

Unsupervised Representation Learning for 3-Dimensional Magnetic Resonance Imaging Super-Resolution with Degradation Adaptation

Jianan Liu, Hao Li, Tao Huang, *Senior Member, IEEE*, Euijoon Ahn, Kang Han, Adeel Razi, Wei Xiang, *Senior Member, IEEE*, Jinman Kim, *Member, IEEE*, and David Dagan Feng, *Life Fellow, IEEE*

Abstract—High-resolution (HR) magnetic resonance imaging is essential in aiding doctors in their diagnoses and image-guided treatments. However, acquiring HR images can be time-consuming and costly. Consequently, deep learning-based super-resolution reconstruction (SRR) has emerged as a promising solution for generating super-resolution (SR) images from low-resolution (LR) images. Unfortunately, training such neural networks requires aligned authentic HR and LR image pairs, which are challenging to obtain due to patient movements during and between image acquisitions. While rigid movements of hard tissues can be corrected with image registration, aligning deformed soft tissues is complex, making it impractical to train neural networks with authentic HR and LR image pairs. Previous studies have focused on SRR using authentic HR images and downsampled synthetic LR images. However, the difference in degradation representations between synthetic and authentic LR images suppresses the quality of SR images reconstructed from authentic LR images. To address this issue, we propose a novel Unsupervised Degradation Adaptation Network (UDEAN). Our network consists of a degradation learning network and an SRR network. The degradation learning network downsamples HR images using the degradation representation learned from misaligned or unpaired LR images. The SRR network then learns to map the downsampled HR images to the original ones. Experimental results show that our method outperforms state-of-the-art networks with an improvement of up to 0.051/3.52 dB in SSIM/PSNR on two public datasets, thus is a promising solution to the challenges in clinical settings.

Impact Statement—Acquiring precisely aligned authentic high-resolution (HR) and low-resolution (LR) image pairs is considerably challenging, making supervised super-resolution (SR) reconstruction unfeasible in clinical settings. Therefore, unsupervised learning has emerged as a promising solution to this issue. This paper introduces an unsupervised network designed to be trained using unpaired or misaligned HR and LR images and enable the reconstruction of high-quality SR images. Additionally,

a comparative analysis is conducted between our network and state-of-the-art unsupervised networks. This investigation delves into the divergent domain transfer strategies employed by these networks, exploring the mechanisms underpinning unsupervised domain transfer. Such insights have facilitated the identification of the most suitable approach for medical image processing.

Index Terms—3D Super-Resolution, Degradation Adaptation, Geometric Deformation, Magnetic Resonance Imaging, Unsupervised Learning.

I. INTRODUCTION

HIGH resolution (HR) magnetic resonance imaging (MRI) provides abundant soft tissue contrast and detailed anatomical structures, which help doctors with accurate diagnosis and image-guided treatment. However, the acquisition of HR images is highly time-consuming and costly. Prolonged acquisition time also leads to considerable patient discomfort, while motion artifacts are inevitable. Therefore, a reduction in the acquisition time of MRI images is highly demanded, and deep learning-based technology has been proposed to tackle this issue in recent years [1], [2], [3], [4], [5], [6], [7], [8], [9], [10].

Advancements in deep learning have facilitated the widespread adoption of super-resolution (SR) techniques across various computer vision applications, each with unique prerequisites and objectives. For instance, SR methods applied to real-world visual images and face hallucinations prioritize visual effect and perceptual performance [11], [12], [13], [14], [15]. Conversely, applications such as remote sensing images [16], [17], satellite videos [18], [19] and medical images [4], [5], [7], [28] require accurate restoration of details via SR for further analysis. Notably, medical images, particularly MRI images, present unique challenges. Unlike remote sensing images and videos, which contain additional information like RGB channels [16], [17], time series data [18], [19], and multi-spectral information [20], [21], [22], MRI SR relies solely on voxel intensity. Furthermore, MRI SR requires three-dimensional (3D) reconstruction, adding further complexities to SR. This is different from standard remote sensing applications that typically involve two-dimensional (2D) image data. Therefore, the development of a dedicated network tailored for the MRI SR task is crucial to address these challenges and ensure high accuracy and reliability in the 3D space.

Training neural networks for supervised super-resolution reconstruction (SRR) can be optimized by utilizing paired

Jianan Liu and Hao Li contribute equally to the work and are co-first authors. Corresponding Authors: Hao Li and Tao Huang.

Jianan Liu is with Vitalent Consulting, Gothenburg, Sweden (email: jianan.liu@vitalent.se).

Hao Li is with the Department of Neuroradiology, University Hospital Heidelberg, Heidelberg, Germany (email: hao.li@med.uni-heidelberg.de).

Tao Huang and Euijoon Ahn are with the College of Science and Engineering, James Cook University, Cairns, Australia (email: tao.huang1@jcu.edu.au; euijoon.ahn@jcu.edu.au).

Kang Han and Wei Xiang are with the School of Computing, Engineering and Mathematical Sciences, La Trobe University, Melbourne, Australia (email: w.xiang@latrobe.edu.au; k.han@latrobe.edu.au).

Adeel Razi is with Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Melbourne, Australia (email: adeel.razi@monash.edu).

Jinman Kim and David Dagan Feng are with the School of Computer Science, University of Sydney, Sydney, Australia (email: jinman.kim@sydney.edu.au; dagan.feng@sydney.edu.au).

low-resolution (LR) and HR MRI images, as reported in previous studies [2], [3], [4], [5], [6], [7], [8], [9], [10]. However, obtaining paired authentic HR and LR images - real images acquired using MRI scanners - remains a challenge in clinical settings. Besides, when the authentic HR and LR MRI image pairs are acquired in a very limited number of cases, the misalignment is introduced due to participant movement during and between the image acquisition, even in the same scan session and with highly experienced participants. The misalignment incurs significant errors in the SR image generation, making it unusable in real-world clinical practice.

Image registration has been suggested as a straightforward solution to correct the transformation between misaligned authentic HR and authentic LR images [23]. However, the limited effectiveness of rigid image registration in handling non-rigid geometric deformation caused by soft tissue movement has been reported [24]. Recent deformable image registration algorithms have improved accuracy, achieving a dice similarity coefficient (DSC) of around 90% [25]. Nevertheless, such accuracy still needs to be improved when considering the size of the field of view and the object.

An alternative solution is to train the network using a combination of authentic HR images and synthetic LR images, with the synthetic LR images generated through deterministic downsampling filters, such as the Gaussian blur filter [2], [3], or K -space truncation from authentic HR images [4], [5], [6], [7], [8], [9], [10]. After training, the trained network is used to reconstruct SR images from authentic LR images acquired separately. Most previous works in MRI SRR have followed this synthetic-LR to authentic-HR routine and have designed several different networks. For example, CSN [2] and SERAN [3] used attention mechanisms to enhance feature fusion during the 2D reconstruction procedure. ReCNN [4], [5], and DCSRN [6] were the first 3D MRI SRR networks proposed. mDCSRN [7], as an extension of DCSRN, further enhanced its capability with increased depth and a discriminator. deepGG [8], SSGNN [26], and SMORE (3D) [27] generated 3D isotropic SR image volume from anisotropic LR images. MCSR [9] and MINet [10] introduced the use of HR reference with different contrast as prior information to achieve outstanding performance. TSRCAN [28] also achieved performance comparable to the MINet in 3D SRR with single-contrast data and significantly reduced demands on computational resources and inference time.

However, in actual clinical settings, the performance of supervised neural networks for super-resolution reconstruction (SRR) is often compromised due to degradation shifts [29]. These shifts occur because of the different degradation presentations between authentic and synthetic LR images. Unsupervised SRR approaches that use only unpaired LR and HR images for training can be employed to address this issue. For example, ZSSR [30] learns the degradation from the LR image to a lower-resolution image further downsampled from the LR image. The learned degradation is then used to fit the inverse process from the LR to the HR image. Similarly, SMORE (3D) [27] follows the same approach in 3D MRI SRR. HR-Ref-ZSSR [31] applies this procedure to MRI image SR with HR reference provided from another modality. ZSSR-GAN

[32] augments the network with a discriminator to implement a similar idea. However, these approaches still suffer from the degradation shift caused by the inconsistency between the learned degradation from the LR image to the lower resolution image and the degradation from the HR image to the LR image, and their performances are still unsatisfactory.

To address the challenge of degradation shift, Xiao *et al.* [17] and Wang *et al.* [33] have proposed self-supervised methods based on contrastive learning for SRR from LR images with unknown degradation. In these methods, the networks are trained with positive and negative samples to push the generated SR images toward positive samples. These methods mitigate the degradation shift between test LR images with various degradation representations and the training LR and HR data. However, building positive and negative samples increases the complexity of the training procedure. Unlike contrastive learning approaches, domain transfer techniques are less complex and more straightforward. The degradation shift could also be addressed by employing domain transfer techniques initially designed to transfer images between different domains. The most straightforward method is CycleGAN [34], which defines the source and target domains and facilitates image transfer between them. PseudoSR [35] uses CycleGAN to transfer the target domain LR image to the source domain and then reconstructs the SR image in the source domain using an SR network. However, these CycleGAN-based methods only transfer images between domains in the image space rather than the latent feature space. Another unsupervised learning method, DASR [36], addresses the degradation shift by calculating a domain distance map. It uses a downsampling network to transfer the HR image from the source domain to the LR image from the target domain and calculate the domain distance map simultaneously. Then, guided by loss functions based on the domain distance map, DASR trains an SR network to transfer the target domain LR image back to the source domain HR image. Although DASR is a state-of-the-art method, it lacks end-to-end training, making it difficult to train the model optimally. However, the approaches mentioned above are designed for general computer vision SR tasks. They may not be appropriate for MRI SRR.

To address the challenges posed by the lack of aligned authentic LR and HR MRI image pairs and the degradation shift, we propose an unsupervised method that can utilize misaligned or unpaired HR and LR images to train the network. The key contributions of our work are summarized below.

- We introduce an end-to-end unsupervised degradation adaptation deep neural network (UDEAN), which employs a degradation adaptation (DA) mechanism to adaptively learn the degradation representation between the misaligned or unpaired authentic LR and HR MRI images in both the image space and the latent feature space. The combination of DA in both spaces significantly enhances the quality of the reconstructed SR images, surpassing the performance of DA in any single space.
- Our proposed DA mechanism outperforms existing domain transfer methods in the literature. Specifically, the DA mechanism transfers the HR image of the source

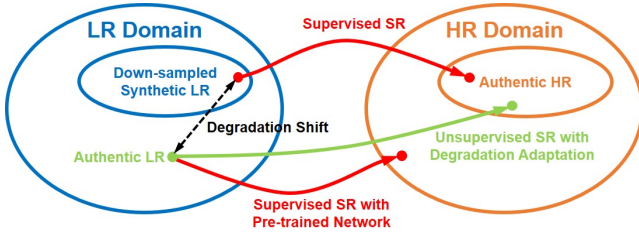


Fig. 1. Domain interpretation of differences between supervised and unsupervised SR. A large degradation shift exists between the SR result and desired HR image, which is caused by applying a supervised network, which is pre-trained with synthetic LR images, to authentic LR images with degradation deviating.

domain to the LR image of the target domain with an unknown degradation representation, minimizing the errors in the reconstructed SR MRI images.

- The proposed method can be trained with either misaligned or unpaired LR and HR MRI images, making it suitable for real-world clinical settings where accurately aligned authentic LR and HR MRI image pairs are not available.
- Our experimental results demonstrate that our proposed method has distinct advantages over existing paired supervised solutions, achieving up to 0.022/1.49 dB improvement in SSIM/PSNR, and outperforming other unsupervised SR training methods by up to 0.051/3.52 dB in SSIM/PSNR across two diverse public datasets and two scale factors.

II. METHODOLOGY

A. Degradation Shift and SR Reconstruction with Unknown Degradation

The degradation shift, also known as the domain gap in the real-world image SRR, was initially noticed in blind SRR tasks, where the downsampling parameters of the LR images were unknown [38]. In previous studies of real-world image SRR, HR images were downsampled with a bicubic algorithm or specific Gaussian blurring kernels [38], [39], [40], [41], [42], [43], [44], [33], [45]. With these downsampling algorithms, the parameters of the algorithms for the LR images were known in both training and inference datasets. They should be identical to guarantee the performance of the trained network [38]. However, the parameters of the downsampling algorithms in the real world were normally unknown and very difficult to model [38]. Therefore, it is impossible to guarantee identical downsampling for the LR images in both the training dataset and the inference dataset, and the difference between the degradation of the training and inference datasets resulted in a downgraded performance of the trained network in inference [38], as shown in Fig. 1.

Similarly, in MRI SRR, K -space truncation is the most widely used downsampling algorithm to generate synthetic LR images [4], [5], [6], [7], [8], [9], [10]. With this algorithm, an HR image is transformed into its fake K -space using the fast Fourier transformation (FFT). The fake K -space is truncated based on the downsampling factor. At last, the remained part

of the fake K -space is transformed back to the synthetic LR image using inverse FFT (iFFT). However, the degradation shift between the synthetic and authentic LR images created in this process still remained [29]. The method of K -space truncation is a highly simplified downsampling model, which simulates the acquisition process of LR images based on the assumption that the K -spaces of both the HR and the LR images were fully sampled. However, to shorten the scan time in the clinical measurements, most of the protocols for both LR and HR images acquire K -spaces, which are under-sampled in several different manners, including partial Fourier in phase-encoding and/or slice-encoding direction, parallel imaging, elliptical K -space, etc. Most of these under-sampling processes are difficult to simulate accurately since many of the key parameters are unknown. In addition, some other factors also lead to unpredictable differences between the degradation representations of synthetic and authentic LR images, such as the varied signal-to-noise ratio caused by the different signal intensity distribution of the K -spaces or different tuning of the transmission/receiving chain.

Therefore, unsupervised SRR can be a promising solution when the degradation shift exists. Several methods have been proposed to address this problem. These methods can be categorized into three types: the generative network-based domain transfer methods [34], [35], [36], the blurring kernel-based methods [38], [39], [40], [41], [42], [43], [44], and contrastive learning [17], [33], [45]. Since the LR images in the MRI were not downsampled with either bicubic or Gaussian blurring kernels, the blurring kernel-based methods are incompatible with the MRI SRR. Contrastive learning requires positive and negative training samples, which are difficult to acquire, making this method difficult to implement in the MRI SRR. Therefore, a network with degradation adaptation based on the domain transfer method is the most straightforward option and was adopted in this paper for SRR from LR images with unknown degradation.

B. Network Architecture of the UDEAN

In terms of unsupervised training, images from the source and target groups are used in the training process. The images from the target group are the LR images that will be upsampled to reconstruct the SR images. The images from the source group provide the HR images with the style that the SR images expect. As shown in Fig. 2, the HR MRI image Y_s from the source group and the LR MRI image X_t from the target group are misaligned or unpaired, and there is no ground truth HR image to calculate any pixel-wise or structural losses with the generated SR image. Noting that misaligned LR and HR images can be considered a special case of unpaired data, where the LR image contains similar structural information as the HR image but is not pixel-wise aligned. Therefore, the training is still not applicable in a supervised manner.

Therefore, we propose an unsupervised degradation adaptation network, UDEAN, whose architecture is shown in Fig. 2. The network contains two components: a degradation learning network and an SR reconstruction network. There are two modules in the degradation learning network. The

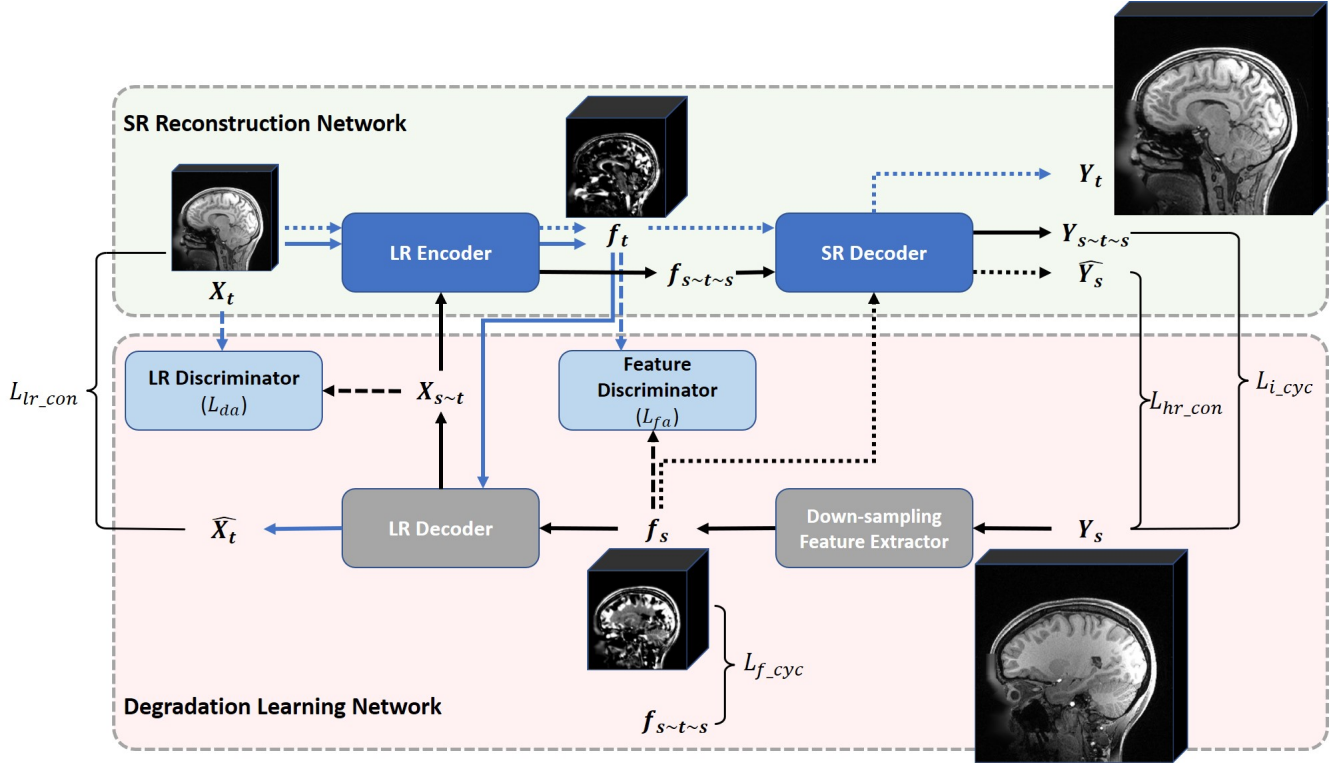


Fig. 2. Pipeline of the UDEAN for 3D MRI super-resolution reconstruction. The network is fed with the unpaired or misaligned source group HR image patch Y_s and the target group LR patch X_t in training. During the inference, only the target group LR image patch X_t is fed in the network, and the SR image patch Y_t is reconstructed.

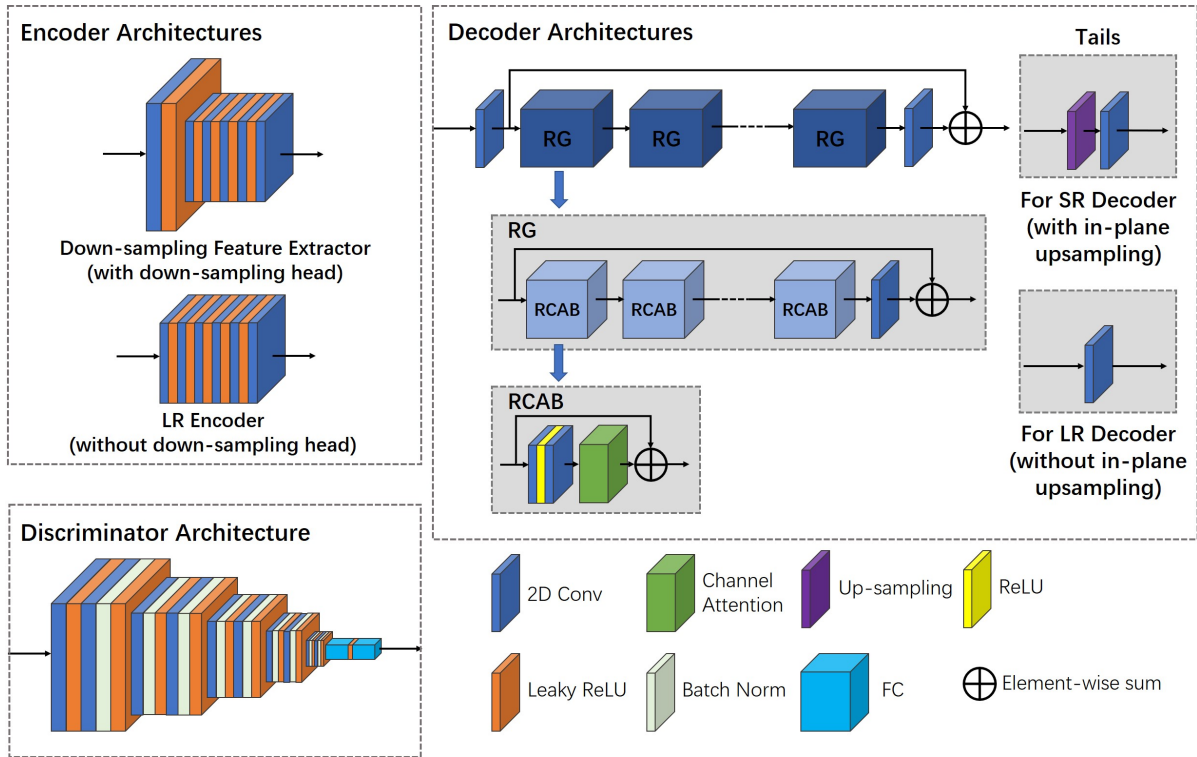


Fig. 3. Detailed structures of network components. The encoders are constructed with 6 convolutional layers, each followed by a Leaky ReLU layer. The decoders adopt the TS-RCAN backbone. The VGG network is employed as the discriminator.

first module is a degradation mapping module, colored gray as shown in Fig. 2. This module contains a downsampling feature extractor and an LR decoder. The second module is an identification module consisting of an LR discriminator and a feature discriminator, as shown in Fig. 2 in light blue color. The SRR network includes an LR encoder and an SR decoder (colored deep blue). It is expected to reconstruct SR images with comparable image quality to the HR images of the source group from the LR images of the target group.

As indicated along the black solid line, $\mathbf{Y}_s$ is fed to a learnable downsampling feature extractor to extract the downsampled feature map f_s of the source group. This feature map contains a generic degradation representation from HR to LR images. Then, the downsampled LR image $X_{s \sim t}$, which contains a comparable degradation representation of the target group LR image $\mathbf{X}_t$, is generated by the LR decoder. Then the feature map $f_{s \sim t \sim s}$ of $X_{s \sim t}$ is extracted using an LR encoder and generates the SR image $Y_{s \sim t \sim s}$ using an SR decoder. As the reconstructed SR image $Y_{s \sim t \sim s}$ and HR image in the source group $\mathbf{Y}_s$ are expected to be identical, and the feature map $f_{s \sim t \sim s}$ and f_s should likewise contain a comparable context, two cycle losses (HR image and feature cycle losses) are calculated for such purposes. Two consistency losses are also involved in making training more stable. The LR encoder and LR decoder are trained by minimizing the target group error between $\mathbf{X}_t$ and the LR image $\hat{X}_t$ decoded from f_t along the blue solid line. Similarly, the downsampling feature extractor and SR decoder are trained by minimizing the source group error between $\mathbf{Y}_s$ and the SR image $\hat{Y}_s$ reconstructed from f_s along the black dotted line.

Since the network is expected to learn the specific degradation representation from the HR images in the source group to the LR images in the target group, an LR discriminator and a feature discriminator are employed. The former is trained to distinguish between the target group LR image $\mathbf{X}_t$ and the LR image $X_{s \sim t}$ reconstructed from source group feature. The second discriminator is trained to distinguish between the feature f_t retrieved from the target group LR and f_s . The discriminators have been trained alternately with the degradation mapping module and the SRR network. The UDEAN can adapt from the source group to the specific degradation representation in the target group as long as the discriminators are deceived.

Only the trained LR encoder and SR decoder are required during the inference step, as shown along the blue dotted line in Fig. 2. The target group LR MRI image $\mathbf{X}_t$ is encoded into the feature map, then the target group SR MRI image $\mathbf{Y}_t$ is generated accordingly.

C. Loss Functions

In our experiments, we employed three types of loss functions, which are the L1 loss, the structural similarity (SSIM) loss [46][56] and the adversarial loss in the training of the generators:

$$L_1(x, y) = \frac{1}{N} \sum_{i=1}^N |x - y| \quad (1)$$

$$L_{SSIM}(x, y) = \frac{1}{N} \sum_{i=1}^N |1 - SSIM(x, y)|^2 \quad (2)$$

$$L_{adv}(x) = \frac{1}{N} \sum_{i=1}^N (D(x) - 1)^2 \quad (3)$$

where the batch size is N , x and y represent the generated SR image and ground truth HR images, and $D(\cdot)$ represents the discriminator. To stabilize the training procedure, we use the least square loss [47] for the adversarial loss in our model instead of the negative log-likelihood [48].

As mentioned in the previous section, the two-cycle losses are used to guide the cross-domain restoration of $Y_{s \sim t \sim s}$. The image cycle loss is calculated as a weighted sum of the L1 loss and SSIM loss between $Y_{s \sim t \sim s}$ and $\mathbf{Y}_s$, and the adversarial loss of $Y_{s \sim t \sim s}$. The feature cycle loss is calculated as the L1 loss between $f_{s \sim t \sim s}$ and f_s :

$$L_{i_cyc} = L_1(Y_{s \sim t \sim s}, \mathbf{Y}_s) + \alpha * L_{SSIM}(Y_{s \sim t \sim s}, \mathbf{Y}_s) + \beta * L_{adv}(Y_{s \sim t \sim s}) \quad (4)$$

$$L_{f_cyc} = L_1(f_{s \sim t \sim s}, f_s) \quad (5)$$

In addition, consistency losses are used to restrict the restoration of $\hat{Y}_s$ and $\hat{X}_t$ within the source and target domain, respectively. They are calculated as a weighted sum of L1, SSIM, and adversarial loss:

$$L_{hr_con} = L_1(\hat{Y}_s, \mathbf{Y}_s) + \alpha * L_{SSIM}(\hat{Y}_s, \mathbf{Y}_s) + \beta * L_{adv}(\hat{Y}_s) \quad (6)$$

$$L_{lr_con} = L_1(\hat{X}_t, \mathbf{X}_t) + \alpha * L_{SSIM}(\hat{X}_t, \mathbf{X}_t) + \beta * L_{adv}(\hat{X}_t) \quad (7)$$

where α and β are the weights of the SSIM and adversarial loss, respectively. We set $\alpha = 0.5$ and $\beta = 0.01$ in our experiments.

Furthermore, the DA losses in image space (L_{da}) and in latent feature space (L_{fa}), which stemmed from the LR discriminator and the feature discriminator, are used to guide the network approaching the distributions of the LR images and the features extracted from different domains:

$$L_{da}(X_{s \sim t}, \mathbf{X}_t) = \frac{1}{N} \sum_{i=1}^N |D_{da}(X_{s \sim t}) - 0.5| + \frac{1}{N} \sum_{i=1}^N |D_{da}(\mathbf{X}_t) - 0.5| \quad (8)$$

$$L_{fa}(f_t, f_s) = \frac{1}{N} \sum_{i=1}^N |D_{fa}(f_t) - 0.5| + \frac{1}{N} \sum_{i=1}^N |D_{fa}(f_s) - 0.5| \quad (9)$$

As a result, the end-to-end training loss for the generators in our network is defined as

$$Loss_G = \lambda_1 * L_{i_cyc} + \lambda_2 * L_{f_cyc} + \lambda_3 * L_{hr_con} + \lambda_4 * L_{lr_con} + \lambda_5 * L_{da} + \lambda_6 * L_{fa} \quad (10)$$

where λ_1 to λ_6 are the weights of the loss components. We set $\lambda_{1,3,4} = 1$ and $\lambda_{2,5,6} = 0.1$ in our experiments for the best performance.

For the training of the LR discriminator (LRD) and the feature discriminator (FD), we also employed the least square loss:

$$L_{LRD}(\mathbf{X}_t, X_{s \sim t}) = \frac{1}{N} \sum_{i=1}^N (D_{da}(\mathbf{X}_t) - 1)^2 + \frac{1}{N} \sum_{i=1}^N (D_{da}(X_{s \sim t}) - 0)^2 \quad (11)$$

$$L_{FD}(f_s, f_t) = \frac{1}{N} \sum_{i=1}^N (D_{fa}(f_s) - 1)^2 + \frac{1}{N} \sum_{i=1}^N (D_{fa}(f_t) - 0)^2 \quad (12)$$

D. Dataset and Data Pre-processing

In this study, we used two public datasets:

1) *Human Connectome Project (HCP) dataset*: The T1w images from the HCP dataset, which consists of images from 1113 healthy participants [49]. The T1w images were acquired in the sagittal plane with 3D MPRAGE, $TR = 2400ms$, $TE = 2.14ms$, $TI = 1000ms$, $FA = 8$, $FOV = 224mm$, $rBW = 210Hz/Px$, the matrix size was $320 \times 320 \times 256$ with isotropic resolution $0.7mm$, and $\times 2$ GRAPPA was activated in the phase encoding direction. To shorten the training time, we randomly selected 300 participants for our study.

2) *Brain Tumor Segmentation Challenge 2020 (BraTS) dataset*: The contrast-enhanced T1w (T1CE) images from the BraTS dataset, which consists of images from 369 participants with brain tumors [50], [51], [52]. The T1CE images were acquired in the axial plane, the matrix size was $240 \times 240 \times 155$ with isotropic resolution $1.0mm$. To shorten the training time, we randomly selected 300 participants for our study.

Our experiments involved four data groups: the source, target, validation, and test groups. The source and target groups were used to train the neural networks, the validation group was used to monitor the network performance during training, and the test group was used to evaluate the networks after training. We randomly selected 120/120/30/30 participants from the HCP dataset for the source, target, validation, and test group. The target group contains only LR images, which were downsampled from HR images by 3D K -space truncation with scale factors of $2 \times 2 \times 1$ and $2 \times 2 \times 2$ [2], [3], [4], [5], [6], [7], [8], [9], [10]. The source group contained only HR images from 120 participants and differed between the experiments with misaligned/unpaired HR and LR images. For the unpaired SRR, the participants of the source group were isolated from the other three groups; for the misaligned SRR, the source group shared the same participants with the target group, and the HR images were distorted with certain deformation patterns.

Specifically, for the SRR from misaligned LR and HR image pairs, the participant movement between the acquisition of HR

and LR images was simulated by adopting mild rigid movement and non-rigid geometric deformation to the HR images. The rigid movement was performed by random rotation of the image volumes by 0 to 2 degrees around the head-feet (H-F) axis and the left-right (L-R) axis and random translation by 0 to 2 voxels in H-F and head-feet L-R directions. The geometric deformation was achieved by randomly shrinking the whole image volumes by 0 to 2 voxels in the anterior-posterior (A-P) and L-R directions, resulting in a 0.7 to 0.9% change in the objects' sizes. Each dataset was scaled to the range of 0 to 1, and to save computation resources, the LR and HR images were cropped into 3D patches of sizes $64 \times 64 \times 3$ and $128 \times 128 \times 3$ for the scale factor of $2 \times 2 \times 1$, and $64 \times 64 \times 3$ and $128 \times 128 \times 6$ for the scale factor of $2 \times 2 \times 2$, respectively.

E. Implementation Details

The detailed structures of our model components are shown in Fig. 3. For the encoders, we used 6 convolution layers with Leaky ReLU between every two layers, and the downsampling feature extractor downsamples the input HR image to the same size as the LR image. We used TS-RCAN [28] with 5 residual groups (RG) and 5 residual blocks (RCAB) in each RG as the backbone of the decoders and VGG [55] as the discriminator for the UDEAN. TS-RCAN is modified from RCAN [53], [54] to conduct 3D MRI SR tasks with low consumption of computation resources and short inference time.

The networks were trained on a workstation equipped with an Nvidia Quadro A6000. For all deep learning experiments, we used Pytorch 1.9 as the back end. In each training batch, eight LR patches were randomly extracted as inputs. We trained our model for 30 epochs using the ADAM optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\epsilon = 10^{-8}$, and a Cosine-decay learning rate was applied from 10^{-4} to 10^{-8} . The image pre-processing of downsampling, deformation, and cropping, the post-processing, and metrics calculation were performed with Matlab 2020a. The rigid image registration was performed on Amira 3D with metrics of extended mutual information.

The image quality was evaluated using the most widely used metrics, the signal-to-noise ratio (PSNR) and the structure similarity index (SSIM) [2], [3], [4], [6], [7], [8], [10]. The higher values of PSNR and SSIM represent better performance. Unlike regular image SRR studied in computer vision research, we didn't adopt any perceptual metrics since structural accuracy is more important than the pure perceptual effect in clinical diagnosis and image-guided treatments.

III. EXPERIMENTS AND RESULTS

A. Illustration of Degradation Adaptation

As shown in Fig. 4, images with similar anatomical structures were selected from different participants in the source group (in the top row) and the target group (in the bottom row). $\mathbf{Y}_s$ was the HR image from the source group and $\mathbf{Y}_t$ was the SR image reconstructed using the LR image (X_t) from the target group. $X_{s \sim t}$ was the LR image downsampled from $\mathbf{Y}_s$ by the degradation learning network of UDEAN. X_t and $X_{s \sim t}$ had highly comparable image quality, this was achieved by the degradation adaptation mechanism of the network. f_{s1}

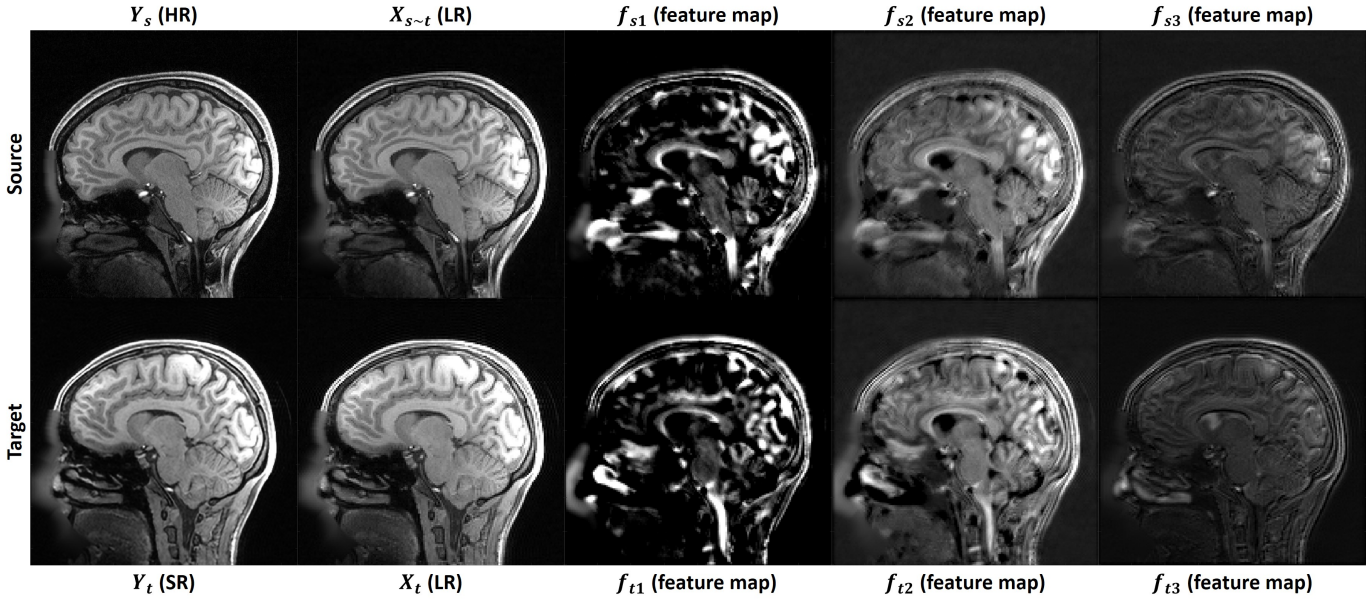


Fig. 4. Illustration of the degradation adaptation and feature extraction. Images with similar anatomical structures were selected from different participants in the source group (in the top row) and the target group (in the bottom row). Y_s was the HR image from the source group, and Y_t was the SR image reconstructed using the LR image (X_t) from the target group. $X_{s~t}$ was the LR image degraded from Y_s by the degradation learning network, and its image quality was comparable to X_t . f_{s1} to f_{s3} and f_{t1} to f_{t3} were examples of the feature maps extracted from Y_s and X_t , respectively. Highly consistent patterns could be found between the feature maps of the source group and the target group, revealing the effectiveness of the degradation adaptation in the latent feature space.

TABLE I
QUANTITATIVE RESULTS OF ABLATION STUDY FOR MRI SRR WITH MISALIGNED LR AND HR IMAGES OF THE HCP DATASET AND SCALE FACTOR OF $2 \times 2 \times 2$.

	Supervised without Rigid-Registration	Supervised with Rigid-Registration	Unsupervised with DA in Feature Space	Unsupervised with DA in Image Space	Unsupervised with DA in Both Spaces
SSIM	0.8939 ± 0.0140	0.9025 ± 0.0116	0.9162 ± 0.0104	0.9229 ± 0.0093	0.9247 ± 0.0097
PSNR	32.7886 ± 1.5546	32.7823 ± 1.5013	33.5224 ± 1.6849	33.5642 ± 1.7719	34.2729 ± 1.8080

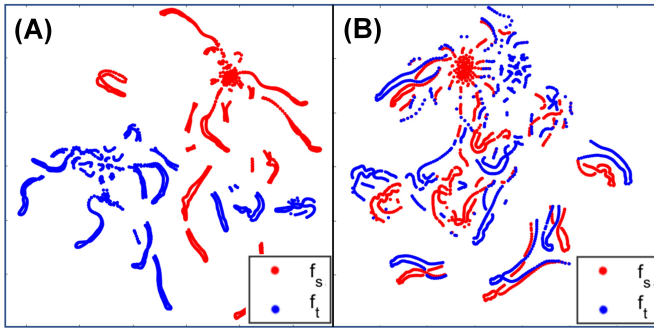


Fig. 5. Illustration of the degradation adaptation process using t-SNE. (A): The distributions of the feature maps from the source domain (f_s) and the target domain (f_t) were differentiable before degradation adaptation was performed. (B): After degradation adaptation, the distributions of the feature maps overlapped.

to f_{s3} and f_{t1} to f_{t3} were the feature maps extracted from Y_s and X_t , respectively. The patterns in the feature maps of the source group and the target group were highly consistent, revealing the effectiveness of the degradation adaptation in the latent feature space. With a certain consistency between the source group and the target group in both the image space and the latent feature space, UDEAN was able to reconstruct SR images from LR images from the target group with a quality

comparable to that of HR images from the source domain. The degradation adaptation process can also be observed using t-SNE to show the distribution of extracted features in the feature space before and after degradation adaptation, as shown in Fig. 5.

B. Ablation Study of the UDEAN

We evaluated our network in various configurations on the same SRR task for misaligned LR and HR datasets to find the best performance. The effect of DA in the image space and latent feature space was evaluated. Furthermore, we compared our network to the same backbone network trained in the supervised strategy with misaligned training data to reveal the effect of the degradation learning modules. As shown in Table I, all the configurations of the UDEAN outperformed the supervised learning using the same generator by over 0.014/0.74 dB in SSIM/PSNR using misaligned LR and HR image from the HCP dataset with the scale factor of $2 \times 2 \times 2$. Among the configurations of the UDEAN, the performance was downgraded by 0.009/0.75 dB or 0.002/0.71 dB in SSIM/PSNR when DA in only latent feature space or image space was applied, respectively.

Furthermore, Fig. 6 shows the visual effect of the reconstructed SR images and the error maps to the HR images.

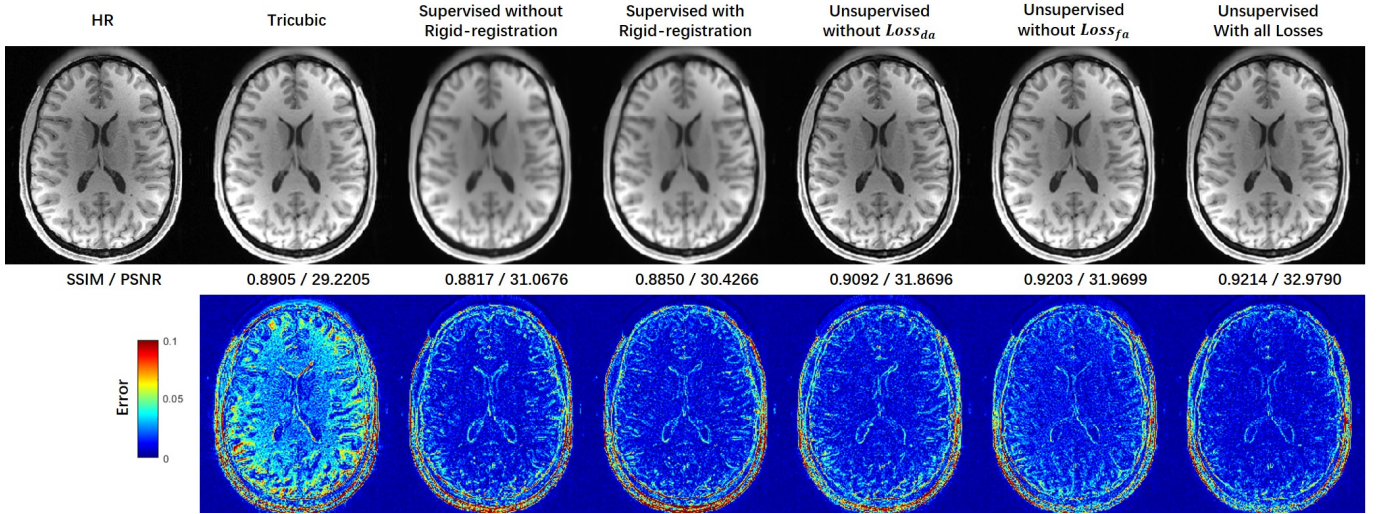


Fig. 6. Comparison of reconstructed SR images and error maps of supervised learning and different configurations of the UDEAN with misaligned LR and HR image pairs of the HCP dataset and the scale factor of $2 \times 2 \times 2$ in the axial plane. The tricubic interpolated image has many errors in the detailed structure. SR images reconstructed using supervised learning are highly blurry, even with a geometric deformation below 1%. The contribution of degradation adaptation can be observed in the images with a single type of degradation adaptation and with both degradation adaptations. The absence of any degradation adaptation leads to lower accuracy of the reconstructed images.

The SR images reconstructed by the supervised methods were highly blurry with or without rigid image registration. On the contrary, the quality of SR images reconstructed by the UDEAN was significantly improved even with DA only in latent feature space or image space. The errors of the reconstructed anatomical structures were further reduced when DA was applied in both domains in the training process.

TABLE II
COMPARISON OF MODEL EFFICIENCIES FOR INFERENCE WITH THE SCALE FACTOR OF $2 \times 2 \times 2$.

	# Params (M)	GFlops	Inference Time (s)
ZSSR [30]	2.457	10.055	2.26 ± 0.04
DASR [36]	2.457	10.055	2.33 ± 0.03
PseudoSR [35]	4.765	19.455	5.27 ± 0.07
Blind-SR [37]	4.765	19.455	5.23 ± 0.06
UDEAN(ours)	2.457	10.055	2.30 ± 0.03

C. Comparison with the State-of-the-Art Unsupervised Methods

We implemented other state-of-the-art unsupervised networks, including ZSSR [30], DASR [36], PseudoSR [35], and Blind-SR [37], as the baseline. All these networks used the same backbone as the UDEAN for both generators and discriminators and trained with identical training setups, including training datasets and hyperparameters, for fair comparisons. For inference, UDEAN and DASR used a single upsampling network with the same structure. As a result, the inference process for UDEAN and DASR consumed identical computational resources, including the same number of network parameters, number of operations, GPU consumption, and inference time. Blind-SR and PseudoSR used another network for style transfer in addition to the upsampling network, thus consuming approximately twice as many computation resources as the other two. More detailed

numerical comparisons on number of parameters, FLOPS, and inference time, are shown in Table II. The numerical results in Table III reveal that the UDEAN outperformed all the above-mentioned networks for MRI SRR from both misaligned and unpaired training data, which are from two public datasets with different contrasts and two scale factors. Specifically, for the HCP dataset and both scale factors, UDEAN outperformed the other networks in the SSIM and PSNR values for both misaligned and unpaired training data, only except the PSNR value of ZSSR with the scale factor of $2 \times 2 \times 1$. For the BraTS dataset, the UDEAN also outperformed all other networks with the highest SSIM and PSNR values in most experiments.

As for the qualitative comparison shown in Fig. 7 for the SRR of the HCP dataset with a scale factor of $2 \times 2 \times 2$, ZSSR generated highly blurry images, although its metrics values are higher than DASR and PseudoSR in some cases. Among all the images, the ones reconstructed by UDEAN achieved the lowest errors as shown in the error maps, and the best accuracy in small anatomical structures pointed out by the red arrows. Also shown in Fig. 8 for the SRR with the BraTS dataset and the scale factor of $2 \times 2 \times 2$, UDEAN achieved better accuracy than the other networks in the structures of lesions, which may improve the accuracy of the assessment of the lesions and the therapy based on the reconstructed SR images.

IV. DISCUSSION

A. Practicality in Clinical Settings

In real clinical settings, perfectly paired and aligned authentic LR and HR images are difficult to acquire due to patient movements. Besides, in-vivo images are mostly acquired with either LR or HR images in clinical settings because of high time consumption and cost. Lack of paired and aligned LR and HR images raises challenges in implementing supervised SRR methods. Our experimental results revealed that supervised methods could not provide acceptable image quality if the LR

TABLE III

QUANTITATIVE COMPARISON WITH THE STATE-OF-THE-ART UNSUPERVISED NETWORKS FOR MRI SRR WITH MISALIGNED AND UNPAIRED LR AND HR IMAGES

Dataset	Scale Factor	Method	Misaligned or Unpaired	SSIM	PSNR
HCP	$2 \times 2 \times 2$	Tricubic	N/A	0.8981 ± 0.0106	31.5862 ± 1.8520
		ZSSR [30]		0.8994 ± 0.0163	33.0213 ± 2.2054
		DASR [36]	Unpaired	0.8931 ± 0.0094	32.3309 ± 1.4671
		Pseudo SR [35]		0.8931 ± 0.0116	31.0407 ± 1.8816
		Blind-SR [37]		0.9132 ± 0.0100	33.0695 ± 1.4656
		UDEAN(ours)		0.9242 ± 0.0083	33.5159 ± 1.7234
		DASR [36]	Misaligned	0.8946 ± 0.0091	32.2544 ± 1.4658
		Pseudo SR [35]		0.8962 ± 0.0086	32.0267 ± 1.5950
		Blind-SR [37]		0.9149 ± 0.0091	33.1597 ± 1.5693
		UDEAN(ours)		0.9247 ± 0.0097	34.2729 ± 1.8080
	$2 \times 2 \times 1$	Tricubic	N/A	0.9227 ± 0.0088	33.6545 ± 1.8492
		ZSSR [30]		0.9412 ± 0.0106	36.0848 ± 2.4098
		DASR [36]	Unpaired	0.9137 ± 0.0084	33.6732 ± 1.4502
		Pseudo SR [35]		0.9244 ± 0.0076	33.8122 ± 1.6044
		Blind-SR [37]		0.9325 ± 0.0081	34.9527 ± 1.7325
		UDEAN(ours)		0.9371 ± 0.0078	35.1688 ± 1.8195
BraTS	$2 \times 2 \times 2$	DASR [36]	Misaligned	0.9192 ± 0.0082	33.9043 ± 1.6386
		Pseudo SR [35]		0.9228 ± 0.0080	34.0500 ± 1.6694
		Blind-SR [37]		0.9395 ± 0.0071	35.0236 ± 1.8130
		UDEAN(ours)		0.9455 ± 0.0072	35.5182 ± 2.2839
		Tricubic	N/A	0.9073 ± 0.0196	31.2470 ± 1.9187
		ZSSR [30]		0.9001 ± 0.0464	29.6156 ± 3.1528
	$2 \times 2 \times 1$	DASR [36]	Unpaired	0.9334 ± 0.0141	30.5198 ± 2.0642
		Pseudo SR [35]		0.8978 ± 0.0129	26.4210 ± 0.8741
		Blind-SR [37]		0.9434 ± 0.0121	31.2900 ± 2.2618
		UDEAN(ours)		0.9471 ± 0.0096	31.4467 ± 2.3570
		DASR [36]	Misaligned	0.9377 ± 0.0151	31.1546 ± 2.0495
		Pseudo SR [35]		0.9378 ± 0.0150	31.0804 ± 1.9391
		Blind-SR [37]		0.9360 ± 0.0120	31.2894 ± 1.9017
		UDEAN(ours)		0.9513 ± 0.0133	32.0435 ± 2.2796
BraTS	$2 \times 2 \times 2$	Tricubic	N/A	0.9299 ± 0.0151	34.8846 ± 2.2722
		ZSSR [30]		0.9514 ± 0.0156	33.2469 ± 3.6422
		DASR [36]	Unpaired	0.9462 ± 0.0149	34.4395 ± 1.8905
		Pseudo SR [35]		0.9397 ± 0.0150	34.0382 ± 1.9015
		Blind-SR [37]		0.9510 ± 0.0156	35.3235 ± 2.0213
		UDEAN(ours)		0.9551 ± 0.0166	35.4973 ± 2.2433
	$2 \times 2 \times 1$	DASR [36]	Misaligned	0.9461 ± 0.0167	34.5953 ± 2.0557
		Pseudo SR [35]		0.9463 ± 0.0162	34.5560 ± 1.9037
		Blind-SR [37]		0.9558 ± 0.0158	36.1029 ± 2.2270
		UDEAN(ours)		0.9629 ± 0.0163	36.7631 ± 2.3787

and HR images in the training datasets were not perfectly aligned. Even with rigid image registration, the inference results were still unsatisfactory if there was a tiny geometric deformation between the LR and HR images. Besides, considering the change in object size in our experiments was between 0.7% and 0.9%, which was far beyond the accuracy of 90% DSC achieved by the latest deep learning-based deformable image registration [25], the deformable image registration algorithms are not helpful in solving such tiny misalignments of soft tissues. Therefore, previous studies adopted synthetic LR images generated from HR images with certain downsampling algorithms as the input of the network [2], [3], [4], [6], [7], [8], [10]. However, the degradation shift between the artificial downsampling and the real degradation from authentic HR to LR images downgrades the performance of the trained networks when they are used for SRR from authentic LR images [29]. Therefore, unsupervised networks can be a possible solution to these problems.

Previously proposed unsupervised networks were designed for 2D real-world images [30], [34], [36], [35]. However, medical images are mostly in 3D form, containing through-slice information. Therefore, we adopted the TS-RCAN [28], which

is a low-cost network for 3D SRR with enhanced performance, as the backbone of the UDEAN to process 3D medical images. And compared to the other unsupervised method, UDEAN learned the degradation representation in both the image space and the latent feature space, thus achieving better performance.

B. Effect of Degradation Representation Learning

The degradation representation learning of the UDEAN was a specific type of domain adaptation, which transferred HR images of the source domain to the LR images of the target domain. As illustrated in the result section, UDEAN was able to downsample the HR image from the source group to the LR image, which had comparable quality to the target group. Consistency in the latent feature space between the source group and the target group further helped the network with improved accuracy. The ablation study also revealed the effects of domain adaptations, which guided the network to learn the degradation representation in both the image and the latent feature space. The experimental results showed the advantage of UDEAN over the supervised learning approach in misaligned LR and HR images. When the LR and HR

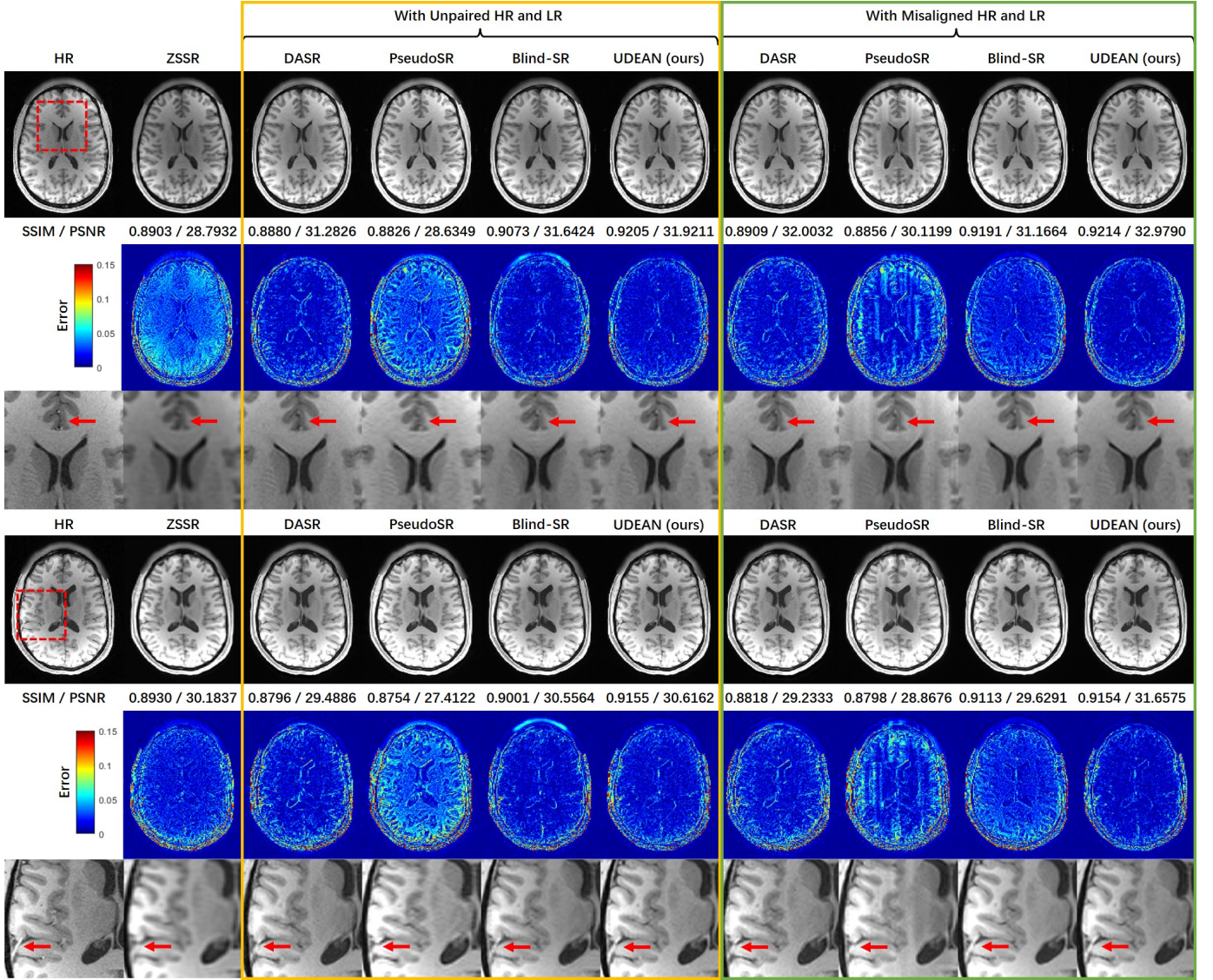


Fig. 7. Qualitative comparison of the UDEAN with the state-of-the-art unsupervised networks in visual effect and error maps in the axial plane with the HCP dataset and the scale factor of $2 \times 2 \times 2$. The arrows point to the small structures that UDEAN reconstructed with higher accuracy than the other networks.

images of the training datasets were not perfectly aligned, the supervised methods failed to provide acceptable image quality. Besides, the ablation study also revealed the effect of degradation representation learning in both image and latent feature space. The network could reconstruct the SR images without either $Loss_{da}$ or $Loss_{fa}$. However, the absence of $Loss_{da}$ or $Loss_{fa}$ downgraded the network's performance, showing the benefit of degradation representation learning in both the image space and the latent feature space.

C. Comparison of Different Domain Transfer Strategies

Regarding the other state-of-the-art unsupervised networks, the ZSSR, as a network without domain transfer, learned the degradation representation between the LR images and the lower-resolution images, which were downsampled from the LR images, and used the learned degradation representation to reconstruct the SR images from the LR images [30]. However, there was a large gap between the learned degradation and the

degradation from HR images to the LR images, thus leading to downgraded performance, particularly with large-scale factors. Although it achieved high metrics values in a few cases, the reconstructed SR images were highly blurry as shown in our results, and cannot meet the quality required in clinics.

Most of the unsupervised networks with domain transfer are derived from CycleGAN. As one of the latest unsupervised SR approaches, the DASR learns the degradation representation by downsampling the HR images of the source domain to LR images in the target domain, and domain distance maps in the image space are generated simultaneously, which are subsequently used to reconstruct SR images [36]. However, this learning strategy was not end-to-end, so the early stop when the downsampling network (DSN) reached the best performance in the validation dataset was not applicable. The DSN could be overfitted to the training dataset, making the whole network difficult to train and optimize, and the learned degradation representation hardly fits the reconstruction proce-

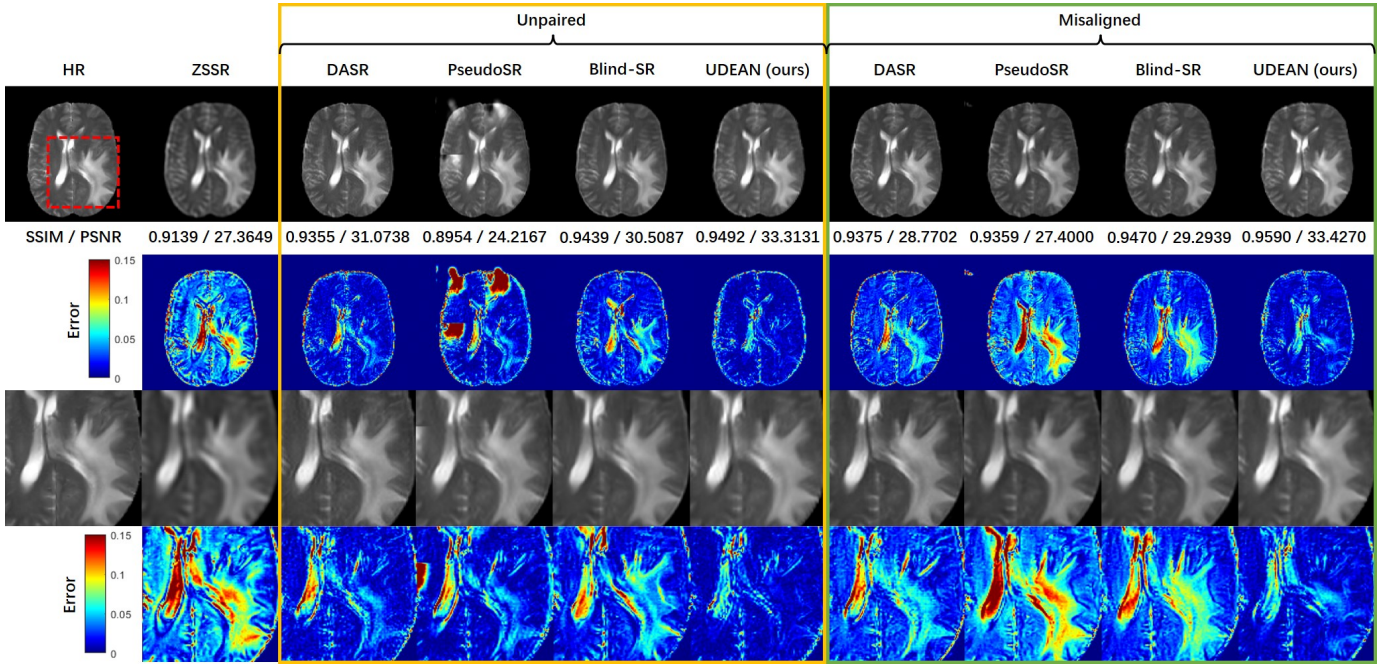


Fig. 8. Qualitative comparison of the UDEAN with the state-of-the-art unsupervised networks in visual effect and error maps in the axial plane with the BraTS dataset and the scale factor of $2 \times 2 \times 2$. UDEAN achieved the lowest error in the region of the lesion.

dure. As a result, the inaccuracies in the DASR’s reconstructed SR MRI images are still massively visible.

PseudoSR is another most recent approach to an unsupervised SRR network. It transfers the LR image of the target domain to the source domain, where the LR images are downgraded from HR images with a predefined algorithm (normally bicubic or Gaussian blurring). Then the SR images are reconstructed from downsampled LR images [35]. The experimental results showed that the PseudoSR could achieve neither high metrics values nor satisfactory qualitative image quality. Despite the adaptation of latent feature space, another main difference between the UDEAN and the PseudoSR was that the UDEAN transfers the source domain HR images to the target domain and learns the mapping of the LR images from the target domain to the HR images in the source domain. On the contrary, the PseudoSR transfers the target domain LR images to the source domain and learns the mapping of the LR images from the source domain to the HR images in the source domain. The ablation study results showed the advantage of the former scheme even with the absence of feature space adaptation. And with the latter approach, more detailed structures were missing in the reconstructed SR images.

Blind-SR is another variation of CycleGAN with simple modifications proposed after UDEAN¹. It shares a similar scheme with the PseudoSR, which transfers the LR image of the target domain to the source domain, whose images are downsampled with bicubic or nearest algorithms. Therefore, it faced a similar problem to PseudoSR, with which much detailed structural information was lost. In addition, the Blind-SR was not trained in an end-to-end fashion either. Its two

components were trained separately, making it face similar difficulties in optimizing the network like DASR. Moreover, Blind-SR showcases performance degradation from UDEAN in both SSIM and PNSR, as well as low quality of MRI SRR in visual effect as shown in Fig.7 and Fig.8. The aforementioned drawbacks make Blind-SR impractical in clinical treatment.

Besides the differences in the domain transfer strategies, UDEAN and DASR utilized a single upsampling network for the inference, whereas PseudoSR and Blind-SR used another style transfer network in addition to the upsampling network. As a result, the computation resources consumed by UDEAN were identical to DASR and approximately 50% less than PseudoSR and Blind-SR.

D. Limitations and Future Works

Despite the promising results of UDEAN, we acknowledge certain limitations in this study. We only tested the UDEAN on brain MRI images, where the geometric deformation is normally mild in real clinics, and the LR images were synthetically generated from HR images. Abdominal and musculoskeletal imaging normally suffer from more severe geometric deformation of soft tissues due to irresistible patient movements. Therefore, the UDEAN can be the best solution in these scenarios and will be investigated in our future study. Besides, public datasets containing authentic LR and HR images of abdominal and musculoskeletal imaging are unavailable. Therefore, the misaligned or unpaired LR and HR images still need to be artificially generated. This study focused on investigating the potential of SRR using networks trained by misaligned or unpaired LR and HR images, the datasets we used in our experiments were still applicable to our experiments. Our future study will investigate the performance of UDEAN trained with authentic LR and HR images.

¹<https://arxiv.org/abs/2205.06891v1>, released on 13 May 2022.

V. CONCLUSION

In this manuscript, we propose the UDEAN as an unsupervised degradation adaptation network that adaptively learns the degradation representation between misaligned or unpaired LR and HR MRI data in both image and latent feature spaces. UDEAN applies SRR in an end-to-end fashion, making the network easy to train and optimize. The DA mechanism adopted by the UDEAN also shows advantages over previously used domain transfer methods, thus minimizing errors in the reconstructed SR images. Our experimental results showed that the UDEAN alleviated the problem of lacking paired authentic LR and HR images, achieved enhanced image quality for SRR, and outperformed the other state-of-the-art networks. Therefore, UDEAN is a promising solution for SRR in clinical settings when perfectly aligned LR and HR image pairs are unavailable. Despite the significant advancements in medical imaging, there is still a shortage of large-scale LR and HR MRI images measured from patients in the clinical environment. As a result, all current state-of-the-art methods, including ours, have been trained using authentic HR images and artificially generated LR images from authentic HR images. However, we plan to incorporate large-scale clinical LR images into our future work to enhance our studies.

REFERENCES

- [1] S. Wang *et al.*, "Accelerating magnetic resonance imaging via deep learning," *Proc. IEEE Int. Symp. Biomed. Imag.*, Prague, Czech Republic, 2016, pp. 514–517.
- [2] X. Zhao, Y. Zhang, T. Zhang, and X. Zou, "Channel splitting network for single MR image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5649–5662, 2019.
- [3] Y. Zhang, K. Li, K. Li, and Y. Fu, "MR image super-resolution with squeeze and excitation reasoning attention network," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Virtual, 2021, pp. 13425–13434.
- [4] C. H. Pham, A. Ducournau, R. Fablet, and F. Rousseau, "Brain MRI super-resolution using deep 3D convolutional networks," *Proc. IEEE Int. Symp. Biomed. Imag.*, Melbourne, Australia, 2017, pp. 197–200.
- [5] C. H. Pham *et al.*, "Multiscale brain MRI super-resolution using deep 3D convolutional networks," *Comput. Med. Imaging. Graph.*, vol. 77, 101647, 2019.
- [6] 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," *Proc. IEEE Int. Symp. Biomed. Imag.*, Washington D.C., USA, 2018, pp. 739–742.
- [7] 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," *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Granada, Spain, 2018, pp. 91–99.
- [8] Y. Sui, O. Afacan, A. Gholipour, and S. K. Warfield, "Learning a gradient guidance for spatially isotropic MRI super-resolution reconstruction," *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Lima, Peru, 2020, pp. 136–146.
- [9] Q. Lyn *et al.*, "Multi-Contrast Super-Resolution MRI Through a Progressive Network," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2738–2749, 2020.
- [10] C. Feng, H. Fu, S. Yuan, and Y. Xu, "Multi-contrast MRI super-resolution via a multi-stage integration network," *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Strasbourg, France, 2021, pp. 140–149.
- [11] Y. Mei, Y. Fan, and Y. Zhou, "Image Super-Resolution with Non-Local Sparse Attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Virtual, 2021, pp. 3516–3525.
- [12] L. Sun, J. Pan, and J. Tang, "Shufflelexer: An efficient convnet for image super-resolution," *Adv Neural Inf Process Syst.*, vol. 35, pp. 17314–17326, 2022.
- [13] J. Su, M. Gan, G. Chen, W. Guo, and C. Chen, "High-Similarity-Pass Attention for Single Image Super-Resolution," *IEEE Trans. Image Process.*, vol. 33, pp. 610–624, 2024.
- [14] K. Jiang, Z. Wang, P. Yi, G. Wang, K. Gu, and J. Jiang, "ATMFN: Adaptive-Threshold-Based Multi-Model Fusion Network for Compressed Face Hallucination," *IEEE Trans. Multimedia*, vol. 22, no. 10, pp. 2734–2747, Oct. 2020.
- [15] K. Jiang, Z. Wang, P. Yi, T. Lu, J. Jiang, and Z. Xiong, "Dual-Path Deep Fusion Network for Face Image Hallucination," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 1, pp. 378–391, Jan. 2022.
- [16] Y. Xiao, Q. Yuan, K. Jiang, J. He, X. Jin, and L. Zhang, "EDiffSR: An Efficient Diffusion Probabilistic Model for Remote Sensing Image Super-Resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–14, 2024.
- [17] Y. Xiao, Q. Yuan, K. Jiang, J. He, Y. Wang, and L. Zhang, "From degrade to upgrade: Learning a self-supervised degradation guided adaptive network for blind remote sensing image super-resolution," *Inf. Fusion*, vol. 96, 297–311, 2023.
- [18] Y. Xiao, X. Su, Q. Yuan, D. Liu, H. Shen, and L. Zhang, "Satellite Video Super-Resolution via Multiscale Deformable Convolution Alignment and Temporal Grouping Projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, 2022.
- [19] Y. Xiao *et al.*, "Local-Global Temporal Difference Learning for Satellite Video Super-Resolution," *IEEE Trans. Circuits Syst. Video Technol.*, doi: 10.1109/TCSVT.2023.3312321, 2023.
- [20] L. Zhang, J. Nie, W. Wei, and Y. Zhang, "Unsupervised Test-Time Adaptation Learning for Effective Hyperspectral Image Super-Resolution With Unknown Degeneration," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2024, doi: 10.1109/TPAMI.2024.3361894.
- [21] J. Zhang, J. Liu, J. Yang, and Z. Wu, "Crossed Dual-Branch U-Net for Hyperspectral Image Super-Resolution," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 2296–2307, 2024.
- [22] X. Zhang, C. Song, T. You, Q. Bai, W. Wei, and L. Zhang, "Dual ODE: Spatial-Spectral Neural Ordinary Differential Equations for Hyperspectral Image Super-Resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–15, 2024.
- [23] Y. Sui, O. Afacan, A. Gholipour, and S. K. Warfield, "MRI super-resolution through generative degradation learning," *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Strasbourg, France, 2021, pp. 430–440.
- [24] C. Komninos *et al.*, "Intra-operative oct (ioct) image quality enhancement: a super-resolution approach using high quality ioct 3d scans," *Ophthalmic Medical Image Analysis (OMIA), in Conjunction with MICCAI*, vol. 12970, 2021, pp. 21–31.
- [25] A. Zakeri *et al.*, "DragNet: Learning-based deformable registration for realistic cardiac MR sequence generation from a single frame," *Medical Image Analysis*, vol. 83, 2023, doi: https://doi.org/10.1016/j.media.2022.102678.
- [26] Z. Sui, O. Afacan, C. Jaimes, A. Gholipour, and S. K. Warfield, "Scan-Specific Generative Neural Network for MRI Super-Resolution Reconstruction," *IEEE Trans. Med. Imag.*, vol. 41, no. 6, pp. 2738–2749, 2022.
- [27] C. Zhao, B. E. Dewey, D. L. Pham, P. A. Calabresi, D. S. Reich, and J. L. Prince, "SMORE: A Self-Supervised Anti-Aliasing and Super-Resolution Algorithm for MRI Using Deep Learning," *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 805–817, 2021.
- [28] H. Li and J. Liu, "3D High-Quality Magnetic Resonance Image Restoration in Clinics Using Deep Learning," 2021, *arXiv:2111.14259*.
- [29] S. Laguna *et al.*, "Super-resolution of portable low-field MRI in real scenarios: integration with denoising and domain adaptation," *Proc. Medical Imaging with Deep Learning*, Zürich, Switzerland, 2022.
- [30] A. Shocher, C. Nadav Cohen, and I. Michal, "zero-shot" super-resolution using deep internal learning," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, USA, 2018, pp. 3118–3126.
- [31] Y. Iwamoto, K. Takeda, Y. Li, A. Shiino, and Y. Chen, "Unsupervised MRI Super-Resolution Using Deep External Learning and Guided Residual Dense Network with Multimodal Image Priors," 2020, *arXiv:2008.11921*.
- [32] J. Cui, K. Gong, P. Han, H. Liu, and Q. Li, "Unsupervised arterial spin labeling image superresolution via multiscale generative adversarial network," *Med. Phys.*, vol. 49, no. 4, pp. 2373–2385, 2022.
- [33] L. Wang *et al.*, "Unsupervised degradation representation learning for blind super-resolution," *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Virtual, 2021, pp. 10581–10590.
- [34] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *Proc.*

- IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, Hawaii, USA, 2017, pp. 2223–2232.
- [35] S. Maeda, “Unpaired image super-resolution using pseudo-supervision,” *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Virtual, 2020, pp. 291–300.
- [36] Y. Wei, S. Gu, Y. Li, R. Timofte, L. jin, and H. Song, “Unsupervised real-world image super resolution via domain-distance aware training,” *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Virtual, 2021, pp. 13385–13394.
- [37] H. Zhou, Y. Huang, Y. Li, Y. Zhou, and Y. Zheng, “Blind Super-Resolution of 3D MRI via Unsupervised Domain Transformation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 3, pp. 1409–1418, 2023.
- [38] A. Liu, Y. Liu, J. Gu, Y. Qiao, and C. Dong, “Blind image super-resolution: A survey and beyond,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022, doi: 10.1109/TPAMI.2022.3203009.
- [39] J. Gu, H. Lu, W. Zuo, and C. Dong, “Blind super-resolution with iterative kernel correction,” *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, 2019, pp. 1604–1613.
- [40] L. Xie, X. Wang, C. Dong, Z. Qi, and Y. Shan, “Finding discriminative filters for specific degradations in blind super-resolution,” *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 51–61.
- [41] Z. Hui, J. Li, X. Wang, and X. Gao, “Learning the non-differentiable optimization for blind super-resolution,” *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Virtual, 2021, pp. 2093–2102.
- [42] M. Yamac, A. Baran, and N. Aakif, “Kernelnet: A blind super-resolution kernel estimation network,” *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Virtual, 2021, pp. 453–462.
- [43] W. Zhang, G. Shi, Y. Liu, C. Dong, and X. Wu, “A Closer Look at Blind Super-Resolution: Degradation Models, Baselines, and Performance Upper Bounds,” *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Louisiana, NO, USA, 2022, pp. 527–536.
- [44] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, “Designing a practical degradation model for deep blind image super-resolution,” *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Virtual, 2021, pp. 4791–4800.
- [45] J. Zhang, S. Lu, F. Zhan, and Y. Yu, “Blind image super-resolution via contrastive representation learning,” 2021, *arXiv:2107.00708*.
- [46] E. M. Masutani, B. Naeim, and H. Albert, “Deep learning single-frame and multiframe super-resolution for cardiac MRI,” *Radiology*, vol. 295, no. 3, pp. 552–561, 2020.
- [47] X. Mao, Q. Li, H. Xie, R. Y. K. Raymond, and Z. Wang, “Multi-class generative adversarial networks with the L2 loss function,” 2016, *arXiv:1611.04076*.
- [48] I. Goodfellow et al., “Generative adversarial networks,” *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [49] D. C. Van Essen et al., “The WU-Minn human connectome project: an overview,” *Neuroimage*, vol. 80 pp. 62–79, 2013.
- [50] B. Menze et al., “The multimodal brain tumor image segmentation benchmark (BRATS),” *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, 2014.
- [51] C. T. Lloyd, S. Alessandro, and A. J. Tatem, “High resolution global gridded data for use in population studies,” *Sci. Data*, vol. 4, no. 1, pp. 1–17, 2017.
- [52] S. Bakas et al., “Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge,” 2018, *arXiv:1811.02629*.
- [53] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, 2018, pp. 286–301.
- [54] Z. Lin et al., “Revisiting RCAN: Improved training for image super-resolution,” 2022, *arXiv:2201.11279*.
- [55] K. Simonyan, and Z. Andrew, “Very deep convolutional networks for large-scale image recognition,” 2014, *arXiv:1409.1556*.
- [56] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.



Jianan Liu received his B.Eng. degree in Electronics and Information Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2007. He received his M.Eng. degree in Telecommunication Engineering from the University of Melbourne, Australia, and his M.Sc. degree in Communication Systems from Lund University, Sweden, in 2009 and 2012, respectively. Jianan has over ten years of experience in software and algorithm design and development. He has held senior R&D roles in the AI consulting, automotive, and telecommunication industries. His research interests include applying statistical signal processing and deep learning for medical image processing, wireless communications, IoT networks, indoor sensing, and outdoor perception using a variety of sensor modalities like radar, camera, LiDAR, WiFi, etc.



Hao Li received his B.Eng. degree in Electronics and Information Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2007. He received his M.Sc. degree in Medical System Engineering from the University Magdeburg, Germany, in 2012. He is a research associate at the University Hospital Heidelberg, Germany. He has over ten years of experience in MRI pulse sequence design and medical image processing. He had held senior R&D roles in the healthcare industry and clinical research. His research interests include MRI

pulse sequence design, conventional and deep learning-based medical image processing, etc.



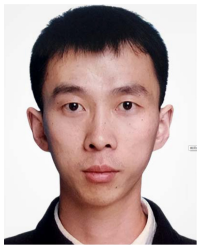
Tao Huang (Senior Member, IEEE) holds a Ph.D. in Electrical Engineering from The University of New South Wales, Sydney, Australia. He also holds an M.Eng. in Sensor System Signal Processing from The University of Adelaide, Adelaide, Australia, and a B.Eng. in Electronics and Information Engineering from Huazhong University of Science and Technology, Wuhan, China. Currently, Dr. Huang is a Senior Lecturer at James Cook University, Cairns, Australia, and serves as the course coordinator for the Master of Engineering (Professional). He was

an Endeavour Australia Cheung Kong Research Fellow, a visiting scholar at The Chinese University of Hong Kong, a research associate at the University of New South Wales, and a postdoctoral research fellow at James Cook University. Before academia, Dr. Huang worked in the industry and held positions such as senior software engineer, senior data scientist, project team lead, and technical lead. Dr. Huang has received the Australian Postgraduate Award, the Engineering Research Award at The University of New South Wales, the Best Paper Award from the IEEE Wireless Communications and Networking Conference, the IEEE Outstanding Leadership Award, and the Citation for Outstanding Contribution to Student Learning at James Cook University. In addition, he is a co-inventor of an international patent on MIMO systems. Dr. Huang is a member of the IEEE Communications Society, IEEE Vehicular Technology Society, IEEE Computational Intelligence Society, and IEEE Industrial Electronics Society. He serves as Vice-Chair of the IEEE Northern Australia Section and local MTT-S/COM Chapter Chair. He also serves in various capacities at international conferences as TPC chair/vice chair, program vice chair, symposium chair, local chair, and TPC member. Dr. Huang is an Associate Editor of the IEEE Open Journal of Communications Society, IEEE Access, and IET Communications. He is also a Topical Advisory Panel Member and Guest Editor of MDPI Electronics. His research interests include electronics systems, deep learning, intelligent sensing, computer vision, pattern recognition, wireless communications, and IoT security.



Euijoon Ahn is a Lecturer at the College of Science and Engineering, James Cook University. Prior to this, He was a postdoctoral research fellow at the Biomedical Multimedia Information Technology (BMIT) group, at the School of Computer Science, The University of Sydney. He obtained his Ph.D. degree in Computer Science (medical image analysis) from The University of Sydney in 2020. He received B. IT degree from The University of Newcastle, Australia, 2009, and M. IT (2014) and MPhil (2016) degree from The University of Sydney. His research

focus is on the development of Machine Learning and Deep Learning, Computer Vision, and more specifically, unsupervised / self-supervised deep learning models for biomedical image analysis, for improving image segmentation, retrieval, quantification, and classification without relying on labeled data. He also works at the coalface of translational health technology research, e.g., health data analytics and telehealth. He has produced top-tier publications in the areas of computer vision and medical image computing, including papers in IEEE T-MI, T-BME, JBHI, MedIA, PR, CVPR, AAAI and MICCAI.



Kang Han received the B.Eng. degree in electronic science and technology from Anhui Polytechnic University, Wuhu, China, in 2014, the M.Eng. degree in circuits and system from Shanghai University, Shanghai, China, in 2017, and the Ph.D. degree from James Cook University, Cairns, Australia, in 2023. He is currently a post-doctoral researcher at La Trobe University, Australia. His research areas include light field image processing, computer vision, and machine learning. In 2017, he won the Certificate of Science and Technology Achievement

of Jiangxi Province, China.



Adeel Razi is an Associate Professor at the Turner Institute for Brain and Mental Health, in the School of Psychological Sciences, Monash University, Australia. He joined Monash, after finishing his post-doctoral studies (2012-2018) at the Wellcome Centre for Human Neuroimaging, UCL, UK. His research is cross-disciplinary, combining engineering, physics, and machine-learning approaches, to model complex, multi-scale, network dynamics of brain structure and function using neuroimaging. He is currently an NHMRC Investigator (Emerging Leadership, 2021-2025), CIFAR Azrieli Global Scholar (2021-2023) in their Brain, Mind, and Consciousness Program, and an ARC DECRA Fellow (2018-2021).

He received a B.E. degree in Electrical Engineering (with a Gold Medal) from the N.E.D. University of Engineering & Technology in Pakistan, the M.Sc. degree in Communications Engineering from the University of Technology Aachen (RWTH), Germany, and the Ph.D. degree in Electrical Engineering from the University of New South Wales, Australia in 2012.



Wei Xiang (S'00-M'04-SM'10) Professor Wei Xiang is Cisco Research Chair of AI and IoT and Director of the Cisco-La Trobe Centre for AI and IoT at La Trobe University. Previously, he was Foundation Chair and Head of the Discipline of IoT Engineering at James Cook University, Cairns, Australia. Due to his instrumental leadership in establishing Australia's first accredited Internet of Things Engineering degree program, he was inducted into Percy Foundation's Hall of Fame in October 2018. He is a TEDx speaker and an elected Fellow of the

IET in the UK and Engineers Australia. He received the TNQ Innovation Award in 2016, and Pearcey Entrepreneurship Award in 2017, and Engineers Australia Cairns Engineer of the Year in 2017. He was a co-recipient of four Best Paper Awards at WiSATS'2019, WCSP'2015, IEEE WCNC'2011, and ICWMC'2009. He has been awarded several prestigious fellowship titles. He was named a Queensland International Fellow (2010-2011) by the Queensland Government of Australia, an Endeavour Research Fellow (2012-2013) by the Commonwealth Government of Australia, a Smart Futures Fellow (2012-2015) by the Queensland Government of Australia, and a JSPS Invitational Fellow jointly by the Australian Academy of Science and Japanese Society for Promotion of Science (2014-2015). He was the Vice Chair of the IEEE Northern Australia Section from 2016-2020. He was an Editor for IEEE Communications Letters (2015-2017) and is currently an Associate Editor for IEEE Communications Surveys & Tutorials, IEEE Internet of Things Journal,

IEEE Access, and Nature Journal of Scientific Reports. He has published over 300 peer-reviewed papers including 3 books and 220 journal articles. He has severed in a large number of international conferences in the capacity of General Co-Chair, TPC Co-Chair, Symposium Chair, etc. His research interest includes the Internet of Things, wireless communications, machine learning for IoT data analytics, and computer vision.



Jinman Kim (Member, IEEE) received the Ph.D. degree in computer science from The University of Sydney, Ultimo, NSW, Australia, in 2006. He was an ARC Post-Doctoral Research Fellow with The University of Sydney and then a Marie Curie Senior Research Fellow with the University of Geneva, Geneva, Switzerland. In 2013, he joined The University of Sydney as a Faculty Member. His research is in the application of machine learning for biomedical image analysis and visualization. His focus is on multimodal data processing and includes

image omics, multimodal data processing, and image data correlation to other health data. He has produced a number of publications in this field and received multiple competitive grants and scientific recognitions. He has actively focused on research translation, where he has worked closely with clinical partners to take his research into clinical practice. He is the Research Director of the Nepean Telehealth Technology Centre (NTTC) at the Nepean hospital, responsible for translational telehealth and digital hospital research. Some of his research has been developed into clinical software that is being used/trialed at multiple hospitals. Dr. Kim is actively involved in his research communities, where he is an Associate Editor of TVCJ and the Vice President of the Computer Graphics Society (CGS). His work on telehealth has been recognized with multiple awards, including the Health Secretary Innovation Award at the NSW Health Innovation Symposium in 2016.



David Dagan Feng (Life Fellow, IEEE) received the M.Eng. degree in electrical engineering and computer science (EECS) from Shanghai Jiao Tong University, Shanghai, China, in 1982, and the M.Sc. degree in biocybernetics and a Ph.D. degree in computer science from the University of California, Los Angeles (UCLA), Los Angeles, CA, USA, in 1985 and 1988, respectively, where he received the Crump Prize for Excellence in Medical Engineering. He is currently the Head of the School of Computer Science, the Director of the Biomedical and Multi-

media Information Technology Research Group, and the Research Director of the Institute of Biomedical Engineering and Technology, The University of Sydney, Sydney, Australia. He has published over 700 scholarly research papers, pioneered several new research directions, and made a number of landmark contributions in his field. More importantly, however, is that many of his research results have been translated into solutions to real-life problems and have made tremendous improvements to the quality of life for those concerned. He is a fellow of the Australian Academy of Technological Sciences and Engineering. He has served as the Chair for the International Federation of Automatic Control (IFAC) Technical Committee on Biological and Medical Systems, has organized/chaired over 100 major international conferences/symposia/workshops, and has been invited to give over 100 keynote presentations in 23 countries and regions.