Generalized Brain Image Synthesis with Transferable Convolutional Sparse Coding Networks

Yawen Huang[1], Feng Zheng[✉2], Xu Sun[1], Yuexiang Li[1], Ling Shao[3], and Yefeng Zheng[✉1]

[1] Tencent Jarvis Lab, Shenzhen, China
[2] Southern University of Science and Technology, Shenzhen, China
[3] Terminus Group, China
{yawenhuang, vicyxli, yefengzheng}@tencent.com, f.zheng@ieee.org, pamixsun@foxmail.com, ling.shao@ieee.org

Abstract. High inter-equipment variability and expensive examination costs of brain imaging remain key challenges in leveraging the heterogeneous scans effectively. Despite rapid growth in image-to-image translation with deep learning models, the target brain data may not always be achievable due to the specific attributes of brain imaging. In this paper, we present a novel generalized brain image synthesis method, powered by our transferable convolutional sparse coding networks, to address the lack of interpretable cross-modal medical image representation learning. The proposed approach masters the ability to imitate the machine-like anatomically meaningful imaging by translating features directly under a series of mathematical processings, leading to the reduced domain discrepancy while enhancing model transferability. Specifically, we first embed the globally normalized features into a domain discrepancy metric to learn the domain-invariant representations, then optimally preserve domain-specific geometrical property to reflect the intrinsic graph structures, and further penalize their subspace mismatching to reduce the generalization error. The overall framework is cast in a minimax setting, and the extensive experiments show that the proposed method yields state-of-the-art results on multiple datasets.

Keywords: Convolutional sparse coding networks, image synthesis

1 Introduction

Neuroimaging techniques like magnetic resonance imaging (MRI) allow assessment of varying physical and chemical tissue properties of brain. Different pulse sequences, used in anatomical MRI, provide diverse and complementary information about the anatomical organization. However, a certain single imaging modality is relatively common in real clinical practice, due to the high inter-equipment variability, expensive examination costs and long acquisition time of multi-modality imaging. The proliferation of multi-modal imaging is urgently

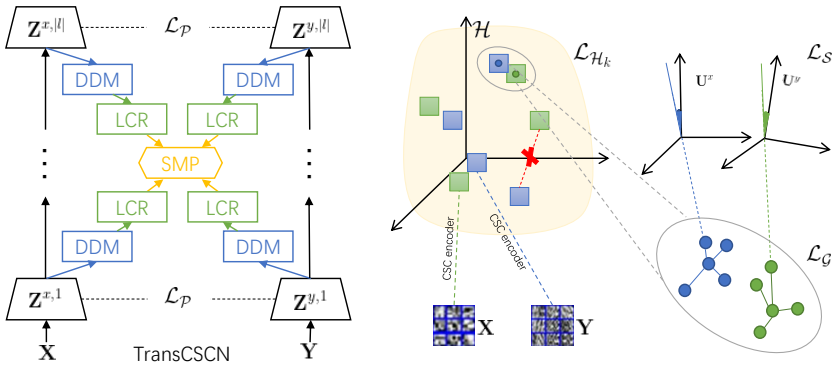

Fig. 1. Architecture of our TransCSCN. The first layer is the globally normalized CSC layer. DDM denotes a module calculating domain discrepancy metric $\mathcal{L}_{\mathcal{H}_k}$, the LCR module enforces Laplacian co-regularization $\mathcal{L}_{\mathcal{G}}$, and SMP enforces subspace mismatch penalization $\mathcal{L}_{\mathcal{S}}$. $\mathcal{L}_{\mathcal{P}}$ is the association loss. The right side shows the feature-leveled operation of each regularizer.

needed for encouraging the comprehensive analysis and making accurate decisions.

Over the last decade, image synthesis technique has enabled transformational advances in various tasks, delivering superior performance on image-to-image translation ubiquitously [13,42]. These methods have also been widely used for medical image analysis [21], including cross-modal MRI synthesis [13], multi-modal image segmentation [26], registration [2], and tracking of anatomical structures [30]. Sparse-representation-based methods [35,34], as an early and trustworthy way, construct a linear function for mapping sparse codes and learning dictionaries jointly. The solution for such a celebrated model can be approximated using greedy algorithms but later known to be sub-optimal, because of highly redundant structure and damaged consistency. Convolutional sparse coding (CSC) [5,12] breaks this dilemma via modeling a shift invariant objective to obtain the coherent and compact representations via convolution. In addition to dictionary-based approaches, deep neural networks [38] have made rapid progress in image generation. The remarkable works are distributed in various applications such as style transfer [42] and sketch-to-photorealism generation [18].

Image synthesis algorithms indeed have achieved promising performance. However, for medical imaging, irrespective of the intrinsic anatomical meaning, requiring a large-scale standardized dataset, at the cost of a large number of parameters or computational complexity, is unacceptable for auxiliary clinical diagnosis and advanced research analysis. Specifically, early methods [7,31] seem to favor a shallow and redundant architecture with descriptors which cannot effectively capture image features, *e.g.*, by finding edges and pooling them. The architecture of recent networks makes feature extraction deeper by imposing a large amount of data and memory overhead in the implementation. However,

collecting multi-modal medical images can be prohibitively hard or even implausible. The other issue is that most medical image synthesis works pursue the superficial consistency, omitting the underlying tissue information. Besides technical challenges, the complexity and heterogeneity of MRI remains a problem in leveraging the heterogeneous scans effectively, for example, imaging by different manufacturers (*e.g.*, Philips Achieva System vs. GE SIGNA system) and abundant sequences (*e.g.*, a turbo spin echo sequence vs. a single-shot EPI sequence). Taken together, a macro perspective expects that these weaknesses can be relieved by a compensatory solution, *i.e.*, constructing a new framework towards standardizing and expanding the synthesis reality for both visual and anatomical significance.

In view of the above challenges, we propose a novel Transferable Convolutional Sparse Coding Network (TransCSCN) that enables the learner to adapt to the target modal. This is done by mapping a latent space to generalize both intra-domain (*i.e.*, multiple imaging manufacturers for one modal) and cross-domains (*i.e.*, multiple modalities) while preserving the domain-specific geometries and their sub-manifolds. An overview of our TransCSCN is shown in Fig. 1. To summarize, this paper makes the following contributions:

- We propose a novel framework, *i.e.*, TransCSCN, for unsupervised brain image synthesis, where multiple objective-specific layers are adapted, resulting in mathematically interpretable formulations and anatomically meaningful results.
- A domain discrepancy metric is provided to embed the globally normalized features in the reproducing kernel Hilbert space to reduce the variant representations of similar tissues in different domains.
- The Laplacian co-regularization term is further devised to optimally preserve the geometric structures underlying the respective domains.
- Finally, a subspace mismatch regularizer is proposed to penalize the generalization error and variation.

2 Related Work

2.1 Domain Adaptation

Domain discrepancy severely degrades the model performance on cross-domain tasks. Luckily, significant effort has been devoted in the literature to provide adapted features or classifiers to new visual domains. Previous methods have tried to learn domain-invariant representations between source and target domains. Of these methods, Zhong *et al.* [40] proposed a transfer cross-validation method, which generalizes a learner across different domains by considering both marginal and conditional distributions. Qiu *et al.* [29] presented a function learning framework by adapting dictionaries learned from one visual domain to the other for smoothly varying domains utilizing regression. Recent work has focused on transferring deep neural network representations from a source dataset to a target domain where the labeled data may be sparse or non-existed. Deep

adaptation network [23] explores feature transferability of deep CNNs in the task-specific layers embedded in a reproducing kernel Hilbert space to reduce the domain discrepancy. In [36], a curriculum manager was proposed as an independent network module to predict the transferability of source domain data and adversarially raise the error rate of a domain discriminator. Yu *et al.* [22] presented dynamic transfer by adapting model parameters to samples to address the domain conflict problem. While many domain adaptation or transformation algorithms for natural images are well explored by minimizing the distribution discrepancy, some disconnections still form non-negligible gaps between the natural and medical images.

2.2 Image-to-Image Translation

Image-to-image translation aims to transfer a source image into the style of a varying reference image. Conventional wisdom and early research [7] tackled this problem using nonparametric settings to resample the feature statistics of a given image texture. Roy *et al.* [31] provided a dictionary-learning-based brain image contrast synthesis approach by assuming that cross-modality patches have similar local geometry to linearly approximate the target image. Vemulapalli *et al.* [33] relaxed the supervision of fully paired data, by jointly maximizing both global mutual information and local spatial consistency to match the similarities across modalities. To circumvent the problem of lacking diversity and good quality, deep generative network was proposed after the introduction of neural style transfer algorithms. The popular works such as CycleGAN [42] are able to transfer rich local texture appearance cross domains, *e.g.*, translating between paintings and photographs. MSGAN [24] designed a mode seeking regularization term for conditional GANs to handle the mode collapse issue. Huang *et al.* [15] relaxed the supervision by matching similarities of both intra- and intre-modal data in feature-level, and then adopting the manifold penalization to handle the brain image synthesis problem. Mainstream image-to-image translation methods are tailored to adapting given modality and target modality; however, these methods have difficulties in modeling complex patterns of irregular distributions with heterogeneous variations.

3 Preliminaries

A natural way to cast the problem of learning a shallow architecture of shift-invariant representations into an optimization problem is a convolutional sparse coding (CSC) method [5]. CSC has gained popularity in computer vision and medical imaging, because of its ability to obtain a structured filter that facilitates a global handling of the image. CSC is remarkable when compared with traditional sparse coding, providing a more elegant way to represent data as the sum of filters convolved with sparsely distributed codes.

Given a set of observations $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_S\}$ in $\mathbb{R}^N$, CSC can be formulated as learning a set of sparse coefficient feature maps $\mathbf{z}_i \in \mathbb{R}^N$ convolved with filters

$\mathbf{f}_i \in \mathbb{R}^M$, $\forall i = \{1, ..., K\}$. Its optimization problem boils down to:

$$\min_{\mathbf{f},\mathbf{z}} \frac{1}{2} \left\| \mathbf{x} - \sum_{i=1}^{K} \mathbf{f}_i * \mathbf{z}_i \right\|_2^2 + \lambda \sum_{i=1}^{K} \|\mathbf{z}_i\|_1 \;\; s.t. \; \|\mathbf{f}_i\|_2^2 \leq 1 \; \forall i = \{1, ..., K\}, \quad (1)$$

where $*$ denotes the 2D convolution operation, λ is the regularization parameter, $\mathbf{x}$ and $\mathbf{z}$ are the vectorized images, and $\mathbf{f}$ is the vectorized filter. The objective in Eq. (1) is difficult to optimize due to the convolutional decomposition mechanism. Motivated by Parseval's theorem and deconvolutional networks, Zeiler *et al.* [37] demonstrated that through an alternation strategy to solve a sequence of convex sub-problems until convergence is an efficient way. As a part of the proximal gradient methods, fast iterative shrinkage thresholding algorithm (FISTA) [3] provided an iterative approach for solving the l_1 penalized least squares problem with fast quadratic convergence. In parallel, the augmented Lagrange methods, such as the alternating direction method of multipliers (ADMM) [4], treated the optimization as sub-problems and computed the convolutions in the Fourier domain. The subsequent CSC based algorithms often rely on the ADMM formulation to circumvent the computational burdens of the inversion of a convolutional linear operation. For example, Heide *et al.* [12] exploited the mask matrices to deal with the incomplete samples, while Choudhury *et al.* [6] leveraged the matrix inverse lemma to achieve a global consensus in each of the estimates.

In this study, we consider a special case, where a source domain training set $\mathbf{X} \in \mathbb{R}^{N \times S}$ of S source modality samples and a target domain training set $\mathbf{Y} \in \mathbb{R}^{N \times T}$ of T target modality samples are given. The image synthesis task is then expected to learn both convolutional feature maps $\mathbf{Z}^x$ and $\mathbf{Z}^y$ over their corresponding filters $\mathbf{F}^x$ and $\mathbf{F}^y$, where the superscript is adopted to distinguish the variate from the source domain x or from the target domain y. The conventional solution following the independent scheme in Eq. (1), results in uncorrelated features. The joint representation learning groups two independent reconstruction errors in a single objective function, leading to a common set of feature maps (*i.e.*, $\mathbf{Z}^x \equiv \mathbf{Z}^y$) shared between source and target domains. The flexible joint learning strategy replaces the common feature assumption by constructing a linear projector $\mathbf{P}$ to calculate $\|\mathbf{Z}^x - \mathbf{P}\mathbf{Z}^y\|_2$, which is more reasonable.

4 Transferable Convolutional Sparse Coding Networks

The challenge of joint learning mainly arises when the target domain has no or only limited data pairing with the source domain. In other words, the assumption of the feature maps from one domain to be identical to those observed at the target domain is no longer valid, let alone the abundance of variations in single domain. In this paper, we address the dilemma of generalizability against multivariate nature of neuroimaging, by providing more flexibility in leveraging the large-scale heterogeneous medical data in an unsupervised manner, such that the learned transferable representations can close the source and target discrepancy.

Following the CSC approximation introduced in Sec. 3, a shallow convolutional structure on the learned matrices is constructed for the purpose of low-level feature extraction. Recent works [8,20] suggested to learn multiple levels of feature representations in a hierarchical architecture to deeply capture both low-level and mid-level features. As expected, CSCNets [15] were proposed to exploit the benefits of depth with convolutional filter learning to convey information with increasing austerity. Given $\mathbf{X}$ and $\mathbf{Y}$, the representations of the multilayered CSC can be formalized as $\mathbf{Z}^{x,|l|} = f(\mathbf{X}, \mathbf{F}^{x,|l-1|}, \lambda)$, $\mathbf{Z}^{y,|l|} = f(\mathbf{Y}, \mathbf{F}^{y,|l-1|}, \lambda)$, where $l \in \{1, 2, \ldots, L\}$ denotes the layer index, f is the feature extractor, $\mathbf{Z}^{x,|l|} \in \mathbb{R}^{N^{|l|} \times h^{|l|} w^{|l|}}$ and $\mathbf{Z}^{y,|l|} \in \mathbb{R}^{N^{|l|} \times h^{|l|} w^{|l|}}$ represent the l-th layered feature maps with tensor properties of height h and width w. Correspondingly, the layerwise projector $\mathbf{P}^{|l|}$ is updated as $\mathcal{L}_{\mathcal{P}}(\mathbf{Z}^{x,|l|}, \mathbf{Z}^{y,|l|}) = \|\mathbf{Z}^{x,|l|} - \mathbf{P}^{|l|}\mathbf{Z}^{y,|l|}\|_F^2 + \alpha\|\mathbf{P}^{|l|}\|_F^2$, where α is association mapping parameter. This can be solved as a set of the least squares problem.

4.1 Domain Discrepancy Metric

Despite the obvious cross-domain divergence, the variations such as different manufacturers and physical parameters in single domain are also harmful to model generalization. To approach this problem, we adopt the single domain unit normalization [15] and begin by a global normalization under $\mathbf{Z}^x$ and $\mathbf{Z}^y$. Then the features are scaled as $\frac{\mathbf{Z}^x}{\max(\|\mathbf{Z}^x\|_2)}$ and $\frac{\mathbf{Z}^y}{\max(\|\mathbf{Z}^y\|_2)}$, respectively. When the maximum of their norms is guaranteed to be unity, we project the features to a unit sphere to eliminate the scaling ambiguity globally as follows: $\hat{\mathbf{Z}}_i^x = \mathbf{Z}_i^x / (\max(\|\mathbf{Z}_i^x\|_2) \sqrt{1 - \left\| \frac{\mathbf{Z}_i^x}{\max(\|\mathbf{Z}_i^x\|_2)} \right\|^2}), \forall i \in \mathbb{R}^S$, $\hat{\mathbf{Z}}_j^y = \mathbf{Z}_j^y / (\max(\|\mathbf{Z}_j^y\|_2) \sqrt{1 - \left\| \frac{\mathbf{Z}_j^y}{\max(\|\mathbf{Z}_j^y\|_2)} \right\|^2}), \forall j \in \mathbb{R}^T$, where the general unit normalization criterion $\left\|\hat{\mathbf{Z}}_i^x\right\|_2^2 = 1, \forall i$ and $\left\|\hat{\mathbf{Z}}_j^y\right\|_2^2 = 1, \forall j$ can be satisfied. The globally normalized convolutional feature maps then become $\mathbf{Z}'^{x,|l|} = f(\hat{\mathbf{Z}}^{x,|l-1|}, \mathbf{F}^{x,|l-1|}, \lambda)$, $\mathbf{Z}'^{y,|l|} = f(\hat{\mathbf{Z}}^{y,|l-1|}, \mathbf{F}^{y,|l-1|}, \lambda)$, where the imposed upper layer of the representation $\hat{\mathbf{Z}}^{x,|l-1|}$ and $\hat{\mathbf{Z}}^{y,|l-1|}$ are treated as the intermediate representations.

The problem of adapting the source domain data to the target domain has been explored [23,14]. Of these methods, bounding the target error by superimposing a discrepancy metric between both domains is a direction to explore, which can be realized by the two-sample test statistics. Theoretically, given two samples coming from different domains following different probability distributions $p(\mathbf{x})$ and $p(\mathbf{y})$, the two-sample testing either accepts or rejects a null hypothesis $p(\mathbf{x}) = p(\mathbf{y})$, based on various metrics, such as the maximum mean discrepancy (MMD) [10]. This prior has motivated us to solve a natural domain variation in a generalized unsupervised way by learning the correlation-relaxed features of different domains more efficiently. On further consideration, the original MMD is restricted by the local generalization leading to the sub-optimal kernel problem, while the extended multi-kernel MMD (MK-MMD) criterion is more applicable to perform unbiased estimation. Suppose the reproducing kernel

Hilbert space (RKHS) $\mathcal{H}_k$ induced with a characteristic kernel k on the vectorized element $\mathbf{Z}$ has a set of positive definite kernels $\{k_u\}_{u=1}^{d}, \forall u \in \{1, \cdots, d\}$. The MK-MMD then can be defined as the squared distance between kernel mean embeddings in $\mathcal{H}_k$ to minimize the domain gap and optimize the kernel selection,

$$\begin{aligned} &\mathcal{L}_{\mathcal{H}_k}(X, Y) = \left\| \mathbb{E}_{p(\mathbf{x})}[f(\mathbf{X})] - \mathbb{E}_{p(\mathbf{y})}[f(\mathbf{Y})] \right\|_{\mathcal{H}_k}^2, \\ &\forall\ k \in \mathcal{K} := \{ \sum_{u=1}^{d} \beta_u k_u : \sum_{u=1}^{d} \beta_u = 1, \beta_u \geq 0 \}, \end{aligned} \tag{2}$$

where $\mathcal{K}$ is the convex combination of u positive definite kernels $\{k_u\}_{u=1}^{d}, \forall u$; β_u denotes the coefficient for constraining the characteristic of $\{k_u\}$; $f(\cdot)$ represents the feature mapping with $k(X, Y) = \langle f(\mathbf{X}), f(\mathbf{Y}) \rangle_{\mathcal{H}_k}$; and $\mathcal{L}_{\mathcal{H}_k}$ can be interpreted as matching all orders of statistics with a property of $p(\mathbf{x}) = p(\mathbf{y})$ iff $\mathcal{L}_{\mathcal{H}_k}(X, Y) = 0$. As principally studied in MK-MMD, we are targeting to boost unpaired cross-modal data underlying the same distributions to be close to each other. Mathematically, the unsupervised method can be established by adding the MK-MMD-based layerwise regularizer $\mathcal{L}_{\mathcal{H}_k}^{l}$:

$$\begin{aligned} \min_{f} \max_{k} & f(\hat{\mathbf{Z}}^{x,|l-1|}, \mathbf{F}^{x,|l-1|}, \lambda) + f(\hat{\mathbf{Z}}^{y,|l-1|}, \mathbf{F}^{y,|l-1|}, \lambda) \\ & + \gamma \left\| \mathbb{E}_{p(\mathbf{x})}[\mathbf{Z}'^{x,|l|}] - \mathbb{E}_{p(\mathbf{y})}[\mathbf{Z}'^{y,|l|}] \right\|_{\mathcal{H}_k}^2, \end{aligned} \tag{3}$$

where γ denotes the penalty parameter. Considering the kernel trick, $\mathcal{L}_{\mathcal{H}_k}^{l}$ can be expressed as the layered expectation of kernel function $\mathcal{L}_{\mathcal{H}_k}^{l} \triangleq \frac{1}{S^2} \sum_{i=1}^{S} \sum_{j=1}^{S} k(\mathbf{Z}_i'^{x,|l|}, \mathbf{Z}_j'^{x,|l|}) + \frac{1}{T^2} \sum_{i=1}^{T} \sum_{j=1}^{T} k(\mathbf{Z}_i'^{y,|l|}, \mathbf{Z}_j'^{y,|l|}) - \frac{2}{ST} \sum_{i=1}^{S} \sum_{j=1}^{T} k(\mathbf{Z}_i'^{x,|l|}, \mathbf{Z}_j'^{y,|l|})$, where $\mathbf{Z}_i'^{x,|l|}, \mathbf{Z}_j'^{x,|l|} \overset{iid}{\sim} p(\mathbf{x})$, and $\mathbf{Z}_i'^{y,|l|}, \mathbf{Z}_j'^{y,|l|} \overset{iid}{\sim} p(\mathbf{y})$, $k \in \mathcal{K}$, $\forall i, j$.

4.2 Laplacian Co-Regularization

The representations learned in Eq. (3) encourage domain-invariant features against cross-modal distribution discrepancy; however, some important low-level details reflecting the domain-specific information are lost. With this limitation, the synthetic may be visually meaningful but lacking practical significance. Recent advances in exploring manifold assumption [41] reflect the geometric structure leading to a realistic and correct approximation. Based on the observation of graph Laplacian (*a.k.a.* manifold learning), we investigate how to preserve the complementary properties by introducing a Laplacian co-regularizer. To be specific, given $\mathbf{Z}'^{x,|l|}$ and $\mathbf{Z}_i'^{y,|l|}$ of $\mathbf{X}$ and $\mathbf{Y}$, respectively, two layerwise q-nearest neighbor graphs $\mathcal{G}^{x,|l|}$ and $\mathcal{G}^{y,|l|}$ can be constructed while each with g vertices [39]. Under the above definition, the Laplacian co-regularization $\mathcal{L}_{\mathcal{G}}(X, Y)$ is given as:

$$\sum_{i,j=1}^{g} \prod_{l \in L} (\mathbf{W}_{i,j}^{x,|l|} \left\| \mathbf{Z}_i'^{x,|l|} - \mathbf{Z}_j'^{x,|l|} \right\|^2 + \mathbf{W}_{i,j}^{y,|l|} \left\| \mathbf{Z}_i'^{y,|l|} - \mathbf{Z}_j'^{y,|l|} \right\|^2), \tag{4}$$

where $\mathbf{W}_{i,j}^{x,|l|}$ and $\mathbf{W}_{i,j}^{y,|l|}$ are the layered weight matrices of $\mathcal{G}^{x,|l|}$ and $\mathcal{G}^{y,|l|}$ having attributions of $\mathbf{W}_{i,j}^{x,|l|} = 1$, $\mathbf{W}_{i,j}^{y,|l|} = 1$ iff any two features $\mathbf{Z}_i'^{x,|l|}$ and $\mathbf{Z}_j'^{x,|l|}$ or $\mathbf{Z}_i'^{y,|l|}$ and $\mathbf{Z}_j'^{y,|l|}$ satisfying $\mathbf{Z}_i'^{x,|l|}$ or $\mathbf{Z}_i'^{y,|l|}$ is among the g-nearest neighbors of $\mathbf{Z}_j'^{x,|l|}$ or $\mathbf{Z}_j'^{y,|l|}$; otherwise, $\mathbf{W}_{i,j}^{x,|l|} = 0$, $\mathbf{W}_{i,j}^{y,|l|} = 0$.

The domain-specific graph structures are encoded into $\mathbf{W}_{i,j}^{x,|l|}$ and $\mathbf{W}_{i,j}^{y,|l|}$ with the corresponding layerwise diagonal matrices $\mathbf{D}^{x,|l|} = \mathrm{diag}(d_1^{x,|l|}, \cdots, d_g^{x,|l|})$ and $\mathbf{D}^{y,|l|} = \mathrm{diag}(d_1^{y,|l|}, \cdots, d_g^{y,|l|})$. The graph Laplacian provides $\mathcal{G} = \mathbf{D} - \mathbf{W}$, such that we can preserve the domain-specific geometrical structures by Eq. (4) updating as $\mathcal{L}_{\mathcal{G}}(X, Y) = \mathrm{Tr}(\mathbf{Z}'^{x,|l|}\mathcal{G}^{x,|l|}{\mathbf{Z}'^{x,|l|}}^T + \mathbf{Z}'^{y,|l|}\mathcal{G}^{y,|l|}{\mathbf{Z}'^{y,|l|}}^T)$.

4.3 Subspace Mismatch Penalization

Considering the heterogeneity of medical images acquired on scanners from different manufacturers and with different physical parameters, all these properties induce conflicted and inconsistent features. The aforementioned formulations bridge the domain gap and enrich the domain-specific representation, but fail to cope with the variational tissue structures across domains. This means that performance may degrade when high-level features are insensitive to tissue boundaries, resulting in over-smoothness of the synthesis and potential scaling-based mismatching. To reduce the generalization error and better preserve the geometry in our synthesis task, we propose a subspace mismatch regularizer to constrain the veritable similar bases in their subspace. As suggested by [32], singular value decomposition (SVD) of the feature matrix can be exploited to enforce the constraint. In this work, we adopt the general SVD to get the layerwise orthogonal matrices $\mathbf{U}^{x,|l|}$ and $\mathbf{U}^{y,|l|}$: $\mathbf{Z}'^{x,|l|} = \mathbf{U}^{x,|l|}\mathbf{\Sigma}^{x,|l|}{\mathbf{V}^{x,|l|}}^T$, $\mathbf{Z}'^{y,|l|} = \mathbf{U}^{y,|l|}\mathbf{\Sigma}^{y,|l|}{\mathbf{V}^{y,|l|}}^T$. Here, $\mathbf{\Sigma}$ is the nonnegative real diagonal matrix, and $\mathbf{V}^T$ is the conjugate transpose of $\mathbf{V}$ denoting the right singular matrix. Following [32], we use principal angles to measure the subspace distance between two domains,

$$\begin{aligned} \mathbf{\Theta}^{|l|} &= \min_{\mathbf{U}^{x,|l|}, \mathbf{U}^{y,|l|}} \arccos\left(\frac{{\mathbf{U}^{x,|l|}}^T \mathbf{U}^{y,|l|}}{\left\|\mathbf{U}^{x,|l|}\right\| \left\|\mathbf{U}^{y,|l|}\right\|}\right), \\ {\mathbf{U}^{x,|l|}}^T \mathbf{U}^{y,|l|} &= \mathbf{A}^{x,|l|}(\mathrm{diag}(\cos\Theta^{|l|}))\mathbf{A}^{y,|l|^T}, \end{aligned} \tag{5}$$

where $\mathbf{\Theta}$ represents the principal angles and $\mathbf{A}$ is the weight matrix. The orthogonal bases are then matched by $\mathcal{L}_{\mathcal{S}}^{|l|} = \||\mathbf{A}|^{x,|l|} - |\mathbf{A}|^{y,|l|}\|_F^2$ in the feature-leveled subspaces.

4.4 Transfer Representation Learning

In our transfer representation learning, we construct the globally normalized features, the penalization of domain discrepancy, the regularization of domain-specific manifold, and the reduction of subspace mismatch. The overall objective

Algorithm 1 Layerwise **F**-Step Optimization

Input: Training data $\mathbf{X}$, $\mathbf{Y}$, ρ^f
Initialize: $\mathbf{Z}_0^x$, $\mathbf{Z}_0^y$, $\mathbf{F}_0^x \in \mathbb{O}$, $\mathbf{F}_0^y \in \mathbb{O}$
$\mathbf{Z}_0^x \to \hat{\mathbf{Z}}_0^x$, $\mathbf{Z}_0^y \to \hat{\mathbf{Z}}_0^y$
while not converged **do**
  for $i = 1$ to B **do**
    $\arg\min_{\mathbf{F}^x,\mathbf{F}^y} \frac{1}{2}(\|\mathbf{X} - \mathbf{F}^x * \hat{\mathbf{Z}}^x\|_2^2 + \|\mathbf{X} - \mathbf{F}^y * \hat{\mathbf{Z}}^y\|_2^2 + \delta(\|\mathbf{F}^x - \tilde{\mathbf{F}}^x + \rho^{fx}\|_2^2 + \|\mathbf{F}^y - \tilde{\mathbf{F}}^y + \rho^{fy}\|_2^2)), s.t.\ \|\mathbf{f}_i^x\|_2^2 \le 1, \|\mathbf{f}_i^y\|_2^2 \le 1, \forall i$
  end for
  $\arg\min_{\tilde{\mathbf{F}}^x,\tilde{\mathbf{F}}^y} \mathrm{ind}_C(\tilde{\mathbf{F}}^x) + \mathrm{ind}_C(\tilde{\mathbf{F}}^y) + \frac{N\delta}{2}(\|\tilde{\mathbf{F}}^x - \bar{\mathbf{F}}^x - \bar{\rho}^{fx}\|_2^2 + \|\tilde{\mathbf{F}}^y - \bar{\mathbf{F}}^y - \bar{\rho}^{fy}\|_2^2)$
  for $i = 1$ to B **do**
    $\rho^{fx\prime} = \rho^{fx} + \mathbf{F}^x - \tilde{\mathbf{F}}^x, \rho^{fy\prime} = \rho^{fy} + \mathbf{F}^y - \tilde{\mathbf{F}}^y$
  end for
end while
Output: $\mathbf{F}^x$, $\mathbf{F}^y$

function is then represented as follows:

$$\min_{f,\mathcal{L}_\mathcal{P},\mathcal{L}_\mathcal{G},\mathcal{L}_\mathcal{S}} \max_k f(\mathbf{X},\lambda) + f(\mathbf{Y},\lambda) + \mathcal{L}_\mathcal{P} + \gamma\mathcal{L}_{\mathcal{H}_k} + \mathcal{L}_\mathcal{G} + \mathcal{L}_\mathcal{S}. \tag{6}$$

The resulting architecture is named as transferable convolutional sparse coding network (TransCSCN). Once the optimization is completed, we can obtain the trained filters $\mathbf{F}^x$, $\mathbf{F}^y$, convolutional feature maps $\mathbf{Z}^x$, $\mathbf{Z}^y$, and their projection matrices $\mathbf{P}$. The learned model is then applied to synthesize images across modalities. For the given test image $\mathbf{X}^t$, the correlated target modality version can be computed as $\mathbf{Y}^t = \mathbf{F}^y\hat{\mathbf{Z}}^{ty}$ with $\hat{\mathbf{Z}}^{ty} \approx \mathbf{P}\mathbf{Z}^{tx}$, where $\mathbf{Z}^{tx} = f(\mathbf{X}^t,\lambda)$.

4.5 Multilevel Optimization

The general CSCNet is convex in each variable of the i-th layer but not jointly convex. The solutions such as the coordinate descent allow to alternately minimize the objective over one block of the variables. Considering the large size of medical images which places great demands on computational efficiency, following [6,12], we reformulate the objective to an unconstrained optimization by introducing an indicator ind_C defined on the convex set of the constraints C,

$$\min_{f,\mathcal{L}_\mathcal{P},\mathcal{L}_\mathcal{G},\mathcal{L}_\mathcal{S}} \max_k f(\mathbf{X},\lambda)+f(\mathbf{Y},\lambda)+\mathrm{ind}_C(\mathbf{F}^x)+\mathrm{ind}_C(\mathbf{F}^y)+\mathcal{L}_\mathcal{P}+\gamma\mathcal{L}_{\mathcal{H}_k}+\mathcal{L}_\mathcal{G}+\mathcal{L}_\mathcal{S}. \tag{7}$$

Eq. (7) then can be solved efficiently by splitting with respect to the filters $\mathbf{F}$, feature maps $\mathbf{Z}$, and the relationship operator $\mathbf{P}$.

F-Step Subproblem: We first exploit the l-th layer filter learning by solving,

$$\begin{aligned}\arg\min_f f(\hat{\mathbf{Z}}^{x,|l|},\mathbf{F}^{x,|l|}) + f(\hat{\mathbf{Z}}^{y,|l|},\mathbf{F}^{y,|l|}) + \mathrm{ind}_C(\tilde{\mathbf{F}}^{x,|l|}) \\ +\mathrm{ind}_C(\tilde{\mathbf{F}}^{y,|l|}), s.t.\ \|\mathbf{f}_k^{x,|l|}\|_2^2 \le 1, \mathbf{f}_k^{y,|l|}\|_2^2 \le 1, \forall k,\end{aligned} \tag{8}$$

Algorithm 2 Layerwise **Z**-Step Optimization

Input: Training data $\mathbf{X}$, $\mathbf{Y}$, ρ^z, λ, γ
1: Initialize: $\mathbf{Z}_0^x$, $\mathbf{Z}_0^y$, $\mathbf{F}_0^x$, $\mathbf{F}_0^y$, $\mathbf{P}_0$
2: $\mathbf{Z}_0^x \to \hat{\mathbf{Z}}_0^x$, $\mathbf{Z}_0^y \to \hat{\mathbf{Z}}_0^y$
3: Let $\hat{\mathbf{Z}}_0^y \leftarrow \hat{\mathbf{Z}}_0^x \mathbf{P}_0$
4: **while** not converged **do**
5: **for** $i = 1$ to B **do**
6: $\arg\min_{f,\mathcal{L}_\mathcal{P},\mathcal{L}_\mathcal{G},\mathcal{L}_\mathcal{S}} \max_k f(\mathbf{X},\lambda) + f(\mathbf{Y},\lambda) + \mathcal{L}_\mathcal{P} + \gamma\mathcal{L}_{\mathcal{H}_k} + \mathcal{L}_\mathcal{G} + \mathcal{L}_\mathcal{S} + \frac{\delta}{2}(\|\hat{\mathbf{Z}}^x - \tilde{\mathbf{Z}}^x + \rho^{zx}\|_2^2 + \|\hat{\mathbf{Z}}^y - \tilde{\mathbf{Z}}^y + \rho^{zy}\|_2^2)$
7: **end for**
8: $\arg\min_{\tilde{\mathbf{Z}}^x,\tilde{\mathbf{Z}}^y} \|\tilde{\mathbf{Z}}^x\|_1 + \|\tilde{\mathbf{Z}}^Y\|_1 + \frac{N\delta}{2}(\|\tilde{\mathbf{Z}}^x - \bar{\mathbf{Z}}^x - \bar{\rho}^{zx}\|_2^2 + \|\tilde{\mathbf{Z}}^y - \bar{\mathbf{Z}}^y - \bar{\rho}^{zy}\|_2^2)$
9: **for** $i = 1$ to B **do**
10: $\rho^{zx\prime} = \rho^{zx} + \hat{\mathbf{Z}}^x - \tilde{\mathbf{Z}}^x, \rho^{zy\prime} = \rho^{zy} + \hat{\mathbf{Z}}^y - \tilde{\mathbf{Z}}^y$
11: **end for**
12: **end while**
Output: $\mathbf{Z}^x$, $\mathbf{Z}^y$, $\mathbf{P}$

where $\tilde{\mathbf{F}}$ means the shared global variable introduced as the slack variable which is subjected to $\mathbf{F} - \tilde{\mathbf{F}} = 0$. Then the optimization with respect to Eq. (8) can be solved by the ADMM strategy derived from the augmented Lagrangian (Lagrange multiplier ρ) with respect to other variables, yielding Algorithm 1.

Z-Step Subproblem: Alternatively, we optimize the layered convolutional least squares with the corresponding filters and other regularization. The subproblem of learning convolutional sparse feature maps then can be written as:

$$\begin{aligned}\min_{f,\mathcal{L}_\mathcal{P},\mathcal{L}_\mathcal{G},\mathcal{L}_\mathcal{S}} \max_k & f(\hat{\mathbf{Z}}^{x,|l|},\mathbf{F}^{x,|l|},\lambda) + f(\hat{\mathbf{Z}}^{y,|l|},\mathbf{F}^{y,|l|},\lambda) + \||\mathbf{A}|^{x,|l|} - |\mathbf{A}|^{y,|l|}\|_F^2 \\ &+\gamma \left\| \mathbb{E}_{p(\mathbf{x})}[\mathbf{Z}'^{x,|l|}] - \mathbb{E}_{p(\mathbf{y})}[\mathbf{Z}'^{y,|l|}] \right\|_{\mathcal{H}_k}^2 + \|\mathbf{Z}^{x,|l|} - \mathbf{P}^{|l|}\mathbf{Z}^{y,|l|}\|_F^2 \\ &+\alpha\|\mathbf{P}^{|l|}\|_F^2 + \mathrm{Tr}(\mathbf{Z}'^{x,|l|}\mathcal{G}^{x,|l|}{\mathbf{Z}'^{x,|l|}}^T + \mathbf{Z}'^{y,|l|}\mathcal{G}^{y,|l|}{\mathbf{Z}'^{y,|l|}}^T).\end{aligned} \tag{9}$$

Like the subproblem of **F**-step, **Z** can be learned in a similar fashion by taking the form of Tikhonov-regularized least squares [19] and facilitating vector-wise manipulations. Through coordinate descent, we derive the **Z**-step subproblem in Algorithm 2.

P-Step Subproblem: Updating the projection matrix, which is only associated with **P**, can be incorporated into the optimization process as:

$$\arg\min_{\mathbf{P}} \|\mathbf{Z}^{x,|l|} - \mathbf{P}^{|l|}\mathbf{Z}^{y,|l|}\|_F^2 + \alpha\|\mathbf{P}^{|l|}\|_F^2. \tag{10}$$

5 Experiments

5.1 Network Architecture

We take the architecture proposed in [11] as the backbone, and construct a nine-layer TransCSCN constrained by different regularizers. All brain volumes

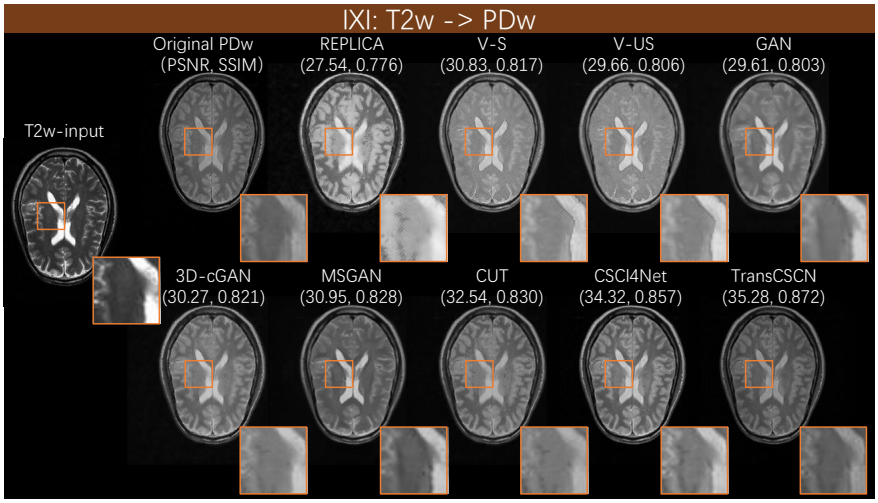

Fig. 2. Visual comparisons of different methods for T2w→PDw on the IXI dataset [1].

are split to 2D slices, and the spatial subsampling operation is fulfilled by our layerwise TransCSCN with a stride of 2 in the last two bottleneck layers, while batch normalization is incorporated after each layer to facilitate the convergence, and the last layer is followed by a global average pooling layer. We train the network for a total of 200 epochs using the Adam solver with a learning rate of 0.0002 and a batch size of 32. The other parameters are set as $\lambda = 0.2$, $\alpha = 0.15$, $\gamma = 1$, and the layered MK-MMD with Gaussian kernels have bandwidths equipped as median pairwise squared distances.

5.2 Experimental Setup

We validate our method on two public multi-modality brain datasets, *viz.* IXI[4] and BraTS[5] databsets, respectively. The IXI dataset involves 578 healthy subjects each imaged using a matrix of 256×256×v (v = 112∼136) scanned from three hospitals (Hammer Smith Hospital, Guy's Hospital, and Institute of Psychiatry) by different Magnetic Resonance Imaging (MRI) systems (Philips and GE). The BraTS dataset, instead, provides multi-modal brain tumor subjects, contributing 225 valid cases. It is worth noting that our experiments are relatively comprehensive since both healthy subjects and pathological data are covered. To be specific, we adopt Proton Density weighted (PDw) and T2w MRI scans (with significant difference) from the IXI dataset, and T1w and Fluid Attenuated Inversion Recovery (FLAIR) acquisitions (with significant difference) from the BraTS dataset. Physically, PDw data recognizes fluid and fat; T2w data reflects intermediate-bright fat and bright fluid; T1w data provides good

[4] https://brain-development.org/ixi-dataset/
[5] https://www.med.upenn.edu/sbia/brats2018/data.html

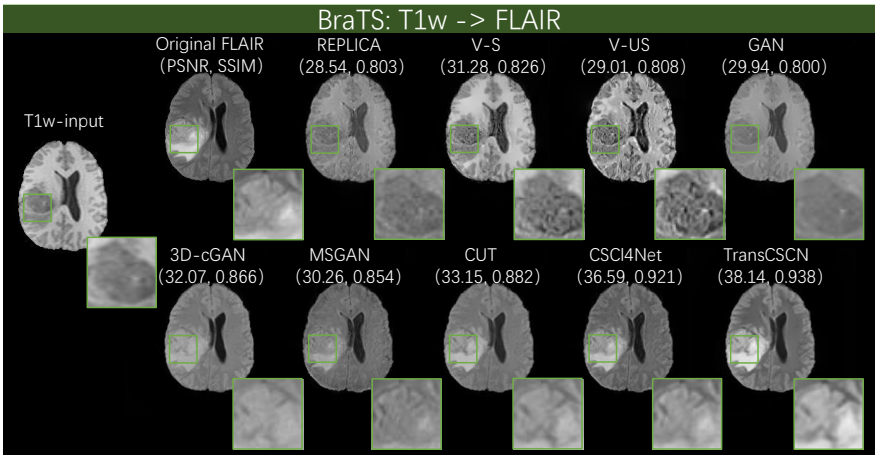

Fig. 3. Visual comparisons of different methods for T1→FLAIR on the BraTS dataset [25].

contrast between Gray Matter (GM) and White Matter (WM); FLAIR data exhibits brighter GM than WM and Cerebrospinal Fluid (CSF) is dark, instead of bright. The conducted evaluations are divided into two parts, resulting in four tasks: (1) generating T2-w images from PD-w acquisitions and *vice versa* on the IXI dataset; (2) synthesizing FLAIR data from T1w images and *vice versa* on the BraTS dataset. We fix the number of test cases, *i.e.*, 80 for the IXI and 45 for the BraTS, respectively, and select 60 samples from the IXI and 20 samples from the BraTS for our validation. We construct the fully unsupervised training data with 219 unpaired PDw & T2w MRI for the IXI and 80 unpaired T2w & FLAIR MRI for the BraTS, respectively, after discarding half of the data pairs. The hyper-parameters of TransCSCN are tuned on our validation set. In addition to the visual effort, the anatomical accuracy needs equal attention. To this end, we calculate the segmentation results of the synthesized data and compare with their ground truths.[6] Both real scans and the synthesized results are fed into the segmentation tool, *i.e.*, FMRIB software library (FSL[7] [16]) to segment major tissue classes (GM, WM, and CSF) of brain, and the yielded results are averagely shown for each brain volume. The tissue prior probability templates are based on averaged multiple automatic segmentation in standard space from the IXI and BraTS datasets, respectively. The evaluation criteria include PSNR, SSIM and Dice score to quantitatively assess the quality of the synthesized results.

5.3 Comparison Methods

We compare our results against several state-of-the-art cross-modality synthesis algorithms including REPLICA [17], V-S and V-US [33], GAN [9], 3D-cGAN [27],

[6] Ground truths are calculated through a well-known segmentation tool on the real scans.

[7] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/

Table 1. Quantitative evaluation of the quality of synthesized images using different methods on the IXI [1] and BraTS [25] datasets.

Metric (avg.)	REPLICA	V-S	V-US	GAN	3D-cGAN	MSGAN	CUT	CSCℓ_4Net	TransCSCN	Improvements ↑
IXI: T2w → PDw										
PSNR (dB)	*31.27*	33.87	32.99	32.25	32.76	32.98	34.06	36.64	**37.18**	0.54∼5.91
SSIM	*0.807*	0.851	0.836	0.831	0.851	0.856	0.876	0.900	**0.904**	0.004∼0.097
Dice (in %)	70.33	68.35	68.02	*66.52*	75.94	72.55	75.64	80.73	**82.03**	1.3∼15.51
IXI: PDw → T2w										
PSNR (dB)	*32.27*	34.28	32.87	33.46	35.08	35.63	36.97	38.08	**39.14**	1.06∼6.87
SSIM	*0.865*	0.919	0.902	0.901	0.899	0.899	0.910	0.959	**0.960**	0.002∼0.095
Dice (in %)	76.13	70.33	*69.66*	69.74	80.25	80.01	82.13	87.62	**88.59**	0.97∼18.93
BraTS: T1w → FLAIR										
PSNR (dB)	*31.60*	32.07	31.85	32.47	33.92	31.85	34.36	37.36	**39.12**	1.76∼7.52
SSIM	*0.811*	0.842	0.833	0.835	0.880	0.870	0.902	0.935	**0.943**	0.008∼0.131
Dice (in %)	70.92	69.89	69.44	*69.26*	73.94	74.02	78.92	84.07	**85.68**	1.61∼16.42
BraTS: FLAIR → T1w										
PSNR (dB)	*31.65*	33.00	31.80	31.93	32.89	33.72	34.96	36.51	**37.44**	0.93∼5.79
SSIM	*0.825*	0.857	0.842	0.847	0.881	0.860	0.887	0.911	**0.924**	0.013∼0.099
Dice (in %)	72.01	70.23	69.90	*69.62*	78.89	77.00	80.06	82.58	**84.02**	1.44∼12.01

MSGAN [24], CUT [28], and CSCℓ_4Net [15]. Note that REPLICA, V-S, GAN and MSGAN are the supervised methods, and we follow the defined rule and input paired data for their training. Others are all unsupervised approaches, thus we input our manually selected unpaired images for training. Moreover, following [17,33,27], the brain MRI scans are bias-field corrected. For fair comparison, we empirically set all methods following the recommended bias correction to obtain the best performance. Except for outer comparison, we also provide the ablation study for measuring the impact of each proposed penalization term.

5.4 Empirical Analysis

We evaluate both visual quality and segmentation performance of the synthesized data, and show the quantitative results along with others. The generality of our TransCSCN is explored by testing on many tasks distributed in two independent datasets with consistent property. Specifically, we demonstrate both visual and quantitative results in Figs. 2-3 and Table 1, respectively. The visual measurements are shown as the average value of the synthesis performance by PSNR and SSIM. The averaged segmentation results (referred as Dice score) potentially reflect the anatomical significance. In Figs. 2-3, we show two sets of synthesized results by different methods and the corresponding ground truths. We found that our method can generate more realistic results with well approximated appearance and better quantitative outcomes. Table 1 demonstrates the summarized performance between TransCSCN and other compared methods over different datasets on different tasks. The last row of Table 1 shows the performance boost over the worst compared results and the best compared results, respectively. In particular, TransCSCN consistently outperforms all advanced approaches and significantly boosts the performances especially in the experiments "T1-w → FLAIR" on the BraTS dataset. Our best case achieves 7.52dB (in PSNR) and

Table 2. Our comprehensive ablation study shows the effects of each proposed regularization on the IXI dataset [1] for T2w → PDw task, and on the BraTS dataset [25] for T1w → FLAIR task.

IXI: T2w → PDw							
CSCNet	GN	$\mathcal{L}_{\mathcal{H}_k}$	$\mathcal{L}_{\mathcal{G}}$	$\mathcal{L}_{\mathcal{S}}$	PSNR (dB)	SSIM	Dice (%)
✓					32.57	0.845	70.71
✓	✓				34.11	0.852	74.54
✓		✓			34.92	0.851	74.76
✓			✓		33.98	0.858	77.23
✓				✓	34.19	0.857	75.60
✓	✓	✓			36.09	0.881	78.03
✓	✓		✓		36.07	0.878	79.63
✓	✓			✓	36.05	0.861	79.62
✓		✓	✓		36.08	0.873	79.78
✓		✓		✓	36.05	0.879	79.82
✓			✓	✓	35.88	0.870	78.89
✓	✓	✓	✓		36.64	0.894	81.25
✓	✓	✓		✓	36.52	0.889	81.29
✓		✓	✓	✓	36.21	0.882	80.34
✓	✓	✓	✓	✓	**37.18**	**0.904**	**82.03**

BraTS: T1w → FLAIR							
CSCNet	GN	$\mathcal{L}_{\mathcal{H}_k}$	$\mathcal{L}_{\mathcal{G}}$	$\mathcal{L}_{\mathcal{S}}$	PSNR (dB)	SSIM	Dice (%)
✓					30.08	0.806	64.19
✓	✓				31.67	0.836	71.08
✓		✓			33.94	0.872	72.33
✓			✓		32.83	0.866	75.06
✓				✓	34.12	0.872	75.82
✓	✓	✓			36.30	0.895	78.23
✓	✓		✓		36.08	0.901	80.23
✓	✓			✓	37.35	0.922	81.66
✓		✓	✓		37.89	0.922	81.87
✓		✓		✓	37.87	0.921	81.58
✓			✓	✓	36.75	0.920	81.29
✓	✓	✓	✓		38.64	0.933	83.83
✓	✓	✓		✓	38.59	0.932	83.29
✓		✓	✓	✓	37.96	0.929	82.03
✓	✓	✓	✓	✓	**39.12**	**0.943**	**85.68**

0.131 (in SSIM) improvements over the worst one (REPLICA), while the performance of segmentation is boosted by 16.42% compared to the worst baseline (GAN). We notice that REPLICA generates visually weaker results but plausible segmentation results. Instead, the appearance quality of GAN seems slightly better than REPLICA, but getting the worst Dice overlap. We also investigate the variants of our models to explore effectiveness of each module. For the T1w → FLAIR experiments on the BraTS, we separately adopt GN, $\mathcal{L}_{\mathcal{H}_k}$, $\mathcal{L}_{\mathcal{G}}$, $\mathcal{L}_{\mathcal{S}}$ and freely combine them upon the baseline CSCNet to investigate the effects in terms of image quality and their segmentation performance, where the detailed results are shown in Table 2 as our comprehensive ablation study (GN means global normalization). We observe that with the assistance of $\mathcal{L}_{\mathcal{H}_k}$, $\mathcal{L}_{\mathcal{G}}$, $\mathcal{L}_{\mathcal{H}_k}$ and $\mathcal{L}_{\mathcal{S}}$, both visual and segmentation results are improved greatly. The appearance score is sensitive to $\mathcal{L}_{\mathcal{G}}$, while the Dice overlap is sensitive to $\mathcal{L}_{\mathcal{G}}$ and $\mathcal{L}_{\mathcal{S}}$.

6 Conclusions

In this paper, we proposed a transferable convolutional sparse coding network for generalizing brain image synthesis task. The proposed method delves into the feature representations that jointly learns the cross-domain transferable features while taking the benefits of both deeper mining and optimal regularization. With the globally normalized convolutional sparse coding net, we exploited the domain discrepancy metric, Laplacian co-regularization, and subspace mismatch penalization for minimizing the domain divergence, preserving the local geometries, and reducing the generalization errors. TransCSCN was evaluated on different datasets, and showed promising results outperforming a number of recent approaches consistently. In future work, we plan to explore the performance of TransCSCN on other medical image processing tasks such as confronting artifacts.

References

1. IXI – Information eXtraction from Images. https://brain-development.org/ixi-dataset/
2. Arar, M., Ginger, Y., Danon, D., Bermano, A.H., Cohen-Or, D.: Unsupervised multi-modal image registration via geometry preserving image-to-image translation. In: IEEE CVPR. pp. 13410–13419 (2020)
3. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
4. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning **3**(1), 1–122 (2011)
5. Bristow, H., Eriksson, A., Lucey, S.: Fast convolutional sparse coding. In: IEEE CVPR. pp. 391–398 (2013)
6. Choudhury, B., Swanson, R., Heide, F., Wetzstein, G., Heidrich, W.: Consensus convolutional sparse coding. In: IEEE ICCV. pp. 4280–4288 (2017)
7. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: SIGGRAPH. pp. 341–346 (2001)
8. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT Press (2016)
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020)
10. Gretton, A., Borgwardt, K., Rasch, M.J., Scholkopf, B., Smola, A.J.: A kernel method for the two-sample problem. arXiv preprint arXiv:0805.2368 (2008)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE CVPR. pp. 770–778 (2016)
12. Heide, F., Heidrich, W., Wetzstein, G.: Fast and flexible convolutional sparse coding. In: IEEE CVPR. pp. 5135–5143 (2015)
13. Huang, Y., Shao, L., Frangi, A.F.: DOTE: Dual convolutional filter learning for super-resolution and cross-modality synthesis in MRI. In: MICCAI. pp. 89–98. Springer (2017)
14. Huang, Y., Shao, L., Frangi, A.F.: Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding. In: IEEE CVPR. pp. 6070–6079 (2017)
15. Huang, Y., Zheng, F., Wang, D., Huang, W., Scott, M.R., Shao, L.: Brain image synthesis with unsupervised multivariate canonical CSCl4Net. In: IEEE CVPR. pp. 5881–5890 (2021)
16. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: FSL. Neuroimage **62**(2), 782–790 (2012)
17. Jog, A., Carass, A., Roy, S., Pham, D.L., Prince, J.L.: Random forest regression for magnetic resonance image synthesis. MIA **35**, 475–488 (2017)
18. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV. pp. 694–711. Springer (2016)
19. Kempen, V., Vliet, V.: The influence of the regularization parameter and the first estimate on the performance of Tikhonov regularized non-linear image restoration algorithms. Journal of Microscopy **198**(1), 63–75 (2000)
20. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
21. Li, Y., Shen, L.: Skin lesion analysis towards melanoma detection using deep learning network. Sensors **18**(2) (2018)

22. Li, Y., Yuan, L., Chen, Y., Wang, P., Vasconcelos, N.: Dynamic transfer for multi-source domain adaptation. In: IEEE CVPR. pp. 10998–11007 (2021)
23. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: ICML. pp. 97–105. PMLR (2015)
24. Mao, Q., Lee, H.Y., Tseng, H.Y., Ma, S., Yang, M.H.: Mode seeking generative adversarial networks for diverse image synthesis. In: IEEE CVPR. pp. 1429–1437 (2019)
25. Menze, B.H., Jakab, A., Bauer, S., et al.: The multimodal brain tumor image segmentation benchmark (BraTS). IEEE TMI **34**(10), 1993–2024 (2015)
26. Mondal, A.K., Dolz, J., Desrosiers, C.: Few-shot 3D multi-modal medical image segmentation using generative adversarial learning. arXiv preprint arXiv:1810.12241 (2018)
27. Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., Shen, D.: Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In: MICCAI. pp. 455–463. Springer (2018)
28. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: ECCV. pp. 319–345. Springer (2020)
29. Qiu, Q., Patel, V.M., Turaga, P., Chellappa, R.: Domain adaptive dictionary learning. In: ECCV. pp. 631–645. Springer (2012)
30. Rosenthal, M., Weeks, S., Aylward, S., Bullitt, E., Fuchs, H.: Intraoperative tracking of anatomical structures using fluoroscopy and a vascular balloon catheter. In: MICCAI. pp. 1253–1254. Springer (2001)
31. Roy, S., Carass, A., Prince, J.L.: Magnetic resonance image example-based contrast synthesis. IEEE TMI **32**(12), 2348–2363 (2013)
32. Van Loan, C.F., Golub, G.: Matrix computations (Johns Hopkins studies in mathematical sciences) (1996)
33. Vemulapalli, R., Van Nguyen, H., Zhou, S.K.: Unsupervised cross-modal synthesis of subject-specific scans. In: IEEE ICCV. pp. 630–638 (2015)
34. Wang, H., Li, Y., He, N., Ma, K., Meng, D., Zheng, Y.: DICDNet: Deep interpretable convolutional dictionary network for metal artifact reduction in CT images. IEEE TMI **41**(4), 869–880 (2021)
35. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE TIP **19**(11), 2861–2873 (2010)
36. Yang, L., Balaji, Y., Lim, S.N., Shrivastava, A.: Curriculum manager for source selection in multi-source domain adaptation. In: ECCV. pp. 608–624 (2020)
37. Zeiler, M.D., Krishnan, D., Taylor, G.W., Fergus, R.: Deconvolutional networks. In: IEEE CVPR. pp. 2528–2535. IEEE (2010)
38. Zhang, H., Mao, H., Long, Y., Yang, W., Shao, L.: A probabilistic zero-shot learning method via latent nonnegative prototype synthesis of unseen classes. IEEE Trans Neural Netw Learn Syst **31**(7), 2361–2375 (2019)
39. Zheng, M., Bu, J., Chen, C., Wang, C., Zhang, L., Qiu, G., Cai, D.: Graph regularized sparse coding for image representation. IEEE TIP **20**(5), 1327–1336 (2010)
40. Zhong, E., Fan, W., Yang, Q., Verscheure, O., Ren, J.: Cross validation framework to choose amongst models and datasets for transfer learning. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 547–562. Springer (2010)
41. Zhu, J.Y., Krähenbühl, P., Shechtman, E., Efros, A.A.: Generative visual manipulation on the natural image manifold. In: ECCV. pp. 597–613. Springer (2016)
42. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE ICCV. pp. 2223–2232 (2017)