Dual-Domain Cascaded Regression for Synthesizing 7T from 3T MRI

Yongqin Zhang[1,2], **Jie-Zhi Cheng**[3], **Lei Xiang**[1], **Pew-Thian Yap**[1], and **Dinggang Shen**[1]

[1]Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[2]School of Information Science and Technology, Northwest University, Xi'an 710127, China

[3]Shanghai United Imaging Intelligence Co., Ltd., Shanghai 201807, China

Abstract

Due to the high cost and low accessibility of 7T magnetic resonance imaging (MRI) scanners, we propose a novel dual-domain cascaded regression framework to synthesize 7T images from the routine 3T images. Our framework is composed of two parallel and interactive multi-stage regression streams, where one stream regresses on spatial domain and the other regresses on frequency domain. These two streams complement each other and enable the learning of complex mappings between 3T and 7T images. We evaluated the proposed framework on a set of 3T and 7T images by leave-one-out cross-validation. Experimental results demonstrate that the proposed framework generates realistic 7T images and achieves better results than state-of-the-art methods.

1 Introduction

Since early 2000s, 3T MRI has become the standard for research and clinical applications. In 2017, the first 7T MRI scanner was approved for clinical use by the United States Food and Drug Administration (FDA)[1]. Compared with 3T MRI, 7T MRI typically affords greater anatomical details and faster image reconstruction, which may benefit the diagnosis of diseases [1]. However, 7T MRI scanners are significantly more expensive and hence less common at hospitals and clinical institutions. To date, there are less than 100 7T MRI scanners worldwide [2]. Accordingly, this motivates the research on 7T image synthesis using the low-field images (e.g., 3T images). In this work, we show how the prediction of 7T MRI from 3T images can be improved by concurrently considering the spatial and frequency domains in a regression framework.

The basic goal of 7T image synthesis is to map low-resolution (LR) 3T images to high-resolution (HR) 7T images. But this is not a simple super-resolution problem because the appearance and contrast of 7T images can be different from those of 3T images. For this purpose, a number of machine learning methods have been proposed in recent years. Bhavsar et al. [3] introduced a group-sparse representation method for resolution

---

[1]https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm580154.htm.

enhancement of CT lung images. Roy et al. [4] presented an example-based super-resolution framework to synthesize HR MR images from multi-contrast atlases. To enhance the quality and resolution of neonatal images, Zhang et al. [5] proposed a super-resolution method with the guidance of longitudinal data. Bahrami et al. [6] proposed a hierarchical sparse representation method with multi-level canonical correlation analysis (M-CCA) for the reconstruction of 7T-like MR images from 3T MR images. A deep learning approach has appeared for resolution enhancement in [7]. Bahrami et al. [8] developed a CNN-based approach that takes into account appearance and anatomical features (CAAF) to predict 7T images from 3T images.

In this paper, we propose a dual-domain cascaded regression (DDCR) framework to synthesize realistic 7T from 3T images with two parallel and interactive multi-stage regression streams based on spatial and frequency domains. Our framework employs complementary cues on both domains to learn complex mappings from 3T to 7T modalities. Comparisons with the existing methods indicate that synthesized 7T images with higher quality can be obtained with DDCR.

2 Method

DDCR (Fig. 1) formulates the mapping between the 7T and 3T images as a regression problem. Specifically, DDCR regresses the local patches of 7T images from local patches of 3T images. DDCR uses intensity and spectral transformations to improve the quality of 7T image synthesis. DDCR consists of two steps: (1) image preprocessing and (2) multi-stage regression.

2.1 Image Preprocessing

An input 3T image $\mathbf{Y}$ and pairs of 3T and 7T exemplar images $\{\mathbf{J}_{3T}, \mathbf{J}_{7T}\}$ are registered to MNI standard space [9] using FLIRT [10,11] to remove pose differences. Specifically, all 7T exemplar images are linearly registered to the MNI standard space with an individual template [9]. The 3T exemplar image is then rigidly aligned to its corresponding 7T image. After registration, bias correction [12] and skull stripping [13] are performed. The image intensity values are normalized using histogram matching and scaled to range of [0, 1]. Histogram matching is performed separately for 3T and 7T images. For 3T images, the histograms of all normalized 3T exemplar images are matched to the histogram of the normalized input 3T image. Following that, the normalized 7T exemplar image whose corresponding 3T exemplar image is nearest to the input 3T image in Euclidean distance is chosen as referenced 7T image for the histogram matching of all remaining 7T images.

2.2 Multi-stage Regression

To model the mapping from 3T (LR) to 7T (HR) modalities, a dual-domain multi-stage regression is developed. As shown in Fig. 1, the regression process is carried out on the two streams of spatial and frequency domains in multiple stages. In this study, the frequency domain is computed with the simple but efficient discrete cosine transform. For each stage of regression, the regression mapping from 3T to 7T modalities is performed separately. The regression results of spatial and frequency domains are then fused together as a new input

for the later regression stage. It is worth noting that input image of the first stage regression is the normalized 3T image, whereas the input images of the remaining stages along each stream are the intermediate synthesized 7T images.

For the first stage, given an input 3T image $\mathbf{X} \in \mathscr{R}^{M \times N \times O}$ and Q pairs of 3T and 7T exemplar images $\mathbf{Z}_{3T}, \mathbf{Z}_{7T} \in \mathscr{R}^{M \times N \times O \times Q}$, we divide the input image into patches, denoted as $\mathbf{x}$, with size $p \times p \times p$ for patch regression. For each input patch, we collect L_1 most similar patches $\{\mathbf{z}_{3T,l} | l = 1, \ldots, L_1\}$ from the 3T exemplar images with block-matching method [14]. The 7T patches $\{\mathbf{z}_{7T,l} | l = 1, \ldots, L_1\}$ with the same locations of 3T exemplar patches are also collected from the corresponding 7T exemplar images. These 3T and 7T patch pairs are employed for the construction of the LR and HR dictionaries: $\mathbf{D}_{LR} = [\mathbf{z}_{3T,1}, \ldots, \mathbf{z}_{3T,L_1}]$ and $\mathbf{D}_{HR} = [\mathbf{z}_{7T,1}, \ldots, \mathbf{z}_{7T,L_1}]$ in the spatial domain, respectively. We propose a linear regression model to represent the mapping from the LR dictionary to the HR dictionary:

$$\mathbf{D}_{HR} = \mathbf{B}_s \mathbf{D}_{LR} + \varepsilon, \quad (1)$$

where $\mathbf{B}_s$ is the projection matrix, and ε is the error. We employ the ridge regression [15] to solve the inverse problem (1) in an optimization form:

$$\widehat{\mathbf{B}}_s = \min_{\mathbf{B}_s} \|\mathbf{D}_{HR} - \mathbf{B}_s \mathbf{D}_{LR}\|_2^2 + \lambda \|\mathbf{B}_s\|_2^2, \quad (2)$$

where λ is a regularization parameter. By taking the first derivative of (2) with respect to the variable $\mathbf{B}_s$ and setting it to zero, the projection matrix can be expressed in a closed-form solution:

$$\widehat{\mathbf{B}}_s = \mathbf{D}_{HR} \mathbf{D}'_{LR} (\mathbf{D}_{LR} \mathbf{D}'_{LR} + \lambda \mathbf{I})^{-1}, \quad (3)$$

where $\mathbf{D}'_{LR}$ denotes the transpose of the matrix $\mathbf{D}_{LR}$, and $\mathbf{I}$ is an identity matrix. According to the inverse matrix identity [16], the estimated projection matrix $\widehat{\mathbf{B}}_s$ in the spatial domain can be rewritten in another form with low computational complexity as follows:

$$\widehat{\mathbf{B}}_s = \mathbf{D}_{HR} (\mathbf{D}'_{LR} \mathbf{D}_{LR} + \lambda \mathbf{I})^{-1} \mathbf{D}'_{LR}. \quad (4)$$

By fixing $\widehat{\mathbf{B}}_s$, the preliminary synthesized $\widehat{\mathbf{y}}_{se}$ of the 7T MR patch $\mathbf{y}$ from the input 3T patch $\mathbf{x}$ becomes a simple matrix projection:

$$\widehat{\mathbf{y}}_{se} = \widehat{\mathbf{B}}_s \mathbf{x}. \quad (5)$$

On the other hand, we can also estimate a synthesized HR dictionary $\widehat{\mathbf{D}}_{se}$ from the LR dictionary $\mathbf{D}_{LR}$ in the following form:

$$\widehat{\mathbf{D}}_{se} = \widehat{\mathbf{B}}_s \mathbf{D}_{LR}, \quad (6)$$

where $\widehat{\mathbf{D}}_{se}$ is denoted as synthesized HR dictionary and can be referenced for the construction of LR dictionary for the next regression stage.

For the first stage of regression in the frequency domain, let α, $\mathbf{U}_{LR}$ and $\mathbf{U}_{HR}$ stand for the respective DCT coefficients of $\mathbf{x}$, $\mathbf{D}_{LR}$ and $\mathbf{D}_{HR}$, respectively. Similar to the regression in the spatial domain, the synthesized 7T component and the synthesized HR dictionary in the frequency domain can be separately computed as

$$\widehat{\mathbf{v}}_{te} = \mathbf{U}_{HR}\big(\mathbf{U}'_{LR}\mathbf{U}_{LR} + \lambda\mathbf{I}\big)^{-1}\mathbf{U}'_{LR}\alpha, \quad (7)$$

and

$$\widehat{\mathbf{U}}_{te} = \mathbf{U}_{HR}\big(\mathbf{U}'_{LR}\mathbf{U}_{LR} + \lambda\mathbf{I}\big)^{-1}\mathbf{U}'_{LR}\mathbf{U}_{LR}. \quad (8)$$

These temporary results can be referenced as the input and the construction of LR dictionary for the next stage.

After regression on both spatial and frequency domain, we further fuse the regression results of $\widehat{\mathbf{y}}_{se}$, $\widehat{\mathbf{D}}_{se}$, $\widehat{\mathbf{v}}_{te}$ and $\widehat{\mathbf{U}}_{te}$ in the spatial and frequency domains as follows:

$$\widehat{\mathbf{y}} = \sqrt{(\widehat{\mathbf{y}}_{se} \circ \widehat{\mathbf{y}}_{se} + \mathrm{idct}(\widehat{\mathbf{v}}_{te}) \circ \mathrm{idct}(\widehat{\mathbf{v}}_{te})) / 2}, \quad (9)$$

$$\widehat{\mathbf{D}}_{HR} = \sqrt{\left(\widehat{\mathbf{D}}_{se} \circ \widehat{\mathbf{D}}_{se} + \mathrm{idct}(\widehat{\mathbf{U}}_{te}) \circ \mathrm{idct}(\widehat{\mathbf{U}}_{te})\right) / 2}, \quad (10)$$

$$\widehat{\mathbf{v}} = \sqrt{(\mathrm{dct}\,(\widehat{\mathbf{y}}_{se}) \circ \mathrm{dct}\,(\widehat{\mathbf{y}}_{se}) + \widehat{\mathbf{v}}_{te} \circ \widehat{\mathbf{v}}_{te}) / 2}, \quad (11)$$

$$\widehat{\mathbf{U}}_{HR} = \sqrt{\left(\mathrm{idct}\left(\widehat{\mathbf{D}}_{se}\right) \circ \mathrm{idct}\left(\widehat{\mathbf{D}}_{se}\right) + \widehat{\mathbf{U}}_{te} \circ \widehat{\mathbf{U}}_{te}\right) / 2}, \quad (12)$$

where dct $(\cdot)$ and idct $(\cdot)$ represent forward and inverse DCT functions, respectively, and the operator $\bigcirc$ is the Hadamard product of two matrices.

With the computed $\hat{\mathbf{y}}$, $\widehat{\mathbf{D}}_{HR}$, $\hat{\mathbf{v}}$ and $\widehat{\mathbf{U}}_{HR}$, the cascaded stages of regression can be further carried out in the streams of respective domains. Specifically, for the cascaded regression in the spatial domain, the synthesized $\hat{\mathbf{y}}$ and dictionary $\widehat{\mathbf{D}}_{HR}$ at the stage k are taken as the input $\mathbf{x}$ and the LR dictionary $\mathbf{D}_{LR}$ of the stage k+1, respectively. Similarly, in the frequency domain, the synthesized HR component $\hat{\mathbf{v}}$ and HR dictionary $\widehat{\mathbf{U}}_{HR}$ at the stage k are treated as the input LR component a and the LR dictionary $\mathbf{U}_{LR}$ of the stage $k+1$. On the other hand, the HR dictionaries $\mathbf{D}_{HR}$ and $\mathbf{U}_{HR}$ from the stage k in the spatial and frequency domains are treated as the HR dictionaries of the stage $k+1$, respectively. With the setup of input, LR and HR dictionaries on both domains, we can perform the equations (5)–(8) to obtain regression results at the current stage. Then we can fuse the regression results with equations (9)–(12) again for the next stage. With K stages of regression, we further collect all synthesized 7T patches to construct the final result.

3 Experiments and Results

3.1 Dataset

With the local institutional review board (IRB), 15 adults were recruited for MR data acquisition in this study. The 3T and 7T brain images of all subjects were acquired with Siemens Magnetom Trio 3T and 7T MRI scanners, respectively. Specifically, for 3T images, T1 images of 224 coronal slices were obtained with the 3D magnetization-prepared rapid gradient-echo (MP-RAGE) sequence. The imaging parameters of 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 = $1 \times 1 \times 1$ mm^3. For 7T images, T1 images of 191 sagittal slices were also obtained with the 3D MP2-RAGE sequence. The imaging parameters of 3D MP2-RAGE sequence were as follows: TR = 6000 ms, TE = 2.95 ms, TI = 800/2700 ms, FA = 4°/4°, and voxel size = $0.65 \times 0.65 \times 0.65$ mm^3. As the gradient echo pulse sequences were used for image acquisition, there is only little distortion between the obtained 3T and 7T MR images, which ensures the imaging consistency across magnetic fields.

3.2 Experimental Setup

Extensive experiments were conducted to illustrate the effectiveness of the proposed method. In all experiments, we adopted leave-one-out cross-validation (LOOCV) for the evaluation. Specifically, in one fold of LOOCV, one 3T MR image was chosen for testing, whereas the remaining paired 3T and 7T MR images were treated as exemplars. The 7T image paired with the testing 3T image was treated as the ground truth image. For simplicity, we chose two stages for the implementation of the multi-stage strategy. The parameters of the proposed method were as follows: $p = 3$, $\lambda = 0.001$, $Q = 14$, $K = 2$, $L_1 = 25$ and $L_2 = 1$ in all experiments, where L_2 is the number of similar patches in the second stage regression. For the parameter settings, we manually tuned these parameters from the first stage to the last to ensure that the proposed method approximates its best performance.

3.3 Results

Several relevant methods like histogram matching (HMAT), M-CCA [6] and CAAF [8] were used as baseline methods for comparison. Meanwhile, to further illustrate the benefit of dual-domain strategy for the cascaded regression, DDCR was also compared with single spatial-domain cascaded regression, denoted as SDCR. For quantitative image quality assessment, we adopted two evaluation metrics: PSNR and Structural SIMilarity (SSIM) index [17]. All synthesized images from baseline methods and our method were compared with the real 7T images for the computation of PSNR and SSIM. Figure 2 shows the box-plots of PSNR and SSIM values of 15 synthesized 7T MR images. As can be observed, DDCR generally achieves higher PSNR and SSIM than the other baseline methods. Even though SDCR almost achieves the same PSNR values as DDCR, SDCR has distinctly lower SSIM values than DDCR. Figure 3 shows the axial, sagittal and coronal views of synthesized 7T MR images for one randomly selected subject. It can also be found from Fig. 3 that the synthesized 7T image by DDCR has better image quality and less distortion.

4 Discussion

Referring to Figs. 2 and 3, the proposed cascaded regression method can synthesize 7T images with better quality. Meanwhile, as can be found in Fig. 2, although the performance of DDCR w.r.t. the PSNR index is similar to that of SDCR, DDCR achieves higher SSIM values than SDCR. It is suggested that the synthesized 7T images by DDCR are more similar to real 7T images and thus validate the effectiveness of the dual-domain strategy.

We convert the image synthesis problem into a regression problem and solve it in a closed form. Multi-stage regression is also employed to further improve the quality of image synthesis. By introducing two complementary domains, two regression streams on respective spatial and frequency domains benefit each other in learning complex mappings between 3T and 7T images. The proposed method is simple and effective with low computational cost, which outperforms sparse representation based methods (e.g., M-CCA [6]) and can even compete with deep learning based methods (e.g., CAAF [8]). The proposed method is free of training process and does not require large amounts of training data. With limited 7T exemplar data, we can still achieve satisfactory 7T image synthesis. Therefore, the dependence of the proposed method on large training data, particularly less available 7T data, is not very strong.

5 Conclusion

In this paper, we have proposed a novel image synthesis method based on dualdomain cascaded regression. With the reference of pairs of 3T and 7T MR exemplar images, the proposed method can synthesize high-quality 7T images from 3T images. The experimental results suggest that the proposed method generally achieves better results than the state-of-the-art methods both qualitatively and quantitatively and thus corroborate the efficacy of the proposed method. For big training data, dual-domain convolutional neural network for image synthesis is left as our future research work.

Acknowledgements.

References

1. Zwaag W, Schafer 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. 29(9), 1274–1288 (2016) [PubMed: 25762497]

2. Forstmann BU, Isaacs BR, Temel Y: Ultra high field MRI-guided deep brain stimulation. Trends Biotechnol. 35(10), 904–907 (2017) [PubMed: 28941469]

3. Bhavsar A, Wu G, Lian J, Shen D: Resolution enhancement of lung 4D-CT via group-sparsity. Med. Phys 40(12), 121717 (2013) [PubMed: 24320503]

4. Roy S, Carass A, Prince JL: Magnetic resonance image example-based contrast synthesis. IEEE Trans. Med. Imaging 32(12), 2348–2363 (2013) [PubMed: 24058022]

5. 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. (2018). 10.1109/TCYB. 2017.2786161

6. Bahrami K, Shi F, Zong X, Shin HW, An H, Shen D: Reconstruction of 7T-like images from 3T MRI. IEEE Trans. Med. Imaging 35(9), 2085–2097 (2016) [PubMed: 27046894]

7. Dong C, Loy CC, He K, Tang X: Learning a deep convolutional network for image super-resolution In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds.) ECCV 2014. LNCS, vol. 8692, pp. 184–199. Springer, Cham (2014). 10.1007/978-3-319-10593-2_13

8. Bahrami K, Shi F, Rekik I, Shen D: Convolutional neural network for reconstruction of 7T-like images from 3T MRI using appearance and anatomical features In: Carneiro G, et al. (eds.) LABELS/DLMIA -2016. LNCS, vol. 10008, pp. 39–47. Springer, Cham (2016). 10.1007/978-3-319-46976-8_5

9. Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC: Enhancement of MR images using registration for signal averaging. J. Comput. Assist. Tomogr. 22(2), 324–333 (1998) [PubMed: 9530404]

10. Jenkinson M, Bannister P, Brady JM, Smith SM: Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17(2), 825–841 (2002) [PubMed: 12377157]

11. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM: FSL. NeuroImage 62, 782–90 (2012) [PubMed: 21979382]

12. Sled JG, Zijdenbos AP, Evans AC: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17(1), 87–97 (1998) [PubMed: 9617910]

13. Shi F, Fan Y, Tang S, Gilmore JH, Lin W, Shen D: Neonatal brain image segmentation in longitudinal MRI studies. NeuroImage 49(1), 391–400 (2010) [PubMed: 19660558]

14. Brunig M, Niehsen W: Fast full-search block matching. IEEE Trans. Circuits Syst. Video Technol. 11(2), 241–247 (2001)

15. Zhang Y, Liu J, Yang W, Guo Z: Image super-resolution based on structure-modulated sparse representation. IEEE Trans. Image Process. 24(9), 2797–2810 (2015) [PubMed: 25966473]

16. Petersen KB, Pedersen MS: The Matrix Cookbook (2012)

17. Wang Z, Bovik A, Sheikh H, Simoncelli E: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004) [PubMed: 15376593]
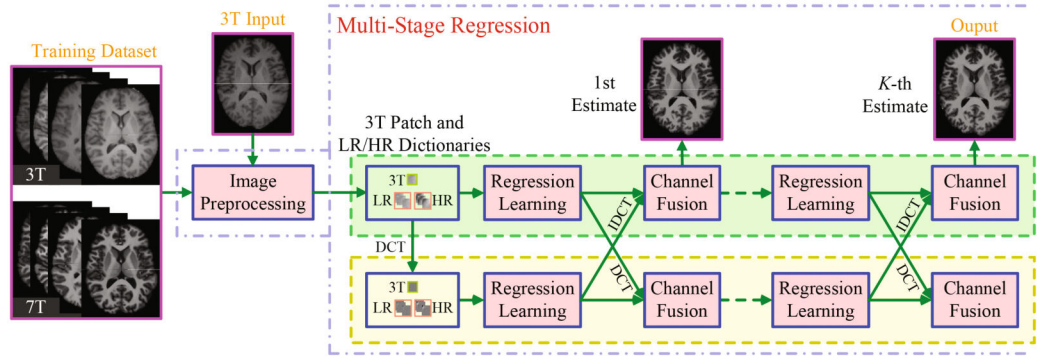
Fig. 1.
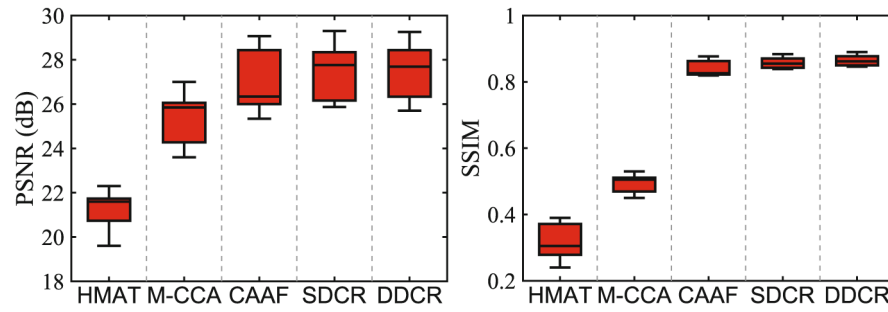Method overview. DCT and IDCT are the forward and inverse discrete cosine transforms, respectively.

Fig. 2.
Box-plots for PSNR and 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 PSNR and SSIM values are as follows: (a) HMAT (PSNR = 21.6 dB, SSIM = 0.30), (b) M-CCA (PSNR = 25.8 dB, SSIM = 0.50), (c) CAAF (PSNR = 26.3 dB, SSIM = 0.83), (d) SDCR (PSNR = 27.7 dB, SSIM = 0.85), and (e) DDCR (PSNR = 27.7 dB, SSIM = 0.86).
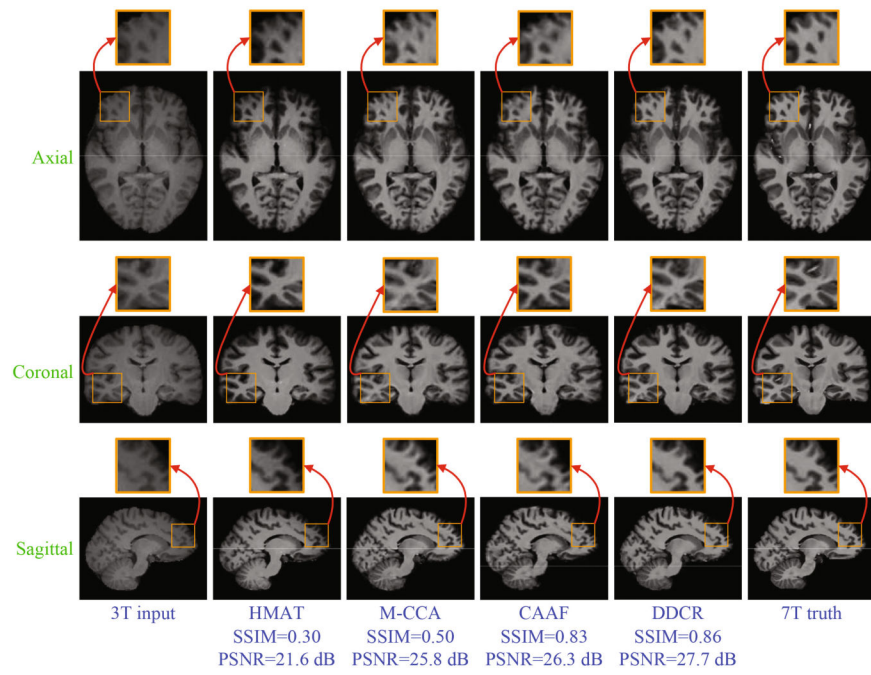
Fig. 3.
Visual comparison of axial, sagittal and coronal views of synthesized 7T images with close-up views of specific regions for one subject.