Class-aware Mutual Mixup with Triple Alignments for Semi-Supervised Cross-domain Segmentation

Zhuotong Cai[1,2,3,4,*] , Jingmin Xin[1,2,3] , Tianyi Zeng[4] , Siyuan Dong[5] , Nanning Zheng[1,2,3] , James S. Duncan[4,5]

[1] National Key Laboratory of Human-Machine Hybrid Augmented Intelligence,
[2] National Engineering Research Center for Visual Information and Applications,
[3] Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University
[4] Department of Radiology & Biomedical Imaging,
[5] Department of Electrical Engineering, Yale University
zhuotongcai@gmail.com

Abstract. Semi-supervised cross-domain segmentation, also referred to as Semi-supervised domain adaptation (SSDA), aims to bridge the domain gap and enhance model performance on the target domain with the limited availability of labeled target samples, lots of unlabeled target samples, and a substantial amount of labeled source samples. However, current SSDA approaches still face challenges in attaining consistent alignment across domains and adequately addressing the segmentation performance for the tail class. In this work, we develop class-aware mutual mixup with triple alignments (CMMTA) for semi-supervised cross-domain segmentation. Specifically, we first propose a class-aware mutual mixup strategy to obtain the maximal diversification of data distribution and enable the model to focus on the tail class. Then, we incorporate our class-aware mutual mixup across three distinct pathways to establish a triple consistent alignment. We further introduce cross knowledge distillation (CKD) with two parallel mean-teacher models for intra-domain and inter-domain alignment, respectively. Experimental results on two public cardiac datasets MM-WHS and MS-CMRSeg demonstrate the superiority of our proposed approach against other state-of-the-art methods under two SSDA settings.

Keywords: Semi-Supervised Domain adaptation · Medical Image Segmentation· Class-imbalance Cross-domain Segmentation· Mutual mixup.

1 Introduction

Recent advancements in deep learning have led to remarkable successes in supervised medical image segmentation, relying heavily on extensive pixel-wise annotations and the presumption that all datasets are independently and identically distributed [11]. However, the performance of the source-domain trained

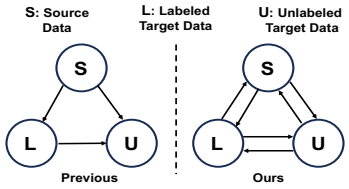

Fig. 1. Left: training model with one-way mixup. Right: Our mutual mixup for triple alignments.

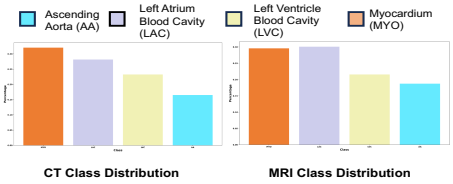

Fig. 2. Imbalanced class distribution of CT (Left) and MRI (Right) for the MM-WHS Dataset.

model can significantly decline in clinical occurrence when facing the test data with large data variances from different scanners[4], scanning protocols[19], or patient cohorts[6]. Unsupervised Domain Adaptation (UDA) seeks to adapt a model trained on a labeled source domain to perform accurately on an unlabeled target domain, without the need for additional target domain supervision. Although many previous works utilized image-translation [10,22] or adversarial learning [6,13,20] methods to bridge the domain gap from feature-level or image-level, it is still hard to achieve promising results without target label supervision to provide domain-specific information.

In comparison with UDA, we observed that some methods [12,29,2] leveraging one-shot or few-shot target labels can yield results that are competitive with, or even surpass state-of-the-art (SOTA) UDA methods. The incorporation of a few target labels embodies a pragmatic and straightforward strategy, harmonizing seamlessly with the practical demands of real-world applications [2]. Motivated by this insight and recent achievements in semi-supervised learning(SSL)[9,28], Semi-Supervised Domain Adaptation (SSDA) has been widely investigated as a promising strategy to further tackle the domain gap issue.

Recently, some approaches[25,27,19] decomposed SSDA as a combination of an SSL task and a UDA task. These two sub-tasks generate pseudo labels independently and participate in a mutually beneficial co-training process, where they learn from each other's outputs. In particular, data augmentation techniques, notably mixup methods such as [31,30,21], play a crucial role in this framework. Liu et.al [19] proposed an asymmetric co-training algorithm to alternately train the UDA and SSL model. Chen et.al [8] adopted dual-level domain mixing to extract domain-invariant representations in both region-level [30] and sample-level [31]. However, these methods still suffer from the following limitations. (1) They only considered one-way mixup for either intra-domain alignment (labeled target→unlabeled target) [19] or inter-domain alignment (labeled source→labeled target) [8], shown in Fig. 1, which may lead to inconsistent alignment with limited diversification of data distribution. (2) They both ignored the class imbalance problem in practical application, shown in Fig. 2, leading to the sub-optimal performance[28].

To address the above challenges, in this study, we present CMMTA, a novel mixup strategy with triple alignments for semi-supervised cross-domain image

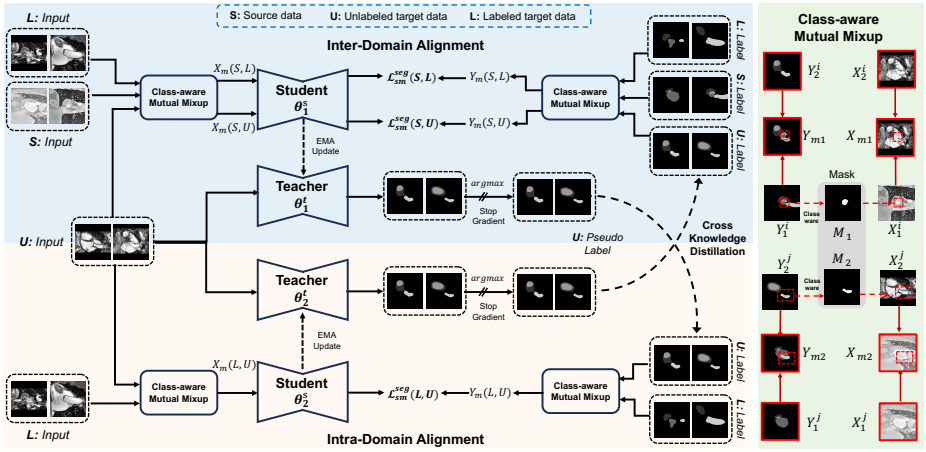

Fig. 3. The framework of our method for semi-supervised domain adaptation

segmentation. Our approach brings several novel contributions to enhance the SSDA framework, as outlined below: (1) We propose a novel class-aware mutual mixup strategy to obtain the maximal diversification of data distribution and enable the model to focus on the tail class. (2) We incorporate our class-aware mutual mixup across three distinct pathways to establish a triple consistent alignment, including labeled source ⇆ labeled target, labeled source ⇆ unlabeled target, labeled target ⇆ unlabeled target. (3) We further introduce cross-knowledge distillation consisting of two parallel segmentation networks with the mean-teacher model for intra-domain and inter-domain alignment, respectively. We evaluate the proposed CMMTA on the public datasets MM-WHS and MS-CMRSeg cardiac [34,33]. Extensive experimental results demonstrated that our method, CMMTA, significantly outperformed previous approaches across two semi-supervised domain adaptation settings.

2 Methodology

In addressing the domain adaptation task, we encounter two datasets characterized by disparate data distributions while sharing a common set of categories, denoted as C. These datasets are identified as the source domain, D_s, and the target domain, D_t. In SSDA setting, we sample an amount of labeled source data $S = \{x_s^i, y_s^i\}_{i=1}^{n_s}$ from D_s, limited labeled target data $L = \{x_l^i, y_l^i\}_{i=1}^{n_l}$ from D_t and unlabeled target data $U = \{x_u^i\}_{i=1}^{n_u}$ from D_t. Typically, n_l is considerably smaller than n_s and n_u. Our goal is to train an SSDA model with S, L, and U to perform well in the target domain.

2.1 Class-aware Mutual Mixup

Following [23,14], we utilize ClassMix[21] for data mixing, which employs the mask with preserved object boundary. Concretely, given an input image X_1 and its corresponding label Y_1, we randomly pick another image X_2 with label Y_2 to generate mixing data X_m. However, the conventional strategy mainly focused on one-way data mixing, cropping the selected area in X_1 and pasting in the X_2. For example, [23,14] sampled X_1 from D_s and sampled X_2 from D_t to build one-way source domain to target domain mixing. Such one-directional consistency (source-to-target) resulted in very limited diversification of data distribution.

In this paper, we extend the previous mixup[21] ($X_1 \rightarrow X_2$) to a mutual mixup with both $X_1 \rightarrow X_2$ and $X_2 \rightarrow X_1$. Specifically, we generate a binary mask $M_1(M_2)$ by randomly selecting half of the classes $k = K/2$ from the label $Y_1^i(Y_2^j)$ with the presented class K. For pixels associated with the selected classes in $M_1(M_2)$, their values are set to 1 and all other pixels are assigned a value of 0. Then, this mask is utilized to perform the data mixing from the image $X_1^i(X_2^j)$ to the image $X_2^i(X_1^j)$, producing the augmented image $X_m(X1, X2) = \{X_{m1}, X_{m2}\}$, which is formulated as:

$$X_{m1} = M_1 \odot X_1^i + (\mathbf{I} - M_1) \odot X_2^i, X_{m2} = M_2 \odot X_2^j + (\mathbf{I} - M_2) \odot X_1^j \quad (1)$$

where $\mathbf{I}$ is the all-ones matrix with the same size of the mask. To form the consistency for the augmented data $X_m(X_1, X_2)$, we also conduct the same operation to obtain the augmented label $Y_m(Y_1, Y_2) = \{Y_{m1}, Y_{m2}\}$, which is computed as:

$$Y_{m1} = M_1 \odot Y_1^i + (\mathbf{I} - M_1) \odot Y_2^i, Y_{m2} = M_2 \odot Y_2^j + (\mathbf{I} - M_2) \odot Y_1^j \quad (2)$$

Finally, the mutual mixup loss $\mathcal{L}_{mm}^{seg}$ contains two consistency losses using cross-entropy for two-way data mixing $\mathcal{L}_{m1}^{seg}$ and $\mathcal{L}_{m2}^{seg}$, which can be computed as:

$$\mathcal{L}_{mm}^{seg} = \underbrace{-\sum Y_{m1} \log(\mathbf{F}(X_{m1}, \theta))}_{\mathcal{L}_{m1}^{seg}} \underbrace{-\sum Y_{m2} \log(\mathbf{F}(X_{m2}, \theta))}_{\mathcal{L}_{m2}^{seg}} \quad (3)$$

The conventional ClassMix approach randomly selected classes present in Y with equal probability for mixing. However, the efficacy of ClassMix may diminish beyond the initial epochs, since the head categories with large amounts contributed less to the segmentation loss [26]. To fully leverage the mixup strategy on the training data, we introduce a class-aware mixup strategy that prioritizes the tail categories. This objective is accomplished by endowing higher probabilities to the selected categories for mixing, thereby compelling the model to intensify its attention on these specific categories.

To implement this, we first ascertain the frequency of all classes, denoted as $\mathcal{F} = \{f_1, ..., f_k\}^K$. A lower frequency value of f_k suggests that the k-th category needs to be further concentrated on. Accordingly, we derive the sampling

probability for the class-aware mixup strategy as: $\hat{\mathcal{P}} = 1 - \mathcal{F}$. This adjustment ensures that the class-aware mixup method preferentially selects less frequent classes, further improving the performance of the mutual mixup strategy.

2.2 Triple-path Alignments

Different from existing strategies that either utilized data mixing between source and target domains for inter-domain alignment [8,23,14], or within target domain between labeled and unlabeled data for intra-domain alignment [19] as seen in many SSL approaches, our method merge these two type of alignments to comprehensively bridge the domain gap. For the SSDA task, we enhance the integration by incorporating the proposed class-aware mutual mixup strategy across three distinct pathways. These include: mixing between the source domain S and the labeled target domain L, between the source domain S and the unlabeled target domain U and within the target domain, between the labeled L and unlabeled U data. Our triple-path alignment strategy establishes the consistent mixing relationship among the three data settings, significantly enhancing the capability of our class-aware mutual mixup in achieving maximal data distribution diversification.

Specifically, taking the implementation between the source domain and unlabeled target domain for example, shown in Fig. 3 green part, we first sample the X_1, Y_1 from S and sample the X_2 from U and Y_2 from P to generate X_m, Y_m, where $P = \{\hat{Y}\}$ represents the pseudo label of unlabeled target data and $\hat{Y} = \arg\max(\mathbf{F}(X_2, \theta))$. According to Eq. 3, the mutual mixup loss between source domain S and unlabeled target domain U are derived as $\mathcal{L}_{mm}^{seg}(S, U)$. For the following two paths, we utilize a similar approach to obtain the mutual mixup loss $\mathcal{L}_{mm}^{seg}(S, L)$ for source domain S and labeled target domain L and $\mathcal{L}_{mm}^{seg}(L, U)$ for labeled target domain L and labeled target domain U. Then the total loss can be calculated as follows:

$$\mathcal{L}_{total}(S, L, U) = \underbrace{\mathcal{L}_{mm}^{seg}(S, L) + \mathcal{L}_{mm}^{seg}(S, U)}_{inter-domain\ alignment} + \underbrace{\mathcal{L}_{mm}^{seg}(L, U)}_{intra-domain\ alignment} \tag{4}$$

The visual representations of the three distinct pathways employing our class-aware mutual mixup strategy are illustrated in Fig. S3.

2.3 Cross Knowledge Distillation

Instead of using the same network with the shared weight, inspired by co-training[27,19] and multi mean-teacher model [18], we introduce cross knowledge distillation (CKD) consisting of two parallel segmentation networks with the mean-teacher model for intra-domain and inter-domain alignment, respectively. We facilitate a collaborative training approach wherein reciprocally instruct each other's student networks, exchanging the generated pseudo target labels to establish consistency. This stands in contrast to the conventional method where each teacher network independently instructs its designated student.

Table 1. Quantitative comparison between our method and the other SOTA methods on the MM-WHS challenge dataset.

Category	Method	Cardiac CT → Cardiac MRI									
		Dice↑					ASSD↓				
		AA	LAC	LVC	MYO	Avg.	AA	LAC	LVC	MYO	Avg.
No DA	Source-only	18.1	25.4	26.4	14.0	21.0±9.9	32.9	31.7	28.9	27.7	30.3±6.0
UDA	SIFA [6]	65.3	62.3	78.9	47.3	63.4±7.8	7.3	7.4	3.8	4.4	5.7±2.8
	ADR [22]	76.0	69.3	81.1	48.3	66.4±6.8	6.8	5.3	3.4	4.0	4.9±1.7
	MPSCL [20]	69.7	74.3	77.1	52.2	68.3±6.1	5.1	3.0	5.8	4.9	4.7±1.6
SSDA 5%Label	DLD [24]	77.7	77.9	88.5	72.4	79.1±6.4	6.0	9.5	4.9	3.7	6.0±2.8
	ACT [19]	74.8	82.1	87.4	71.7	79.0±5.6	5.7	6.2	5.7	4.7	4.6±3.8
	SLA [8]	77.9	**81.7**	88.2	70.9	79.7±6.0	4.2	**4.5**	4.5	2.5	3.9±2.0
	GFDA [2]	75.8	78.5	84.4	69.0	75.8±5.5	5.3	7.6	5.1	4.4	5.6±2.5
	CMMTA	**79.5**	79.9	**91.2**	**74.4**	**81.3±6.1**	**3.3**	5.7	**2.5**	**1.6**	**3.3±1.9**
SSDA 10%Label	DLD [24]	77.9	81.6	90.3	74.3	81.0±4.9	4.0	8.0	2.9	3.2	4.5±2.4
	ACT [19]	77.0	82.3	88.0	73.4	80.2±5.1	5.7	3.5	5.7	4.4	4.8±2.5
	SLA [8]	78.3	83.7	89.0	75.8	81.9±5.1	5.2	3.8	5.0	2.6	4.2±2.2
	GFDA [2]	73.4	78.6	88.9	69.4	77.6±9.1	7.2	6.7	3.7	5.6	5.8±3.7
	CMMTA	**80.8**	**83.7**	**91.6**	**76.1**	**83.1±4.6**	**3.9**	**2.7**	**2.2**	**1.6**	**2.6±0.9**
Supervised	Source+Target	83.0	85.3	92.2	80.1	85.2±3.3	3.4	2.7	1.9	1.7	2.4±1.2

Specifically, for the inter-domain alignment, the student model, denoted as θ_1^s processed mixed data X_m from both the source and labeled target domains (S,L) as well as the source and unlabeled target domains (S,U), generating predictions $\mathbf{F}(X_m, \theta_1^s)$. Concurrently, the teacher model θ_1^t takes the original image $x_u \in U$ as input to generate the pseudo label $\hat{y}_{u1} = \arg\max(\mathbf{F}(x_u, \theta_1^t))$ for establishing the data mixing consistency. A similar mechanism is applied for intra-domain alignment, where the student model θ_2^s and teacher model θ_2^t generate predictions $\mathbf{F}(X_m, \theta_2^s)$ and pseudo labels $\hat{y}_{u2} = \arg\max(\mathbf{F}(x_u, \theta_2^t))$ for data mixing within the labeled and unlabeled target domains (L,U). Finally, the pseudo label $\hat{y}_{u1}$ from θ_1^t and pseudo label $\hat{y}_{u2}$ from θ_2^t are cross utilized to form the consistency losses $\mathcal{L}_{mm}^{seg}(L, U)$ and $\mathcal{L}_{mm}^{seg}(S, U)$, respectively, leading to a more robust and effective framework for semi-supervised domain adaptation. The final optimization for the two student networks is described as follows:

$$\theta_1^s \leftarrow \theta_1^s - \eta\nabla(\mathcal{L}_{mm}^{seg}(S, L) + \mathcal{L}_{mm}^{seg}(S, U)), \qquad \theta_2^s \leftarrow \theta_2^s - \eta\nabla(\mathcal{L}_{mm}^{seg}(L, U)) \quad (5)$$

where η indicates the learning rate. Then θ_1^t and θ_2^t are updated by exponentially moving average (EMA) of θ_1^s and θ_2^s. We use average prediction from two networks during the evaluation process. The overall training procedure is summarized in Algorithm 1 (Appendix).

3 Experiment

3.1 Dataset

Cardiac substructure segmentation. 1) Multi-Modality Whole Heart Segmentation Challenge 2017 dataset (MM-WHS) [34] comprised a cardiac segmentation dataset from two modalities: MRI and CT. Each modality included 20 volumes,

Table 2. Quantitative comparison between our method and the other SOTA methods on the MS-CMRSeg dataset.

Category	Method	bSSFP CMR → LGE CMR							
		Dice↑				ASSD↓			
		MYO	RVC	LVC	Avg.	MYO	RVC	LVC	Avg.
No DA	Source-only	19.5	50.0	54.5	41.3±25.0	10.5	5.0	4.1	6.5±9.5
UDA	SIFA [6]	51.4	65.0	64.6	60.4±32.7	2.9	3.6	0.9	2.5±3.1
	ADR [22]	42.0	67.7	56.0	55.2±22.8	2.6	2.1	4.5	3.1±1.8
	MPSCL [20]	61.3	80.2	71.3	70.9±21.5	1.9	1.2	1.1	1.4±0.6
SSDA 5%Label	DLD [24]	55.6	78.4	70.7	68.2±17.9	1.9	1.9	1.8	1.9±0.9
	ACT [19]	57.1	82.5	71.7	70.5±11.6	2.8	2.2	1.8	2.3±1.0
	SLA [8]	61.1	83.1	75.0	73.0±10.4	2.0	2.0	1.6	1.8±0.8
	GFDA [2]	57.9	74.2	62.3	64.8±19.8	2.6	2.3	5.6	3.5±3.7
	CMMTA	**63.6**	**84.1**	**79.8**	**75.9±7.2**	**1.5**	**1.3**	**1.3**	**1.4±0.6**
SSDA 10%Label	DLD [24]	61.4	83.3	76.1	73.6±11.9	1.6	1.5	1.5	1.5±0.6
	ACT [19]	64.4	84.1	76.6	75.1±7.5	1.8	1.6	1.6	1.6±0.5
	SLA [8]	66.5	86.0	79.3	77.3±8.8	1.4	1.4	1.3	1.3±0.6
	GFDA [2]	65.4	84.1	68.7	72.7±3.7	1.8	1.5	2.9	2.1±0.8
	CMMTA	**69.4**	**85.6**	**82.9**	**79.4±6.4**	**1.4**	**1.4**	**1.1**	**1.3±0.5**
Supervised	Source+Target	82.3	92.1	86.3	86.9±5.7	0.8	0.7	0.8	0.8±0.4

each sourced from different sites, with no direct pairing between the modalities. For our study, we focused on four cardiac structures: ascending aorta (AA), left atrium blood cavity (LAC), left ventricle blood cavity (LVC), and myocardium of the left ventricle (MYO). Following [6], we constructed the training dataset from 16 CT scans as the source data and 16 MRI scans as the target data. The test dataset consisted of 4 MRI scans. 2) The Multi-Sequence Cardiac MR Segmentation (MS-CMRSeg) [33] included three CMR sequences (LGE, bSSFP, and T2) aiming to segment the RV, LV, and MYO. Our study used bSSFP images as source data and LGE images as target data. The dataset comprised 45 paired bSSFP and LGE CMR subjects with annotations. In our experiment, we used 40 subjects for training and 5 subjects for testing. In the SSDA setting, 5% and 10% of labels from the target domain were utilized for both datasets.

Implementation Detail: Our model utilized the DeepLab-V2 [7] architecture as the backbone, pre-trained on source data. Training was executed using Pytorch 1.6.0 with Python 3.8.8 on an NVIDIA V100 16G GPU for 5×10^4 iterations with a batch size of 4. For the MM-WHS dataset, we employed an SGD optimizer, setting the learning rate, momentum, and weight decay to $2.5e^{-4}, 9e^{-1}, 1e^{-4}$, respectively. Conversely, for the MS-CMRSeg dataset, training was conducted using an Adam optimizer with a learning rate of $1e^{-4}$. To assess the model's performance, we employed the Dice coefficient and Average Symmetric Surface Distance (ASSD) as pivotal metrics.

3.2 Results

Comparison with State-of-the-art Methods. We compared our proposed approach against SOTA methods with UDA, including SIFA[6], ADR[22], MPSCL[20], and SOTA SSDA approaches such as DLD[24], ACT[19], SLA[8], and GFDA[2]. Comparisons also included a source-only model (without adaptation) and a fully

Table 3. Quantitative results of ablation study for the influence of different components on the MM-WHS dataset with 10% labeled target data.

Cardiac MRI → Cardiac CT									
Experiment	$\mathcal{L}_{m1}^{seg}$	$\mathcal{L}_{m2}^{seg}$	Class-aware	CKD	Dice↑				
					AA	LAC	LVC	MYO	Avg.
#1	✓				78.4	80.8	91.1	72.6	80.7
#2		✓			75.3	80.7	89.0	73.1	79.5
#3	✓	✓			79.4	80.8	91.1	75.0	81.6
#4	✓	✓	✓		80.4	82.8	90.4	75.6	82.3
#5	✓	✓		✓	80.1	83.6	91.4	73.9	82.3
#6	✓	✓	✓	✓	**80.8**	**83.7**	**91.6**	**76.1**	**83.1**

supervised training scenario. Table 1 shows the result of the MM-WHS dataset and Table 2 shows the result of the MS-CMRSeg dataset. Our principal findings were summarized as follows: (1) **MM-WHS**: Table 1 demonstrates that while UDA methods surpass the Baseline (source-only) performance, they fall short of reaching the benchmarks set by the supervised training due to the domain gap and insufficient target domain supervision. In the SSDA paradigm, our CMMTA yielded better performance over existing SSDA approaches on all settings, particularly on the tail class AA. Fig. S1 (Appendix) further indicated the superior performance of the proposed method against all other competitors with more complete boundary regions preserved. (2) **MS-CMRSeg**: As Tabel 2 shown, our CMMTA achieved the best segmentation performance in terms of Dice and ASSD, consistently outperforming the previous SSDA methods across two labeled settings. Fig. S2 (Appendix) revealed our model's capability to precisely segment cardiac structure boundaries, even in the most challenging areas such as the RV and MYO.

Ablation Study. We performed ablation study to verify the effectiveness of each component of the proposed method on the MM-WHS dataset with a 10% label ratio of the target domain. In Table 3, $\mathcal{L}_{m1}^{seg}$ and $\mathcal{L}_{m2}^{seg}$ indicated that we only selected the one-way from mutual mixup. **Mutual Mixup Impact**: experiment #3 underscored the pivotal importance of our mutual mixup strategy. These two mutual mixup processes served complementary roles, achieving improvements of +0.9 and +2.1 in the Dice, respectively. This synergy significantly enhanced the diversification of data distribution. **Class-aware Strategy**: From Experiment #4, incorporating the class-aware strategy will lead to improved performance on the tail class(AA) as well as head classes(LAC and MYO). **CKD Enhancement**: The comparison between experiments #3 and #5 demonstrated that incorporating the CKD component notably enhanced model performance. In addition, we verified the effectiveness of our class-aware mutual mixup, particularly with the CutMix strategy shown in Table S1 (Appendix). Besides, in Table S2 (Appendix), we provided a detailed comparison of CKD against single pseudo supervision, which used pseudo target label from model θ_1^t and θ_2^t.

4 Conclusion

In this paper, we propose class-aware mutual mixup with triple alignments for semi-supervised cross-domain segmentation. Motivated by the goals of achieving consistent domain alignment and addressing class imbalance, we develop a novel class-aware mutual mixup strategy and comprehensively incorporated it across three distinct pathways. Further enhancement of our model's capabilities is achieved through the introduction of cross knowledge distillation. Extensive experiments and comprehensive ablations indicate that our model has consistently achieved superior performance than the prior semi-supervised domain adaptation approaches across different target label ratios. Future extension work will combine our method with more effective segmentation backbones and verify within other organs/modalities.

Acknowledgments. The research reported in this publication was supported by the Natural Science Foundation of China under Grant/Award Number 62088102.

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

References

1. Bai, Y., et al.: Bidirectional copy-paste for semi-supervised medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11514–11524 (2023)
2. Basak, H., Yin, Z.: Semi-supervised domain adaptive medical image segmentation through consistency regularized disentangled contrastive learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 260–270. Springer (2023)
3. Cai, Z., et al.: Dstunet: Unet with efficient dense swin transformer pathway for medical image segmentation. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2022)
4. Cai, Z., et al.: Unsupervised domain adaptation by cross-prototype contrastive learning for medical image segmentation. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 819–824. IEEE (2023)
5. Cai, Z., et al.: Symmetric consistency with cross-domain mixup for cross-modality cardiac segmentation. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1536–1540. IEEE (2024)
6. Chen, C., et al.: Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. IEEE transactions on medical imaging **39**(7), 2494–2505 (2020)
7. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2017)
8. Chen, S., et al.: Semi-supervised domain adaptation based on dual-level domain mixing for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11018–11027 (2021)

9. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2613–2622 (2021)
10. Cheng, Y., Wei, F., Bao, J., Chen, D., Zhang, W.: Adpl: Adaptive dual path learning for domain adaptation of semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(8), 9339–9356 (2023)
11. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2021)
12. Guo, S., Xu, L., Feng, C., Xiong, H., Gao, Z., Zhang, H.: Multi-level semantic adaptation for few-shot segmentation on cardiac image sequences. Medical Image Analysis **73**, 102170 (2021)
13. Han, X., Qi, L., Yu, Q., Zhou, Z., Zheng, Y., Shi, Y., Gao, Y.: Deep symmetric adaptation network for cross-modality medical image segmentation. IEEE Transactions on Medical Imaging **41**(1), 121–132 (2022)
14. Hoyer, L., et al.: Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9924–9935 (2022)
15. Hu, H., Wei, F., Hu, H., Ye, Q., Cui, J., Wang, L.: Semi-supervised semantic segmentation via adaptive equalization learning. Advances in Neural Information Processing Systems **34**, 22106–22118 (2021)
16. Jiang, P., Wu, A., Han, Y., Shao, Y., Qi, M., Li, B.: Bidirectional adversarial training for semi-supervised domain adaptation. In: IJCAI. pp. 934–940 (2020)
17. Kim, D., Seo, M., Park, K., Shin, I., Woo, S., Kweon, I.S., Choi, D.G.: Bidirectional domain mixup for domain adaptive semantic segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 1114–1123 (2023)
18. Li, K., Wang, S., Yu, L., Heng, P.A.: Dual-teacher++: Exploiting intra-domain and inter-domain knowledge with reliable transfer for cardiac segmentation. IEEE Transactions on Medical Imaging **40**(10), 2771–2782 (2020)
19. Liu, X., et al.: Act: Semi-supervised domain-adaptive medical image segmentation with asymmetric co-training. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 66–76. Springer (2022)
20. 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)
21. Olsson, V., et al.: Classmix: Segmentation-based data augmentation for semi-supervised learning. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1369–1378 (2021)
22. Sun, X., Liu, Z., Zheng, S., Lin, C., Zhu, Z., Zhao, Y.: Attention-enhanced disentangled representation learning for unsupervised domain adaptation in cardiac segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VII. pp. 745–754. Springer (2022)
23. Tranheden, W., Olsson, V., et al.: Dacs: Domain adaptation via cross-domain mixed sampling. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1379–1389 (2021)
24. Wang, Z., Wei, Y., Feris, R., Xiong, J., Hwu, W.M., Huang, T.S., Shi, H.: Alleviating semantic-level shift: A semi-supervised domain adaptation method for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 936–937 (2020)

25. Xia, Y., Yang, D., Yu, Z., Liu, F., Cai, J., Yu, L., Zhu, Z., Xu, D., Yuille, A., Roth, H.: Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. Medical image analysis **65**, 101766 (2020)
26. Yang, L., Jiang, H., Song, Q., Guo, J.: A survey on long-tailed visual recognition. International Journal of Computer Vision **130**(7), 1837–1872 (2022)
27. Yang, L., et al.: Deep co-training with task decomposition for semi-supervised domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8906–8916 (2021)
28. You, C., Dai, W., Min, Y., Staib, L., Sekhon, J., Duncan, J.S.: Action++: Improving semi-supervised medical image segmentation with adaptive anatomical contrast. arXiv preprint arXiv:2304.02689 (2023)
29. Yu, Y.C., Lin, H.T.: Semi-supervised domain adaptation with source label adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 24100–24109 (2023)
30. Yun, S., et al.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6023–6032 (2019)
31. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)
32. Zhou, Q., et al.: Domain adaptive semantic segmentation via regional contrastive consistency regularization. In: 2022 IEEE International Conference on Multimedia and Expo (ICME). pp. 01–06 (2022)
33. 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)
34. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. Medical image analysis **31**, 77–87 (2016)