The following publication Y. Zhang, Z. -A. Huang, J. Wu and K. C. Tan, "Asymmetric Source-Free Unsupervised Domain Adaptation for Medical Image Diagnosis," 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore, Singapore, 2024, pp. 234-239 is available at https://doi.org/10.1109/CAI59869.2024.00051.

Asymmetric Source-Free Unsupervised Domain Adaptation for Medical Image Diagnosis

1st Yajie Zhang
Department of Computing
The Hong Kong Polytechnic University
Hong Kong SAR, China
yajie.zhang@connect.polyu.hk

2nd Zhi-An Huang*
Research Office
City University of Hong Kong (Dongguan)
Dongguan, China
City University of Hong Kong
Shenzhen Research Institute
Shenzhen, China
huang.za@cityu-dg.edu.cn

3rd Jibin Wu
Department of Computing
The Hong Kong Polytechnic University
Hong Kong SAR, China
jibin.wu@polyu.edu.hk

4th Kay Chen Tan*
Department of Computing
The Hong Kong Polytechnic University
Hong Kong SAR, China
kctan@polyu.edu.hk

***Abstract*—Existing source-free unsupervised domain adaptation (SFUDA) methods primarily focus on addressing the domain gap issue for single-modal data, overlooking two crucial aspects: 1) In medical scenarios, clinicians often rely on multi-modal information for disease diagnosis. Consequently, emphasizing single-modal (symmetric modality) SFUDA algorithms neglect the complementary information from other modalities (asymmetric modalities). 2) Restricting SFUDA to a single modality limits downstream institutions's ability to handle diverse modalities beyond that singular modality. To tackle these challenges, we propose an Asymmetric Source-Free Unsupervised Domain Adaptation (A-SFUDA) algorithm. This method leverages source model and unlabeled data from both symmetric and asymmetric modalities in the target domain for disease diagnosis. A-SFUDA adopts a two-stage training approach. In the first stage, A-SFUDA employs knowledge distillation (KD) to obtain two models capable of handling symmetric and asymmetric data in the target domain, facilitating preliminary diagnosis ability. In the second stage, A-SFUDA optimizes the target models through a pseudo-label correction mechanism based on multi-modal prediction correction and class-centered distance correction. Incorporating the two pseudo-label correction modules effectively mitigates noise within the training data, thereby facilitating the learning of the target models. We validate the performance of the proposed A-SFUDA algorithm on a large chest X-ray dataset, demonstrating its excellent performance for disease diagnosis in the target domain.**

***Index Terms*—source-free, unsupervised domain adaptation, pseudo-labeling, asymmetric modality**

I. Introduction

The powerful capability of deep learning in feature learning has significantly propelled the development of computer-aided diagnostic systems, such as medical image segmentation [1], medical image classification [2]–[4], image report generation [5], etc. However, when there is a distribution discrepancy between the training set and the test set, the performance of deep models can significantly degrade due to the domain gap [6], limiting their practical applications. Unsupervised domain adaptation algorithms [7], [8] leverage fully annotated source domain dataset and unlabeled target domain dataset to address the domain gap between the source and target domains, thereby achieving high performance on the target domain. However, in the medical field, source domain data is often private and non-disclosable, making source-free unsupervised domain adaptation (SFUDA) [9] a more feasible solution to address the domain gap in medical scenarios. SFUDA solely utilizes the source model and unlabeled target domain data to enhance the performance in the target domain, offering high security and strong applicability.

Existing SFUDA methods primarily concentrate on alleviating the domain gap issue inherent in single-modality data, thus they exhibit limitations when confronted with partially different modalities in target domain. CSDA [10] aims to address the domain gap issue in X-ray modality data for diagnosing lung diseases. DPL [11] proposes a pseudo-labeling method to reduce the discrepancy between different domains for single-modality data of fundus images. However, single-modality data may only provide local or specific information, while the integration of multi-modality data can offer more comprehensive and multi-faceted information [12], thus aiding doctors in gaining a holistic understanding of a patient's condition. For instance, the diagnosis of brain diseases typically requires the combination of various modalities of medical imaging data to provide comprehensive information about brain structure, function, and metabolism, such as magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET). For cancer diagnosis, a combination of various modalities is needed to provide information about tu-

*Corresponding Author

mor morphology, histological features, molecular markers, and genetic variations, such as tissue biopsy, blood tests, imaging, and genomic data. In light of this, the ubiquity of diverse modalities within patient medical information underscores the challenge of devising solutions for the recognition of multi-modal data in the target domain.

An intuitive approach to implementing multi-modal SFUDA is to mitigate the domain gap between multi-modal source domain and multi-modal target domain [14]. It requires that the data modalities of the source domain and the target domain are highly consistent, otherwise it would be incapable of handling inconsistent modality data in the target domain. The modality consistence between domains is difficult to achieve in reality for several reasons: 1) Hospitals with different resources use different equipment to acquire medical data. For instance, large comprehensive hospitals are typically equipped with a variety of medical devices, such as MRI, CT, ultrasound, etc., while smaller medical institutions may only be equipped with basic medical devices, such as X-ray machines and ultrasound equipment. 2) Personnel in different hospitals have preferences for reading and analyzing various modalities, and hospitals will select appropriate equipment based on the expertise of their staff. Therefore, there exists an asymmetry in data between different institutions, i.e., the data modalities between two medical institutions are either completely inconsistent or only partially consistent. How to address the issue of asymmetric modalities between the source and target domains in SFUDA is an unexplored but important problem.

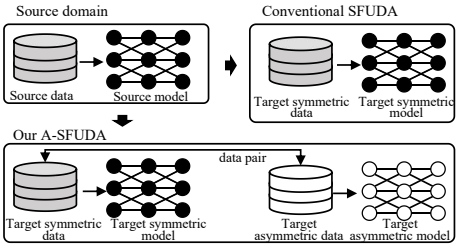

Fig. 1. Illustrations of conventional SFUDA and the proposed A-SFUDA. Conventional SFUDA aims to mitigate the domain gap between symmetric source data and target data. A-SFUDA can learn to predict for both symmetric and asymmetric target data.

Given the complete non-overlap modalities between source and target domains, source domain models become ineffective. Therefore, we aim to explore solutions for the challenge posed by partial non-overlap between the source and target modalities, which is displayed in Fig. 1. Motivated by this consideration, we propose a two-stage asymmetric source-free unsupervised domain adaptation (A-SFUDA) approach. In the first stage, we employ knowledge distillation to preliminarily train on asymmetric modality, thereby facilitating the handling of multi-modal data in the target domain. In the second stage, we introduce a pseudo-label correction mechanism [13] based on multi-modal information to enhance the diagnostic performance on multi-modal data. The pseudo-label correction method stratifies target domain data into high-confident, medium-confident, and low-confident samples by multi-modal prediction correction and class-centered distance correction. Then, the cross-entropy and center loss functions are utilized to train the target models accordingly. Additionally, A-SFUDA employs a consistency loss function based on center distance to enforce consistency in data topology among different modalities and mitigate inter-modal disparities. Experiment on a large public X-ray dataset involves comparisons with six state-of-the-art SFUDA methods. Our proposed approach demonstrates superior performance on symmetric modality data and exhibits scalability in addressing asymmetric modality data, showcasing the versatility of our method. Our contributions can be summarized as follows:

- An A-SFUDA method is proposed for medical diagnosis on both symmetric and asymmetric target data, which is the first time to tackle asymmetric modalities in SFUDA.
- A-SFUDA leverages knowledge distillation and pseudo-label correction to handle asymmetric modality data respectively, while harnessing multimodal information to facilitate the processing of asymmetric data.
- Experiment conducted on A-SFUDA and six SFUDA baseline methods demonstrates that the proposed A-SFUDA is superior for both symmetric and asymmetric modalities.

II. Method

In this section, the proposed asymmetric source-free unsupervised domain adaptation (A-SFUDA) will be introduced. A-SFUDA provides a source model f^{s_M1} trained with labeled source data $D^{s\_M1} = \{\mathbf{x}_i^{s\_M1}, \mathbf{y}_i^{s\_M1}\}_{i=1}^{n_s}$, where D^{s_M1} belongs to modality $M1$, n_s is the number of samples in D^{s_M1}, $\mathbf{y}_i^{s\_M1} \in \{0,1\}^C$ is the label of $\mathbf{x}_i^{s\_M1}$ and C is the number of classes. Concurrently, A-SFUDA provides unlabeled multi-modal target data $D^{t\_M1} = \{\mathbf{x}_i^{t\_M1}\}_{i=1}^{n_t}$ and $D^{t\_M2}\{\mathbf{x}_i^{t\_M2}\}_{i=1}^{n_t}$, where D^{t_M1} is the symmetric target data with the same modality as D^{s_M1}, D^{t_M2} is the asymmetric target data with a different modality from D^{s_M1}, $\mathbf{x}_i^{t\_M1}$ and $\mathbf{x}_i^{t\_M2}$ represent a data pair coming from the same patient, and n_t is the number of target training data. The goal of A-SFUDA is to train target models f^{t_M1} and f^{t_M2} for the two modalities based on the information carried by the source model f^{s_M1}, enabling them to perform diagnosis with the corresponding data.

A-SFUDA consists of two stages, as shown in Fig. 2. The first stage aims to initialize the asymmetric model with the knowledge from the source model. To this end, the knowledge distillation is utilized to inject the information of source model f^{s_M1} into the target asymmetric model f^{t_M2}. The second stage involves enhancing target models using a pseudo-label correction mechanism, employing both probability-defined and class-center-defined pseudo-labeling. The pseudo labels are refined by multi-modal information, which helps to increase the reliability. Ultimately, we advocate for alleviating the center loss discrepancy between the two modalities to enhance cross-modality consistency.

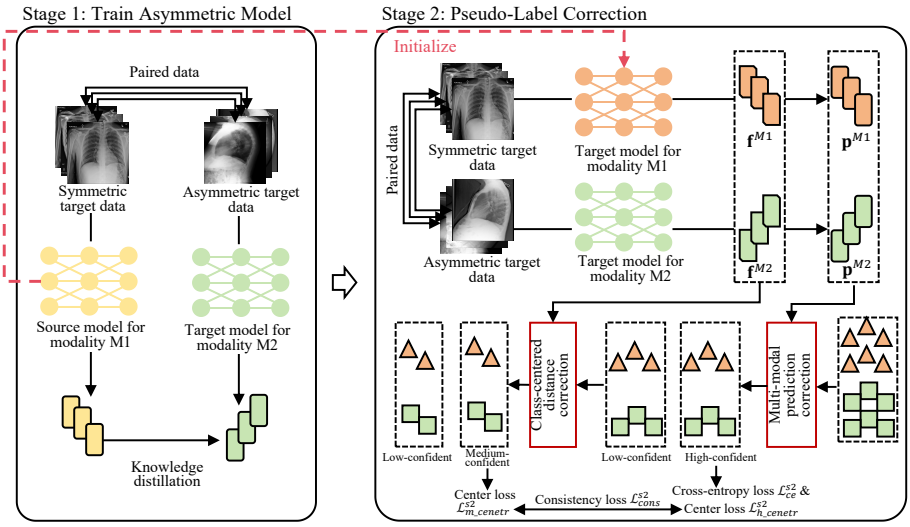

Fig. 2. Framework of A-SFUDA, consisting of two stages. In the first stage, asymmetric model is trained based on knowledge distillation. In the second stage, symmetric and asymmetric target models are optimized by pseudo-label correction. The target samples are divided into high-confident, medium-confident, and low-confident samples by multi-modal prediction correction and class-centered distance correction, respectively. Finally, consistency loss is applied to reduce the discrepancy of the two modalities.

A. Train Asymmetric Model

At this stage, our objective is to preliminarily acquire the symmetric target model f^{t_M1} and the asymmetric target model f^{t_M2}, which can respectively handle symmetric and asymmetric data in the target domain. Given that the source model f^{s_M1} can directly process the symmetric target data, the symmetric target model f^{t_M1} can be initialized from the source model f^{s_M1}, which can be formulated as:

$$\begin{aligned} f^{t\_M1} &= h^{t\_M1}(g^{t\_M1}(\cdot, \theta_g^{t\_M1}), \theta_h^{t\_M1}) \\ \Leftrightarrow f^{s\_M1} &= h^{s\_M1}(g^{s\_M1}(\cdot, \theta_g^{s\_M1}), \theta_h^{s\_M1}), \end{aligned} \tag{1}$$

where "$\cdot$" denotes the input image of a function, $g^{s/t\_M1}(\cdot)$ and $h^{s/t\_M1}(\cdot)$ are feature extractor and classifier with parameters θ_g^{s/t_M1} and θ_h^{s/t_M1}, respectively. Then, the results of symmetric target data can be calculated as $\mathbf{p}^{t\_M1} = f^{t\_M1}(\mathbf{x}^{t\_M1})$, and $\mathbf{p}^{t\_M1}$ can be regarded as the soft label of corresponding paired asymmetric target data. With the soft labels of target data, the asymmetric target model can be trained by knowledge distillation with Kullback–Leibler (KL) divergence [15] as:

$$\mathcal{L}_{kl}^{s1} = \sum KL(\mathbf{p}^{t\_M1}, \mathbf{p}^{t\_M2}) = \sum \mathbf{p}^{t\_M1} log(\frac{\mathbf{p}^{t\_M1}}{\mathbf{p}^{t\_M2}}), \tag{2}$$

where $\mathbf{p}^{t\_M2} = f^{t\_M2}(\mathbf{x}^{t\_M2})$ is the predicted label of $\mathbf{x}^{t\_M2}$.

B. Pseudo-Label Correction

Due to the distribution discrepancy between the source and target domains, direct initialization of the symmetric target model with the source model and training the asymmetric target model through knowledge distillation often fails to demonstrate satisfactory diagnostic performance on the target domain. To address this domain gap, in the second stage, we propose a simultaneous optimization approach for both target models by pseudo-label correction. Firstly, we partition the target data into high-confident and low-confident samples by utilizing multi-modal prediction correction. The joint prediction outcomes from multiple modalities encompass more comprehensive information in comparison to single-modal predictions, resulting in more accurate pseudo-labels. Secondly, to enhance the utilization of low-confident samples, we further subdivide them into medium-confident and low-confident samples based on the class-centered distance correction in the multi-modal feature space. The selection of medium-confident samples reduces the noise ratio within the training data, thereby enhancing the robustness of target models. Finally, we employ a modality alignment mechanism by decreasing the discrepancy between topological structures of the two modalities, where the topological structure represents the distance between class centers and samples for each modality. The specific steps of the entire second phase are as follows.

Multi-modal prediction correction. To divide high-confident and low-confident samples, we utilize the mean value of symmetric and asymmetric predictions as the final result for each sample, and the final result is $\mathbf{p}^t = 0.5 \times (\mathbf{p}^{t\_M1} + \mathbf{p}^{t\_M2})$. Within each class, we select $a\%$ samples with highest prediction values as high-confident samples, as higher prediction values typically indicate more reliable predictions. Regarding the high-confident samples as fully-labeled, we apply the negative cross-entropy loss to optimize the two target models, which can be expressed as:

$$\begin{aligned} \mathcal{L}_{ce}^{s2} &= \mathcal{L}_{ce}^{s2\_M1} + \mathcal{L}_{ce}^{s2\_M2} \\ &= -\sum \mathbf{y}(log(\mathbf{p}^{t\_M1}) + log(\mathbf{p}^{t\_M2})), \end{aligned} \tag{3}$$

where $\mathbf{y} \in \{0,1\}^C$ is the one-hot label calculated from $\mathbf{p}^t$. Additionally, to enhance the intra-class compactness and inter-

class separability of the learned features, we employ the center loss function to optimize the two target models within the feature space. The feature $\mathbf{f}^{h\_M1/2}$ of high-confident target data $\mathbf{x}^{h\_M1/2}$ is formulated as:

$$\mathbf{f}^{h\_M1/2} = g^{t\_M1/2}(\mathbf{x}^{h\_M1/M2}, \theta_g^{t\_M1/2}) \in \mathbb{R}^d, \quad (4)$$

where d represents the dimension of $\mathbf{f}^{h\_M1/2}$. The high-confident class centers are the average value of high-confident samples in each class, which can be formulated as:

$$\mathbf{k}_i^{h\_M1/2} = \frac{1}{n_{h_i}} \sum_{j=1}^{n_{h_i}} \mathbf{f}_j^{h\_M1/2}, \quad (5)$$

where $i = \{1, \cdots, C\}$, $\mathbf{k}_i^{h\_M1/2}$ is the i-th class center for modality $M1/2$, and n_{h_i} represents the number of high-confident samples in the i-th class. Subsequently, the center loss function can be expressed as:

$$\begin{aligned}\mathcal{L}_{h\_center}^{s2} &= \mathcal{L}_{h\_center}^{s2\_M1} + \mathcal{L}_{h\_center}^{s2\_M2} \\ &= \|\hat{\mathbf{F}}^{h\_M1}(\hat{\mathbf{K}}^{h\_M1})^T - \mathbf{Y}\|^2 + \|\hat{\mathbf{F}}^{h\_M2}(\hat{\mathbf{K}}_h^{h\_M2})^T - \mathbf{Y}\|^2,\end{aligned} \quad (6)$$

where $\hat{\mathbf{F}}^{h\_M1/2}$ and $\hat{\mathbf{K}}^{h\_M1/2}$ are the normalized $\mathbf{F}^{h\_M1/2}$ and $\mathbf{K}^{h\_M1/2}$, respectively.

Class-centered distance correction. To improve the usage of low-confident samples, the medium-confident samples are filtered from them by evaluating the class-centered distance with multi-modal information. Firstly, the low-confident samples are ranked by the cosine distances between themselves and corresponding high-confident class center. Then, the top $b\%$ samples with highest cosine similarity are selected from each modality as the initial medium-confident samples, which are symbolized as $Q^{M1/2} = \{\mathbf{X}^{l\_M1/2}\}$. Next, we obtain the final set of medium-confident samples $Q = \{\mathbf{X}^m\}$ by taking the intersection of Q^{M1} and Q^{M2}. This intersection approach allows us to filter out noisy data within the initial medium-confident sample sets, thereby enhancing the robustness of the model. Finally, we utilize the center loss to make medium-confident samples be discriminative. The medium-confident center loss is formulated as:

$$\begin{aligned}\mathcal{L}_{m\_center}^{s2} &= \mathcal{L}_{m\_center}^{s2\_M1} + \mathcal{L}_{m\_center}^{s2\_M2} \\ &= \|\hat{\mathbf{F}}^{m\_M1}(\hat{\mathbf{K}}^{m\_M1})^T - \mathbf{Y}\|^2 + \|\hat{\mathbf{F}}^{m\_M2}(\hat{\mathbf{K}}^{m\_M2})^T - \mathbf{Y}\|^2,\end{aligned} \quad (7)$$

where $\hat{\mathbf{K}}^{m\_M1/2}$ is the normalized $\mathbf{K}^{m\_M1/2}$, and $\mathbf{K}^{m\_M1/2}$ is the medium-confident class centers calculated from medium-confident samples as Eq. (5).

Modality alignment mechanism. Given that the model training in the aforementioned steps solely exploits multi-modal information, disregarding the consistency between the two modalities, we propose to calculate the consistency between samples and class centroids for mitigating the dissimilarity in data topological structures. The consistency of topological structures for two modality can be formulated as:

$$\begin{aligned}\mathcal{L}_{cons}^{s2} = &\|\hat{\mathbf{F}}^{h\_M1}(\hat{\mathbf{K}}^{h\_M1})^T - \hat{\mathbf{F}}^{h\_M2}(\hat{\mathbf{K}}^{h\_M2})^T\|^2 \\ &+ \|\hat{\mathbf{F}}^{m\_M1}(\hat{\mathbf{K}}^{m\_M1})^T - \hat{\mathbf{F}}^{m\_M2}(\hat{\mathbf{K}}^{m\_M2})^T\|^2.\end{aligned} \quad (8)$$

This loss function underscores that samples from both modalities within a data pair should exhibit equal distances to their respective class centers, thereby ensuring consistency across the feature spaces of the two modalities.

To sum up the above steps, the overall loss function of the second stage is:

$$\mathcal{L}^{s2} = \mathcal{L}_{ce}^{s2} + \alpha\mathcal{L}_{h\_center}^{s2} + \beta\mathcal{L}_{m\_center}^{s2} + \sigma\mathcal{L}_{cons}^{s2}, \quad (9)$$

where α, β, and σ are hyper-parameters to balance the three components.

III. Experiment

In this section, we conduct the experiments to assess the proposed A-SFUDA. Firstly, we introduce the experimental datasets and evaluation protocols. Then, the compared baseline methods and experimental details are provided. Finally, we present and analyze the experimental results.

A. Dataset and Evaluation Protocols

Source dataset. ChestX-Ray14 [16] serves as the source data in our experiment. It consists of 112,120 frontal-view X-ray images obtained from 30,805 distinct patients. The dataset is annotated with fourteen commonly occurring disease labels, extracted through text-mining techniques. We train the source model based on the officially designated training, validation, and test sets. The training, validation, and test sets comprise 78,468, 11,219, and 22,433 images, respectively. Images depicting the presence of diseases are considered abnormal, while those without any pathology are categorized as normal.

Target dataset. We evaluate A-SFUDA on MIMIC-CRX [17] dataset as the target dataset, a large-scale dataset of chest X-ray images. MIMIC-CRX consists of frontal and lateral chest X-ray images and corresponding clinical reports from multiple hospitals in the Boston area. The dataset covers lung disease information from approximately 65,000 patients, with a total of more than 350,000 chest X-ray images. Among these, there are about 250,000 frontal X-ray images and 89,000 lateral X-ray images. The dataset includes various types of chest diseases and pathologies, including pneumonia, pulmonary edema, tuberculosis, and normal cases, spanning 14 categories. According to our statistics, there are 89,497 patients with both frontal and lateral X-ray images. To save training time, we randomly sampled data from these 89,497 patients at a ratio of 0.2, resulting in an experimental dataset of 17,900 images. Subsequently, we randomly split the extracted dataset into a training set (12,530 patients) and a test set (5,370 patients) using an 8:2 ratio. If a patient's report states "no findings" the patient is classified as normal; otherwise, the patient is classified as abnormal. Frontal-view images are regarded as symmetrical modality (modality $M1$) data, while lateral-view images are considered as asymmetrical modality (modality $M2$) data in the target domain.

For the evaluation protocols, we employ Accuracy, Precision, Recall, and F1-Score simultaneously for comprehensive

classification evaluation. These metrics offer distinct perspectives, aiding in the assessment of the performance across various aspects.

B. Baselines and Experimental Details

Baselines. To comprehensively analyze and evaluate our framework, we select six state-of-the-art SFUDA methods for comparison, i.e., UB^{2}DA [18], NRC_SFDA [19], DINE [20], CoWA_JMDS [21], BPDA [22], and GPL_UE [23]. They encompass pseudo-labeling [18], [21], [23], knowledge distillation [20], adversarial training [22], and metric learning [19]. Such a comprehensive integration enables a more holistic evaluation of our proposed approach.

Experimental Details. In our experimental setup, ResNet18 serves as the baseline model, and all experiments are conducted on an RTX 3090 GPU. The images used in the experiments are scaled to 128x128 pixels. Mini-batch SGD is employed with a learning rate of 1e-3. The batch size is set to 512, and the training epoch is set to 15. The hyper-parameters α, β, and σ are respectively set to 0.2, 0.2, and 0.2. The ratio of high-confident samples is set as 30%, and the ratio of medium-confident samples is set as 40%.

TABLE I
THE ACCURACY (ACC), PRECISION (P), RECALL (R), AND F1-SCORE (F1) RESULTS OF A-SFUDA AND SIX BASELINES AT MIMIC-CRX DATASET . THE **BEST RESULTS** AND SECOND-BEST RESULTS ARE MARKED WITH CORRESPONDING FORMATS. "M" MEANS MODALITY.

Method	M	ACC(%)	P(%)	R(%)	F1(%)	AVG(%)
Source only	$M1$	68.52	61.78	72.26	66.61	67.29
UB^{2}DA	$M1$	68.51	60.61	**78.56**	68.43	69.03
NRC_SFDA	$M1$	71.38	66.93	72.01	69.37	69.92
DINE	$M1$	71.52	65.07	74.58	69.50	70.17
CoWA_JMDS	$M1$	68.10	67.16	51.99	58.61	61.47
BPDA	$M1$	71.08	66.64	71.58	69.02	69.58
GPL_UE	$M1$	71.45	**68.48**	63.52	65.91	67.34
A-SFUDA	$M1$	**71.54**	64.32	77.45	**70.28**	**70.90**
KD	$M2$	67.24	59.65	**75.99**	66.84	67.43
A-SFUDA	$M2$	**70.68**	**63.62**	75.95	**69.24**	**69.87**

TABLE II
ABLATION STUDIES OF A-SFUDA. "M" MEANS MODALITY.

Method	M	ACC(%)	P(%)	R(%)	F1(%)	AVG(%)
A-SFUDA w/o C1	$M1$	69.98	63.12	74.32	68.26	68.92
	$M2$	69.06	62.17	73.55	67.38	68.04
A-SFUDA w/o C2	$M1$	70.03	62.58	77.15	69.11	69.72
	$M2$	67.89	60.99	72.39	66.20	66.87
A-SFUDA w/o C3	$M1$	70.52	62.82	78.56	69.85	70.44
	$M2$	70.56	64.85	72.69	68.55	69.16
A-SFUDA	$M1$	71.54	64.32	77.45	70.28	70.90
	$M2$	70.68	63.62	75.95	69.24	69.87

C. Experimental Results

Comparison with State-of-the-arts. Table I shows the experimental results of six baselines and A-SFUDA in both frontal and lateral modalities across four evaluation metrics. In terms of Modality $M1$, A-SFUDA achieves the best average performance of four metrics. Notably, it achieves the optimal results in both Accuracy and F1-score, underscoring the superior effectiveness of the A-SFUDA in handling symmetric modality. However, in terms of precision, A-SFUDA lags behind the best-performing method by 4%, suggesting the need for further improvement in precision. Additionally, A-SFUDA exhibits the capability to handle asymmetric modality data compared to other methods. Furthermore, it can be observed that the performance of A-SFUDA in Modality $M2$ is slightly diminished by 1% compared to Modality $M1$. This discrepancy may be attributed to the fact that the model for $M1$ undergoes initialization from the source model, leading to comparatively better results than Modality $M2$. Despite this, when compared to other best-performing methods in Modality $M1$, A-SFUDA in $M2$ is only 0.3% lower. This affirms the ability of A-SFUDA to effectively diagnose target data from both modalities simultaneously.

Ablation Study. A-SFUDA comprises three main components: 1) C1: correction of pseudo-labels for high-confident samples based on multi-modal prediction correction; 2) C2: filtration of medium-confident samples utilizing class-centered distance correction; and 3) C3: a modality alignment mechanism based on data topological structure. To assess the effectiveness of these three modules, we conducted experiments by individually removing each component from A-SFUDA. The experimental results are presented in Table II. In a holistic assessment, the performance of A-SFUDA exhibited a noticeable decline when each of the three modules is individually excluded, underscoring the efficacy of these components. Specifically, for Modality $M1$, the removal of the C1 module resulted in the most significant performance deterioration, emphasizing the supplementary role of multi-modal information in handling symmetrical modality. For Modality $M2$, the exclusion of the C2 module leads to a 3% decrease in performance, which may be attributed to the absence of pre-training for asymmetric target model. It suggests that the inclusion of medium-confident samples in the training set enhances the generalization of the asymmetric model. Moreover, the integration of the C3 module into A-SFUDA yielded performance improvements in both modalities, highlighting the significance of modality alignment.

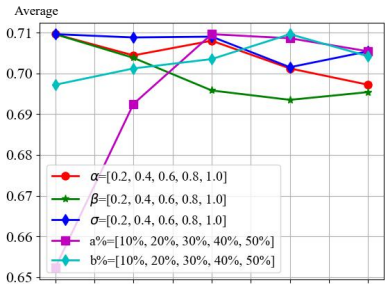

Fig. 3. Parameters analysis for A-SFUDA.

Parameter Analysis. There are five parameters in A-SFUDA, i.e., α, β, σ in Eq. (9), and the ratios a% and b% of high-confident and medium-confident samples. We conducted experiments on the frontal-view image dataset to

determine the optimal parameter values. The ranges for α, β, σ are [0.2, 0.4, 0.6, 0.8, 1.0], and for a% and b%, the ranges are [10%, 20%, 30%, 40%, 50%]. Experimental results, illustrated in Fig. 3, indicate that the best performance is achieved when α, β, σ, a%, and b% are set to 0.2, 0.2, 0.2, 30%, and 40%, respectively. Furthermore, it is noteworthy that the selection of a% significantly influences the performance of A-SFUDA, underscoring the importance of choosing an appropriate proportion of high-confident samples.

IV. Conclusion

A-SFUDA addresses the challenge of partial non-overlap between source and target modalities in the context of medical diagnosis. The first stage of A-SFUDA utilizes knowledge distillation to preliminarily train on asymmetric modalities, laying the groundwork for effectively handling multi-modal data in the target domain. Subsequently, in the second stage, we implement a pseudo-label correction mechanism based on multi-modal prediction correction and class-centered distance correction. This mechanism categorizes target domain data into high-confident, medium-confident, and low-confident samples, ensuring the refinement of pseudo labels. Experimental results on a large public X-ray dataset validate the superior performance of our proposed A-SFUDA. A-SFUDA not only excels on symmetric modality data but also showcases its scalability in addressing the challenges posed by asymmetric modalities. In the future, we plan to explore A-SFUDA methodology in the context of non-paired data across different modalities.

Acknowledgment

This work was supported by the Research Grants Council of the Hong Kong SAR (Grant No. PolyU11211521, PolyU15218622, PolyU15215623, and PolyU25216423), The Hong Kong Polytechnic University (Project IDs: P0039734, P0035379, P0043563, and P0046094), and the National Natural Science Foundation of China (Grant No. U21A20512, 62306259, and 62202399).

References

[1] G. Wang et al., "Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning," IEEE Trans. Med. Imaging, vol. 37, no. 7, pp. 1562-1573, July 2018, doi: 10.1109/TMI.2018.2791721.

[2] Z. -A. Huang, Z. Zhu, C. H. Yau and K. C. Tan, "Identifying Autism Spectrum Disorder From Resting-State fMRI Using Deep Belief Network," IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 7, pp. 2847-2861, July 2021, doi: 10.1109/TNNLS.2020.3007943.

[3] R. Liu, Z. -A. Huang, Y. Hu, Z. Zhu, K. -C. Wong and K. C. Tan,"Spatial–Temporal Co-Attention Learning for Diagnosis of Mental Disorders From Resting-State fMRI Data," IEEE Trans. Neural Netw. Learn. Syst., doi: 10.1109/TNNLS.2023.3243000.

[4] Z. -A. Huang, R. Liu, Z. Zhu and K. C. Tan, "Multitask Learning for Joint Diagnosis of Multiple Mental Disorders in Resting-State fMRI," IEEE Trans. Neural Netw. Learn. Syst., doi: 10.1109/TNNLS.2022.3225179.

[5] M. Li, R. Liu, F. Wang et al, "Auxiliary Signal-Guided Knowledge Encoder-Decoder for Medical Report Generation." in World Wide Web, vol. 1, pp. 253–270, 2023.

[6] Y. Hu et al., "Source Free Semi-Supervised Transfer Learning for Diagnosis of Mental Disorders on fMRI Scans," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 11, pp. 13778-13795, 1 Nov. 2023, doi: 10.1109/TPAMI.2023.3298332.

[7] T. Sun, C. Lu, H. Ling, "Prior Knowledge Guided Unsupervised Domain Adaptation," in Proc. Eur. Conf. Comput. Vis. (ECCV), vol 13693, pp. 639-655, 2022.

[8] J. Huang, D. Guan, A. Xiao, S. Lu and L. Shao, "Category Contrast for Unsupervised Domain Adaptation in Visual Tasks," in IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, 2022, pp. 1193-1204, doi: 10.1109/CVPR52688.2022.00127.

[9] J. Pei, Z. Jiang, A. Men, L. Chen, Y. Liu and Q. Chen, "Uncertainty-Induced Transferability Representation for Source-Free Unsupervised Domain Adaptation," IEEE Trans. Image Process., vol. 32, pp. 2033-2048, 2023, doi: 10.1109/TIP.2023.3258753.

[10] S. Qiu, "Causality-Inspired Source-Free Domain Adaptation for Medical Image Classification," in Proc. Int. Conf. Image Graphic, pp. 68-80, 2023.

[11] C. Chen et al., "Source-Free Domain Adaptive Fundus Image Segmentation with Denoised Pseudo-Labeling," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), Strasbourg, France, September 27–October 1, 2021.

[12] R. Liu, Z. -A. Huang, Y. Hu, Z. Zhu, K. -C. Wong and K. C. Tan, "Attention-Like Multimodality Fusion With Data Augmentation for Diagnosis of Mental Disorders Using MRI," IEEE Trans. Neural Netw. Learn. Syst., 2022, doi: 10.1109/TNNLS.2022.3219551.

[13] Y. Hu, Z. -A. Huang, R. Liu, X. Xue, L. Song and K. Chen Tan, "A Dual-Stage Pseudo-Labeling Method for the Diagnosis of Mental Disorder on MRI Scans," in Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN), Padua, Italy, 2022, pp. 1-8, doi: 10.1109/IJCNN55064.2022.9892792.

[14] Z. Wang, et al., "M-MSSEU: Source-Free Domain Adaptation for Multi-Modal Stroke Lesion Segmentation Using Shadowed Sets and Evidential Uncertainty," Health Information Science and Systems, no.1, pp. 46, 2023.

[15] G. Hinton, O. Vinyals, J. Dean, "Distilling the knowledge in a neural network," arXiv preprint arXiv:1503.02531, 2015.

[16] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri and R. M. Summers, "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," in IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, pp. 3462-3471, 2017, doi: 10.1109/CVPR.2017.369.

[17] A. Johnson et al., "MIMIC-CXR, a De-Identified Publicly Available Database of Chest Radiographs with Free-Text Reports," Scientific Data, vol, 6, no. 1, pp. 317, 2019.

[18] B. Deng, Y. Zhang, H. Tang, C. Ding, K. Jia, "On Universal Black-Box Domain Adaptation," arXiv preprint arXiv:2104.04665, 2021.

[19] S. Yang et al., "Exploiting the Intrinsic Neighborhood Structure for Source-Free Domain Adaptation," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 34, pp. 29393-29405, 2021.

[20] J. Liang, D. Hu, J. Feng and R. He, "DINE: Domain Adaptation from Single and Multiple Black-box Predictors," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, 2022, pp. 7993-8003, doi: 10.1109/CVPR52688.2022.00784.

[21] J. Lee et al., "Confidence Score for Source-Free Unsupervised Domain Adaptation," in Proc. Int. Conf. Mach. Learn. (ICML), pp. 12365-12377, 2022.

[22] Y. Shi et al., "Source-free and Black-box Domain Adaptation via Distributionally Adversarial Training," Pattern Recognit., p. 109750, 2023.

[23] M. Litrico, A. Del Bue, P. Morerio, "Guiding Pseudo-Labels With Uncertainty Estimation for Source-Free Unsupervised Domain Adaptation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 7640-7650, 2023.