ADVERSARIAL EXAMPLES GUIDED PSEUDO-LABEL REFINEMENT FOR DECENTRALIZED DOMAIN ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Unsupervised domain adaptation (UDA) methods usually assume data from multiple domains can be put together for centralized adaptation. Unfortunately, this assumption impairs data privacy, which leads to the failure of traditional methods in practical scenarios. To cope with the above issue, we present a new approach named Adversarial Examples Guided Pseudo-label Refinement for Decentralized Domain Adaptation (AGREE), which conducts target adaptation in an iterative training process during which only models can be delivered across domains. More specifically, to train a promising target model, we leverage Adversarial Examples (AEs) to filter out error prone predictions of source models towards each target sample based on both robustness and confidence, and then treat the most frequent prediction as the pseudo-label. Besides, to improve central model aggregation, we introduce Knowledge Contribution (KC) to compute reasonable aggregation weights. Extensive experiments conducted on several standard datasets verify the superiority of the proposed AGREE. Especially, our AGREE achieves the new state-of-the-art performance on the DomainNet and Office-Caltech10 datasets. The implementation code will be publicly available.

1 INTRODUCTION

Deep Learning has drawn surging attention over the past decade. To solve the problem that deep models usually suffer from significant performance degradation when applied to an unseen target domain due to domain shift, unsupervised domain adaptation (UDA) (Long et al., 2015; Tzeng et al., 2017; Zhang et al., 2017) has been introduced to transfer knowledge from a fully labeled source domain to an unlabeled target domain. UDA has enabled several advances in various applications, such as image classification (Chen et al., 2020), object detection (Zhang et al., 2021b), semantic segmentation (Zhang et al., 2021a), and so on. A common strategy in domain adaptation is to minimize the distribution discrepancy across domains by matching the statistical moments of distributions (Chen et al., 2020). Another popular paradigm employs adversarial training to lead the learned source and target features to be indistinguishable from each other (Ganin et al., 2016; Dong et al., 2020a). However, most current UDA methods assume that source data are merely drawn from a single domain, which neglects the more practical scenarios where labeled samples are typically collected from multiple domains, $e.g.$, different weather or lighting conditions, different visual cues, different modalities, etc. Therefore, Multi-Source Domain Adaptation (MSDA) (Li et al., 2018; Zhao et al., 2019; Lin et al., 2020) is proposed to transfer knowledge from multiple distinct domains to one unlabeled target domain. Specifically, MSDA explores complementarily transferable knowledge from the multi-source domains for target prediction (Zhao et al., 2018; Peng et al., 2019; Bai et al., 2021). For instance, Peng et al. (2019) matches the features across domains and then quantifies the contributions of source domains.

Unfortunately, many MSDA methods (Li et al., 2021; Russo et al., 2019) work under the strict condition that source data are always available when adapting source domains to the target domain. However, such a condition could make them unpractical in real-world applications where source data from different domains cannot be shared for joint training, due to data privacy policies and storage or transmission concerns. To this end, in this paper, we study the recently introduced problem of decentralized UDA (Peng et al., 2020), which aims to perform decentralized domain adaptation with

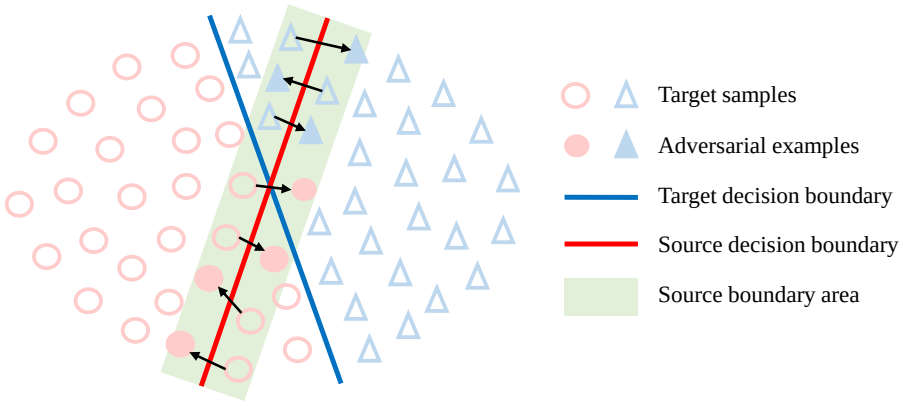


Figure 1: Identification of target samples in the source boundary area. For image classification, the predicted label of the adversarial example is different from that of the original sample when they are fed into the same network, which means that the original sample is located in the decision boundary area of the model. If the adversarial attack is easy to succeed, the prediction is not robust.

non-shared data from multiple domains. Up to now, some works (Feng et al., 2021; Wu & Gong, 2021) have been researched for decentralized UDA, resorting to federated learning and pseudo-labeling. Nevertheless, they do not take robustness into account when generating pseudo-labels. Moreover, the importance of target model weight has been overlooked during the global model aggregation to some extent in KD3A (Feng et al., 2021).

To effectively tackle the decentralized UDA problem, we propose an alternative solution, Adversarial Examples Guided Pseudo-label Refinement for Decentralized Domain Adaptation (AGREE), which conducts target adaptation in an iterative training process during which only models can be delivered across domains. Specifically, we leverage adversarial examples to filter out error prone predictions of source models towards each target sample and then treat the most frequent prediction as the pseudo-label in the remaining predictions for the optimization of target model. Adversarial examples refer to the images which are added with small noise that cannot be perceived by human but make the outputs of neural network change explicitly (Szegedy et al., 2013), and are commonly used in optimizing the adversarial robustness of neural networks (Goodfellow et al., 2014) despite being originally utilized to attack neural networks. As we all know, boundary samples are more easily misclassified (Han et al., 2005), which will be aggravated by domain shift. Therefore, we exploit adversarial examples to identify boundary samples (see Figure 1) and then eliminate the predictions of boundary samples. Instead of adopting the same attack intensity for all source models, we impose stronger attacks on models with low confidence to take both robustness and confidence into consideration. Furthermore, we introduce a knowledge contribution mechanism to facilitate global model aggregation. Especially, the target model weight is determined based on degree of domain bias between source and target domains, which is reflected by the quantity of pseudo-labels. The main contributions of this work are summarized as follows:

- We propose a novel approach called AGREE to tackle the decentralized UDA problem in a privacy-preserving way, in which all the data and computations on source domains are kept decentralized during the whole adaptation process.
- We present an adversarial examples guided refinement strategy for producing confident pseudo-labels on top of robustness and confidence to train a promising target model.
- We introduce knowledge contribution to generate reasonable target and source model weights for promoting global model aggregation.
- We conduct extensive experiments on three benchmarks verifying the efficacy of our approach. In particular, our AGREE achieves the new state-of-the-art performance on the DomainNet and Office-Caltech10 datasets.

2 RELATED WORK

Unsupervised Domain Adaptation. Unsupervised domain adaptation (UDA) aims to bridge the domain gap between labeled source samples and unlabeled target samples. There are two prevalent strategies: adversarial learning-based methods (Ganin et al., 2016; Dong et al., 2020b; Zhang et al., 2021c; Chi et al., 2021) have been proposed to perform adaptation in feature space or pixel space, and moment matching-based methods (Chen et al., 2020; Fang et al., 2020; Liu et al., 2020) are proposed to reduce the distribution discrepancy by matching the statistical moments of different orders. In addition, the methods based on reconstruction (Bousmalis et al., 2016; Ghifary et al., 2016), classifier discrepancy (Saito et al., 2018; Lee et al., 2019) and batch normalization (Wang et al., 2019) are also designed to address the domain discrepancy. If there are multiple source domains, the main idea is to incorporate and transfer the knowledge learned from different source domains to the target domain (Peng et al., 2019; Hoffman et al., 2018; Li et al., 2018; Lin et al., 2020). Upon this, all these methods follow a centralized training paradigm with sharing source data across domains, which poses a threat to data privacy.

Federated Learning. Federated learning (FL) (Konečnỳ et al., 2016; McMahan et al., 2017; Wang et al., 2020) is a distributed machine learning paradigm for optimizing a global model across multiple decentralized datasets without sharing local data. Federated average (FedAvg) (McMahan et al., 2017) is the most representative federated learning method, which iteratively aggregates the updates of models from different clients to build a global model that is hereafter distributed to selected clients for the next communication round if requisite. Under the limited communication costs, FedAvg enables decentralized training in a privacy-preserving way. In sharp contrast with different data partitions from the same dataset in FedAvg, source data in our AGREE are separately collected from different domains with various domain shifts.

Federated Domain Adaptation. In recent years, there are some works (Peng et al., 2020; Feng et al., 2021; Wu & Gong, 2021) studying decentralized UDA problem. FADA (Peng et al., 2020) introduces feature disentanglement to resolve domain shift without accessing data. KD3A (Feng et al., 2021) is a knowledge distillation based decentralized UDA method with pronounced communication efficiency. COPA (Wu & Gong, 2021) aims at optimizing a generalized target model for decentralized UDA via collaborative optimization and aggregation. Besides, there are some one-shot federated domain adaptation approaches (Ahmed et al., 2021; Dong et al., 2021; Liang et al., 2022), which can be broadly categorized into reconstruction-based (Li et al., 2020; Kurmi et al., 2021) and self-training (Huang et al., 2021; Liang et al., 2020). As self-explanatory by the name, one-shot federated domain adaptation allows only one communication round, which is also known as multi-source-free domain adaptation (Ahmed et al., 2021).

3 METHODOLOGY

Problem Setting. In decentralized UDA problem, we are given n source domains $\{D_S^i\}_{i=1}^n$ where each source domain contains N_i labeled samples for K classes as $D_S^i = \{(X_j^i, y_j^i)\}_{j=1}^{N_i}$ with $y_j^i \in \{1, ..., K\}$ and a target domain D_T with N_T unlabeled samples as $D_T = \{X_j^T\}_{j=1}^{N_T}$. The goal is to learn a promising model on the target domain without sharing data from each domain.

Overview. As shown in Figure 2, we illustrate an overview of the proposed approach. Specifically, our base model is composed of a feature extractor F and a classifier C. We train each source model $\{F_S^i, C_S^i\}$ separately on each source domain D_S^i for several local epochs. Then source models are uploaded to target domain for adaptation and global aggregation. To distill knowledge of source domains, we employ source models to produce pseudo-labels refined by Adversarial Examples techniques for target samples. To facilitate model aggregation, we obtain source weights via their respective contributions in the process of producing pseudo-labels. For target weight, we acquire it through degree of domain shift between source and target domains measured by the quantity of pseudo-labels. After training target model $\{F_T, C_T\}$ on the target domain D_T with pseudo-labels for several epochs, the feature extractors $\{F_S^i\}_{i=1}^n$ and F_T are aggregated to construct a global feature extractor F_G, which is used to update $\{F_S^i\}_{i=1}^n$ and F_T for the next round of training.

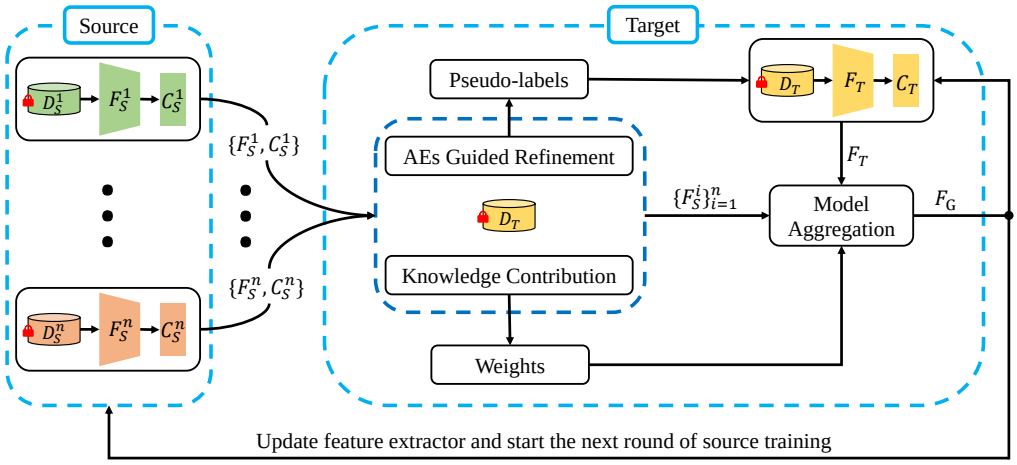


Figure 2: An overview of the proposed approach to decentralized unsupervised domain adaptation (UDA) problem. Note that, due to data privacy and storage or transmission limitations, data from different domains are non-shared and only models can be delivered. AEs: Adversarial Examples.

3.1 Model Optimization Separately on each Source Domain

We optimize each source model $\{(F_S^i, C_S^i)\}$ on the corresponding source domain D_S^i with the cross-entropy loss as follows:

$$\mathcal{L}_{ce} = -\mathbb{E}_{X_j^i \in D_S^i}[\bar{y}_j^i \log \sigma(C_S^i(F_S^i(X_j^i)))], \tag{1}$$

where $\bar{y}_j^i$ is a one-hot vector of y_j^i and $\sigma(\cdot)$ is the softmax operation. After learning each source model separately for R local epochs, we send them to the target domain for adaptation and central aggregation. In decentralized UDA, the setting of local epochs R is paramount. We must aggregate model at the appropriate frequency since the different source models have different convergence rates. Similar to Feng et al. (2021), we set $R = 1$ by default.

3.2 Target Adaptation and Global Aggregation

Adversarial Examples Guided Refinement. After receiving source models, we employ them to respectively produce prediction p_j^i for each sample X_j^T in D_T as:

$$p_j^i = \sigma(C_S^i(F_S^i(X_j^T))). \tag{2}$$

Next we use adversarial examples to filter out some potential noisy predictions. Specifically, based on prediction p_j^i, we generate the corresponding adversarial example via the projected gradient descent (PGD) (Madry et al., 2017) as follows:

$$x^{t+1} = \Pi_{\mathcal{B}[x^0]}\left(x^t + \eta \,\mathrm{sgn}(\nabla_{x^t}\mathcal{L}(\{\theta_F^i, \theta_C^i\}, x^t, \hat{y}_j^i))\right), \tag{3}$$

where x^0 denotes the original example X_j^T, $\hat{y}_j^i$ is the most confident class of p_j^i, $\{\theta_F^i, \theta_C^i\}$ denotes the model parameters of $\{F_S^i, C_S^i\}$ and $\Pi_{\mathcal{B}[x^0]}(\cdot)$ is the projection function. Though PGD is a multi-step method, it is more than enough that we only attack one step in all experiments. Instead of adopting the same attack intensity for all source models, we impose stronger attacks on models with low confidence. Therefore, we set η as follows:

$$\eta = \frac{\epsilon \cdot (1 - \beta_j^i)}{10}, \tag{4}$$

where ϵ gradually increases from low (e.g., $\frac{6}{255}$) to high (e.g., $\frac{10}{255}$) in the training process and β_j^i refers to the confidence score of the i-th source model towards the target sample X_j^T. When the prediction of $\{F_S^i, C_S^i\}$ towards the adversarial example of X_j^T conflicts with p_j^i, we abandon

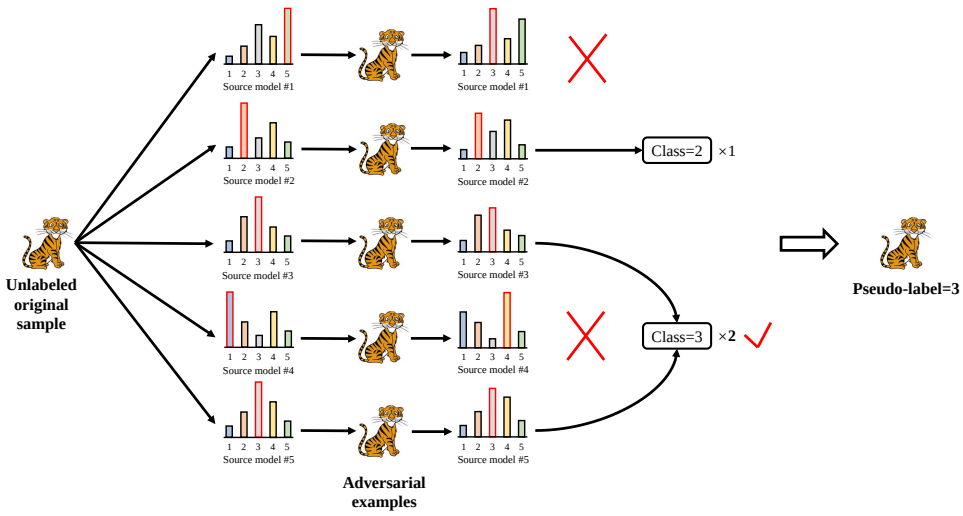


Figure 3: The graphical illustration of Adversarial Examples Guided Refinement. Adversarial examples make the predictions of some source models diverge. We filter out error prone predictions towards a target sample and then treat the most frequent prediction as the pseudo-label.

the predictions of $\{F_S^i, C_S^i\}$ for X_j^T and its adversarial example. In other words, predictions with low confidence in 'width' minima are likely to be retained and predictions with high confidence in 'sharp' minima are likely to be discarded. In the remaining predictions, we take the prediction with the highest frequency as a pseudo-label of X_j^T. We depict the workflow of adversarial examples guided pseudo-label refinement in Figure 3. If there is more than one prediction with the highest frequency or no prediction left, we do not generate pseudo-label for the target sample.

Knowledge Contribution. During the adversarial examples guided refinement, we record the contribution of each source model with Q_i which is zero initially: If a prediction of the $i-th$ source model overcomes adversarial attack and is consistent with the pseudo label, then we add 1 to Q_i. Meanwhile, we denote the number of pseudo-labels as Z. Since the process of generating pseudo-labels is the same, we utilize the number of pseudo-labels to measure degree of domain shift between source and target domains. Specifically, less pseudo-labels means larger domain bias. When encountering large domain bias, we assign a relatively large weight to the target model and relatively small weights to the source models. First, we compute the target weight α_T as

$$\alpha_T = \gamma \cdot (1 - \frac{Z}{N_T}), \tag{5}$$

where γ is set to 1 by default. In practice, we only obtain Z in the first communication round and gradually increase α_T from $\frac{1}{N}$, where $N = n + 1$, to $\gamma \cdot (1 - \frac{Z}{N_T})$ in the training process. Then we compute the source weights as

$$\alpha_S^i = (1 - \alpha_T) \cdot \frac{e^{\widetilde{Q}_i}}{\sum_{j=1}^{n} e^{\widetilde{Q}_j}}, \tag{6}$$

where $\widetilde{Q}_i = \frac{Q_i}{\sum_{j=1}^{n} Q_j}$. After pseudo-label refinement, we optimize the target model with pseudo-labels and perform model aggregation as follows:

$$\theta_F^G \leftarrow \alpha_T \cdot \theta_F + \sum_{i=1}^{n} \alpha_S^i \cdot \theta_F^i, \tag{7}$$

where θ_F denotes the parameters of F_T. Ultimately, $\{F_S^i\}_{i=1}^n$ and F_T are updated via θ_F^G for the next round of training.

To summarize, our approach works out decentralize UDA iteratively without collecting data from different domains together for centralized training. The complete process of the proposed method is shown in Algorithm 1.

Algorithm 1 The Training Process of Our Method

Input: Labeled source data $\{\mathcal{D}_S^i\}_{i=1}^n$; Unlabeled target data $\mathcal{D}_T$; Target model $\{F_T, C_T\}$
for $i = 1$ to n **do**
    $\{F_S^i, C_S^i\} \leftarrow \{F_T, C_T\}$
end for
for $e = 1$ to E **do**
    for $i = 1$ to n **do**
        Train $\{F_S^i, C_S^i\}$ on $\mathcal{D}_S^i$
    end for
    Upload $\{F_S^i, C_S^i\}$ to the target domain
    Obtain pseudo-labels for $\mathcal{D}_T$
    Compute aggregation weights α_T and $\{\alpha_S^i\}_{i=1}^n$
    Train $\{F_T, C_T\}$ on $\mathcal{D}_T$ with pseudo-labels
    Construct F_G via aggregating $\{F_S^i\}_{i=1}^n$ and F_T
    for $i = 1$ to n **do**
        $F_S^i \leftarrow F_G$
    end for
    $F_T \leftarrow F_G$
end for
Return F_G and C_T

4 Experimental Results

4.1 Experimental Setup

Datasets. We perform experiments on three datasets including Digit-Five (Peng et al., 2019), Office-Caltech10 (Gong et al., 2012), DomainNet (Peng et al., 2019). **Digit-Five** is a digit classification dataset containing five domains: MNIST (mt), MNIST-M (mm), SVHN (sv), SYN (sy), USPS (up). **Office-Caltech10** consists of ten object categories from four domains, namely Amazon (A), Caltech (C), DSLR (D) and Webcam (W), with 2,533 images in total. **DomainNet** is a relatively large-scale dataset which contains 345 classes and six domains, i.e. Clipart (Clp), Infograph (Inf), Painting (Pnt), Quickdraw (Qdr), Real (Rel) and Sketch (Skt). Following previous methods (Peng et al., 2019; 2020), we fix one of the domains as the target domain with unlabeled training data and the rest as the source domains with labeled training data.

Implementation Details. Following (Peng et al., 2019), we utilize a 3-layer CNN as backbone for Digit-Five while ResNet-101 for Office-Caltech10 and DomainNet. For Digit-Five and Office-Caltech10, We set batch size to 64 and global epoch $E = 50$. For DomainNet, we set batch size to 48 and global epoch $E = 80$. To make each source domain comparable in sample size, we randomly choose 30000 images from each domain for training on DomainNet in each epoch. We apply SGD with momentum as the optimizer and decay learning rate from high (i.e., 0.05 for Digit-Five and 0.002 for Office-Caltech10 and DomainNet) to zero with a cosine annealing rule. Following (Wu & Gong, 2021), We report top-1 accuracy averaged over five runs for each experiment.

4.2 Comparison with State-of-the-Art Methods

Baseline Methods. To evaluate the efficiency of the proposed method, we conduct extensive comparison experiments with the state-of-the-art centralized or decentralized approaches. Concretely, centralized approaches include DAN (Long et al., 2015), MDAN (Zhao et al., 2018), MCD (Saito et al., 2018), M^{3}SDA (Peng et al., 2019), DANN (Ganin & Lempitsky, 2015), DCTN (Xu et al., 2018), MoE (Guo et al., 2018), DSBN (Chang et al., 2019), CMSS (Yang et al., 2020), CMSDA (Scalbert et al., 2021), MOST (Nguyen et al., 2021a) and STEM (Nguyen et al., 2021b), and decentralized approaches comprise FADA (Peng et al., 2020), SHOT (Liang et al., 2020), KD3A (Feng et al., 2021) and COPA (Wu & Gong, 2021). Furthermore, two baselines without domain adaptation are reported, namely oracle that directly performs supervised learning on target domains and source-only that naively combines source domains to train a single model.

Table 1: Comparison with state-of-the-art UDA methods on Digit-Five (the left part) using a convolutional backbon and on Office-Caltech10 (the right part) using ResNet-101. Our AGREE achieves competitive performance with SOTA methods in terms of average overall accuracy on Digit-Five and new state-of-the-art results on Office-Caltech10. Paradigm, Decentralized and Source-only are abbreviated as Pa., Decentra. and Src-only due to space limitations.

Pa.	Method	Digit-Five						Office-Caltech10				
		mt	mm	sv	sy	up	Avg	A	C	D	W	Avg
-	Oracle	99.5	95.4	92.3	98.7	99.2	97.0	99.7	98.4	99.8	99.7	99.4
	Src-only	92.3	63.7	71.5	83.4	90.7	80.3	88.7	85.4	98.2	99.1	92.9
Centralized	DAN	96.3	63.8	72.5	85.4	94.2	82.4	91.6	89.2	99.1	99.5	94.8
	MDAN	98.0	69.5	69.2	87.4	92.5	83.3	95.4	91.8	98.6	98.9	96.1
	MCD	99.2	80.7	81.9	95.4	98.3	91.1	92.1	91.5	99.1	99.5	95.6
	M^3SDA	98.4	72.8	81.3	89.6	96.2	87.7	94.5	92.2	99.2	99.5	96.4
	DANN	97.6	71.3	63.5	85.4	92.3	82.1	92.6	91.3	99.1	99.4	95.6
	DCTN	96.2	70.5	77.6	86.8	92.8	84.8	93.2	91.5	99.1	99.2	95.7
	MoE	97.1	70.8	78.7	87.6	95.2	85.8	94.1	95.8	99.1	99.6	97.2
	DSBN	97.2	71.6	77.9	88.7	96.1	86.3	93.2	91.6	98.9	99.3	95.8
	CMSS	99.0	75.3	88.4	93.7	97.7	90.8	96.0	93.7	99.3	99.6	97.2
	MOST	99.6	91.5	90.9	96.4	98.4	95.4	96.4	96.0	100	100	98.1
	STEM	99.4	89.7	89.9	97.5	98.4	95.0	98.4	94.2	100	100	98.2
Decentra.	FADA	92.5	64.5	63.5	82.8	91.7	79.0	84.2	88.7	87.1	88.1	87.1
	SHOT	98.2	80.2	84.5	91.1	97.1	90.2	96.4	96.2	98.5	99.7	97.7
	KD3A	99.2	87.3	85.6	89.4	98.5	92.0	97.4	**96.4**	98.4	99.7	97.9
	COPA	99.4	89.8	**91.0**	**97.5**	**99.2**	**95.4**	95.8	94.6	99.6	99.8	97.5
	AGREE	**99.5**	**92.0**	88.9	92.9	98.4	94.3	**99.0**	**96.4**	**100**	**100**	**98.9**

Digit-Five. As shown in the left of Table 1, our method achieves competitive performance with SOTA methods in terms of average overall accuracy. Moreover, the proposed approach achieves the oracle performance on MNIST and outperforms all the baselines when MNIST-M as target.

Office-Caltech10. As summarized in the right of Table 1, our AGREE outperforms both centralized and decentralized methods, and achieves new state-of-the-art results on Office-Caltech10. In particular, when adapting to DSLR (D) and Webcam (W), AGREE yields 100% accuracy.

DomainNet. From Table 2, we can clearly observe that our approach achieves new state-of-the-art results across all tasks on DomainNet, and even outperforms all compared centralized methods in the light of mean overall accuracy, although the DomainNet dataset is evidently challenging.

4.3 Ablation Study

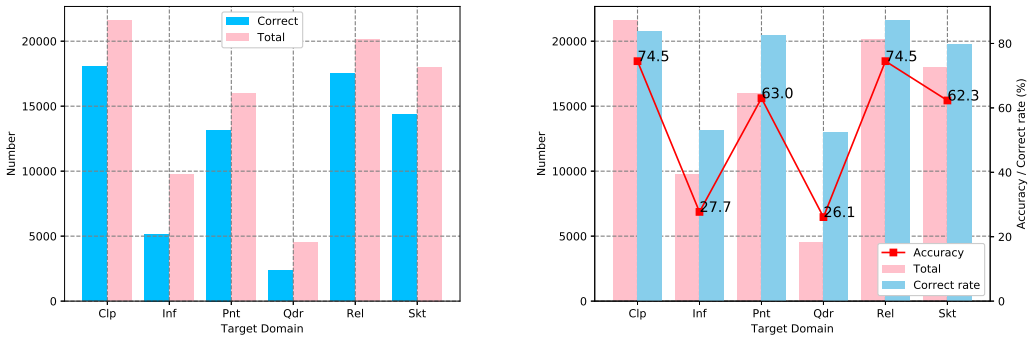
Component Effectiveness. As shown in the bottom of Table 2, without any component, the performance will markedly decline, while Adversarial Examples Guided Refinement has a major impact. It is noteworthy that without using Knowledge Contribution, our method still outperforms all compared decentralized methods and is comparable to the state-of-the-art centralized method.

Effect of Pseudo-Labels. In Figure 4(a), we present the quantities of pseudo-labels and correct pseudo-labels in each task on DomainNet. In Figure 4(b), we show the quantity and correct rate of pseudo-labels and the relationship between them and adaptation performance. Apparently, the test accuracy is positively related to the quantity and correct rate of pseudo-labels.

Target Weight Analysis. As shown in Table 3, we evaluate the influence of target weight. Different from setting a fixed value, our strategy can adaptively determine an appropriate target weight, which provides better performance. Too small a target weight will affect the protection of knowledge in the target domain, while too large will hinder the learning of knowledge in the source domains.

Table 2: Comparison with state-of-the-art UDA methods on DomainNet using ResNet-101. Our AGREE achieves new state-of-the-art results on the DomainNet dataset. Component effectiveness evaluation results in the bottom two lines show that every proposed component contributes to the adaptation performance in all target domains. AGREE w/o AEs: AGREE without Adversarial Examples Guided Refinement and AGREE w/o KC: AGREE without Knowledge Contribution.

Paradigm	Method	→ Clp	→ Inf	→ Pnt	→ Qdr	→ Rel	→Skt	Avg
W/o DA	Oracle	69.3	34.5	66.3	66.8	80.1	60.7	63.0
	Source-only	47.6	13.0	38.1	13.3	51.9	33.7	32.9
Centralized	DAN	48.4	14.8	40.2	15.3	53.9	34.0	34.5
	M^3SDA	58.6	26.0	52.3	6.3	62.7	49.5	42.6
	DANN	52.5	11.1	42.0	14.7	52.9	38.1	35.2
	DCTN	48.6	23.5	48.8	7.2	53.5	47.3	38.2
	MoE	55.8	21.3	46.2	9.2	63.3	46.3	40.4
	DSBN	57.2	25.6	52.3	6.5	62.7	47.6	42.0
	CMSDA	71.0	26.6	57.6	21.3	68.1	59.5	50.4
	STEM	72.0	28.2	61.5	25.7	72.6	60.2	53.4
Decentralized	FADA	52.3	16.3	41.9	13.9	52.7	36.8	35.7
	SHOT	58.8	19.8	49.9	9.7	66.0	46.3	41.3
	KD3A	72.5	23.4	60.9	16.4	72.7	60.6	51.1
	AGREE	**74.5**	**27.7**	**63.0**	**26.1**	**74.5**	**62.3**	**54.7**
	AGREE w/o AEs	69.0	21.4	56.0	18.6	68.9	58.1	48.7
	AGREE w/o KC	73.9	26.6	62.0	23.1	73.7	61.1	53.4



(a) (b)

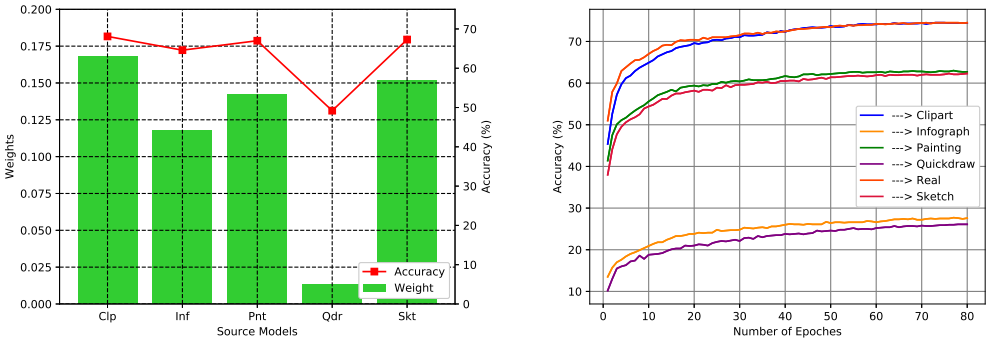
Figure 4: Quantitative comparison of pseudo-labels in the last communication round on DomainNet. (a) We present the quantities of pseudo-labels and correct pseudo-labels in each task. (b) we show the quantity and correct rate of pseudo-labels and the relationship between them and accuracy.

Source Weights and Convergence Performance. From Figure 5(a), we can see that the source weights properly reflect which source model performs better on the target domain. In Figure 5(b), we evaluate convergence performance of our method on DomainNet. Note that the proposed approach converges to a stable value with small fluctuations.

Variants of Ensemble. As presented in Table 4, without combining source models, the average overall accuracies decrease slightly, which means that there may be some knowledge in the source models that has not been learned by the target model. Even though ensemble methods employ multiple models to predict, our global model still performs best.

Table 3: Evaluation of different target weights on DomainNet. Here, $Adaptive$ denotes our strategy that sets a large target weight when encountering large domain shift and vice versa.

Target Weight	→ Clp	→ Inf	→ Pnt	→ Qdr	→ Rel	→ Skt	Avg
$[\frac{1}{N} : \frac{1}{N}]$	74.2	26.8	62.5	23.2	73.9	61.7	53.7
$[\frac{1}{N} : \frac{\sqrt{N}}{N}]$	**74.5**	27.4	**63.0**	25.0	**74.5**	62.2	54.4
$[\frac{1}{N} : \frac{N}{N}]$	72.4	27.1	62.1	26.0	74.2	60.8	53.8
$[\frac{1}{N} : Adaptive]$	**74.5**	**27.7**	**63.0**	**26.1**	**74.5**	**62.3**	**54.7**



(a) Source weights (b) Convergence curves

Figure 5: Analysis of source weights and convergence performance. (a) The source weights obtained by our method on DomainNet correlates positively with each source model performance. (b) Our method converges to a stable state with small fluctuations.

Table 4: Comparison with ensemble methods on DomainNet. Here, $TS\ ensemble$ refers to using the weighted logit ensemble of target and source models as the prediction, $TG\ ensemble$ refers to using the average logit ensemble of target and global models as the prediction and global model consists of the global feature extractor F_G and the target classifier C_T.

Component	→ Clp	→ Inf	→ Pnt	→ Qdr	→ Rel	→ Skt	Avg
Target model	73.6	27.6	62.4	25.9	73.8	61.9	54.2
TS ensemble	74.1	27.5	62.1	26.0	73.8	62.2	54.3
TG ensemble	74.3	**27.7**	62.7	26.0	74.3	62.1	54.5
Global model (ours)	**74.5**	**27.7**	**63.0**	**26.1**	**74.5**	**62.3**	**54.7**

5 Conclusion and Future Work

In this paper, we propose an effective approach to address decentralized UDA problem. To be specific, we filter out noisy predictions with adversarial examples to produce confident pseudo-labels for the optimization of target model. Furthermore, a knowledge contribution mechanism is introduced to generate reasonable model aggregation weights. Extensive experiments on three UDA benchmark datasets demonstrate the effectiveness of our method. Since the adaptation performance depends heavily on the quality of pseudo-labels and there are still some noisy pseudo-labels after refinement, we will manage to find a more effective pseudo-label calculation strategy in the future.

REFERENCES

Sk Miraj Ahmed, Dripta S Raychaudhuri, Sujoy Paul, Samet Oymak, and Amit K Roy-Chowdhury. Unsupervised multi-source domain adaptation without access to source data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10103–10112, 2021.

Zechen Bai, Zhigang Wang, Jian Wang, Di Hu, and Errui Ding. Unsupervised multi-source domain adaptation for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12914–12923, 2021.

Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. *Advances in neural information processing systems*, 29, 2016.

Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 7354–7362, 2019.

Chao Chen, Zhihang Fu, Zhihong Chen, Sheng Jin, Zhaowei Cheng, Xinyu Jin, and Xian-Sheng Hua. Homm: Higher-order moment matching for unsupervised domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 3422–3429, 2020.

Haoang Chi, Feng Liu, Wenjing Yang, Long Lan, Tongliang Liu, Bo Han, William Cheung, and James Kwok. Tohan: A one-step approach towards few-shot hypothesis adaptation. *Advances in Neural Information Processing Systems*, 34:20970–20982, 2021.

Jiahua Dong, Yang Cong, Gan Sun, Yuyang Liu, and Xiaowei Xu. Cscl: Critical semantic-consistent learning for unsupervised domain adaptation. In *European Conference on Computer Vision*, pp. 745–762. Springer, 2020a.

Jiahua Dong, Yang Cong, Gan Sun, Yunsheng Yang, Xiaowei Xu, and Zhengming Ding. Weakly-supervised cross-domain adaptation for endoscopic lesions segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(5), 2020b.

Jiahua Dong, Zhen Fang, Anjin Liu, Gan Sun, and Tongliang Liu. Confident anchor-induced multi-source free domain adaptation. *Advances in Neural Information Processing Systems*, 34:2848–2860, 2021.

Zhen Fang, Jie Lu, Feng Liu, Junyu Xuan, and Guangquan Zhang. Open set domain adaptation: Theoretical bound and algorithm. *IEEE transactions on neural networks and learning systems*, 32(10):4309–4322, 2020.

Haozhe Feng, Zhaoyang You, Minghao Chen, Tianye Zhang, Minfeng Zhu, Fei Wu, Chao Wu, and Wei Chen. Kd3a: Unsupervised multi-source decentralized domain adaptation via knowledge distillation. In *ICML*, pp. 3274–3283, 2021.

Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning*, pp. 1180–1189. PMLR, 2015.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.

Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, David Balduzzi, and Wen Li. Deep reconstruction-classification networks for unsupervised domain adaptation. In *European conference on computer vision*, pp. 597–613. Springer, 2016.

Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE conference on computer vision and pattern recognition*, pp. 2066–2073. IEEE, 2012.

Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

Jiang Guo, Darsh Shah, and Regina Barzilay. Multi-source domain adaptation with mixture of experts. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 4694–4703, 2018.

Hui Han, Wen-Yuan Wang, and Bing-Huan Mao. Borderline-smote: a new over-sampling method in imbalanced data sets learning. In *International conference on intelligent computing*, pp. 878–887. Springer, 2005.

Judy Hoffman, Mehryar Mohri, and Ningshan Zhang. Algorithms and theory for multiple-source adaptation. *Advances in Neural Information Processing Systems*, 31, 2018.

Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Model adaptation: Historical contrastive learning for unsupervised domain adaptation without source data. *Advances in Neural Information Processing Systems*, 34:3635–3649, 2021.

Jakub Konečnỳ, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*, 2016.

Vinod K Kurmi, Venkatesh K Subramanian, and Vinay P Namboodiri. Domain impression: A source data free domain adaptation method. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 615–625, 2021.

Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10285–10295, 2019.

Keqiuyin Li, Jie Lu, Hua Zuo, and Guangquan Zhang. Multi-source contribution learning for domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9641–9650, 2020.

Yitong Li, David E Carlson, et al. Extracting relationships by multi-domain matching. *Advances in Neural Information Processing Systems*, 31, 2018.

Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning*, pp. 6028–6039. PMLR, 2020.

Jian Liang, Dapeng Hu, Jiashi Feng, and Ran He. Dine: Domain adaptation from single and multiple black-box predictors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8003–8013, 2022.

Chuang Lin, Sicheng Zhao, Lei Meng, and Tat-Seng Chua. Multi-source domain adaptation for visual sentiment classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 2661–2668, 2020.

Feng Liu, Wenkai Xu, Jie Lu, Guangquan Zhang, Arthur Gretton, and Danica J Sutherland. Learning deep kernels for non-parametric two-sample tests. In *International conference on machine learning*, pp. 6316–6326. PMLR, 2020.

Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pp. 97–105. PMLR, 2015.

Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. PMLR, 2017.

Tuan Nguyen, Trung Le, He Zhao, Quan Hung Tran, Truyen Nguyen, and Dinh Phung. Most: Multi-source domain adaptation via optimal transport for student-teacher learning. In *Uncertainty in Artificial Intelligence*, pp. 225–235. PMLR, 2021a.

Van-Anh Nguyen, Tuan Nguyen, Trung Le, Quan Hung Tran, and Dinh Phung. Stem: An approach to multi-source domain adaptation with guarantees. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9352–9363, 2021b.

Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1406–1415, 2019.

Xingchao Peng, Zijun Huang, Yizhe Zhu, and Kate Saenko. Federated adversarial domain adaptation. In *International Conference on Learning Representations*, 2020.

Paolo Russo, Tatiana Tommasi, and Barbara Caputo. Towards multi-source adaptive semantic segmentation. In *International Conference on Image Analysis and Processing*, pp. 292–301. Springer, 2019.

Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3723–3732, 2018.

Marin Scalbert, Maria Vakalopoulou, and Florent Couzinié-Devy. Multi-source domain adaptation via supervised contrastive learning and confident consistency regularization. *arXiv preprint arXiv:2106.16093*, 2021.

Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7167–7176, 2017.

Hongyi Wang, Mikhail Yurochkin, Yuekai Sun, Dimitris Papailiopoulos, and Yasaman Khazaeni. Federated learning with matched averaging. In *International Conference on Learning Representations (ICLR)*, 2020.

Ximei Wang, Ying Jin, Mingsheng Long, Jianmin Wang, and Michael I Jordan. Transferable normalization: Towards improving transferability of deep neural networks. *Advances in neural information processing systems*, 32, 2019.

Guile Wu and Shaogang Gong. Collaborative optimization and aggregation for decentralized domain generalization and adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6484–6493, 2021.

Ruijia Xu, Ziliang Chen, Wangmeng Zuo, Junjie Yan, and Liang Lin. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3964–3973, 2018.

Luyu Yang, Yogesh Balaji, Ser-Nam Lim, and Abhinav Shrivastava. Curriculum manager for source selection in multi-source domain adaptation. In *European Conference on Computer Vision*, pp. 608–624. Springer, 2020.

Pan Zhang, Bo Zhang, Ting Zhang, Dong Chen, Yong Wang, and Fang Wen. Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12414–12424, 2021a.

Yang Zhang, Philip David, and Boqing Gong. Curriculum domain adaptation for semantic segmentation of urban scenes. In *Proceedings of the IEEE international conference on computer vision*, pp. 2020–2030, 2017.

Yixin Zhang, Zilei Wang, and Yushi Mao. Rpn prototype alignment for domain adaptive object detector. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12425–12434, 2021b.

Yiyang Zhang, Feng Liu, Zhen Fang, Bo Yuan, Guangquan Zhang, and Jie Lu. Clarinet: a one-step approach towards budget-friendly unsupervised domain adaptation. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 2526–2532, 2021c.

Han Zhao, Shanghang Zhang, Guanhang Wu, José MF Moura, Joao P Costeira, and Geoffrey J Gordon. Adversarial multiple source domain adaptation. *Advances in neural information processing systems*, 31, 2018.

Sicheng Zhao, Bo Li, Xiangyu Yue, Yang Gu, Pengfei Xu, Runbo Hu, Hua Chai, and Kurt Keutzer. Multi-source domain adaptation for semantic segmentation. *Advances in Neural Information Processing Systems*, 32, 2019.