Unsupervised Domain Adaptation using Soft-Labeled Contrastive Learning with Reversed Monte Carlo Method for Cardiac Image Segmentation

Mingxuan Gu[1], Mareike Thies[1], Siyuan Mei[1], Fabian Wagner[1], Mingcheng Fan[1], Yipeng Sun[1], Zhaoya Pan[3], Sulaiman Vesal[2], Ronak Kosti[1], Dennis Possart[3], Jonas Utz[3], and Andreas Maier[1]

[1] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[2] Department of Radiology, Stanford University, Stanford, CA, USA
[3] Department Artificial Intelligence in Biomedical Engineering, Erlangen, Germany
mingxuan.gu@fau.de

Abstract. Recent unsupervised domain adaptation methods in medical image segmentation adopt centroid/prototypical contrastive learning (CL) to match the source and target features for their excellent ability of representation learning and semantic feature alignment. Of these CL methods, most works extract features with a binary mask generated by similarity measure or thresholding the prediction. However, this hard-threshold (HT) strategy may induce sparse features and incorrect label assignments. Conversely, while the soft-labeling technique has proven effective in addressing the limitations of the HT strategy by assigning importance factors to pixel features, it remains unexplored in CL algorithms. Thus, in this work, we present a novel CL approach leveraging soft pseudo labels for category-wise target centroid generation, complemented by a reversed Monte Carlo method to achieve a more compact target feature space. Additionally, we propose a centroid norm regularizer as an extra magnitude constraint to bolster the model's robustness. Extensive experiments and ablation studies on two cardiac data sets underscore the effectiveness of each component and reveal a significant enhancement in segmentation results in Dice Similarity Score and Hausdorff Distance 95 compared with a wide range of state-of-the-art methods.

Keywords: Soft-Labeling · Cardiac Image Segmentation · Contrastive Learning

1 Introduction

Accurate cardiac segmentation is crucial for various medical applications. In clinical settings, multi-modality medical images are extensively utilized to aid diagnosis. However, automatic cardiac image segmentation often suffers from

performance degradation due to a lack of labels. Unsupervised domain adaptation (UDA) avoids laborious manual annotation by transferring the knowledge learned from a label-rich source domain to the label-deficient target domain. With the success of the generative adversarial network, adversarial learning (AL) has been widely used for solving UDA problems [15,17,16,3,18,1]. However, AL-based methods primarily focus on globally aligning the domain gap without considering the semantic information shared between the source and target domain [21,10]; thus, the performance of AL-based methods are constrained due to the misalignment of similar regions with different semantic meanings.

Contrastive learning (CL), on the other hand, pushes the pixel features of different classes apart and pulls pixel features of the same class closer. This pixel-to-pixel (P2P) alignment enforces semantic discrimination between individual features and benefits cross-domain feature alignment. Alternative centroid-to-pixel (C2P) approaches leverage centroids, *i.e.*, the average of the category-wise features, markedly lowering the memory need. Since ground-truth labels are only available in the source domain, both approaches require target pseudo-labels for contrastive pairs, posing challenges, particularly in the early stages of training, when the model exhibits poor generalizability to the target domain. Incorrect predictions on target samples can lead to over-confident or erroneous features [27,9]. To address this issue, Liu *et al.* [9] and Lee *et al.* [7] proposed involving only target features with high similarity to the paired source centroid, but this correspondence can be violated for datasets with large domain gaps. Other works screened reliable target features using entropy thresholds or high-certainty prediction scores [23,10]. Both similarity-based and threshold-based methods take one hot-encoded target pseudo-labels, which is called hard-threshold (HT) strategy in the following text. However, C2P with HT drastically reduces the number of features exposed to the network. Moreover, correctly predicted features with low certainty may not be utilized for the whole training process, resulting in severe class imbalance and degradation of the segmentation performance [19]. In contrast, the soft-labeling (SL) [22,13,14,27,11] strategy has been used to circumvent the drawbacks of HT by assigning importance factors to the pseudo-labels. In classification tasks, label smoothing has been observed to diminish the misleading impact of ambiguous pseudo-labels [27,12]. Similarly, label fusion has been shown to enhance prediction calibration and alleviate overconfidence on out-of-distribution data [14]. Drawing inspiration from these methods, we introduce the SL strategy for the unsupervised target domain within the CL technique for UDA in cardiac image segmentation.

Our contributions are summarized as follows: 1) We propose centroid-to-centroid (C2C) contrastive learning with a soft-labeling strategy (SLCL) to alleviate the misclassification and sparse feature space in the conventional C2P contrastive learning with a hard-threshold strategy (HTCL). 2) As an optimal category-wise centroid does not guarantee optimal individual features (Fig. 2a), we introduce the reversed Monte Carlo method (rMC) (Fig. 2b) for a more compact target feature space. To the best of our knowledge, it is the first time that the SL strategy and the Monte Carlo method [4] have been utilized

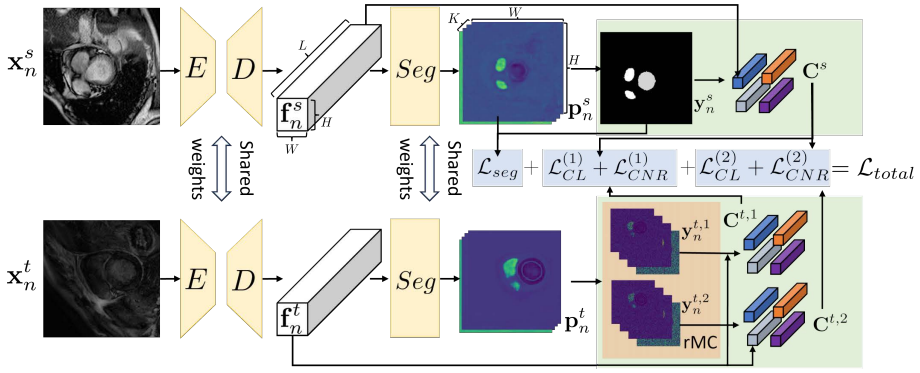


Fig. 1: An overview of the proposed method.

to boost the task performance of CL. 3) Centroid Norm Regularizer (CNR) is proposed as a complementary regularizer to force the magnitudes of the source and target features to be consistent. The code is available at `https://github.com/MingxuanGu/Soft-Labeled-Contrastive-Learning`.

2 Method

Due to distribution shifts between the source and target domains, models trained solely on source data struggle to generalize to target data. Given a set of labeled source data $\{\mathbf{x}_i^s, \mathbf{y}_i^s\}_{i=1}^{N^s}$, and unlabeled target data $\{\mathbf{x}_i^t\}_{i=1}^{N^t}$, where N^s, N^t refer to the number of source and target samples, respectively, UDA aims to bridge the performance gap between the source and the target domain. Our workflow is shown in Fig. 1. The segmentation network contains an encoder E, a decoder D, and a segmentation layer Seg. First, the category-wise source centroids are calculated as the mean of the source features for each class. Subsequently, rMC is applied to evenly split the target prediction mask, followed by the calculation of the target sub-centroids with SL strategy. Finally, the contrastive loss and the CNR are applied between the source centroid and each target sub-centroid.

2.1 Soft-Labeled Contrastive Learning (SLCL)

Current CL methods [9,7,10,6,8,21] generate the source centroid ($\mathbf{C}^s$) and target centroid ($\mathbf{C}^t$) with a masked average pooling of the class features. The features are filtered by either similarity-based or threshold-based criteria. This binary classification may introduce incorrect pseudo-labels and reduce features available during training, resulting in scattered classification in cardiac image segmentation and inaccurate diagnosis in clinical applications. In contrast, SL strategy converts this binary classification into a regression problem by putting soft importance weights on the features, which smooths the negative effect of the ambiguous pseudo-labels and makes full use of all the features. We formulate our

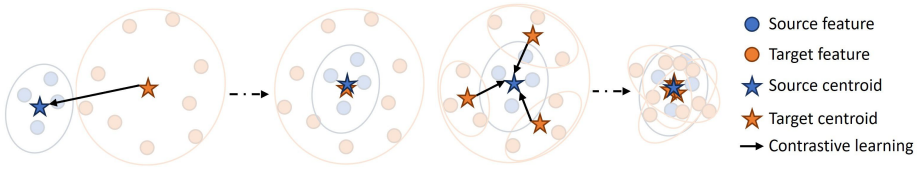


Fig. 2: Illustration of feature space before and after centroid-to-centroid (C2C) alignment. (a) Conventional C2C aligns the source and target centroids, whereas the individual features are still scattered. (b) A schematic of the proposed rMC method. Each target centroid from the subspace with the source centroid helps to generate a more compact target space.

proposed target and source centroid as

$$\mathbf{C}^t[k] = \frac{\sum_{n=1}^{B}\sum_{l}^{H\times W}\mathbf{f}_n^t[l]\mathbf{p}_n^t[l,k]}{\sum_{n=1}^{B}\sum_{l}^{H\times W}\mathbf{p}_n^t[l,k]}, \mathbf{C}^s[k] = \frac{\sum_{n=1}^{B}\sum_{l}^{H\times W}\mathbf{f}_n^s[l]\mathbf{y}_n^s[l,k]}{\sum_{n=1}^{B}\sum_{l}^{H\times W}\mathbf{y}_n^s[l,k]}, \quad (1)$$

where k is the class index, B denotes the batch size, l is the pixel location, $\mathbf{f}$ refers to the decoder feature, $\mathbf{y}^s$ is the source ground-truth label, and $\mathbf{p}^t$ is the softmax prediction of the target domain which is taken as the importance weight for each target feature. The choice of the importance weight keeps the algorithm lightweight and easy to implement. Moreover, including $\mathbf{p}^t$ in the calculation of the target centroids opens a second path for the network to optimize the target feature, *i.e.*, the prediction score, providing another degree of freedom when updating the model parameters. Since the ground-truth labels are provided for the source domain, we calculate the source centroid with HT. We progressively refine the source centroid by the exponential moving average with a momentum of 0.9.

2.2 Reversed Monte-Carlo Method (rMC)

When applying the C2C contrastive learning, once the centroids are aligned, the network can no longer provide enough gradient to further update the pixel features (Fig. 2a). To solve this problem, we consider using the Monte Carlo method [4] but in a reversed way. The Monte Carlo method can be used to estimate the expectation with a limited number of samples.

$$\mathbb{E}[X] = \int_{\Omega} xp(x)dx \rightarrow \mathbb{E}[X] \approx \bar{X} = \sum_{i=1}^{N} x_i/N, \quad (2)$$

where X indicates the random variable, Ω is the variable space, N denotes the number of samples, x_i refers to the sampled objects. The variance of this estimation is calculated as $\mathrm{Var}(\bar{\mathrm{X}}) = \mathrm{Var}(\mathrm{X})/\mathrm{N}$. It can be reduced by increasing

the number of samples N. In other words, if we reduce the number of samples, we can increase the chance of inducing a higher deviation from the expectation, *i.e.*, the category-wise centroid in CL. Then, the individual features can be further optimized. As we can see in Fig. 2b, we reduce the number of samples by splitting the pixel features into P partitions. The sub-centroid $C^{t,i}$ significantly deviates from the source centroid, which induces higher gradients during back-propagation. After optimization, the feature space is more compact. Finally, the contrastive loss is applied to each sub-centroid:

$$\mathcal{L}_{CL}^{(i)} = \mathcal{L}_{CL}(\mathbf{C}^s, \mathbf{C}^{t,i}), \tag{3}$$

where $i \in \{1, 2, ..., P\}$. The best result is achieved when $P = 2$. A parameter study on P is provided in Fig. 4a. The contrastive loss is defined as a modified InfoNCE [5,6]:

$$\mathcal{L}_{CL}(\mathbf{C}^s, \mathbf{C}^t) = -\frac{1}{K}\sum_{k}^{K} log\frac{\mathrm{h}(\mathbf{C}^{\mathrm{s}}[\mathrm{k}], \mathbf{C}^{\mathrm{t}}[\mathrm{k}])}{\mathrm{h}(\mathbf{C}^{\mathrm{s}}[\mathrm{k}], \mathbf{C}^{\mathrm{t}}[\mathrm{k}]) + \sum_{\substack{r\neq k \\ q\in\{s,t\}}}^{\mathrm{K}} \mathrm{h}(\mathbf{C}^{\mathrm{t}}[\mathrm{k}], \mathbf{C}^{\mathrm{q}}[\mathrm{r}])}, \tag{4}$$

where K is the total number of classes, and $\mathrm{h}(\mathbf{u}, \mathbf{v}) = \exp(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|_2\|\mathbf{v}\|_2}/\tau)$ is the cosine similarity. We empirically set τ to 0.1 [2].

2.3 Centroid Norm Regularizer (CNR)

The cosine similarity only measures the angle between two vectors (Eq. 4), while the impact of the magnitude is ignored, which may result in separate source and target feature spaces and sub-optimal performance. In this work, we propose the CNR to regularize the magnitude of the target (sub-) centroids:

$$\mathcal{L}_{CNR}^{(i)} = \mathcal{L}_{CNR}(\mathbf{C}^s, \mathbf{C}^{t,i}) = \frac{1}{K}\sum_{k=1}^{K}(\|\mathbf{C}_k^s\|_2 - \left\|\mathbf{C}_k^{t,i}\right\|_2)^2, \tag{5}$$

where i is the partition index. We consider the source centroids to be the ground truth of the target centroids. Therefore, the gradient of the CNR only flows through the target path. Moreover, cross-entropy and Jaccard loss are used for supervised learning in the source domain. We sum up the overall loss function as:

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \lambda_{CL}\sum_{i=1}^{P}\mathcal{L}_{CL}^{(i)} + \lambda_{CNR}\sum_{i=1}^{P}\mathcal{L}_{CNR}^{(i)}. \tag{6}$$

λ_{CL} and λ_{CNR} are the regularization weights of the corresponding loss terms.

3 Experiments and Results

Datasets Cardiac segmentation is conducted on two public datasets, namely MS-CMRSeg [24] and CT-MR (MMWHS) dataset [25]. Specifically, the MS-CMRSeg includes 45 MR volume pairs from short-axis balanced Steady-State

Table 1: The quantitative comparison results on the MS-CMRSeg dataset [24]. **Src-Only** is the baseline trained only with $\mathcal{L}_{seg}$. **Supervised** refers to the supervised learning on the target data. The best scores are highlighted in bold.

	Volumetric DSC ↑				Volumetric HD95 (mm)↓			
Methods	MYO	LV	RV	AVG	MYO	LV	RV	AVG
Src-Only	0.434	0.614	0.538	0.529	13.251	20.427	19.390	17.689
Supervised	0.806	0.913	0.834	0.851	4.342	4.807	8.440	5.863
AdaptSeg [15]	0.629	0.826	0.716	0.724	9.084	9.317	11.433	9.945
Advent [17]	0.660	0.824	0.748	0.744	7.988	8.949	11.806	9.581
AdaptEvery [16]	0.677	0.859	0.788	0.775	8.007	7.189	11.809	9.002
MPSCL [9]	0.677	0.856	0.750	0.761	7.524	8.016	11.773	9.104
BCL [7]	0.722	0.874	0.797	0.798	**6.568**	8.035	10.076	8.226
SLCL (ours)	**0.743**	**0.884**	**0.820**	**0.816**	8.169	**6.397**	**9.634**	**8.067**

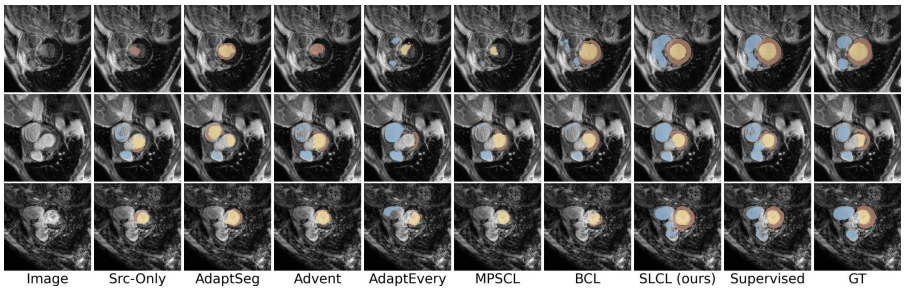


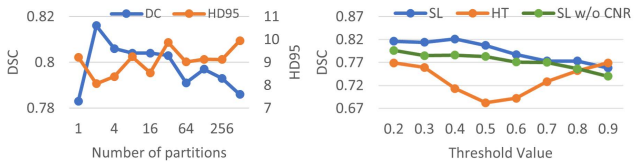
Fig. 3: The qualitative results of the cardiac segmentation of all the comparison methods on the MS-CMRSeg dataset [24]. The raw images are enhanced for better visualization. (Best viewed in color)

Free Procession (bSSFP) and Late-gadolinium enhancement (LGE), which are utilized as source and target domain, respectively; Five samples of MS-CMRSeg are for validation, and the remaining 40 samples are split evenly for training and testing. The CT-MR dataset contains 20 CT (source) and MR (target) volumes with ground-truth labels, whereas the remaining 32 CT and 25 MR volumes were automatically labeled [26] and only used for training. The dataset sampled 16 slices from each CT and MR volume. Gold standard MR images are evenly split for training and testing. Two-fold cross-validation is utilized for evaluation. For both datasets, three categories are segmented: the myocardium (MYO), the left ventricle (LV), and the right ventricle (RV).

Implementation details All the experiments are implemented with Python 3.9, PyTorch 1.10, CUDA 12.2, and executed on NVIDIA A100. The same UNet architecture (pretrained ResNet50 on ImageNet for the encoder) is used for all the reference methods for a fair comparison. We conduct a random search for all the methods and set the hyperparameters empirically for the best results. We use an SGD optimizer and a batch size of 32. Only flipping and rotation $[n \times 90°]$ are used for data augmentation. Table 3 shows the detailed hyperparameter settings.

Table 2: The quantitative comparison results on CT-MR dataset [25]. The best scores are highlighted in bold.

	DSC ↑				HD95 (mm)↓			
Methods	MYO	LV	RV	AVG	MYO	LV	RV	AVG
Src-Only	0.191	0.694	0.341	0.409	29.574	34.498	29.543	31.205
Supervised	0.724	0.872	0.821	0.806	15.563	15.852	19.168	16.861
AdaptSeg [15]	0.582	0.786	0.603	0.657	22.563	24.972	20.615	22.717
Advent [17]	0.608	0.854	0.758	0.740	13.848	12.618	16.031	14.166
AdaptEvery [16]	**0.645**	0.853	0.779	0.759	**12.026**	11.829	15.155	13.003
MPSCL [9]	0.615	0.866	0.728	0.736	13.126	12.828	16.968	14.307
BCL [7]	0.570	0.794	0.686	0.683	27.891	22.767	24.563	25.074
SLCL (ours)	0.606	**0.869**	**0.807**	**0.761**	13.054	**11.737**	**13.287**	**12.693**



(a) (b)

Fig. 4: Experiment results on MS-CMRSeg dataset [24]. (Best viewed in color) (a) Parameter study on the partition number P of rMC. $P = 1$ refers to Soft-Labeled Contrastive Learning (SLCL) without rMC. (b) Ablation study on the threshold.

3.1 Performance on Cardiac Image Segmentation

We evaluate the segmentation performance with Dice Similarity Score (DSC) and Haudorff Distance 95 (HD95). The quantitative comparison of different reference methods on MS-CMRSeg [24] is provided in Table 1. The AL-based methods (AdaptSeg [15], Advent [17], AdaptEvery [16]) generally achieved lower DSC and higher HD95 than CL-based methods due to their semantic agnostic characteristics. Both CL-based methods (MPSCL [9], BCL [7]) generate centroids with C2P and HT, which induces a sparse feature space, and limits the information exposed to the network. Our method outperforms the other feature-alignment methods by a large margin both in DSC and HD95. Similar results can be observed qualitatively in Fig. 3. More qualitative results are provided in the supplementary file.

Quantitative results on the CT-MR dataset [25] are shown in Table 2. The models perform worse on this dataset due to a larger domain gap and less data. The performance of BCL [7] drops most significantly. BCL [7] proposed a dynamic pseudo-label to correct the error between the source and target centroid, while it is not guaranteed that the dynamic labels are denser and more accurate. This also raises the value of finding a more robust pseudo-labeling strategy for CL. In comparison, the proposed method is robust across different datasets with different degrees of domain shift. The qualitative results of the CT-MR dataset [25] are provided in the supplementary file.

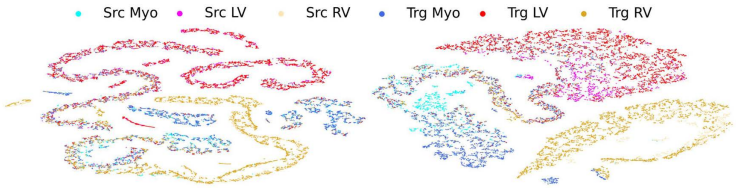


Fig. 5: t-SNE visualization. Left: t-SNE of SLCL without rMC. Right: t-SNE of the proposed SLCL with rMC. (Best viewed in color)

Table 3: The detailed hyperparameter settings of the proposed method during training.

	MS-CMRSeg [24]	CT-MR [20]
learning rate	2e-2	8e-4
λ_{CL}	1	1
λ_{CNR}	2e-3	4e-5

Table 4: The results of the ablation study on different components on MS-CMRSeg dataset [24].

	DSC ↑	HD95 (mm) ↓
Src-Only	0.529	17.689
HTCL (HT+rMC+CNR)	0.769	11.429
SLCL w/o CNR (SL+rMC)	0.796	9.179
Proposed SLCL (SL+rMC+CNR)	**0.816**	**8.067**

Fig. 4a shows a parameter study on the number of partitions. When we apply rMC, as we increase the partition number, the number of samples for each partition drastically decreases, which results in a large bias in the centroid and induces a performance drop. To interpret this observation from another perspective, when P equals the number of samples, C2C is degraded to C2P, which impairs the rectification ability of the SL strategy. The proposed algorithm reaches the optimum performance under the trade-off between the feature compactness and centroid stability when P equals 2. Fig. 5 shows the t-SNE plots for the proposed SLCL with and without rMC. It is illustrated that with rMC, the feature space is more compact and aligned between source and target feature space.

3.2 Ablation Study on MS-CMRSeg

Although the proposed SLCL discards the threshold used in the HT strategy, we conduct an ablation study (Fig. 4b) on MS-CMRSeg [24] to show the impact of the threshold on both SL and HT pseudo-labeling strategies. For HTCL, as we increase the threshold, the feature space gets sparse, and the noise dominates the performance. As we further increase the threshold, the uncertain pseudo-labels are gradually removed, resulting in a better performance. For the proposed SLCL, as we decrease the threshold, the model gets more information from the features. Meanwhile, the SL strategy makes the model more robust to the noisy pseudo-labels, thus achieving better performance. CNR helps the model to achieve even better results compared to the model without CNR (SL w/o CNR).

Table 4 shows the quantitative results of an ablation study on the proposed components. Performance degrades as we either switch SL to HT or remove CNR from the proposed SLCL, showing the effectiveness of each component.

4 Conclusion

We propose a simple yet effective CL method that combines SLCL with a rMC and a CNR. Its simplicity enables the combination with any UDA method to improve the performance further. Abundant experiments and ablation studies on the two public cardiac datasets validate the effectiveness and robustness of the proposed method.

Acknowledgments. The work is supported by the European Research Council (ERC Grant No. 810316) and HPC resources are provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg. The hardware is funded by the German Research Foundation.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cai, J., Xia, Y., Yang, D., Xu, D., Yang, L., Roth, H.: End-to-end adversarial shape learning for abdomen organ deep segmentation. In: Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10. pp. 124–132. Springer (2019)
2. Chaitanya, K., Erdil, E., Karani, N., Konukoglu, E.: Contrastive learning of global and local features for medical image segmentation with limited annotations. Advances in Neural Information Processing Systems **33**, 12546–12558 (2020)
3. Haq, M.M., Huang, J.: Adversarial domain adaptation for cell segmentation. In: Medical Imaging with Deep Learning. pp. 277–287. PMLR (2020)
4. Harrison, R.L.: Introduction to monte carlo simulation. In: AIP conference proceedings. vol. 1204, pp. 17–21. American Institute of Physics (2010)
5. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9729–9738 (2020)
6. Kuang, S., Woodruff, H.C., Granzier, R., van Nijnatten, T.J., Lobbes, M.B., Smidt, M.L., Lambin, P., Mehrkanoon, S.: Mscda: Multi-level semantic-guided contrast improves unsupervised domain adaptation for breast mri segmentation in small datasets. Neural Networks **165**, 119–134 (2023)
7. Lee, G., Eom, C., Lee, W., Park, H., Ham, B.: Bi-directional contrastive learning for domain adaptive semantic segmentation. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXX. pp. 38–55. Springer (2022)
8. Liang, C., Cheng, B., Xiao, B., Dong, Y., Chen, J.: Multilevel heterogeneous domain adaptation method for remote sensing image segmentation. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–16 (2023)
9. Liu, Z., Zhu, Z., Zheng, S., Liu, Y., Zhou, J., Zhao, Y.: Margin preserving self-paced contrastive learning towards domain adaptation for medical image segmentation. IEEE Journal of Biomedical and Health Informatics **26**(2), 638–647 (2022)

10. Marsden, R.A., Bartler, A., Döbler, M., Yang, B.: Contrastive learning and self-training for unsupervised domain adaptation in semantic segmentation. In: 2022 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2022)
11. Mugnai, D., Pernici, F., Turchini, F., Del Bimbo, A.: Soft pseudo-labeling semi-supervised learning applied to fine-grained visual classification. In: Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part IV. pp. 102–110. Springer (2021)
12. Müller, R., Kornblith, S., Hinton, G.E.: When does label smoothing help? Advances in neural information processing systems **32** (2019)
13. Tanaka, D., Ikami, D., Yamasaki, T., Aizawa, K.: Joint optimization framework for learning with noisy labels. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5552–5560 (2018)
14. Thulasidasan, S., Chennupati, G., Bilmes, J.A., Bhattacharya, T., Michalak, S.: On mixup training: Improved calibration and predictive uncertainty for deep neural networks. Advances in Neural Information Processing Systems **32** (2019)
15. Tsai, Y., Hung, W., Schulter, S., Sohn, K., Yang, M., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: CVPR. pp. 7472–7481 (2018)
16. Vesal, S., Gu, M., Kosti, R., Maier, A., Ravikumar, N.: Adapt everywhere: unsupervised adaptation of point-clouds and entropy minimization for multi-modal cardiac image segmentation. IEEE Transactions on Medical Imaging **40**(7), 1838–1851 (2021)
17. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2517–2526 (2019)
18. Wang, S., Yu, L., Li, K., Yang, X., Fu, C.W., Heng, P.A.: Boundary and entropy-driven adversarial learning for fundus image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. pp. 102–110. Springer (2019)
19. Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., Wu, L., Zhao, R., Le, X.: Semi-supervised semantic segmentation using unreliable pseudo-labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4248–4257 (2022)
20. Wu, F., Zhuang, X.: Cf distance: a new domain discrepancy metric and application to explicit domain adaptation for cross-modality cardiac image segmentation. IEEE Transactions on Medical Imaging **39**(12), 4274–4285 (2020)
21. Xie, B., Li, S., Li, M., Liu, C.H., Huang, G., Wang, G.: Sepico: Semantic-guided pixel contrast for domain adaptive semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
22. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
23. Zhao, X., Vemulapalli, R., Mansfield, P.A., Gong, B., Green, B., Shapira, L., Wu, Y.: Contrastive learning for label efficient semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10623–10633 (2021)
24. Zhuang, X.: Multivariate mixture model for myocardial segmentation combining multi-source images. IEEE transactions on pattern analysis and machine intelligence **41**(12), 2933–2946 (2018)

25. Zhuang, X., Li, L., Payer, C., Štern, D., Urschler, M., Heinrich, M.P., Oster, J., Wang, C., Örjan Smedby, Bian, C., Yang, X., Heng, P.A., Mortazi, A., Bagci, U., Yang, G., Sun, C., Galisot, G., Ramel, J.Y., Brouard, T., Tong, Q., Si, W., Liao, X., Zeng, G., Shi, Z., Zheng, G., Wang, C., MacGillivray, T., Newby, D., Rhode, K., Ourselin, S., Mohiaddin, R., Keegan, J., Firmin, D., Yang, G.: Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge. Medical Image Analysis **58**, 101537 (2019)
26. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. Medical image analysis **31**, 77–87 (2016)
27. Zou, Y., Yu, Z., Liu, X., Kumar, B., Wang, J.: Confidence regularized self-training. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5982–5991 (2019)