DOTE: Dual cOnvolutional filTer lEarning for Super-Resolution and Cross-Modality Synthesis in MRI

Yawen Huang[1(✉)], Ling Shao[2], and Alejandro F. Frangi[1]

[1] Department of Electronic and Electrical Engineering,
The University of Sheffield, Sheffield, UK
{yhuang36,a.frangi}@sheffield.ac.uk
[2] School of Computing Sciences, University of East Anglia, Norwich, UK
ling.shao@uea.ac.uk

Abstract. Cross-modal image synthesis is a topical problem in medical image computing. Existing methods for image synthesis are either tailored to a specific application, require large scale training sets, or are based on partitioning images into overlapping patches. In this paper, we propose a novel Dual cOnvolutional filTer lEarning (DOTE) approach to overcome the drawbacks of these approaches. We construct a closed loop joint filter learning strategy that generates informative feedback for model self-optimization. Our method can leverage data more efficiently thus reducing the size of the required training set. We extensively evaluate DOTE in two challenging tasks: image super-resolution and cross-modality synthesis. The experimental results demonstrate superior performance of our method over other state-of-the-art methods.

Keywords: Dual learning · Convolutional sparse coding · 3D · Multi-modal · Image synthesis · MRI

1 Introduction

In medical image analysis, it is sometimes convenient or necessary to infer an image from one modality or resolution from another image modality or resolution for better disease visualization, prediction and detection purposes. A major challenge of cross-modality image segmentation or registration comes from the differences in tissue appearance or spatial resolution in images arising from different physical acquisition principles or parameters, which translates into the difficulty to represent and relate these images. Some existing methods tackle this problem by learning from a large amount of registered images and constraining pairwise solutions in a common space. In general, one would desire to have high-resolution (HR) three-dimensional Magnetic Resonance Imaging (MRI) with near isotropic voxel resolution as opposed to the more common image stacks of multiple 2D slices for accurate quantitative image analysis and diagnosis. Multi-modality

© Springer International Publishing AG 2017
M. Descoteaux et al. (Eds.): MICCAI 2017, Part III, LNCS 10435, pp. 89–98, 2017.
DOI: 10.1007/978-3-319-66179-7_11

imaging can generate tissue contrast arising from various anatomical or functional features that present complementary information about the underlying organ. Acquiring low-resolution (LR) single-modality images, however, is not uncommon.

To solve the above problems, super-resolution (SR) [1,2] reconstruction is carried out for recovering an HR image from its LR counterpart, and cross-modality synthesis (CMS) [3] is proposed for synthesizing target modality data from available source modality images. Generally, these methods have explored image priors from either internal similarities of image itself [4] or external data support [5], to construct the relationship between two modalities. Although these methods achieve remarkable results, most of them suffer from the fundamental limitations associated with large scale pairwise training sets or patch-based overlapping mechanism. Specifically, a large amount of multi-modal images is often required to learn a sufficiently expressive dictionaries/networks. However, this is impractical since collecting medical images is very costly and limited by many factors. On the other side, patch-based methods are subjected to inconsistencies introduced during the fusion process that takes place in areas where patches overlap.

To deal with the bottlenecks of training data and patch-based implementation, we develop a dual convolutional filter learning (DOTE) method with an application to neuroimaging that investigates data (in both source and target modalities from the same set of subjects) in a more effective way, and solves image SR and CMS problems respectively. The contributions of this work are mainly in four aspects: (1) We present a unified model (DOTE) for any cross-modality image synthesis problem; (2) The proposed method can efficiently reduce the amount of training data needed from the model, by generating abundant feedbacks from dual mapping functions during the training process; (3) Our method integrates feature learning and mapping relation in a closed loop for self-optimization. Local neighbors are preserved intrinsically by directly working on the whole images; (4) We evaluate DOTE on two datasets in comparison with stat-of-the-art methods. Experimental results demonstrate superior performance of DOTE over these approaches.

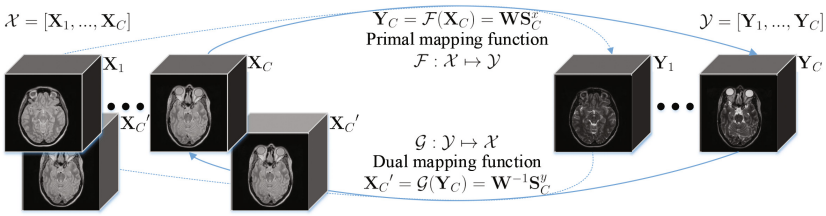

Fig. 1. Flowchart of the proposed method for MRI cross-modality synthesis.

2 Method

2.1 Background

Convolutional Sparse Coding (CSC) remedies a fundamental drawback of conventional patch-based sparse representation methods by modeling shift invariance for consistent approximation of local neighbors on whole images. Instead of decomposing the vector as the multiplication of dictionary atoms and the coded coefficients, CSC provides a more elegant way to model local interactions. That is, by representing an image as the summation of convolutions of the sparsely distributed feature maps and the corresponding filters. Concretely, given an $m \times n$ image $\mathbf{x}$ in vector form, the problem of learning a set of vectorized filters for sparse feature maps is solved by minimizing the objective function that combines the convolutional least-squares term and the l_1-norm penalty on the representations:

$$\arg\min_{\mathbf{f},\mathbf{s}} \frac{1}{2} \left\| \mathbf{x} - \sum_{k=1}^{K} \mathbf{f}_k * \mathbf{s}_k \right\|_2^2 + \lambda \sum_{k=1}^{K} \|\mathbf{s}_k\|_1 \quad s.t.\ \|\mathbf{f}_k\|_2^2 \leq 1 \ \ \forall k = \{1, ..., K\}, \tag{1}$$

where $\mathbf{f}_k \in \mathbf{f} = \left[\mathbf{f}_1^T, ..., \mathbf{f}_K^T\right]^T$ is the k-th $d \times d$ filter, $*$ denotes the 2D convolution operator, $\mathbf{z}_k \in \mathbf{z} = \left[\mathbf{z}_1^T, ..., \mathbf{z}_K^T\right]^T$ refers to the sparse feature map corresponding to $\mathbf{f}_k$ with size $(m + d - 1) \times (n + d - 1)$ to approximate $\mathbf{x}$, and λ is a regularization parameter. The problem in Eq. (1) can be efficiently and explicitly solved in the Fourier domain, derived within an Alternating Direction Method of Multipliers (ADMM) framework [6].

Dual Learning (DL) [7] is a new learning paradigm that translates the input model by forming a closed loop between source and target domains to generate informative feedbacks. Specifically, for any dual tasks (e.g., $A \leftrightarrow B$) DL strategy appoints $A \rightarrow B$ as the primary task and the other $A \leftarrow B$ as the dual task, and forces them learning from each other to produce the pseudo-input A'. It can achieve the comparable performance through iteratively updating and minimizing the reconstruction error $A - A'$ that helps maximize the use of data. Therefore, making the learning-based methods have less dependent on the large number of training data.

Problem Formulation: The cross-modality image synthesis problem can be formulated as: given an 3D image $\mathbf{X}$ of modality $\mathcal{M}_1$, the task is to infer from $\mathbf{X}$ a target 3D image $\mathbf{Y}$ that approximates to the ground truth of modality $\mathcal{M}_2$. Let $\mathcal{X} = [\mathbf{X}_1, ..., \mathbf{X}_C] \in \mathbb{R}^{m \times n \times z \times C}$ be a set of images of modality $\mathcal{M}_1$ in the source domain, and $\mathcal{Y} = [\mathbf{Y}_2, ..., \mathbf{Y}_C] \in \mathbb{R}^{m \times n \times z \times C}$ be a set of images of modality $\mathcal{M}_2$ in the target domain. m, n are the dimensions of axial view of the image, and z denotes the size of image along the z-axis, while C is the numbers of elements in the training sets. Each pair of $\{\mathbf{X}_i, \mathbf{Y}_i\}$ $\forall i = \{1, ..., C\}$ are registered. To bridge image appearances across different modalities while preserving the

intrinsic local interactions (i.e., intra-domain consistency), we propose a method based on CSC to jointly learn a pair of filters $\mathbf{F}^x$ and $\mathbf{F}^Y$. Moreover, inspired by the DL strategy, we form a closed loop between both domains and assume that there exists a primal mapping function $\mathcal{F}(\cdot)$ from $\mathcal{X}$ to $\mathcal{Y}$ for relating and predicting from one another. We also assume there exists a dual mapping function $\mathcal{G}(\cdot)$ from $\mathcal{Y}$ to $\mathcal{X}$ to generate feedbacks for model self-optimization. Experimentally, we investigate human brain MRI and apply our method to two cross-modality synthesis tasks, i.e., image SR and CMS. An overview of our method is depicted in Fig. 1. **Notation**: Matrices and 3D images are written in bold uppercase (e.g., image $\mathbf{X}$), vectors and vectorized 2D images in bold lowercase (e.g., filter $\mathbf{f}$) and scalars in lowercase (e.g., element k).

2.2 Dual Convolutional Filter Learning

Inspired by CSC (cf. Sect. 2.1) and the benefits of conventional coupled sparsity, we propose a dual convolutional filter learning (DOTE) model, which extends the original CSC formulation into a DL strategy and joint representation into a unified framework. More specifically, given $\mathcal{X}$ together with the corresponding $\mathcal{Y}$ for training, in order to facilitate a joint mapping, we associate the sparse feature maps of each registered data pair $\{\mathbf{X}_i, \mathbf{Y}_i\}_{i=1}^{C}$ by constructing a forward mapping function $\mathcal{F} : \mathcal{X} \mapsto \mathcal{Y}$ with $\mathbf{Y} = \mathcal{F}(\mathbf{X})$. Since such cross-modality synthesis problem satisfies a dual-learning mechanism, we further leverage the duality of the bidirectional transformation between the two domains. That is, by establishing a dual mapping function $\mathcal{G} : \mathcal{Y} \mapsto \mathcal{X}$ with $\mathbf{Y} = \mathcal{G}(\mathbf{X})$. Incorporating feature maps representing and the above closed-loop mapping functions, we can thus derive the following objective function:

$$\begin{aligned}
&\arg \min_{\mathbf{F}^x, \mathbf{F}^y, \mathbf{S}^x, \mathbf{S}^y, \mathbf{W}} \frac{1}{2} \left\| \mathbf{X} - \sum_{k=1}^{K} \mathbf{F}_k^x * \mathbf{S}_k^x \right\|_F^2 + \frac{1}{2} \left\| \mathbf{Y} - \sum_{k=1}^{K} \mathbf{F}_k^y * \mathbf{S}_k^y \right\|_F^2 + \gamma \sum_{k=1}^{K} \|\mathbf{W}_k\|_F^2 \\
&+ \lambda \left(\sum_{k=1}^{K} \|\mathbf{S}_k^x\|_1 + \sum_{k=1}^{K} \|\mathbf{S}_k^y\|_1 \right) + \beta \left(\sum_{k=1}^{K} \|\mathbf{S}_k^y - \mathbf{W}_k \mathbf{S}_k^x\|_F^2 + \sum_{k=1}^{K} \left\|\mathbf{S}_k^x - \mathbf{W}_k^{-1} \mathbf{S}_k^y\right\|_F^2 \right) \\
&\qquad s.t.\ \|\mathbf{f}_k^x\|_2^2 \leq 1 \ \ \|\mathbf{f}_k^y\|_2^2 \leq 1 \ \ \forall k = \{1, ..., K\}.
\end{aligned} \tag{2}$$

where $\mathbf{S}_k^x$ and $\mathbf{S}_k^y$ take the role of the k-th sparse feature maps that approximate data $\mathbf{X}$ and $\mathbf{Y}$ when convolved with the k-th filters $\mathbf{F}_k^x$ and $\mathbf{F}_k^y$ of a fixed spatial support, $k = 1, ..., K$. $\|\cdot\|_F$ is a Frobenius norm chosen to induce the convolutional least squares approximation, and $*$ is represented as a 3D convolution operator, while λ, β, γ are the regularization parameters. Particularly, dual mapping functions $\mathcal{F}(\mathbf{S}_k^x, \mathbf{W}_k) = \mathbf{W}_k \mathbf{S}_k^x$ and $\mathcal{G}\left(\mathbf{S}_k^y, \mathbf{W}_k^{-1}\right) = \mathbf{W}_k^{-1} \mathbf{S}_k^y$ are used to relate the sparse feature maps of $\mathbf{X}$ and $\mathbf{Y}$ over $\mathbf{F}^x$ and $\mathbf{F}^y$. They are done by solving two sets of least squares terms (i.e., $\sum_{k=1}^{K} (\|\mathbf{S}_k^y - \mathbf{W}_k \mathbf{S}_k^x\|_F^2 + \sum_{k=1}^{K} \left\|\mathbf{S}_k^x - \mathbf{W}_k^{-1} \mathbf{S}_k^y\right\|_F^2$) with respect to the linear projections.

2.3 Optimization

Similar to classical dictionary learning methods, the objective function in Eq. (2) is not simultaneously convex with respect to the learned filter pairs, the sparse feature maps and the mapping. Instead, we divide the proposed method into three sub-problems: learning $\mathbf{S}^x$, $\mathbf{S}^y$, training $\mathbf{F}^x$, $\mathbf{F}^y$, and updating $\mathbf{W}$.

Computing sparse feature maps: We first initialize the filters $\mathbf{F}^x$, $\mathbf{F}^y$ as two random matrices and the mapping $\mathbf{W}$ as an identity matrix, then fix them for calculating the solutions of sparse feature maps $\mathbf{S}^x$, $\mathbf{S}^y$. As a result, the problem of Eq. (2) can be converted into two optimization sub-problems. Unfortunately, this cannot be solved under l_1 penalty without breaking rotation invariance. The resulting alternating algorithms [6] by introducing two auxiliary variables $\mathbf{U}$ and $\mathbf{V}$ enforce the constraint inherent in the splitting. In this paper, we follow [6] and solve the convolution subproblems in the Fourier domain within an ADMM optimization strategy:

$$\begin{aligned}
&\min_{\mathbf{S}^x} \frac{1}{2}\left\|\hat{\mathbf{X}} - \sum_{k=1}^{K} \hat{\mathbf{F}}_k^x \odot \hat{\mathbf{S}}_k^x\right\|_F^2 + \lambda \sum_{k=1}^{K} \|\mathbf{U}_k^x\|_1 + \beta \sum_{k=1}^{K} \left\|\hat{\mathbf{S}}_k^y - \mathbf{W}_k \hat{\mathbf{S}}_k^x\right\|_F^2 \\
&\qquad s.t.\ \|\mathbf{V}_k^x\|_2^2 \leq 1,\ \mathbf{V}_k^x = \mathbf{V}\mathbf{\Phi}^T \hat{\mathbf{F}}_k^x,\ \mathbf{U}_k^x = \mathbf{S}_k^x\ \ \forall k = \{1, ..., K\}, \\
&\min_{\mathbf{S}^y} \frac{1}{2}\left\|\hat{\mathbf{Y}} - \sum_{k=1}^{K} \hat{\mathbf{F}}_k^y \odot \hat{\mathbf{S}}_k^y\right\|_F^2 + \lambda \sum_{k=1}^{K} \|\mathbf{U}_k^y\|_1 + \beta \sum_{k=1}^{K} \left\|\hat{\mathbf{S}}_k^x - \mathbf{W}_k^{-1} \hat{\mathbf{S}}_k^y\right\|_F^2 \\
&\qquad s.t.\ \|\mathbf{V}_k^y\|_2^2 \leq 1,\ \mathbf{V}_k^y = \mathbf{V}\mathbf{\Phi}^T \hat{\mathbf{F}}_k^y,\ \mathbf{U}_k^y = \mathbf{S}_k^y\ \ \forall k = \{1, ..., K\},
\end{aligned} \tag{3}$$

where ˆ applied to any symbol denotes the frequency representations (i.e., Discrete Fourier Transform (DFT)). For instance, $\hat{\mathbf{X}} \leftarrow f(\mathbf{X})$ where $f(\cdot)$ is the Fourier transform operator. $\odot$ represents the component-wise product. $\mathbf{\Phi}^T$ is the inverse DFT matrix, and $\mathbf{V}$ projects a filter onto the small spatial support. The auxiliary variables $\mathbf{U}_k^x$, $\mathbf{U}_k^y$, $\mathbf{V}_k^x$ and $\mathbf{V}_k^y$ relax each of the CSC problems under dual mapping constraint by leading to several subproblem decompositions.

Learning convolutional filters: Like when solving for sparse feature maps, filter pairs can be learned similarly by setting $\mathbf{S}^x$, $\mathbf{S}^y$ and $\mathbf{W}$ fixed, and then learning $\mathbf{F}^x$ and $\mathbf{F}^y$ by minimizing

$$\begin{aligned}
&\min_{\mathbf{F}^x, \mathbf{F}^y} \frac{1}{2}\left\|\hat{\mathbf{X}} - \sum_{k=1}^{K} \hat{\mathbf{F}}_k^x \odot \hat{\mathbf{S}}_k^x\right\|_F^2 + \frac{1}{2}\left\|\hat{\mathbf{Y}} - \sum_{k=1}^{K} \hat{\mathbf{F}}_k^y \odot \hat{\mathbf{S}}_k^y\right\|_F^2 \\
&\qquad s.t.\ \|\mathbf{f}_k^x\|_2^2 \leq 1, \|\mathbf{f}_k^y\|_2^2 \leq 1\ \ \forall k = \{1, ..., K\},
\end{aligned} \tag{4}$$

Equation (4) can be solved by a one-by-one update strategy through an augmented Lagrangian method [6].

Updating mapping: With fixed $\mathbf{F}^x$, $\mathbf{F}^y$, $\mathbf{S}^x$ and $\mathbf{S}^y$, we solve the following ridge regression problem for updating mapping $\mathbf{W}$:

$$\min_{\mathbf{W}} \sum_{k=1}^{K} \|\mathbf{S}_k^y - \mathbf{W}_k \mathbf{S}_k^x\|_F^2 + \left\|\mathbf{S}_k^x - \mathbf{W}_k^{-1} \mathbf{S}_k^y\right\|_F^2 + \left(\frac{\gamma}{\beta}\right) \sum_{k=1}^{K} \|\mathbf{W}_k\|_F^2 . \tag{5}$$

Particularly, the primal mapping function $\|\mathbf{S}_k^y - \mathbf{W}_k \mathbf{S}_k^x\|_F^2$ constructs an intrinsic mapping while the corresponding dual mapping function $\left\|\mathbf{S}_k^x - \mathbf{W}_k^{-1} \mathbf{S}_k^y\right\|_F^2$ is utilized to give feedbacks and further optimize the relationship between $\mathbf{S}_k^x$ and $\mathbf{S}_k^y$. Ideally (as the final solution), $\mathbf{S}_k^y = \mathbf{W}_k \mathbf{S}_k^x$, such that the problem in Eq. (5) is reduced to $\min_{\mathbf{W}_k} \sum_{k=1}^{K} \|\mathbf{S}_k^y - \mathbf{W}_k \mathbf{S}_k^x\|_F^2 + \left(\frac{\gamma}{\beta}\right) \sum_{k=1}^{K} \|\mathbf{W}_k\|_F^2$ with the solution $\mathbf{W} = \mathbf{S}_k^y \mathbf{S}_k^{x^T} (\mathbf{S}_k^x \mathbf{S}_k^{x^T} + \frac{\gamma}{\beta} \mathbf{I})^{-1}$, where $\mathbf{I}$ is an identity matrix. We summarize the proposed DOTE method in the following Algorithm 1.

Algorithm 1. DOTE algorithm

Input: Training data $\mathbf{X}$ and $\mathbf{Y}$, parameters λ, γ, σ.
Initialize $\mathbf{F}_0^x$, $\mathbf{F}_0^y$, $\mathbf{S}_0^x$, $\mathbf{S}_0^y$, $\mathbf{W}_0$, $\mathbf{U}_0^x$, $\mathbf{U}_0^y$, $\mathbf{V}_0^x$, $\mathbf{V}_0^y$.
Perform FFT $\mathbf{S}_0^x \rightarrow \hat{\mathbf{S}}_0^x$, $\mathbf{S}_0^y \rightarrow \hat{\mathbf{S}}_0^y$, $\mathbf{F}_0^x \rightarrow \hat{\mathbf{F}}_0^x$, $\mathbf{F}_0^y \rightarrow \hat{\mathbf{F}}_0^y$, $\mathbf{U}_0^x \rightarrow \hat{\mathbf{U}}_0^x$, $\mathbf{U}_0^y \rightarrow \hat{\mathbf{U}}_0^y$, $\mathbf{V}_0^x \rightarrow \hat{\mathbf{V}}_0^x$, $\mathbf{V}_0^y \rightarrow \hat{\mathbf{V}}_0^y$.
Let $\hat{\mathbf{S}}_0^y \leftarrow \mathbf{W}\hat{\mathbf{S}}_0^x$.
while *not converged* **do**
  Solve for $\hat{\mathbf{S}}_{k+1}^x$, $\hat{\mathbf{S}}_{k+1}^y$, $\hat{\mathbf{U}}_{k+1}^x$ and $\hat{\mathbf{U}}_{k+1}^y$ using (3) with fixed filters and $\mathbf{W}_k$.
  Train $\hat{\mathbf{F}}_{k+1}^x$, $\hat{\mathbf{F}}_{k+1}^y$, $\hat{\mathbf{V}}_{k+1}^x$ and $\hat{\mathbf{V}}_{k+1}^y$ by (4) with fixed feature maps and $\mathbf{W}_k$.
  Update $\mathbf{W}_{k+1}$ by (5).
  Inverse FFT $\hat{\mathbf{F}}_{k+1}^x \rightarrow \mathbf{F}_{k+1}^x$, $\hat{\mathbf{F}}_{k+1}^y \rightarrow \mathbf{F}_{k+1}^y$.
end
Output: $\mathbf{F}^x$, $\mathbf{F}^y$, $\mathbf{W}$.

2.4 Synthesis

Once the optimization is completed, we can obtain the learned filters $\mathbf{F}^x$, $\mathbf{F}^y$ and the mapping $\mathbf{W}$. We then apply the proposed model to synthesize images across different modalities (i.e., LR $\rightarrow$ HR and $\mathcal{M}_1 \rightarrow \mathcal{M}_2$, respectively). Given a test image $\mathbf{X}^t$, we compute the sparse feature maps $\mathbf{S}^{tx}$ related to $\mathbf{F}^x$ by solving a single CSC problem like Eq. (1): $\mathbf{S}^{tx} = \arg\min_{\mathbf{S}^{tx}} \frac{1}{2} \left\|\mathbf{X}^t - \sum_{k=1}^{K} \mathbf{F}_k^x * \mathbf{S}_k^{tx}\right\|_2^2 + \lambda \sum_{k=1}^{K} \|\mathbf{S}_k^{tx}\|_1$. After that, we can synthesize the target modality image of $\mathbf{X}^t$ by the sum of K target feature maps $\mathbf{S}_k^{ty} = \mathbf{W}\mathbf{S}_k^{tx}$ convolved with $\mathbf{F}_k^y$, i.e., $\mathbf{Y}^t = \sum_{k=1}^{K} \mathbf{F}_k^y \mathbf{S}_k^{ty}$.

3 Experimental Results

Experimental Setup: The proposed DOTE is validated on two datasets: IXI[1] (including 578 $256 \times 256 \times p$ MR healthy subjects) and NAMIC[2] (involving 20 $128 \times 128 \times q$ subjects). In our experiments, we perform 4-fold cross-validation for testing. That is, selecting 144 subjects from IXI and 5 subjects from NAMIC, respectively, as our test data. Following [1], the regularization parameters σ, λ, β, and γ are empirically set to be 1, 0.05, 0.10, 0.15, respectively. The number of filters is set as 800 according to [8]. Convergence towards primal feasible solution is proved in [6] by first converting Eq. (2) into two optimization sub-problems that involve two proxies $\mathbf{U}$, $\mathbf{V}$ and then solving them alternatively. DOTE converges after ca. 10 iterations. For the evaluation criteria, we adopt PSNR and SSIM indices to objectively assess the quality of our results.

MRI Super-Resolution. As we introduced in Sect. 1, we first address image SR as one of cross-modality image synthesis. In this scenario, we investigate the T2-w images of the IXI dataset for evaluating and comparing DOTE with ScSR [1], A+ [2], NLSR [4], Zeyde [5], ANR [9], and CSC-SR [8]. Generally, LR images are generated by down-sampling HR ground-truth images using bicubic interpolation. We perform image SR with scaling factor 2, and show visual results in Fig. 2. The quantitative results are reported in Fig. 3, while the average PSNRs and SSIMs for all 144 test subjects are shown in Table 1. The proposed model achieves the best PSNRs and SSIMs. Moreover, to validate our argument that DL-based self-optimization strategy is beneficial and requires less training data, we compare $\text{DOTE}_{\text{nodual}}$ (removing dual mapping term) and DOTE under different training data size (i.e., $\frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ of the original dataset). The results are listed in Table 2. From Table 2, we see that DOTE is always better than $\text{DOTE}_{\text{nodual}}$ especially with few training samples.

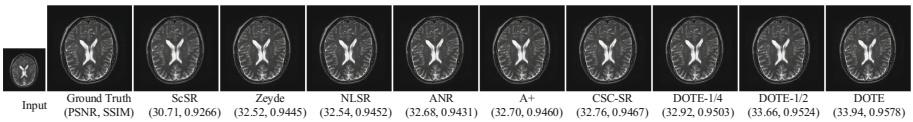

Fig. 2. Example SR results and the corresponding PSNRs and SSIMs.

Cross-Modality Synthesis. For the problem of CMS, we evaluate DOTE and the relevant algorithms on both datasets involving six groups of experiments: (1) synthesizing T2-w image from PD-w acquisition and (2) *vice versa*; (3) generating T1-w image from T2-w input, and (4) *vice versa*. We conduct (1–2) experiments on the IXI dataset, while (3–4) are explored on the NAMIC dataset. The representative and state-of-the-art CMS methods, including Vemulapalli's

[1] http://brain-development.org/ixi-dataset/.
[2] http://hdl.handle.net/1926/1687.

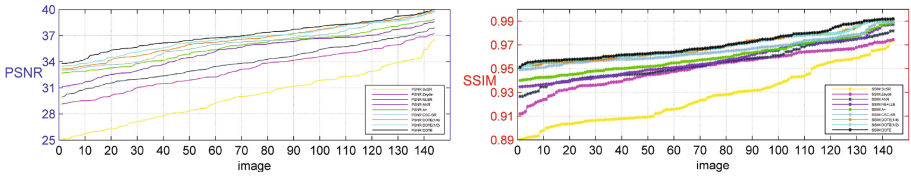

Fig. 3. Error measures of SR results on the IXI dataset.

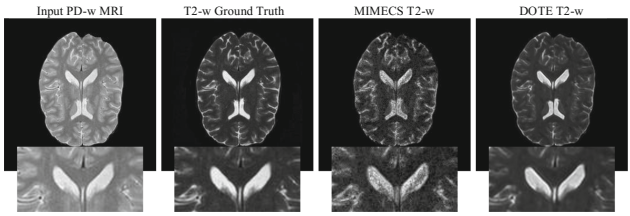

Fig. 4. Visual comparison of synthesized results using MIMECS and DOTE.

Table 1. Quantitative evaluation: DOTE vs. other SR methods.

Avg.	ScSR	Zeyde	NLSR	ANR	A+	CSC-SR	DOTE
PSNR	29.98	33.10	33.97	35.23	35.72	36.18	**37.07**
SSIM	0.9265	0.9502	0.9548	0.9568	0.9600	0.9651	**0.9701**

Table 2. Quantitative evaluation: DOTE vs. $\text{DOTE}_{\text{nodual}}$.

Avg.	$\text{DOTE}_{\text{nodual}}\,\frac{1}{4}$	$\text{DOTE}_{\text{nodual}}\,\frac{1}{2}$	$\text{DOTE}_{\text{nodual}}\,\frac{3}{4}$	DOTE $\frac{1}{4}$	DOTE $\frac{1}{2}$	DOTE $\frac{3}{4}$
PSNR	31.23	33.17	36.09	**36.56**	**36.68**	**37.07**
SSIM	0.9354	0.9523	0.9581	**0.9687**	**0.9690**	**0.9701**

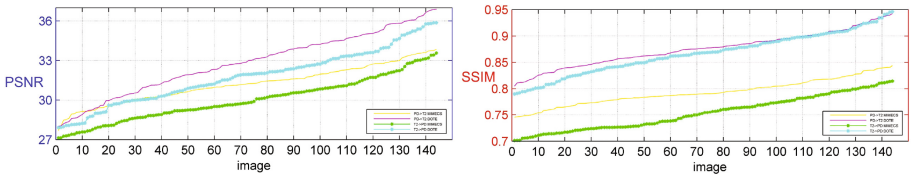

Fig. 5. CMS results: DOTE vs. MIMECS on the IXI dataset.

method [3] and MIMECS [10] are employed to compare with our DOTE approach. We demonstrate visual and quantitative results in Figs. 4, 5 and Table 3, respectively. Our algorithm yields the best results against MIMECS and Vemulapalli for two datasets validating our claim of being able to synthesize better results through the expanded dual optimization.

Table 3. CMS results: DOTE vs. other synthesis methods on the NAMIC dataset.

Metric(avg.)	NAMIC					
	T1 − > T2			T2 − > T1		
	MIMECS	Vemulapalli	DOTE	MIMECS	Vemulapalli	DOTE
PSNR	24.98	27.22	**29.83**	27.13	28.95	**32.03**
SSIM	0.8821	0.8981	**0.9013**	0.9198	0.9273	**0.9301**

4 Conclusion

We presented a dual convolutional filter learning (DOTE) method which directly decomposes the whole image based on CSC, such that local neighbors are preserved consistently. The proposed dual mapping functions integrated with joint learning model form a closed loop that leverages the training data more efficiently and keeps a very stable mapping between image modalities. We applied DOTE to both image SR and CMS problems. Extensive results showed that our method outperforms other state-of-the-art approaches. Future work could concentrate on extending DOTE to higher-order imaging modalities like diffusion tensor MRI and to other modalities beyond MRI.

Acknowledgments. This work has been partially supported by the European Commission FP7 project VPH-DARE@IT (FP7-ICT-2011-9-601055).

References

1. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE TIP **19**(11), 2861–2873 (2010)
2. Timofte, R., De Smet, V., Van Gool, L.: A+: adjusted anchored neighborhood regression for fast super-resolution. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9006, pp. 111–126. Springer, Cham (2015). doi:10.1007/978-3-319-16817-3_8
3. Vemulapalli, R., Van Nguyen, H., Zhou, S.K.: Unsupervised cross-modal synthesis of subject-specific scans. In: IEEE ICCV, pp. 630–638 (2015)
4. Rousseau, F.: Alzheimer disease neuroimaging initiative: a non-local approach for image super-resolution using intermodality priors. MIA **14**(4), 594–605 (2010)
5. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Boissonnat, J.-D., Chenin, P., Cohen, A., Gout, C., Lyche, T., Mazure, M.-L., Schumaker, L. (eds.) Curves and Surfaces 2010. LNCS, vol. 6920, pp. 711–730. Springer, Heidelberg (2012). doi:10.1007/978-3-642-27413-8_47
6. Bristow, H., Eriksson, A., Lucey, S.: Fast convolutional sparse coding. In: IEEE CVPR, pp. 391–398 (2013)
7. He, D., Xia, Y., Qin, T., Wang, L., Yu, N., Liu, T., Ma, W.: Dual learning for machine translation. In: NIPS, pp. 820–828 (2016)

8. Gu, S., Zuo, W., Xie, Q., Meng, D., Feng, X., Zhang, L.: Convolutional sparse coding for image super-resolution. In: IEEE ICCV, pp. 1823–1831 (2015)
9. Timofte, R., De Smet, V., Van Gool, L.: Anchored neighborhood regression for fast example-based super-resolution. In: IEEE ICCV, pp. 1920–1927 (2013)
10. Roy, S., Carass, A., Prince, J.L.: Magnetic resonance image example-based contrast synthesis. IEEE TMI **32**(12), 2348–2363 (2013)