MRI Restoration Using Edge-Guided Adversarial Learning

YAQIONG CHAI1,2,*, BOTIAN XU1,*, KANGNING ZHANG3, NATASHA LEPORE1,2, AND JOHN C. WOOD1,4

1Department of Biomedical Engineering, University of Southern California, Los Angeles, CA 90007, USA
2CIBORG Laboratory, Department of Radiology, Children’s Hospital Los Angeles, Los Angeles, CA 90027, USA
3Department of Electrical and Computer Engineering, University of California Davis, Davis, CA 95616, USA
4Division of Cardiology, Children’s Hospital Los Angeles, Los Angeles, CA 90027, USA

Corresponding author: Botian Xu (botianxu@usc.edu)

*Yaqiong Chai and Botian Xu contributed equally to this work.

This work was supported in part by the National Heart Lung and the Blood Institute under Grant 1U01HL117718-01 and Grant 1RO1HL136484-A1, in part by the National Center for Research Resources under Grant UL1 TR001855-02, and in part by the University of Southern California’s Center for High-Performance Computing.

ABSTRACT Magnetic resonance imaging (MRI) images acquired as multislice two-dimensional (2D) images present challenges when reformatted in orthogonal planes due to sparser sampling in the through-plane direction. Restoring the “missing” through-plane slices, or regions of an MRI image damaged by acquisition artifacts can be modeled as an image imputation task. In this work, we consider the damaged image data or missing through-plane slices as image masks and proposed an edge-guided generative adversarial network to restore brain MRI images. Inspired by the procedure of image inpainting, our proposed method decouples image repair into two stages: edge connection and contrast completion, both of which used general adversarial networks (GAN). We trained and tested on a dataset from the Human Connectome Project to test the application of our method for thick slice imputation, while we tested the artifact correction on clinical data and simulated datasets. Our Edge-Guided GAN had superior PSNR, SSIM, conspicuity and signal texture compared to traditional imputation tools, the Context Encoder and the Densely Connected Super Resolution Network with GAN (DCSRN-GAN). The proposed network may improve utilization of clinical 2D scans for 3D atlas generation and big-data comparative studies of brain morphometry.

INDEX TERMS Artifact correction, edge, generative adversarial network, image restoration, imputation, magnetic resonance imaging.

I. INTRODUCTION Magnetic resonance imaging (MRI), as an indispensable tool for medical diagnosis and imaging research, offers detailed visualization of the human torso, extremities and brain. However, artifacts often occur, reducing image quality, diagnostic utility, and scientific relevance [1]. While a plethora of two-dimensional (2D) MRI scans are acquired in hospitals, retrieving missing information due to image artifacts, or due to large slice thickness is of great importance, especially for downstream meta-analyses.

Types and manifestations of artifacts were reviewed by Stadler and Ba-ssalalah [2] and Zhuo and Gullapalli [3], of which some are obvious, and some are subtle, leading to misinterpretation or misdiagnosis. The most common one is motion artifact due to respiration or other movement of the imaging subject [4], [5]. It appears as blurring or coherent ghosting, and in more severe case, it smears the image. The other frequently encountered artifacts are equipment-related ones, such as spike (herringbone) artifacts, appearing as dark stripes overlaid on the image, or zipper artifacts, exhibiting increased noise that extends throughout the image slices [6], [7]. Finally, many of them exhibit as voids in the images. All of these often leads to the discarding of the affected slice, and hence a loss of potentially crucial information, especially in cases of pathologic conditions. Therefore, correcting the artifact affected slices is of great importance for both clinical and research work.

Another important image restoration application lies in retrieving anatomic information coded in the through-plane direction of 2D images, which can be formulated as missing slices. This is particularly imperative for T2-weighted...
FIGURE 1. A stack of MRI 2D images. (a) The resolution in the through-plane direction (coronal and sagittal) is usually much lower than that in the in-plane (axial) direction. (b) Initially the slices to be restored are modeled as 1-valued masks in the through plane direction, while shown as masked rows in the other two planes. Note that the masked region shown in coronal and sagittal view in (b) is one slice out of every three slices, different than the mask size that we implemented in this paper (see Section III.A). It is a schematic illustration of how missing slices are represented in the other two orthogonal planes.

imaging, due to its long signal recovery time and vulnerability to motion artifacts [4], [8], [9]. Specifically, inter-slice spacing is three to six times larger than the in-plane resolution of individual slices. This results in a resolution that is much higher in-plane than in the through-plane direction, as shown in Fig 1.(a).

Both multi-slice 2D acquisitions, with their large inter-slice spacings, and artifact-corrupted images, are poorly suited for downstream segmentation and shape analyses, such as skull stripping, deformable registration, and surface or shape construction [10]–[12]. Therefore, retrieving missing slices to achieve isotropic-resolution, and correcting artifact affected regions are crucial steps in obtaining as much relevant information as possible out of the images.

Prior MRI restoration methods proposed to go from anisotropic 2D MRI images to isotropic ones can be grouped under two main categories: model-based and data-driven. The earliest methods include piecewise interpolation such as nearest neighbor, linear, polynomial and spline interpolations [13]. Mahmoudzadeh and Kashou [14] registered three 2D orthogonal scanning planes to a high-resolution grid, and combined three interpolated volumes to achieve the high-resolution image. On the other hand, using data-driven methods, Greenspan et al. [15] and Yang et al. [16] extended the iterative back projection (IBP) method proposed by Irani and Peleg [17], and modeled sparse representation and over-complete dictionary learning to restore missing slices. Dalca et al. [18] employed an expectation maximization algorithm to train a Gaussian mixture model, imputing the missing structures by learning solely from the available collection of sparsely sampled images. The work also investigated the effects of various slice thicknesses on the performance [18].

Mathematically, image restoration can be modeled as an ill-posed problem to find $f^{-1}(\cdot)$, the inverse of the image degrading mapping, and to minimize the difference between estimated results $\hat{X}$ and the desired but unknown images $X$, in forward model $Y = f(X)$, where $Y$ represents the observed images, and $f$ is the image degrading mapping. Efforts have been made using IBP [17], non-local means, and matrix completion algorithms [19]–[21], from which the reconstruction and correction were iterated in a multi-scaled manner. More recently, deep learning approaches have enjoyed explosive popularity and a powerful capability to improve reconstruction results. Convolution neural networks (CNN) [22] have achieved satisfactory results compared with previously applied methods [23]–[25]. However, the widely used optimization methods of CNNs minimize voxel-wise error between estimated and the ground truth images without regard for the underlying structure. This leads to an overall blurring and lower perceptual image quality [26], suggesting that CNN based methods struggle to retain high frequency information. Evidence from other imaging applications suggest that generative adversarial networks (GAN) better preserve the edges and image texture essential to perceptual quality [27]–[29], but suffer from poorer voxel-wise performance because of their emphasis on learned patterns [24].
In this work, we propose a new method to improve the voxel-wise performance of GANs by cascading two networks focusing on specific tasks. We explicitly concentrate on restoring missing slices due to image artifacts or 2D scanning schemes, to generate anatomically plausible and consistent 3D volumes by imputing the missing slices. Our framework was inspired by image inpainting by artists, with the goal of "reconstituting the missing or damaged portions of the work, in order to make it more legible and to restore its unity" [30]. In image inpainting, an artist composes a drawing by initially delineating the spaces and shapes, using a “lines first, color second” principle [31]. Indeed, in both de-novo creation of an artistic painting or during image restoration, a completed sketch or edge recovery plays a vital role and comes before paint is applied to the canvas [32]. Attempts have been made to develop image inpainting using natural images [33]–[35].

Our proposed edge-guided image restoration network decouples the recovery of high and low-frequency components of the missing information, to generate coherent anatomical details from adjacent slices. We first apply a GAN to recover edge information based upon existing image context. A contrast completion GAN subsequently combines the “sketched” edges from the CNN to fill in appropriate image contrasts.

**II. METHODS**

**A. DATA REPRESENTATION AND PROPOSED FRAMEWORK**

In this work, we model the damaged or missing through-plane slices (axial slices, illustrated in Fig. 1(a)) as binarized masks, where the masked regions are set to value 1. This in turn will be visualized as masked rows in the other two planes, as presented in Fig. 1(b); retrieving missing slices is achieved by estimating the masked rows in the other two orthogonal planes. To mimic image inpainting, the first step is to connect the broken edges in the masked rows. Then a contrast completion network utilizes the connected edges from the first step and estimates the voxel intensities in the missing rows. This approach is achieved by edge-guided GAN (EG-GAN) to enhance perceptually consistent results.

Our proposed method consists of two steps: edge connection and contrast completion; both steps follow an adversarial model, consisting a generator and a discriminator (Fig 2). EG-GAN firstly connects the missing edges of the affected artifacts or low-resolution images by edge generator, taking 2D scans, and masks generated from missing slices in through-plane as input, supervised by the edges generated from the original images. As input and ground truth, edges of 2D images and original, isotropic resolution images, respectively, are extracted by a Canny edge detector [36]. In the second step of our method, a contrast generator fills the intensities based on the original contrast from 2D images, guided by edges generated from the first step and supervised by the original images.

Our framework was inspired by Nazeri et al. [37], which has achieved impressive results in image restoration for natural images. We started building our network from context encoder, an image semantic inpainting net implemented by Pathak et al. [38]. The design of two networks will be presented in detail in the following sections.
B. DESIGN OF LOSS FUNCTIONS FOR EG-GAN

1) LOSS FUNCTION IN EDGE CONNECTION

The loss function of edge connection is extended from [39] and designed as:

\[ \text{loss}_{\text{EG}} = \lambda_{\text{GAN}} \cdot \text{loss}_{\text{GAN1}} + \lambda_f \cdot \text{loss}_f + \lambda_{\text{DSC}} \cdot \text{loss}_{\text{DSC}} \]  

where \( \text{loss}_{\text{GAN1}} \) is the adversarial loss, and \( \text{loss}_f \) is the feature-matching loss [27], used to stabilize the network. To enhance the voxel-wise precision of the generated edge, the Dice similarity coefficient (DSC) loss [40] \( \text{loss}_{\text{DSC}} \) is included. Finally, \( \lambda_{\text{GAN}}, \lambda_f \) and \( \lambda_{\text{DSC}} \) are regularization parameters.

We denote \( I_{\text{GT}} \) as the original image (ground truth), and \( I_{\text{GT}}^M \) the masked image: \( I_{\text{GT}}^M = I_{\text{GT}} \odot M \), where \( M \) is the mask that is designed to mimic the thick slices. By the same token, \( C_{\text{GT}} \) is the image contour and \( C_{\text{GT}}^M \) its masked or degraded edges. The edge mapping performed by generator can be represented as:

\[ C_{\text{pred}} = G_E \left( C_{\text{GT}}^M, M \right), \]  

where \( C_{\text{pred}} \) is the predicted image contour. \( C_{\text{pred}} \) and \( C_{\text{GT}} \) are designed as paired inputs of the discriminator to distinguish if \( C_{\text{pred}} \) is real.

The adversarial loss for the edge connection network can be written as:

\[ \text{loss}_{\text{GAN1}} = \mathbb{E}_{(I_{\text{GT}}, C_{\text{GT}})} \left[ \log \mathbb{D}_E \left( C_{\text{GT}}, I_{\text{GT}} \right) \right] + \mathbb{E}_{I_{\text{GT}}} \left[ \log \left( 1 - \mathbb{D}_E \left( C_{\text{pred}}, I_{\text{GT}} \right) \right) \right], \]

where \( I_{\text{GT}} \) and \( C_{\text{GT}} \) are ground truth images and their contours.

The feature-matching loss, \( \text{loss}_f \), calculated by the difference of the activation maps generated by the hidden layers in the discriminator, is defined as:

\[ \text{loss}_f = \mathbb{E} \left[ \sum_{i=1}^{L} \frac{1}{N_i} \left\| D_E^{(i)} (C_{\text{gt}}) - D_E^{(i)} (C_{\text{pred}}) \right\|_1, \]

where \( L \) is the number of hidden layers in the discriminator, and \( N_i \) is the number of elements in the map of the \( i \)-th layer. \( D_i \) represents the \( i \)-th activation of the discriminator.

The dice similarity coefficient calculates a spatial overlap index that ranges between 0 and 1 [41]. It has been widely used in evaluating two sets of binary segmentation results. In edge generator, DSC loss, \( \text{loss}_{\text{DSC}} \), is designed to restrict the bias towards background in learning [40], [42]:

\[ \text{loss}_{\text{DSC}} = -\frac{2 \sum_i c_i^G \cdot c_i^C}{\sum_i^N c_i^G + \sum_i^N c_i^C}, \]

where \( c_i^G \in C_{\text{GT}} \) and \( c_i^C \in C_{\text{pred}} \). The summation runs over the \( N \) voxels.

2) LOSS FUNCTION IN CONTRAST COMPLETION

The contrast completion network takes incomplete image \( I_{\text{GT}}^M \) as input, enhanced by the connected contour from the previous step. The contrast mapping \( G_c \) can be represented as:

\[ I_{\text{pred}} = G_c \left( I_{\text{GT}}^M, C_{\text{pred}} \right), \]

where \( I_{\text{pred}} \) is the prediction of the restored image. Similar to (3), the adversarial loss of the contrast completion network is defined as:

\[ \text{loss}_{\text{GAN2}} = \mathbb{E}_{(I_{\text{GT}}, C_{\text{pred}})} \left[ \log \mathbb{D}_C \left( I_{\text{GT}}, C_{\text{pred}} \right) \right] + \mathbb{E}_{C_{\text{pred}}} \left[ \log \left( 1 - \mathbb{D}_C \left( C_{\text{pred}}, I_{\text{pred}} \right) \right) \right], \]

where \( D_C \) is the corresponding discriminator. To ensure that the reconstructed image has both high voxel-wise accuracy and good perceptual quality, the network is trained utilizing a combined loss function, including \( I_2 \) loss, perceptual loss [43], \( \text{loss}_p \), and style loss \( \text{loss}_s \) [44], where \( I_2 \) loss is the most common reconstruction loss; perceptual loss is included so that results that are not perceptually similar to the ground truth will be refined; style loss [45] is chosen to ameliorate “checkerboard” artifacts due to transpose convolutional layers [26]. Taken together, the overall loss function of the contrast completion is:

\[ \text{loss}_{\text{GC}} = \lambda_f \cdot \text{loss}_f + \lambda_{\text{GAN2}} \cdot \text{loss}_{\text{GAN2}} + \lambda_p \cdot \text{loss}_p + \lambda_s \cdot \text{loss}_s. \]

C. ARCHITECTURE OF PROPOSED NETWORK

Fig.3 illustrates the structural details of each component. Our generators are adapted from the method from Johnson et al. [43], which was demonstrated to perform well in image-to-image translation and super-resolution [44], [46], [47]. The generators of both networks include encoders, eight residual blocks and decoders. Both discriminators follow a PatchGAN architecture [46], [48].

To further stabilize networks by scaling weight matrices using their respective largest singular values, spectral normalization is applied in both the generator and discriminator [49]. Note that the edge generator has spectral normalization and instance normalization across all the layers [49], [50], whereas the contrast generator only uses instance normalization, as learning high frequency information such as edges requires more restrictions to maintain the stability of the network [50]. However, for low frequency contrast information, spectral normalization is not necessary and might slow the training procedure. Therefore, spectral normalization is removed from the contrast generator.

III. EXPERIMENTS

In this section, we will introduce our experimental datasets, training and testing schemes, two state-of-art methods for similar applications for comparison, and our evaluation methods.

A. DATA DESCRIPTION

To demonstrate the generalization of EG-GAN, we used a large publicly available T1-weighted brain image dataset in the Human Connectome Project (HCP) S1200 collection [51]. We downloaded images after preprocessing
FIGURE 3. Individual architecture of edge connection and contrast completion networks. The hyperparameters of each layer are labelled as K (kernel size), N (number of channels), and S (stride). Both of the generators are built from ResNet with different normalization strategies. While spectral normalization is applied after each layer in the edge generator to enhance the stability, several layers in the contrast generator are designed without spectral normalization to speed up the training procedure. Note that the discriminator for both networks follows the same hyperparameters.

pipelines, including distortion correction and brain extraction. We randomly chose 600 subjects for our experiments. The images come in 0.7 mm isotropic high resolution, and we removed boundary all-zero slices and fit them to 256 × 256 × 256. The whole dataset is split into 50% training, 25% validation, and 25% testing, without overlapping subjects. The predicted results from the testing set only were used in our final performance evaluation and comparison.

The original images were used as ground truth images $I_{GT}$ and their counterpart thick-slice images, $I_{GT}^M$, were artificially generated mimicking damaged slices or 2D MRI scans. To stimulate typical clinical 2D scans to the greatest extent, the slice thickness is four times larger in through plane, i.e., three slices were masked (filled with the value 1, as illustrated in Fig1.b) for every four slices in through plane direction. Therefore, the thick slice images have the same size as the reference images (256 × 256 × 256).

In addition, to investigate the image quality difference of the imputed image using k-space or image-space resampling, we generated low resolution images using similar way as in Chen et al. [52]: we applied FFT to convert the original image into k-space; truncated the outer part of 3D k-space data; filled in the truncated data with zero; and finally, applied inverse FFT to convert to image-space. High frequency truncation and zero padding in k-space effects low-pass filtration with interpolation to restore the original image size. The interpolated slices (representing three out of every four slices) were masked in through-plane direction to mimic 2D MRI scenarios.

To evaluate the intra-subject generalization of our network and slice consistency, we chose the axial plane as the thick slice direction to conduct the following experiments.

B. TRAINING PROCEDURE

The models were implemented in PyTorch on High-Performance Computing clusters with NVIDIA Tesla P100 GPUs. Learning rate was set to $10^{-4}$ for generators and $10^{-4}$ for discriminators. The models were optimized by ADAM optimizer. Our training image size was 256 × 256, with a batch size of eight, and the Canny edge detection threshold was set to 0.5. The standard deviation of the Gaussian filter used in Canny detector was one. We set the maximum training iteration as 20,000 for the edge connection stage, and 40,000 for the image completion stage as no significant improvement afterward.
Our training time is about 230 ms/iteration and sums up to less than two hours to finish training edge connection stage, and less than 4 hours for image completion stage. The prediction time is about 2 minutes for each volume.

C. COMPARISON METHODS
To evaluate the competence of our proposed method, we first assessed the baseline of the performance of the simplest and fastest interpolation methods: nearest neighbor (NN) and cubic. In order to evaluate the improvement of the edge guidance, we then compared a variant of context encoder [38] (the contrast completion part of our proposed network) with the EG-GAN.

Due to very limited work on image imputation, we compared with a similar application, the GAN-based 3D MRI reconstruction in [53], coined DCSRN-GAN, which outperformed previously proposed methods in this task [23], [52]. We used the same data preparation steps as [52] and followed the same hyperparameters to build mDCSRN-GAN, as proposed by the authors [53]. Note that we did not implement our proposed method in 3D as in [52] and [53], due to memory constraints and because patch-based implementations degrade performance of the edge connection stage, which is essential for our method. To evaluate the performance of our method for super-resolution tasks, we inverse-Fourier transformed the images and performed linear down-sampling in Fourier space (in a way of zero-padding), as a linear up-sampling method. Similarly, we used pseudo-k-space down-sampled data as “low-resolution” images for input data while comparing DCSRN-GAN, the context encoder, and EG-GAN.

We also tested our model on two kinds of artifacts affected images, spike artifact and zipper artifact, which are the most relevant extension of our proposed mask model. To simulate the spike artifact, we randomly added spike gradients at different frequencies and angles in the pseudo-k-space of the HCP data. For the experiment of zipper artifact removal, we visually screened a large multi-center pediatric clinical dataset [54] including T1-weighted axial readout (256 × 256 × 22, 0.86mm × 0.86mm × 6mm) and T1-weighted sagittal readout (24 × 256 × 256, 6mm × 0.86mm × 0.86mm) under the supervision of a certified on-site radiologist. Two subjects (age = 9.5 and 13) were identified with images containing zipper artifacts only and were tested using our model trained on the HCP dataset.

D. EVALUATION
To evaluate the similarity between original images and our results, we compared the voxel-wise intensity value accuracy using peak signal-to-noise ratio (PSNR). Furthermore,
we measured structural similarity index (SSIM) [55], which is considered to reflect perceptual image quality. Both metrics are used to evaluate each comparison independently.

IV. EXPERIMENTAL RESULTS

A. IMPUTATION

Fig.4 demonstrates the restored images in sagittal and coronal views from a random subject using the four methods. The top panel shows the results from the basic interpolation methods: NN and cubic. Either from the whole brain view or the regional zoomed-in one, we observe the stairway effects as expected from the results using NN in Fig.4 (a). The cubic interpolation yields more aesthetically pleasing results, but at the cost of some edge blurring (Fig.4b). White matter disconnection can be seen in the second zoomed image due to the interpolated values closer to the neighboring grey matter voxels, when comparing to the reference image in Fig.4 (e). The bottom panel compares the results of the context encoder and our proposed method, EG-GAN in Fig.4 (c) and Fig.4 (d), respectively. Neither context encoder nor EG-GAN exhibit the stairway effects or broken white matter tracts of the interpolation methods, and both exhibit many of the fine details of the brain anatomy. However, the result of EG-GAN appears much more visually plausible, and the method is capable of restoring fine details of small vessels (the second magnified region in Fig.4 (d)) as well as distinguishable and smooth boundaries for the cortex. Note that the result of the context encoder shows faint traces of the masked row (missing slices in the axial direction), which can be observed in all three magnified figures in Fig.4 (c). To demonstrate the inter-slice consistency, the axial slices that were filled in the through plane using EG-GAN are presented in Supplemental Fig. 1.

Quantitative results are summarized in Table. 1. Consistent with Fig. 4, the context encoder, which is widely used for image inpainting, achieved higher PSNR and SSIM than NN or cubic. However, EG-GAN outperformed traditional methods by roughly 10% higher PSNR and 5% higher SSIM. Note that the standard deviation of the PSNR in EG-GAN is only a third of those using other methods.

### TABLE 1. Evaluation results of four different methods for image imputation presented as mean ± standard deviation. PSNR: peak signal-to-noise ratio; SSIM: structural similarity index.

| Method          | PSNR      | SSIM      |
|-----------------|-----------|-----------|
| Nearest neighbor| 36.5402 ± 1.5988 | 0.9177 ± 0.0055 |
| Bi-cubic        | 37.8756 ± 1.7133 | 0.9410 ± 0.0050 |
| Context Encoder | 38.5449 ± 1.3537 | 0.9528 ± 0.0325 |
| EG-GAN          | 40.4849 ± 0.5101 | 0.9745 ± 0.0022 |
FIGURE 6. Clinical axial (a) and sagittal (d) scans with zipper artifact and their corresponding zoomed-in regions, masks overlapped on the artifact corrupted rows (b, e), and artifact corrected images by EG-GAN (c, f) with the magnified areas, respectively.

B. SUPER-RESOLUTION FROM PSEUDO-K-SPACE
Fig. 5 exhibits the reconstructed images in sagittal and coronal view of another random subject using pseudo-k-space down-sampled data. We compared the basic linear interpolation method, a state-of-art super-resolution method, DCSRN-GAN, and two image inpainting-based methods, context encoder and our EG-GAN in Fig.5 (a)-(d), respectively. Both linear interpolation and DCSRN-GAN avoid stairway effects but show overall blurring when comparing three magnified regions to the reference image in Fig.5 (e). We observed areas that DCSRN-GAN failed to restore, as seen on the top of the second magnified image in Fig.5 (b). When comparing the results of context encoder and EG-GAN, the context encoder is not able to restore regions where the folds of gyri and sulci are more complex, nor the blood vessel (the second magnified image in Fig.5 (c), whereas EG-GAN can (bright voxels between sulcus folds in coronal view of Fig.5 (e), and the second magnified image). The reconstructed axial slices are shown in the Supplemental Fig.2.

Table 2 shows the quantitative evaluation results of linear, DCSRN-GAN, context encoder, and EG-GAN methods. Note that DCSRN-GAN and the context encoder achieved comparable performance to EG-GAN in both PSNR and SSIM. However, when comparing the performance of super-resolution and image imputation (Table 1), both PSNR and SSIM are higher in image imputation task.

C. ARTIFACT CORRECTION
Fig.6 (a) and (d) present axial and coronal views of images with mild zipper artifacts from two subjects. The mask that we used in the training model can perfectly cover the artifact.
FIGURE 7. Spike artifacts removal: representative coronal slices of two subjects from the HCP dataset. The artifacts are simulated by adding two different random spike gradients (a, d) in pseudo k-space. The corrected images and their corresponding references are shown in (b, e) and (c, f), respectively.

TABLE 2. Evaluation results of four methods for super-resolution using pseudo-k-space down-sampled data presented as mean ± standard deviation. PSNR: peak signal-to-noise ratio; SSIM: structural similarity index.

| Method       | PSNR       | SSIM       |
|--------------|------------|------------|
| Linear       | 37.8820 ± 2.4217 | 0.9332 ± 0.0052 |
| DCSRN-GAN    | 39.8407 ± 2.3747 | 0.9604 ± 0.0118 |
| Context Encoder | 39.4249 ± 1.4223 | 0.9447 ± 0.0089 |
| EG-GAN       | 40.1075 ± 1.0164 | 0.9559 ± 0.0052 |

corrupted rows (Fig.6 (b) and (e)). Fig.6 (c) and (f) demonstrate that our EG-GAN model can recover the zipper artifact corrupted rows. The magnified images present the detail of the recovered region, proving that the recovered rows removed the artifact while preserving the anatomical contrast both in the ventricular area (the magnified image of Fig.6 (c)) and cortex (the magnified image of Fig.6 (f)).

Our simulated spike artifact on a random subject in the HCP dataset is displayed in Fig.7 (a) and (d). Due to the different frequencies and angles in pseudo-space, they manifest different directions and brightnesses. The artifact corrected images shown in Fig.7 (b) and (e) restored the affected lines in (a) and (d).

V. DISCUSSION AND CONCLUSION
While true 3D images have many advantages, such as higher SNR and the ease of image reformatting and registration, the lengthy acquisitions required for 3D imaging are vulnerable to image motion. Hence, multiplanar 2D acquisitions are the norm in pediatric imaging, particularly for T2-weighted images. In this work, the process of restoring missing slices due to image artifacts or 2D scanning scheme was first modeled as image imputation problem. With this novel data representation, we employed the context encoder [38], which was initially designed to solve image inpainting problem, to solve our specific MRI restoration tasks (imputation, super-resolution, and artifact removal). However, the original context encoder is not capable of retrieving the missing slices due to lack of constraints. Therefore, we sought to edge as a constrain to boost the network performance. We hypothesized that an intact edge could enhance the overall system performance due to its intrinsic high frequency information and its ability to constrain contrast matching. The performance of our proposed edge-guided image restoration network supports this hypothesis by demonstrating higher PSNR
and SSIM than the widely used context encoder, which was initially designed for generating the contents of an arbitrary image region conditioned on its surroundings [27], [38], [56].

One of the differences between our work and Nazeri’s method [37] is that our inpainting model was rooted from context encoder, and we only predicted the masked region. However, the network in [37] predicts the whole image, modifying physically acquired data. This difference is also embedded into the loss functions. The reconstruction loss from [37] is only normalized by the mask size, while the context encoder, as well as our work directly calculates the reconstruction loss within the masked region. Despite excluding the primarily acquired data from image optimization, we did not observe any boundary artifacts.

In image imputation, the restored image using NN showed stairway effects due to duplication from neighboring slices (Fig. 4 (a)). Although cubic interpolation outperformed NN (PSNR, SSIM, and aesthetics), the smoothness of interpolated regions led to both underestimation and overestimation in the target voxels. The context encoder fitted the regional voxel distributions in an image, producing better PSNR and SSIM than either NN or cubic interpolation. However, due to the lack of structural constraints, the masked region could not be fully recovered, and images were left with an abnormal texture and boundary effects for the mask sizes used in our study. By conditioning on edges, EG-GAN improved contrast generalization and alleviated model collapse near the boundaries [57], providing quantitatively and aesthetically superior results.

While EG-GAN outperformed all methods in both the imputation (Fig. 4 and Table 1) and super-resolution (Fig. 5 and Table 2) tasks, EG-GAN exhibited half of the standard deviation during imputation compared with super-resolution.

In imputation, uncorrupted adjacent slices were used to train the network, while during super-resolution adjacent slices suffer from through-plane blurring that may compromise edge reconstruction. In addition, for both the context encoder and DCSRN-GAN methods [38], [53], voxel-wise intensity similarity between the synthesized and real images were enforced (DCSRN-GAN yields the same SSIM as EG-GAN in Table 2), however, the structure of image content, such as the complex anatomical details were not emphasized. This is partially because context-encoder aims at reducing reconstruction loss ($l_2$ loss), and it lacks constraints to prevent model collapsing in prediction. In contrast, mDCSRN combines voxel-wised reconstruction loss and adversarial loss to stabilize the network and further improves the structural similarity. However, adversarial loss matching the generated and real distribution may produce artificial structures [58], which causes the higher standard deviation in SSIM comparing with EG-GAN. Close inspection of Fig. 5 demonstrates superior image sharpness and structural texture in EG-GAN compared with the DCSRNGAN method. However, if we focus on comparing Fig. 5 (d) and (e), our method does not correctly retrieve the shape of the small vessel in the second zoom-in figure (location is highlighted in coronal view from (a)). It is because that the most portion of the vessel falls in the masked region, resulting the failure of connecting the correct counter of the vessel. Note that none of the other comparing methods are able to restore the exact shape of the vessel due to their intrinsic properties, yet this defect in restoring small anatomical structures could be alleviated by adjusting mask size or randomly shuffling the starting masked slice.

Other than task type, mask size also plays a critical role in network performance. Both SSIM and PSNR decrease up to 15% when increasing the mask size from 20% to 50% [59]. Our EG-GAN method collapsed when the mask size increased from 75% to 80%, primarily because edge continuity could not be unambiguously restored. We postulate that exploitation of information provided by two orthogonal 2D scans (which is commonly used in clinical practice) could potentially improve the performance of EG-GAN. There are ongoing efforts to combine orthogonal 2D images using Gaussian mixture models, or including various priors [60], [61]. Potentially those methods could be imbedded in EG-GAN when combining two orthogonal 2D scans in future work. For example, slice to volume image registration usually suffers from the lack of geometry information and may require image fusion process when handling multi-slices [62]. Additionally, the performance of directly registering 2D images to 3D volume might be affected by viewpoint angle. Although registration can be performed, the different 2D feature points might cause inaccurate results [63]. Therefore, building a superior isotropic 3D volume could benefit the downstream registration and multi-site image group analysis.

Image artifacts such as zipper and gradient spike artifacts are disruptive to clinical workflow because they are sporadic and do not always resolve when a sequence is repeated. In this paper, we demonstrate that EG-GAN offers a simple method for image restoration from these artifacts that could be easily implemented on clinical workstations. As zipper artifacts usually only affect a few lines/columns of the image, our method can correct the artifact with high fidelity. Spike artifacts, whose mask pattern can be irregular, are more challenging because they affect the entire image (Fig. 6) and the mask percentage is very close to the upper limits of restoration (75%). While the exhibited spike artifact in Fig. 7 could be fully restored, satisfactory results cannot be guaranteed when the spike artifact affects a larger fraction of the image. The same situation could apply to zipper artifacts, if the “zipper” affects more than 75% of the slices, which is the maximum masked area that our model could restore. Note that this mask size can be counted locally than globally, meaning that even if only five or six rows/columns of the entire image are corrupted, the masked ratio would be 5/6 or 6/7 locally, and hence prohibit full image recovery. This inference roots to the model that we presented: at least a quarter of the adjacent slices are needed to recover the missing edge, providing a completed and effective constraint, and further to complete the image with the enhanced consistency of the contrast-matching network.
In addition, for artifact correction, mask generation is required, which may be challenging for tasks such as motion artifact and cardiac tagging, whose masks could be highly irregular. Therefore, an automatic and accurate method might be required to extract the mask of the areas to be restored. Another limitation of artifact correction using our method is that lesions lying exclusively in the masked area could be painted over and not recovered.

As discussed in several places, edge completion plays a critical role in improving the image restoration results. This does not advance our network to utilize an end-to-end learning fashion. We had to visually ensure that edge connection stage was well-trained before starting contrast completion. In practice, using task array on GPUs could make our proposed network an end-to-end training, but an intermediate check is recommended. Secondly, in our work, we used a fairly old and primitive technique for constructing edge maps (Canny edge detection) [36]. However, learning-based methods, such as holistically nested edge detection could potentially be used in combination with the Canny detector in the future work [64], [65] and promote a more efficient learning. In addition, more robust edge-detection and completion could potentially overcome the limitations encountered with 3D patch implementation.

In summary, our proposed edge-guided image restoration network decouples the recovery of high and low-frequency components of the missing information to generate coherent anatomical details from adjacent slices. The network proved to effectively restore the missing image detail either due to 2D scanning schemes, or due to image artifacts. We propose that EG-GAN could improve utilization of clinical 2D scans for 3D atlas generation and big-data comparative studies of brain morphometry.

REFERENCES

[1] K. Somasundaram and P. Kalavathi, “Analysis of imaging artifacts in MR brain images,” *Oriental J. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 135–141, 2012.

[2] A. Stadler, W. Schima, A. Ba-Ssalamah, J. Kettenbach, and E. Eisenhuber, “Artifacts in body MR imaging: Their appearance and how to eliminate them,” *Eur. Radiol.*, vol. 17, no. 5, pp. 1242–1255, Apr. 2007.

[3] J. Zhuo and R. P. Gullapalli, “MR artifacts, safety, and quality control,” *RadioGraphics*, vol. 26, no. 1, pp. 275–297, Jan. 2006.

[4] E. B. Welch, J. P. Felmlee, R. L. Ehman, and A. Manduca, “Motion correction using thek-space phase difference of orthogonal acquisitions,” *Magn. Reson. Med.*, vol. 48, no. 1, pp. 147–156, Jul. 2002.

[5] J. Maclellan, M. Herbst, O. Speck, and M. Zaitsev, “Prospective motion correction in brain imaging: A review,” *Magn. Reson. Med.*, vol. 69, no. 3, pp. 621–636, Mar. 2013.

[6] A. D. Elster, “Artifacts in 3-T MRI: Physical background and reduction strategies,” *Eur. J. Radiol.*, vol. 65, no. 1, pp. 29–35, Jan. 2008.

[7] S. Heiland, “From a as in aliasing to z as in zipper: Artifacts in MRI,” *Clin. Neuroradiol.*, vol. 18, no. 1, pp. 25–36, Mar. 2008.

[8] E. Lavdas, I. Tsougos, S. Kogia, G. Gratsias, P. Svolos, V. Roka, I. V. Fezoulidis, and E. Kapsali, “T2 FLAIR artifacts at 3-T brain magnetic resonance imaging,” *Clin. Imag.*, vol. 38, no. 2, pp. 85–90, Mar. 2014.

[9] M. Zaitsev, J. Maclellan, and M. Herbst, “Motion artifacts in MRI: A complex problem with many partial solutions,” *J. Magn. Reson. Imag.*, vol. 42, no. 4, pp. 887–901, Oct. 2015.

[10] I. Despotovic, B. Goosens, and W. Philips, “MRI segmentation of the human brain: Challenges, methods, and applications,” *Comput. Math. Methods Med.*, vol. 2015, pp. 1–23, Mar. 2015.

[11] P. J. Kostelec and S. Periaswamy, “Image Registration for MRI,” in *Modern Signal Processing (MSRI Book Series)*, vol. 46. Oakland, CA, USA: Mathematical Sciences Research Institute, 2003, pp. 161–184.

[12] F. P. Oliveira and J. M. R. Tavares, “Medical image registration: A review,” *Comput. Methods Biomech. Biom. Eng.*, vol. 17, no. 2, pp. 73–93, Jan. 2014.

[13] T. M. Lehmann, C. Conner, and K. Spitzer, “Survey: Interpolation methods in medical image processing,” *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1049–1075, Nov. 1999.

[14] A. P. Mahbod and N. H. Khosh, “Interpolation-based super-resolution reconstruction: Effects of slice thickness,” *J. Med. Imag.*, vol. 1, no. 3, Dec. 2014, Art. no. 034007.

[15] H. Greenspan, G. Oz, N. Kiyati, and S. Peled, “MRI inter-slice reconstruction using super-resolution,” *Magn. Reson. Imag.*, vol. 20, no. 5, pp. 437–446, Jun. 2002.

[16] 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.

[17] M. Irani and S. Peleg, “Motion analysis for image enhancement: Resolution, occlusion, and transparency,” *J. Vis. Commun. Image Represent.*, vol. 4, no. 4, pp. 324–335, Dec. 1993.

[18] A. V. Dalca, K. L. Bouman, W. T. Freeman, N. S. Rost, M. R. Sabuncu, and P. Golland, “Medical image imputation from image collections,” *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 504–514, Feb. 2019.

[19] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 60–65.

[20] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, “An optimized blockwise nonlocal means denoising filter for sparse + low-rank decomposition of annihilating filter-based Hankel matrix,” *Magn. Reson. Med.*, vol. 78, no. 1, pp. 327–340, Jul. 2017.

[21] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[22] X. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 391–407.

[23] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[24] C.-H. Pham, C. Tor-Díez, H. Meunier, N. Bednarek, R. Fablet, N. Passat, and F. Rousseau, “Multiscale brain MRI super-resolution using deep 3D convolutional networks,” *Comput. Med. Imag. Graph.*, vol. 77, Oct. 2019, Art. no. 101821.

[25] A. Odena, V. Dumoulin, and C. Olah, “Deconvolution and checkerboard artifacts,” *Distill*, vol. 1, no. 10, p. e3, Oct. 2016.

[26] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, “High-resolution image synthesis and semantic manipulation with conditional GANs,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7898–8807.

[27] D. Ravi, A. B. Sitzotka, S. P. Pereira, and T. Vercauteren, “Adversarial training with cycle consistency for unsupervised super-resolution in endomicroscopy,” *Med. Image Anal.*, vol. 53, pp. 123–131, Apr. 2019.

[28] I. Sanchez and V. Vilaplana, “Brain MRI super-resolution using 3D generative adversarial networks,” 2018, arXiv:1812.11440 [Online]. Available: http://arxiv.org/abs/1812.11440

[29] G. Émile-Mâle, in *The Restorer’s Handbook of Easel Painting (PSG AIC)*, 2nd ed. New York, NY, USA: Van Nostrand, 1976.

[30] G. Émile-Mâle, in *Practical Notebooks of Easel Painting (PSG AIC)*, vol. 9, Int. J. Comput. Appl., 2016, pp. 2861–2873, Nov. 2010.
A. Bugeau, M. Bertalmio, V. Caselles, and G. Sapiro, “A comprehensive framework for image inpainting,” IEEE Trans. Image Process., vol. 19, no. 10, pp. 2634–2645, Oct. 2010.

M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, “Simultaneous structure and texture image inpainting,” IEEE Trans. Image Process., vol. 12, no. 8, pp. 882–889, Aug. 2003.

J. Canny, “A computational approach to edge detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, “EdgeConnect: Structure guided image inpainting using edge prediction,” in Proc. IEEE Int. Conf. Comput. Vis. Workshop, Nov. 2019.

D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, “Context encoders: Feature learning by inpainting,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2536–2544.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672–2680.

F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in Proc. 4th Int. Conf. 3D Vis. (3DV), Oct. 2016, pp. 565–571.

A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool,” BMC Med. Imag., vol. 15, no. 1, Dec. 2015, Art. no. 29.

C. Shen, H. R. Roth, H. Oda, M. Oda, Y. Hayashi, K. Misawa, and K. Mori, “On the influence of Dice loss function in multi-class organ segmentation of abdominal CT using 3D fully convolutional networks,” (2018), arXiv:1801.05912. [Online]. Available: https://arxiv.org/abs/1801.05912

J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in Proc. Eur. Conf. Comput. Vis. (ECCV), Cham, Switzerland: Springer, Oct. 2016, pp. 694–711.

R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 586–595.

L. A. Gatys, A. S. Ecker, and M. Bethge, “Image style transfer using convolutional neural networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2414–2423.

J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image translation using cycle-consistent adversarial networks,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2242–2251.

M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch, “EnhanceNet: Single image super-resolution through automated texture synthesis,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 4501–4510.

P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image translation with conditional adversarial networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1125–1134.

T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” 2018, arXiv:1802.05957. [Online]. Available: http://arxiv.org/abs/1802.05957.

D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 6924–6932.

D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil, “The WU-minn human connectome project: An overview,” Neuroimage, vol. 80, pp. 62–79, Oct. 2013.

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. (ISBI ), Apr. 2018, pp. 739–742.

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. Intervent. (MICCAI), Cham, Switzerland: Springer, Sep. 2018, pp. 91–99.

J. F. Casella, A. A. King, B. Barton, D. A. White, M. J. Noetzel, R. N. khord, C. Terrill, D. Hirtz, R. C. McKinstry, J. J. Strouse, T. H. Howard, T. D. Coates, C. P. Minniti, A. D. Campbell, B. A. Vendt, H. Lehmann, and M. R. DeBaun, “Design of the silent cerebral infarct transfusion (SIT) trial,” Pediatric Hematol. Oncol., vol. 27, no. 2, pp. 69–89, Mar. 2010.
KANGNING ZHANG received the B.S. degree in electrical engineering from Beijing Jiaotong University, Beijing, China, in 2015, and the M.S. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 2018. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of California Davis. His research interests include image signal processing and compressive sensing.

NATASHA LEPORE received the B.Sc. degree in physics and mathematics from the University of Montreal, the master’s degree in applied mathematics and general relativity from the University of Cambridge, and the Ph.D. degree in theoretical physics from Harvard University.

She was a Postdoctoral Fellow in medical imaging (with Prof. Paul Thompson) with the Laboratory of Neuroimaging, UCLA. Since 2009, she has been a Faculty Member in radiology and biomedical engineering with the Children’s Hospital Los Angeles and the University of Southern California. She is currently the Director of the Computational Imaging of Brain Organization Research Group (CIBORG), specializes in mathematical and numerical methods to study brain anatomy and function through magnetic resonance imaging. These methods are applied to furthering the understanding of different neurological disorders and normal and abnormal brain development.

JOHN C. WOOD received the B.S. degree in electrical engineering from the University of California, Davis, CA, USA, in 1984, and the M.D./Ph.D. degree in bioengineering from the University of Michigan, Ann Arbor, MI, USA, in 1994, with a focus in time-frequency transform analysis.

He performed the Residency and Fellowship in pediatric cardiology, Yale. He joined the Children’s Hospital Los Angeles/USC Keck School of Medicine, in 1999, studying wavelet-packet denoising applications with MRI. He was studying in cardiovascular consequences of hemoglobinopathies for almost a decade. He was a Principle investigator for the NIH-Sponsored Early Detection of Iron Cardiomyopathy Trial, whose goal is to identify earlier markers of cardiac dysfunction. He is currently the Director of the Cardiovascular MRI and specializes with the MRI assessment of congenital heart disease and noninvasive assessment of iron burden with MRI. He is one of the pioneers of the MRI-based cardiac and liver iron measurements and also studying oral chelation strategies in animals. He also has funded projects examining pancreatic and pituitary iron burden with MRI and their functional correlates. He received the ARRA Challenge Grant to study the role of iron overload and other factors in sickle cell vasculopathy. He is exploring the links between abnormal red cell mechanics and vascular dynamics in the hemoglobinopathies. His research interests include relationship between cerebrovascular reserve, anemia, and white matter loss in chronic anemia patients at particularly high risk for silent stroke. He also explores white matter quantification and super-resolution image using machine learning methods.

Dr. Wood memberships, include the American Medical Association, the American Academy of Pediatrics, the Society for Cardiovascular Magnetic Resonance, and the International Society for Magnetic Resonance in Medicine. His awards and honors, include the Tau Beta Pi (National Engineering Honor Society), in 1982, the Alpha Omega Alpha, in 1993, the Alfred F. Towsley Award for Pediatrics, in 1994, the RSNA Scholar, in 2001, and the Russell Smith Award for Innovation in Pediatric Research, in 2009.