Leveraging Anatomical Consistency for Multi-Object Detection in Ultrasound Images via Source-free Unsupervised Domain Adaptation

Bin Pu[1], Xingguo Lv[2], Jiewen Yang[1], Xingbo Dong[2], Yiqun Lin[1], Shengli Li[3], Kenli Li[4†], Xiaomeng Li[1†]

[1]The Hong Kong University of Science and Technology, HKSAR, China
[2]Anhui University, Hefei, China
[3]Shenzhen Maternity and Child Healthcare Hospital, Southern Medical University, Shenzhen, China
[4]Hunan University, Changsha, China
{eebinpu, eexmli}@ust.hk, xingbo.dong@ahu.edu.cn, lkl@hnu.edu.cn

Abstract

Source-free unsupervised domain adaptation aims to eliminate domain shifts when data from the source domain and annotation from the target domain are not available. The multi-object detection tasks in medical image analysis are constrained by patient privacy and extremely huge annotation consumption. Hence, Source-free UDA is considered a more practical approach for eliminating the domain gap. However, relevant research that explores this topic is a dearth. In this paper, we design an ***A***natomy-aware ***A***lignment ***T***eacher-***S***tudent learning method using topological consistency based on a mean-teacher framework for Source-free UDA in multiple medical object detection named ***AATS***, including Unsupervised Structure Refinement (USR) and Graph-aware Morphology Alignment (GMA). To match the student and teacher at the low-level and visual features, we propose the USR via an unsupervised clustering algorithm to group organs in ultrasound images. Based on USR, we obtain a graph with organ relations on the teacher branch. While in the student branch, we acquire visual features to construct graphical space and optimize the model with graph propagation. Finally, to match the student and teacher, GMA is designed to align the teacher and student based on both topology and morphology information that is derived from prior medical knowledge. Four groups of adaptation experiments were conducted on available medical datasets, and the outcomes demonstrate that our approach not only achieves state-of-the-art performance but also provides substantial advantages over existing methods.

Code — https://github.com/xmed-lab/AATS

Introduction

Medical multi-object detection deep models, which is essential for downstream tasks such as imaging quality control (Lin et al. 2019), standardized automatic screening (Baumgartner et al. 2017), and auxiliary diagnosis (Nguyen et al. 2021; Yang et al. 2024), are often deployed across several different domains/hospitals. However, there are wide domain shifts between the data collected by different hospitals due to the use of different medical devices and the skills of the scanning physicians. Therefore, d1irectly applying a model trained within medical center *A* to data from medical center *B* may result in diminished detection performance or render the approach ineffective. Meanwhile, obtaining a large number of detailed box-based annotations for different domains is usually costly/ unavailable. Domain adaptation techniques have been developed to address the cross-domain adaptation problem in the absence of labeled target domain data, such as unsupervised domain adaptation object detection (UDAOD) (Oza et al. 2023) and Source-free UDAOD (Li et al. 2024).

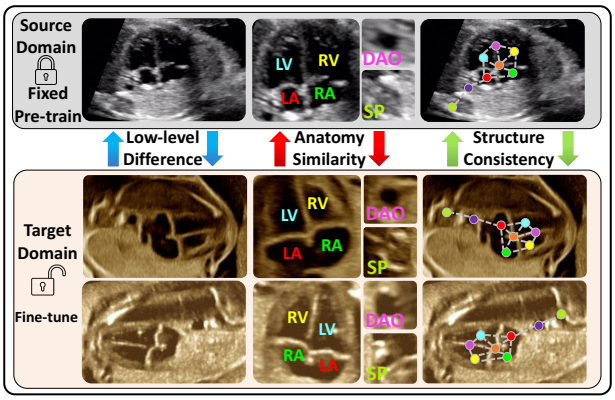

Figure 1: Motivation of our AATS. Left Ventricle (LV), Right Ventricle (RV), Left Atrium (LA), Right Atrium (RA), Spine (SP), and Descending Aorta (DAO). Our adaptation technique obviates the need for source domain data and leverages solely the pre-trained weights that suffice to safeguard the privacy of the source domain. Upon observations, it becomes evident that capitalizing on the inherent topological consistency of anatomical structures across various domains can effectively bridge domain gaps.

UDAOD requires source-domain data to be trained together with target-domain data, often utilizing techniques such as invariant feature learning (Vs et al. 2021), self-training (Zhao et al. 2020), image translation (Chen et al. 2020), and domain randomization (Kim et al. 2019) to improve detection performance. In some specific real-world medical scenarios, patient privacy concerns or data storage limitations may prevent the availability of the source data.

†Corresponding authors.

Contrary to traditional UDA approaches, Source-free UDA eliminates the necessity for access to source domain data, relying instead on lightweight pre-trained weights from the source domain. This characteristic suggests that Source-free UDA methods may offer greater applicability and practicality in real-world medical scenarios.

However, most previous source-free UDA methods (Shen et al. 2023; Yang et al. 2022; Qiu et al. 2021; Kundu et al. 2022) have focused on natural scene. While pioneering works (Liu, Li, and Yuan 2023; Liu and Yuan 2022) on Source-free UDA have been explored in single-object detection in medical scenarios, their application in clinical settings underscores a critical gap. In the realm of healthcare, the significance of multi-object detection extends beyond mere identification; it serves as a vital tool for training novice physicians and enhancing the accuracy of assisted diagnoses (He et al. 2024). The interdependence of medical objects means that the absence or anomaly of one can indicate complications or deficiencies in others, highlighting the indispensable role of multi-object detection in addressing such complexities. Furthermore, a body of research has consistently highlighted the necessity and benefits of multi-object detection in medical imaging, advocating for its integration to improve diagnostic processes and patient outcomes (Lin et al. 2019). The previous neglect and limited exploration of medical multi-object cross-domain adaptation detection underscore the unique challenges in this field. Unlike Sourc-free UDAOD for natural images, medical images come from human body scanning and contain a lot of a priori knowledge, such as extensive anatomical knowledge. For example, in natural image detection scenarios, where targets are irregular and cluttered, the very popular SIGMA (Li, Liu, and Yuan 2022) and SIGMA++ (Li, Liu, and Yuan 2023) may not be applicable in medical image analysis due to their failure to capture anatomical knowledge well. In medical scenarios, the prototype-anchored feature alignment (Yu et al. 2023) may not provide alignment of the topological information of invariant anatomical structures, leading to adaptation performance degradation. In this work, to exploit the anatomical knowledge between the medical multi-objects, we propose a ***A***natomy-aware ***A***lignment ***T***eacher-***S***tudent (AATS) learning framework for Sourc-free UDA detection task. This represents the first application of a source-free UDA method in medical detection.

Based on our observations, medical images exhibit a wide variation of the same object/organ across different domains (Pu et al. 2021), especially in medical ultrasound images (Chen et al. 2017). However, Fig. 1 illustrates that medical objects from various domains exhibit visual variability while maintaining topological consistency. Therefore, relying on prior knowledge of topological graphs is expected to be more reliable. To this end, we propose an Unsupervised Structure Refinement (USR) module based on the source-free mean-teacher learning framework (Tarvainen and Valpola 2017), which constructs at the center of each type of organ cluster to match visual features, local organ features, and low-level features of different domains. Subsequently, we build both a student-graph and a teacher-graph by transforming visual features into the graphical space and performing graph propagation. Finally, we align the a priori topological knowledge of organ structures from the source and target domains by optimizing and matching the teacher-graph and student-graph. The main contributions can be summarized as follows.

- We propose the first source-free UDA method for the medical multi-target detection task called AATS, which can serve as a first benchmark for the multi-object adaptation detection task in medical applications.
- We propose an Unsupervised Structure Refinement (USR) method by organ structure clustering to align the low-level, visual, and organ local features between different domains.
- We construct the student graph and teacher graph and propose a Graph-aware Morphology Alignment (GMA) method to align consistent topological knowledge from source and target domains.
- Extensive adaptation experiments are conducted on the collected and publicly available datasets, and the experiment results demonstrate that our AATS achieves state-of-the-art performance with a clear margin.

Related Works

Source-Free UDA Object Detection in Nature Images

Before the exploration of Source-Free Unsupervised Domain Adaptation Object Detection (SFUDAOD) topic, investigators focused on unsupervised domain adaptation object detection (UDAOD) (Chen et al. 2018; Vs et al. 2021; Wang et al. 2021; Zhao and Wang 2022; Yu et al. 2022; Chen et al. 2021; Kim et al. 2019; Deng et al. 2021; Li et al. 2022b; Cao et al. 2023; Deng et al. 2023). With the urgent need for data privacy protection, SFUDAOD has emerged as an emerging branch of data protection detection. Due to the complexity of the object detection task and the unavailability of data from the source domain, the direct use of UDAOD to SFUDAOD does not yield satisfactory results (Li et al. 2022c; Yuan et al. 2022). Recently, a series of UDAOD methods (Li et al. 2022a; Chu et al. 2023; VS, Oza, and Patel 2023a,b; Oza et al. 2023; Liu et al. 2023; Hegde and Patel 2024) have emerged for object detection in natural scenarios. (Li et al. 2021) defined it as a learning problem with noisy labels and generates pseudo-labels from source domain pretrained networks for target domain optimization by finding a suitable confidence threshold for reliable pseudo-label generation. (Chen, Wang, and Zhang 2023) introduced the high and low confidence thresholds together to take full advantage of pseudo-labels.

However, previous approaches have explored the SFUDAOD in natural scenes, and the source-free domain multi-object detection in medical images has not been fully explored. The medical image contains unique a priori information, so the results may be unsatisfactory if it is directly used in natural scenario approaches to medical scenarios.

Source-Free UDA in Medical Scenarios

Recently, medical image classification (Song et al. 2024; Qiu 2023; Dai et al. 2022), especially segmentation (Zheng

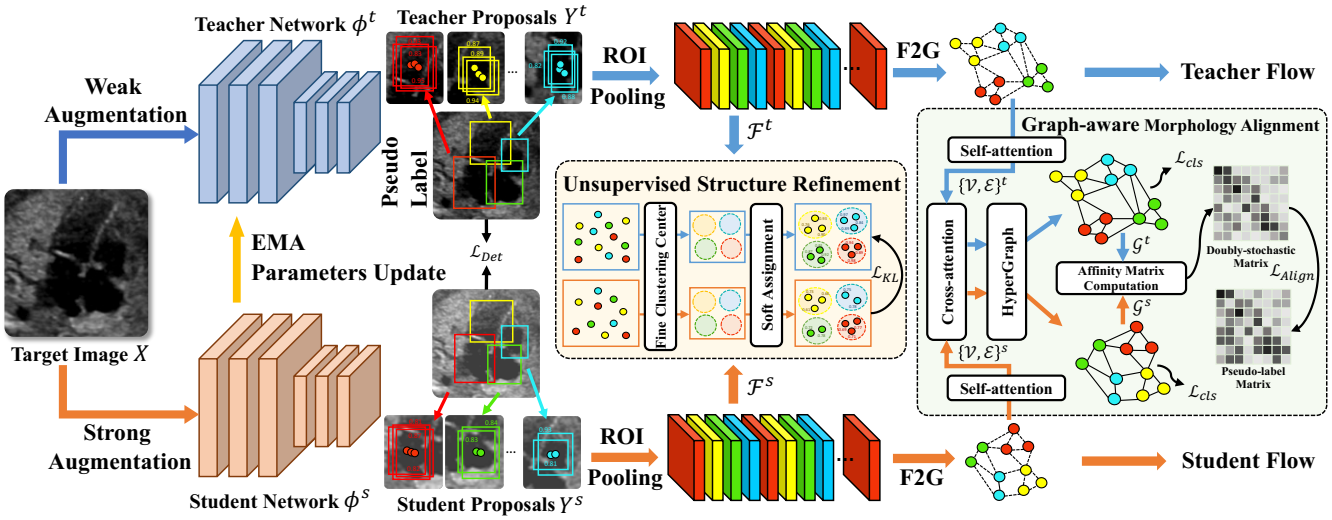

Figure 2: **Overview of our proposed AATS.** During the training phase, the teacher network generates pseudo-labels to train the student while the student updates the teacher using Exponential Moving Average (EMA). The proposed USR uses unsupervised clustering to obtain the alignment of each organ between networks. GMA constructs a graph representation of organ morphology to achieve teacher-student alignment. During the inference phase, only the trained teacher network is utilized. F2G represents the transformation of the vision features into graph nodes.

et al. 2023; Wang et al. 2023; Wu et al. 2023; Dinsdale, Jenkinson, and Namburete 2023; Yu et al. 2023; Yang, Liu, and Yuan 2023; Huai et al. 2023; Yang et al. 2023) in the source-free domain adaptation task, has drawn a lot of attention, which limits its application. (Song et al. 2024) proposed an uncertainty-aware adaptive distillation for the multi-source-free UDA setting by knowledge distillation.

To date, only two studies (Liu, Li, and Yuan 2023; Liu and Yuan 2022) have explored SFUDAOD in 2D medical medical scenarios. However, both of these methods are about single object detection, i.e., polyp detection. Unlike the single object detection task, multiple objects share a topological relationship with each other in medical scenarios from the source and target domains. This work will serve as the first sour-free domain adaptation multi-object detection benchmark in medical image analysis by alignment of the priori information between different domains based on graph-aware topological knowledge.

Methodology

Fig. 2 shows an overview of our method. In an efficient mean-teacher framework, the teacher network provides stable and consistent predictions for unlabeled data unaffected by fluctuations in the data from a single training batch with weak augmentation, thus offering a stable learning target for the student network. The student network enhances the learning ability of low-level features, improving the generalization ability by strong data augmentation. The teacher network is not updated parameters by backpropagation during the training but through EMA with the parameters of the student. Unsupervised Structure Refinement (USR) keeps the information consistent by matching the sample distribution between the student and teacher networks by refining the clustering centers of the organs in an unsupervised manner and iteratively optimizing the clusters. Graph-aware Morphology Alignment (GMA) obtains proposal features from both teacher and student through USR and builds graph representations for teacher-student inter-organ morphological alignment. This is achieved by computing the affinity matrix using intra-domain self-attention and inter-domain cross-attention.

Unsupervised Structure Refinement

As illustrated in Fig. 2, we use the superscript s and t to denote the elements belonging to student and teacher flow, respectively. Firstly, the detected bounding boxes $Y_b^t \in \mathbb{R}^{n^t \times 4}$, $Y_b^s \in \mathbb{R}^{n^s \times 4}$ and corresponding class of bounding boxes $Y_c^t \in \mathbb{R}^{n^t}$, $Y_c^s \in \mathbb{R}^{n^s}$ are predicted by the teacher netowrk $\phi^t(\dot{X})$ and student netowrk $\phi^s(\ddot{X})$, respectively, where $n^t, n^s \in \mathbb{Z}^+$ are the detected organ number. The student network takes the detected objects of Y^t with maximum detected confidence as the ground truth label for supervised learning. The loss that is formulated with the conventional object detection as:

$$\mathcal{L}_{\text{Det}} = \text{IoU}(Y_b^s, Y_b^t) + \text{CE}(Y_c^s, Y_c^t), \tag{1}$$

where $\text{IoU}(\cdot,\cdot)$ measures the spatial overlap between the predicted and the true bounding boxes and $\text{CE}(\cdot,\cdot)$ denotes the cross-entropy loss. With the predicted results, we use the region of interest pooling to generate in total $M^{t/s}$ averaged features $\mathcal{F} = \{f_i^{t/s}\}_{i=0}^{M^{t/s}} \subset \mathbb{R}^d$ with d channels of all detected region. In this step, we utilize the single feature map for all the proposals generated by the region proposal network (Ren et al. 2015) in a single pass.

As shown in Fig. 2, $\mathcal{F}$ are regarded as the represented feature of organs and represent their overall morphology

information for both teacher and student networks. Hence, we treat these features as visual representations of organs, which should make consistency between the teacher and student networks based on prior knowledge. Moreover, due to the non-matching number and unordered features $\mathcal{F}$ of different organs from both networks during training, the pair-wise distance is incapable of measuring the discrepancy. To tackle this problem, we propose unsupervised clustering for acquiring the alignment of each organ between networks, which allows iterative refinement via soft assignment.

In our approach, we incorporate the medical prior knowledge, where the soft assignment in the distribution of detected organs in the student network matches the teacher network. This alignment helps fine-tune the teacher-student network on the target domain data by incorporating the medical prior knowledge. Given the unordered features $\mathcal{F}^{t/s}$ of different classes from teacher and student networks. In the teacher network, we apply the K-means clustering algorithm with the K set to the total class number N and compute each cluster centroid as $c^t \in \mathbb{R}^d$. For both networks, we first let the number of features in $\mathcal{F}^t$ and $\mathcal{F}^s$ equally by randomly dropping the elements of $\mathcal{F}$ to the quantity of M. Then, we compute the probability distribution for each $f^{t/s}$ in $\mathcal{F}^{t/s}$ as:

$$p_{ij}^{t/s} = \frac{(1 + ||f_i^{t/s} - c_j^t||^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{j=0}^{N}(1 + ||f_i^{t/s} - c_j^t||^2/\alpha)^{-\frac{\alpha+1}{2}}}, \tag{2}$$

where p_{ij} define the probability of assigning i-th feature to j-th cluster, α is the degree of the freedom of the disturibution. In our experiment, we set α=1 (Van Der Maaten 2009).

To iteratively refine the clusters by learning from the high-confidence assignments, the probability distributions between the student and teacher network remain consistent due to the structural information being the same in both $\dot{X}$ and $\ddot{X}$. We hence define our objective as the distribution alignment, where the p_{ij}^s align to the target distribution p_{ij}^t, by using the Kullback–Leibler divergence minimization as:

$$\mathcal{L}_{\text{KL}} = \sum\nolimits_{i=0}^{M} \sum\nolimits_{j=0}^{N} \left(p_{ij}^s log \frac{p_{ij}^s}{p_{ij}^t} \right). \tag{3}$$

According to Eq. (3), we are able to refine the teacher-student framework by presenting a more cohesive representation for the detected regions of the same class. Also, the distribution alignment allows the teacher and student to reduce the accumulation of noise by using only one high-confidential pseudo label for supervising the student.

Graph-aware Morphology Alignment

The above discussion illustrates that we match the probability distribution of soft assignments across all detected regions in the student to the teacher network. In this section, we highlight how we construct the graph representation for the morphology information alignment. The other detected regions with lower confidence than the optimal choice are also able to provide reference information for training. Moreover, the involvement of such information can also prevent teacher-student training from being biased against noise that is produced by sparse pseudo-labels. Since the input $\dot{X}$ and $\ddot{X}$ for teacher $\phi^t(\cdot)$ and student $\phi^s(\cdot)$ networks share the same structural information both locally and globally. For example, the representation of the overall structure of the organ and each sub-structure of the organ are consistent.

To construct such representation, we hence employ the graph to bridge the inner connection of input as shown in Fig. 2. In USR, we acquire the represented feature $\mathcal{F}^{t/s}$ for both teacher and student networks. To implement the inner interaction, we use the self-attention mechanism for $\mathcal{F}^{t/s}$ and compute the attention mapping, which is formulated as:

$$\mathcal{E}^{t/s} = \mathcal{F}^{t/s}\mathcal{W}_q(\mathcal{F}^{t/s}\mathcal{W}_k)^{\mathsf{T}} \in \mathbb{R}^{M^{t/s} \times M^{t/s}}, \tag{4}$$

where $\mathcal{E}$ is the attention map computed by the pair-wise distance between each feature of $\mathcal{F}$, and $\mathcal{W}_q, \mathcal{W}_v \in \mathbb{R}^{d \times d}$ denote the weight of linear layer that project the feature f of $\mathcal{F}$ to latent space as query and key vectors. To finish the inner interaction, we then have the following equation:

$$\mathcal{V}^{t/s} = \lambda \cdot \mathcal{E}^{t/s}(\mathcal{F}^{t/s}\mathcal{W}_v) \in \mathbb{R}^{M^{t/s} \times d}, \tag{5}$$

where the $\mathcal{V}$ is the inner interacted feature of the $\mathcal{F}$, λ is the scaling factor that set as $1/\sqrt{d}$ and $\mathcal{W}_v \in \mathbb{R}^{d \times d}$ is the weight of linear layer that project the $\mathcal{F}$ to latent space as the value vector. Furthermore, since the graphs constructed from teacher and student networks are essentially homogeneous, we also implement the cross-attention for $V^{t/s}$ to build a more robust representation, which is denoted as:

$$\hat{\mathcal{V}}^{t/s} = \lambda \cdot \left((\mathcal{V}^{t/s}\hat{\mathcal{W}}_q)(\mathcal{V}\hat{\mathcal{W}}_k)^{\mathsf{T}} \right) (\mathcal{V}\hat{\mathcal{W}}_v);\ \mathcal{V} = \mathcal{V}^t \oplus \mathcal{V}^s, \tag{6}$$

where $\hat{\mathcal{V}} \in \mathbb{R}^{M^{t/s} \times d}$ are the cross interacted feature, $\hat{\mathcal{W}}_q, \hat{\mathcal{W}}_k, \hat{\mathcal{W}}_v \in \mathbb{R}^{d \times d}$ share the same attribution as $\mathcal{W}$ in Eq. (5), λ is the scaling factor as the same as Eq. (5) that set as $1/\sqrt{d}$, the $\oplus$ and the $\mathcal{V} \in \mathbb{R}^{(M^t + M^s) \times d}$ denotes the concatenation operation and concatenated feature.

To this end, the visual graph construction for the feature of both teacher and student networks can then be formulated as $\{\hat{\mathcal{V}}, \mathcal{E}\}^{t/s}$, where $\hat{\mathcal{V}}$ is the feature transformed by Equations 5 and 6, while $\mathcal{E}$ is the connectivity of the feature $\mathcal{F}$. With the visual graph $\{\hat{\mathcal{V}}, \mathcal{E}\}^{t/s}$, we complete the graphical representation by introducing Graph Convolutional Network (GCN) and the formulation of the complete graph is:

$$\mathcal{G}^{t/s} = \theta(\{\hat{\mathcal{V}}, \mathcal{E}\}^{t/s}) \in \mathbb{R}^{M^{t/s} \times d}, \tag{7}$$

where $\theta(\cdot)$ denotes the GCN, and in our experiment, we compare different types of GCN and adopt the Hypergraph (Feng et al. 2019) as our solution. Since each graph node has the corresponding class as the label, we thus use the cross-entropy loss to measure whether graph nodes are classified to the correct class, which can be formulated as:

$$\mathcal{L}_{\text{cls}} = \text{CE}(\delta(\mathcal{G}^{t/s}), Y_c^{t/s}), \tag{8}$$

where $\delta(\cdot)$ is the classifier consisting of a normalization and a linear layer for both teacher and student networks.

After we complete the graphical representation construction, the optimal target of the graph $\mathcal{G}^t$ and $\mathcal{G}^s$ is to minimize their discrepancy. However, the graph is the topological structure and graph nodes from both $\mathcal{G}^t$ and $\mathcal{G}^s$ are non-matching and not equal in quantity. Hence, to measure their

discrepancy, we convert these two graphs as different distributions and introduce optimal transport as our measurement. To this end, we use the Wasserstein distance as:

$$\mathcal{L}_{\text{Align}} = \sum\nolimits_{i=0}^{M^s} \sum\nolimits_{j=0}^{M^t} \left\| \mathcal{G}_i^s - \mathcal{G}_{\pi(j)}^t \right\|_2^2, \tag{9}$$

where $\mathcal{G}_i^s$ denotes the i-th node in graph $\mathcal{G}^s$, and the $\mathcal{G}_{\pi(j)}^t$ is the overall mapping $\pi(\cdot)$ across all nodes of graph $\mathcal{G}^t$. Due to the asymmetrical and high-dimensional optimization, we thus introduce the Sinkhorn iteration (Cuturi 2013) to compute the optimal solution as the transport distance. Minimizing the distance between graphs is to find the optimized mapping $\pi(\cdot)$. We thus compute the affinity matrix to represent the mapping $\pi(\cdot)$ between two graphs. Since each node of the graphs has a label of the corresponding detected class, we expect the nodes with the same class to be more similar to the nodes with different ones. To further simplify the Eq. (9), we thus have the following formulation:

$$\mathcal{L}_{\text{Align}} = \sum\nolimits_{i=0}^{M^s} \sum\nolimits_{j=0}^{M^t} \mathbb{I}(Y_c^i = Y_c^j) \cdot \left\| \mathcal{G}_i^s - \mathcal{G}_j^t \right\|_2^2, \tag{10}$$

where Y_c^i and Y_c^j are the corresponding label between nodes $\mathcal{G}_i^s$ and $\mathcal{G}_j^t$, and $\mathbb{I}(\cdot)$ is the indicator function.

Experiments

Datasets and Evaluation

Fetal Heart (FH) (Pu et al. 2024a) is collected a heart view from two medical centers for routine screening of congenital heart diseases. All datasets were annotated by senior sonographers with extensive clinical experience, and each fetal image was labeled with a view category as well as category and location annotation information for anatomical structures, including 9 organs. We carried out experiments focused on adaptation tasks, primarily between medical center 1 and medical center 2. For instance, adapting from center 1 (source) to center 2 (target) is denoted as center 1→2.

AATS-EPV dataset is the same view and structure of Early Pregnancy View (EPV) (Lin et al. 2022) collected from our partner medical center, and also its data is from different machine devices such as SonoScape and Samsung. The gestational age range is from 11-14 weeks. The EPV is adaptive to AATS-EPV, denoted as EPV→AATS-EPV, similarly AATS-EPV→EPV. EPV (Lin et al. 2022) is a highly challenging early pregnancy ultrasound dataset collected comprising 1131 images with a gestational age range of 10-14 weeks. EPV includes a fetal view containing nine key structures. During the evaluation process, we report the mean Average Precision (mAP) in the target domain with an Intersection over Union (IoU) threshold of 0.5.

Implementation Details

We utilized VGG-16 as the feature extractor and employed Faster R-CNN as the detection head. All our code implementations are based on the Detectron2 framework. During training, we employed the Stochastic Gradient Descent (SGD) optimizer with an initial learning rate of 0.04, a batch size of 6, and a total of 10,000 iterations. We performed proportional scaling on input images, resizing the shorter edge to 600 pixels. The dataset is divided into training, validation, and test sets in a 7:1:2 ratio. To compare more Source-free UDA methods, we compare AT-SF, a Source-free version of AT (Li et al. 2022b).

Comparison with State-of-the-arts

Center 1 → 2 on FH. Table 1 reports the comparison detection results for multiple anatomical structures of all methods when adapting from center 1 to center 2. It is clear that our method significantly outperforms both the UDA and Source-free UDA detection methods. Compared to the best UDA method, i.e., AT, our method improved the mAP by 4.69%. Similarly, compared to the best Source-free UDA AT-SF, our method exceeded performance by 11.74% mAP. Meanwhile, our method achieves the best performance in detecting both large structures, such as R, and small structures, such as DAO, demonstrating its comprehensive performance.

Center 2→1 on FH. As shown in Table 1, it is observed that the proposed method continues to surpass all UDA and Source-free UDA baselines, achieving the best performance. Compared to AT-SF and A^2SFOD, our method improved the detection results with a clear margin of 10.33% and 10.59% mAP, respectively. Similarly, our approach achieved state-of-the-art results in the adaptive detection of all 9 types of structures. Further, these reported results may indicate that our method can almost eliminate the domain gap on some datasets through the graph-aware topological knowledge alignment of the teacher and student.

AATS-EPV→EPV. As listed in Table 2, our approach achieves SOTA performance among all UDA and Source-free UDA baseline methods and surpasses the second-best AT method by a margin of 3.77% mAP. Compared to the *Source Only* approach, our method achieves an improvement of 43.72%, underscoring the significant enhancements afforded by the utilization of domain adaptation algorithms. Compared to FH, the adaptive detection of structures in EPV presents a more significant challenge. This is primarily due to the incomplete development of structures within the EPV in the early pregnancy period, where certain structures are smaller, less distinct, and more ambiguous, thus increasing the difficulty of detection.

EPV→AATS-EPV. Table 2 also reports the results of all methods on EPV→AATS-EPV. Our method still outperforms all the comparison methods, and the results fully demonstrate the relevance of our designed algorithm by graph-aware morphology alignment and unsupervised structure refinement to solve the problem of Source-free domain adaptation for structure detection.

Further Empirical Analysis

Effect of USR and GMA. As shown in Table 3, compared with the baseline (none of USR and GMA) on 4 group experiments, our method has a large improvement for adaptation detection. In each adaptation experiment, our method improves over the baseline method by more than 10% mAP. In the adaptation Center 1→2 detection task, GMA boosts 9.19% mAP and achieves a detection accuracy of 75.16% mAP, which exceeds most of the methods in Table 1. On

Methods	Center 1→2										Center 2→1									
	LA	RA	LV	RV	CR	R	VS	SP	DAO	mAP (%)	LA	RA	LV	RV	CR	R	VS	SP	DAO	mAP (%)
Source Only	35.06	38.77	34.46	36.67	52.86	52.51	42.19	54.66	48.01	43.90	70.58	80.29	76.08	65.34	79.98	61.64	75.96	88.07	79.21	75.23
UDA Methods																				
AT (Li et al. 2022b)	63.47	72.99	77.28	68.14	72.31	79.00	77.28	80.91	65.84	73.02	83.56	94.43	93.13	90.50	84.94	70.69	90.45	96.16	**92.76**	88.52
SIGMA (Li, Liu, and Yuan 2022)	50.11	62.10	49.52	51.30	58.94	55.61	46.68	54.00	47.85	52.04	83.86	86.42	83.60	78.08	82.47	68.87	78.06	90.46	85.64	81.94
LRA (Piao, Tang, and Zhao 2023)	36.04	46.06	51.56	37.66	42.95	76.10	44.19	65.79	45.21	49.50	88.43	71.82	79.30	84.66	84.83	64.46	83.76	82.33	80.53	80.01
CMT (Cao et al. 2023)	62.43	71.72	75.49	70.48	71.61	80.52	78.12	81.59	59.77	72.41	91.65	**95.50**	94.54	92.98	88.26	55.69	90.66	91.46	88.84	87.73
SIGMA++ (Li, Liu, and Yuan 2023)	57.12	56.97	60.21	58.32	56.05	58.87	55.46	59.08	60.11	57.83	86.88	88.19	82.83	79.55	87.17	66.66	82.20	82.62	86.48	82.50
M^3-UDA (Pu et al. 2024b)	57.54	57.16	60.85	56.71	58.15	75.43	68.88	69.07	65.36	63.24	**94.62**	90.05	94.09	88.65	**90.73**	74.69	90.59	90.17	89.63	89.24
ToMo-UDA (Pu et al. 2024a)	64.23	75.62	70.40	64.32	66.69	75.03	75.48	77.24	**72.99**	71.33	93.91	90.25	94.72	88.20	89.92	69.86	90.66	90.91	90.06	88.71
Source-free Methods																				
AT-SF (Li et al. 2022b)	52.07	67.22	74.77	64.34	65.01	70.92	73.88	66.24	59.24	65.97	74.66	85.33	88.46	85.49	85.15	64.71	89.00	87.39	77.35	81.95
LODS (Li et al. 2022a)	29.83	41.89	48.72	32.20	33.94	49.36	39.56	57.72	31.68	40.55	85.47	85.85	87.08	83.21	76.91	60.33	78.60	86.97	66.02	78.94
SMPT++ (Liu and Yuan 2022)	50.05	60.19	66.73	56.68	51.18	77.69	56.68	70.36	62.88	61.80	74.66	85.33	88.46	85.49	85.15	**85.62**	89.00	87.39	77.35	84.27
A^2SFOD (Chu et al. 2023)	15.43	30.76	28.20	28.50	18.07	48.58	24.27	45.02	43.30	31.35	87.73	88.49	88.52	87.00	75.69	67.92	88.09	89.82	67.92	81.69
IRG (VS, Oza, and Patel 2023a)	37.73	44.25	60.43	44.89	47.34	69.41	47.42	64.77	52.53	52.09	83.86	90.19	84.50	84.46	65.71	65.49	84.81	92.48	60.64	79.13
Ours	**67.63**	**79.75**	**81.02**	**75.22**	**73.39**	**84.35**	**82.90**	**86.01**	69.08	**77.71**	90.35	94.68	**95.75**	**95.12**	**85.02**	**59.21**	**95.39**	**96.86**	92.27	**89.41**
Target Only	67.87	89.20	85.50	84.13	80.34	85.84	86.05	88.61	76.55	82.68	97.42	97.52	97.65	97.06	93.25	74.29	97.40	97.77	95.48	94.21

Table 1: Domain adaptation results on the FH. The best results are in bold, and the second-best is underlined.

Methods	AATS-EPV→EPV						EPV→AATS-EPV					
	T	N	P	RH	NT	mAP (%)	T	N	P	RH	NT	mAP (%)
Source Only	27.39	36.40	35.64	35.28	25.70	32.08	86.69	56.15	87.13	91.84	36.40	71.64
UDA Methods												
AT (Li et al. 2022b)	88.04	53.06	67.76	88.13	31.21	65.64	**97.13**	78.06	87.81	93.46	58.34	82.96
SIGMA (Li, Liu, and Yuan 2022)	72.82	53.21	70.55	72.36	27.37	59.26	89.23	72.52	89.13	86.68	54.07	78.32
LRA (Piao, Tang, and Zhao 2023)	**89.82**	66.95	79.98	77.96	41.84	71.31	82.48	74.65	67.73	88.30	54.97	73.62
CMT (Cao et al. 2023)	81.31	63.36	72.48	84.70	39.28	68.23	96.06	79.68	93.53	93.65	48.55	82.29
SIGMA++ (Li, Liu, and Yuan 2023)	60.72	53.44	78.04	73.66	35.30	60.23	89.31	79.63	86.07	86.46	51.47	78.58
M^3-UDA (Pu et al. 2024b)	76.86	54.94	83.74	66.48	40.12	64.43	86.40	79.53	89.66	84.72	59.77	80.01
ToMo-UDA (Pu et al. 2024a)	68.75	54.28	75.39	73.27	34.89	61.32	90.32	87.85	90.59	89.59	63.83	84.43
Source-free Methods												
AT-SF (Li et al. 2022b)	82.48	37.77	61.21	88.94	40.69	62.23	95.07	58.59	92.04	92.75	37.73	75.23
LODS (Li et al. 2022a)	86.96	55.92	70.84	89.55	36.51	67.96	82.43	77.39	89.11	89.95	55.49	78.87
SMPT++ (Liu and Yuan 2022)	78.44	64.17	60.76	88.21	39.94	66.31	93.80	63.44	**93.88**	93.74	57.51	80.47
A^2SFOD (Chu et al. 2023)	88.64	34.80	70.06	84.46	12.31	58.05	90.28	82.45	89.81	90.88	**71.52**	84.99
IRG (VS, Oza, and Patel 2023a)	76.82	65.20	64.70	85.14	18.37	62.05	88.38	68.13	87.47	91.89	44.79	76.13
Ours	86.49	**72.41**	**85.38**	**89.92**	**44.79**	**75.80**	94.66	**88.64**	91.48	**93.81**	61.62	**86.04**
Target Only	81.52	77.70	94.81	94.35	73.38	84.35	93.39	89.26	94.45	91.65	90.33	91.82

Table 2: Domain adaptation results on AATS-EPV→EPV and EPV→AATS-EPV.

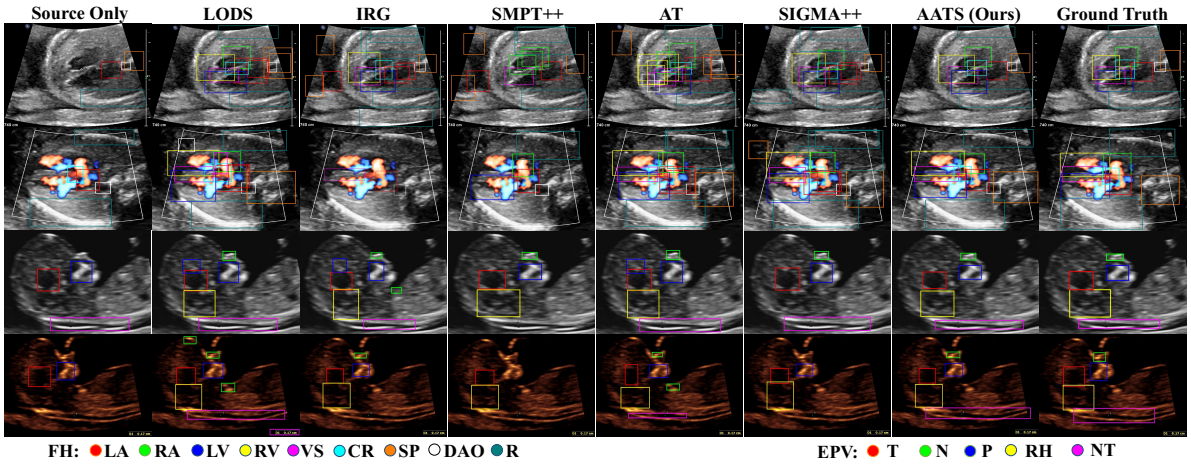

Figure 3: Qualitative result comparison on center 1→2 on FH (first row), center 2→1 on FH (second row), AATS-EPV→EPV (third row), and EPV→AATS-EPV (last row).

AATS-EPV→EPV, USR improves by 8.44%, and GMA increases by 7.31% mAP. Our method (both USR and GMA) is a significant 13.57% mAP improvement over baseline. The experimental results in Table 3 further demonstrate the effectiveness of our approach to align the student and teacher through the proposed USR and GMA, hence eliminating the domain gap without source data.

Effect of loss function on GMA. We further explored the implications of $\mathcal{L}_{cls}$ and $\mathcal{L}_{Align}$ of GMA for our model. As reported in Table 4, we observe that $\mathcal{L}_{cls}$ and $\mathcal{L}_{Align}$

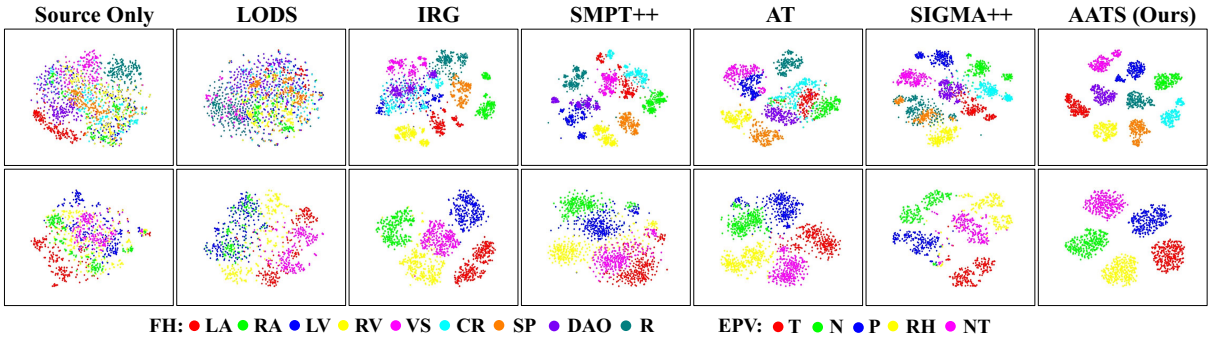

Figure 4: Feature visualization by t-SNE technique is performed on center 1→2 of FH (first row), EPV→AATS-EPV (last row).

	Center 1→2 on FH			Center 2→1 on FH		
Method	*USR*	*GMA*	mAP (%)	*USR*	*GMA*	mAP (%)
Baseline	-	-	65.97	-	-	81.95
Ours	✓	✗	73.27 (+7.30)	✓	✗	84.23 (+2.28)
	✗	✓	75.16 (+9.19)	✗	✓	88.91 (+6.96)
	✓	✓	**77.71** (+11.7)	✓	✓	**92.28** (+10.3)
	EPV→AATS-EPV			AATS-EPV→EPV		
Baseline	-	-	75.23	-	-	62.23
Ours	✓	✗	79.57 (+4.34)	✓	✗	70.67 (+8.44)
	✗	✓	82.90 (+7.67)	✗	✓	69.54 (+7.31)
	✓	✓	**85.37** (+10.1)	✓	✓	**75.80** (+13.6)

Table 3: Ablation results of USR and GMA.

Center 1→2 on FH			Center 2→1 on FH		
$\mathcal{L}_{cls}$	$\mathcal{L}_{Align}$	mAP (%)	$\mathcal{L}_{cls}$	$\mathcal{L}_{Align}$	mAP (%)
✗	✗	65.97	✗	✗	81.95
✓	✗	73.59 (+7.62)	✓	✗	83.92 (+1.97)
✗	✓	74.24 (+8.27)	✗	✓	87.89 (+5.94)
✓	✓	**75.16** (+9.19)	✓	✓	**88.91** (+6.96)
EPV→AATS-EPV			AATS-EPV→EPV		
✗	✗	75.23	✗	✗	62.23
✓	✗	79.01 (+3.78)	✓	✗	63.34 (+1.11)
✗	✓	77.73 (+2.50)	✗	✓	66.90 (+4.67)
✓	✓	**82.90** (+7.67)	✓	✓	**69.54** (+7.31)

Table 4: Ablation results of loss function on GMA.

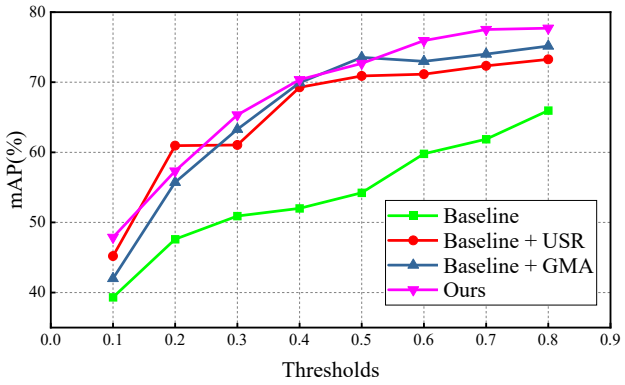

Figure 5: Detection results for different thresholds on center 1→2 of FH dataset.

improve the performance of detection on all experimental datasets. For example, on EPV→AATS-EPV, $\mathcal{L}_{cls}$ enhances 3.78% mAP, $\mathcal{L}_{Align}$ improves 2.5% mAP, and adding both $\mathcal{L}_{cls}$ and $\mathcal{L}_{Align}$ boost 7.67% mAP. In GMA, each graph node represents an organ structure, and $\mathcal{L}_{cls}$ effectively classifies organs in the student and teacher branches to improve detection accuracy. Then $\mathcal{L}_{Align}$ matches the topological knowledge in teacher and student by graph alignment. The all superior results reported in Table 4 demonstrate the effectiveness of proposed $\mathcal{L}_{cls}$ and $\mathcal{L}_{Align}$.

Parameter Analysis. We further discuss the sensitivity influence of the threshold of our method. The threshold determines the number of proposals that meet the requirements in the ROI head output for both the teacher and student. The quality and quantity of proposals will affect the pooling operations, which in turn will affect the performance of the USR and GMA modules. As shown in Fig. 5, the best detection performance is obtained with a threshold of 0.8. It is worth noting that when the threshold is set at 0.9, the confidence levels of the proposals generated by the student network often fail to meet the requirements, resulting in the entire network being unable to train.

Visualization Analysis. As shown in Fig. 3, It can be observed intuitively that our method is consistent with the ground truth, which further demonstrates the effectiveness of our AATS. However, other comparison methods are subject to miss and false detections. Fig. 4 demonstrates a comparison of the classification feature visualizations, which can directly show that our method recognizes better.

Conclusion

Our paper proposed a novel source-free unsupervised domain adaptation object detection method named AATS in medical image analysis. AATS, as the first Source-free method for medical multi-object detection, utilizes prior knowledge from ultrasound images that focuses on exploring the morphological information of each organ and the topological relationships among different organs. We use the graphical representation to formulate the above knowledge and conduct the experiment across four groups of adaptation experiments. Extensive experimental results show that our method outperforms the previous baseline methods and achieves state-of-the-art results.

Acknowledgements

This work is supported by two grants from the National Natural Science Foundation of China (Grant Nos. 62306254 and 62306003), a grant from the Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone (Grant No. HZQB-KCZYB-2020083), and the Open Research Fund from the Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) under Grant No. GML-KF-24-29.

References

Baumgartner, C. F.; Kamnitsas, K.; Matthew, J.; Fletcher, T. P.; Smith, S.; Koch, L. M.; Kainz, B.; and Rueckert, D. 2017. SonoNet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. *IEEE Trans. Med. Imaging*, 36(11): 2204–2215.

Cao, S.; Joshi, D.; Gui, L.-Y.; and Wang, Y.-X. 2023. Contrastive Mean Teacher for Domain Adaptive Object Detectors. In *CVPR*, 23839–23848.

Chen, C.; Zheng, Z.; Ding, X.; Huang, Y.; and Dou, Q. 2020. Harmonizing transferability and discriminability for adapting object detectors. In *CVPR*, 8869–8878.

Chen, H.; Wu, L.; Dou, Q.; Qin, J.; Li, S.; Cheng, J.-Z.; Ni, D.; and Heng, P.-A. 2017. Ultrasound standard plane detection using a composite neural network framework. *IEEE Transactions on Cybernetics*, 47(6): 1576–1586.

Chen, Y.; Li, W.; Sakaridis, C.; Dai, D.; and Van Gool, L. 2018. Domain adaptive faster r-cnn for object detection in the wild. In *CVPR*, 3339–3348.

Chen, Y.; Wang, H.; Li, W.; Sakaridis, C.; Dai, D.; and Van Gool, L. 2021. Scale-aware domain adaptive faster r-cnn. *Int. J. Comput. Vis.*, 129(7): 2223–2243.

Chen, Z.; Wang, Z.; and Zhang, Y. 2023. Exploiting Low-confidence Pseudo-labels for Source-free Object Detection. In *ACM MM*, 5370–5379.

Chu, Q.; Li, S.; Chen, G.; Li, K.; and Li, X. 2023. Adversarial alignment for source free object detection. In *AAAI*, volume 37, 452–460.

Cuturi, M. 2013. Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances. arXiv:1306.0895.

Dai, W.; Li, X.; Ding, X.; and Cheng, K.-T. 2022. Cyclical self-supervision for semi-supervised ejection fraction prediction from echocardiogram videos. *IEEE Transactions on Medical Imaging*, 42(5): 1446–1461.

Deng, J.; Li, W.; Chen, Y.; and Duan, L. 2021. Unbiased mean teacher for cross-domain object detection. In *CVPR*, 4091–4101.

Deng, J.; Xu, D.; Li, W.; and Duan, L. 2023. Harmonious Teacher for Cross-Domain Object Detection. In *CVPR*, 23829–23838.

Dinsdale, N. K.; Jenkinson, M.; and Namburete, A. I. 2023. Sfharmony: Source free domain adaptation for distributed neuroimaging analysis. In *ICCV*, 11494–11505.

Feng, Y.; You, H.; Zhang, Z.; Ji, R.; and Gao, Y. 2019. Hypergraph neural networks. In *AAAI*, volume 33, 3558–3565.

He, J.; Yang, L.; Liang, B.; Li, S.; and Xu, C. 2024. Fetal cardiac ultrasound standard section detection model based on multitask learning and mixed attention mechanism. *Neurocomputing*, 127443.

Hegde, D.; and Patel, V. M. 2024. Attentive prototypes for source-free unsupervised domain adaptive 3d object detection. In *WACV*, 3066–3076.

Huai, Z.; Ding, X.; Li, Y.; and Li, X. 2023. Context-Aware Pseudo-label Refinement for Source-Free Domain Adaptive Fundus Image Segmentation. In *MICCAI*, 618–628. Springer.

Kim, T.; Jeong, M.; Kim, S.; Choi, S.; and Kim, C. 2019. Diversify and match: A domain adaptive representation learning paradigm for object detection. In *CVPR*, 12456–12465.

Kundu, J. N.; Kulkarni, A. R.; Bhambri, S.; Mehta, D.; Kulkarni, S. A.; Jampani, V.; and Radhakrishnan, V. B. 2022. Balancing discriminability and transferability for source-free domain adaptation. In *ICML*, 11710–11728. PMLR.

Li, J.; Yu, Z.; Du, Z.; Zhu, L.; and Shen, H. T. 2024. A comprehensive survey on source-free domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.*

Li, S.; Ye, M.; Zhu, X.; Zhou, L.; and Xiong, L. 2022a. Source-free object detection by learning to overlook domain style. In *CVPR*, 8014–8023.

Li, W.; Liu, X.; and Yuan, Y. 2022. Sigma: Semantic-complete graph matching for domain adaptive object detection. In *CVPR*, 5291–5300.

Li, W.; Liu, X.; and Yuan, Y. 2023. SIGMA++: Improved Semantic-complete Graph Matching for Domain Adaptive Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.*

Li, X.; Chen, W.; Xie, D.; Yang, S.; Yuan, P.; Pu, S.; and Zhuang, Y. 2021. A free lunch for unsupervised domain adaptive object detection without source data. In *AAAI*, volume 35, 8474–8481.

Li, Y.-J.; Dai, X.; Ma, C.-Y.; Liu, Y.-C.; Chen, K.; Wu, B.; He, Z.; Kitani, K.; and Vajda, P. 2022b. Cross-domain adaptive teacher for object detection. In *CVPR*, 7581–7590.

Li, Z.; Zhao, L.; Chen, W.; Yang, S.; Xie, D.; and Pu, S. 2022c. Target-aware auto-augmentation for unsupervised domain adaptive object detection. In *ICASSP*, 3848–3852. IEEE.

Lin, Q.; Zhou, Y.; Shi, S.; Zhang, Y.; Yin, S.; Liu, X.; Peng, Q.; Huang, S.; Jiang, Y.; Cui, C.; et al. 2022. How much can AI see in early pregnancy: A multi-center study of fetus head characterization in week 10–14 in ultrasound using deep learning. *Computer Methods and Programs in Biomedicine*, 226: 107170.

Lin, Z.; Li, S.; Ni, D.; Liao, Y.; Wen, H.; Du, J.; Chen, S.; Wang, T.; and Lei, B. 2019. Multi-task learning for quality assessment of fetal head ultrasound images. *Medical Image Analysis*, 58: 101548.

Liu, Q.; Lin, L.; Shen, Z.; and Yang, Z. 2023. Periodically exchange teacher-student for source-free object detection. In *ICCV*, 6414–6424.

Liu, X.; Li, W.; and Yuan, Y. 2023. Decoupled Unbiased Teacher for Source-Free Domain Adaptive Medical Object Detection. *IEEE Transactions on Neural Networks and Learning Systems*.

Liu, X.; and Yuan, Y. 2022. A source-free domain adaptive polyp detection framework with style diversification flow. *IEEE Trans. Med. Imaging*, 41(7): 1897–1908.

Nguyen, E. H.; Yang, H.; Deng, R.; Lu, Y.; Zhu, Z.; Roland, J. T.; Lu, L.; Landman, B. A.; Fogo, A. B.; and Huo, Y. 2021. Circle representation for medical object detection. *IEEE Trans. Med. Imaging*, 41(3): 746–754.

Oza, P.; Sindagi, V. A.; Sharmini, V. V.; and Patel, V. M. 2023. Unsupervised domain adaptation of object detectors: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*

Piao, Z.; Tang, L.; and Zhao, B. 2023. Unsupervised domain-adaptive object detection via localization regression alignment. *IEEE Transactions on Neural Networks and Learning Systems*.

Pu, B.; Li, K.; Li, S.; and Zhu, N. 2021. Automatic fetal ultrasound standard plane recognition based on deep learning and IIoT. *IEEE Transactions on Industrial Informatics*, 17(11): 7771–7780.

Pu, B.; Lv, X.; Yang, J.; Guannan, H.; Dong, X.; Lin, Y.; Shengli, L.; Ying, T.; Fei, L.; Chen, M.; Jin, Z.; Li, K.; and Li, X. 2024a. Unsupervised Domain Adaptation for Anatomical Structure Detection in Ultrasound Images. In *ICML*, volume 235, 41204–41220. PMLR.

Pu, B.; Wang, L.; Yang, J.; He, G.; Dong, X.; Li, S.; Tan, Y.; Chen, M.; Jin, Z.; Li, K.; et al. 2024b. M3-UDA: A New Benchmark for Unsupervised Domain Adaptive Fetal Cardiac Structure Detection. In *CVPR*, 11621–11630.

Qiu, S. 2023. Causality-Inspired Source-Free Domain Adaptation for Medical Image Classification. In *ICIG*, 68–80. Springer.

Qiu, Z.; Zhang, Y.; Lin, H.; Niu, S.; Liu, Y.; Du, Q.; and Tan, M. 2021. Source-free domain adaptation via avatar prototype generation and adaptation. *arXiv preprint arXiv:2106.15326*.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *NeurIPS*, 28.

Shen, M.; Lu, Y.; Hu, Y.; and Ma, A. J. 2023. Collaborative Learning of Diverse Experts for Source-free Universal Domain Adaptation. In *ACM MM*, 2054–2065.

Song, Y.; Fan, J.; Liu, D.; and Cai, W. 2024. Multi-source-free Domain Adaptation via Uncertainty-aware Adaptive Distillation. *arXiv preprint arXiv:2402.06213*.

Tarvainen, A.; and Valpola, H. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *NeurIPS*, 30.

Van Der Maaten, L. 2009. Learning a parametric embedding by preserving local structure. In *Artificial Intelligence and Statistics*, 384–391. PMLR.

Vs, V.; Gupta, V.; Oza, P.; Sindagi, V. A.; and Patel, V. M. 2021. Mega-cda: Memory guided attention for category-aware unsupervised domain adaptive object detection. In *CVPR*, 4516–4526.

VS, V.; Oza, P.; and Patel, V. M. 2023a. Instance relation graph guided source-free domain adaptive object detection. In *CVPR*, 3520–3530.

VS, V.; Oza, P.; and Patel, V. M. 2023b. Towards online domain adaptive object detection. In *WACV*, 478–488.

Wang, Y.; Cheng, J.; Chen, Y.; Shao, S.; Zhu, L.; Wu, Z.; Liu, T.; and Zhu, H. 2023. Fvp: Fourier visual prompting for source-free unsupervised domain adaptation of medical image segmentation. *IEEE Trans. Med. Imaging*.

Wang, Y.; Zhang, R.; Zhang, S.; Li, M.; Xia, Y.; Zhang, X.; and Liu, S. 2021. Domain-specific suppression for adaptive object detection. In *CVPR*, 9603–9612.

Wu, J.; Wang, G.; Gu, R.; Lu, T.; Chen, Y.; Zhu, W.; Vercauteren, T.; Ourselin, S.; and Zhang, S. 2023. UPL-SFDA: Uncertainty-aware Pseudo Label Guided Source-Free Domain Adaptation for Medical Image Segmentation. *IEEE Trans. Med. Imaging*.

Yang, C.; Liu, Y.; and Yuan, Y. 2023. Transferability-Guided Multi-source Model Adaptation for Medical Image Segmentation. In *MICCAI*, 703–712. Springer.

Yang, J.; Ding, X.; Zheng, Z.; Xu, X.; and Li, X. 2023. GraphEcho: Graph-Driven Unsupervised Domain Adaptation for Echocardiogram Video Segmentation. In *ICCV*, 11878–11887.

Yang, J.; Lin, Y.; Pu, B.; Guo, J.; Xu, X.; and Li, X. 2024. Cardiacnet: Learning to reconstruct abnormalities for cardiac disease assessment from echocardiogram videos. In *ECCV*, 293–311. Springer.

Yang, S.; Jui, S.; van de Weijer, J.; et al. 2022. Attracting and dispersing: A simple approach for source-free domain adaptation. *NeurIPS*, 35: 5802–5815.

Yu, F.; Wang, D.; Chen; et al. 2022. Sc-uda: Style and content gaps aware unsupervised domain adaptation for object detection. In *WACV*, 382–391.

Yu, Q.; Xi, N.; Yuan, J.; Zhou, Z.; Dang, K.; and Ding, X. 2023. Source-Free Domain Adaptation for Medical Image Segmentation via Prototype-Anchored Feature Alignment and Contrastive Learning. In *MICCAI*, 3–12. Springer.

Yuan, P.; Chen, W.; Yang, S.; Xuan, Y.; Xie, D.; Zhuang, Y.; and Pu, S. 2022. Simulation-and-mining: Towards accurate source-free unsupervised domain adaptive object detection. In *ICASSP*, 3843–3847. IEEE.

Zhao, G.; Li, G.; Xu, R.; and Lin, L. 2020. Collaborative training between region proposal localization and classification for domain adaptive object detection. In *ECCV*, 86–102. Springer.

Zhao, L.; and Wang, L. 2022. Task-specific inconsistency alignment for domain adaptive object detection. In *CVPR*, 14217–14226.

Zheng, Z.; Yang, J.; Ding, X.; Xu, X.; and Li, X. 2023. GL-Fusion: Global-Local Fusion Network for Multi-view Echocardiogram Video Segmentation. In *MICCAI*, 78–88. Springer.