

A NOVEL IMAGE-SPECIFIC TRANSFER APPROACH FOR PROSTATE SEGMENTATION IN MR IMAGES

Pinzhuo Tian¹, Lei Qi¹, Yinghuan Shi¹, Luping Zhou², Yang Gao¹, Dinggang Shen³

¹State Key Laboratory for Novel Software Technology, Nanjing University, China

²School of Computing and Information Technology, University of Wollongong, Australia

³Department of Radiology and BRIC, UNC Chapel Hill, USA

ABSTRACT

Prostate segmentation in Magnetic Resonance (MR) Images is a significant yet challenging task for prostate cancer treatment. Most of the existing works attempted to design a global classifier for all MR images, which neglect the discrepancy of images across different patients. To this end, we propose a novel transfer approach for prostate segmentation in MR images. Firstly, an image-specific classifier is built for each training image. Secondly, a pair of dictionaries and a mapping matrix are jointly obtained by a novel Semi-Coupled Dictionary Transfer Learning (SCDTL). Finally, the classifiers on the source domain could be selectively transferred to the target domain (*i.e.* testing images) by the dictionaries and the mapping matrix. The evaluation demonstrates that our approach has a competitive performance compared with the state-of-the-art transfer learning methods. Moreover, the proposed transfer approach outperforms the conventional deep neural network based method.

Index Terms— Prostate Segmentation, Image-Specific Transfer Approach, Semi-Coupled Dictionary Transfer Learning.

1. INTRODUCTION

Nowadays, prostate cancer is considered as one of the most leading diseases for cancer-related death for males over the world. According to the report from the National Cancer Institute, in 2017 about 161,360 new cases are estimated and 26,730 deaths are caused by prostate cancer in the US ¹. In order to cure prostate cancer, the cancer tissues should be precisely killed by high-energy X-rays from different directions in clinical practice. Therefore, prostate segmentation in prostate MR image for accurate localization is clinically significant. However, prostate segmentation by asking the physician to provide the manual delineation is very time-consuming and laborious. Thus, in recent years, there are

increasing demands to develop computer-aided segmentation methods for localizing the prostate region in MR images.

Recently, many efforts for developing automatic or semi-automatic methods for prostate segmentation in MR images have been presented [1–4]. Previous automatic prostate segmentation methods could be roughly classified into non-learning methods (*e.g.* multi-atlas methods [5], the deformable methods [4]) and learning-based methods (*e.g.* using marginal space learning [6], feature learning [7] and deep convolution neural networks [1, 3, 8]). In particular, learning-based methods from a data-driven perspective have obtained more attention in prostate segmentation.

Most existing methods focus on learning one unique global model on all training MR images. However, the prostate MR images of different patients are largely different in appearance, shape and size, *etc.*, which lead to a large data distribution discrepancy. Even for a same patient, the same prostate region may appear very different in different treatment days, due to irregular, unpredictable prostate motion and shape distortion, *etc* [7]. These conventional learning paradigms, which normally train a global model by collecting all the samples from the training images, might suffer from the difference in data distributions across different patients (domains). This might also pose a major obstacle in adapting predictive model for the target task [9]. Therefore, one global model cannot perform well on each MR image. Basically, the goal of transfer learning is to eliminate or relieve the data distribution discrepancy between different domains [10–12]. In this case, the model learned on the source domain could perform well on the target domain. Thus, this could be naturally borrowed to tackle the data distribution discrepancy issue in prostate segmentation.

2. OUR METHOD

In this paper, we propose a novel image-specific transfer model for prostate segmentation in MR images. The framework of the proposed approach is illustrated in Fig. 1. Firstly, an individual image-specific classifier is trained for each training MR image. Secondly, the proposed Semi-Coupled Dictionary Transfer Learning (SCDTL) algorithm is used to learn a pair of dictionaries and a mapping matrix. By the dictionaries and the mapping matrix, the correspondence between image fea-

This work was supported by NSFC (61432008, 61673203), NIH Grant (CA206100), ARC (DE160100241), and Young Elite Scientists Sponsorship Program by CAST (YESS 20160035).

¹<https://seer.cancer.gov/statfacts/html/prost.html>

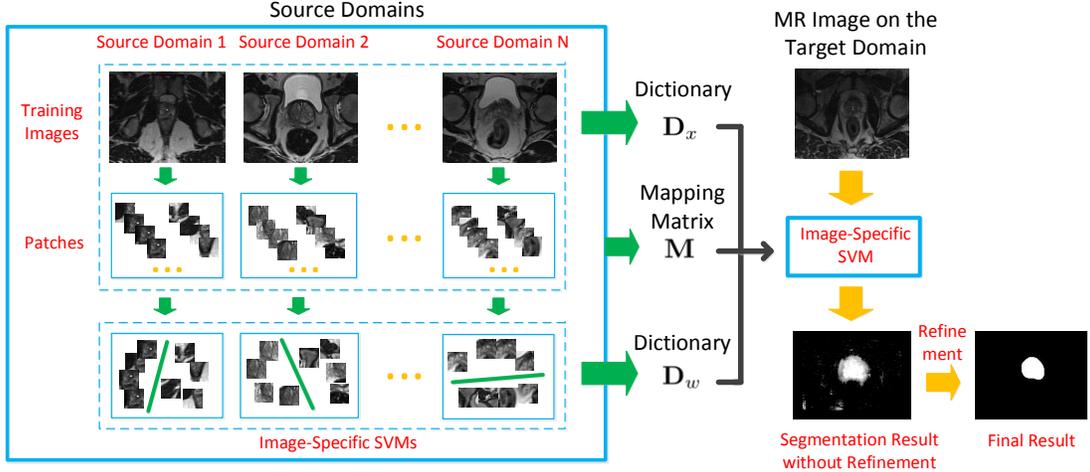


Fig. 1. The framework of the proposed method for prostate segmentation in MR images.

ture space and classifier weight space could be established. Thus, the classifier on source domain could be transferred to target domain. Finally, we perform the result refinement to obtain the final results.

Image Preprocessing. Since our prostate segmentation method is performed on patches of MR images [7, 8, 13], we extract the patches from each MR image on the source domain. For each patch, its label is assigned according to the label of the center voxel. Typically, if the center voxel is in prostate region, the label of this patch is set to 1 (positive), otherwise, it will be set to -1 (negative). The size of each patch is 32×32 . Considering the fact that non-prostate region is much larger than prostate region, we employ the sampling method [8] to roughly get equal amounts of positive and negative patches. Specifically, about $300 \sim 400$ training patches are sampled in each training slice according to the size of prostate region, thus about $6 \times 10^3 \sim 1.6 \times 10^4$ training patches are sampled for each patient.

Image-Specific SVM Learning. Given the training image set $\mathcal{X}_s = \{\mathbf{x}_i\}_{i=1}^N$, where $\mathbf{x}_i \in \mathbb{R}^m$ represents the i -th image, each MR image is treated as a single domain in this paper. Because of the data distribution discrepancy, each image (each domain) should have a specific classifier.

Inspired by the existing learning-based methods [7, 14], we attempt to judge the location of each voxel in MR image by the label of patch. Based on a classifier \mathcal{F} , if a patch is predicted to be positive, the center voxel of this patch is in the prostate region. Otherwise, it will not belong to the prostate region.

Considering that the linear SVM [15] is always with the good interpretability [16], we model the image-specific SVM for the i -th image as follows:

$$\begin{aligned} \min_{\mathbf{w}_i, b, \xi_j} \quad & \frac{1}{2} \|\mathbf{w}_i\|^2 + C \sum_{j=1}^h \xi_j, \\ \text{s.t.} \quad & y_j (\mathbf{w}_i^T \mathbf{p}_{ij} + b) \geq 1 - \xi_j, \\ & \xi_j \geq 0, \quad j = 1, 2, \dots, h. \end{aligned} \quad (1)$$

where $C \in \mathbb{R}$ is a regularization parameter, and $\xi_j \in \mathbb{R}$ denotes the slack variable, \mathbf{p}_{ij} is the feature vector of the j -th patch of the i -th image, respectively.

Note that, by solving Eq. (1), we could obtain the specific classifier weight vector $\mathbf{w}_i \in \mathbb{R}^l$ for the i -th image (*i.e.* \mathbf{x}_i and its SVM weight vector \mathbf{w}_i is one to one correspondence). Basically, similar images should have similar directions of the weight vectors, thus, the similarity of different images could be used to guide to transfer the image-specific classifiers from multiple source domains to the target domain.

SCDTL. Our goal is to infer the weight vector for each image on the target domain according to its image feature, which can be realized by learning an intrinsic relationship between image feature space and classifier weight space. Inspired by [16–18], in this paper, the special relationship between source and target domain is constructed by dictionary learning. Our SCDTL aims to learn a pair of dictionaries in feature space and classifier space, as well as a mapping matrix between the two aforementioned dictionaries. In particular, the two dictionaries could represent the intrinsic structures of the two spaces, and also the mapping matrix characterizes the relationship between them. Thus, an image-specific classification model on the target domain can be obtained by these dictionaries and the mapping matrix.

We introduce $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{m \times N}$ to represent the training feature matrix with each column denoting a feature vector of one training image. $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N] \in \mathbb{R}^{l \times N}$ indicates the matrix of classifier weight, where a weight vector in the \mathbf{W} corresponds to an image. $\mathbf{D}_x \in \mathbb{R}^{m \times k}$ and $\mathbf{D}_w \in \mathbb{R}^{l \times k}$ are feature dictionary and classifier weight dictionary, respectively. $\mathbf{M} \in \mathbb{R}^{k \times k}$ is the mapping matrix. k denotes the dictionary size.

Following the assumption that the dictionaries are over-complete, many existing methods [17, 18] impose ℓ_1 -norm regularization for coding coefficients to choose a few atoms from the learned dictionary to describe a sample. In addition, since feature dimension is usually much larger than the

number of images, Frobenius norm regularization is adopted to capture the correlation structure of data with large variations [16]. However, the aforementioned methods assume that feature space and classifier weight space have the same dimensionality. In prostate segmentation, the feature dimensionality of the whole image is much larger than the number of images, while the dimensionality of classifier weight based on patches is smaller than the number of image. It is not suitable for all coding coefficients to apply a same constraint. Thus, our proposed SCDTL is mathematically formulated as follows:

$$\begin{aligned} \min_{\{\mathbf{D}_x, \mathbf{D}_w, \mathbf{M}\}} & \|\mathbf{X} - \mathbf{D}_x \mathbf{\Lambda}_x\|_F^2 + \|\mathbf{W} - \mathbf{D}_w \mathbf{\Lambda}_w\|_F^2 + \\ & \gamma \|\mathbf{\Lambda}_w - \mathbf{M} \mathbf{\Lambda}_x\|_F^2 + \lambda \|\mathbf{\Lambda}_x\|_F^2 + \lambda \|\mathbf{\Lambda}_w\|_1 \\ & + \lambda \|\mathbf{M}\|_F^2, \quad (2) \\ \text{s.t. } \forall i, & \|\mathbf{d}_{x,i}\| \leq 1, \|\mathbf{d}_{w,i}\| \leq 1. \end{aligned}$$

where $\mathbf{\Lambda}_x \in \mathbb{R}^{k \times N}$ and $\mathbf{\Lambda}_w \in \mathbb{R}^{k \times N}$ denote the coding coefficients of feature and weight dictionaries, respectively. $\|\cdot\|_F$ is the Frobenius norm to obtain the correlation structure of data with large variations, and $\|\cdot\|_1$ is ℓ_1 -norm to impose the coding coefficients to be sparse. γ and λ are defined as two regularization parameters to balance the terms in Eq. (2), respectively. $\mathbf{d}_{x,i} \in \mathbb{R}^m$ and $\mathbf{d}_{w,i} \in \mathbb{R}^l$ denote the atoms of \mathbf{D}_x and \mathbf{D}_w , respectively.

Note that Eq. (2) is not jointly convex to $\mathbf{D}_x, \mathbf{D}_w, \mathbf{M}$. However, the objective function *w.r.t.* only one variable is a convex function. Therefore, the problem can be solved by the alternating optimization strategy.

- **Update $\mathbf{\Lambda}_x$:** After parameter initialization, $\mathbf{M}, \mathbf{D}_x, \mathbf{D}_w, \mathbf{M}$ and $\mathbf{\Lambda}_w$ are fixed to calculate the coding coefficients $\mathbf{\Lambda}_x$ as follows:

$$\min_{\{\mathbf{\Lambda}_x\}} \|\mathbf{X} - \mathbf{D}_x \mathbf{\Lambda}_x\|_F^2 + \gamma \|\mathbf{\Lambda}_w - \mathbf{M} \mathbf{\Lambda}_x\|_F^2 + \lambda \|\mathbf{\Lambda}_x\|_F^2. \quad (3)$$

Note that Eq. (3) could be solved by its closed-form solution, as:

$$\mathbf{\Lambda}_x = (\mathbf{D}_x^T \mathbf{D}_x + \gamma \mathbf{M}^T \mathbf{M} + \lambda \mathbf{I})^{-1} (\mathbf{D}_x^T \mathbf{X} + \gamma \mathbf{M}^T \mathbf{\Lambda}_w). \quad (4)$$

- **Update $\mathbf{\Lambda}_w$:** After solving Eq. (3), we fix $\mathbf{D}_x, \mathbf{D}_w, \mathbf{M}$ and $\mathbf{\Lambda}_x$ to calculate the sparse coding coefficients $\mathbf{\Lambda}_w$ as follows:

$$\min_{\{\mathbf{\Lambda}_w\}} \|\mathbf{W} - \mathbf{D}_w \mathbf{\Lambda}_w\|_F^2 + \gamma \|\mathbf{M} \mathbf{\Lambda}_x - \mathbf{\Lambda}_w\|_F^2 + \lambda \|\mathbf{\Lambda}_w\|_1. \quad (5)$$

Eq. (5) is a multi-task LASSO problem, thus we perform LARS [19] for its solution.

- **Update \mathbf{D}_x and \mathbf{D}_w :** We fix $\mathbf{\Lambda}_x, \mathbf{\Lambda}_w$ and \mathbf{M} to optimize the dictionary pair \mathbf{D}_x and \mathbf{D}_w , which could be formulated as:

$$\begin{aligned} \min_{\{\mathbf{D}_x, \mathbf{D}_w\}} & \|\mathbf{X} - \mathbf{D}_x \mathbf{\Lambda}_x\|_F^2 + \|\mathbf{W} - \mathbf{D}_w \mathbf{\Lambda}_w\|_F^2 \\ \text{s.t. } \forall i, & \|\mathbf{d}_{x,i}\| \leq 1, \|\mathbf{d}_{w,i}\| \leq 1. \end{aligned} \quad (6)$$

Algorithm 1 The Optimization of SCDTL

Input: image matrix \mathbf{X} , classifier weight matrix \mathbf{W} , parameters γ, λ .

Output: feature dictionary \mathbf{D}_x , weight dictionary \mathbf{D}_w and mapping matrix \mathbf{M} .

Initialize: $\mathbf{D}_x, \mathbf{D}_w, \mathbf{M}, \mathbf{\Lambda}_x$ and $\mathbf{\Lambda}_w$.

Repeat:

- 1: Fix $\mathbf{D}_x, \mathbf{D}_w, \mathbf{M}, \mathbf{\Lambda}_w$, update $\mathbf{\Lambda}_x$ by Eq. (4).
- 2: Fix $\mathbf{D}_x, \mathbf{D}_w, \mathbf{M}, \mathbf{\Lambda}_x$, update $\mathbf{\Lambda}_w$ by Eq. (5).
- 3: Fix $\mathbf{\Lambda}_x, \mathbf{\Lambda}_w, \mathbf{M}$, update $\mathbf{D}_x, \mathbf{D}_w$ by Eq. (6).
- 4: Fix $\mathbf{D}_x, \mathbf{D}_w, \mathbf{\Lambda}_x, \mathbf{\Lambda}_w$, update \mathbf{M} by Eq. (8).

Until: convergency.

Eq. (6) is a quadratically constrained quadratic program problem (QCQP). Inspired by [20], we adopt a column by column update strategy to solve Eq. (6).

- **Update \mathbf{M} :** Finally, with the coding coefficients and dictionaries fixed, we update the mapping matrix \mathbf{M} as:

$$\min_{\{\mathbf{M}\}} \|\mathbf{\Lambda}_w - \mathbf{M} \mathbf{\Lambda}_x\|_F^2 + \frac{\lambda}{\gamma} \|\mathbf{M}\|_F^2. \quad (7)$$

Similar to Eq. (3), Eq. (7) can also be solved by a closed-form solution, as:

$$\mathbf{M} = \mathbf{\Lambda}_w \mathbf{\Lambda}_x^T (\mathbf{\Lambda}_x \mathbf{\Lambda}_x^T + \frac{\lambda}{\gamma} \mathbf{I})^{-1}. \quad (8)$$

The whole optimization algorithm for solving SCDTL is summarized in Algorithm 1.

Transferring Classifiers to Testing Images. With the obtained dictionary pair $\mathbf{D}_x, \mathbf{D}_w$ and the mapping matrix \mathbf{M} , given a prostate MR image \mathbf{x}_t on the target domain, its corresponding classifier weight vector \mathbf{w}_t can be obtained by the following steps.

- **Step 1:** The coding coefficient $\alpha_{x_t} \in \mathbb{R}^k$ of \mathbf{x}_t can be obtained as:

$$\min_{\{\alpha_{x_t}\}} \|\mathbf{x}_t - \mathbf{D}_x \alpha_{x_t}\|_F^2 + \lambda \|\alpha_{x_t}\|_F^2. \quad (9)$$

The optimal solution of Eq. (9) is calculated as:

$$\alpha_{x_t} = (\mathbf{D}_x^T \mathbf{D}_x + \lambda \mathbf{I})^{-1} \mathbf{D}_x^T \mathbf{x}_t. \quad (10)$$

After α_{x_t} is obtained, the coding vector $\alpha_{w_t} \in \mathbb{R}^k$ of \mathbf{w}_t is derived by:

$$\alpha_{w_t} = \mathbf{M} \alpha_{x_t}. \quad (11)$$

- **Step 2:** The classifier weight vector \mathbf{w}_t of \mathbf{x}_t can be inferred:

$$\mathbf{w}_t = \mathbf{D}_w \alpha_{w_t}. \quad (12)$$

- **Step 3:** The specific weight vector w_t of the MR image x_t could be used to predict the labels of all the patches from testing image x_t . Through D_x , D_w and M , the classifier of each MR image on the source domain is selectively transferred to the target domain.

To obtain the binary segmentation results, we employ the level-set [21] model to refine the segmentation results.

3. EXPERIMENT

In the experiment, 3D series prostate MR images from 22 different patients are employed to validate our method. Each patient contains 20 ~ 40 2D MR image slices. Meanwhile, all MR images have already been manually delineated by the physicians which can be treated as ground-truth. In our experiment, we perform 2-fold cross validation for segmenting the prostate MR images of the 22 patients. Specifically, in each cross validation, the prostate MR images from 11 patients are used as source domains and the rest of the MR images are considered as target domains. The two criteria including average Dice similarity coefficient (DSC) which is the most popular metric in tissue segmentation [1, 2, 7] and centroid distance (CD) along from two directions are employed as the evaluation metrics.

In the experiment, we compared the proposed method with several the state-of-the-art transfer learning paradigms including Laplacian SVM (LapSVM) [12], Transfer Component Analysis (TCA) [10] and Subspace Alignment (SA) [11]. These 3 methods obtaining competitive performance in different fields are designed to deal with transfer problem from different views including feature, subspace and model, respectively. Specifically, since the time complexity of LapSVM, TCA and SA is more than quadratic with the number of samples, we randomly sampled 10000 patches from the source domains as labeled data and all the patches from one testing image as unlabeled data to obtain LapSVM, TCA-SVM and SA-SVM for this testing MR image on the target domain. For all these methods, we adopt level-set model [21] to refine the results. Also, linear kernel, deep features are used for all the methods.

In addition, our method is compared with a global deep model, *LeNet* [22]. The patches sampled from all the source domains (about $6.6 \times 10^4 \sim 1.7 \times 10^5$) are utilized to train the global *LeNet*. Specifically, we use the global *LeNet* and *VGG16* model [23] trained on the *ImageNet*² as patch feature extractor and whole image feature extractor, respectively. Also, we set the dictionary size to 150.

The results of all the experiments are reported in Table 1. Fig. 2 shows the typical results of different methods. From Table 1, we could observe: (1) Due to the time complexity of LapSVM, TCA-SVM and SA-SVM, these transfer methods could not fully use all the information on the source domain. Moreover, they cannot selectively transfer useful knowledge from source to target domain. This flaw leads to the negative transfer. Thus, the global *LeNet* outperforms these transfer

Table 1. Evaluation results of all the methods. Note that highlighted values represent the best results.

Method	Dice	CD-x(mm)	CD-y(mm)
LapSVM [12]	0.80 (0.05)	1.10 (0.54)	1.12 (0.81)
TCA-SVM [10]	0.79 (0.04)	1.13 (0.58)	1.18 (1.07)
SA-SVM [11]	0.71 (0.10)	3.79 (1.83)	2.00 (0.75)
<i>LeNet</i> [22]	0.81 (0.02)	1.08 (0.62)	1.06 (0.81)
SCDTL	0.83 (0.02)	0.73 (0.46)	1.06 (0.75)

methods. (2) Although deep model could relieve the discrepancy between different domains, it cannot totally eliminate the issue. However, our proposed method leverages the model information of similar image to largely eliminate the data distribution discrepancy between source and target domains. Meanwhile, with the deep feature, the performance of our method is better than the specific *LeNet* on the target domain. (3) What's more, our method could selectively transfer the knowledge from many different source domains, which effectively avoids the possible negative transfer. These aforementioned analyses support that the performance of SCDTL is better than the mentioned state-of-the-art transfer paradigms.

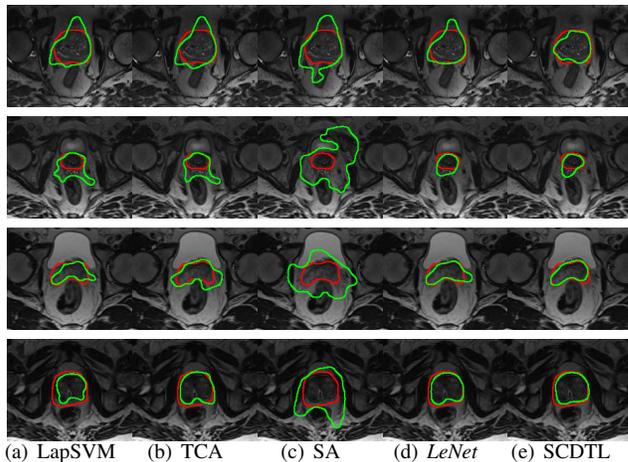


Fig. 2. The visualized results of the proposed method and baselines. Note that the red line represents groundtruth and the green line denotes segmentation result.

4. CONCLUSION

In this paper, inspired by the observation that prostate regions in different MR images show their specific characteristic, we propose a novel image-specific transfer approach for prostate segmentation. By our proposed SCDTL, a pair of dictionaries for image feature space and classifier weight space, and the mapping matrix between them are obtained. According to the obtained dictionaries and the mapping matrix, the classifiers of images on the source domain can be selectively transferred to the target domain with less negative transfer. Experimental results show the proposed method performs better than several state-of-the-art transfer methods. Furthermore, our method is more effective than the global *LeNet* model.

²<http://image-net.org/index>

References

- [1] Lequan Yu, Xin Yang, Hao Chen, Jing Qin, and Pheng-Ann Heng, “Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images,” in *AAAI*, 2017, pp. 66–72.
- [2] Yanrong Guo, Yaozong Gao, and Dinggang Shen, “Deformable mr prostate segmentation via deep feature learning and sparse patch matching,” in *Deep Learning for Medical Image Analysis*, pp. 197–222. Elsevier, 2017.
- [3] Tyler Clark, Alexander Wong, Masoom A Haider, and Farzad Khalvati, “Fully deep convolutional neural networks for segmentation of the prostate gland in diffusion-weighted mr images,” in *International Conference Image Analysis and Recognition*. Springer, 2017, pp. 97–104.
- [4] Zhiqiang Tian, LiZhi Liu, and Baowei Fei, “A supervoxel-based segmentation method for prostate mr images,” in *Medical Imaging 2015: Image Processing*. International Society for Optics and Photonics, 2015, vol. 9413, p. 941318.
- [5] Stefan Klein, Uulke A Van Der Heide, Irene M Lips, Marco Van Vulpen, Marius Staring, and Josien PW Pluim, “Automatic segmentation of the prostate in 3d mr images by atlas matching using localized mutual information,” *Medical physics*, vol. 35, no. 4, pp. 1407–1417, 2008.
- [6] Yefeng Zheng and Dorin Comaniciu, *Marginal Space Learning for Medical Image Analysis*, Springer, 2014.
- [7] Yinghuan Shi, Yaozong Gao, Shu Liao, Daoqiang Zhang, Yang Gao, and Dinggang Shen, “Semi-automatic segmentation of prostate in ct images via coupled feature representation and spatial-constrained transductive lasso,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 11, pp. 2286–2303, 2015.
- [8] Yinghuan Shi, Wanqi Yang, Yang Gao, and Dinggang Shen, “Does manual delineation only provide the side information in ct prostate segmentation?,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 692–700.
- [9] Sinno Jialin Pan and Qiang Yang, “A survey on transfer learning,” *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [10] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang, “Domain adaptation via transfer component analysis,” *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.
- [11] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars, “Unsupervised visual domain adaptation using subspace alignment,” in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2960–2967.
- [12] Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani, “Manifold regularization: A geometric framework for learning from labeled and unlabeled examples,” *Journal of machine learning research*, vol. 7, no. Nov, pp. 2399–2434, 2006.
- [13] Shu Liao, Yaozong Gao, Jun Lian, and Dinggang Shen, “Sparse patch-based label propagation for accurate prostate localization in ct images,” *IEEE transactions on medical imaging*, vol. 32, no. 2, pp. 419–434, 2013.
- [14] Yaozong Gao, Shu Liao, and Dinggang Shen, “Prostate segmentation by sparse representation based classification,” *Medical physics*, vol. 39, no. 10, pp. 6372–6387, 2012.
- [15] Corinna Cortes and Vladimir Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [16] Ying Zhang, Baohua Li, Huchuan Lu, Atshushi Irie, and Xiang Ruan, “Sample-specific svm learning for person re-identification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1278–1287.
- [17] Shenlong Wang, Lei Zhang, Yan Liang, and Quan Pan, “Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2216–2223.
- [18] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma, “Image super-resolution via sparse representation,” *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [19] Bradley Efron, Trevor Hastie, Iain Johnstone, Robert Tibshirani, et al., “Least angle regression,” *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [20] Meng Yang, Lei Zhang, Jian Yang, and David Zhang, “Metaface learning for sparse representation based face recognition,” in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 1601–1604.
- [21] Pablo Marquez-Neila, Luis Baumela, and Luis Alvarez, “A morphological approach to curvature-based evolution of curves and surfaces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 2–17, 2014.
- [22] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [23] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.