A Novel Negative-Transfer-Resistant Fuzzy Clustering Model With a Shared Cross-Domain Transfer Latent Space and its Application to Brain CT Image Segmentation

Yizhang Jiang, Xiaoqing Gu, Dongrui Wu, Wenlong Hang, Jing Xue, Shi Qiu, and Chin-Teng Lin

Abstract—Traditional clustering algorithms for medical image segmentation can only achieve satisfactory clustering performance under relatively ideal conditions, in which there is adequate data from the same distribution, and the data is rarely disturbed by noise or outliers. However, a sufficient amount of medical images with representative manual labels are often not available, because medical images are frequently acquired with different scanners (or different scan protocols) or polluted by various noises. Transfer learning improves learning in the target domain by leveraging knowledge from related domains. Given some target data, the performance of transfer learning is determined by the degree of relevance between the source and target domains. To achieve positive transfer and avoid negative transfer, a negative-transfer-resistant mechanism is proposed by computing the weight of transferred knowledge. Extracting a negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space (called NTR-FC-SCT) is proposed by integrating negative-transfer-resistant and maximum mean discrepancy (MMD) into the framework of fuzzy c-means clustering. Experimental results show that the proposed NTR-FC-SCT model outperformed several traditional non-transfer and related transfer clustering algorithms.

Index Terms—Medical image segmentation, fuzzy clustering, transfer learning, negative transfer

1 Introduction

With the development of electronic information and computer technology, medical imaging and image processing technology have developed rapidly. Medical imaging equipments collect images for a short time, and are less affected by external factors. Today, medical imaging technology has become a powerful tool and core technology for modern clinical diagnosis and treatment. Commonly used medical imaging techniques include Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Computed Radiography (CR), Ultrasound, and so on. CT scans produce clearer images than conventional x-ray for internal organs, bone and soft tissues. MRI scans furnish greater clearness and higher resolution than CT scans with lower resolution [1]. Medical image segmentation is a basic and important step in medical image processing and analysis. It is also the basis of medical image registration, medical image information fusion and 3D visualization. In the current clinical practice, manual segmentation based on visual recognition and empirical judgment by doctors is still the most typical and common segmentation method. However, manual segmentation is tedious, time consuming and subjective. For example, in the Isointense Infant Brain Segmentation Challenge (ISEG2017), manual segmentation of each brain MRI scan took an average of one week for neuroradiologists [2]. Furthermore, the differences in physician experience and uncertain factors such as visual fatigue will affect the correct analysis of segmentation results.

With the rapid growth of the image processing technology, many automatic image processing techniques have appeared in recent years [3], [4], [5]. Image segmentation methods can be broadly classified into four categories: graph based methods, classification methods, deep learning methods and clustering methods [6]. A medical image in graph based image segmentation is presented as a weighted undirected graph [7]. Each pixel or region in the image is treated

- *Y. Jiang is with the Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi, Jiangsu 214122, China and is also with the School of Digital Media, Jiangnan University, Wuxi, Jiangsu 214122, China. E-mail: yzjiang@jiangnan.edu.cn.*
- *X. Gu is with the School of Information Science and Engineering, Changzhou University, Changzhou, Jiangsu 213164, China. E-mail: czxqgu@163.com.*
- *D. Wu is with the Key Laboratory of the Ministry of Education for Image Processing and Intelligent Control, School of Automation, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China. E-mail: drwu09@gmail.com.*
- *W. Hang is with the School of Computer Science and Technology, Nanjing Tech University, Nanjing, Jiangsu 211816, China. E-mail: hangwenlongnj@sina.com.*
- *J. Xue is with the Department of Nephrology, the Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, Jiangsu 214023, China. E-mail: 475809491@qq.com.*
- *S. Qiu is with the Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shaanxi 710119, China. E-mail: qiushi215@163.com.*
- *C.-T. Lin is with the Centre of Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia. E-mail: chinteng.lin@uts.edu.au.*

Manuscript received 14 Oct. 2019; revised 14 Dec. 2019; accepted 30 Dec. 2019. Date of publication 3 Jan. 2020; date of current version 3 Feb. 2021. (Corresponding author: X. Gu).
Digital Object Identifier no. 10.1109/TCBB.2019.2963873

as a vertex of a graph, and the set of edges can be connected by adjacent pixels or two adjacent regions. Then the image is divided into several parts according to the relationship of the adjacent pixels. The second category, also called supervised methods, use labeled segmented images to extract features and train a segmentation model, such as k-nearest neighbor (KNN) [8], neural network [9], support vector machine (SVM) [10], and so on. A major drawback of classification methods is that they require sufficient labeled training images. But in the areas of medical imaging, it is relatively easy and inexpensive to obtain a large amount of unlabeled data [11]. Deep learning learns the feature representation of tissue contour based on deep convolutional neural networks [12], [13]. Deep learning methods have successfully applied for medical image segmentation in recent years. However, deep learning methods usually need a large number of training dataset and special hardware devices. It is known that the medical image segmentation problem can be considered as classifying the pixels of images into homogeneous regions. This process can be viewed as clustering problem. Clustering method, as an unsupervised machine learning approach, is the process of grouping a set of data points into subsets so that data points in the same subset are similar (according to some criteria). Widely used clustering methods include expectation-maximization, spectral clustering and fuzzy clustering, and so on. In the last decade, clustering-based methods have attracted great interest in the segmentation of medical images [14]. Portela *et al.* [15] proposed clustering based semi-supervised classification for brain image segmentation, which used K-means clustering as initial processing to select brain slices. Ortiz *et al.* [16] improved brain image segmentation by using self-organizing maps (SOMs) based voxel clustering to extract features, and then used entropy-gradient clustering to segment brain images. Based on the idea of multiobjective optimization, Saha *et al.* [17] proposed semi-supervised clustering in the intensity space for medical image segmentation. Abdel-Maksouda *et al.* [18] combined K means and FCM clustering to propose a hybrid clustering method for tumour segmentation from brain image.

To make cluster-based segmentation methods perform better, the training medical image data needs to be representative of the target data. However, medical images are often collected with different scanners and scanning parameters, and medical images may have large differences in image quality due to machine performance or scanning technology, such as varying degrees of rotation, noise, etc. The requirement of training and target data under the same distribution prevents the use of clustering methods in larger research and clinical practice. Since the above scenarios exist in a large number of real-world environments, this leads to unsatisfactory segmentation results and the risk of algorithm failure.

To solve this problem, researchers have introduced the idea of transfer learning into clustering methods [19], [20]. With the help of some knowledge of auxiliary domain (called source domain), transfer learning handles the cases where the distribution, feature space or tasks are different between source domain and testing domain (called target domain). In transfer learning, the auxiliary knowledge from the source domain involves the data sample, feature representations, parameters and relationships [21], [22], [23]. The knowledge is usually obtained from certain precise procedures and reliable theory through some specific perspectives. Jiang *et al.* [24] proposed a transfer spectral clustering method, which used both data manifold and feature manifold between related clustering tasks. Deng *et al.* [25] proposed a transfer prototype-based fuzzy clustering method, which incorporated prototype knowledge induced from source domain to implement the clustering in the target domain. Qian *et al.* [20] proposed a cross-domain maximum entropy clustering method, which utilized the auxiliary knowledge from cluster centers and fuzzy memberships belonging to source data. However, these methods have a common assumption that source domain and target domain must have the same number of clusters. Moreover, most existing transfer clustering methods are not developed for noisy scenarios. Thus, these methods may be not suitable for noisy medical image segmentation.

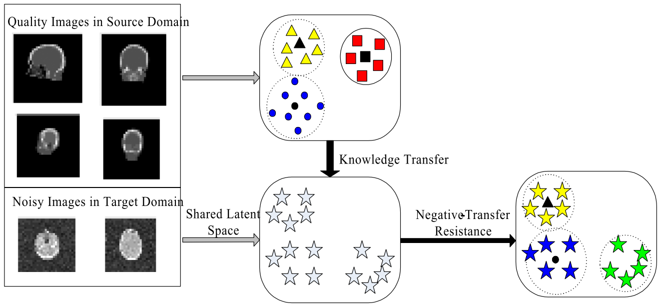


Fig. 1. The motivation of NTR-FC-SCT.

Since the medical images of different domains may have variations due to changes caused by noise, field offset and bias field, in this paper, we study the problem of medical image segmentation in a noisy scenario by transferring medical images collected from related scenarios. We consider the new noisy scenario as the target domain and the existing medical image dataset from related scenario elsewhere as the source domain, and then use the learning on clean images of source data to improve the clustering in target data. To improve the transfer learning performance, we consider learning the negative-transfer-resistant mechanism, so that the influence of positive transfer knowledge is reinforced and the influence of negative transfer knowledge is reduced or even eliminated. Meanwhile, we think medical images in different scenarios may share certain common representations such as bone and soft tissue, and the shared representations could be preserved in a shared space. Inspired of maximum mean discrepancy (MMD) [26], we learn the shared latent space for source and target domains such that the distributions in different domains are close to each other. We investigate transferring ability of each cluster belonging to source domain in the shared latent space for medical image segmentation modeling. We use the clustering centers in the source domain as the transfer knowledge, regardless of whether the number of clusters in the source domain is the same as that in the target domain. With the above ideas, we propose a negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space (called NTR-FC-SCT). The motivation of NTR-FC-SCT is shown in Fig. 1. Two cluster centers presented as black triangle and circle have positive transfer influence to the clustering in the target domain, while the cluster center presented as Black Square has negative influence to the clustering in the target domain. NTR-FC-SCT will automatically resist black square participating in the clustering in the target domain by using the negative-transfer-resistance mechanism. We evaluate the proposed model on real world datasets, and compare with several non-transfer and related transfer clustering methods. The results on real world brain CT dataset demonstrate that NTR-FC-SCT is more robust than the comparison methods.

The novelty of this study is as follows. 1) We formulate the problem of insufficient and noisy medical image segmentation as a model of transfer clustering task. To the best of our knowledge, our study is the first attempt to address this issue. 2) The negative-transfer-resistance mechanism is proposed to identify and resist negative source transfer knowledge. 3) The MMD is introduced into NTR-FC-SCT to unify the representation of image data of different domains in the shared transfer latent space, which helps transferring knowledge across different domains. 4) Clustering centers based transfer matching scheme is used to deal with the inconsistency problem of clustering numbers between source and target domains, so that more robust cluster performance can be promoted.

The rest of this paper is organized as follows. Concepts related to FCM, transfer learning and MMD are reviewed in Section 2. In Section 3, the negative-transfer-resistant mechanism and new proposed algorithm is introduced. Its parameter learning based on iteratively optimization strategy is then presented accordingly. The experimental results in real-world brain CT image datasets are reported in Section 4. Conclusions are given in the last section.

2 Related Work

2.1 Conventional FCM for Image Segmentation

Fuzzy *C*-means (FCM) clustering [27] is one of most commonly used fuzzy clustering methods. FCM allows data points to belong to more than one cluster defined by a membership matrix. Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ be a given dataset where d and N are data dimension and capacity, respectively. Suppose C clusters exist in $\mathbf{X}$, FCM derives the following objection function:

$$\min_{\mathbf{U},\mathbf{V}} J = \sum_{i=1}^{N}\sum_{j=1}^{C}\mu_{ij}^{m}\left\|\mathbf{x}_i - \mathbf{v}_j\right\|^2$$
$$s.t. 0 \le \mu_{ij} \le 1, \sum_{j=1}^{C}\mu_{ij} = 1, \tag{1}$$

where $\mathbf{U} = [\mu_{ij}]_{N\times C}$ denotes the fuzzy membership matrix, and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_C]^T$ denotes the clustering center matrix. m ($m > 1$) denotes the fuzzy index.

FCM finds an optimal group of sets to explain the data samples into C clusters via matrixes $\mathbf{Y}$ and $\mathbf{U}$. FCM minimizes the total membership weighted distance of each sample $\mathbf{x}_i$ to the clustering center $\mathbf{v}_j$. FCM can easily optimize the objective functions by an iterative technique. Among center-based clustering methods, FCM was simple, efficient and high popularity. It is widely used in transfer clustering methods. For example, a FCM-based transfer learning was proposed in [28], which combined with Gini-Simpson diversity index and quadratic weights on membership. A knowledge-leveraged transfer FCM (KL-TFCM) was proposed in [29], which used three-interlinked framework of knowledge extraction, knowledge matching, and knowledge utilization to leverage source information to help clustering in the target domain. However, the above FCM-based transfer learning clustering methods are completed in the original space, and while they do not consider resisting negative transfer.

2.2 Transfer Learning and Maximum Mean Discrepancy

Currently, when the training data is not enough to represent the data in the current domain, transfer learning, multi-task learning and co-clustering are three effective techniques that can enhance the clustering performance in the current domain. Multi-task learning performs multiple learning tasks together through by sharing certain knowledge among all tasks [30], [31]. Co-clustering performs clustering on both objects and features to exploit the clear duality between rows and columns of a contingency table [32]. Transfer learning clustering enhances the clustering performance in the target domain by leveraging useful knowledge from different but related domains. Many researches show that the transfer clustering methods have better learning ability to obtain an effective model with the idea of transfer learning [28], [29], [33]. In real applications, due to the existences of noise and field offset etc, the insufficient medical images are inadequate to complete image segmentation. Therefore, we think transfer learning clustering can be used as an effect technology to promote the segmentation of insufficient and noisy medical images in the new domain.

In transfer learning, a fundamental problem is to evaluate the distribution difference between source domain and target domain. Many criteria, like Kullback-Leibler (KL) divergence can be used for distribution estimation. But some criteria need density estimation, some are parametric and not suitable for high-dimensional data [34], [35]. In these cases, maximum mean discrepancy (MMD) as a nonparametric estimate criterion receives is widely used for comparing distributions. Let $\mathbf{X}_s = \{\mathbf{x}_{1,s}, \mathbf{x}_{2,s}, \ldots, \mathbf{x}_{N_s,s}\}$ and $\mathbf{X}_t = \{\mathbf{X}_{1,t}, \mathbf{X}_{2,t}, \ldots \mathbf{X}_{N_t,t}\}$ denote the samples from distributions $P_{source}(\mathbf{X}_s)$ and $P_{target}(\mathbf{X}_t)$ belonging to source and target domains, respectively. The MMD for comparing distributions between $P_{source}(\mathbf{X}_s)$ and $P_{target}(\mathbf{X}_t)$ is defined as

$$Dist\big(P_{source}(\mathbf{X}_s), P_{target}(\mathbf{X}_t)\big) = \left\|\frac{1}{N_s}\sum_{i=1}^{N_s} f(\mathbf{x}_{i,s}) - \frac{1}{N_t}\sum_{i=1}^{N_t} f(\mathbf{x}_{i,t})\right\|^2. \tag{2}$$

MMD is based on reproducing kernel Hilbert space (RKHS). Suppose $f : \mathbf{X} \to \mathbf{H}$, $\mathbf{H}$ is a universal RKHS. By inducing nonlinear mapping ϕ, function evaluation can be represented $f(\mathbf{x}) = <\phi(\mathbf{x}), f>$, then equation (2) can be rewritten as

$$Dist(P_{source}(\mathbf{X}_s), P_{target}(\mathbf{X}_t)) = \left\|\frac{1}{N_t}\sum_{i=1}^{N_t} \phi(\mathbf{x}_{i,t}) - \frac{1}{N_s}\sum_{i=1}^{N_s} \phi(\mathbf{x}_{i,s})\right\|^2. \tag{3}$$

When the difference between source and target domains is small, the relationship between two domains is strong and the transfer knowledge can be fully utilized. However, when the

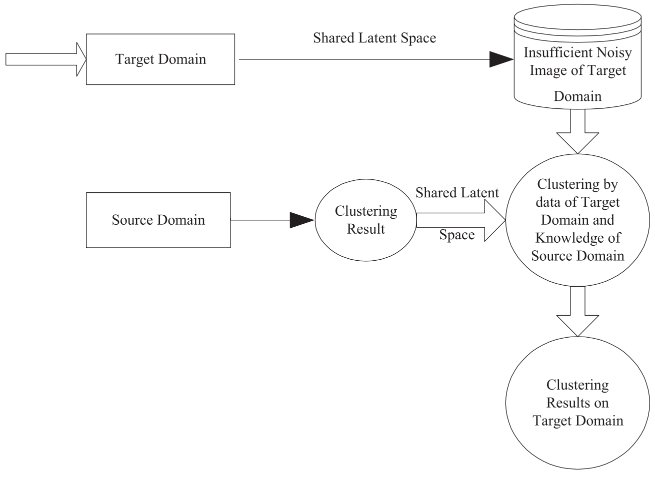


Fig. 2. The schematic diagram of NTR-FC-SCT.

data in the source domain is not sufficiently related, the clustering performance in the target domain may not only fail to promote, it may even actually decrease. Thus, transfer learning would resist negative transfer when source and target domains are not a good match. One strategy of resisting negative transfer is to identify and reject unhelpful knowledge from source domain. Some data selection and source selection methods have been proposed. The former implements some rules to select data samples to reconstruct the training set of source domain. For example, Rosenstein *et al.* [36] proposed a detecting negative transfer algorithm based on naive Bayes classification model. Croonenborghs *et al.* [37] proposed an option-based transfer in reinforcement learning algorithm to achieve a balance between positive and negative transfer. The source selection methods are applicable for multiple source scenarios, which select the best source domain (task) for transfer learning. An example of this strategy is Talvitie and Singh [38] proposed a Markov decision process to select the proper source task.

3 Negative-Transfer-Resistant Fuzzy Clustering Model With a Shared Cross-Domain Transfer Latent Space

The schematic diagram of NTR-FC-SCT is shown in Fig. 2. In the noisy transfer learning scenario, compared with data sample, feature representations, parameters and relationships are usually considered as being more insightful and more resistant to noise. In this study, we use the cluster centers in the source domain as auxiliary knowledge. The reason is that cluster centers are computed by certain reliable theories and rigorous procedures; such that the obtained cluster centers can well represent a cluster and all affiliated samples in a cluster.

3.1 Negative-Transfer-Resistant Mechanism

To resist the negative transfer, our objective is to discard bad cluster centers in the source domain and select helpful cluster centers that can help the clustering task satisfactorily. Let we have a total of N^{SD} training images in the source domain $\mathbf{x}_{i,s}(\mathbf{x}_{i,s} \in \mathbf{R}^{d\times 1}, i = 1, 2, \ldots, N^{SD})$ and N^{TD} training images in the target domain $\mathbf{x}_{i,t}(\mathbf{x}_{i,t} \in \mathbf{R}^{d\times 1}, i = 1, 2, \ldots, N^{TD})$, where $N^{TD} \ll N^{SD}$. We consider there exists a shared latent space, spanned by a projection matrix $\Theta \in \mathbf{R}^{r\times d}$, where r is the dimensions of the shared latent space. In this way, the known cluster centers $\hat{\mathbf{v}}_h^{SD}(h = 1, 2, \ldots, C^{SD})$ in the source domain obtained by a certain clustering method could be represented as $\Theta\hat{\mathbf{v}}_h^{SD}$. The projection of a target domain sample $\mathbf{x}_{i,t}$ could be represented as $\Theta\mathbf{x}_{i,t}$. Suppose $\tilde{\mathbf{v}}_j^{TD}(j = 1,\ 2, \ldots, C^{TD})$ is the unsolved cluster centers in the target domain in the shared latent space. We consider the following optimization term for resisting negative transfer:

$$\min_{\tilde{\mathbf{v}}}\ J(\tilde{\mathbf{V}}) = \sum_{j=1}^{C^{TD}} \sum_{h=1}^{C^{SD}} \left\| \tilde{\mathbf{v}}_j^{TD} - S_{jh}\Theta\hat{\mathbf{v}}_h^{SD} \right\|^2, \tag{4}$$

where the parameter S_{jh} is called the weight of transfer knowledge, and its value is in the range $[0, 1]$. S_{jh} denotes the matching degree between the *j*th cluster center of the target domain and the *h*th cluster center of the source domain. To find useful transfer knowledge from the cluster centers in the source domain, it is needed to devise a strategy to set the values of S_{jh} to high values for positive transfer and low values for negative transfer. In Eq. (4), we set S_{jh} as follows:

$$S_{jh} = 1/\sum_{h'=1}^{C^{\mathrm{SD}}} \frac{\left\| \tilde{\mathbf{v}}_j^{TD} - \Theta\hat{\mathbf{v}}_h^{SD} \right\|^2}{\left\| \tilde{\mathbf{v}}_j^{TD} - \Theta\hat{\mathbf{v}}_{h'}^{SD} \right\|^2}, \tag{5}$$

When S_{jh} tends to 1, $\tilde{\mathbf{v}}_j^{TD}$ exactly matches $\Theta\hat{\mathbf{v}}_h^{SD}$. In this case, the clustering results will be coherent and these two centers clusters of different domains are much closer to each other. When S_{jh} tends to 0, $\tilde{\mathbf{v}}_j^{TD}$ does not match $\Theta\hat{\mathbf{v}}_h^{SD}$. In this case, we make $\Theta\hat{\mathbf{v}}_h^{SD}$ participate in transfer learning as little as possible. That is to say, the transfer performance will be at least no worse than performing the target clustering without transfer. In other words, if the source clustering transfer ability of $\Theta\hat{\mathbf{v}}_h^{SD}$ decreases cluster performance of target domain data, it means the partial source clustering may be not related or the relationship is not sufficiently leveraged, then the negative transfer has occurred. By adjusting the value of $\Theta\hat{\mathbf{v}}_h^{SD}$, Eq. (4) can help the transfer model make positive transfer when two domains are appropriately matched and resist negative transfer when two domains are not matched. At one extreme, S_{jh} is set to be 1, the transferred knowledge from the source domain are completely helpful, such that the cluster centers in the source and target domains are coincided with each other, and all transferred knowledge are completely adopted. At the other extreme, when S_{jh} tends to 0, it means the transferred knowledge from the source domain are unhelpful. The clustering on the target domain will disregard the transferred knowledge. But in most cases part transferred knowledge are selectively keep and the other parts are disregard.

3.2 The Proposed NTR-FC-SCT

To find a proper projection matrix Θ, we think the difference between source and target domains in the shared latent space should be as small as possible, such that the relationship between domains is strengthened, and the transfer knowledge of data in the source domain will be more helpful to complete medical segmentation in the target domain. Based on the definition of MMD, the difference between two domains in the shared latent space can be computed as follows

$$\begin{aligned} d(P_{source}, P_{target}) &= \left\| \frac{1}{N^{TD}} \sum_{i=1}^{N^{TD}} \Theta \mathbf{x}_{i,t} - \frac{1}{N^{SD}} \sum_{i=1}^{N^{SD}} \Theta \mathbf{x}_{i,s} \right\|^2 \\ &= \frac{1}{(N^{TD})^2} \sum_{i=1}^{N^{TD}} \sum_{j=1}^{N^{TD}} \Theta \mathbf{x}_{i,t} \mathbf{x}_{j,t}^T \Theta^T + \frac{1}{(N^{SD})^2} \sum_{i=1}^{N^{SD}} \sum_{j=1}^{N^{SD}} \Theta \mathbf{x}_{i,s} \mathbf{x}_{j,s}^T \Theta^T \\ &- \frac{2}{N^{TD} N^{SD}} \sum_{i=1}^{N^{TD}} \sum_{j=1}^{N^{SD}} \Theta \mathbf{x}_{i,t} \mathbf{x}_{j,s}^T \Theta^T. \end{aligned} \tag{6}$$

The distribution difference between source and target domains is simply the distance between the two mean in the shared latent space. Let

$$\begin{aligned} \Omega =& \frac{1}{(N^{TD})^2} \sum_{i=1}^{N^{TD}} \sum_{j=1}^{N^{TD}} \mathbf{x}_{i,t} \mathbf{x}_{j,t}^T + \frac{1}{(N^{SD})^2} \sum_{i=1}^{N^{SD}} \sum_{j=1}^{N^{SD}} \mathbf{x}_{i,s} \mathbf{x}_{j,s}^T \\ &- \frac{2}{N^{TD} N^{SD}} \sum_{i=1}^{N^{TD}} \sum_{j=1}^{N^{SD}} \mathbf{x}_{i,t} \mathbf{x}_{j,s}^T. \end{aligned} \tag{7}$$

The optimization of $d(P_{source}, P_{target})$ can be simplified as

$$\begin{aligned} &\min_{\Theta} \ d(P_{source}, P_{target}) = \min_{\Theta} \Theta \Omega \Theta^T, \\ &s.t. \ \ \Theta \Theta^T = \mathbf{I}_{r \times r}. \end{aligned} \tag{8}$$

where $\mathbf{I}_{r\times r}$ is a $r \times r$ identity matrix, such that the projection matrix $\mathbf{H}$ is orthogonal.

Coming to the transfer learning tasks, we incorporate Eqs. (4) and (8) into the FCM framework. We obtain the objective function of NTR-FC-SCT as follows:

$$\begin{aligned} J =& \sum_{i=1}^{N^{TD}} \sum_{j=1}^{C^{TD}} \left(\mu_{ij}^{TD}\right)^m \left(\left\| \Theta \mathbf{x}_{i,t} - \tilde{\mathbf{v}}_j^{TD} \right\|^2 \right) \\ &+ \lambda_1 \sum_{j=1}^{C^{TD}} \sum_{h=1}^{C^{SD}} \left\| \tilde{\mathbf{v}}_j^{TD} - S_{jh} \Theta \hat{\mathbf{v}}_h^{SD} \right\|^2 + \lambda_2 \Theta \ \Omega \Theta^T, \\ & s.t. \ \Theta \Theta^T = \mathbf{I}_{r \times r}, \sum_{j=1}^{C^{TD}} \mu_{ij}^{TD} = 1, \end{aligned} \tag{9}$$

where parameters $\lambda_1 > 0$ and $\lambda_2 > 0$ are the coefficients of transfer optimization term and MMD term, respectively. In NTR-FC-SCT, the parameter λ_1 is used to control the influence of transfer optimization term $\sum_{j=1}^{C^{TD}} \sum_{h=1}^{C^{SD}} \|\tilde{\mathbf{v}}_j^{TD} - S_{jh} \Theta \hat{\mathbf{v}}_h^{SD}\|^2$ to the entire objective function. The larger the λ_1 value, the greater the contribution of the transfer term will be. In this case, the unsolved cluster centers $\tilde{\mathbf{v}}_j^{TD}$ in the target domain should be close to $S_{jh}\Theta\hat{\mathbf{v}}_h^{SD}$ in the shared latent space. Conversely, when λ_1 tends to 0, the contribution of the transfer term is weakened, the difference between the unsolved cluster centers and known cluster centers in two different domains can be relaxed.

3.3 Optimization of NTR-FC-SCT

The solution of objective function in Eq. (9) relates to the matrixes Θ, $\mathbf{U}$ and $\tilde{\mathbf{V}}$. In the following, we optimize them one by one using the iteratively optimization strategy. In terms of the Lagrange optimization, the minimization of J in Eq. (9) by introducing the Lagrangian multiplier α in Eq. (9) can be converted to the following unconstrained minimization problem:

$$\begin{aligned} L =& \sum_{i=1}^{N^{TD}} \sum_{j=1}^{C^{TD}} \left(\mu_{ij}^{TD}\right)^m \left(\left\| \Theta \mathbf{x}_{i,t} - \tilde{\mathbf{v}}_j^{TD} \right\|^2 \right) \\ &+ \lambda_1 \sum_{j=1}^{C^{TD}} \sum_{h=1}^{C^{SD}} \left\| \tilde{\mathbf{v}}_j^{TD} - S_{jh} \Theta \hat{\mathbf{v}}_h^{SD} \right\|^2 \\ &+ \lambda_2 \Theta \Omega \Theta^T + \sum_{i=1}^{N^{TD}} \alpha_i \left(1 - \sum_{j=1}^{C^{TD}} \mu_{ij}^{TD} \right). \end{aligned} \tag{10}$$

In the first step, we fix parameters Θ and $\mathbf{U}$, and only consider $\tilde{\mathbf{V}}$. To minimize this objective on parameter $\tilde{\mathbf{V}}$, we set the derivative with regard to $\tilde{\mathbf{V}}$ to zero:

$$\begin{aligned} \frac{\partial L}{\partial \tilde{\mathbf{v}}_j^{TD}} =& - \sum_{i=1}^{N^{TD}} \left(\mu_{ij}^{TD}\right)^m (\Theta \mathbf{x}_{i,t} - \tilde{\mathbf{v}}_j^{TD}) \\ &+ \lambda_1 \sum_{h=1}^{C^{SD}} \left(\tilde{\mathbf{v}}_j^{TD} - S_{jh} \Theta \hat{\mathbf{v}}_h^{SD} \right) = 0 \\ \Leftrightarrow & \tilde{\mathbf{v}}_j^{TD} \left(\sum_{i=1}^{N^{TD}} \left(\mu_{ij}^{TD}\right)^m + \lambda_1 C^{SD} \right) \\ =& \sum_{j=1}^{N^{TD}} \left(\mu_{ij}^{TD}\right)^m \Theta \mathbf{x}_{i,t} + \lambda_1 \sum_{h=1}^{C^{SD}} S_{jh} \Theta \hat{\mathbf{v}}_h^{SD}. \end{aligned} \tag{11}$$

We can get $\tilde{\mathbf{V}}$ in a closed form as follows

$$\begin{aligned} \tilde{\mathbf{v}}_j^{TD} =& \left(\sum_{j=1}^{N^{TD}} \left(\mu_{ij}^{TD}\right)^m \Theta \mathbf{x}_{i,t} + \lambda_1 \sum_{h=1}^{C^{SD}} S_{jh} \Theta \hat{\mathbf{v}}_h^{SD} \right) / \\ & \left(\sum_{i=1}^{N^{TD}} \left(\mu_{ij}^{TD}\right)^m + \lambda_1 C^{SD} \right). \end{aligned} \tag{12}$$

Likewise, in the next step, we fix parameters Θ and $\tilde{\mathbf{V}}$, and only consider $\mathbf{U}$. The minimization problem of Eq. (10) with respect to $\mathbf{U}$ can be equivalent to the following problem,

$$\begin{aligned} \frac{\partial L}{\partial \mu_{ij}^{TD}} &= m(\mu_{ij}^{TD})^{m-1} \left\| \Theta \mathbf{x}_{i,t} - \tilde{\mathbf{v}}_j^{TD} \right\|^2 - \alpha_i = 0 \\ \Leftrightarrow \mu_{ij}^{TD} &= \left(\alpha_i \Big/ m \left\| \Theta \mathbf{x}_{i,t} - \tilde{\mathbf{v}}_j^{TD} \right\|^2 \right)^{\frac{1}{m-1}}. \end{aligned} \tag{13}$$

In light of $\sum_{j=1}^{C^{TD}} \mu_{ij}^{TD} = 1$, we can obtain

$$\mu_{ij}^{TD} = \left(\frac{1}{\left\| \Theta \mathbf{x}_{i,t} - \tilde{\mathbf{v}}_j^{TD} \right\|^2} \right)^{\frac{1}{m-1}} / \sum_{k=1}^{C^{TD}} \left(\frac{1}{\left\| \Theta \mathbf{x}_{i,t} - \tilde{\mathbf{v}}_k^{TD} \right\|^2} \right)^{\frac{1}{m-1}}. \tag{14}$$

In the next step, we update matrix Θ and fix parameters $\tilde{\mathbf{V}}$ and $\mathbf{U}$. Let

$$\tilde{\mathbf{U}}_1 = [\tilde{\mu}_{11}, \ldots, \tilde{\mu}_{i1}, \ldots, \tilde{\mu}_{N^{\mathrm{TD}}1}] \in \mathbf{R}^{1 \times N^{\mathrm{TD}}}, , \tag{15}$$

where $\tilde{\mathbf{U}} = [\tilde{\mathbf{U}}_1, \ldots \tilde{\mathbf{U}}_{C^{\mathrm{TD}}}] \in \mathbf{R}^{1\times C^{\mathrm{TD}} N^{\mathrm{TD}}}$, $\widehat{\mathbf{U}} = diag(\tilde{\mathbf{U}}) \in \mathbf{R}^{C^{\mathrm{TD}} N^{\mathrm{TD}} \times C^{\mathrm{TD}} N^{\mathrm{TD}}}$. Let

$$\omega = [\underbrace{\mathbf{E}, \ldots, \mathbf{E}}_{C^{\mathrm{TD}}}] \in \mathbf{R}^{N^{\mathrm{TD}} \times C^{\mathrm{TD}} N^{\mathrm{TD}}}, \tag{16}$$

where $\mathbf{E} \in \mathbf{R}^{N^{\mathrm{TD}} \times N^{\mathrm{TD}}}$. Let $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_{C^{\mathrm{TD}}}] \in \mathbf{R}^{r\times C^{\mathrm{TD}} N^{\mathrm{TD}}}$, where $\mathbf{q}_i = [\underbrace{\mathbf{v}_i^{\mathrm{TD}}, \ldots, \mathbf{v}_i^{\mathrm{TD}}}_{N^{\mathrm{TD}}}] \in \mathbf{R}^{r\times N^{\mathrm{TD}}}$. Substituting Eqs. (16), (17) and (18) back to Eq. (9), the minimization problem of Eq. (10) with respect to Θ can be equivalent to the following problem,

$$\begin{aligned} \ell(\mathbf{\Theta}) &= tr\Big((\mathbf{\Theta}\mathbf{x}t\omega - \mathbf{Q})\widehat{\mathbf{U}}(\mathbf{\Theta}\mathbf{x}_t\omega - \mathbf{Q})^T\Big) \\ &+\lambda_1 tr\bigg((\mathbf{\Theta}\hat{\mathbf{V}}^{\mathrm{SD}}\mathbf{S}^T - \tilde{\mathbf{V}}^{\mathrm{TD}})(\mathbf{\Theta}\hat{\mathbf{V}}^{\mathrm{SD}}\mathbf{S}^T - \tilde{\mathbf{V}}^{\mathrm{TD}})^T\bigg) + \lambda_2 tr(\mathbf{\Theta}\Omega\mathbf{\Theta}^T), \end{aligned} \tag{17}$$

where

$$\begin{aligned} \mathbf{\Omega} = &\tfrac{1}{(N^{TD})^2}\mathbf{x}_t[\mathbf{1}]^{N^{TD}\times N^{TD}}(\mathbf{x}_t)^T + \tfrac{1}{(N^{SD})^2}\mathbf{x}_s[\mathbf{1}]^{N^{SD}\times N^{SD}}(\mathbf{x}_s)^T \\ &- \tfrac{1}{N^{TD}N^{SD}}\mathbf{x}_t[\mathbf{1}]^{N^{TD}\times N^{SD}}(\mathbf{x}_s)^T - \tfrac{1}{N^{TD}N^{SD}}\mathbf{x}_s[\mathbf{1}]^{N^{SD}\times N^{TD}}(\mathbf{x}_t)^T. \end{aligned}$$

Likewise, the minimization problem of Eq. (10) with respect to Θ can be equivalent to the following problem,

$$\begin{aligned} \frac{\partial \ell}{\partial \mathbf{\Theta}} =& (\mathbf{\Theta}\mathbf{x}_t\omega\ \widehat{\mathbf{U}}(\omega)^T(\mathbf{x}_t)^T - \mathbf{Q}\ \widehat{\mathbf{U}}(\omega)^T(\mathbf{x}_t)^T) \\ &+ \lambda_1\bigg(\mathbf{\Theta}\hat{\mathbf{V}}^{\mathrm{SD}}\mathbf{S}^T\mathbf{S}\Big(\hat{\mathbf{V}}^{\mathrm{SD}}\Big)^T - \tilde{\mathbf{V}}^{\mathrm{TD}}\mathbf{S}\Big(\hat{\mathbf{V}}^{\mathrm{SD}}\Big)^T\bigg) + \lambda_2\mathbf{\Theta}\Omega. \end{aligned} \tag{18}$$

In this study, the widely gradient descent method is adopted to compute the optimal Θ. By setting the initial value Θ as Θ^0, the gradient descent method successively optimal Θ as follows

$$\mathbf{\Theta}^l = \mathbf{\Theta}^{l-1} - \eta\frac{\partial \ell}{\partial \mathbf{\Theta}}|_{\Theta=\Theta^{l-1}}, \tag{19}$$

where η is the learning rate and l is the iteration number. Considering the constraint of $\Theta\Theta^T = \mathbf{I}$. After each updating step of Θ^l, let $\Theta^l = \Theta^{l'}\mathbf{R}$ be the QR decomposition of Θ^l, where $\Theta^{l'}$ has orthogonal columns and $\mathbf{R}$ is an upper triangle. Then we replace Θ^l with $\Theta^{l'}$ for the next iteration. Eq. (19) will be iteratively solved until the convergence condition is satisfied.

Based on the above analysis, the proposed NTR-FC-SCT model is presented in Algorithm 1.

4 Experiments

4.1 Data Sets and Settings

We use ultrashort echo time (UTE) and modified Dixon brain image datasets [39], [40]. It consists of 256 brain CT image slices of 10 patients, with each image of a resolution of 256×256 pixels. All CT images with corresponding manual segmentation are segmented into three classes: bone, water and soft issues. These class labels are assigned by physicians or technicians. We randomly select 20 brain CT images as the original target domain data, and the rest 236 brain CT images as source domain data. We consider the application of NTR-FC-SCT in the scenario of target images polluted by noise. To this aim, all target images were corrupted by 5, 10, 15, 20, 25 and 30 percent Gaussian noise. The example images in the source and target domains are shown in Fig. 3. Following the training protocol established in [41], we construct a total training data set combining 236 source brain images and random 8 target brain images, while the remaining 12 target brain images are used as testing brain images. We repeat the experiment for 10 runs and record the experimental results.

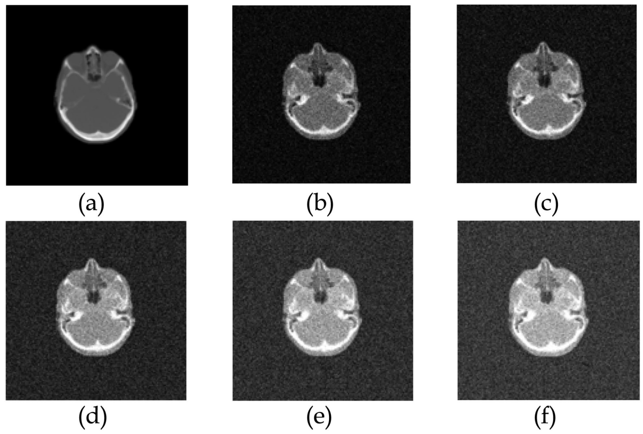


Fig. 3. The example brain CT images in source and target domains, (a) images in source domain, (b) Subject1 in the target domain with 5 percent noise, (c) Subject1 in the target domain with 10 percent noise, (d) Subject1 in the target domain with 15 percent noise, (e) Subject1 in the target domain with 20 percent noise, (f) Subject1 in the target domain with 25 percent noise, (g) Subject1 in the target domain with 30 percent noise.

Algorithm 1. NTR-FC-SCT Model

Initialize Set the maximum number of iterations t_{max}, the fuzzy index m, the regularization parameters λ_1 and λ_2, and the learning rate η.
Repeat:
Exacting transfer knowledge form the source domain;
Perform soft-partition clustering methods in the source domain, such as FCM, and obtain the cluster centers of data in the source domain;
$t = t + 1$;
Initialize the clustering centers of data in the target domain;
Compute the weight of transfer knowledge S_{jh} using Eq. (10);
Fix $\mathbf{U}(t)$ and $\Theta(t)$, obtain $\tilde{\mathbf{V}}^{TD}(t)$ using Eq. (12);
Fix $\tilde{\mathbf{V}}^{TD}(t)$ and $\Theta(t)$, obtain $\mathbf{U}(t)$ using Eq. (14);
Fix $\mathbf{U}(t)$ and $\tilde{\mathbf{V}}^{TD}(t)$, obtain $\Theta(t)$ using Eqs. (18) and (19);
Compute $J(t)$ using Eq. (9);
Until $\|J(t) - J(t-1)\| \le \delta$ or $t \ge t_{max}$;

To compare the segmentation performance of NTR-FC-SCT with that of existing methods, complete image segmentation is obtained and compared with segmentations obtained by FCM [27], transfer spectral clustering (TSC) [24], and type-I knowledge-transfer-oriented c-means (T1-KT-FCM) [28]. As introduced in Section 2, FCM is the baseline algorithm of NTR-FC-SCT. TSC performs transfer learning based on bipartite graph co-clustering, which adopts both the data manifold and sample manifold shared among different domains. T1-KT-FCM makes the cluster centers in the source domain as the transfer information to control the knowledge transfer in the test images. T1-KT-FCM incorporates this idea into FCM to

TABLE 1
NMI Performance of all Comparison Methods on 5 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0112	0.0041	0.5085	0.2249	0.5802	0.0254	**0.7146**	0.0056
Subject2	0.1515	0.1468	0.5801	0.1632	0.6359	0.0148	**0.7526**	0.0142
Subject3	0.1583	0.1040	0.5573	0.1591	0.6438	0.0290	**0.7253**	0.0059
Subject4	0.0137	0.0062	0.5149	0.1868	0.6091	0.0590	**0.7256**	0.0138
Subject5	0.2333	0.1355	0.5649	0.1840	0.6329	0.0254	**0.7432**	0.0094
Subject6	0.1823	0.1094	0.5659	0.1446	0.6505	0.0234	**0.7441**	0.0116
Subject7	0.0086	0.0046	0.5283	0.2242	0.6096	0.0212	**0.7352**	0.0110
Subject8	0.1324	0.1157	0.5664	0.1917	0.6603	0.0245	**0.7586**	0.0108
Subject9	0.1188	0.1240	0.5845	0.1638	0.6700	0.0399	**0.7693**	0.0111
Subject10	0.0185	0.0083	0.5765	0.2749	0.6512	0.0258	**0.7781**	0.0127
Subject11	0.1178	0.2362	0.5882	0.2124	0.6580	0.0179	**0.7862**	0.0068
Subject12	0.0642	0.0376	0.5719	0.1542	0.6631	0.0358	**0.7759**	0.0079

TABLE 3
NMI Performance of all Comparison Methods on 15 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0130	0.0106	0.5053	0.2261	0.4470	0.0181	**0.5684**	0.0082
Subject2	0.2346	0.1282	0.5815	0.1536	0.5631	0.0172	**0.6452**	0.0118
Subject3	0.2155	0.0679	0.5541	0.2003	0.4881	0.0175	**0.5978**	0.0024
Subject4	0.1789	0.2392	0.5076	0.2182	0.4509	0.0171	**0.5849**	0.0157
Subject5	0.2061	0.0862	0.5621	0.1394	0.5341	0.0121	**0.6220**	0.0030
Subject6	0.2592	0.1368	0.5691	0.1526	0.5079	0.0163	**0.6261**	0.0096
Subject7	0.0963	0.1848	0.5274	0.1991	0.4664	0.0168	**0.5939**	0.0130
Subject8	0.1826	0.1223	0.5674	0.1458	0.5087	0.0223	**0.6198**	0.0101
Subject9	0.2245	0.1570	0.5869	0.1486	0.5118	0.0120	**0.6419**	0.0157
Subject10	0.0735	0.1450	0.5697	0.2271	0.4975	0.0197	**0.6098**	0.0169
Subject11	0.0937	0.1846	0.5826	0.2770	0.5267	0.0242	**0.6107**	0.0075
Subject12	0.2454	0.1816	0.5720	0.2135	0.5328	0.0129	**0.6537**	0.0117

achieve automatic image segmentation. To obtain the optimal parameters in all four methods, the common used grid search is conducted. Fuzzy index m in all fuzzy clusters is set within the grid $\{1.1, 1.5, 2, 2.5\}$. The K-nearest parameters in TSC are set within the grid $\{0, 0.005, 0.1, 0.5, 0.7, 1, 1.5, 10, 50, 100\}$. The λ, γ parameters in T1-KT-FCM are set within the grid $\{10\mathrm{e}-4, 10\mathrm{e}-3, \ldots, 10\mathrm{e}4\}$. The parameters λ_1, λ_2 in NTR-FC-SCT are set within the grid $\{0, 10\mathrm{e}-4, 10\mathrm{e}-3, \ldots, 10\mathrm{e}6\}$, learning rate η in NTR-FC-SCT are set within the grid $\{1\mathrm{e}-4, 1\mathrm{e}-3, 1\mathrm{e}-2\}$, and the maximum number of iterations is 10e5.

In this study, the performance of image segmentation by clustering methods is evaluated in terms of two validity indicators: normalized mutual information (NMI) [42] and adjusted rand index (ARI) [43]. NMI and ARI can efficiently evaluate the agreement degree between the known clusters and the estimated data structure. Both NMI and ARI take values from 0 to 1, and larger value means better cluster performance. Experimental environment is Intel Core i3-4170 3.7 GHz CPU and 12 GM RAM, Windows 10, and MATLAB R2016a in this study.

4.2 Performance Comparison

The clustering performance of four methods is reported in the following. The mean and standard deviation of NMI and ARI for all compared clustering methods are displayed in Tables 1, 2, 3, 4, 5, 6, respectively. The experimental results show that three transfer learning methods are superior to FCM. The introduction of transfer knowledge from source domain has indeed promoted the cluster performance of data in the target domain. FCM is not a transfer learning cluster method which simply combines the source domain and target domain data as the training data. Due to underlying noise or outliers in the target domain, the distribution difference between source and target domains are significant different. Thus, FCM can not obtain good clustering performance in terms of NMI and ARI. Our model achieves the best performance in all datasets. TSC may be not suitable for the transfer scenario in noisy medical image segmentation, since the character of medical image are usually different in noisy scenario, while the manifold and sample manifold shared among different domains can not resist negative transfer for TSC. T1-KT-FCM exploits the transfer knowledge across domains in the original data space; however, the limited transfer knowledge can not be fully exploited in such original space. NTR-FC-SCT has shown better performance than the other comparison methods in terms of NMI and ARI. Both the reliable knowledge obtained in the source domain and the ability of resisting negative transfer has the important influence on the segmentation performance of NTR-FC-SCT. To better observe the behavior of all

TABLE 2
NMI Performance of all Comparison Methods on 10 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0155	0.0098	0.5063	0.2137	0.4748	0.0095	**0.6337**	0.0091
Subject2	0.2943	0.1666	0.5797	0.1689	0.5792	0.0194	**0.7230**	0.0045
Subject3	0.2182	0.0813	0.5502	0.1716	0.5224	0.0366	**0.6908**	0.0117
Subject4	0.0196	0.0163	0.5061	0.1966	0.4706	0.0110	**0.6475**	0.0127
Subject5	0.1959	0.1420	0.5635	0.1591	0.5273	0.0205	**0.7056**	0.0122
Subject6	0.2868	0.0777	0.5683	0.1421	0.5310	0.0162	**0.6894**	0.0177
Subject7	0.0076	0.0038	0.5336	0.2612	0.4696	0.0106	**0.6710**	0.0162
Subject8	0.2006	0.1377	0.5680	0.1448	0.5357	0.0217	**0.7077**	0.0158
Subject9	0.2052	0.1502	0.5923	0.1901	0.5409	0.0165	**0.7274**	0.0104
Subject10	0.0153	0.0136	0.5755	0.2609	0.5103	0.0186	**0.7244**	0.0046
Subject11	0.0814	0.1664	0.5846	0.2263	0.5231	0.0215	**0.7203**	0.0122
Subject12	0.1534	0.1036	0.5694	0.1864	0.5409	0.0062	**0.7144**	0.0149

TABLE 4
NMI Performance of all Comparison Methods on 20 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0246	0.0289	0.5095	0.1908	0.4782	0.0069	**0.6000**	0.0055
Subject2	0.3092	0.1299	0.5796	0.1631	0.5596	0.0082	**0.6348**	0.0128
Subject3	0.2597	0.0992	0.5556	0.1588	0.5029	0.0108	**0.5926**	0.0051
Subject4	0.0102	0.0061	0.5136	0.2026	0.4778	0.0066	**0.5254**	0.0133
Subject5	0.3071	0.0852	0.5654	0.1471	0.5333	0.0119	**0.6237**	0.0084
Subject6	0.2998	0.0697	0.5662	0.1376	0.5247	0.0176	**0.5875**	0.0099
Subject7	0.0994	0.1742	0.5298	0.2377	0.4804	0.0088	**0.5719**	0.0063
Subject8	0.2557	0.1427	0.5683	0.1438	0.5426	0.0160	**0.6023**	0.0114
Subject9	0.3098	0.0342	0.5852	0.1508	0.5366	0.0089	**0.6433**	0.0062
Subject10	0.0966	0.2007	0.5788	0.2565	0.5309	0.0066	**0.5991**	0.0080
Subject11	0.1196	0.2075	0.5818	0.2101	0.5525	0.0080	**0.6516**	0.0129
Subject12	0.1530	0.1197	0.5707	0.1767	0.5466	0.0167	**0.5869**	0.0122

TABLE 5
NMI Performance of all Comparison Methods on 25 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0107	0.0058	0.5078	0.2154	0.4969	0.0043	**0.5469**	0.0104
Subject2	0.1708	0.1001	0.5792	0.1741	0.5842	0.0076	**0.6572**	0.0182
Subject3	0.2460	0.0981	0.5503	0.1685	0.5232	0.0145	**0.5940**	0.0093
Subject4	0.0121	0.0081	0.5177	0.2390	0.5045	0.0112	**0.5717**	0.0197
Subject5	0.1503	0.1169	0.5621	0.1599	0.5580	0.0122	**0.6090**	0.0083
Subject6	0.2199	0.0660	0.5662	0.1533	0.5434	0.0311	**0.5935**	0.0142
Subject7	0.0284	0.0285	0.5331	0.2528	0.5156	0.0137	**0.5677**	0.0095
Subject8	0.1320	0.1375	0.5705	0.1577	0.5547	0.0156	**0.6272**	0.0136
Subject9	0.2594	0.1054	0.5859	0.1542	0.5682	0.0124	**0.6437**	0.0127
Subject10	0.1980	0.0293	0.5763	0.2203	0.5488	0.0073	**0.6416**	0.0157
Subject11	0.2156	0.2697	0.5830	0.2349	0.5745	0.0184	**0.6732**	0.0117
Subject12	0.3935	0.0908	0.5724	0.1827	0.5617	0.0122	**0.6386**	0.0078

TABLE 6
NMI Performance of all Comparison Methods on 30 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0120	0.0093	0.5086	0.2028	0.5569	0.0124	**0.6567**	0.0084
Subject2	0.2553	0.1770	0.5833	0.1497	0.6333	0.0113	**0.7439**	0.0109
Subject3	0.2010	0.0804	0.5526	0.1537	0.5872	0.0206	**0.6602**	0.0110
Subject4	0.0891	0.1752	0.5143	0.2144	0.5664	0.0050	**0.6720**	0.0114
Subject5	0.1820	0.1376	0.5663	0.1463	0.6172	0.0193	**0.7080**	0.0028
Subject6	0.2316	0.1094	0.5695	0.1928	0.6062	0.0120	**0.6872**	0.0086
Subject7	0.1080	0.1686	0.5316	0.2091	0.5769	0.0172	**0.6966**	0.0054
Subject8	0.1755	0.1045	0.5696	0.1777	0.6231	0.0097	**0.7037**	0.0103
Subject9	0.2584	0.0893	0.5913	0.1901	0.6295	0.0216	**0.7356**	0.0200
Subject10	0.0113	0.0135	0.5751	0.2901	0.6247	0.0198	**0.7441**	0.0115
Subject11	0.2121	0.2838	0.5827	0.2652	0.6343	0.0212	**0.7491**	0.0042
Subject12	0.1547	0.1170	0.5702	0.1941	0.6226	0.0045	**0.7256**	0.0095

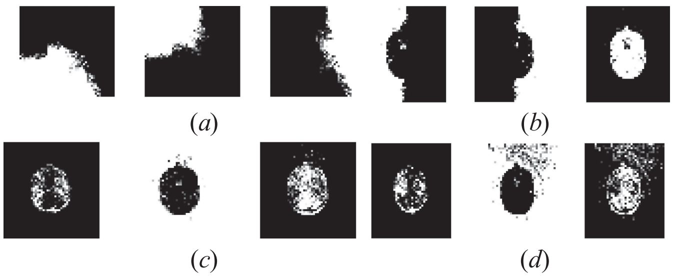


Fig. 4. Clustering segmentations on subject1+5 percent noise, (a)FCM, (b)TSC, (c)T1-KT-FCM, (d)NTR-FC-SCT.

algorithms, Figs. 4, 5, 6, 7, 8, 9 graphically shows the segmentation results of all comparison methods obtained on subject1 with different noise. Similar to the results in the Tables 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, NTR-FC-SCT obtains the best segmentation results for distinguishing the bone, water and soft issues. The boundaries between different organizations are smooth, and obvious are relatively clearer than the other three methods.

4.3 Flexibility Evaluation of NTR-FC-SCT

To validate the effect of two regularization terms on the performance of NTR-FC-SCT, we present two comparison methods NTR-FC-SCT ($\lambda_1 = 0$) and NTR-FC-SCT ($\lambda_2 = 0$),

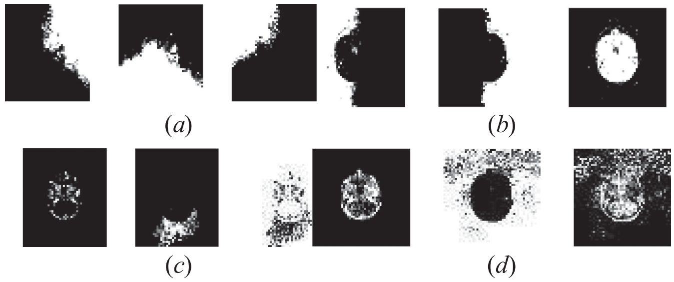


Fig. 5. Clustering segmentations on subject1+10 percent noise, (a)FCM, (b)TSC, (c)T1-KT-FCM, (d)NTR-FC-SCT.

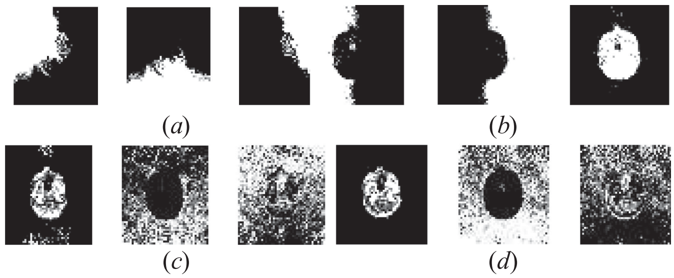


Fig. 6. Clustering segmentations on subject1+15 percent noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)NTR-FC-SCT.

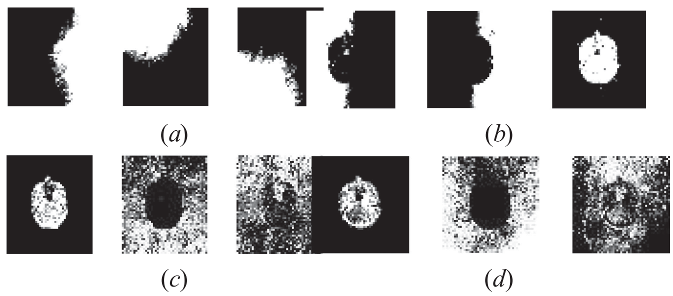


Fig. 7. Clustering segmentations on subject1+20 percent noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)NTR-FC-SCT.

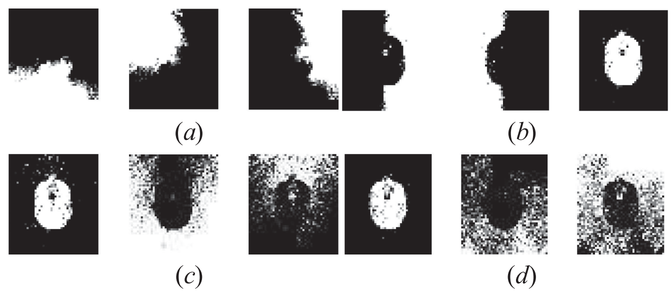


Fig. 8. Clustering segmentations on subject1+25 percent noise, (a)FCM, (b)TSC, (c) T1-KT-FCM, (d)NTR-FC-SCT.

obtained with the parameter $\lambda_1 = 0$ and $\lambda_2 = 0$ in NTR-FC-SCT, respectively. We compare them with FCM and NTR-FC-SCT on Subjects 1-8 and show their mean and standard deviation of NMI and ARI in Tables 13, 14, respectively. The experimental results show that the performances of both NTR-FC-SCT ($\lambda_1 = 0$) and NTR-FC-SCT ($\lambda_2 = 0$) are better than baseline method FCM. The regularization term in NTR-FC-SCT ($\lambda_2 = 0$) is $\sum_{j=1}^{C^{TD}} \sum_{h=1}^{C^{SD}} \|\tilde{\mathbf{v}}_j^{TD} - S_{jh}\Theta\hat{\mathbf{v}}_h^{SD}\|^2$, which can effectively resist negative transform by using the transfer optimization strategy and improve the segmentation performance in

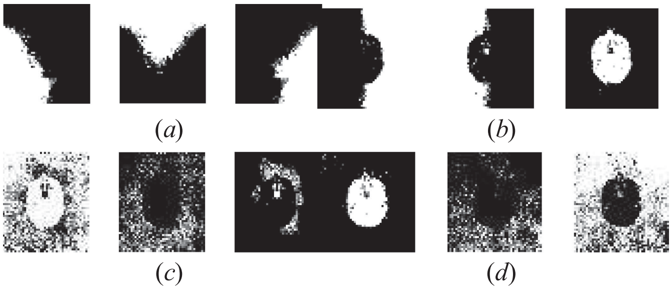


Fig. 9. Clustering segmentations on subject1+30 percent noise, (a) FCM, (b)TSC, (c) T1-KT-FCM, (d)NTR-FC-SCT.

TABLE 7
ARI Performance of all Comparison Methods on 5 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0021	0.0043	0.3399	0.1756	0.7579	0.0301	**0.8880**	0.0028
Subject2	0.0854	0.0814	0.3931	0.1561	0.8032	0.0160	**0.9111**	0.0109
Subject3	0.0861	0.0521	0.4144	0.1478	0.8072	0.0318	**0.8839**	0.0031
Subject4	0.0013	0.0041	0.3466	0.1701	0.7730	0.0629	**0.8913**	0.0107
Subject5	0.1186	0.0877	0.3884	0.1761	0.8085	0.0316	**0.9080**	0.0078
Subject6	0.0839	0.0560	0.4153	0.1393	0.8069	0.0264	**0.8950**	0.0108
Subject7	0.0012	0.0026	0.3593	0.1594	0.7728	0.0243	**0.8922**	0.0119
Subject8	0.0651	0.0732	0.3902	0.1753	0.8293	0.0276	**0.9169**	0.0095
Subject9	0.0658	0.0682	0.4153	0.1524	0.8305	0.0378	**0.9187**	0.0102
Subject10	0.0071	0.0055	0.3806	0.2090	0.8036	0.0304	**0.9242**	0.0106
Subject11	0.0771	0.1614	0.3858	0.1903	0.8089	0.0215	**0.9299**	0.0053
Subject12	0.0094	0.0109	0.3744	0.1365	0.8131	0.0424	**0.9269**	0.0081

TABLE 8
ARI Performance of all Comparison Methods on 10 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0028	0.0113	0.3403	0.1791	0.5742	0.0284	**0.8155**	0.0056
Subject2	0.1798	0.1182	0.3955	0.1613	0.5515	0.0292	**0.8812**	0.0056
Subject3	0.1158	0.0542	0.4121	0.1696	0.6163	0.0635	**0.8515**	0.0148
Subject4	0.0046	0.0063	0.3433	0.1729	0.5650	0.0237	**0.8281**	0.0134
Subject5	0.1051	0.0953	0.3848	0.1535	0.6339	0.0317	**0.8777**	0.0093
Subject6	0.1523	0.0812	0.4168	0.1373	0.6359	0.0297	**0.8467**	0.0136
Subject7	0.0008	0.0012	0.3607	0.1920	0.5573	0.0211	**0.8482**	0.0125
Subject8	0.1140	0.0988	0.3915	0.1311	0.6423	0.0301	**0.8522**	0.0149
Subject9	0.1202	0.1143	0.4210	0.1781	0.6382	0.0300	**0.8839**	0.0101
Subject10	0.0011	0.0089	0.3800	0.1963	0.5939	0.0328	**0.8849**	0.0037
Subject11	0.0493	0.1062	0.3848	0.1608	0.6045	0.0401	**0.8742**	0.0118
Subject12	0.0484	0.0401	0.3737	0.1681	0.6302	0.0112	**0.8738**	0.0125

noisy scenario. The regularization term in NTR-FC-SCT ($\lambda_1 = 0$) is $\Theta^T\Omega\Theta$, which finds a shared latent space for data cross domains, so that the projection data distributions of the source and target domains are close to each other. NTR-FC-SCT has the advantages of both NTR-FC-SCT ($\lambda_1 = 0$) and NTR-FC-SCT ($\lambda_2 = 0$). It can exploit more transfer knowledge; meanwhile, and achieve a good balance between making use of positive transfer and resisting negative transfer.

Next, we discuss the influence of the number of samples in the source domain on the performance of NTR-FC-SCT. We randomly select 10, 30, 50, 70, 90 and 100 percent proportion of training samples in the source domain as the source training dataset. To make the results fair, we repeat the above sampling 10 times for each sample size. NMI and ARI performances of NTR-FC-SCT on Subject1 and Subject2 with 5 percent noisy are shown in Figs. 10 and 11, respectively. The experimental results show that the values of NMI and ARI increase with the increase of the number of samples in the source domain. The reason is that NTR-FC-SCT can

TABLE 9
ARI Performance of all Comparison Methods on 15 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0009	0.0031	0.3389	0.1753	0.3607	0.0262	**0.6833**	0.0089
Subject2	0.1206	0.1058	0.3935	0.1414	0.4404	0.0126	**0.6679**	0.0107
Subject3	0.1083	0.0540	0.4127	0.1807	0.4323	0.0238	**0.5997**	0.0032
Subject4	0.1239	0.1680	0.3439	0.1746	0.3620	0.0344	**0.7034**	0.0137
Subject5	0.0847	0.0405	0.3859	0.1306	0.4408	0.0298	**0.7278**	0.0046
Subject6	0.1421	0.1255	0.4182	0.1430	0.4259	0.0204	**0.6636**	0.0077
Subject7	0.0639	0.1304	0.3588	0.1519	0.3662	0.0287	**0.5863**	0.0161
Subject8	0.0729	0.0824	0.3902	0.1349	0.4120	0.0355	**0.5671**	0.0136
Subject9	0.1302	0.0900	0.4174	0.1444	0.4186	0.0102	**0.6287**	0.0148
Subject10	0.0398	0.0848	0.3778	0.2008	0.4098	0.0241	**0.5256**	0.0122
Subject11	0.0593	0.1239	0.3843	0.2049	0.4230	0.0250	**0.6560**	0.0087
Subject12	0.1478	0.1185	0.3750	0.1692	0.4119	0.0303	**0.6518**	0.0128

TABLE 10
ARI Performance of all Comparison Methods on 20 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0019	0.0080	0.3440	0.1486	0.3241	0.0059	**0.6624**	0.0098
Subject2	0.1754	0.1208	0.3933	0.1522	0.3859	0.0059	**0.6005**	0.0116
Subject3	0.1443	0.0831	0.4147	0.1533	0.3742	0.0102	**0.5570**	0.0066
Subject4	0.0015	0.0060	0.3464	0.1713	0.3270	0.0090	**0.5008**	0.0139
Subject5	0.1679	0.0931	0.3883	0.1370	0.3730	0.0056	**0.6152**	0.0068
Subject6	0.1701	0.0682	0.4176	0.1304	0.3838	0.0091	**0.5775**	0.0079
Subject7	0.0625	0.1226	0.3602	0.1912	0.3309	0.0064	**0.4983**	0.0083
Subject8	0.1520	0.1066	0.3913	0.1339	0.3809	0.0088	**0.5055**	0.0102
Subject9	0.1619	0.0404	0.4164	0.1424	0.3871	0.0056	**0.5725**	0.0053
Subject10	0.0620	0.1394	0.3812	0.1927	0.3571	0.0027	**0.5494**	0.0065
Subject11	0.0713	0.1457	0.3838	0.1649	0.3695	0.0065	**0.5819**	0.0122
Subject12	0.0657	0.0780	0.3746	0.1617	0.3641	0.0048	**0.4376**	0.0106

TABLE 11
ARI Performance of all Comparison Methods on 25 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0023	0.0020	0.3409	0.1786	0.3419	0.0056	**0.5128**	0.0139
Subject2	0.0687	0.0687	0.3948	0.1626	0.4354	0.0132	**0.6609**	0.0156
Subject3	0.1447	0.0664	0.4126	0.1659	0.4043	0.0113	**0.5219**	0.0102
Subject4	0.0125	0.0047	0.3484	0.1856	0.3738	0.0138	**0.5349**	0.0204
Subject5	0.0701	0.0553	0.3862	0.1545	0.4182	0.0235	**0.6445**	0.0077
Subject6	0.1147	0.0392	0.4186	0.1502	0.4153	0.0305	**0.5271**	0.0151
Subject7	0.0076	0.0109	0.3611	0.1885	0.3654	0.0088	**0.5132**	0.0116
Subject8	0.0656	0.0666	0.3914	0.1499	0.4143	0.0176	**0.5902**	0.0148
Subject9	0.1248	0.0841	0.4173	0.1462	0.4233	0.0152	**0.6459**	0.0121
Subject10	0.1157	0.0035	0.3804	0.1958	0.3796	0.0046	**0.5953**	0.0165
Subject11	0.1382	0.1889	0.3842	0.1686	0.4277	0.0200	**0.6537**	0.0123
Subject12	0.2421	0.0869	0.3739	0.1749	0.4058	0.0177	**0.6031**	0.0091

TABLE 12
ARI Performance of all Comparison Methods on 30 percent Noisy CT Image Datasets

Dataset	FCM		TSC		T1-KT-FCM		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0008	0.0052	0.3401	0.1505	0.5432	0.0294	**0.7743**	0.0062
Subject2	0.1663	0.1180	0.3941	0.1505	0.5887	0.0309	**0.8122**	0.0136
Subject3	0.0947	0.0389	0.4132	0.1584	0.5700	0.0397	**0.7414**	0.0125
Subject4	0.0601	0.1242	0.3471	0.1760	0.5526	0.0094	**0.7876**	0.0146
Subject5	0.0867	0.0967	0.3886	0.1351	0.5858	0.0488	**0.7966**	0.0044
Subject6	0.1193	0.0774	0.4173	0.1744	0.5833	0.0131	**0.7834**	0.0053
Subject7	0.0675	0.1173	0.3611	0.1547	0.5637	0.0408	**0.7874**	0.0079
Subject8	0.0684	0.0490	0.3911	0.1728	0.6018	0.0277	**0.7363**	0.0144
Subject9	0.1227	0.0787	0.4184	0.1823	0.6060	0.0214	**0.8233**	0.0176
Subject10	0.0047	0.0076	0.3800	0.2041	0.5840	0.0421	**0.8140**	0.0134
Subject11	0.1420	0.1947	0.3839	0.2030	0.6010	0.0318	**0.8226**	0.0052
Subject12	0.0588	0.0640	0.3745	0.1792	0.5793	0.0106	**0.7996**	0.0093

TABLE 13
NMI Performance of all Comparison Methods on 5 percent Noisy CT Image Datasets

Dataset	FCM		NTR-FC-SCT ($\lambda_1 = 0$)		NTR-FC-SCT ($\lambda_2 = 0$)		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0112	0.0041	0.1312	0.0087	0.7009	0.0101	**0.7146**	0.0056
Subject2	0.1515	0.1468	0.4187	0.0121	0.7483	0.0135	**0.7526**	0.0142
Subject3	0.1583	0.1040	0.3852	0.0099	0.7011	0.0074	**0.7253**	0.0059
Subject4	0.0137	0.0062	0.1474	0.0137	0.7088	0.0062	**0.7256**	0.0138
Subject5	0.2333	0.1355	0.3566	0.0173	0.7304	0.0121	**0.7432**	0.0094
Subject6	0.1823	0.1094	0.4366	0.0075	0.7231	0.0071	**0.7441**	0.0116
Subject7	0.0086	0.0046	0.2283	0.0108	0.7184	0.0069	**0.7352**	0.0110
Subject8	0.1324	0.1157	0.4831	0.0078	0.7389	0.0082	**0.7586**	0.0108

TABLE 14
ARI Performance of all Comparison Methods on 5 percent Noisy CT Image Datasets

Dataset	FCM		NTR-FC-SCT ($\lambda_1 = 0$)		NTR-FC-SCT ($\lambda_2 = 0$)		NTR-FC-SCT	
	Means	Std	Means	Std	Means	Std	Means	Std
Subject1	0.0021	0.0043	0.4671	0.0069	0.8621	0.0042	**0.8880**	0.0028
Subject2	0.0854	0.0814	0.6893	0.0114	0.9017	0.0089	**0.9111**	0.0109
Subject3	0.0861	0.0521	0.7022	0.0093	0.8702	0.0047	**0.8839**	0.0031
Subject4	0.0013	0.0041	0.4705	0.0078	0.8856	0.0122	**0.8913**	0.0107
Subject5	0.1186	0.0877	0.6244	0.0086	0.8901	0.0099	**0.9080**	0.0078
Subject6	0.0839	0.0560	0.5921	0.0077	0.8901	0.0058	**0.8950**	0.0108
Subject7	0.0012	0.0026	0.5156	0.0143	0.8815	0.0103	**0.8922**	0.0119
Subject8	0.0651	0.0732	0.5702	0.0084	0.8977	0.0127	**0.9169**	0.0095

not mine enough transfer knowledge from source domain when training samples in the source domain are too few. On the other hand, exploiting clear and concise transfer knowledge need a certain amount of high quality samples in the source domain. Thus, it can be inferred that the more samples in the source domain, the more helpful the knowledge obtained in the source domain and the more efficient the NTR-FC-SCT will be in the target domain.

4.3 Parameter Sensitive

In the experiments, the parameters λ_1 and λ_2 are determined in a given search grid. In the following, we discuss the

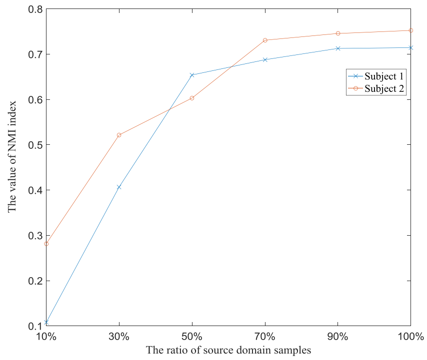


Fig. 10. NMI performance of NTR-FC-SCT with different proportion of samples in the source domain on 5 percent noisy Subject1 and Subject2

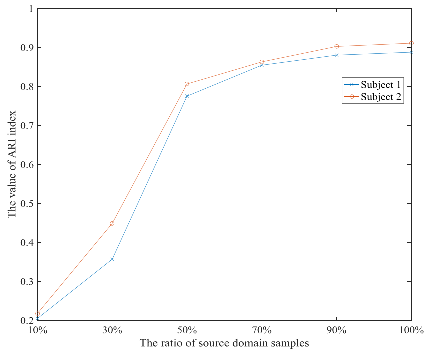


Fig. 11. ARI performance of NTR-FC-SCT with different proportion of samples in the source domain on 5 percent noisy Subject1 and Subject2

performance of NTR-FC-SCT using different parameters. Tables 15, 16 show the means of NMI and ARI on the subject using different λ_1 and λ_2, while fixing the parameter $m = 2$.

1) NTR-FC-SCT is sensitive to parameters λ_1 and λ_2. Different λ_1 and λ_2 lend to different cluster performance of NTR-FC-SCT in terms of NMI and ARI. It can be found that in most situations when the value of NMI is better, the value of ARI is also better. Thus, it is feasible to use NMI and ARI as performance criterions to determine the suitable parameters.
2) Fixed the value of m, NTR-FC-SCT obtains the worst NMI and ARI when $\lambda_1 = 0$ and $\lambda_2 = 0$. The clustering performance of NTR-FC-SCT is improved when λ_1 and λ_2 are not equal to 0. Since when $\lambda_1 = 0$ and $\lambda_2 = 0$ NTR-FC-SCT is degenerated to the classical FCM clustering.
3) We can find that when the value of λ_1 is large, NTR-FC-SCT obtains the satisfactory performance in terms of NMI and ARI. This further demonstrates that the proposed negative-transfer-resistance mechanism has played an effective role. Thus, in the subsequent experiments, we can reduce the search grid of λ_1 in the range $\{0, 10\mathrm{e}-4, \ldots, 10\mathrm{e}6\}$. We can't find the rule to select parameter λ_2. We think it is reasonable to select optimal λ_2 within the search grid. The range $\{0, 10\mathrm{e}-4, \ldots, 10\mathrm{e}6\}$ is appropriate.

5 Conclusion

In this study, we have addressed the problem of medical image segmentation with insufficient and noisy samples,

TABLE 15
Means of NMI by NTR-FC-SCT on the Subject1+5 Percent Noise Using Different λ_1 and λ_2, While Fixing $m = 2$

λ_2 \ λ_1	0	10e-4	10e-3	10e-2	10e-1	1	10e1	10e2	10e3	10e4	10e5	10e6
0	0.4801	0.5006	0.5181	0.5501	0.6011	0.6250	0.6091	0.6375	0.6788	0.6397	0.6427	0.6378
10e-4	0.4915	0.5326	0.5592	0.5662	0.6109	0.6161	0.6469	0.6499	0.6468	0.6493	0.6431	0.6425
10e-3	0.5094	0.5436	0.5572	0.5866	0.7007	0.6540	0.6413	0.6456	0.6465	0.6457	0.6449	0.6438
10e-2	0.5054	0.5605	0.5449	0.6980	**0.7146**	0.7006	0.6971	0.7075	0.6905	0.6766	0.7017	0.7106
10e-1	0.5036	0.5025	0.5036	0.5017	0.5070	0.5140	0.5138	0.5216	0.5184	0.5140	0.5147	0.5195
1	0.3540	0.3162	0.3118	0.3022	0.3340	0.3467	0.3500	0.3545	0.3517	0.3456	0.3630	0.3538
10e1	0.2511	0.2533	0.2691	0.2531	0.2482	0.2995	0.3116	0.3147	0.3098	0.3054	0.3077	0.3213
10e2	0.2104	0.2058	0.2204	0.2286	0.2035	0.2682	0.2794	0.2775	0.2786	0.2834	0.2765	0.2647
10e3	0.1872	0.1861	0.1964	0.1803	0.1866	0.2560	0.2657	0.2775	0.2550	0.2749	0.2682	0.2654
10e4	0.1727	0.1741	0.1636	0.1687	0.1650	0.2231	0.2297	0.2329	0.2270	0.2272	0.2208	0.2359
10e5	0.1425	0.1313	0.1294	0.1229	0.1282	0.1500	0.1597	0.1594	0.1544	0.1574	0.1565	0.1541
10e6	0.1386	0.1319	0.1385	0.1274	0.1313	0.1360	0.1375	0.1365	0.1325	0.1314	0.1391	0.1392

TABLE 16
Means of ARI by NTR-FC-SCT on the Subject 1+5 Percent Noise Using Different λ_1 and λ_2, While Fixing $m = 2$

λ_2 \ λ_1	0	10e-4	10e-3	10e-2	10e-1	1	10e1	10e2	10e3	10e4	10e5	10e6
0	0.7254	0.7452	0.7332	0.7778	0.7948	0.7880	0.7879	0.7876	0.7997	0.7905	0.7989	0.7845
10e-4	0.7534	0.7997	0.7590	0.8253	0.8656	0.8506	0.8670	0.8615	0.8579	0.8600	0.8460	0.8406
10e-3	0.7618	0.7943	0.8057	0.8553	0.8715	0.8733	0.8780	0.8717	0.8767	0.8755	0.8781	0.8769
10e-2	0.7586	0.8079	0.8055	0.8616	**0.8880**	0.8840	0.8840	0.8878	0.8837	0.8871	0.8849	0.8838
10e-1	0.6952	0.6842	0.6858	0.6788	0.7073	0.7172	0.7190	0.7139	0.7167	0.7215	0.7187	0.7135
1	0.6621	0.6716	0.6685	0.6733	0.6787	0.6788	0.6785	0.6797	0.6742	0.6813	0.6740	0.6748
10e1	0.6514	0.6555	0.6601	0.6612	0.6671	0.6615	0.6668	0.6665	0.6668	0.6647	0.6613	0.6649
10e2	0.6599	0.6637	0.6743	0.6744	0.6702	0.6724	0.6732	0.6746	0.6737	0.6720	0.6789	0.6771
10e3	0.6621	0.6625	0.6721	0.6583	0.6612	0.6633	0.6613	0.6650	0.6696	0.6696	0.6694	0.6666
10e4	0.6567	0.6497	0.6595	0.6536	0.6576	0.6562	0.6581	0.6597	0.6537	0.6537	0.6561	0.6512
10e5	0.6495	0.6397	0.6462	0.6357	0.6378	0.6446	0.6451	0.6448	0.6408	0.6426	0.6412	0.6388
10e6	0.6456	0.6403	0.6553	0.6402	0.6409	0.6506	0.6429	0.6419	0.6489	0.6466	0.6438	0.6439

and proposed NTR-FC-SCT model for leveraging source knowledge to improve the segmentation performance of target domain. We explore the negative-transfer-resistant mechanism to reinforce the influence of positive transfer and reduce, or even eliminate, the negative transfer. In particular, we find a shared latent space based on the idea of MMD, in which the mapped data distributions of source domain and target domain are close to each other. The experiments focus on noisy brain CT images. The experimental results show that with insufficient and noisy medical images, it is possible to build an efficient segmentation model with the help of medical images from the related scenarios. Future work will extend our algorithm to other medical image segmentation applications. We will extend the framework so as to apply various clustering algorithms in order to obtain more satisfactory medical image segmentation results. We will also study how many images in the source domain can be considered sufficient, and how to select the important images to further improve the transfer. In addition, how to speed up NTR-FC-SCT is worthy to be studied in the future.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61806026, 61702225, 61772241, and 61711540041, in part by the Natural Science Foundation of Jiangsu Province under Grants BK20180956, BK20161268, and BK20160187, in part by the Fundamental Re-search Funds for the Central Universities under Grants JUSRP51614A and JUSRP11737, in part by 2016 Qinglan Project of Jiangsu Province, and in part by Six Talent Peaks Project of Jiangsu Province under Grant XYDXX-127.

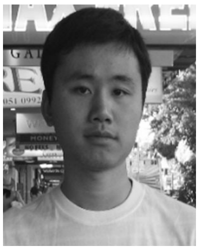
References

[1] A. Ribbens, J. Hermans, F. Maes, D. Vandermeulen, and P. Suetens, "Unsupervised segmentation, clustering, and groupwise registration of heterogeneous populations of brain MR images," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 201–224, Feb. 2014.

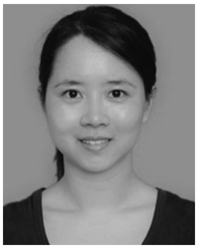
[2] Y. Ren, K. Hu, X. Dai, L. Pan, S. C. Hoi, and Z. Xu, "Semi-supervised deep embedded clustering. "*Neurocomputing*, vol. 325, no. 1, pp. 121–130, Jan. 2019.

[3] V. Borges, F. D. O. M. Cristina, T. Silva, A. Vieira, and B. Hamann, "Region growing for segmenting green microalgae images," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 1, pp. 257–270, Oct. 2018.

[4] A. Raj, A. K. Tiwari, and M. G. Martini, "Fundus image quality assessment: Survey, challenges, and future scope," *IET Image Process.*, vol. 13, no. 8, pp. 1–14, Apr. 2019.
[5] P. Poudel, A. Illanes, E. J. Ataide, N. Esmaeili, S. Balakrishnan, and M. Friebe, "Thyroid ultrasound texture classification using autoregressive features in conjunction with machine learning approaches," *IEEE Access*, vol. 7, pp. 79354–79365, Jun. 2019.
[6] G. Wang *et al.*, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018.
[7] S. Wazarkar, B. N. Keshavamurthy, and A. Hussain, "Region-based segmentation of social images using soft KNN algorithm." *Procedia Comput. Sci.*, vol. 125, no. 1, pp. 93–98, Jan. 2018.
[8] M. D. Steenwijk *et al.*, "Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs)," *Neuroimage Clin.*, vol. 3, pp. 462–469, Oct. 2013.
[9] A. Feng-Ping and L. Zhi-Wen, "Medical image segmentation algorithm based on feedback mechanism convolutional neural network," *Biomed. Signal Process. Control*, vol. 53, Aug. 2019, Art. no. 101589.
[10] X. Y. Wang, T. Wang, and J. Bu, "Color image segmentation using pixel wise support vector machine classification," *Pattern Recognit.*, vol. 44, no. 4, pp. 777–787, Apr. 2011.
[11] J. Enguehard, P. O'Halloran, and A. Gholipour, "Semi-supervised learning with deep embedded clustering for image classification and segmentation," *IEEE Access*, vol. 7, pp. 11093–11104, Jan. 2019.
[12] Z. Guo, X. Li, H. Huang, N. Guo, and Q. Li, "Deep learning-based image segmentation on multi-modal medical imaging," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 162–169, Mar. 2019.
[13] Y. Zhou, W. Huang, P. Dong, Y. Xia, and S. Wang, "D-UNET: A dimension-fusion U shape network for chronic stroke lesion segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, to be published, doi: 10.1109/TCBB.2019.2939522.
[14] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2-3, pp. 191–203, Dec. 1984.
[15] N. Dhanachandra and Y. J. Chanu, "A survey on image segmentation methods using clustering techniques," *Eur. J. Eng. Res. Sci.*, vol. 2, no. 1, pp. 15–20, Jan. 2017.
[16] N. M. Portela, G. D. Cavalcanti, and T. I. Ren, "Semi-supervised clustering for MR brain image segmentation," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1492–1497, Mar. 2014.
[17] A. Ortiz, J. M. Gorriz, J. Ramirez, and D. Salas-Gonzalez, "Improving MR brain image segmentation using self-organising maps and entropy-gradient clustering," *Inf. Sci.*, vol. 262, no. 20, pp. 117–136, Mar. 2014.
[18] S. Saha, A. K. Alok, and A. Ekbal, "Brain image segmentation using semi-supervised clustering," *Expert Syst. Appl.*, vol. 52, no. 15, pp. 50–63, Jun. 2016.
[19] E. Abdel-Maksoud, M. Elmogy, and R. Al-Awadi, "Brain tumor segmentation based on a hybrid clustering technique," *Egyptian Inform. J.*, vol. 16, no. 1, pp. 71–81, Mar. 2015.
[20] X. Wang, B. Qian, and I. Davidson, "On constrained spectral clustering and its applications," *Data Mining Knowl. Discovery*, vol. 28, no. 1, pp. 1–30, Jan. 2014.
[21] P. Qian *et al.*, "Cluster prototypes and fuzzy memberships jointly leveraged cross-domain maximum entropy clustering," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 181–193, Jan. 2016.
[22] H. Venkateswara, S. Chakraborty, and S. Panchanathan, "Deep-learning systems for domain adaptation in computer vision: Learning transferable feature representations," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 117–129, Nov. 2017.
[23] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, Mar. 2013.
[24] C. Williams, "Transfer in context: Replication and adaptation in knowledge transfer relationships," *Strategic Manage. J.*, vol. 28, no. 9, pp. 867–899, May 2007.
[25] W. Jiang and F. L. Chung, "Transfer spectral clustering," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2012, pp. 789–803.
[26] Z. Deng, Y. Jiang, F. L. Chung, H. Ishibuchi, K. S. Choi, and S. Wang, "Transfer prototype-based fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 5, pp. 1210–1232, Oct. 2016.
[27] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," Norwell, MA, USA: Kluwer Academic Publishers, 1981.
[28] P. Qian *et al.*, "Cross-domain, soft-partition clustering with diversity measure and knowledge reference," *Pattern Recognit.*, vol. 50, no. 2, pp. 155–177, Feb. 2016.
[29] P. Qian *et al.*, "Knowledge-leveraged transfer fuzzy C-Means for texture image segmentation with self-adaptive cluster prototype matching," *Knowl.-Based Syst.*, vol. 130, no. 8, pp. 33–50, May. 2017.
[30] A. Charuvaka and H. Rangwala, "Classifying protein sequences using regularized multi-task learning," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 11, no. 6, pp. 1087–1098. Nov. 2014.
[31] W. Gu, Z. Zhang, X. Xie, and Y. He, "An improved muti-task learning algorithm for analyzing cancer survival data," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, to be published, doi: 10.1109/TCBB.2019.2920770
[32] S. F. Hussain and M. Ramazan, "Biclustering of human cancer microarray data using co-similarity based co-clustering," *Expert Syst. Appl.*, vol. 55, no. 8, pp. 520–531, Aug. 2016.
[33] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[34] X. Zhang, X. Zhang, H. Liu, and X. Liu, "Multi-task clustering through instances transfer," *Neurocomputing*, vol. 251, pp. 145–155, Apr. 2017.
[35] A. Van Opbroek, H. C. Achterberg, M. W. Vernooij, and M. De Bruijne, "Transfer learning for image segmentation by combining image weighting and kernel learning," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 213–224, Jan. 2019.
[36] M. T. Rosenstein, Z. Marx, L. P. Kaelbling, and T. G. Dietterich, "To transfer or not to transfer," in *Proc. Neural Inf. Process. Syst. Workshop Inductive Transfer: 10 Years Later*, 2005, pp. 1–4.
[37] T. Croonenborghs, K. Driessens, and M. Bruynooghe, "Learning relational options for inductive transfer in relational reinforcement learning," in *Proc. Int. Conf. Inductive Logic Program.*, 2007, pp. 88–97.
[38] E. Talvitie and S. P. Singh, "An experts algorithm for transfer learning," in *Proc. 20th Int. Joint Conf. Artif. Intell.*, 2007, pp. 1065–1070.
[39] M. T. Hooijmans *et al.*, "Fast multistation water/fat imaging at 3T using DREAM-based RF shimming," *J. Magn. Reson. Imag.*, vol. 42, no. 1, pp. 217–223, Oct. 2014.
[40] A. Kalemis, B. M. Delattre, and S. Heinzer, "Sequential whole-body PET/MR scanner: Concept, clinical use, and optimisation after two years in the clinic. The manufacturer's perspective," *Magn. Reson. Materials Phys. Biol. Med.*, vol. 26, no. 1, pp. 5–23, Feb. 2013.
[41] A. van Opbroek, M. A. Ikram, M. W. Vernooij, and M. de Bruijne, "Supervised image segmentation across scanner protocols: A transfer learning approach," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2012, pp. 160–167.
[42] F. Nie, D. Xu, and X. Li, "Initialization independent clustering with actively self-training method," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 42, no. 1, pp. 17–27, Feb. 2012.
[43] J. Liu, J. Mohammed, J. Carter, S. Ranka, T. Kahveci, and M. Baudis, "Distance-based clustering of CGH data," *Bioinformatics*, vol. 22, no. 16, pp. 1971–1978, Aug. 2006.



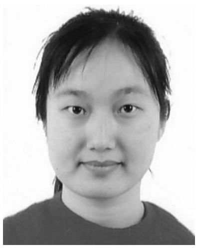
Yizhang Jiang (M'12–SM'19) received the PhD degree from Jiangnan University, in 2015. He has also been a research assistant in the Department of Computing, Hong Kong Polytechnic University, for two years. His research interests include pattern recognition, intelligent computation and their applications. He is the author/co-author of more than 40 research papers in international/national journals, including the *IEEE Trans. on Fuzzy Systems*, *IEEE Trans. on Neural Networks and Learning Systems*, *IEEE Trans. on Cybernetics*, *and Information Sciences*. He is an associate editor of the *IEEE Access* (2019-). He has served as a reviewer or co-reviewer of several international conferences and journals, such as ICDM, *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Fuzzy Systems*, the *IEEE Transactions on Neural Networks and Learning Systems*, *Pattern Recognition*, *Neurocomputing*, and *Neural Computing & Applications*. He has also served as a leader guest editor or guest editor of several international journals, such as journal of the *Ambient Intelligence and Humanized Computing*, *Computational and Mathematical Methods in Medicine*, and *Frontiers in Neuroscience*. He is a Senior member of the IEEE.



Xiaoqing Gu received the PhD degree from Jiangnan University, in 2017. She is currently an associate professor with the School of Information Science and Engineering, Changzhou University, Changzhou, China. She has published more than 30 papers in international/national authoritative journals. Her current research interests include pattern recognition and machine learning.

Dongrui Wu (S'05–M'09–SM'14) received the BE degree in automatic control from the University of Science and Technology of China, in 2003, the ME degree in electrical engineering from the National University of Singapore, in 2005, and the PhD degree in electrical engineering from the University of Southern California, in 2009. He is currently a professor with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China, and the deputy director of the Key Laboratory of Image Processing and Intelligent Control, Ministry of Education. His research interests include affective computing, brain computer interfaces, computational intelligence, and machine learning. He has more than 120 publications, including a book entitled Perceptual Computing (Wiley-IEEE Press, 2010). He received the IEEE Computational Intelligence Society Outstanding PhD Dissertation Award, in 2012, *IEEE TRANSACTIONS ON FUZZY SYSTEMS* Outstanding Paper Award, in 2014, NAFIPS Early Career Award, in 2014, *IEEE Systems, Man and Cybernetics (SMC)* Society Early Career Award, in 2017, and the IEEE SMC Society Best Associate Editor Award, in 2018. He was also a finalist of another three Best Paper Awards. He was/is an associate editor of the *IEEE TRANSACTIONS ON FUZZY SYSTEMS* (2011-2018), *IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS* (2014-), *IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE* (2017-), and *IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING* (2019). He is a Senior member of the IEEE.

Wenlong Hang received the PhD degree from Jiangnan University, in 2017. He is currently a lecturer with the School of Computer Science and Technology, Nanjing Tech University, Nanjing, China. He has published more than 10 papers in international/national authoritative journals. His current research interests include pattern recognition and machine learning.



Jing Xue received the MM degree from the Department of Nephrology, Nanjing Medical University, China. She received the MD (doctor of medicine) degree from the Department of Nephrology, Nanjing Medical University, China, in 2019. Her research interests include machine learning and its applications on medicine. She is a doctor with the Affiliated Wuxi People's Hospital of Nanjing Medical University, China.

Shi Qiu received the MS degree from Xidian University, Xi'an, P. R. China, in 2012, the PhD degree from the University of Chinese Academy of Sciences, Xi'an, P. R. China, in 2016. He is currently an associate research fellow with the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, P. R. China. His research interests include image restoration, sparse representation, and object detection.

Chin-Teng Lin (S'88–M'91–SM'99–F'05) received the BS degree in control engineering from the National Chiao-Tung University (NCTU), Hsinchu, Taiwan, in 1986, and the master's and PhD degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989 and 1992, respectively. He is currently a distinguished professor of the Faculty of Engineering and Information Technology, University of Technology Sydney. He also holds an Honorary chair professorship with Electrical and Computer Engineering, NCTU, International Faculty of University of California at San-Diego, La Jolla, CA, USA, and an honorary professorship with the University of Nottingham, Nottingham, UK. He is the co author of Neural Fuzzy Systems (Prentice-Hall, 1996), and the author of Neural Fuzzy Control Systems with Structure and Parameter Learning (World Scientific, 1994). He has published more than 220 journal papers (Total Citation: 18 236, H-index: 57, i10-index: 295) and holds 97 patents in the areas of computational intelligence, fuzzy neural networks, natural cognition, brain–computer interface, intelligent system, multimedia information processing, machine learning, robotics, and intelligent sensing and control, including approximately 108 IEEE journal papers. He was elevated to a fellow of the IEEE for his contributions to biologically inspired information systems, in 2005, and was elevated to an International Fuzzy Systems association fellow, in 2012. He received the Merit National Science Council Research Fellow Award, Taiwan, in 2009, Outstanding Electrical and Computer Engineer, Purdue University, in 2011, Outstanding Achievement Award by Asia Pacific Neural Network Assembly, in 2013, and IEEE Fuzzy Systems Pioneer Award, in 2016. He served as the deputy editor-in-chief of the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II* in 2006–2008 and as editor in-chief of the *IEEE Transactions on Fuzzy Systems* from 2011 to 2016. He also served on the Board of Governors of the IEEE Systems, Man, Cybernetics Society during 2003–2005, IEEE Circuits and Systems (CAS) Society during 2005–2008, and IEEE Computational Intelligence Society (CIS) during 2008–2010. He was the chair of the IEEE Taipei Section during 2009–2010. He was a distinguished lecturer of the IEEE CAS Society from 2003 to 2005 and the CIS Society from 2015 to 2017. He was the program chair of the IEEE International Conference on Systems, Man, and Cybernetics, in 2005 and the general chair of the 2011 IEEE International Conference on Fuzzy Systems.