Magnetic Resonance Imaging 64 (2019) 90–100

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

journal homepage: www.elsevier.com/locate/mri

Original contribution

Dual-domain convolutional neural networks for improving structural information in 3 T MRI

Yongqin Zhang[a,b], Pew-Thian Yap[b], Liangqiong Qu[b], Jie-Zhi Cheng[c], Dinggang Shen[b,d,*]

[a] School of Information Science and Technology, Northwest University, Xi'an 710127, China
[b] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA
[c] Shanghai United Imaging Intelligence Co., Ltd., Shanghai 201807, China
[d] Department of Brain and Cognitive Engineering, Korea University, Seoul 136713, South Korea

ARTICLE INFO

Keywords:
Image synthesis
Image super-resolution
Magnetic resonance imaging
Deep learning
Convolutional neural network

ABSTRACT

We propose a novel dual-domain convolutional neural network framework to improve structural information of routine 3 T images. We introduce a parameter-efficient butterfly network that involves two complementary domains: a spatial domain and a frequency domain. The butterfly network allows the interaction of these two domains in learning the complex mapping from 3 T to 7 T images. We verified the efficacy of the dual-domain strategy and butterfly network using 3 T and 7 T image pairs. Experimental results demonstrate that the proposed framework generates synthetic 7 T-like images and achieves performance superior to state-of-the-art methods.

1. Introduction

Magnetic resonance imaging (MRI) scanners use strong magnetic fields to generate soft-tissue images and are widely used for medical diagnosis and disease monitoring. It is often desirable to acquire images with fine voxel resolution for greater anatomical details, but this often results in lower signal-to-noise ratio (SNR) and longer scan times. To improve the resolution of magnetic resonance (MR) images while retaining sufficient SNR [12], efforts have been dedicated in the past three decades to enhance the magnetic field strength of MRI scanners.

The world's first commercial MRI scanner was manufactured using a permanent magnet with low field strength in 1980 [20]. In 1982, the first MRI scanner operating at 1.5T with superconducting magnet was introduced. By 1985, 1.5T MRI scanners have become standard in clinical practices and remained so even to 2010. 3 T MRI emerged in the early 2000s has been becoming more common in clinical settings. In 2017, the first 7 T MRI scanner was approved for clinical use by the United States Food and Drug Administration (FDA)[1].

Fig. 1 shows a visual comparison of 3 T and 7 T MR images obtained from the same subject. Compared with 3 T MRI, 7 T MRI typically affords greater anatomical details, which can benefit disease diagnosis [28]. However, 7 T MRI scanners are significantly more expensive and are hence less common in hospitals and clinical institutions. To date, there are < 100 7 T MRI scanners worldwide [6]. In contrast, there are > 20,000 3 T MRI scanners conducting > 60 million examinations annually.

In this paper, we propose a method for synthesizing high-resolution (HR) 7 T-like images from low-resolution (LR) 3 T images. Note that this is not a simple super-resolution problem because the appearance and contrast of 7 T images can be different from 3 T images. We demonstrate that the reconstruction of 7 T-like images from 3 T MRI can be improved by concurrently considering the spatial and frequency domains in a convolutional neural network (CNN) framework.

2. Related work

To acquire high-quality synthetic images, quite a few machine learning methods have been proposed in recent years. Yang et al. [30] proposed a single image super-resolution method using coupled dictionaries trained from paired LR and HR image patches. Zhang et al. [34] presented a hierarchical patch-based sparse representation method for computed tomography (CT) resolution enhancement. Bhavsar et al. [4] introduced a group-sparse representation method for resolution enhancement of CT lung images. Rueda et al. [23] proposed a sparsity-based super-resolution method using coupled LR and HR dictionaries for brain MR images. Roy et al. [22] presented an example-based super-resolution framework to synthesize high-quality MR images from multi-contrast atlases. By joint modeling of complementary image priors on

* Corresponding author at: Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA.
E-mail addresses: zhangyongqin@nwu.edu.cn (Y. Zhang), ptyap@med.unc.edu (P.-T. Yap), liangqiong@med.unc.edu (L. Qu), dgshen@med.unc.edu (D. Shen).
[1] https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm580154.htm.

https://doi.org/10.1016/j.mri.2019.05.023
Received 14 December 2018; Received in revised form 19 May 2019; Accepted 19 May 2019
0730-725X/

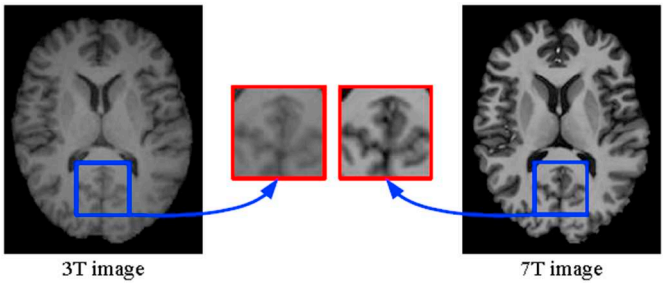

Fig. 1. Comparison of 3 T and 7 T images acquired from the same subject.

the gradients, self-similarity and sparsity, Zhang et al. [32] proposed a structure-modulated sparse representation method for image super-resolution. Bahrami et al. [3] presented a hierarchical sparse representation method based on multi-level canonical correlation analysis (M-CCA) for the reconstruction of 7 T-like MR images from 3 T MR images. To enhance the resolution of neonatal images, Zhang et al. [33] proposed a neonatal image super-resolution method with the guidance of longitudinal data of an identical subject. Subsequently, they [35] proposed a residual structured sparse representation method for neonatal image super-resolution using longitudinal data across subjects. However, restricted by hand-crafted features, these model-based methods cannot be automatically adapted to handling complex image synthesis problems. Besides, they also require tricky parameter tuning and time-consuming calculations for testing.

Recently, deep learning achieves impressive results and attracts great attention in the community of artificial intelligence. Dong et al. [5] introduced a deep learning approach for resolution enhancement. Mao et al. [18] presented a deep convolutional encoder-decoder network with symmetric skip connections for the improvement of training convergence and local optimal solutions. Bahrami et al. [2] developed a CNN-based approach that takes into account the appearance and anatomical features (CAAF) to predict 7 T images from 3 T images. By integrating adversarial loss and content loss as perceptual loss, Ledig et al. [16] proposed a generative adversarial network to generate super-resolved photo-realistic images. Tai et al. [26] proposed a deep recursive residual network via residual learning in a global and local sense. McDonagh et al. [19] proposed a context-sensitive upsampling method using a residual CNN model to produce sharp HR images. By generating more data for model training, Krizhevsky et al. [13,14] proved that data augmentation can benefit the generalization ability of deep learning models for image classification. By introducing data augmentation, Li et al. [17] proposed a fully convolutional network based on deformable image registration for hippocampus segmentation. Touloupou et al. [27] employed data augmentation to estimate the marginal likelihood for the selection between competing statistical models. However, these data-driven methods are still unsatisfactory due to their limited modeling capabilities, even without consideration of specific attributes of MRI.

In this paper, we propose a dual-domain convolutional neural network (DDCNN) framework to solve the above-mentioned problems. DDCNN synthesizes 7 T-like images from 3 T MRI with two parallel and interactive multi-layer streams based on spatial and frequency domains. Our contribution is threefold: 1) We propose a novel dual-domain convolutional neural network framework to learn complex mappings from 3 T to 7 T modalities; 2) We introduce both two complementary domains and parameter-efficient butterfly modules to assist deep learning; and 3) We implement numerous experiments to verify the efficacy of the proposed framework for improving structural information in 3 T MRI. Comparisons with existing methods indicate that synthesized 7 T-like images with higher quality can be obtained with DDCNN.

In our previously published conference paper [31], we proposed a dual-domain cascaded regression (DDCR) method for synthesizing 7 T-like images from routine 3 T images. As linear regression has limited ability to model complex mappings between 3 T and 7 T modalities, in this paper, we generalize DDCR to DDCNN using convolutional neural networks due to their high non-linear modeling ability. This paper serves three purposes. The first purpose is to further verify the efficacy of the dual-domain strategy and butterfly modules for deep learning. The second purpose is to provide additional method details, examples, results, discussion, and insights that are not presented in our conference publication [31]. The third purpose is to introduce a modification for the fusion of feature maps and domain transform, which uses fast Fourier transform (FFT), instead of discrete cosine transform (DCT), because FFT has wide applications. In experiments, we found the performance of DDCNN in this paper was often better than that of DDCR in our conference paper [31].

The remainder of this paper is organized as follows. Section 3 elaborates the proposed method. Section 4 presents experimental results and analysis. Section 5 provides discussions. Section 6 concludes the paper.

3. Method

In this section, given an input 3 T image, we present a DDCNN model to directly learn a mapping function for improving its structural information and reconstructing its corresponding 7 T-like image. We will elaborate the network architecture, image preprocessing and implementation details of the proposed model below.

3.1. Network architecture

Fig. 2 illustrates the architecture of DDCNN. DDCNN uses five types of operations: forward fast Fourier transform (FFT), inverse fast Fourier transform (IFFT), convolution (Conv), concatenation (Concat), and

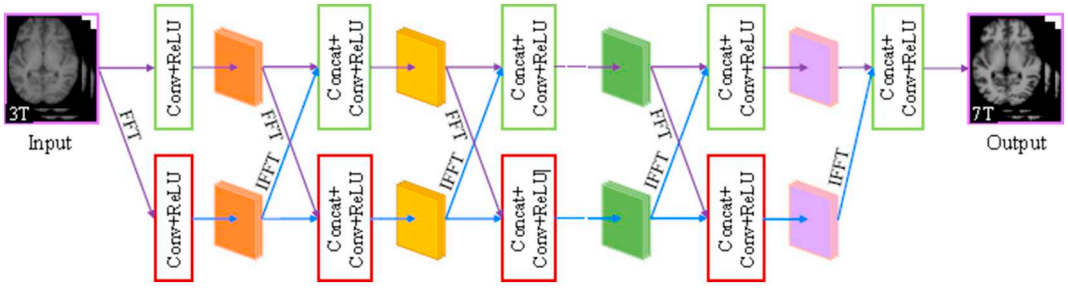

Fig. 2. Network architecture of DDCNN.

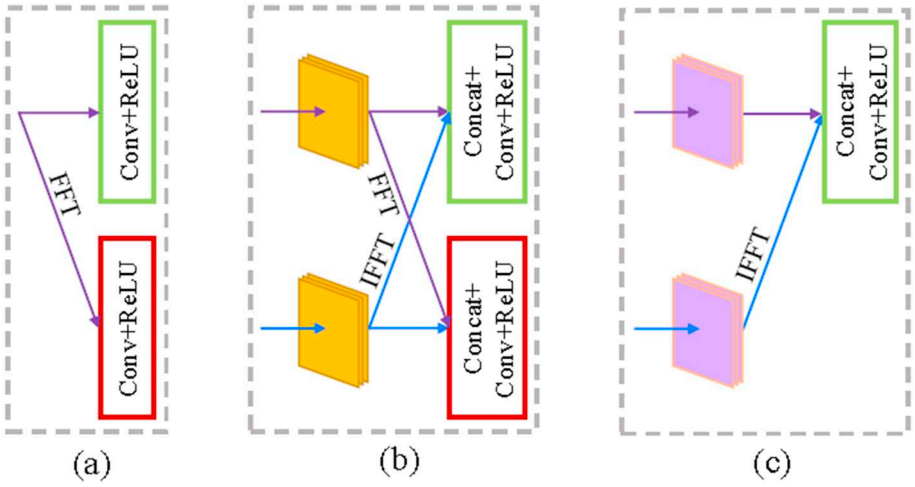

Fig. 3. The respective architectures of three layers. From left to right: (a) the input feature extraction layer, (b) one butterfly module in the middle feature mapping layer, and (c) the output feature reconstruction layer.

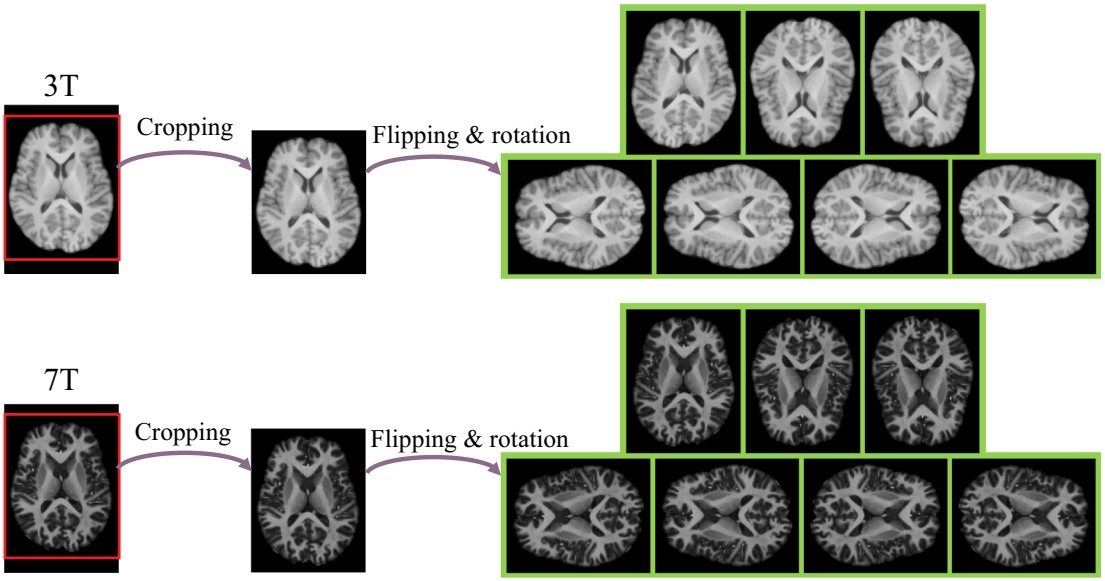

Fig. 4. Visual comparison of the augmented training samples from one subject.

rectified linear unit (ReLU) as the activation function. FFT and IFFT are used for signal conversion between spatial and frequency domains. Compared with the spatial domain, the frequency domain describes distribution features of signals from another aspect. By exploiting complementary information between spatial and frequency domains, we introduce the dual-domain strategy and butterfly modules to achieve high-quality synthesized images.

As shown in Fig. 2, the proposed model consists of two parallel and interactive multi-layer network streams. These two streams complement each other on respective spatial and frequency domains, which help model the complex mappings between 3 T and 7 T images. We employ complex-to-complex Fourier transforms for the interconversion between spatial and frequency domains. As the MR images are usually magnitude images being widely used for diagnosis in clinical MRI, we take an input 3 T image block as the real part and its corresponding zero matrix as the imaginary part of complex-valued network input. On the other hand, we extract the magnitude component of complex-valued network output in the spatial domain as the synthesized 7 T-like image block.

With the 3 T image block of size $H \times W \times c$ as the input, DDCNN with depth L consists of three types of layers: input feature extraction layer, middle feature mapping layer, and output feature reconstruction layer. Fig. 3 shows the respective architectures of these layers. In each layer, both real and imaginary parts of complex-valued features separately go through the same operations, i.e., Conv + ReLU or Concat + Conv + ReLU. To verify the efficacy of the dual-domain strategy and butterfly modules, all layers are assumed to have the same number $2n_f$ of convolution filters on respective spatial and frequency domains. In the following, we will detail these three layers. For simplicity, we only describe Concat, Conv and ReLU operations on real parts of the complex-valued feature maps.

(a) **The input feature extraction layer**: this input layer consists of FFT and two Conv + ReLU operations on respective spatial and

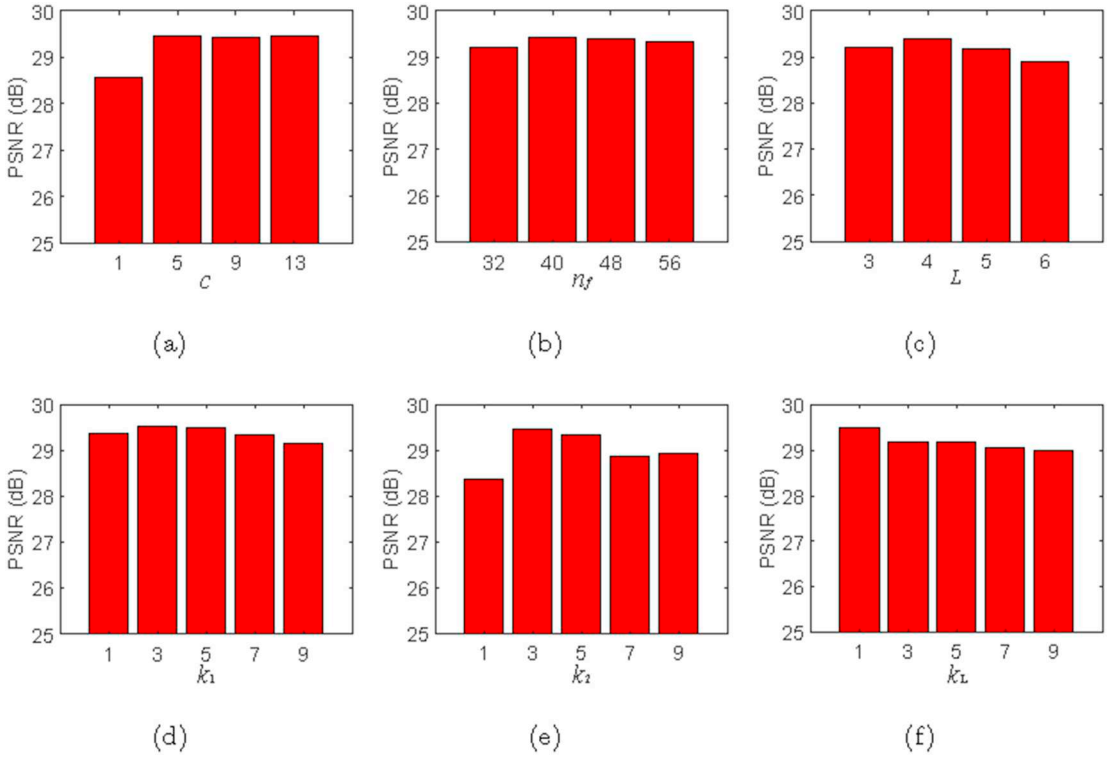

Fig. 5. PSNR values given by DDCNN with respect to its key parameters.

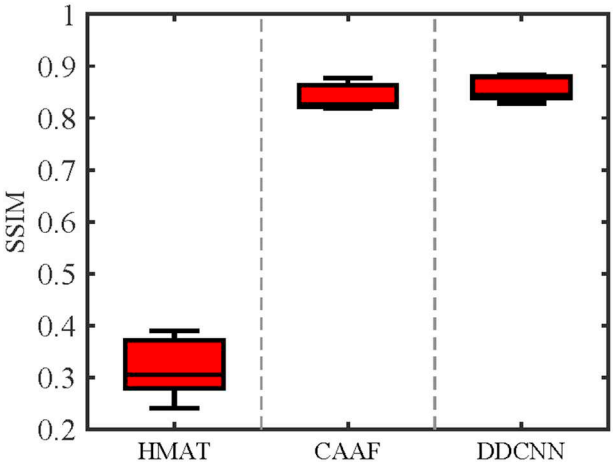

Fig. 6. Box plots for SSIM values. The middle line of each box is the median, the edges mark the 25th and 75th percentiles, and the whiskers extend to the minimum and the maximum. For all methods, the respective medians of SSIM values are as follows: (a) HMAT (SSIM = 0.3050), (b) CAAF (SSIM = 0.8258), and (c) DDCNN (SSIM = **0.8438**).

Table 1
p-values of two-sample *t*-tests of SSIM results between DDCNN and the competing methods. The asterisks indicate *p*-values less than the threshold $\alpha_t = 0.05$ chosen for statistical significance.

Index	HMAT	CAAF	SDCNN	PDCNN
SSIM	< 0.0001*	0.0856	0.0097*	0.0176*

frequency domains, which is used to extract the feature maps from the original 3 T images in two different complementary domains. In this layer, FFT is used to convert the input 3 T image block from the spatial domain to the frequency domain. n_f convolution filters of size $k_1 \times k_1 \times c$ are applied to produce n_f feature maps and then ReLU is used for nonlinear mapping and training acceleration in the spatial domain. Similarly, another n_f convolution filters of size $k_1 \times k_1 \times c$ and ReLU are applied successively in the frequency domain.

(b) **The middle feature mapping layer**: as shown in Figs. 2 and 3(b), this middle layer consists of $L - 2$ butterfly modules. Each butterfly module is composed of FFT, IFFT and two Concat + Conv + ReLU operations on respective spatial and frequency domains. Specifically, FFT is used to convert the output feature maps of the previous module (or layer) from the spatial domain to the frequency domain. On the other hand, IFFT is used to convert the output feature maps of the previous module (or layer) from the frequency domain to the spatial domain. For butterfly module $L - 2l \in \{2, \cdots, L-1\}$, in the spatial domain, after concatenating two groups of output feature maps from the previous module (or layer), n_f convolution filters of size $k_l \times k_l \times 2n_f$ and ReLU are applied sequentially to generate n_f feature maps. Similarly, Concat, Conv and ReLU are applied sequentially to produce new feature maps in the frequency domain.

3.2. Image preprocessing

A set of 3 T and 7 T image pairs are registered to the Montreal Neurological Institute (MNI) standard space [8] using FLIRT [9,10] to remove pose differences. Specifically, all 7 T images are linearly registered to the MNI standard space with an individual template [8]. Each

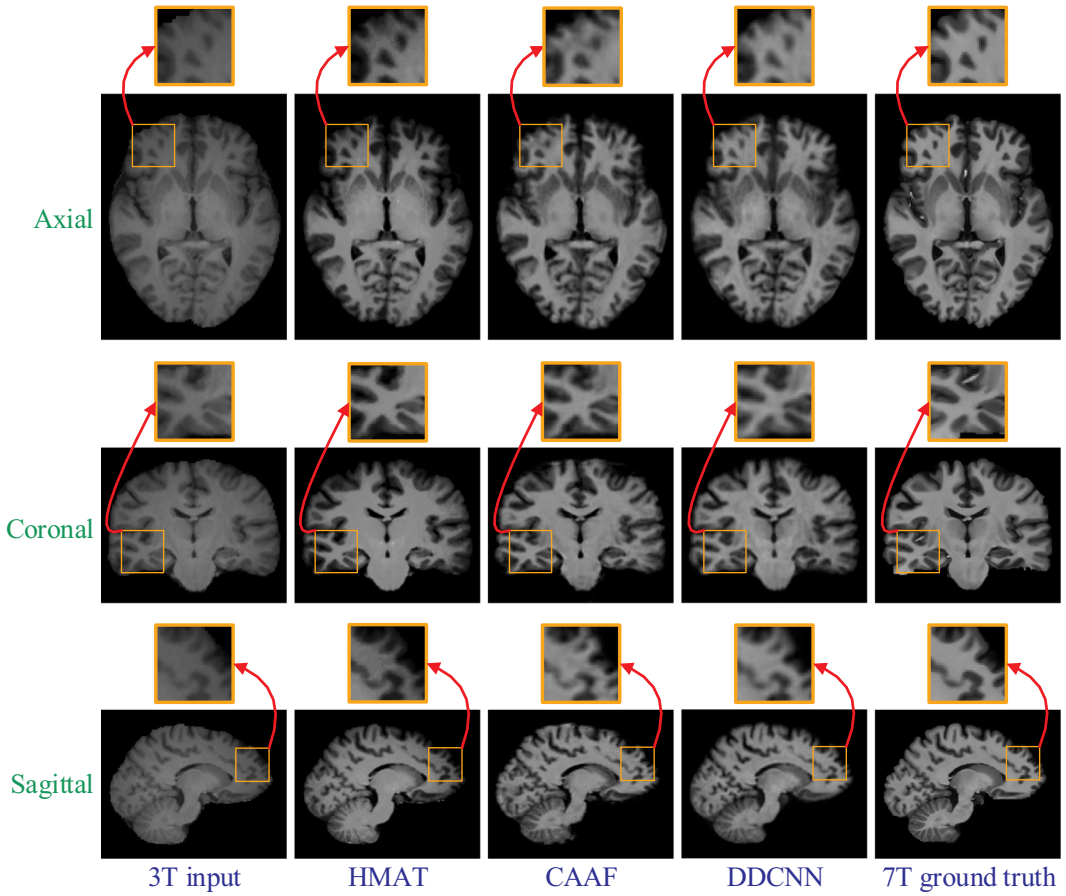

Fig. 7. Visual comparison of axial, sagittal and coronal views of synthesized 7 T-like images with close-up views of specific regions for one subject. For all methods, SSIM values are as follows: (a) 3 T input, (b) HMAT (SSIM = 0.3025), (c) CAAF (SSIM = 0.8590), (d) DDCNN (SSIM = **0.8792**), and (e) 7 T ground truth.

3 T image is then rigidly aligned to its corresponding 7 T image. After registration, bias correction [25] and skull stripping [24] are performed.

To reduce adverse impact of signal variation across different scanners and sites, we first normalize all MR images and then use histogram matching to make them similar in the intensity range. Specifically, first, an input 3 T image and all pairs of aligned 3 T and 7 T training images are normalized to the range $[0,1]$ using $z = (z - z_{\min})/(z_{\max} - z_{\min})$. Here z is the voxel value, $z_{\min}$ is the minimum value, and $z_{\max}$ is the maximum value of an image. Then all normalized 3 T training images are matched to the normalized input 3 T image by three-dimensional (3D) histogram matching, ensuring that all matched 3 T training images have similar contrast ranges. Correspondingly, by 3D histogram matching, all normalized 7 T training images are matched to the corresponding 7 T training image, whose 3 T training image is closest to the input 3 T image. Thus, these matched 7 T training images also have similar intensity and contrast ranges.

As the training data is small in this study, we use several data augmentation techniques [7,13,14], such as cropping, flipping and rotation, enlarge the dataset to avoid over-fitting. We first crop the brains from the original training images. Then we use horizontal flipping, vertical flipping, 90-degree left rotation and 90-degree right rotation for data augmentation. Fig. 4 shows the augmented samples from one randomly selected subject. Data augmentation increases variability in the training dataset to avoid overfitting.

3.3. Implementation details

For an input 3 T image $\mathbf{X}$, DDCNN aims to learn a mapping function $\mathcal{F}(\mathbf{X};\Theta)$ for the prediction of the latent real 7 T image $\mathbf{Y}$. To learn the model parameters Θ, the loss function is enforced by minimizing the averaged mean squared error (MSE) between the synthesized 7 T-like blocks and the ground-truth 7 T blocks as follows:

$$\ell(\Theta) = \frac{1}{2N}\sum_{i=1}^{N} \|\mathbf{y}_i - \mathcal{F}(\mathbf{x}_i;\Theta)\|_F^2, \tag{1}$$

where $\{(\mathbf{x}_i,\mathbf{y}_i)\}_{i=1}{}^{N}$ denotes N pairs of 3 T and 7 T training blocks, and $\|\cdot\|_F$ represents the Frobenius norm.

In our model, each 3 T or 7 T training block consists of several consecutive two-dimensional (2D) slices, which are extracted from corresponding 3D 3 T or 7 T training images. Once the model is properly trained, it is applied to an input 3 T test image block by block and then the results are assembled to construct the whole synthesized 7 T-like image. Due to the limitations of memory capacity and small data sets, it is not feasible to directly train a full 3D model in this study. As several consecutive 2D slices contain rich contextual information, the full 3D modeling is not necessary for this task.

We adopt an adaptive moment estimation (Adam) solver [11] with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $eps = 1e-08$ for optimization. The learning rate starts at $lr = 0.0001$ and decays by a factor of 0.1

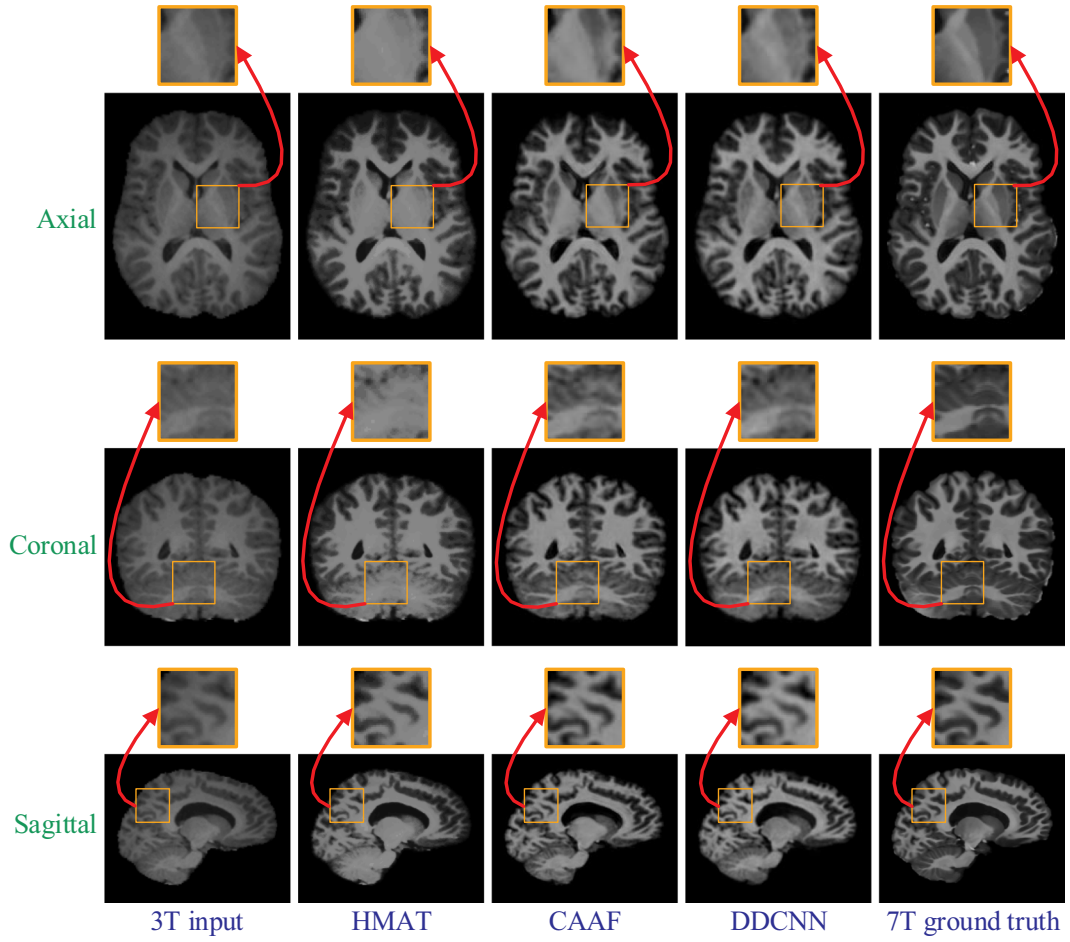

Fig. 8. Visual comparison of axial, sagittal and coronal views of synthesized 7 T-like images with close-up views of specific regions for another subject. For all methods, SSIM values are as follows: (a) 3 T input, (b) HMAT (SSIM = 0.3948), (c) CAAF (SSIM = 0.8261), (d) DDCNN (SSIM = **0.8359**), and (e) 7 T ground truth.

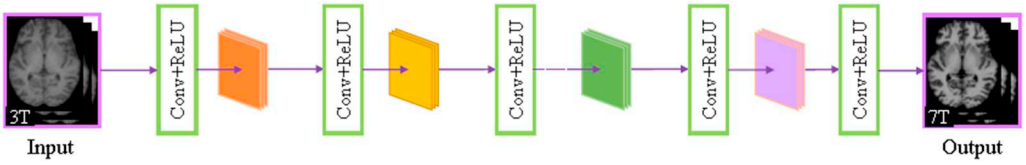

Fig. 9. Network architecture of SDCNN.

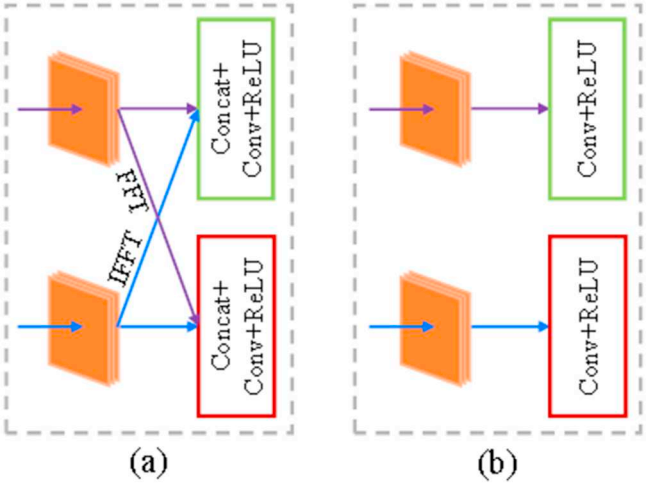

Fig. 10. Structural comparison of the butterfly module and the parallel module. From left to right: (a) butterfly module, and (b) parallel module.

every 50 epochs. To reconstruct high-quality 7 T-like images, we employ the cross-validation [1] for model selection by checking whether the model has been overfitted. The best trained model is applied to test images for synthesizing 7 T-like images.

4. Experimental results

4.1. Dataset

Approved by the local institutional review board (IRB), 15 subjects were recruited for MR data acquisition in a large study of brain MRI measurement. They consist of 5 healthy subjects, 8 patients with epilepsy and 2 patients with mild cognitive impairment (MCI). The 3 T and 7 T brain images of all subjects were acquired with Siemens Magnetom Trio 3 T and 7 T MRI scanners, respectively. Specifically, for 3 T MRI, T1 images of 224 coronal slices were obtained with the 3D magnetization-prepared rapid gradient-echo (MP-RAGE) sequence. The imaging parameters of the 3D MP-RAGE sequence were as follows: repetition time (TR) = 1900 ms, echo time (TE) = 2.16 ms, inversion time (TI) = 900 ms, flip angle (FA) = 9°, and voxel size = 1 × 1 × 1mm^3.

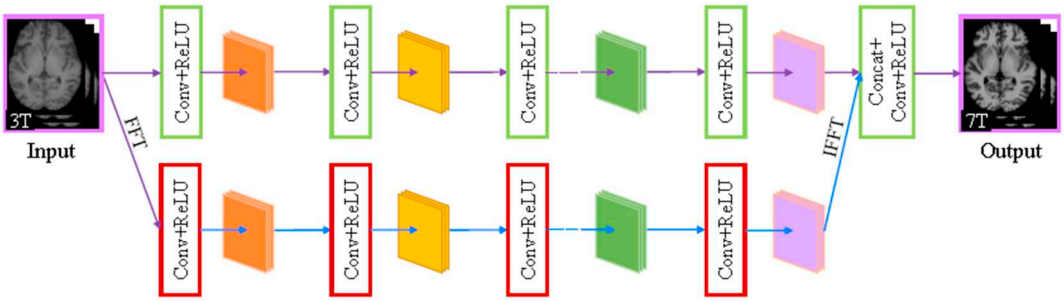

Fig. 11. Network architecture of PDCNN.

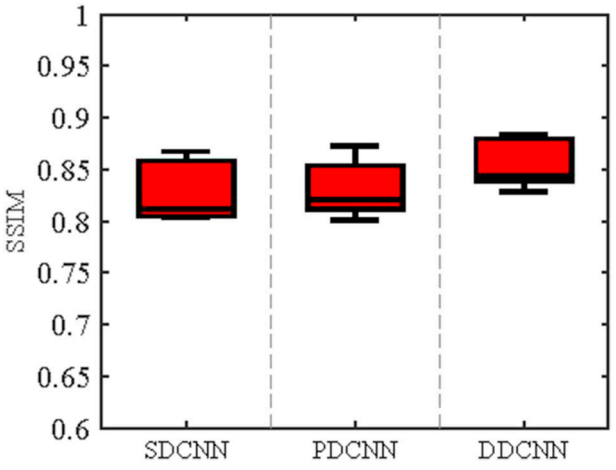

Fig. 12. Box plots for SSIM values. For all methods, the respective medians of SSIM values are as follows: (a) SDCNN (SSIM = 0.8113), (b) PDCNN (SSIM = 0.8206), and (c) DDCNN (SSIM = **0.8438**).

For 7 T MRI, T1 images of 191 sagittal slices were also obtained with the 3D MP2-RAGE sequence. The imaging parameters of the 3D MP2-RAGE sequence were as follows: TR = 6000 ms, TE = 2.95 ms, TI = 800/2700 ms, FA = 4°, and voxel size = 0.65 × 0.65 × 0.65mm^3. All 3 T images were resampled to the same spatial resolution of corresponding 7 T images. As the gradient echo pulse sequences were used for image acquisition, there is only little distortion between the acquired 3 T and 7 T MR images, which ensures the imaging consistency across magnetic fields.

4.2. Experimental setup

Extensive experiments were conducted to demonstrate the effectiveness of the proposed method. In all experiments, we adopted the leave-one-out cross-validation (LOOCV) for the evaluation. In each fold of LOOCV, one 3 T image was chosen for testing, whereas the remaining paired 3 T and 7 T images were divided into 10 image pairs for training and 4 image pairs for validation. The real 7 T image paired with the testing 3 T image was treated as the ground-truth image. For avoiding overfitting in the cross-validation, the training and validation images were used to select the best model from different network architectures and training parameters. In network configuration, the network depth $L = 4$, the batch size $b = 6$, the input channel number $c = 9$, filter number $n_f = 48$, the convolution kernel size $k_1 = 5$ in the input feature extraction layer, $k_2 = 3$ in the middle feature mapping layer, and $k_L = 1$ in the output feature reconstruction layer. The proposed models were implemented in Python with PyTorch and took approximately 2 days to train on an NVIDIA Titan Xp (i.e. graphics processing unit abbreviated as GPU).

For parameter settings, we manually tuned the parameters from the first layer to the last one, ensuring DDCNN close to its best performance. We also analyzed the robustness of DDCNN by changing a parameter while keeping all other parameters at their current values. As a function of the parameter value, the peak signal-to-noise ratio (PSNR) values were calculated between synthesized images and ground-truth images.

For a randomly selected subject, Fig. 5 shows PSNR values by DDCNN against the changes of key parameters: c, n_f, L, k_1, k_2 and k_L. Given training data and optimization algorithms, PSNR first increases faster and then hardly increases as c or n_f grows, whereas PSNR decreases as k_L increases. With the increase of other parameters (i.e., L, k_1, k_2), PSNR basically increases first and then decreases. Further explanation of these parameters can be seen in the Discussion section.

4.3. Method comparison

Several relevant methods, such as histogram matching (HMAT) and CAAF [2], were used as baseline methods for comparison both qualitatively and quantitatively. We adopted Structural SIMilarity (SSIM) index [29] for quantitative mage quality assessment. All synthesized images achieved by baseline methods and our method were compared with the ground-truth images in terms of SSIM.

Fig. 6 shows the box plots of SSIM values of 15 synthesized 7 T-like images obtained by different methods. As can be observed, DDCNN generally achieves higher SSIM than the other baseline methods. For further evaluation, the two sample t-test [15,33] was used to evaluate whether or not DDCNN outperforms the competing methods in SSIM. Table 1 gives p-values of two-sample t-tests of SSIM results between DDCNN and the competing methods. The p-values marked with an asteroid imply that DDCNN achieves significantly larger SSIM means than the competing methods.

Besides quantitative evaluation with respect to SSIM, Fig. 7 and 8 separately show axial, sagittal and coronal views of synthesized 7 T-like images for two randomly selected subjects, indicating that DDCNN achieves better visual effect and less distortion than the competing methods, i.e., HMAT and CAAF [2].

4.4. Ablation study

For further elaboration on the benefit of the dual-domain strategy for deep learning, we also compared DDCNN with single spatial-domain

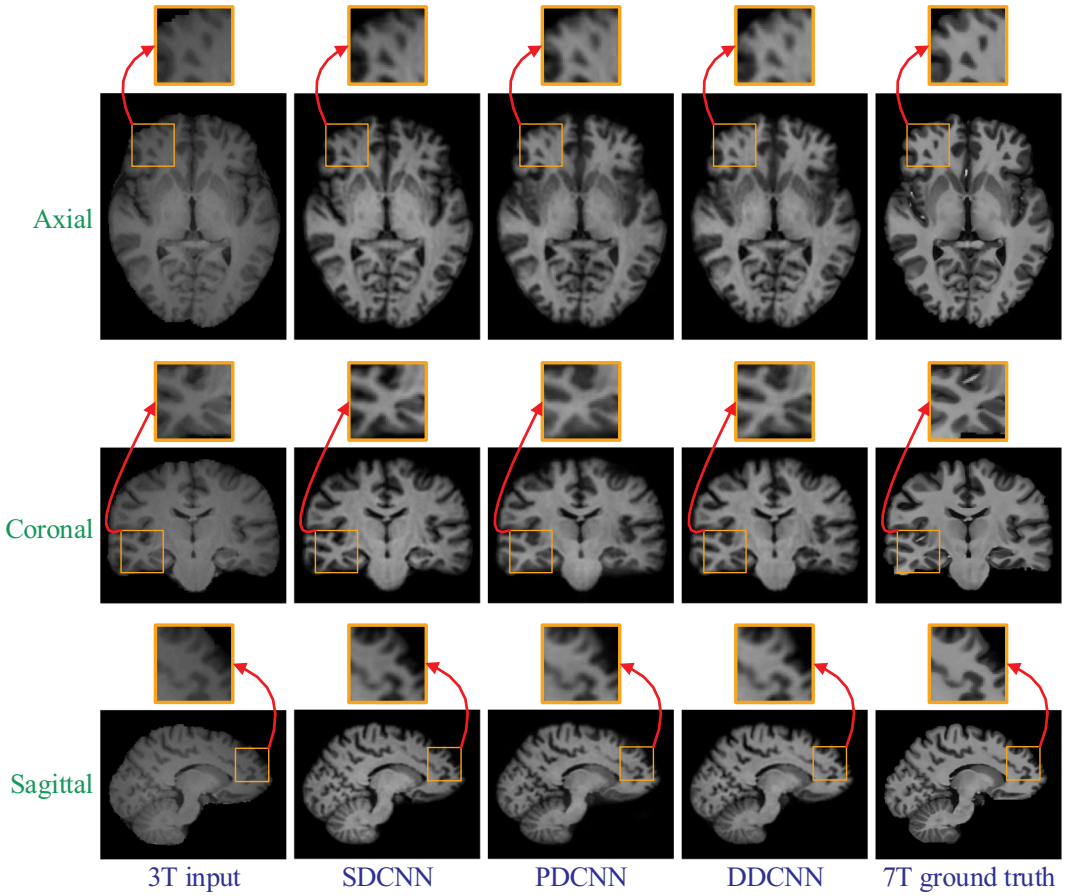

Fig. 13. Visual comparison of axial, sagittal and coronal views of synthesized 7 T-like images with close-up views of specific regions for one subject. For all methods, SSIM values are as follows: (a) 3 T input, (b) SDCNN (SSIM = 0.8584), (c) PDCNN (SSIM = 0.8531), (d) DDCNN (SSIM = **0.8792**), and (e) 7 T ground truth.

convolutional neural network (SDCNN) constructed by choosing the same parameters in the spatial domain. Fig. 9 shows the network architecture of SDCNN.

To evaluate the superiority of the butterfly modules, we introduced a parallel-domain convolutional neural network (PDCNN) as a competing method. PDCNN was constructed by replacing the butterfly modules (see Fig. 10(a)) with the parallel modules (see Fig. 10(b)) and keeping the others invariant to DDCNN. Fig. 11 shows the network architecture of PDCNN.

To verify the necessity of two domains and butterfly modules, we separately synthesized 7 T-like images from the same set of 3 T images with SDCNN, PDCNN and DDCNN. Fig. 12 shows the box plots of SSIM values of 15 synthesized 7 T-like images by these different methods. From Fig. 12, DDCNN achieves much higher SSIM than SDCNN and PDCNN. Table 1 indicates that DDCNN achieves significantly larger SSIM means than SDCNN and PDCNN, judging based on two-sample t-tests.

Besides the quantitative evaluation with respect to SSIM, Figs. 13 and 14 separately show axial, sagittal and coronal views of synthesized 7 T-like images for two randomly selected subjects, indicating that DDCNN has better visual results than SDCNN and PDCNN. Compared with SDCNN and PDCNN with parallel modules, DDCNN with butterfly modules makes great contributions to performance improvement of image synthesis. This indicates the necessity and superiority of the dual-domain strategy and butterfly modules.

4.5. Brain tissue segmentation

We further evaluated the influence of the 15 synthesized 7 T-like MR images on the post-processing procedures, such as brain tissue segmentation. We adopted a two-dimensional (2D) U-Net [21] with default parameters for brain tissue segmentation. The brain tissues were separated into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). The hand-segmented results of 7 T brain images were used as the ground truth. We compared the segmentation results of the original 3 T images, and the synthesized 7 T-like images by the competing HMAT, CAAF [2], SDCNN, PDCNN, and our DDCNN.

The segmentation results of synthesized images by U-Net [21] were evaluated using Dice similarity coefficient (DSC). Fig. 15 shows the box plots of DSC metrics for three brain tissues (i.e., GM, WM and CSF). Fig. 16 provides the visual comparison of segmentation results for one randomly selected subject. From Figs. 15 and 16, the segmentation results of synthesized images obtained by DDCNN are closer to the ground truth than those obtained by the competing methods, indicating that DDCNN improves the accuracy of brain MRI measurement.

5. Discussion

We have introduced a convolutional neural network for two complementary domains and parameter-efficient butterfly modules. The network learns an end-to-end mapping between pairs of 3 T and 7 T

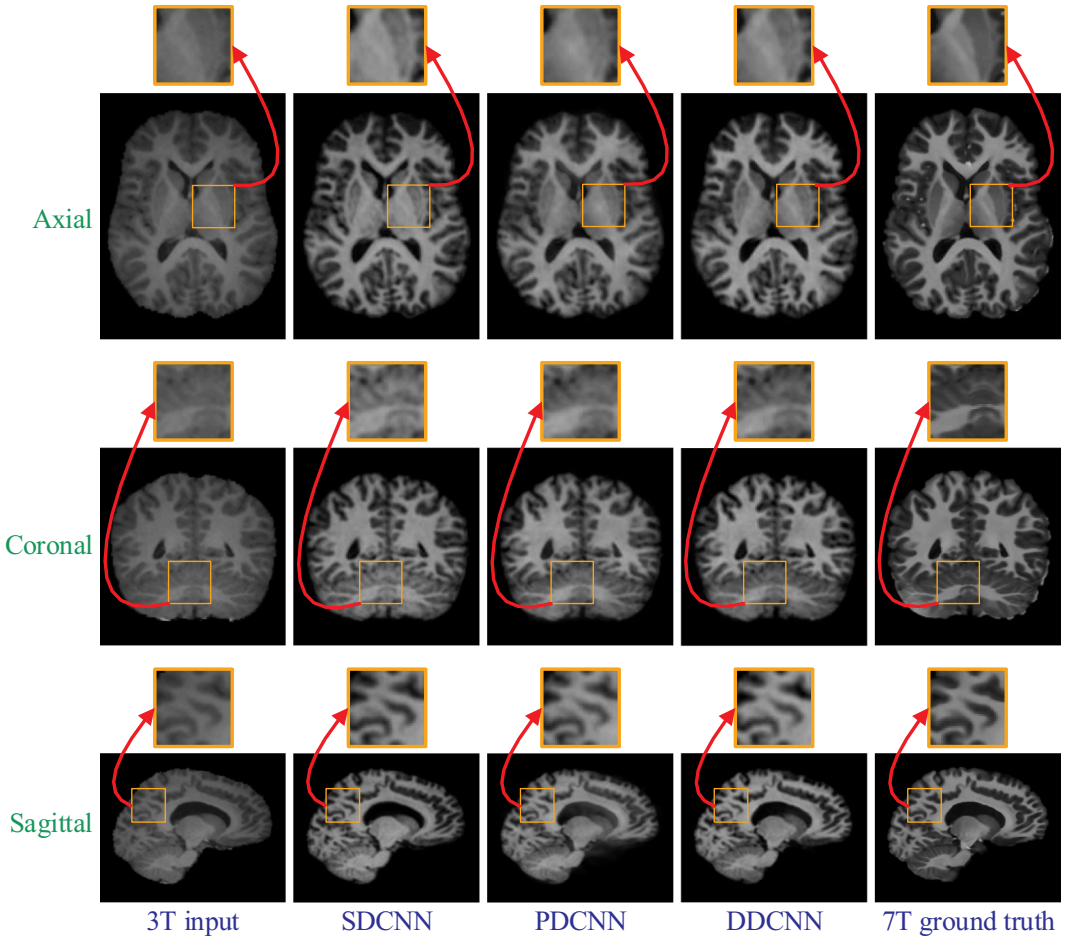

Fig. 14. Visual comparison of axial, sagittal and coronal views of synthesized 7 T-like images with close-up views of specific regions for another subject. For all methods, SSIM values are as follows: (a) 3 T input, (b) SDCNN (SSIM = 0.8134), (c) PDCNN (SSIM = 0.8073), (d) DDCNN (SSIM = **0.8359**), and (e) 7 T ground truth.

training images for image synthesis. Fig. 5 shows the robustness of the proposed method in terms of PSNR. Fig. 5 also shows that the depth and width of the network do not obey the rule of "bigger is better", but depend on training data and optimization algorithms. This is similar for the kernel size of the convolutional layer. The performance of DDCNN depends on its network architecture, the optimization algorithm, hyperparameter selection, the number of model parameters, and the amount of training data. Figs. 12 to 14 show that PDCNN achieves higher SSIM values than SDCNN, but lower than DDCNN. From Figs. 6 to 8, DDCNN with simple architecture outperforms the state-of-the-art methods (e.g., CAAF [2]) both qualitatively and quantitatively, suggesting that the synthesized images obtained by DDCNN resemble closer real 7 T images due to the effectiveness of the dual-domain strategy and butterfly modules.

We showed that the proposed method works well on the 3 T and 7 T images with spatial resolution 0:5 ~ 1 mm^3. In this paper, we assume that pairs of 3 T and 7 T training images are well aligned using FLIRT [9,10]. However, the registration error between 3 T and 7 T training images, especially at tissue boundaries, may affect the quality of the synthesized images.

6. Conclusions

In this paper, we have proposed a novel image synthesis method based on dual-domain convolutional neural networks. By introducing two complementary domains and parameter-efficient butterfly modules, two interactive streams of DDCNN on respective spatial and frequency domains benefit each other in learning complex mappings between 3 T and 7 T training images. With the well-trained model, the proposed method can synthesize high-quality 7 T-like images from input 3 T images. The experimental results suggest that the proposed method generally achieves better results than the state-of-the-art methods both qualitatively and quantitatively. The benefit of the dual-domain strategy and butterfly modules is also corroborated by the convincing results. To evaluate image synthesis comprehensively, we will research on functional information and disease diagnosis of synthesized images in the future.

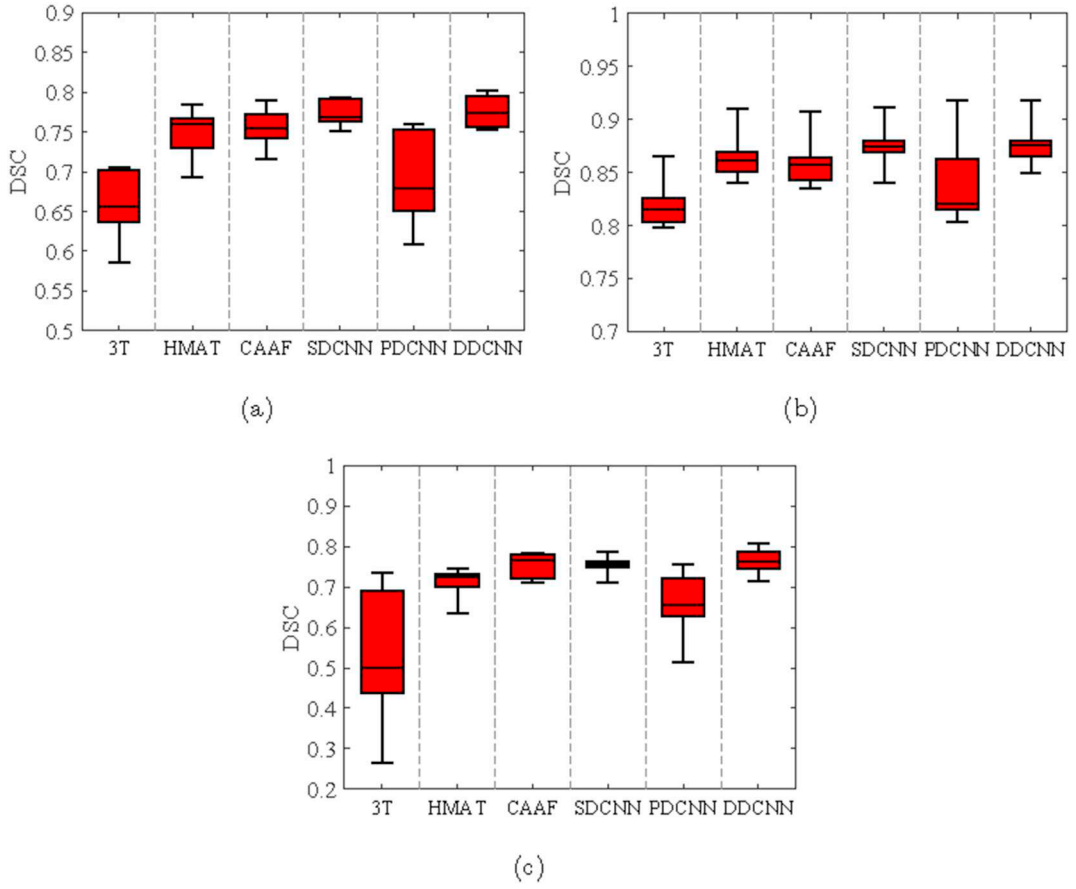

Fig. 15. Box plots of DSC values for segmentation of (a) GM, (b) WM, and (c) CSF from the reconstructed brain images by the competing methods and our DDCNN.

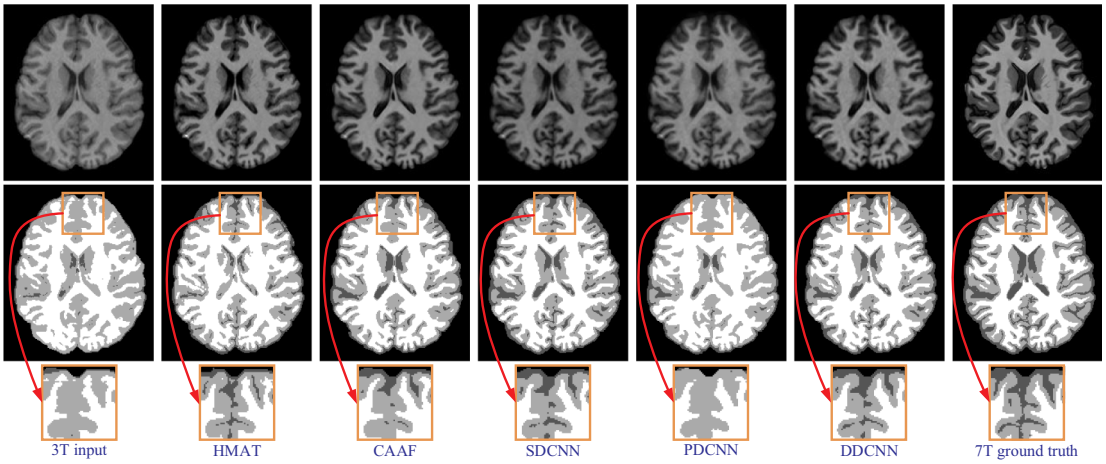

Fig. 16. Brain tissue segmentation results of reconstructed images by the competing methods and our DDCNN. From top to bottom, reconstructed images, segmentation results and close-up views.

Acknowledgements

Pew-Thian Yap, Liangqiong Qu, and Dinggang Shen were supported in part by NIH grants (EB006733 and AG053867). Yongqin Zhang was supported by Natural Science Basic Research Program of Shaanxi (Program No. 2019JM-103) and NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

[1] Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Stat Surv 2010;4:40–79.
[2] Bahrami K, Shi F, Rekik I, Shen D. Convolutional neural network for reconstruction of 7 T-like images from 3 T MRI using appearanceand anatomical features. Deep learning and data labeling for medical applications. DLMIA 2016, LABELS 2016. 2016. p. 39–47.
[3] Bahrami K, Shi F, Zong X, Shin HW, An H, Shen D. Reconstruction of 7 T-like images from 3 T MRI. IEEE Trans Med Imaging 2016;35(9):2085–97.

[4] Bhavsar A, Wu G, Lian J, Shen D. Resolution enhancement of lung 4D-CT via group-sparsity. Med Phys 2013;40(12).
[5] Dong C, Chen CL, He K, Tang X. Learning a deep convolutional network for image super-resolution. European conference on computer vision. 2014. p. 184–99.
[6] Forstmann BU, Isaacs BR, Temel Y. Ultra high field MRI-guided deep brain stimulation. Trends Biotechnol 2017;35(10):904–7.
[7] Han D, Liu Q, Fan W. A new image classification method using CNN transfer learning and web data augmentation. Expert Syst Appl 2018;95:43–56.
[8] Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. Enhancement of MR images using registration for signal averaging. J Comput Assist Tomo 1998;22(2):324–33.
[9] Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 2002;17(2):825–41.
[10] Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. NeuroImage 2012;62(2):782–90.
[11] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization.arXiv:1412.6980, 2014.
[12] Kraff O, Fischer A, Nagel AM, Moenninghoff C, Ladd ME. MRI at 7 tesla and above: demonstrated and potential capabilities. J Magn Reson Imaging 2015;41(1):13–33.
[13] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems. 2012. p. 1106–14.
[14] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90.
[15] Krzywinski M, Altman N. Significance, P values and *t*-tests. Nat Methods 2013;10(11):1041–2.
[16] Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. IEEE conference on computer vision and pattern recognition. 2017. p. 105–14.
[17] Li H, Yang J, Xiao Y, Fan Y. Accurate hippocampus segmentation using fully convolutional networks and deformable image registration based data augmentation. Annual meeting of American Association of Physics in Medicine. 2018. p. E362.
[18] Mao XJ, Shen C, Yang YB. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. Advances in neural information processing systems. 2016. p. 2802–10.
[19] Mcdonagh S, Hou B, Alansary A, Oktay O, Kamnitsas K, Rutherford M, et al. Context-sensitive super-resolution for fast fetal magnetic resonance imaging. Molecular imaging, reconstruction and analysis of moving body organs, and stroke imaging and treatment. 2017. p. 116–26. RAMBO 2017, CMMI 2017, SWITCH 2017.
[20] Rogalla H, Kes PH. 100 years of superconductivity. Taylor & Francis; 2011.
[21] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. International conference on medical image computing and computer assisted intervention. 2015. p. 234–41.
[22] Roy Snehashis, Carass Aaron, Prince Jerry L. Magnetic resonance image example-based contrast synthesis. IEEE Trans Med Imaging 2013;32(12):2348–63.
[23] Rueda A, Malpica N, Romero E. Single-image super-resolution of brain MR images using overcomplete dictionaries. Med Image Anal 2013;17(1):113–32.
[24] Shi F, Fan Y, Tang S, Gilmore JH, Lin W, Shen D. Neonatal brain image segmentation in longitudinal MRI studies. NeuroImage 2010;49(1):391–400.
[25] Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 1998;17(1):87–97.
[26] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network. IEEE conference on computer vision and pattern recognition. 2017. p. 1–9.
[27] Touloupou P, Alzahrani N, Neal P, Spencer SEF, McKinley TJ. Efficient model comparison techniques for models requiring large scale data augmentation. Bayesian Anal 2018;13(2):437–59.
[28] van der Zwaag W, Schaefer A, Marques JP, Turner R, Trampel R. Recent applications of UHF-MRI in the study of human brain function and structure: a review. NMR Biomed 2016;29(9):1274–88.
[29] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13(4):600–12.
[30] Yang J, Wright J, Huang TS, Ma Y. Image super-resolution via sparse representation. IEEE Trans Image Process 2010;19(11):2861–73.
[31] Zhang Y, Cheng JZ, Xiang L, Yap PT, Shen D. Dual-domain cascaded regression for synthesizing 7 T from 3 T MRI. International conference on medical image computing and computer assisted intervention. 2018. p. 410–7.
[32] Zhang Y, Liu J, Yang W, Guo Z. Image super-resolution based on structure-modulated sparse representation. IEEE Trans Image Process 2015;24(9):2797–810.
[33] Zhang Y, Shi F, Cheng J, Wang L, Yap PT, Shen D. Longitudinally guided super-resolution of neonatal brain magnetic resonance images. IEEE Trans Cybern 2019;49(2):662–74.
[34] Zhang Y, Wu G, Yap PT, Feng Q, Lian J, Chen W, et al. Hierarchical patch-based sparse representation-a new approach for resolution enhancement of 4D-CT lung data. IEEE Trans Med Imaging 2012;31(11):1993–2005.
[35] Zhang Y, Yap P-T, Chen G, Lin W, Wang L, Shen D. Super-resolution reconstruction of neonatal brain MR images via residual structured sparse representation. Med Image Anal 2019;55:76–87.