

MRI Super-Resolution With Ensemble Learning and Complementary Priors

Qing Lyu^{ID}, Student Member, IEEE, Hongming Shan^{ID}, Member, IEEE, and Ge Wang^{ID}, Fellow, IEEE

Abstract—Magnetic resonance imaging (MRI) is a widely used medical imaging modality. However, due to the limitations in hardware, scan time, and throughput, it is often clinically challenging to obtain high-quality MR images. The super-resolution approach is potentially promising to improve MR image quality without any hardware upgrade. In this article, we propose an ensemble learning and deep learning framework for MR image super-resolution. In our study, we first enlarged low resolution images using five commonly used super-resolution algorithms and obtained differentially enlarged image datasets with complementary priors. Then, a generative adversarial network (GAN) is trained with each dataset to generate super-resolution MR images. Finally, another GAN is used for ensemble learning that synergizes the outputs of GANs into the final MR super-resolution images. According to our results, the ensemble learning results outperform any single GAN output component. Compared with some state-of-the-art deep learning-based super-resolution methods, our approach is advantageous in suppressing artifacts and keeping more image details.

Index Terms—Deep learning, ensemble learning, generative adversarial network (GAN), magnetic resonance imaging (MRI), super-resolution.

I. INTRODUCTION

MAGNETIC resonance imaging (MRI) is one of the most commonly used medical imaging modalities over the world. It noninvasively depicts structural and functional features inside a patient in rich contrasts. Compared with other medical imaging modalities such as computed tomography (CT) and nuclear imaging, MRI does not involve ionizing radiation. Furthermore, MRI plays a dominating role in neurological imaging especially brain research.

However, the strength and homogeneity of the MRI background field are both limited, and can be only improved at a high engineering cost. Also, physiological motion introduces image blurring and fundamentally limits achievable spatial resolution. Hence, the pixel size of a clinical MR image is often constrained to an order of millimeters. To obtain higher resolution MR images, either a more complicated MRI system with stronger

Manuscript received June 15, 2019; revised September 25, 2019 and December 6, 2019; accepted December 6, 2019. Date of publication January 6, 2020; date of current version January 23, 2020. This work was supported in part by NIH/NCI under Award numbers R01CA233888 and R01CA237267, and in part by NIH/NIBIB under Award number R01EB026646. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mathews Jacob. (*Corresponding author: Ge Wang.*)

The authors are with the Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: lyuq@rpi.edu; shanh@rpi.edu; wangg6@rpi.edu).

Digital Object Identifier 10.1109/TCI.2020.2964201

background magnetic field or a longer scan time needs to be used, which will dramatically increase expenses and is rarely clinically applicable. To address this problem, the super-resolution (SR) approach holds a great promise that needs to change neither hardware nor scanning protocol [1], [2].

SR has been a hot topic in the computational vision field over the past decades. A large number of methods were proposed to improve image resolution retrospectively. Roughly speaking, super-resolution algorithms can be divided into the two categories: model-based and learning-based. Interpolation algorithms [3], [4], like bilinear, bicubic, and nearest neighbor interpolation techniques, are representatives of simple model-based approaches, which can be directly used to enlarge images. Wiener filtering and iterative deblurring algorithms utilize knowledge on the system point spread function (PSF) to recover image resolution [5], [6], and are also considered as model-based algorithms. Dictionary learning-based super-resolution techniques [7]–[9] are examples of learning-based algorithms.

Recently, with the rapid evolution of machine learning, especially deep learning, deep neural networks have become popular among SR studies. In this context, diverse neural network architectures were designed and tested. Successful architectures include convolutional neural networks (CNNs) [10]–[16], generative adversarial networks (GANs) [17]–[20], residual neural networks (ResNets) [21]–[24], and recurrent neural networks (RNNs) [25]–[27]. Specifically, for MRI SR studies, models based on CNN, ResNet, and GAN were reported [28]–[33]. Although there were a plenty of decent SR results in the literature, bothersome artifacts or distortions were often observed in the SR results produced by these methods. Therefore, the challenge remains for achieving superior SR on MR images. In other words, MRI SR research must deliver images with faithful details and little artifacts in order to be clinically relevant. Since the existing methods cannot perfectly satisfy such a requirement, it is necessary to develop new SR methods based on all encouraging results.

Among important deep learning strategies, ensemble deep learning is a meta-method for further enhancement of deep learning results [34], [35]. By combining multiple models, ensemble learning is expected to produce synergistic results better than results produced by any single GAN model. Merits of ensemble learning were successfully demonstrated to improve image quality [36], [37]. In this MRI SR study, by combining multiple GAN SR results trained on different image prior datasets, we demonstrate better MRI SR image quality than that associated with any single model. In this paper, we describe a GAN-based

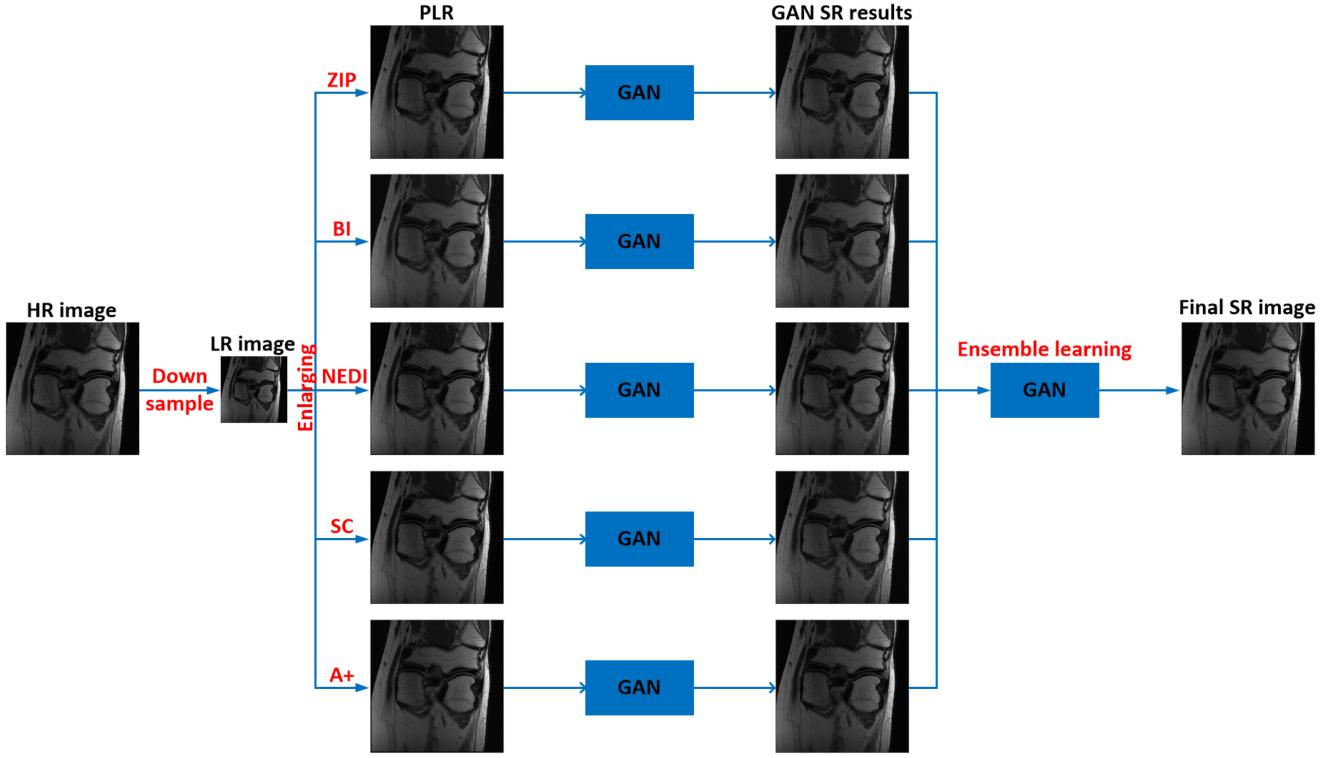


Fig. 1. Overall workflow of the SR method proposed in this paper.

deep ensemble learning framework to integrate complementary SR results from individual GAN models. Our method produces better image quality than state-of-the-art SR models, being capable of detailing textures and avoiding artifacts.

II. METHODOLOGY

Assuming that $I_{LR} \in \mathbb{R}^{p \times q}$ is a low-resolution (LR) MR image of the size $p \times q$, and $I_{HR} \in \mathbb{R}^{m \times n}$ the corresponding high-resolution (HR) MR image of the size $m \times n$. The relationship between them can be expressed as

$$I_{LR} = \|f(e^{iP_{HR}} I_{HR})\|, \quad (1)$$

where $P_{HR} \in \mathbb{R}^{m \times n}$ is the phase term, $f : I_{HR} \in \mathbb{R}^{m \times n} \rightarrow I_{LR} \in \mathbb{R}^{p \times q}$ denotes a complex-to-complex mapping which is a k -space truncating/down-sampling/blurring process to create an LR counterpart from an HR image. The overall SR process is to implement an approximate inverse function $g \approx f^{-1}$

$$g \approx f^{-1} : I_{HR} \approx g(I_{LR}), \quad (2)$$

assuming that the phase term is zero, $P_{HR} = 0$, without loss of generality.

As an ill-posed problem, it is impossible to find this inverse function exactly. Instead, priors can be used to acquire an approximation of f^{-1} . Multiple types of priors are used in this study to approximate f^{-1} as closely to the ideal function as possible.

Fig. 1 presents the proposed overall workflow of SR method. In our scheme, we first down-sample original HR images and obtain LR counterparts, and then use five different SR algorithms

to process LR images and obtain five types of processed LR (PLR) datasets. These PLR datasets are then separately fed into the corresponding GAN models for training and then to produce five GAN SR counterparts. The final step is to use another GAN for ensemble learning and obtain the final MRI SR result.

A. Down-Sampling

Differing from commonly used down-sampling strategy that down samples data in the image domain, we chose to down-sample data in the frequency domain. The original HR images were converted into the frequency domain via the 2D Fourier transform. Only the central 25% of the data points were kept in the k -space, while all peripheral data points were zeroed out. Then, the cropped dataset was converted back into a magnitude image through the 2D inverse Fourier transform as the LR image. Using this down-sampling method, we obtained LR images with 2-fold resolution degradation, which means $p = m/2$ and $q = n/2$.

B. Five Processing Algorithms to Enlarge LR Images

Before LR images are fed into the neural network, we adopt existing SR algorithms to enlarge a LR image to the size of the target HR image. This processing step was also used in some SR studies like VDSR [13] and SRCNN [11]. With such a processing step, the quality of enlarged images can be somehow improved so that it is easier for neural networks to extract image features and obtain better prediction results. Moreover, we can establish datasets with complementary priors through using multiple SR algorithms in this processing step. In this paper, in order

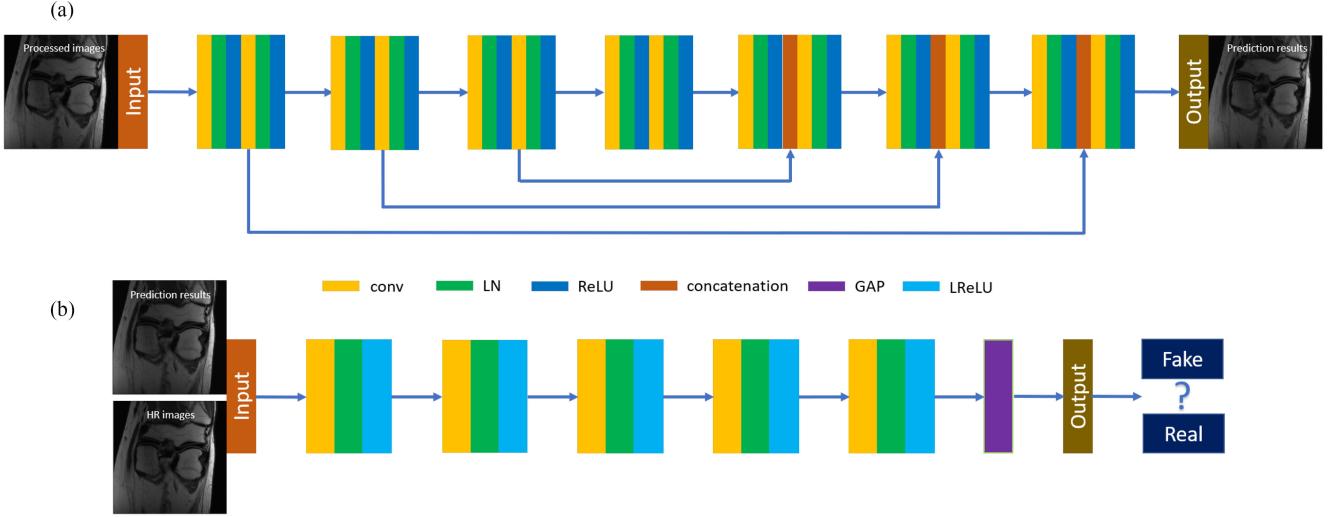


Fig. 2. The architecture of the proposed GAN. (a) The generator and (b) the discriminator.

to distinguish enlarged images with the latter neural network results, we call enlarged images processed LR images (PLR) rather than SR images. In this work, 5 classic SR algorithms were used to enlarge LR images and obtain the corresponding PLRs, which are zero interpolation filling (ZIP) [38], bicubic interpolation (BI) [3], new edge directed interpolation (NEDI) [4], sparse coding (SC) [7], and adjusted anchored neighborhood regression (A+) [9]. ZIP is a simplest super-resolution method that adds zeros in the peripheral region in the frequency domain. It is widely used to enlarge MR images. BI and NEDI are two improved interpolation-based algorithms, while SC and A+ are two dictionary-learning-based algorithms. In the A+ algorithm, 1,024 dictionary atoms were used, which were obtained from a training dataset of 0.5 million natural images. Similarly, the SC method also used a dictionary contained 1,024 atoms. Using these different SR algorithms, different enlarged images were created, each with its unique priors. Next, we used these enlarged images as the inputs to train the corresponding GAN models separately. Each enlarging/deblurring process can be expressed as

$$I_{PLR,i} = h_i(I_{LR}), \quad (3)$$

where $I_{PLR,i} \in \mathbb{R}^{m \times n}$ is a deblurred/enlarged image. $h_i : I_{LR} \in \mathbb{R}^{p \times q} \rightarrow I_{PLR,i} \in \mathbb{R}^{m \times n}$ denotes the processing procedure. As there were 5 SR algorithms used in this processing step, $i = 1, 2, \dots, 5$.

C. GAN Model

The GAN model with Wasserstein distance and gradient penalty (WGAN-GP) [39] is selected in this study. The generator used in this study is a modified version of the encoder-decoder network that was originally used for CT denoising [40]. The architectures of the generator and the discriminator are shown in Fig. 2. For the generator, it contains seven blocks with three skip connections. Each block involves convolution, layer normalization (LN), and ReLU activation operations sequentially. As mentioned in [39], batch normalization cannot be used in

TABLE I
STRUCTURAL DETAILS OF THE GENERATOR. NOTE THAT S1N1 MEANS 1 CONVOLUTIONAL KERNELS WITH A STRIDE OF 1

Block1	Block2	Block3	Block4	Block5	Block6	Block7
n32s1	n64s1	n128s1	n256s1	n128s1	n64s1	n32s1
n32s1	n64s1	n128s1	n256s1	n128s1	n64s1	n32s1

TABLE II
STRUCTURAL DETAILS OF THE DISCRIMINATOR. NOTE THAT S1N1 MEANS 1 CONVOLUTIONAL KERNELS WITH A STRIDE OF 1

Block1	Block2	Block3	Block4	Block5
n64s2	n128s2	n256s2	n512s1	n1s1

the WGAP-GP model, instead, we utilize layer normalization. With skip connections, feature maps in the previous blocks are concatenated to the feature maps in the latter blocks. Because of the structural similarity between LR and HR images, adding such skip connections can greatly boost the training process and keep the generated images structurally similar to the input LR images. Similarly, the discriminator consists of five blocks followed by a global average pooling (GAP) layer. Each block includes convolution, layer normalization, and leaky-ReLU (LReLU). Note that we replace the fully-connected layer with a GAP layer to reduce the amount of parameters in the neural network and facilitate the training process [41]. Each block in the discriminator performs convolution, layer normalization and ReLU activation function similar to what the block in the generator does. The details of the generator and the discriminator are listed in Table I and Table II.

The objective function of the generator is as follows:

$$\min_{\theta_G} \mathcal{L}_G = \mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{gra} + \lambda_2 \mathcal{L}_{mse} + \lambda_3 \mathcal{L}_{per}, \quad (4)$$

$$\mathcal{L}_{adv} = -\mathbb{E}_{I_{PLR}}[D(G(I_{PLR}))], \quad (5)$$

$$\mathcal{L}_{gra} = \mathbb{E}_{I_{PLR}, I_{HR}} \|\nabla[G(I_{PLR})] - \nabla(I_{HR})\|^2, \quad (6)$$

$$\mathcal{L}_{per} = \mathbb{E}_{I_{PLR}, I_{HR}} \|\Phi[G(I_{PLR})] - \Phi(I_{HR})\|^2, \quad (7)$$

where θ_G stands for the generator parameters. \mathcal{L}_{adv} is the adversarial loss. \mathcal{L}_{gra} is the gradient loss, which is the L2 norm difference between gradients of the generated image and its ground truth. ∇ stands for the gradient calculated in both x and y axis. \mathcal{L}_{mse} and \mathcal{L}_{per} are the mean squared error (MSE) and the perceptual loss respectively. MSE reflects the image content similarity between generated output images and processed input images at the pixel level, while the perceptual loss shows the similarity in a high-level feature space. We use the pre-trained VGG19 model to calculate the perceptual loss [40], [42], and Φ represents the feature map of the 16th convolutional layer in the VGG19 model. As the mean-squared-error loss tends to blur images, the gradient loss is introduced to penalize any blurring and enhance the edge. The objective function of the discriminator is the standard WGAN-GP equation shown in [39].

Overall, this GAN SR process is then expressed as

$$I_{SR,i} = G_i(I_{PLR,i}) = \varphi(I_{PLR,i}, \theta_{G_i}), \quad (8)$$

where $I_{SR,i} \in \mathbb{R}^{m \times n}$ is an output image from generator in GAN, $\varphi : I_{PLR,i} \in \mathbb{R}^{m \times n} \rightarrow I_{SR,i} \in \mathbb{R}^{m \times n}$ denotes a GAN deblurring process for an SR image, θ_{G_i} represents the parameters of a GAN model. As there are 5 SR algorithms used in this step, $i = 1, 2, \dots, 5$.

D. Ensemble Learning

Ensemble learning is to synergize diverse information sources for the best results. In this study, SR images individually obtained can be integrated via ensemble learning to improve SR MR image quality. Specifically, the final SR results are derived from different types of GAN SR predictions. In the processing step, we utilized the five SR algorithms to process LR images and obtain five sets of enlarged/deblurred images. Then, we train five GAN models (with the same structure) separately corresponding to each of these five algorithms and then obtain five sets of prediction results. After that, another GAN model which is the same as the model used in the previous step is used to integrate all these images through ensemble learning. Images from each type of GAN SR results are concatenated before being input to the neural network for training, validation, and testing. The objective function of the ensemble learning training is also the same as the objective function in the previous step. The ensemble learning process can be expressed as

$$I_{SR} = \phi(\{I_{SR,i}\}_{i=1}^5, \Theta), \quad (9)$$

where $I_{SR} \in \mathbb{R}^{m \times n}$ denotes ensemble learning results which are also our SR outcomes, $\phi : I_{SR,i} \in \mathbb{R}^{m \times n} \rightarrow I_{SR} \in \mathbb{R}^{m \times n}$ denotes the ensemble learning process, Θ represents the parameters of the GAN model.

III. RESULTS

Our ensemble learning SR approach was streamlined and evaluated for MRI SR. It produced SR images with more structural and textural details, less artifacts and distortion.

In the following, we will show our PLR results, GAN SR results, and final ensemble learning results, and compare our SR results with state-of-the-art methods. To evaluate image quality,

TABLE III
COMPARISON OF DIFFERENT OBJECTIVE FUNCTIONS

	SSIM	PSNR
$\mathcal{L}_{adv} + \mathcal{L}_{per}$	0.848 ± 0.062	33.634 ± 2.986
$\mathcal{L}_{adv} + \mathcal{L}_{per} + \mathcal{L}_{mse}$	0.866 ± 0.059	33.723 ± 2.931
$\mathcal{L}_{adv} + \mathcal{L}_{per} + \mathcal{L}_{mse} + \mathcal{L}_{gra}$	0.870 ± 0.051	33.732 ± 2.876

the metrics we used include structural similarity (SSIM) and peak signal-to-noise ratio (PSNR). For easier comparing the differences between each image, from Fig. 5 to 9, zoomed-in subfigures were displayed. In addition, image contrast for each zoomed-in subfigure was also enhanced by uniforming the pixel values in each zoomed-in region between 0 and 1. Noticeably, SSIM and PSNR scores below each subfigure measured the whole image not the zoomed-in region.

A. Experimental Implementation and Neural Network Convergence

The NYU fastMRI singlecoil dataset was used in this study [43]. All images were reconstructed from proton density weighted knee scans under 1.5 or 3 Tesla. The size of original HR images is 320×320 . Totally, 159 patient scans with 5,744 slices were used for training (80%) and validation (20%), and additional 43 patient scans with 1,500 slices for testing. Before training, each PLR result was decomposed into forty-nine 80×80 patches. As far as the weight for each loss term in the objective function is concerned, we first tuned the perceptual loss. Then, we tuned the weight of the MSE loss and the gradient loss. Based on the results shown in Fig. 3, the weights in the objective function were selected to be 0.01, 0.01, and 1 for λ_1 , λ_2 , and λ_3 respectively. It can be found that the perceptual loss \mathcal{L}_{per} contributes most to the superior quality of SR images. Compared to the perceptual loss \mathcal{L}_{per} , the gradient loss \mathcal{L}_{gra} and the MSE loss \mathcal{L}_{mse} have a less effect on SR results. Nevertheless, they do contribute to the improvement of SR results, as shown in Table III. In the GAN SR process, the training continued for 50 epochs with the learning rate of 1×10^{-4} and batch size of 64 for each GAN model. During the training process, to visualize the convergence of the involved models, the values of MSE and perceptual loss are recorded and showed in Fig. 4. For ensemble learning, the training continued 50 epochs with the learning rate of 1×10^{-4} and batch size of 4. The experiment was conducted in Tensorflow on a single GTX 1080Ti GPU.

According to the MSE and perceptual loss curves in Fig. 4, all the models converge well during the training process since all the curves become gradually steady with the number of training epochs. Compared with the other MSE curves, it can be found that the curve of A+ is lower than the other four curves, which shows that the A+ based GAN model can achieve the best SR image with the highest similarity at the pixel level. In the high-level feature space, as extracted by the perceptual loss, both the A+ and ZIP curves show the lowest values while the NEDI curve is the highest. However, all the curves are pretty close to each other after 30 epochs.

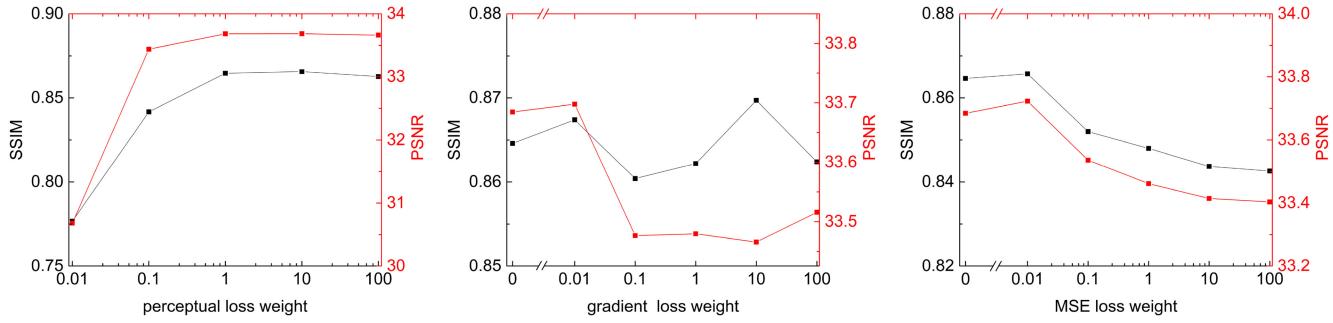


Fig. 3. SSIM and PSNR scores with respect to the weights of the loss terms.

TABLE IV
COMPARISON OF THE PLR RESULTS, GAN SR RESULTS, AND ENSEMBLE LEARNING RESULTS BASED ON A TEST DATASET CONSISTING OF 43 PATIENT SCANS

	PLR					GAN SR					Ensemble learning
	ZIP	BI	NEDI	SC	A+	ZIP	BI	NEDI	SC	A+	
SSIM	0.834	0.785	0.810	0.769	0.829	0.883	0.842	0.870	0.872	0.898	0.920
PSNR	30.552	26.287	29.211	28.931	28.655	34.263	33.219	33.732	33.279	34.849	36.060

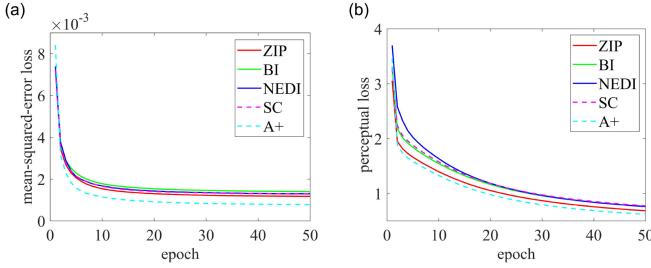


Fig. 4. Comparisons between the involved model types during the training process in terms of (a) MSE and (b) perceptual loss.

B. Processed LR Results

It is shown in the upper row of Fig. 5 that PLR images had smoother edges and clearer shapes compared with the corresponding LR image. As different SR algorithms were used in the processing step, PLR images obtained from these SR algorithms were somehow different. For example, the SC PLR are dramatically different from the other four PLR results with less-smoothed texture. Statistically, the SSIM and PSNR scores of ZIP, NEDI, and A+ PLR results are higher than BI and SC PLR results. After the processing step, we obtained diverse enlarged/deblurred images.

C. Super-Resolution Results of PLR Images Through GAN

After training the GAN models separately, five types of GAN SR prediction results were obtained, as shown in the bottom row in Fig. 5. Compared with the PLR results, the image quality of GAN SR results was much improved. Generally, these SR images included more textural details, clearer edges, and less blurs, which are much closer to the original HL image than the PLR versions. For example, in the zoomed-in regions, not only the boundary but also the texture of the meniscal and bone structures are clearer in GAN SR images than they are in PLR images. As the GANs were trained on different PLR datasets,

when GAN SR results are compared, differences among them are clearly detectable for the same ground truth. Taking the shape of the dark region in the bone as an example, as pointed by the red arrow in Fig. 5 it can be found that the dark region shape in each GAN SR result was slightly different from each other and from the ground truth. Quantitative analysis of PLR and GAN SR prediction results is summarized in Table IV. It can be seen that the PSNR and SSIM values of the GAN SR results were higher than the corresponding PLR results.

D. Super-Resolution Results Through Ensemble Learning

Compared with GAN SR results, the most distinct characteristic of the ensemble learning results is that they represent details more faithfully and become much closer to the ground truth. For example, the line pointed by the red arrow in the middle row of Fig. 6 can be clearly detected in the ensemble learning result while they are unclear in other GAN SR prediction results. As quantitatively shown in Table IV, the ensemble learning results achieved the highest PSNR and SSIM scores among all the results.

To investigate which type of GAN SR results contributed the most in the ensemble learning process, we conducted an ablation study by removing one type of GAN SR results at each time. The ablation study results are shown in the upper part of Table V. It can be found that after we deleted A+ SR results, the ensemble learning results obtained based on the other four GAN SR results were with the least SSIM and PSNR scores among all 5 types of ablation results, indicating that A+ SR results contributes the most in the ensemble learning process. On the contrary, BI SR results contributes the least as the highest SSIM and PSNR scores achieved when there are no BI SR results used in the ensemble learning process among the 5 types of ablation results. Then, we conducted experiments by only using a few the most important GAN SR results in the ensemble learning process to show the degradation of final results when less GAN SR results are involved. The results are shown in the bottom part of Table V.

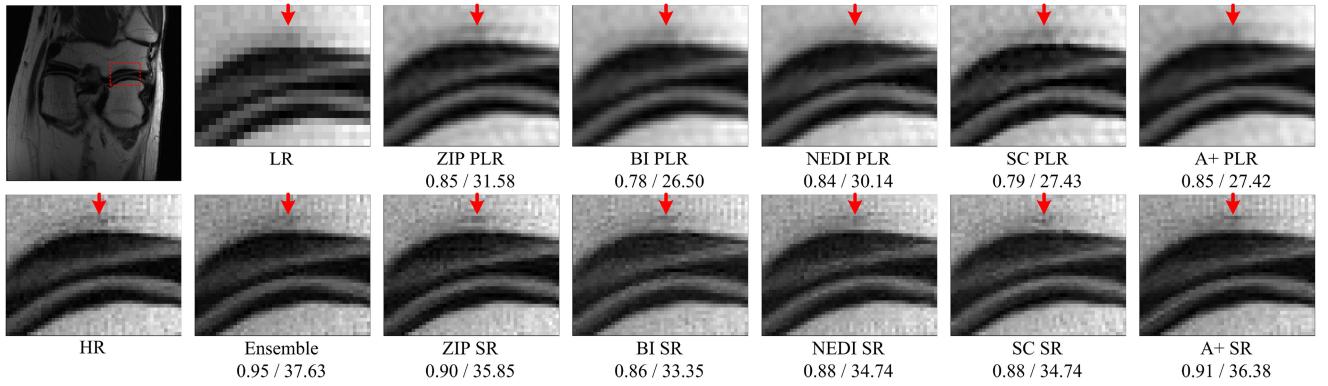


Fig. 5. Upper: LR image and 5 PLR images obtained using the ZIP, BI, NEDI, SC, and A+ algorithms respectively. Bottom: The HR image, ensemble learning result, and 5 different GAN SR results. The red boxed region is zoomed in for clarity. The numbers below each subfigure are SSIM and PSNR scores respectively.

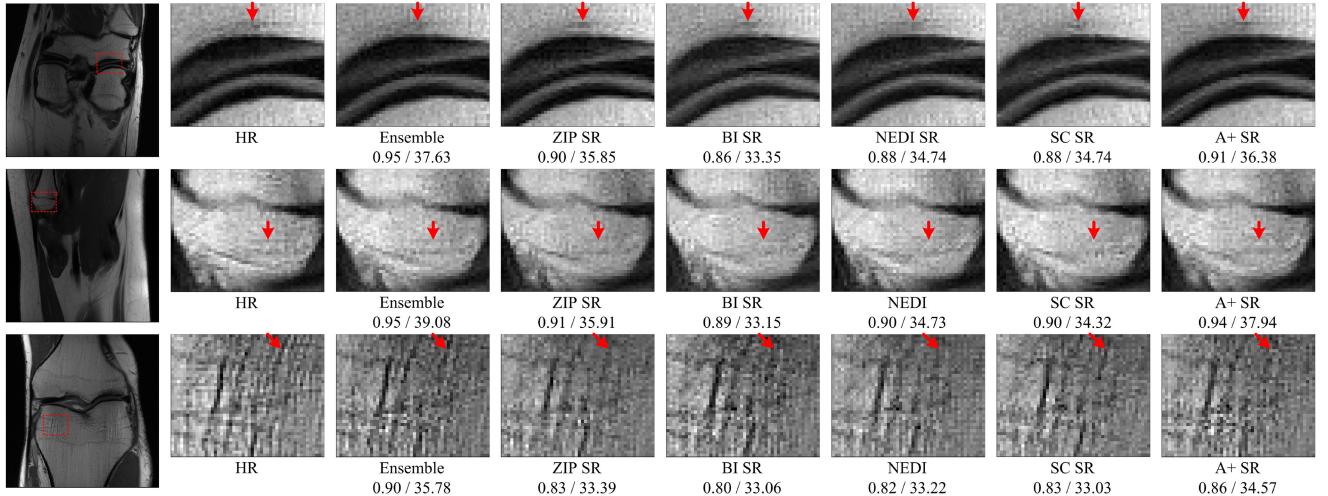


Fig. 6. GAN SR prediction results and the ensemble learning results. The red boxed regions are zoomed in for clarity. The numbers below each subfigure are SSIM and PSNR scores respectively.

TABLE V
COMPARISON OF ENSEMBLE LEARNING RESULTS TRAINED USING DIFFERENT GAN SR RESULTS. RESULTS ARE BASED ON TEST DATASET FROM 43 PATIENT SCANS

	SSIM	SSIM drops (↓)	PSNR	PSNR drops (↓)
Lack-A+	0.859 ± 0.022	0.061	34.689 ± 2.371	1.371
Lack-NEDI	0.891 ± 0.022	0.029	35.030 ± 2.408	1.030
Lack-ZIP	0.909 ± 0.021	0.011	35.735 ± 2.373	0.325
Lack-SC	0.913 ± 0.021	0.007	35.750 ± 2.442	0.310
Lack-BI	0.916 ± 0.020	0.004	35.935 ± 2.345	0.125
EN-2	0.904 ± 0.042	0.016	35.435 ± 3.027	0.625
EN-3	0.911 ± 0.031	0.009	35.863 ± 2.778	0.197
EN-4	0.916 ± 0.020	0.004	35.935 ± 2.345	0.125
EN-5	0.920 ± 0.018	-	36.060 ± 2.155	-

EN-2 means two mostly contributed GAN SR results (A+ GAN SR and NEDI GAN SR) were used in the ensemble learning process. For EN-3, A+, NEDI, and ZIP GAN SR results were used. For EN-4, A+, NEDI, ZIP, and SC GAN SR results were

used. It can be seen that instead combining all 5 types of GAN SR results for ensemble learning, partially using our GAN SR results will decrease the image quality evidenced by decreased SSIM and PSNR scores. Representative results are presented in Fig. 7. Compared with EN-5 results, structural details become less detectable in the results from partially reduced GAN SR data. All these results demonstrate that ensemble learning results can be further improved by combining more GAN SR predictions.

To investigate the contribution of the neural network used in the ensemble learning process, we compared our ensemble learning results with the results obtained from simply averaging all 5 types of GAN SR results. For the simple averaging method, the value of each pixel was obtained by averaging pixel values in all 5 types of GAN SR results. There was no involvement of any neural network in combining GAN SR results when the simple averaging method was used. As shown in Fig. 8, using a GAN during the ensemble learning process can produce better results than simply averaging all GAN SR results.

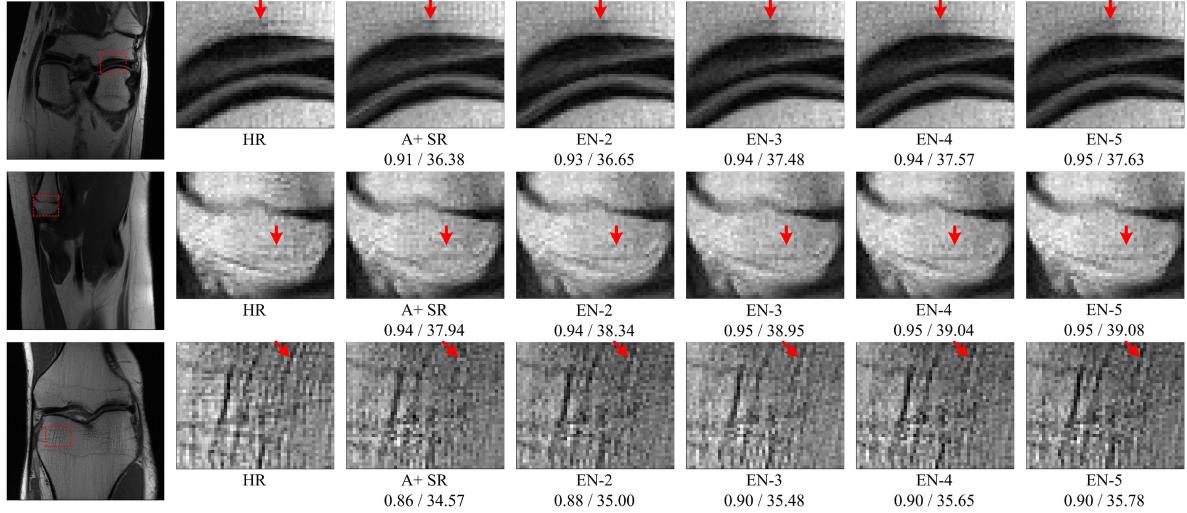


Fig. 7. Comparison of ensemble learning results obtained by combining various numbers of GAN SR results. En-*k* means ensemble learning results from *k* types of GAN SR results. The red boxed regions are zoomed in for clarity. The numbers below each subfigure are SSIM and PSNR scores respectively.

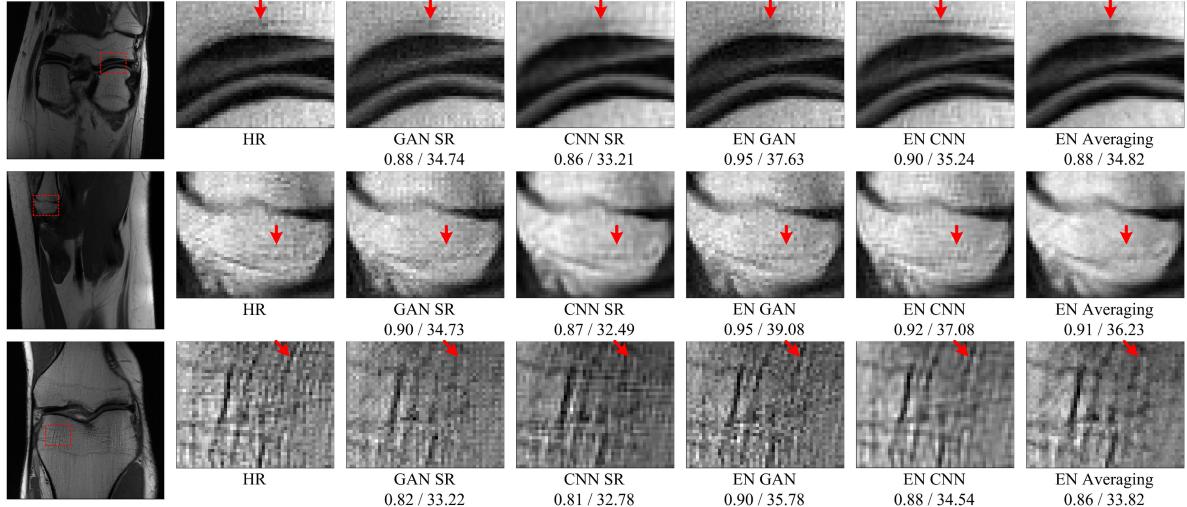


Fig. 8. Comparison of GAN and CNN results. “GAN SR” and “CNN SR” show the NEDI SR results produced by GAN and CNN (GAN generator) respectively. “EN GAN” and “EN CNN” show the ensemble learning results produced by GAN and CNN (GAN generator) respectively. “EN Averaging” shows the ensemble learning result by simply averaging the 5 GAN SR results. The red boxed regions are zoomed in for clarity. The numbers below each subfigure are SSIM and PSNR scores respectively.

E. Comparison With GAN and CNN in the Proposed SR Framework

In our proposed SR framework, GANs are involved in two steps: SR of PLR results and ensemble learning. To investigate the effect of GANs on the proposed SR framework, we substituted GAN with a CNN that had the same architecture as the GAN generator. The loss function of the CNN is the same as the objective function of the GAN without the adversarial loss term. CNN experiments were conducted using the same setting as GAN experiments. In Fig. 8, it can be seen that when substituting the GAN with a CNN for SR of PLR results, the CNN SR results were more blurry and had lower SSIM and PSNR scores than the GAN SR results. Substituting GAN with CNN in the ensemble learning step achieved similar results. Compared with GAN based ensemble learning results, CNN

based results tend to be more blurry with less clear details and lower SSIM and PSNR scores.

F. Comparison With Other Deep Learning-Based Methods

Fig. 9 and Table VI compare our results with the counterparts obtained using other state-of-the-art SR methods. These well-known benchmark SR methods include FRSCNN [12], LapSRN [16], ESPCN [15], and VDSR [13]. Encouragingly, our method outperformed these methods as evidenced by the resultant SR images with higher PSNR and SSIM values. More importantly, our results faithfully present more details in HR images. For example, the line pointed by the red arrows in the middle row of Fig. 9 can only be seen in our SR results. Also, our results greatly reduced artifacts and distortions. Taking the line pointed by the red arrows as an example, in the bottom row

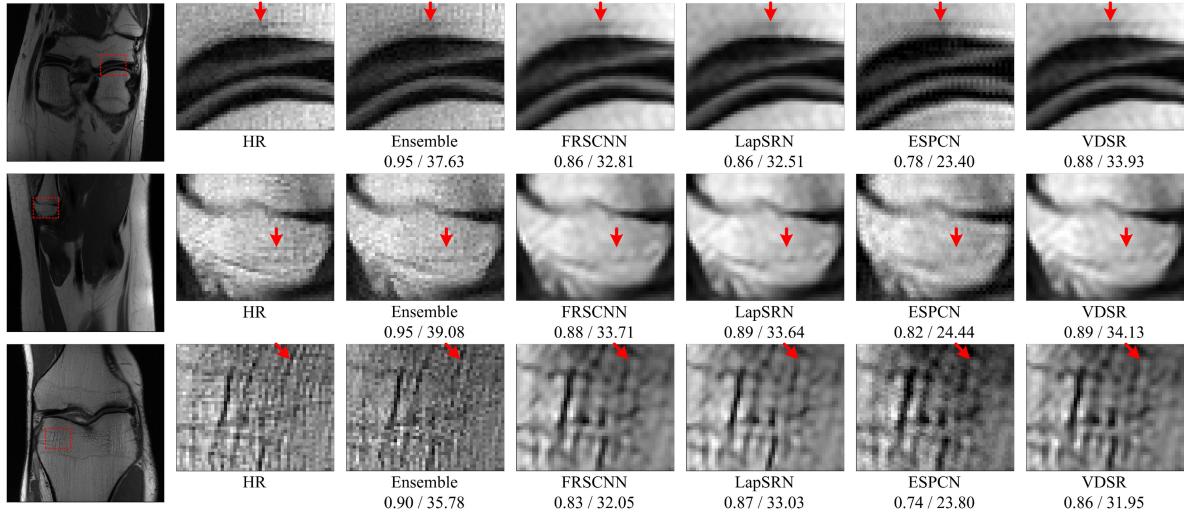


Fig. 9. Comparison of the proposed SR method and state-of-the-art methods. The red boxed regions are zoomed in for clarity. The numbers below each subfigure are SSIM and PSNR scores respectively.

TABLE VI
COMPARISON OF OUR METHOD WITH OTHER STATE-OF-THE-ART METHODS.
RESULTS ARE BASED ON TEST DATASET FROM 43 PATIENT SCANS

	SSIM	PSNR
FSRCNN	0.873 ± 0.024	28.877 ± 2.331
LapSRN	0.884 ± 0.028	28.804 ± 2.384
ESPCN	0.831 ± 0.030	24.933 ± 3.303
VDSR	0.883 ± 0.021	28.997 ± 2.258
Ours	0.920 ± 0.018	36.060 ± 2.155

of Fig. 9, the shape of edges in our SR images is much closer to the ground truth than the counterparts.

IV. DISCUSSIONS

Based on the results described above, it can be seen that through the ensemble learning superior MRI SR images can be achieved. In comparison with other state-of-the-art methods, our method produced impressive results without significant artifacts, distortions, and blurry details, giving the highest PSNR and SSIM values.

In the processing LR images step, 5 different SR algorithms were used to enlarge LR images. As a result, 5 types of enlarged PLR results were obtained. It can be found in Fig. 5 that each PLR image is slightly different from each other and from the ground truth. These differences can be seen as information in the original image being preserved to different degrees. These diversified PLR images provide complementary priors for subsequent ensemble learning. According to our design, by jointly using multiple SR algorithms to process LR images, we are able to create a PLR dataset with much more image priors compared with only using any single PLR results. Then, we can obtain better results in the ensemble learning process. Results in Fig. 5 and Table IV are consistent to the goal of this ensemble learning network.

Results in Fig. 7 and Table V show that when increasing the number of GAN SR results in the ensemble learning, the

gain in the quality of final results is gradually diminishing, indicating that less information is added when overly more GAN SR results are involved. Based on this observation, we believe that the dataset we created by combining 5 types of GAN SR results contains enough information or complementary priors for ensemble learning. In terms of how many types of GAN SR results are needed for decent ensemble learning results, generally speaking, more types of GAN SR results involved in the ensemble learning will result in better SR results. There is no doubt that enlarging the size of the ensemble learning input dataset by adding more SR results like FRSCNN and LapSRN results can further improve ensemble learning results. However, we must take into account the computational resources. More GAN SR results involved means more computations and longer training time. To balance the ensemble learning results and computational resources, in this study we only combined 5 types of GAN SR results and obtained satisfactory outcomes.

Since introduced in 2014 [44], GAN has been extensively used in many applications due to its compelling performance originated from the adversarial training mechanism. Such a mechanism makes GAN models achieve higher quality of SR results than CNN models. However, the GAN model could potentially induce artifacts and distortions in prediction results, which may be due to the nature of GAN and the prevalence of background noise. On one hand, during the training process of a GAN model, the generator is trained to learn the true distribution of training data. Compared with CNN-based supervised training, the training process of GAN is not strongly supervised as predictions are controlled indirectly through a discriminator, not directly anchored with labels. Owing to this mechanism, it is not easy to train the model so that it can perfectly learn the true distribution. As a result, the generator can hardly produce a fake result that is exactly the same as a real result. As shown in the bottom row of Fig. 6, the GAN results suffer from edge distortions. On the other hand, during the MR scanning process, noise is inevitably recorded [45], by which clinical MR images are greatly influenced. Noise hampers the training process of

GAN so that it cannot be perfectly trained. This phenomenon is very evident when noise is prevalent that features cannot be easily extracted. Fortunately, these artifacts are mostly unique and can be corrected by ensemble learning. By combining multiple GAN SR results, the effect of an artifact in any type of GAN SR results can be corrected by other GAN SR results without such an artifact. In addition, jointly using multiple GAN SR results expands the training dataset, and as a result the robustness of the whole SR framework can be strengthened.

Ensemble learning synergistically combines all GAN SR results together. Different from the simple average algorithm, the GAN-based integrator learns to perform ensemble learning in a data-driven fashion and is task-specific. Roughly speaking, the GAN can be assumed as a black box. Through the GAN model, pixels in each input image is assigned for a weight for summation and nonlinear activation. During the training process, the weight of each pixel in each type of GAN SR input can be learned according to high-level features and contexts. This data-driven and context-based mechanism guarantees that the GAN-based ensemble learning method outperforms the commonly used averaging method.

Differing from other medical imaging modalities such as CT, the down-sampling process of MRI is happened in the frequency domain. Multiple down-sampling methods were used to create blurry LR images. Among all the methods, low frequency data in the center region is often preferred to be kept as these data are collectable with a low-field MRI scanner and contribute to reconstruct major biological structures and features. On the contrary, high frequency data in the peripheral region will be highly down-sampled which mainly contribute to form details of images. In this study, we only keep the central low frequency data and discard all peripheral high frequency data. In addition, as data is truncated uniformly, the generated LR images are isotropic and the artifacts brought about in the down-sampling process are more predictable compared with other down-sampling methods such pseudo-random down-sampling [46] and random down-sampling in the phase-encoding direction [47].

Currently, there are two strategies for implementation of deep-learning-based MRI image reconstruction. The first strategy uses a neural network to map LR images to HR ones directly in the image domain. Both fidelity and regularization terms are jointly learned by the neural network. Examples are VDSR [13] and LapSRN [16]. The other strategy features a network to map k -space data to an MRI image directly, such as combining traditional iterative algorithms with neural networks by unrolling the iterative scheme into several individual steps. With this strategy, neural networks are sometimes trained to only learn the regularizer in the objective function. Examples of this method are variational networks [48] and CRNN [47]. In the two strategies, the first one, which is also the approach proposed in this paper, is conceptually simple, computationally efficient, and has attracted a major attention recently.

Our proposed SR framework has a potential utility to enable a faster MRI scan and achieve a higher image quality. According to the design presented in the paper, we can shorten the scan time by 2-fold and still keep the image quality essentially intact, with clear structural and textural details comparable to the full-data-derived images.

Our proposed SR framework currently works on magnitude images with phase information ignored. In the clinical applications of MRI, both k -space data and reconstructed images are generally complex numbers, indicating that there are not only magnitude but also phase images. If both magnitude and phase images are used together, the super-resolution results could be further improved. We are actively working along this direction in our follow-up study.

V. CONCLUSION

We have proposed an ensemble learning based SR method to improve the image quality of MR images. We have first used 5 existing SR algorithms to enlarge LR images and obtained PLR images with complementary priors. Then, a GAN framework has been used to produce SR prediction results based on different PLR images. Next, we can obtain superior SR images through deep ensemble learning. The SR images obtained after ensemble learning is better than any individual PLR images or GAN SR prediction results. Compared with other state-of-the-art methods, our method enjoys advantages of minimal artifacts, little distortions, and rich details.

REFERENCES

- [1] E. Plenge, D. H. Poot, M. Bernsen, G. Kotek, G. Houston, and P. Wielopolski, "Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time?", *Magn. Reson. Med.*, vol. 68, no. 6, pp. 1983–1993, 2012.
- [2] E. Van Reeth, I. W. Tham, C. H. Tan, and C. L. Poh, "Super-resolution in magnetic resonance imaging: A review," *Concepts Magn. Reson. Part A*, vol. 40, no. 6, pp. 306–325, 2012.
- [3] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.
- [4] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.
- [5] R. Hardie, "A fast image super-resolution algorithm using an adaptive Wiener filter," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2953–2964, Dec. 2007.
- [6] J. V. Manjón, P. Coupé, A. Buades, V. Fonov, D. L. Collins, and M. Robles, "Non-local MRI upsampling," *Med. Image Anal.*, vol. 14, no. 6, pp. 784–792, 2010.
- [7] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [8] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.*, 2010, pp. 711–730.
- [9] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. Asian Conf. Comput. Vision*, 2014, pp. 111–126.
- [10] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vision*, 2014, pp. 184–199.
- [11] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.
- [12] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 391–407.
- [13] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 1646–1654.
- [14] J. Kim, J. Kwon Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 1637–1645.
- [15] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 1874–1883.

- [16] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2599–2613, Nov. 2019.
- [17] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, and A. Acosta, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 4681–4690.
- [18] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 63–79.
- [19] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin, "Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2018, pp. 701–710.
- [20] C. You *et al.*, "CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 188–203, Jan. 2020.
- [21] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2017, pp. 136–144.
- [22] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 4799–4807.
- [23] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 2472–2481.
- [24] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 286–301.
- [25] W. Han, S. Chang, D. Liu, M. Yu, M. Witbrock, and T. S. Huang, "Image super-resolution via dual-state recurrent networks," in *Proc. IEEE Conf. Computer Vision Pattern Recognit.*, 2018, pp. 1654–1663.
- [26] M. S. Sajjadi, R. Vemulapalli, and M. Brown, "Frame-recurrent video super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 6626–6634.
- [27] Y. Huang, W. Wang, and L. Wang, "Bidirectional recurrent convolutional networks for multi-frame super-resolution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 235–243.
- [28] A. S. Chaudhari *et al.*, "Super-resolution musculoskeletal MRI using deep learning," *Magn. Reson. Med.*, vol. 80, no. 5, pp. 2139–2154, 2018.
- [29] K. Zeng, H. Zheng, C. Cai, Y. Yang, K. Zhang, and Z. Chen, "Simultaneous single-and multi-contrast super-resolution for brain MRI images based on a convolutional neural network," *Comput. Biol. Med.*, vol. 99, pp. 133–141, 2018.
- [30] Y. Chen, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, and D. Li, "Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network," in *Proc. Int. Conf. Med. Image Comput. Comput-Assist. Intervention*, 2018, pp. 91–99.
- [31] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. G. Christodoulou, and D. Li, "Brain MRI super resolution using 3D deep densely connected neural networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, 2018, pp. 739–742.
- [32] C. Liu, X. Wu, X. Yu, Y. Tang, J. Zhang, and J. Zhou, "Fusing multi-scale information in convolution network for MR image super-resolution reconstruction," *Biomed. Eng. Online*, vol. 17, no. 1, 2018, Art. no. 114.
- [33] Q. Lyu, C. You, H. Shan, Y. Zhang, and G. Wang, "Super-resolution MRI and CT through GAN-CIRCLE," in *Developments in X-Ray Tomography XII*, vol. 11113, p. 1111130X, International Society for Optics and Photonics, 2019.
- [34] C. Zhang and Y. Ma, *Ensemble Machine Learning: Methods and Applications*. Berlin, Germany: Springer, 2012.
- [35] T. G. Dietterich, "Ensemble methods in machine learning," in *Proc. Int. Workshop Multiple Classifier Syst.*, 2000, pp. 1–15.
- [36] L. Wang, Z. Huang, Y. Gong, and C. Pan, "Ensemble based deep networks for image super-resolution," *Pattern Recognit.*, vol. 68, pp. 191–198, 2017.
- [37] R. Liao, X. Tao, R. Li, Z. Ma, and J. Jia, "Video super-resolution via deep draft-ensemble learning," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 531–539.
- [38] M. A. Bernstein, S. B. Fain, and S. J. Riederer, "Effect of windowing and zero-filled reconstruction of MRI data on spatial resolution and acquisition strategy," *J. Magn. Reson. Imag.: An Official J. Int. Soc. Magn Reson. Medicine*, vol. 14, no. 3, pp. 270–280, 2001.
- [39] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
- [40] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, and L. Sun, "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018.
- [41] M. Lin, Q. Chen, and S. Yan, "Network in network," in *Proc. 2nd Int. Conf. Learn. Represent.*, Banff, AB, Canada, Apr. 2014, *arXiv:1312.4400*.
- [42] Q. Yang *et al.*, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.
- [43] J. Zbontar *et al.*, "fastMRI: An open dataset and benchmarks for accelerated MRI," 2018, *arXiv:1811.08839*.
- [44] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 2672–2680, 2014.
- [45] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Reson. Med.*, vol. 34, no. 6, pp. 910–914, 1995.
- [46] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model based deep learning architecture for inverse problems," *CoRR*, vol. abs/1712.02862, 2017.
- [47] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 280–290, Jan. 2019.
- [48] K. Hammernik *et al.*, "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.



Qing Lyu (S'20) received the Bachelor of Science degree and Master of Science degree from the Department of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2013, and 2016, respectively. He is currently working toward the Ph.D. degree in biomedical engineering with Rensselaer Polytechnic Institute, Troy, NY, USA. His main research interests include the MRI postprocessing and signal optimization.



Hongming Shan (M'20) received the B.S. degree from the Shandong University of Technology, Zibo, China, in 2011, and the Ph.D. degree from Fudan University, Shanghai, China, in 2017. He is currently a Postdoctoral Scholar with Rensselaer Polytechnic Institute, Troy, NY, USA. His research interests include machine/deep learning, computer vision, dimension reduction, and biomedical imaging.



Ge Wang (F'03) is the Clark & Crossan Endowed Chair Professor and Director of the Biomedical Imaging Center, Rensselaer Polytechnic Institute, Troy, NY, USA. He authored the first spiral/helical cone-beam/multi-slice CT algorithm in 1991 and has since then systematically contributed to more than 100 papers to theory, algorithms, artifact reduction, and biomedical applications in this important area of CT research. Currently, there are 100+ million medical CT scans yearly with a majority in the spiral/helical cone-beam/multi-slice mode. His group developed interior tomography theory and algorithms to solve the long-standing "interior problem" for high-fidelity local reconstruction, enabling omni-tomography ("all-in-one") with CT-MRI as an example. He initiated the area of bioluminescence tomography with a significant impact on biophotonics. He has authored or coauthored 460+ journal publications, covering imaging-related diverse topics at depth and receiving a high number of citations and academic awards. His results were featured in *Nature*, *Science*, *PNAS*, and various news media. In 2016, he authored the publication the first perspective on deep-learning-based tomographic imaging as the new direction of machine learning and the *Nature Machine Intelligence* paper on the superiority of deep learning over iterative reconstruction. His team has been continuously well-funded by federal agencies and major imaging companies, actively translating deep learning techniques into imaging products. His research interests include X-ray CT, MRI, optical tomography, multimodality fusion, and machine learning. He is the Lead Guest Editor for five *IEEE TRANSACTIONS ON MEDICAL IMAGING* (TMI) Special Issues, Founding Editor-in-Chief for *International Journal of Biomedical Imaging*, a board member of the IEEE ACCESS, and an Associate Editor for the TMI (recognized as "Outstanding Associate Editor" by TMI), Medical Physics, and *Machine Learning Science and Technology*.