Discriminative Consistent Domain Generation for Semi-supervised Learning

Jun Chen[1], Heye Zhang[2], Yanping Zhang[1(✉)], Shu Zhao[1], Raad Mohiaddin[3,4], Tom Wong[3,4], David Firmin[3,4], Guang Yang[3,4(✉)], and Jennifer Keegan[3,4]

[1] School of Computer Science and Technology, Anhui University, Hefei, China
zhangyp2@gmail.com
[2] School of Biomedical Engineering, Sun Yat-Sen University, Shenzhen, China
[3] Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, UK
[4] National Heart and Lung Institute, Imperial College London,
London SW7 2AZ, UK
g.yang@imperial.ac.uk

Abstract. Deep learning based task systems normally rely on a large amount of manually labeled training data, which is expensive to obtain and subject to operator variations. Moreover, it does not always hold that the manually labeled data and the unlabeled data are sitting in the same distribution. In this paper, we alleviate these problems by proposing a discriminative consistent domain generation ($DCDG$) approach to achieve a semi-supervised learning. The discriminative consistent domain is achieved by a double-sided domain adaptation. The double-sided domain adaptation aims to make a fusion of the feature spaces of labeled data and unlabeled data. In this way, we can fit the differences of various distributions between labeled data and unlabeled data. In order to keep the discriminativeness of generated consistent domain for the task learning, we apply an indirect learning for the double-sided domain adaptation. Based on the generated discriminative consistent domain, we can use the unlabeled data to learn the task model along with the labeled data via a consistent image generation. We demonstrate the performance of our proposed $DCDG$ on the late gadolinium enhancement cardiac MRI (LGE-CMRI) images acquired from patients with atrial fibrillation in two clinical centers for the segmentation of the left atrium anatomy (LA) and proximal pulmonary veins (PVs). The experiments show that our semi-supervised approach achieves compelling segmentation results, which can prove the robustness of $DCDG$ for the semi-supervised learning using the unlabeled data along with labeled data acquired from a single center or multicenter studies.

J. Chen and H. Zhang—These authors contributed equally to this work.

---

1   Introduction

Fitting the possible differences of distributions between labeled data and unlabeled data is of high importance for the semi-supervised learning. The usage of unlabeled data can overcome the limitation of insufficient labeled data, which is normally a hurdle in medical image analysis problems that lack labeled data. In practice, incorporating unlabeled data may fail due to the domain shift between labeled data and unlabeled data [1].

The domain adaptation can learn generically adaptive representation domain but is subject to the limited discriminative feature domain for the task model learning. On the one hand, recent domain adaptation approaches usually introduce a discriminator to encourage data from one domain to generate a feature domain that is similar to the other one by keeping inter representation invariant between the two domains. They are based on a single adaptation direction. Thus the generated feature space is limited by one of the two domains. Because the original feature space of the other domain is lost, it can result in a reduced feature space of the two domains. On the other hand, the widespread domain adaptation approaches work with only the labeled data for the task model learning. The unlabeled data is only used to generate the domain adaptation space along with the labeled data based on an adversarial learning. Therefore, the discriminativeness of adapted features for subsequent task model still completely rely on the labeled data. Consequently, it is hard to guarantee the discriminativeness of adapted features for task learning of the unlabeled data.
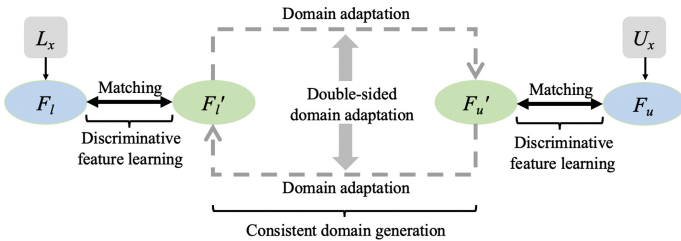

Fig. 1. Indirect double-sided domain adaptation. The labeled data (L_x) and the unlabeled data (U_x) are mapped to the adapted consistent feature domain $F(F_l$ and $F_u)$. The F is generated based on the double-sided domain adaptation via an indirect adversarial learning that the domain discriminator is applied to the feature domain F', which is obtained from the predicted label. Feature matching is then used to keep the consistence between F and F'.

In this work, we propose a discriminative consistent domain generation method based on double-sided domain adaptation (as shown in Fig. 1) to learn a task model (TM) in a semi-supervised manner. In our proposed $DCDG$, the available labeled data $L = (L_x, L_y)$ come from domain D_l and the unlabeled data $U = (U_x)$ come from domain D_u. The D_l and D_u are from the same or a similar

domain. We adopt the double-sided domain adaptation to generate a consistent feature domain F that fuses the feature spaces of D_l and D_u instead of extracting the common parts of two domains or making one domain to adapt the other one. $DCDG$ shares a feature representation generator $FG \rightarrow F$ that maps L_x and U_x to the consistent feature domain F (F_l and F_u). For the purpose of discriminative feature generation, we map F to the predicted label domain via TM. Then the domain discriminator maps predicted label domain to another feature domain (F') to make an identification via indirect double-sided domain adaptation for F. During the indirect double-sided domain adaptation, the parameters of TM are fixed and we constrain F' to match the F generated by FG. We can adapt the F indirectly by adapting the F' to guarantee the discriminativeness of feature domain F for the subsequent learning of TM. During the discriminative consistent feature domain generation, we can learn the TM although there are no available labels in U_x for us to directly learn the TM. The F' is matched with F, which is produced by both the L_x and U_x. We can further map F' to the generated labeled and unlabeled data (L'_x and U'_x) to use the image consistency as the semi-supervised information to learn the TM. We demonstrate the performance of our proposed $DCDG$ for the left atrium segmentation [2,3] on a LGE-CMRI dataset, which plays an important role in the management of atrial fibrillation and myocardial infarction [4,5].

2 Method

The proposed $DCDG$ tries to generate the discriminative consistent feature domain with fused feature space from the labeled data L_x and unlabeled data
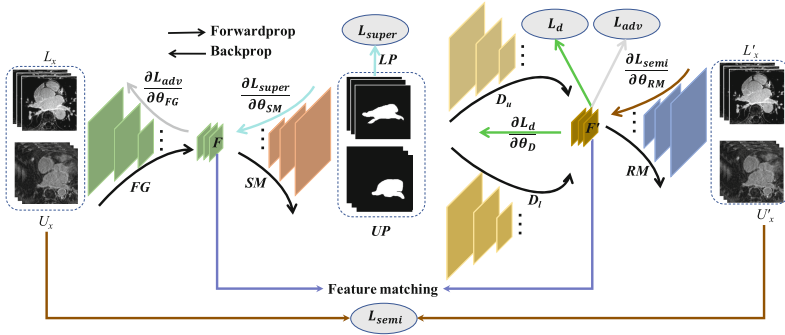

Fig. 2. Illustration of the proposed $DCDG$. L_x and U_x represent the labeled data and unlabeled data respectively. They share a common feature generator (FG) to generate the consistent features F. Then the F is mapped to the LP and UP of LA and PVs mask for L_x and U_x by the segmentation model (SM). Next, two domain discriminators with shared weights map the LP and UP to a feature domain (F') to make an identification via indirect double-sided adversarial learning for F. During the adversarial learning, the parameters of SM are fixed and we constrain F' to match the F generated by FG. Finally, the F' is mapped to the L'_x and U'_x matched with L_x and U_x.

U_x by the indirect double-sided domain adaptatipon. Meanwhile, we introduce an extra image consistent generation as the semi-supervised learning. Then the segmentation model can also be trained by the U_x along with the L_x. Detailed network configuration can be found in the supplementary materials.

Discriminative Feature Extraction via Indirect Learning. We aim to generate the consistent feature domain via double-sided domain adaptation to learn a segmentation model using both the labeled data L_x and the unlabeled data U_x. It is important to maintain the discriminativeness of feature representations for the generated consistent feature domain. Hence, we introduce the indirect learning for the discriminative feature domain extraction. In our proposed $DCDG$, we use a feature generator FG to generate the consistent feature domain F (F_l and F_u) from the L_x and U_x without using any knowledge of the source of images during testing [6]. In order to guarantee the discriminativeness of the generated features for subsequent segmentation model, we introduce the indirect domain adaptation instead of direct domain adaptation for F. We map the generated feature F to the estimated LA and PVs with probability maps (LP and UP) of L_x and U_x via the segmentation model (SM). Then the domain discriminator (D) generates the feature domain F' (F'_l and F'_u) by the final convolutional layer from LP and UP, and finally produces a scalar value to identify the F'_l and F'_u. During the indirect learning, we fix the parameters of SM and match the feature domains between F' and F via the squared L_2-norm that is defined as:

$$L_{fm} = ||F_l - F'_l||_2^2 + ||F_u - F'_u||_2^2 \tag{1}$$

where the F_l and F_u is produce by the FG (L_x) and FG (U_x) respectively. The F'_l and F'_u is produce by the final convolutional layer of the domain discriminator.

Therefore, we identify the F indirectly by identifying the F' to guarantee the discriminativeness of feature domain F for the learning of segmentation network.

Consistent Feature Domain Generation via Double-Sided Adaptation. In order to generate the consistent feature domain, we introduce two discriminators (D_l and D_u) to achieve a double-sided domain adaptation which enables the features produced by the L_x and U_x to adapt each other as shown in Fig. 2. D_l is used to encourage the L_x to generate the feature domain that is similar to the ones produced by U_x. D_u is used to force the U_x to generate the feature domain that is similar to the ones produced by L_x. Hence, a double-sided adversarial training is used to achieve the double-sided domain adaptation. During the double-sided adversarial training, for the learning of domain discriminator, F'_l is used as the *fake* feature and F'_u is used as the *real* feature to learn the D_l, while F'_l is used as the *real* feature and F'_u is used as the *fake* feature to learn the D_u simultaneously. For the learning of feature generator FG, D_l tries to identify the F'_l as *real* features and D_u tries to identify the F'_u as *real* features. In order to reduce the parameters of network, we make the D_l and D_u to share a discriminator D_{lu} to achieve the double adversarial learning directly. Inspired by this, we take the F'_l and F'_u as *real* features and *false* features respectively to directly learn the D_{lu}. When we learn the generator, the F'_l and the F'_u are

assigned with the $False$ label and $True$ label respectively. Overall, during the double-sided adaptation, we fix the parameters of segmentation model SM and aim to optimize the following L_d and L_{adv} for learning the domain discriminator and feature generator respectively:

$$L_d(\theta_{D_{lu}}) = \sigma(D_{lu}(SM(FG(L_x))), 1) + \sigma(D_{lu}(SM(FG(U_x))), 0) \qquad (2)$$

$$L_{adv}(\theta_{FG}) = \sigma(D_{lu}(SM(FG(L_x))), 0) + \sigma(D_{lu}(SM(FG(U_x))), 1) \qquad (3)$$

where the L_x and U_x represent the input of the labeled and unlabeled data, respectively. σ is the binary cross-entropy loss. SM represents the segmentation model. During the domain adaptation, the parameters of the SM are fixed.

Semi-supervised Segmentation Model Learning. During the generation of the discriminative consistent feature domain, we learn the segmentation model SM by L_x and U_x. Since there is no available labels for U_x, U_x cannot be directly used for training the segmentation model along with the L_x. However, during the double-sided domain adaptation, we get the matched feature domains between F and F', while the F is generated from the L_x and U_x. Hence, we perform a reverse mapping (RM) that mapping the F' to the L'_x and U'_x which are matched with L_x and U_x to achieve the consistent image generation. Then the U_x can use a consistent image loss as semi-supervised loss L_{semi} to train the SM along with the supervised loss L_{super} from the L_x. Finally, we use L_x and U_x to train SM with the loss defined as follows:

$$L^l_{super,semi}(\theta_{SM}, \theta_{RM}, L_x) = ||L_x - L'_x||_2^2 + DL(LP, L_y) \qquad (4)$$

$$L^u_{semi}(\theta_{SM}, \theta_{RM}, U_x) = ||U_x - U'_x||_2^2 \qquad (5)$$

where LP is the estimated LA and PVs map. L_y is the ground truth. $|| \cdot ||_2^2$ is the squared L_2-norm. F and F' are the matched features. $DL(\cdot)$ represents the Dice loss function. We train the SM after each epoch of the domain adaptation.

Evaluation Criteria. We use the region-based metrics: the Dice coefficient $(Dice = 2\frac{|P \bigcap G|}{|P|+|G|})$ and intersection-over-union $(IoU = \frac{|P \bigcap G|}{|P \bigcup G|})$, which validate the predicted LA and PVs (P) against the ground-truth (G). And the surface-based metric of mean surface distance defined as $MSD = \frac{1}{2}[\bar{d}(S, S') + \bar{d}(S', S)]$, which $\bar{d}(S, S')$ is the mean of the distances between every surface voxel in predicted mesh S and the closest surface voxel in ground-truth mesh S'.

3 Experimental Results and Discussion

In our experiments, we used two centers dataset (detailed imaging parameters can be found in the supplementary materials) for the LA and PVs segmentation to validate the proposed $DCDG$. The final segmentation model was obtained using 'early stopping' on validation data. To demonstrate the performance of our proposed $DCDG$, we compare our proposed $DCDG$ with full supervised methods, e.g., $2D$ $UNet$ [7], $SegNet$ [8] and a recent state-of-the-art 3D segmentation

architecture namely 3D *DenseNet* [9], and also make a comparision between *DCDG* and a semi-supervised [10] (*AR*) method along with a domain adaptation [11] (*ASOS*) method with the single-level adversarial learning. In addition, we also compared *DCDG* to itself but used in a fully supervised manner, namely full supervised segmentation (*FSS*).

Table 1. Comparison of the performance of our proposed *DCDG* on C1.

Method	IoU	MSD (mm)	Dice
2D UNet	0.8695 ± 0.0294	1.1149 ± 0.4745	0.9299 ± 0.0170
SegNet	0.8418 ± 0.0265	1.2788 ± 0.2666	0.9139 ± 0.0157
3D DenseNet	0.8414 ± 0.0258	1.4548 ± 0.5463	0.9136 ± 0.0153
FSS	0.8719 ± 0.0272	1.0318 ± 0.2776	0.9314 ± 0.0157
AR (25%)	0.8215 ± 0.0330	1.5829 ± 0.4103	0.9016 ± 0.0200
ASOS (25%)	0.8252 ± 0.0312	1.5645 ± 0.4455	0.9039 ± 0.0189
DCDG (25%)	0.8474 ± 0.0338	1.2823 ± 0.3524	0.9170 ± 0.0202
AR (50%)	0.8349 ± 0.0377	1.4538 ± 0.4038	0.9096 ± 0.0288
ASOS (50%)	0.8529 ± 0.0341	1.1741 ± 0.3647	0.9203 ± 0.0184
DCDG (50%)	0.8709 ± 0.0276	1.0236 ± 0.2749	0.9307 ± 0.0161
AR (75%)	0.8541 ± 0.0255	1.1536 ± 0.3531	0.9211 ± 0.0149
ASOS (75%)	0.8602 ± 0.0238	1.1176 ± 0.2805	0.9247 ± 0.0139
DCDG (75%)	$\mathbf{0.8803 \pm 0.0188}$	$\mathbf{0.9285 \pm 0.2566}$	$\mathbf{0.9362 \pm 0.0107}$

Experiments on a Single Center Dataset. We performed multiple experiments on data with the same image domain based on different ratios of labeled data acquired at center 1 denoted as $C1$. The total number of $C1$ data is 175, in which 140 samples are randomly selected and used to train the model. We randomly selected 15 samples for model validation (7 pre-ablation and 8 post-ablation samples). We also randomly selected 20 samples for independent testing (10 pre-ablation and 10 post-ablation samples). During the experiment, we randomly selected different ratio r of labeled cases ($r = 25\%, 50\%, 75\%$) from 140 samples along with the ratio $(1 - \mathrm{r})$ of the unlabeled data for the semi-supervised learning of *DCDG* and *AR*, while the *ASOS* is learned by the labeled data of ratio r with the test data. The full supervised methods performed with the standard supervised training manner based on all the labeled data. The quantitative results are summarized in Table 1. As shown in Table 1, when we use 50% labeled data, the performance of *DCDG* is superior to 2D *UNet*, *SegNet*, 3D *DenseNet*. When we use 75% labeled data, the performance of *DCDG* is superior to *FSS*, which use 100% labeled data with fully supervised learning. Compared to those methods, we can use much less labeled data to obtain better results. It has great significance to avoid costly manual labeling when there is limited expert availability. Furthermore, compared with *AR* and *ASOS*, our proposed *DCDG* also achieves the best results.

Table 2. Comparison of the performance of our proposed $DCDG$ on C1 and C2

Method	IoU	MSD (mm)	Dice
2D UNet	0.8125 ± 0.0355	1.7503 ± 0.6119	0.8961 ± 0.0216
SegNet	0.7420 ± 0.0474	3.1581 ± 1.4469	0.8510 ± 0.0310
3D DenseNet	0.7922 ± 0.0339	$\mathbf{1.1819 \pm 0.4967}$	0.8836 ± 0.0212
AR	0.6748 ± 0.0610	3.6165 ± 1.0855	0.8043 ± 0.0434
ASOS	0.7878 ± 0.0353	1.8757 ± 0.5038	0.8808 ± 0.0222
DCDG	$\mathbf{0.8315 \pm 0.0254}$	1.4031 ± 0.4948	$\mathbf{0.9078 \pm 0.0151}$
FSS	0.8154 ± 0.0401	2.5606 ± 2.1998	0.8978 ± 0.0245
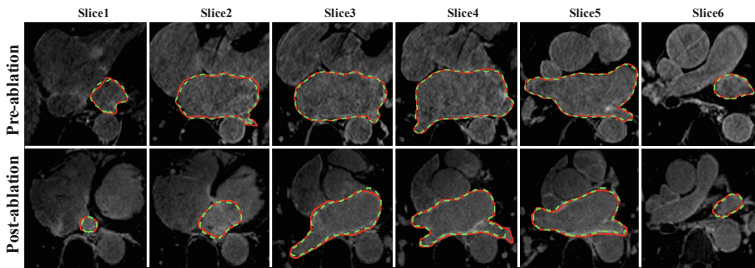
Fig. 3. Qualitative visualization of LA and PVs segmentation results compared to the manual delineation for representative slices on a pre-ablation and a post-ablation 3D LGE-CMRI images. Each estimated segmentation is represented as a dashed green contour, and its corresponding manual delineation is represented as a red contour. (Color figure online)

Experiments on a Two Center Dataset. For the experiment performed at $C1$ data was labelled as opposed to $C2$ which was unlabelled. The total number of $C2$ data is 94. We randomly selected 20 samples for testing (including 10 pre-ablation samples and 10 post-ablation samples). The remained 74 samples with no labels were used to train the $DCDG$ along with the 140 samples with available labels from $C1$. We also apply the same 15 validation dataset used in the learning from the single center data to validate the segmentation model during training. For $2D$ $UNet$, $SegNet$, $3D$ $DenseNet$ and FSS, supervised learning were performed on 140 samples with available labels from $C1$. As shown in Table 2, although $2D$ $UNet$, $SegNet$, $3D$ $DenseNet$ have achieved great success in many medical image segmentation applications, they can not be directly applied to learn a good model from the distribution of one domain to the other domain thus obtained sub-performance. In addition, our $DCDG$ also obtained the best results compared to the AR and $ASOS$. Figure 3 visually gives a further illustration to demonstrate the good performance of $DCDG$.

Model Variation Study. To verify the effectiveness of each component in our proposed $DCDG$, we perform Model variation study on $C1$ and $C2$. We take
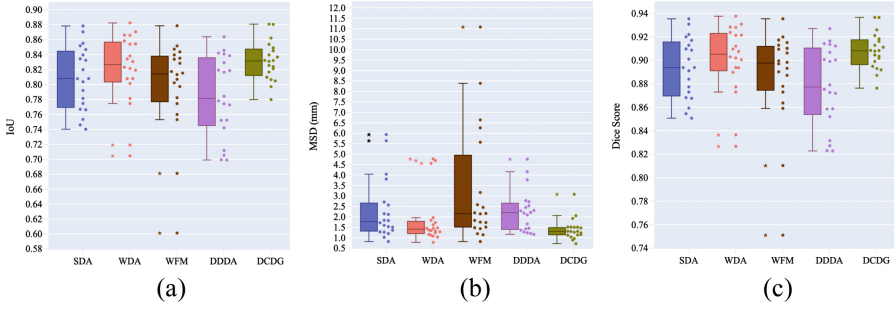
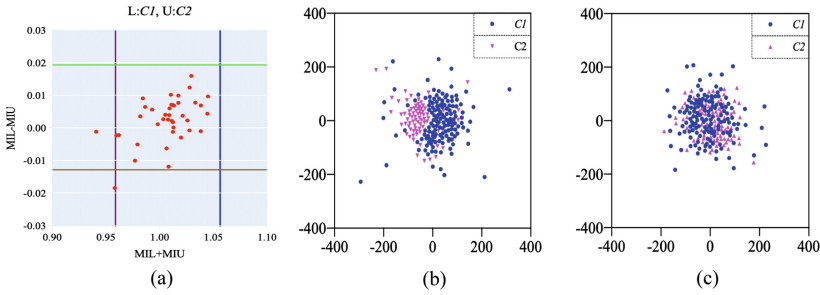**Fig. 4.** Boxplots of IoU, MSD and $Dice$ evaluations for ablation tests.

Fig. 5. (a) The stability of double-sided domain adaptation. The MIL and MIU are the mean identification value of the discriminator on labeled samples and unlabeled samples respectively after each learning of the double-sided domain adaptation. The intervals between the two vertical lines and two horizontal lines are the confidence intervals of 95% limits of agreement for MIL-MIU and MIL+MIU. (b) and (c): The t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization of feature distribution on two-center data via double-sided adaptation (c) compared to the ones without adaptation (b). The red and blue points represent the samples from $C2$ and $C1$ respectively. (Color figure online)

the structure of $DCDG$ as the baseline. Then we train the model with single domain adaptation (SDA), without domain adaptation (WDA), without feature matching (WFM) and with the direct double-sided domain adaptation ($DDDA$) for F. We compare these models with the $DCDG$ and the results are shown in Fig. 4. As shown in the Fig. 4, our proposed $DCDG$ achieves the best results across these models on the measures of IoU, MSD and $Dice\ Score$, which proves the effectiveness of our $DCDG$.

Effectiveness of Double-Sided Domain Adaptation. In our proposed $DCDG$, the double-sided domain adaptation aims to generate the discriminative consistent feature domain. Its effectiveness can be demonstrated by the identified accuracy of discriminator for L_x and U_x. Ideally, the probability values are identified by the discriminator for L_x and U_x are both 0.5. In our experiment, we record the MIL and MIU as shown in Fig. 5(a). Ideally, the value of MIL plus

MIU is 1, while the value of MIL minus MIU is 0. In this situation, the identification values of discriminator on labeled data and unlabeled data are both 0.5, which illustrates that the model achieves the double-sided domain adaptation. As shown in Fig. 5(a), in each epoch during double-sided domain adaptation, the value of MIL plus MIU is close to 1. Meanwhile, the value of MIL minus MIU is close to 0. Furthermore, the adapted features show the fused features of L_x and U_x (Fig. 5(c)) compared to no adapted feature distribution (Fig. 5(b)).

4   Conclusion

In this paper, we proposed a discriminative consistent domain generation for the semi-supervised learning. In our proposed $DCDG$, we investigate the double-sided domain adaptation based on an indirect adversarial learning to fit the differences between labeled data and unlabeled data and generate the discriminative feature domain with fused feature space. Validation of our framework has been performed against manually delineated ground truth of the LA and PVs segmentation task. Compared to other supervised, semi-supervisd and domain adaptation methods, our $DCDG$ has demonstrated superior performance. In conclusion, our proposed $DCDG$ makes it possible to create a robust semi-supervised learning model using the unlabeled data along with labeled data collected from a single center or a multicenter studies that can be well extended to solve other medical image analysis problems.

References

1. Patel, V.M., Gopalan, R., Li, R., Chellappa, R.: Visual domain adaptation: a survey of recent advances. IEEE Sign. Process. Mag. **32**(3), 53–69 (2015)
2. Chen, J., et al.: Multiview two-task recursive attention model for left atrium and atrial scars segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 455–463. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_51
3. Yang, G., et al.: Multiview sequential learning and dilated residual learning for a fully automatic delineation of the left atrium and pulmonary veins from late gadolinium-enhanced cardiac MRI images. In: EMBC, pp. 1123–1127 (2018)
4. Yang, G., et al.: Fully automatic segmentation and objective assessment of atrial scars for long-standing persistent atrial fibrillation patients using late gadolinium-enhanced MRI. Med. Phys. **45**(4), 1562–1576 (2018)
5. Zhang, N., et al.: Deep learning for diagnosis of chronic myocardial infarction on nonenhanced cardiac cine MRI. Radiology **291**(3), 606–617 (2019)
6. Zhang, Y., Zhang, Y., Wang, Y., Tian, Q.: Domain-Invariant Adversarial Learning for Unsupervised Domain Adaption. arXiv preprint arXiv: 1811.12751 (2018)

7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
8. Badrinarayanan, V., et al.: SegNet: a deep convolutionalencoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
9. Bui, T.D., Shin, J., Moon, T.: 3D densely convolution networks for volumetric segmentation. arXiv preprint arXiv: 1709.03199 (2017)
10. Koziński, M., et al.: An adversarial regularisation for semi-supervised training of structured output neural networks. arXiv preprint arXiv: 1702.02382 (2017)
11. Tsai, Y.H., et al.: Learning to adapt structured output space for semantic segmentation. In: CVPR, pp. 7472–7481 (2018)