric of the conventional Gaussian method and proposed method were 0.553 ± 0.039 and 0.669

± 0.021, respectively.

Key Words: super-resolution, MRI, deep learning, GAN, kernel estimation

1. Introduction

Magnetic resonance imaging (MRI) is widely used for various clinical purposes. Spatial resolution is one of the key parameters of MRI. High spatial resolution MR images consist of rich structural details that benefit treatment planning(Liu *et al.*, 2015), diagnosis, and image analysis(Van Reeth *et al.*, 2012). However, image resolution is limited by several factors(Van Reeth *et al.*, 2012; Plenge *et al.*, 2012), including the scanning time, MRI hardware, and desired signal-to-noise ratio (SNR). For example, it may be difficult for some people to hold their breath over a long period during an HR abdominal scan. Thus, image resolution is a trade-off between the scanning time and patients' comfort. Over the years, various post-processing techniques known as super-resolution (SR)(Van Reeth *et al.*, 2012) have been developed to improve the spatial resolution of MRI without modifying the hardware and scanning protocol. The aim of these post-processing techniques is to reconstruct HR images from a single or a set of low-resolution (LR) images to improve the visibility of the regions of interest (ROIs). This study will mainly focus on single image SR (SISR).

Prior to the emergence of deep learning, interpolation- and regularization-based methods had been the representative approaches to MRI SR research. Traditional interpolation-based methods(Lehmann *et al.*, 1999) resize the LR images to obtain HR images. Though these simple approaches are intuitive and can be implemented quickly, they show limited potential in reconstructing high frequency information. Some sophisticated interpolation-based methods incorporate non-local means (NLM) to improve performance(Manjón *et al.*, 2010; Jafari-Khouzani, 2014). Nevertheless, these methods do not apply in inhomogeneous regions because they assume some smooth priors(Luo *et al.*, 2017). Regularization-based methods(Shi *et al.*, 2015; Rueda *et al.*, 2013; Tourbier *et al.*, 2015; Zhang *et al.*, 2015) solve an optimization problem with a cost function consisting of fidelity and regularization terms. The fidelity term penalizes differences between the LR and degraded HR images, while various regularization terms, including low-rank(Shi *et al.*, 2015), total variation(Shi *et al.*, 2015; Tourbier *et al.*, 2015), non-local similarity(Zhang *et al.*, 2015; Jafari-Khouzani, 2014), and global regularization(Rueda *et al.*, 2013), are incorporated into the cost function. These methods share common drawbacks: they do not exploit information from the vast amount of training data, and it is time-consuming to iteratively solve the minimization problem for every single image.

With the rapid development of machine learning, especially deep learning (DL), convolutional neural network (CNN) has become the major component of state-of-the-art SR approaches to natural images(Dong *et al.*, 2015; Tai *et al.*, 2017; Lim *et al.*, 2017; Ledig *et al.*, 2017; Wang *et al.*, 2018; Wang *et al.*, 2020). Additionally, CNN has yielded success in MRI SR research(Chaudhari *et al.*, 2018; Pham *et al.*, 2017; Shi *et al.*, 2018b). Many previously developed SR techniques are supervised and pay attention to the network design(Shi *et al.*, 2018b; Shi *et al.*, 2018a; Lyu *et al.*, 2019). Some methods have shown effective network structures in recovering high-frequency content(Lim *et al.*, 2017; Shi *et al.*, 2018a). Most of these methods require strictly matched LR/HR pairs as the training data for the network(Lim *et al.*, 2017; Shi *et al.*, 2018b; Shi *et al.*, 2018a; Ledig *et al.*, 2017; Pham *et al.*, 2017). These methods consist of two steps: (i) generating matched LR and HR image pairs by a simple translation model (e.g., Gaussian blur and k-space zero-filling) and (ii) training the network by minimizing pixel-wise differences (e.g., mean absolute error).

To train an ideal SR network, the best way is to prepare strictly matched real LR and HR images, which can be achieved easily for brain MRI because patients' motion can be well-controlled. However, the acquisition of matched LR/HR pairs is almost impossible in abdominal MRI because the positions of the internal organs are affected by breathing(Cai *et al.*, 2008; Yang *et al.*, 2014) and other motions. Therefore, the alternative is to generate training LR images using simple translation models such as Gaussian blur, which have been commonly used in previous studies(Zeng *et al.*, 2018; Dong *et al.*, 2014; Zheng *et al.*, 2020). These models simplify the degradation process; thus, they form unreal features in the training LR images(Chun *et al.*, 2019). These unrealities cause the gap between the training LR and real LR domains. Given the limited generalizability of current SR models, this domain gap could cause the degradation of SR performance on real LR images(Lei *et al.*, 2020). Although good results have been observed with simulated LR images, the actual performance on real LR images has been overlooked. In summary, the clinically applicable SR methods should address blind SR problem, which means SR with unknown downsampling kernel.

Internal statistics of natural images have been reported in the computer vision (CV) community(Zontak and Irani, 2011; Michaeli and Irani, 2013). Patches extracted from a natural image tend to recur much more frequently inside the same image. Thus, the patches from a single image could follow a specific distribution. Due to the various contents of natural images, the internal statistics is often limited to a single image. In medical imaging, however, the patches from a series of scans could follow a specific distribution because the imaging objects are similar and the scanning protocols are consistent. Inspired by this idea, the gap between training LR and real LR images can be narrowed by making training LR images simulate real LR images in terms of patch distribution. Some generative adversarial network(Goodfellow *et al.*, 2014) (GAN) variants have shown a potential for domain transfer(Isola *et al.*, 2017; Zhu *et al.*, 2017), which attempted to address the challenge of learning the distribution of target images.

Therefore, the GAN structure could be a good fit for LR data construction.

In this study, we aimed to present a blind MRI SR framework to achieve better enhancement on real LR MR images. In contrast to conventional simple translation models, we attempted to improve the training set construction. This method comprises two steps. (a) In training set construction, we trained a downsampling GAN with images from HR and real LR domains. The generator learned to translate HR to LR images that were closer to the real LR domain. A deep linear network was used as the generator. A fully convolutional network was used as the discriminator with a 7 × 7 receptive field. (b) In super-resolution training, we trained an enhanced deep super resolution network (EDSR) with the previously constructed dataset. Finally, the SR network was evaluated with real LR images.

2. Materials and Methods

2.A LR Data Construction

2.A.1. Downsampling GAN

The first step was to synthesize LR images from HR images in the absence of strictly matched LR/HR pairs. The synthetic LR images were expected to share the same features with the real LR images and match the input HR images in anatomy structure. We developed a GAN to accomplish this task (**Figure 2**). Once the generator was trained, it synthesized the LR images from the HR input. The discriminator distinguished between the patches from the synthetic and real LR images. During adversarial training, the generator maximized patch-level similarity between real and synthetic LR images.

Building upon the recent success of real-world SR in the CV community, we used the deep linear network(Bell-Kligler *et al.*, 2019; Ji *et al.*, 2020) as the generator. The first three layers are 7 × 7, 5 × 5, and 3 × 3 convolutions, followed by three 1 × 1 convolutions (**Figure 2(b)**). These layers form a receptive field of 13 × 13, constituting a downsampling kernel of 13 × 13. And regularization terms penalize the peripheral values of that kernel, which makes it resemble a low-pass filter. It’s natural to recognize that such a small kernel will not introduce unnecessary distortions to the synthetic LR images. Thus, the linear network was itself a constraint in preserving structural consistency. The non-linear activation layers were removed to reduce unnecessary distortions to the synthetic LR images. If the non-linear activation layers were incorporated into the architecture, the generator tried to cheat the discriminator by creating images containing real patches but with a distorted global structure.

The discriminator, illustrated in **Figure 2(c)**, is fully convolutional (also referred to as patchGAN in the style transfer task(Isola *et al.*, 2017)). The first layer is a 7 × 7 convolution followed by six 1 × 1 convolutions. Spectral normalization was used to improve the convergence. This structure made it possible to predict the fidelity of each 7 × 7 patch independently. The purpose for the independent discrimination of patches was to make the discriminator learn the patch distributions of real LR images, thus guiding the generator to synthesize LR images whose patches are indistinguishable from those of real LR images. Another benefit of the small receptive field was that the limitations on data acquisition could be relaxed; the training LR and HR images were not required to be strictly paired, since the discriminator did not recognize global features of the input images.

Edge maps were also incorporated into the GAN architecture. Edges can reflect changes in local intensity, thus showing the sharpness of the image(Yu *et al.*, 2019).

Since LR images generated by different methods vary in sharpness, the discrimination of edge maps can increase the power of the discriminator. The edge maps were computed using commonly used Sobel operators in horizontal and vertical directions:

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix},\ S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \qquad [1]$$

The operators were convolved with the output of the generator (synthetic LR) and the real LR images. Subsequently, the two maps were concatenated with the LR images as three channels before being fed into the discriminator. The discriminator outputted a probability map, where each pixel indicated the fidelity of a 7 × 7 sliding window of the concatenated triplet.

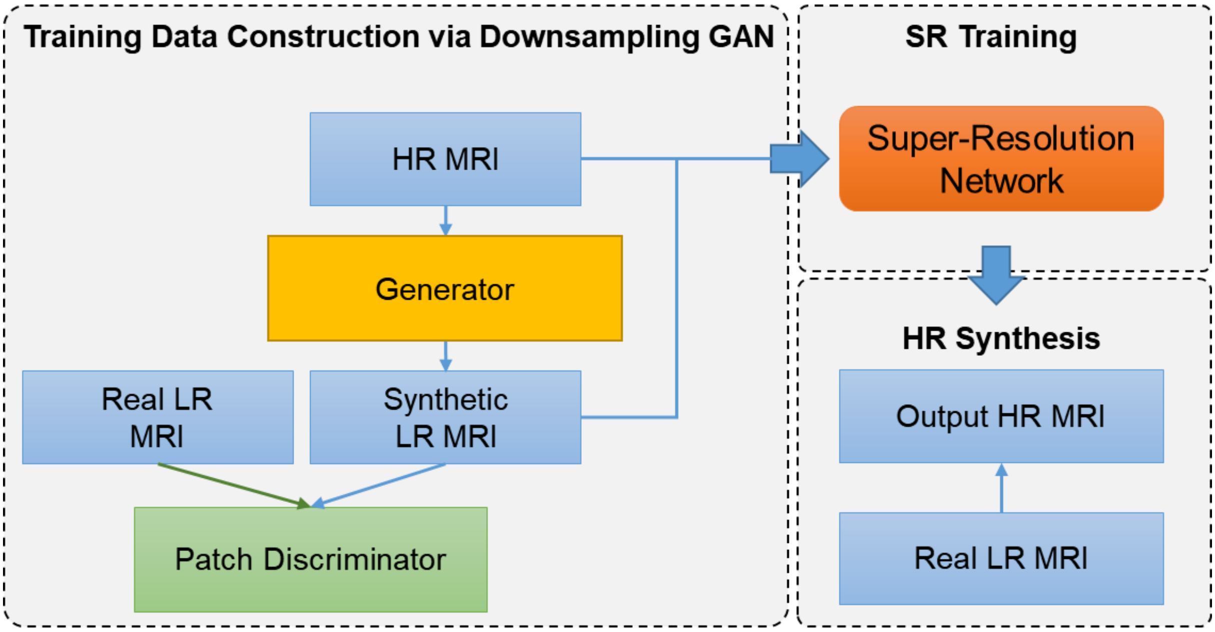

Figure 1. The overall framework of our proposed method. The major modification relative to the conventional supervised SR methods was made to the training set construction. A more realistic training set is prepared by the proposed downsampling GAN, after which the LR and HR pairs are fed into the SR network training.

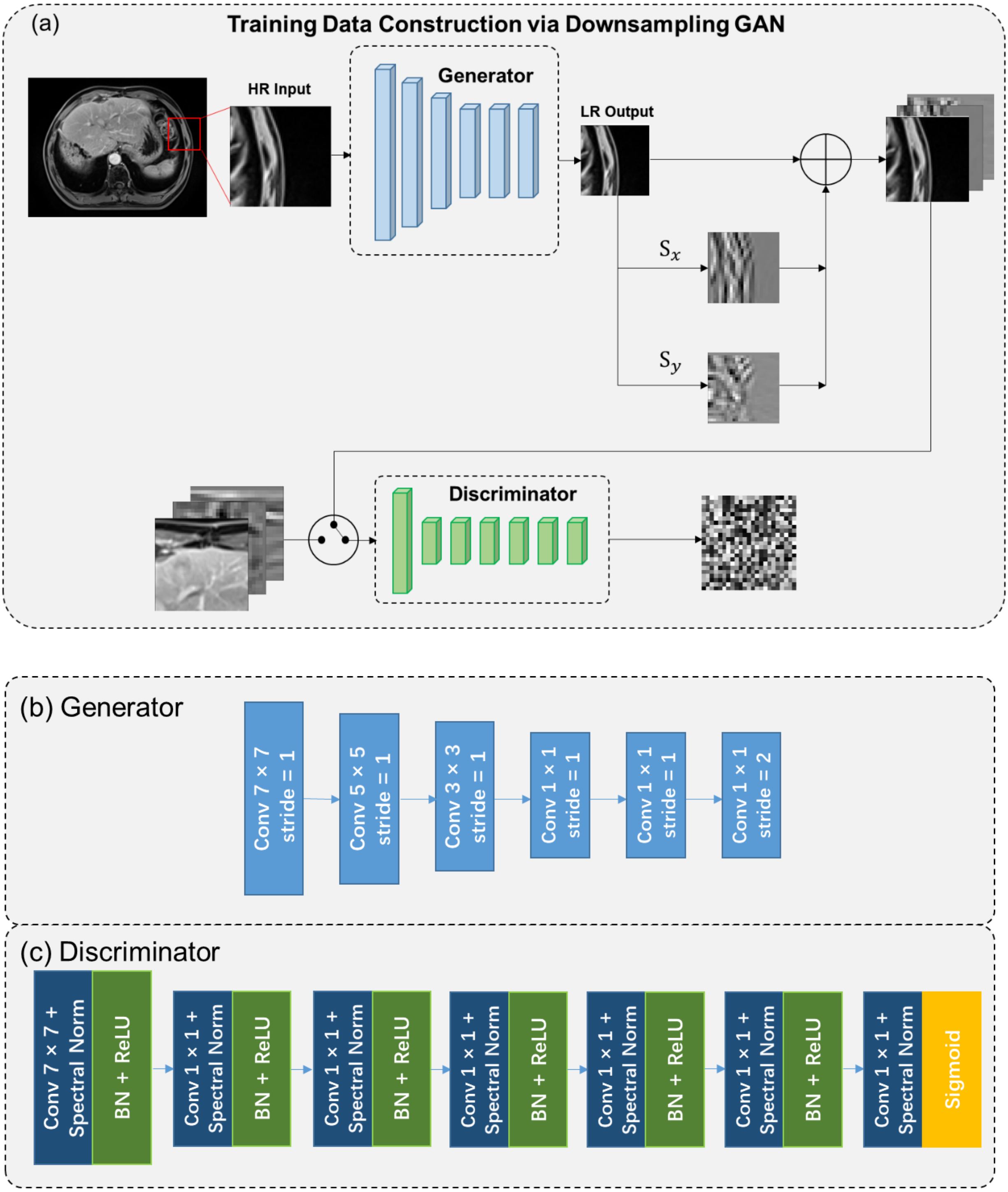

Figure 2. Framework and network architecture of the downsampling GAN. (a) shows the framework of adversarial training. Patches are randomly cropped from the HR images as the input to the generator in different iterations. The output synthetic LR images are convolved with Sobel filters in the horizontal and vertical directions. These maps are concatenated as three channels before being fed into the discriminator. The

architecture of the generator and discriminator is shown in (b) and (c). A deep linear network without a non-linear activation layer is the generator. The receptive field of the generator is 13 × 13, which is also the size of the downsampling kernel simulated by the generator. (c) shows the patchGAN structure of the discriminator. The discriminator outputs a probability map, in which each pixel indicates the fidelity of 7 × 7 patches from the input LR image.

2.A.2. Loss Function

These two networks were trained simultaneously in an adversarial manner by solving the following equation:

$$\underset{\theta_G}{argmin}\,\underset{\theta_D}{max}\left\{\mathbb{E}_{y\sim p_{lr}}\left[\left|D\big(y,S(y)\big)-1\right|\right]+\mathbb{E}_{x\sim p_{hr}}\left[\left|D\big(G(x),S(G(x))\big)\right|\right]+\mathcal{R}\right\},$$

[2]

where G and D denote the generator and the discriminator, respectively; x and y denote the HR and real LR images, respectively; $S(\cdot)$ denotes convolution with a Sobel filter; and $\mathcal{R}$ denotes the regularization term enforcing constraints on the parameters of the generator. The regularization term $\mathcal{R}$ can be formulated as:

$$\mathcal{R}=\lambda_{bicubic}|G(x)-I_{bicubic}|+\lambda_{sum2one}\left|1-\sum\nolimits_{i,j}k_{ij}\right|+$$

$$\lambda_{boundaries}\sum\nolimits_{i,j}\left|k_{i,j}\cdot m_{ij}\right|+\lambda_{sparse}\sum\nolimits_{i,j}\left|k_{i,j}\right|^{0.2}+\lambda_{center}\left\|(x_c,y_c)-\frac{\sum_{i,j}k_{i,j}\cdot(i,j)}{\sum_{i,j}k_{i,j}}\right\|_2, \quad [3]$$

where $k_{ij}=G(x_{delta})$ represents the generator response to a delta function x_{delta}.

1. $|G(x)-I_{bicubic}|$ is the downscale loss, in which the $I_{bicubic}$ denotes the LR images resized by bicubic interpolation. Minimizing this term encourages the

kernel to converge to a low-pass filter in the first several hundred epochs.

2. $\left|1 - \sum_{i,j} k_{ij}\right|$ encourages k to sum to 1.

3. $\sum_{i,j}\left|k_{i,j} \cdot m_{ij}\right|$ penalizes non-zero values close to the boundaries, where m is a mask with non-zero values on the periphery.

4. $\sum_{i,j}\left|k_{i,j}\right|^{0.2}$ encourages sparsity to prevent the network from over-smoothing.

5. $\left\|\left(x_{c,}y_{c}\right) - \frac{\sum_{i,j} k_{i,j}\cdot(i,j)}{\sum_{i,j} k_{i,j}}\right\|_2$ encourages k's center of mass to remain at the center of the kernel. x_c and y_c are the center indices.

2.A.2 Training Details of the Downsampling GAN

The training of a downsampling GAN requires real LR and HR images. The whole workflow was implemented in a 2D manner since the HR and LR images used in this study had the same slice thickness. The GAN was trained separately for each axial LR image. For each training, one LR and one HR image from the same slice location were input into the framework. The training comprised 3000 iterations. For each iteration, one 64 × 64 patch from the HR image and one 26 × 26 (size of the generator output) patch from the LR image were randomly cropped. The probability of patch selection was based on the magnitude of the gradient maps of the HR and LR images, as the regions with greater intensity variance better reflected the effects of low-pass filtration. The selection of LR and HR patches was independent, thus preparing unmatched LR and HR patches for each iteration.

The weights of the abovementioned 5 regularization terms were adjusted dynamically. At the beginning of the training, $\lambda_{bicubic} = 5, \lambda_{sum2one} = 0.5, \lambda_{boundries} = 0.5, \lambda_{center} = 0$, and $\lambda_{sparse} = 0$. Centralized loss and sparse loss were not applied, and bicubic loss was emphasized to accelerate the convergence to low-pass filtration. When the bicubic loss was lower than 0.4, $\lambda_{bicubic}$ started to decay at every 200 epochs with a decay rate of 0.01. Meanwhile, when $\lambda_{bicubic} < 0.005$, λ_{center} and λ_{sparse} grew to 1 and 5, respectively. The training was performed in PyTorch using a GeForce RTX 2080Ti GPU (NVIDIA, Santa Clara, CA, USA).

2.B. Super-Resolution Network

An EDSR was chosen as the SR network due to its state-of-the-art performance.(Lim *et al.*, 2017) An EDSR is a ResNet-based network that removes batch normalization to improve performance. Different from other popular architectures like U-Net, the EDSR architecture does not contain any downsampling block; thus, it focuses more on local features than on global features. This focus on local features also corresponds to maximizing the patch-level similarity of the synthetic LR and real LR when constructing the training data. Other networks (e.g., GAN-based networks with ResNets as the generator) can also be used in this method. The EDSR was trained by minimizing the mean absolute error (MAE) function, since optimizing L1 loss rather than L2 loss works better in SR tasks.(Zhao *et al.*, 2015; Lim *et al.*, 2017) During the training, the input LR images were cropped into 48 × 48, and the output HR patches were enlarged to 96 × 96 by a pixel shuffle module. The number of residual blocks was set to 16. The DL framework and hardware were the same as those in section 2.A.2.

2.C Experiments

2.C.1 Data

All data used in the following two experiments were collected via a 3T MR instrument (MAGNETOM Skyra, Siemens Healthineers) in Beijing Cancer Hospital, resulting in 1728 LR/HR pairs from 24 patients. The serial scans were performed in a single breath-hold, of which the HR scan took 11.0 s and the LR scan took 4.6 s. The LR images were registered to their HR counterparts by b-spline method. The imaging parameters are shown in **Table 1**.

Table 1. Imaging parameters.

	LR	HR
Sequence	TWIST-VIBE	TWIST-VIBE
Resolution	2.7 mm × 2.7 mm × 2.7 mm	1.25 mm × 1.25 mm × 2.7 mm
Field of view (FOV) (mm)	430	400
Slices per slab	72	72
TR (ms)	3.44	3.93
TE (ms)	1.23	1.26
Flip angle (°)	5	5

2.C.2 Comparison with Conventional Methods

This controlled experiment was designed to compare our proposed method with conventional methods that use gaussian blur and k-space zero-filling to construct data. We trained three EDSRs via three training data construction methods, including the proposed framework, gaussian blur and k-space zero-filling. Those EDSRs were evaluated by the same real LR images introduced above. Thus, the sole difference between the conventional and proposed frameworks was how the LR training data were constructed. A Gaussian kernel with a standard variance of 1.5 was selected because HR images downsampled using this kernel achieved the best similarity in both numerical metrics and visual comparison (**Figure 3**) with real LR images. The k-space-zero-filled LR images were generated by zero-filling the peripheral 75% components of k-space, after which a Fermi filter was used as the anti-ringing window.(Bernstein *et al.*, 2001)

The proposed framework was implemented with data acquired from serial LR and HR liver scans. It is worth noting that although the patients were instructed to hold their breath, motion still existed. Cardiac motion, intestinal peristalsis, and failure to maintain breath-hold were likely to cause mismatches between the LR and HR scans. The LR images were resized to 2.5 mm × 2.5 mm via bicubic interpolation, as the EDSR magnifies the size of input images by a ratio with an integer value. The slice thickness of all HR and LR images was 2.7 mm, making it valid to implement the workflow in a 2D manner. The LR images were denoised by non-local means(Buades *et al.*, 2011) (NLM) prior to the downsampling training because the texture of the noise could destabilize the downsampling training and degrade the SR performance(Zhao *et al.*, 2020). All of the image pairs were used in the downsampling training. We used eight rotations, where $\theta = n\pi/4$ and $n = 0, \dots, 7$, to augment the training data for the subsequent training of EDSR. Five-fold cross-validation was performed to ensure robustness.

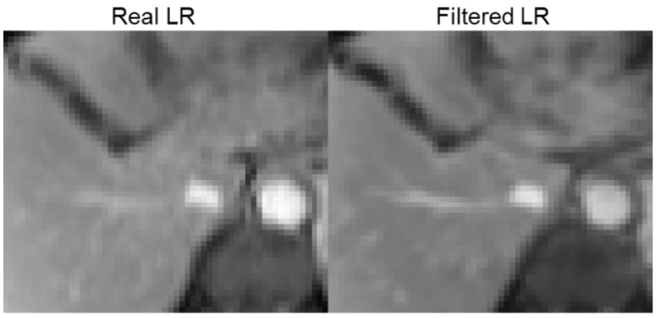

Figure 3. Visual comparison of the Gaussian blurred ($\sigma = 1.5$) and real LR images. The filtered LR image on the left denotes the training LR images used in the baseline training of the EDSR.

2.C.3 Comparison with a Blind SR Method

Prior to this study, a relevant study(Chun *et al.*, 2019) also emphasized data construction. The authors proposed a cascaded framework consisting of a denoising autoencoder (DAE), a downsampling network (DSN), and a super-resolution generative (SRG) model. The resulting fully convolutional DSN served the same purpose as our downsampling GAN. However, in contrast to the adversarial training developed in our study, the authors used a simple pixel-wise loss to train their DSN. This experiment was designed for this comparison. The denoising was performed analytically without using a DAE. And the SRG was also replaced by our EDSR network to make the experiment a fair comparison between our proposed downsampling GAN and the DSN.

3. Results

3.A. Comparison with Conventional Methods

Figure 4 shows the comparison of the SR output between our proposed framework and conventional frameworks with Gaussian blurred LR and k-space-zero-filled LR. The EDSR trained with Gaussian-blurred LR images produced a considerable amount of over-sharpening artifacts, and the network trained with k-space-zero-filled LR delivered almost no improvement compared with the HR images resized by bicubic interpolation. However, the EDSR from our proposed framework demonstrated fewer artifacts and better resolution enhancement on real LR images. SSIM and PSNR values were also presented in **Figure 4**. The absolute value was relatively low due to the inevitable mismatches between the LR and HR scans (the mismatches will be shown in **Figure 6** in the next section). Besides the low numeric value, SSIM and PSNR did not fully reflect the opinion from human observers in the comparison between ‘Bicubic’, ‘ZF + EDSR’, and ‘DSGAN + EDSR’, as they are approximately equal. But the artifacts in the ‘GB + EDSR’ group were successfully reflected by SSIM and PSNR. The SSIM and PSNR values of ‘GB + EDSR’ and ‘DSGAN + EDSR’ were further evaluated by Student’s t test. The p-values were less than 0.05, which indicated a significant difference in the image quality between these two groups.

Since PSNR and SSIM could not fully reflect the image quality, we also introduced the Blind/Reference-less Image Spatial Quality Evaluator(Mittal *et al.*, 2012; Chow and Rajagopal, 2017) (BRISQUE) metric to evaluate performance in the absence of the ground truth. This evaluator is a regression module trained by a support vector machine (SVM) regressor (SVR)(Schölkopf *et al.*, 2000). This regression module maps the statistical features of mean subtracted contrast normalized (MSCN) coefficients to a quality score. A total of 900 labeled MR images consisting of reference HR images, real LR images, and images distorted by Rician noise, Gaussian noise, Gaussian blur, and discrete cosine transform were used in the training. The label of the training images was the differential mean opinion score (DMOS) obtained from human subjects.

In order to investigate the intermediate results of LR construction, we used the MSCN coefficients from the BRISQUE metric to investigate the difference between the statistical features of the proposed LR, gaussian blurred LR, and real LR images. MSCN coefficients were calculated for the above three groups using the following equations:

$$\hat{I}(m,n) = \frac{I(m,n) - \mu(m,n)}{\sigma(m,n) + C}, m \epsilon 1,2, \ldots, M, n \epsilon 1,2, \ldots, N, \qquad [4]$$

where

$$\mu(m,n) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} I_{k,l}(m,n), \quad [5]$$

$$\sigma(m,n) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} (I_{k,l}(m,n) - \mu(m,n))^2}. \quad [6]$$

$I(m,n)$ is the intensity image, $\mu(m,n)$ is the local mean, and $\sigma(m,n)$ is the local variance. M and N represent the size of the images. $C = 1$ is introduced to avoid a zero denominator. $w_{k,l}, k = -3, \dots, 3, l = -3, \dots, 3$ is a 2D Gaussian weighting function with a standard variance of 1 pixel. The distribution of $\hat{I}(m,n)$ and pairwise products of adjacent MSCN coefficients $D(m,n) = \hat{I}(m,n)\hat{I}(m-1,n-1)$ are shown in **Figure 6**. The distribution of our synthetic LR images is denoted by the dashed line. This dashed line is closer to the solid line than the dotted line, implying that our synthetic LR images were closer to the real LR images in terms of the distribution of MSCN coefficients.

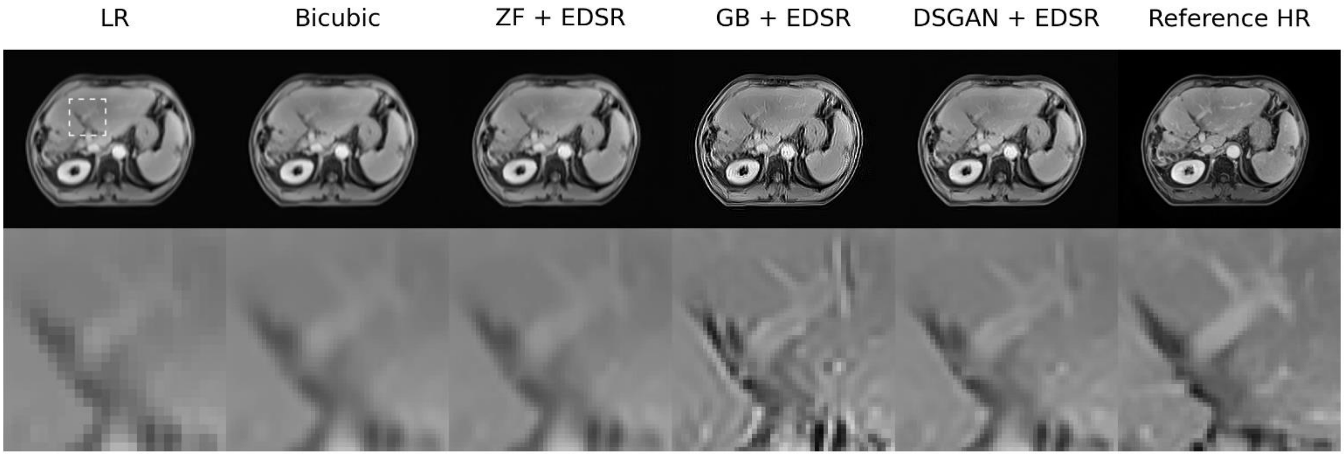

Figure 4. Output of EDSRs trained with different data. LR: the input LR images to the EDSR. Bicubic: bicubic interpolation used to resize the image directly. ZF + EDSR: EDSR trained with LR images prepared by k-space zero-filling. GB + EDSR: EDSR trained with Gaussian blurred images. DSGAN + EDSR: LR generation and EDSR

training using the proposed method. Reference HR: images from the corresponding HR scans.

Table 2. Quantitative evaluations of the proposed method and conventional methods. The abbreviations here are the same as those in **Figure 4**.

	Bicubic	ZF + EDSR	GB + EDSR	DSGAN + EDSR
BRISQUE	58.0 ± 2.5	54.2 ± 3.4	46.6 ± 4.2	34.1 ± 2.4
SSIM	0.605 ± 0.066	0.601 ± 0.065	0.516 ± 0.045	0.627 ± 0.050
PSNR	22.2 ± 0.94	22.2 ± 0.91	19.5 ± 0.94	22.3 ± 0.98

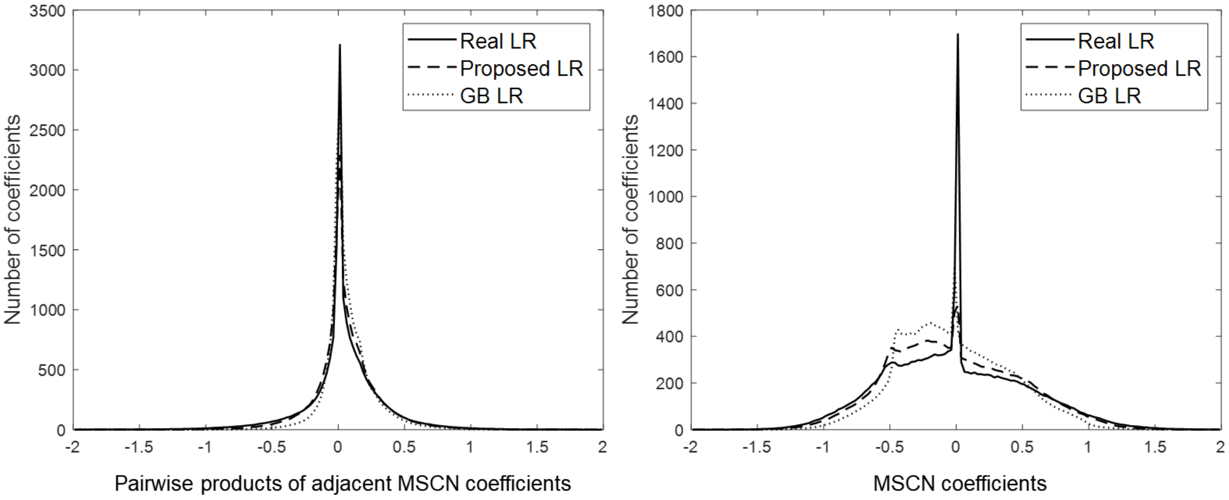

Figure 5. The distribution of MSCN coefficients and pairwise products of adjacent MSCN coefficients of real LR images, proposed synthetic LR images, and Gaussian blurred LR images.

3.B. Comparison with a Blind SR Method

The mismatches between real LR and HR images were significant, which is shown by 1D profile in **Figure 6(a)**. The DSN training guided by pixel-wise loss was presumably unsuccessful. First, as in the 1D profile presented, the LR images generated by DSN did not match well with the HR images in pixel value. By comparison, the proposed downsampling GAN generated LR images matched well with the HR ones as a result of the fully linear generator. Second, there were some structures missing in the output LR images (**Figure 6(b)**). Those missing structures would encourage the SR network to learn something from nowhere. The final SR results proved the failure of DSN training. In **Figure 7**, the combination of DSN and EDSR delivered severe checkboard artifacts and nearly zero enhancement in details.

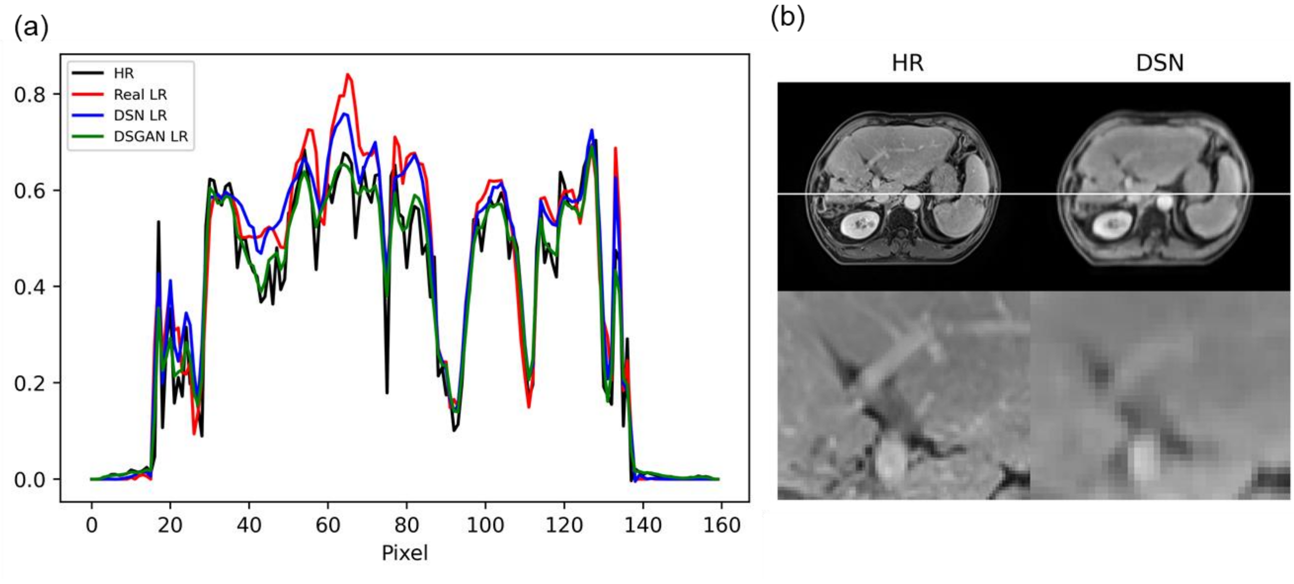

Figure 6. (a) shows the 1D profile (white line in (b)) of HR and different LR images. The real LR images (Real LR) and the output LR images of DSN (DSN LR) have significant mismatches with the HR images. However, the LR images generated by downsampling GAN (DSGAN LR) matched well with the HR images. (b) shows the input HR image to the DSN and the corresponding output. Some details are missing and the output images are over-blurred.

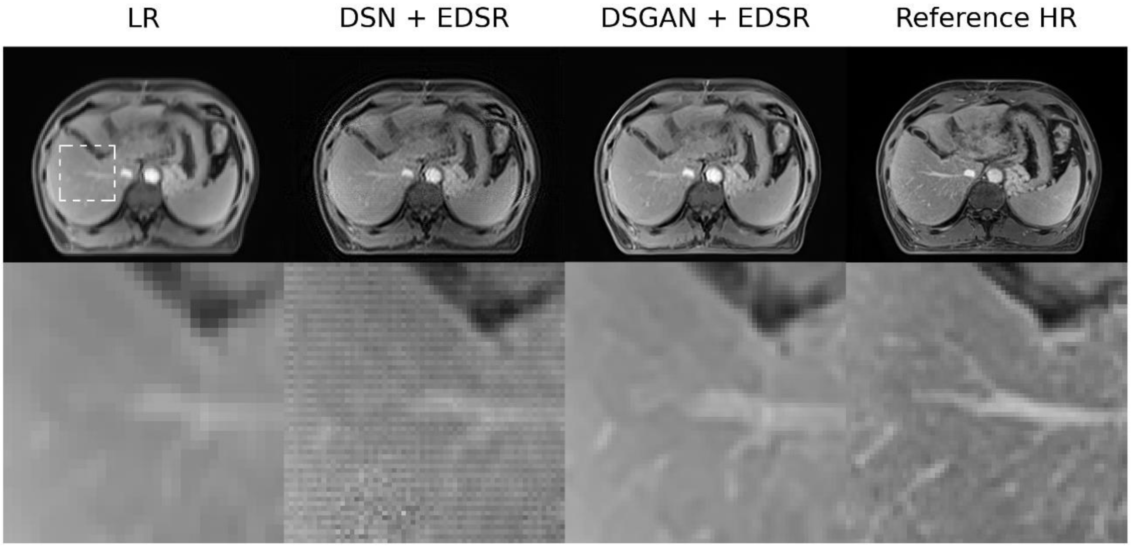

Figure 7. Comparison of the SR results of the relevant blind SR method (DSN + EDSR) and our proposed method (DSGAN + EDSR). The combination of DSN and EDSR produced significant checkboard artifacts, reflecting the failure of DSN training.

Table 3. Quantitative evaluations of the proposed method and another blind SR method.

	DSN + EDSR	DSGAN + EDSR
BRISQUE	85.4 ± 5.1	34.1 ± 2.4
SSIM	0.489 ± 0.051	0.627 ± 0.050
PSNR	21.7 ± 0.93	22.3 ± 0.98

4. Discussion

In this study, we proposed a novel blind SR method for MR images and compared it with conventional methods and a relevant blind SR framework. The main innovation of this two-step workflow can be summarized as follows: (a) we proposed a method to solve the blind SR challenge, which means SR with an unknown downsampling kernel. In contrast to other SR methods for MRI, our proposed method focused on achieving better enhancement on real LR images other than those generated by simple translation models. (b) We developed a novel GAN architecture to construct the data. A deep linear generator was used to perform the downsampling. The discriminator was designed to have a patch-level receptive field. This GAN maximized the patch-level similarity between synthetic LR and real LR images. (c) We relaxed the limitation on data acquisition. Since the discrimination was on the patch level, it was not necessary to use strictly matched LR/HR pairs, which are nearly impossible to acquire in a clinical setting. During the acquisition of serial LR and HR images, we relaxed the limitation on structural consistency between HR and LR images to make the process easier to implement.

Most previous MRI SR studies introduced data construction without solid interpretation. Simple translation models, including Gaussian blur(Shi *et al.*, 2018b), k-space zero-filling(Lyu *et al.*, 2020), and bicubic interpolation(Lim *et al.*, 2017), have been reported. These different models construct LR data containing disparate features; however, few studies have justified the reason for selecting a specific model and the related parameters. For Gaussian blur, for example, 1 voxel is the commonly chosen standard variance. However, the LR images generated by this Gaussian kernel were much clearer and sharper than our real LR images.

This study also raises attention to the limited generalizability of current SR network and the gap between different data in SR tasks. Though deep learning has led to remarkable improvement in the SR tasks of medical imaging, the models still suffer from poor generalizability. The gap not only exists between real LR images and LR images generated by simple models, as investigated in this study, but also exists between different clinical datasets. The features of real LR images can vary with different protocols, magnetic field strengths, and MRI scanners. Therefore, the excellent performance of the state-of-the-art network cannot be guaranteed when using different data. For clinical applications, we should focus on the blind SR problem that enhances real LR images. In this study, rather than improving the generalizability of the SR model, we chose to solve the blind SR problem by fitting the model to clinical data by data construction. Further study is needed to quantitatively investigate and improve the generalizability of the SR networks.

In the comparison with the relevant method, we demonstrated how mismatches between LR and HR images could impact the SR training. The pixel-wise loss required strong structural consistency between the LR and HR images, which is not always available in clinical applications. Although our patients were instructed to hold their breath, mismatches in structure and contrast inevitably existed due to motion and different imaging protocols, which are common in clinical practice. The non-linear structure of DSN provided unnecessary freedom to the downsampling process, which resulted in the missing of important details and the intensity mismatches. Our solution to this problem was proved to be successful. The linear structure of the generator enforced constraints on the downsampling process. And the pixel-wise loss was replaced by adversarial loss, which learned the patch distribution of the generator's output. The small receptive field of the discriminator was a key to overcoming the mismatches.

The clinical potential of this framework lies in reducing the scanning time for HR MRI. This framework can improve the image quality for patients who have difficulties in holding their breath during a standard HR scan. Taking the data used in this study as an example, the HR scan took 11.0 s and the LR scan took 4.6 s. 4.6 s is a more comfortable period for patients who are physically incapable of holding their breath over 10s. It can also be used to enhance the resolution of standard HR scan. Higher resolution scans can be performed on healthy volunteers who are able to hold their breath for longer time. Thus, a SR network can be trained to reconstruct the higher resolution version of HR scans. Apart from this, real time imaging is another potential scenario for this framework. Once the training is finished, the time HR reconstruction costs will be negligible compared with iterative reconstruction algorithms. Reduced scanning time and fast reconstruction make it a potential solution in real time imaging.

The underlying features of LR images were somewhat elusive, as demonstrated in the experiment described in section 2.C.3: the SR network trained with Gaussian-blurred LR data delivered severe artifacts with real LR data, but the training and testing LR images were almost indistinguishable in a visual comparison. Therefore, we introduced the discriminator to read the patches of LR images and expected the generator to learn the patch distribution automatically during the adversarial training. The experiment in section 2.C.3 demonstrated that our downsampling GAN was able to construct more realistic LR images. However, the improvement in data construction was validated to a greater extent by the final outcome of the SR network than by the intermediate results of LR construction. Though we calculated the MSCN coefficients for the LR images, this statistical method may not reflect the features seen by the neural network. We still lack a method to further investigate these features in different LR images and interpret the disparity in SR performance.

The mismatches between LR/HR pairs were allowed due to the novel network design and served as the root for the relatively low SSIM and PSNR on the other hand. These mismatches cause difficulties in numerical evaluation since SSIM and PSNR measure the pixel-wise difference. Thus, those pixel-wise differences may not able to fully quantify the image quality. More data with better structural consistency between LR and HR images are needed to facilitate evaluations in future.

In this framework, we developed a 2D network to change only the in-plane resolution. For scenarios where the resolutions of HR and LR images are different in both in- and through-plane directions, this method may not be applicable because HR images need to be downsampled in 3D, which introduces more instability to the training. For the subsequent SR training, much more data will be required and the training will be difficult. In future, this downsampling GAN could be extended to 3D when more data are collected. The 3D version is expected to learn the 3D patch distribution. Besides extending to 3D, the current 2D method can also be combined with some dedicated methods(Zhao *et al.*, 2020; Chaudhari *et al.*, 2018) focusing on through-plane resolution enhancement; here, in-plane SR with the current framework will be the first step prior to through-plane SR.

5. Conclusion

In this study, we proposed a novel SR framework to reconstruct HR images from clinical LR images. Compared with simple translation models like Gaussian blur and k-space zero-filling, our proposed downsampling GAN was better able to synthesize LR images that matched real LR images. Our method outperformed the conventional methods by demonstrating better resolution enhancement and fewer artifacts on real LR MR images. Compared with another blind SR method whose downsampling training was driven by pixel-wise loss, this framework had better performance on the training data with mismatches.

Acknowledgement

This research was partly supported by the following research grants: GRF 151021/18M, GRF 151022/19M, HMRF 07183266, and NIH R01 CA226899.

References:

Bell-Kligler S, Shocher A and Irani M *Advances in Neural Information Processing Systems,2019)*, vol. Series*)* pp 284-93

Bernstein M A, Fain S B and Riederer S J 2001 Effect of windowing and zero‐filled reconstruction of MRI data on spatial resolution and acquisition strategy *Journal of magnetic resonance imaging: an official journal of the international society for magnetic resonance in medicine* **14** 270-80

Buades A, Coll B and Morel J-M 2011 Non-local means denoising *Image Processing On Line* **1** 208-12

Cai J, Read P W, Larner J M, Jones D R, Benedict S H and Sheng K 2008 Reproducibility of interfraction lung motion probability distribution function using dynamic MRI: statistical analysis *International Journal of Radiation Oncology* Biology* Physics* **72** 1228-35

Chaudhari A S, Fang Z, Kogan F, Wood J, Stevens K J, Gibbons E K, Lee J H, Gold G E and Hargreaves B A 2018 Super‐resolution musculoskeletal MRI using deep learning *Magnetic resonance in medicine* **80** 2139-54

Chow L S and Rajagopal H 2017 Modified-BRISQUE as no reference image quality assessment for structural MR images *Magnetic resonance imaging* **43** 74-87

Chun J, Zhang H, Gach H M, Olberg S, Mazur T, Green O, Kim T, Kim H, Kim J S and Mutic S 2019 MRI super‐resolution reconstruction for MRI‐guided adaptive radiotherapy using cascaded deep learning: In the presence of limited training data and unknown translation model *Medical physics* **46** 4148-64

Dong C, Loy C C, He K and Tang X *European conference on computer vision,2014)*, vol. Series*)*: Springer) pp 184-99

Dong C, Loy C C, He K and Tang X 2015 Image super-resolution using deep convolutional networks *IEEE transactions on pattern analysis and machine intelligence* **38** 295-307

Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial networks *arXiv preprint arXiv:1406.2661*

Isola P, Zhu J-Y, Zhou T and Efros A A *Proceedings of the IEEE conference on computer vision and pattern recognition,2017)*, vol. Series*)* pp 1125-34

Jafari-Khouzani K 2014 MRI upsampling using feature-based nonlocal means approach *IEEE transactions on medical imaging* **33** 1969-85

Ji X, Cao Y, Tai Y, Wang C, Li J and Huang F *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops,2020)*, vol. Series*)* pp 466-7

Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J and Wang Z *Proceedings of the IEEE conference on computer vision and pattern recognition,2017)*, vol. Series*)* pp 4681-90

Lehmann T M, Gonner C and Spitzer K 1999 Survey: Interpolation methods in medical image processing *IEEE transactions on medical imaging* **18** 1049-75

Lei K, Mardani M, Pauly J M and Vasanawala S S 2020 Wasserstein GANs for MR imaging: from paired to unpaired training *IEEE transactions on medical imaging* **40** 105-15

Lim B, Son S, Kim H, Nah S and Mu Lee K *Proceedings of the IEEE conference on computer vision and pattern recognition workshops,2017)*, vol. Series*)* pp 136-44

Liu Y, Yin F F, Chen N k, Chu M L and Cai J 2015 Four dimensional magnetic resonance imaging with retrospective k‐space reordering: A feasibility study *Medical physics* **42** 534-41

Luo J, Mou Z, Qin B, Li W, Yang F, Robini M and Zhu Y 2017 Fast single image super-resolution using estimated low-frequency k-space data in MRI *Magnetic resonance imaging* **40** 1-11

Lyu Q, Shan H and Wang G 2020 MRI super-resolution with ensemble learning and complementary priors *IEEE Transactions on Computational Imaging* **6** 615-24

Lyu Q, You C, Shan H, Zhang Y and Wang G *Developments in X-Ray Tomography XII,2019)*, vol. Series 11113*)*: International Society for Optics and Photonics) p 111130X

Manjón J V, Coupé P, Buades A, Fonov V, Collins D L and Robles M 2010 Non-local MRI upsampling *Medical image analysis* **14** 784-92

Michaeli T and Irani M *Proceedings of the IEEE International Conference on Computer Vision,2013)*, vol. Series*)* pp 945-52

Mittal A, Moorthy A K and Bovik A C 2012 No-reference image quality assessment in the spatial domain *IEEE Transactions on image processing* **21** 4695-708

Pham C-H, Ducournau A, Fablet R and Rousseau F *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017),2017),* vol. Series*)*: IEEE) pp 197-200

Plenge E, Poot D H, Bernsen M, Kotek G, Houston G, Wielopolski P, van der Weerd L, Niessen W J and Meijering E 2012 Super‐resolution methods in MRI: can they improve the trade‐off between resolution, signal‐to‐noise ratio, and acquisition time? *Magnetic resonance in medicine* **68** 1983-93

Rueda A, Malpica N and Romero E 2013 Single-image super-resolution of brain MR images using overcomplete dictionaries *Medical image analysis* **17** 113-32

Schölkopf B, Smola A J, Williamson R C and Bartlett P L 2000 New Support Vector Algorithms *Neural Comput.* **12** 1207–45

Shi F, Cheng J, Wang L, Yap P-T and Shen D 2015 LRTV: MR image super-resolution with low-rank and total variation regularizations *IEEE transactions on medical imaging* **34** 2459-66

Shi J, Li Z, Ying S, Wang C, Liu Q, Zhang Q and Yan P 2018a MR image super-resolution via wide residual networks with fixed skip connection *IEEE journal of biomedical and health informatics* **23** 1129-40

Shi J, Liu Q, Wang C, Zhang Q, Ying S and Xu H 2018b Super-resolution reconstruction of MR image with a novel residual learning network algorithm *Physics in Medicine & Biology* **63** 085011

Tai Y, Yang J and Liu X *Proceedings of the IEEE conference on computer vision and pattern recognition,2017),* vol. Series*)* pp 3147-55

Tourbier S, Bresson X, Hagmann P, Thiran J-P, Meuli R and Cuadra M B 2015 An efficient total variation algorithm for super-resolution in fetal brain MRI with adaptive regularization *NeuroImage* **118** 584-97

Van Reeth E, Tham I W, Tan C H and Poh C L 2012 Super‐resolution in magnetic resonance imaging: a review *Concepts in Magnetic Resonance Part A* **40** 306-25

Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y and Change Loy C *Proceedings of the European Conference on Computer Vision (ECCV) Workshops,2018),* vol. Series*)* pp 63-79

Wang Z, Chen J and Hoi S C 2020 Deep learning for image super-resolution: A survey *IEEE transactions on pattern analysis and machine intelligence*

Yang J, Cai J, Wang H, Chang Z, Czito B G, Bashir M R, Palta M and Yin F-F 2014 Is diaphragm motion a good surrogate for liver tumor motion? *International Journal of Radiation Oncology* Biology* Physics* **90** 952-8

Yu B, Zhou L, Wang L, Shi Y, Fripp J and Bourgeat P 2019 Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis *IEEE transactions on medical imaging* **38** 1750-62

Zeng K, Zheng H, Cai C, Yang Y, Zhang K and Chen Z 2018 Simultaneous single-and multi-contrast super-resolution for brain MRI images based on a convolutional neural network *Computers in biology and medicine* **99** 133-41

Zhang D, He J, Zhao Y and Du M 2015 MR image super-resolution reconstruction using sparse representation, nonlocal similarity and sparse derivative prior *Computers in biology and medicine* **58** 130-45

Zhao C, Dewey B E, Pham D L, Calabresi P A, Reich D S and Prince J L 2020 SMORE: A Self-supervised Anti-aliasing and Super-resolution Algorithm for MRI Using Deep Learning *IEEE transactions on medical imaging* **40** 805-17

Zhao H, Gallo O, Frosio I and Kautz J 2015 Loss functions for neural networks for image processing *arXiv preprint arXiv:1511.08861*

Zheng Y, Zhen B, Chen A, Qi F, Hao X and Qiu B 2020 A hybrid convolutional neural network for super‐resolution reconstruction of MR images *Medical physics* **47** 3013-22

Zhu J-Y, Park T, Isola P and Efros A A *Proceedings of the IEEE international conference on computer vision,2017),* vol. Series*)* pp 2223-32

Zontak M and Irani M *CVPR 2011,2011),* vol. Series*)*: IEEE) pp 977-84