Discriminative Consistent Domain Generation for Semi-supervised Learning

Jun Chen*1, Heye Zhang*2, Yanping Zhang1(✉), Shu Zhao1, Raad Mohiaddin3,4, Tom Wong3,4, David Firmin3,4, Guang Yang3,4(✉), Jennifer Keegan3,4

1School of Computer Science and Technology, Anhui University, Hefei, China
2School of Biomedical Engineering, Sun Yat-Sen University, Shenzhen, China
3Cardiovascular Research Centre, Royal Brompton Hospital, London SW3 6NP, UK
4National Heart & Lung Institute, Imperial College London, London SW7 2AZ, 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 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.

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.

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 $TM$.

---

**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 $F'$. 
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 \(U_x\) by the indirect double-sided domain adaptation. 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. 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 segmentation model ($SM$) and match the feature domains between $F'$ and $F$ via the squared $L_2$-norm that is defined as:

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

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:

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

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

where the $L_x$ and $U_x$ represent the input of the labeled and unlabeled data, respectively. $\sigma$ 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:

$$L_{super,semi}^l(θ_SM,θ_{RM},L_x) = ||L_x - L'_x||_2^2 + DL(LP,L_y)$$

(4)

$$L_{semi}^u(θ_SM,θ_{RM},U_x) = ||U_x - U'_x||_2^2$$

(5)

where $LP$ is the estimated LA and PVs map. $L_y$ is the ground truth. $||·||_2^2$ is the squared $L_2$-norm. $F$ and $F'$ are the matched features. $DL(·)$ 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 ($Dice = 2|P ∩ G|/|P| + |G|$) and intersection-over-union ($IoU = |P ∩ G|/|P| + |G| - |P ∩ G|$), which validate the predicted LA and PVs ($P$) against the ground-truth ($G$). And the surface-based metric of mean surface distance defined as $MSD = \frac{1}{2}[\bar{d}(S,S') + \bar{d}(S',S)]$, where $\bar{d}(S,S')$ 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 comparison 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).

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
achieves the best results. Furthermore, compared with methods, we can use much less labeled data to obtain better results. It has great performance when we use 75% 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.

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.

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

| Method       | IoU | MSD (mm) | Dice          |
|--------------|-----|----------|---------------|
| 2D UNet      | 0.8825 ± 0.0384 | 1.249 ± 0.0372 | 0.9129 ± 0.0176 |
| SegNet       | 0.8439 ± 0.0205 | 1.276 ± 0.0266 | 0.9139 ± 0.0157 |
| 3D DenseNet  | 0.8414 ± 0.0229 | 1.415 ± 0.5403 | 0.9138 ± 0.0153 |
| FSS          | 0.8739 ± 0.0272 | 1.031 ± 0.2779 | 0.9131 ± 0.0157 |
| AR(25%)      | 0.8235 ± 0.0336 | 1.543 ± 0.4031 | 0.9038 ± 0.0200 |
| ASOS(25%)    | 0.8252 ± 0.0312 | 1.564 ± 0.4055 | 0.9039 ± 0.0199 |
| DCDG(25%)    | 0.8474 ± 0.0238 | 1.293 ± 0.5224 | 0.9176 ± 0.0202 |
| AR(50%)      | 0.8420 ± 0.0377 | 1.431 ± 0.4038 | 0.9098 ± 0.0248 |
| ASOS(50%)    | 0.8520 ± 0.0341 | 1.741 ± 0.3647 | 0.9203 ± 0.0194 |
| DCDG(50%)    | 0.8799 ± 0.0276 | 1.123 ± 0.2740 | 0.9397 ± 0.0161 |
| AR(75%)      | 0.8540 ± 0.0325 | 1.133 ± 0.9375 | 0.9215 ± 0.0149 |
| ASOS(75%)    | 0.8602 ± 0.0238 | 1.117 ± 0.2905 | 0.9247 ± 0.0139 |
| DCDG(75%)    | 0.8663 ± 0.0188 | 0.9265 ± 0.2566 | 0.9362 ± 0.0107 |

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

| Method       | IoU | MSD (mm) | Dice          |
|--------------|-----|----------|---------------|
| 2D UNet      | 0.8125 ± 0.0375 | 1.753 ± 0.0419 | 0.8981 ± 0.0206 |
| SegNet       | 0.7420 ± 0.0474 | 3.158 ± 1.4809 | 0.8510 ± 0.0336 |
| 3D DenseNet  | 0.7922 ± 0.0319 | 1.183 ± 0.6907 | 0.8838 ± 0.0212 |
| AR           | 0.7479 ± 0.0010 | 3.845 ± 1.0975 | 0.8943 ± 0.0348 |
| ASOS         | 0.7529 ± 0.0153 | 1.972 ± 0.5938 | 0.8668 ± 0.0222 |
| DCDG         | 0.8035 ± 0.0254 | 1.403 ± 0.4916 | 0.9079 ± 0.0151 |
| FSS          | 0.8114 ± 0.0481 | 2.500 ± 2.1999 | 0.8973 ± 0.0243 |
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. Fig. 3 visually gives a further illustration to demonstrate the good performance of DCDG.

![Fig. 3](image)

**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.

![Fig. 4](image)

**Fig. 4.** Boxplots of IoU, MSD and Dice evaluations for ablation tests.

**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 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-supervised 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.

5 Acknowledgments

This work was supported in part by the Young Scientists Fund of the National Natural Science Foundation of China (61602003), in part by the Natural Science
Foundation of China (61771464 and U1801265) and in part by the Guangdong Science and Technology (2018A050506031 and 2019B010110001).

References

[1] Patel, V.M., Gopalan, R., Li, R., Chellappa, R.: Visual domain adaptation: A survey of recent advances. IEEE signal processing magazine. 32(3), 53-69 (2015).
[2] Chen, Jun., et al.: Multiview two-task recursive attention model for left atrium and atrial scars segmentation. In: MICCAI, pp. 455–463. Springer (2018).
[3] Yang, Guang., 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, Guang., et al.: Fully automatic segmentation and objective assessment of atrial scars for long-standing persistent atrial fibrillation patients using late gadolinium-enhanced MRI. Medical physics. 45(4), 1562-1576 (2018).
[5] Zhang, Nan., et al.: Deep Learning for Diagnosis of Chronic Myocardial Infarction on Nonenhanced Cardiac Cine MRI. Radiology. 291(3), pp. 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: MICCAI, pp. 234-241. Springer (2015).
[8] Badrinarayanan, V., et al.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence. 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).