

# SEP-REP-NET: MULTI-SOURCE FREE DOMAIN ADAPTATION VIA MODEL SEPARATION AND REPARAMETERIZATION

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

We consider multi-source free domain adaptation, the problem of adapting multiple existing models to a new domain without accessing the source data. This is a practical problem, which often arises in commercial settings but remains an open question despite the advances in recent years. Previous methods, *e.g.*, model ensemble, are effective, but they also incur significantly increased computational costs. Conventional solutions for efficiency, such as distillation, are limited in preserving source knowledge, *i.e.*, maintaining generalizability. In this work, we propose a novel framework called **SepRep-Net**, which tackles multi-source free domain adaptation via model **Separation** and **Reparameterization**. Concretely, SepRep-Net reassembled multiple existing models to a unified network, while maintaining separate pathways (**Separation**). During training, separate pathways are optimized in parallel with the information exchange regularly performed via an additional feature merging unit. With our specific design, these pathways can be further reparameterized into a single one to facilitate inference (**Reparameterization**). SepRep-Net is characterized by **1) effectiveness**: competitive performance on the target domain, **2) efficiency**: low computational costs, and **3) generalizability**: maintaining more source knowledge than existed solutions. As a general approach, SepRep-Net can be seamlessly plugged into various methods. Extensive experiments validate the performance of SepRep-Net on mainstream benchmarks. We promise to release our code towards reproducibility.

## 1 INTRODUCTION

Domain Adaptation Pan & Yang (2010) aims at adapting models from source domains to a target domain with unlabeled target data. Previous works Tzeng et al. (2014); Long et al. (2015); Ganin et al. (2016) mostly rely on source data, *i.e.*, samples from the source domain, during adaptation. However, the source data may become inaccessible due to storage and privacy concerns in commercial settings, which hinders the wide applications of domain adaptation. Towards this obstacle, great interests have been invoked to explore source-free domain adaptation (SFDA) Liang et al. (2020), where models pre-trained on the labeled source data are adapted to the unlabeled target domain without accessing source data. One extension of SFDA is to leverage knowledge from multiple existing models that are pre-trained on different source domains, which is formulated as multi-source free domain adaptation (MSFDA).

Among mainstream MSFDA approaches Wang et al. (2020b); Ahmed et al. (2021); Dong et al. (2021), model ensemble Wang et al. (2020b) serves as a simple yet strong baseline, where multiple source models are adapted to the target domain respectively. Then these models vote for correct predictions. These methods Ahmed et al. (2021); Dong et al. (2021) based on model ensemble bring remarkable improvements, but they inevitably lead to high computational overheads. To alleviate such a burden, knowledge distillation is usually leveraged to train a single model via mimicking the ensembled predictions on the target domain. Although distillation preserves the accuracy on the target domain and reduces computation costs, it incurs another problem: severe source knowledge forgetting, which harms the generalizability in real-world applications Yang et al. (2021).

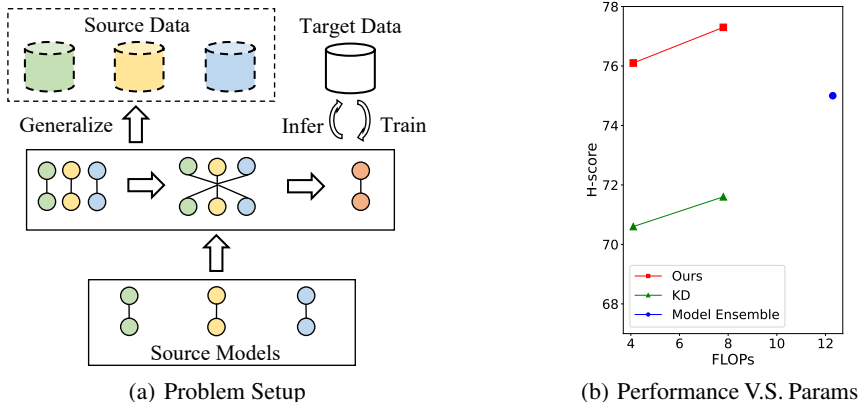


Figure 1: **Problem Setup and Method Performance.** (a) We reassemble multiple existing source models to one target model that is adapted to a novel unlabeled target domain. The target model is expected to be 1) effective for target data, 2) efficient in inference and 3) able to generalize well on source data, *i.e.* preserving more source knowledge. (b) Compared to previous methods on Office-Home, our framework enjoys better effectiveness, generalizability, and efficiency (H-score  $\frac{2 \cdot Acc_{source} \cdot Acc_{target}}{Acc_{source} + Acc_{target}}$  (%) evaluates effectiveness and generalizability jointly).

Considering all the challenges mentioned above, this paper targets on a more practical setting, Multi-Source Free Domain Adaptation with Effectiveness, Efficiency, and Generalizability. As shown in Table 1, we focus on source-free domain adaptation with multiple existing source models, aiming at preserving model efficiency and source knowledge, *i.e.*, generalizability, as well as achieving competitive target domain accuracy. To achieve this goal, an intuitive solution is to **reassemble multiple source models**, which have the same network architecture, into a single model. The reassembled model will be further adapted to the target domain. The main challenge in reassembling multiple source models lies in the trade-off between preserving more parameters and pursuing model efficiency. In this paper, we propose **SepRep-Net** (Figure. 1(a)), a framework for multi-source free domain adaptation via Separation and Reparameterization. For training, SepRep-Net inherits all the parameters from source models and reassembles them via Separation. While during inference, thanks to our carefully designed framework, the model can be further simplified via reparameterization.

Concretely, SepRep-Net reassembles models by formulating multiple separate pathways in parallel (Separation). Specifically, in each Conv-BN unit, the input passes through separate pathways. The multiple outputs are then integrated into a single one by a feature merging unit. The unified feature serves as the input of the next Conv-BN units in different pathways. With this design, during inference, we can convert multiple pathways into a single one via reparameterization to enhance model efficiency. In other words, we eventually convert multiple source models into a single target model. Moreover, when ensembling final classifier outputs, previous works Ahmed et al. (2021) adopt learnable combination weights, resulting in bias to target data. In this paper, we revise the reweighting strategy to an uncertainty-based one to further boost generalizability. Our method is characterized by 1) Effectiveness: reassembled model is ready to be adapted to the target data, 2) Efficiency: largely reduces computational costs during inference via reparameterization, 3) Generalizability: parameters inherited from multiple source models naturally preserve source knowledge.

Our proposed SepRep-Net can be readily integrated into various existing methods. Extensive experiments are performed on several benchmarks to evaluate SepRep-Net. Experimental results (as illustrated in Figure 1(b) as an example) show that SepRep-Net achieves a better trade-off among effectiveness, efficiency, and generalizability.

## 2 RELATED WORK

**Domain Adaptation** Aiming at mitigating the distribution gap between the source and target domains, mainstream domain adaptation methods mostly fall into two paradigms, *i.e.*, moment match-

Table 1: **Comparison of different domain adaptation settings.** Adapting multiple source domains to a novel target domain without source data, our method takes effectiveness, efficiency, and preserving source knowledge (*i.e.* generalizability) into consideration.

Method	No source data	Multiple source domains	Model efficiency	Source knowledge
UDA Pan & Yang (2010)	✗	✗	-	-
SFDA Liang et al. (2020)	✓	✗	-	-
MSDA Peng et al. (2019)	✗	✓	✗	-
G-SFDA Yang et al. (2021)	✓	✗	-	✓
MSFDA Dong et al. (2021)	✓	✓	✗	✗
MSFDA + KD Ahmed et al. (2021)	✓	✓	✓	✗
<b>SepRep-Net (Ours)</b>	✓	✓	✓	✓

ing Gretton et al. (2012); Long et al. (2015); Tzeng et al. (2014); Long et al. (2017); Kang et al. (2019); Lee et al. (2019) and adversarial training Ganin et al. (2016); Tzeng et al. (2017); Pei et al. (2018); Long et al. (2018); Hoffman et al. (2018); Zhang et al. (2019). The former alleviates domain shift by minimizing feature discrepancy while the latter borrows the spirit of Generative Adversarial Networks (GANs) Goodfellow et al. (2014); Mirza & Osindero (2014) to learn domain invariant features. Recent works enlighten some novel perspectives for domain adaptation, such as clustering Pan et al. (2019); Tang et al. (2020), self-training Shu et al. (2018); Zou et al. (2018; 2019); Jin et al. (2020); Saito et al. (2020), network architecture design Carlucci et al. (2017); Li et al. (2018); Wang et al. (2019) and feature norm Xu et al. (2019). In order to leverage the source knowledge effectively, accessing source data during training is necessary for these methods, which may be unavailable under some circumstances.

**Source-free Domain Adaptation** Adapting source models to a novel unlabeled target domain without accessible source data, namely Source-free Domain Adaptation Liang et al. (2020) (SFDA), is a highly practical problem. Recent works have explored diverse scenarios in this field, *e.g.* closed-set Liang et al. (2020); Li et al. (2020); Liang et al. (2021); Yang et al. (2020b), open-set Kundu et al. (2020b) and universal SFDA Kundu et al. (2020a). Our work is related to multi-source SFDA Ahmed et al. (2021) (MSFDA) and Generalized SFDA Yang et al. (2021) (G-SFDA), which pay attention to adapting multiple models to the novel target domain and the model performance on source domains, *i.e.* maintaining source knowledge, respectively.

**Multi-source Free Domain Adaptation** Previous works Xu et al. (2018); Peng et al. (2019); Wang et al. (2020a); Yang et al. (2020a); Nguyen et al. (2021) widely investigate multi-source domain adaptation Sun et al. (2015) (MSDA) which adapts multiple source domains with different distributions. When MSDA meets the source-free setting, the problem becomes multi-source free domain adaptation (MSFDA). Model ensemble works as a simple yet effective baseline. Recent advances Ahmed et al. (2021); Dong et al. (2021) push it to a stronger level but still suffer from heavy computational costs. To reduce the computation overheads, knowledge distillation (KD) is adopted as an effective solution. It preserves the target performance but leads to severe source knowledge forgetting, harming the generalizability.

### 3 METHODOLOGY

This work aims at proposing a framework for adapting multiple existing models from different source domains to a novel target domain without accessing source data. In this setting, we are given  $K$  source models pre-trained on different source data and  $n_t$  unlabeled samples  $\mathcal{X}_t = \{x_t^i\}_{i=1}^{n_t}$  from the target domain. Both the source domains and the target domain have  $C$  categories. Thus, it is a  $C$ -way classification task. The  $k^{th}$  source model consists of one feature extractor  $g_s^k$  and one classifier  $h_s^k$ . And all source models are homogeneous in network architecture. Our goal is to reassemble these multiple source models to a new model on the target domain using unlabeled target data, pursuing effectiveness, efficiency, and generalizability.

To smooth the presentation of our framework, we start from the preliminaries. Then, we propose SepRep-Net, an efficient framework for multi-source free domain adaption.

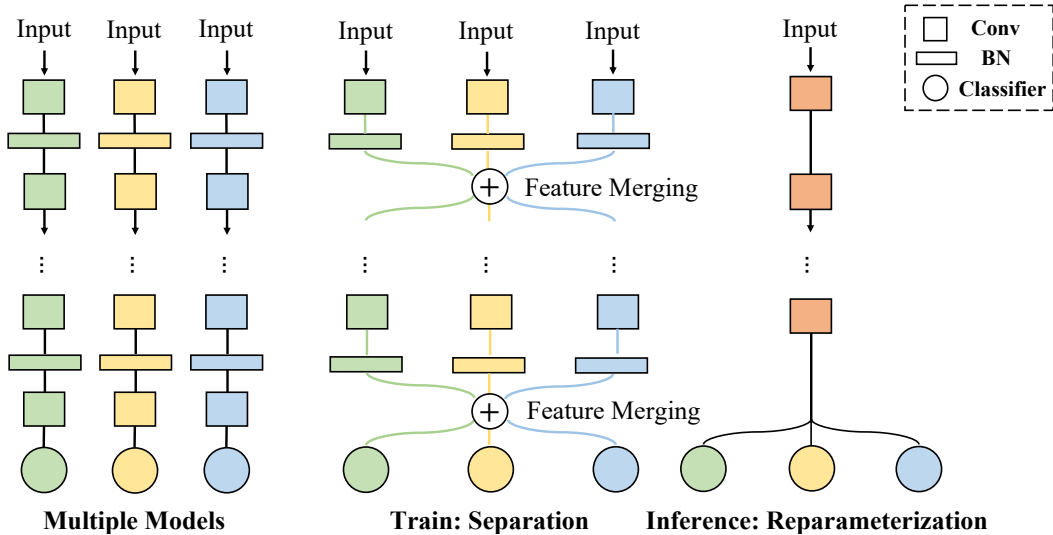


Figure 2: **Method Overview.** Take three source models as an example. **Train:** In each Conv-BN unit, separate pathways are applied in parallel, forming a structure with multiple pathways. The outputs are then integrated into a unified output via the feature merging unit. **Inference:** Multiple pathways are re-parameterized to a unique one during inference. An uncertainty-based weighting strategy ensembles multiple classifier heads to obtain the final prediction. (*Best viewed in color*)

### 3.1 PRELIMINARIES

**Model Ensemble** We target on source-free domain adaptation (SFDA) from multiple models, which is also coined Multi-Source Free Domain Adaptation (MSFDA). Some methods are designed specifically for this problem and proved to be effective. Except for them, another simple yet robust solution for this problem is model ensemble, where we utilize the existing SFDA methods (e.g. SHOT Liang et al. (2020)) to adapt each source model to the target data. As a result, we obtain  $K$  adapted target models  $\{g_t^k, h_t^k\}_{k=1}^K$  and can ensemble predictions from these adapted models as the final result,

$$\hat{y}_t = \sum_{k=1}^K \alpha_k \cdot h_t^k(g_t^k(x_t)), \quad (1)$$

where  $\alpha_k$  denotes the importance weight of the  $k^{th}$  model. We may simply take  $\alpha_k = \frac{1}{K}$ . Though effective, model ensemble passes each data sample through all  $K$  models in inference, bringing around  $K$  times computational and parameter costs.

### 3.2 SEPREP-NET

In this paper, we propose a framework for integrating multiple source models, namely SepRep-Net, to boost efficiency, generalizability, as well as effectiveness. As shown in Figure 2, SepRep-Net reassembles multiple source models to a single model by formulating separate pathways during training (Separation). In each Conv-BN unit, the input data passes through multiple pathways, and then the corresponding outputs are merged into a single one. During inference, such a design enables us to convert the multiple pathways to a single one via reparameterization.

**Train: Separation** During the training period, SepRep-Net inherits the parameters from source models and reassembles them into one network. As shown in Figure 2, the weights from source models (square: Conv, rectangle: BN) form multiple separate pathways for model forward. Afterward, the outputs from different pathways are converted into a unified one via the feature merging unit, which integrates them via weighted sum. Concretely, take one Conv-BN unit as an example, denote the input as  $f^{input}$  and the corresponding output as  $f^{output}$ , we have

$$f^{output} = \sum_k BN_k(Conv_k(f^{input})) \cdot w_k, \quad (2)$$

where  $Conv_k$  and  $BN_k$  indicates the convolution and normalization layer from the  $k^{th}$  model respectively, and  $w_k$  is the weight hyper-parameter of the  $k^{th}$  model. Through this framework, we can obtain multiple classifier outputs.

Next, we need to ensemble these classifier outputs. Here, we consider that different pathways cannot be viewed as equal during training. Therefore, a re-weighting strategy shall be introduced. Concretely, we state that the models with lower uncertainty on the target data are more likely to have stronger performance. To highlight these models, we first quantify the model uncertainty on the target data by Entropy.

$$M(g_s^k, h_s^k) = -\mathbb{E}_{x_t \in \mathcal{X}_t} \sum_{c=1}^C h_s^k(g_s^k(x_t))^c \log(h_s^k(g_s^k(x_t))^c), \quad (3)$$

where  $h_s^k(\cdot)^c$  indicates the prediction on the  $c^{th}$  category of classifier  $h_s^k$ , and  $M(g_s^k, h_s^k)$  is a score that indicates the prediction uncertainty of the  $k^{th}$  source model on target data  $x_t$ . A lower score demonstrates lower prediction uncertainty, which shows that the model is potentially more capable of adapting to the target data.

Our proposed SepRep-Net leverages the uncertainty metric in Eq. 3 with Softmax function to re-weight the losses on different classifiers,

$$L_{total} = \sum_{k=1}^K \frac{e^{-M(g_s^k, h_s^k)}}{\sum_{i=1}^K e^{-M(g_s^i, h_s^i)}} L_k(f_k; \mathcal{X}_t), \quad (4)$$

where  $L_k(f_k; \mathcal{X}_t)$  denotes the loss function of the  $k^{th}$  model. The loss functions Liang et al. (2020) in previous works for SFDA can be readily served as  $L_k$  in SepRep-Net.

**Inference: Reparameterization** For inference, as shown in Figure 2, our framework design enables us to merge these separate pathways into a unified one with model reparameterization, which enhances the model efficiency.

Concretely, take one Conv-BN unit as an instance, suppose one input feature feeds to this unit is  $x^{(1)} \in \mathbb{R}^{H_1 \times W_1 \times C_1}$ , where  $H_1 \times W_1$  shows the spatial resolution and  $C_1$  is the channel size. For the convolution layer, take the one in the  $k^{th}$  pathway as an instance, suppose there are  $C_2$  convolution filters in total, the  $j^{th}$  filter can be denoted as  $F^{(k,j)} \in \mathbb{R}^{U \times V \times C_1}$ , the corresponding output of this convolution layer  $x^{(2)} \in \mathbb{R}^{H_2 \times W_2 \times C_2}$  follows

$$x_{:::,j}^{(2)} = \sum_{i=1}^{C_1} x_{:::,i}^{(1)} * F_{:::,i}^{(k,j)}, \quad (5)$$

where  $x_{:::,i}^{(1)}$  is the  $i^{th}$  channel of  $x^{(1)}$ ,  $F_{:::,i}^{(k,j)}$  is the  $i^{th}$  channel of  $F^{(k,j)}$ , and  $*$  is the 2D convolution operator. Afterward, it passes through the batch normalization layer in this pathway, and the output  $o \in \mathbb{R}^{H_2 \times W_2 \times C_2}$  becomes

$$o_{:::,j} = (x_{:::,j}^{(2)} - \tilde{\mu}_j^k) \frac{\gamma_j^k}{\tilde{\sigma}_j^k} + \beta_j^k, \quad (6)$$

where  $\tilde{\mu}_j^k$  and  $\tilde{\sigma}_j^k$  are the  $j^{th}$  element of the mean and standard deviation statistics in batch normalization,  $\gamma_j^k$  and  $\beta_j^k$  are the  $j^{th}$  element of the learned scaling factor and bias respectively. Therefore, the weighted sum of the outputs of multiple pathways is

$$x_{:::,j}^{(3)} = \sum_k w_k \cdot \left( \left( \sum_{i=1}^{C_1} x_{:::,i}^{(1)} * F_{:::,i}^{(k,j)} - \tilde{\mu}_j^k \right) \frac{\gamma_j^k}{\tilde{\sigma}_j^k} + \beta_j^k \right). \quad (7)$$

We can reparameterize the output as follows,

$$F^{(j)} = \frac{w_1 \gamma_j^1}{\sigma_j^1} F^{(1,j)} \oplus \frac{w_2 \gamma_j^2}{\sigma_j^2} F^{(2,j)} \oplus \dots \oplus \frac{w_K \gamma_j^K}{\sigma_j^K} F^{(K,j)}, \quad (8)$$

$$b_j = -w_1 \frac{\tilde{\mu}_j^1 \gamma_j^1}{\tilde{\sigma}_j^1} - \dots - w_K \frac{\tilde{\mu}_j^K \gamma_j^K}{\tilde{\sigma}_j^K} + w_1 \beta_j^1 + \dots + w_K \beta_j^K, \quad (9)$$

$$x_{:::,j}^{(3)} = \sum_{k=1}^C x_{:::,i}^{(1)} * F_{:::,i}^{(j)} + b_j, \quad (10)$$

where  $\oplus$  indicates element-wise addition on the corresponding positions. Now, the multiple pathways in one Conv-BN unit, with  $K$  convolution layers and  $K$  batch normalization layers, are converted to one convolution layer, as shown in Eq. 10.

As shown in Figure. 2, multiple models are finally reassembled into a unified architecture with multiple classifier heads. Since the parameter and computation overheads of classifier heads are negligible compared to the feature extractor, the computation and parameter costs of SepRep-Net are very close to a single model.

**Importance Reweighting** Finally, akin to the spirit of model ensemble, we take the weighted sum of multiple classifiers as the final prediction,

$$\hat{y}_t = \sum_{k=1}^K \alpha_k \cdot h_t^k(g_t(x_t)), \quad (11)$$

where  $\alpha_k$  is the ensemble weight which plays an important role in model ensemble. A recent work Ahmed et al. (2021) proposed to learn a combination weight during the training procedure. However, such a strategy inevitably causes bias to target data, which hurts the model generalizability. Here, we revise the reweighting strategy to a parameter-free one. We take the entropy of the trained models to formulate our reweighting strategy, through which we can highlight the model that is more certain on the test data,

$$\alpha_k = \frac{e^{-M(g_t, h_t^k)}}{\sum_{i=1}^K e^{-M(g_t, h_t^i)}}. \quad (12)$$

Notably,  $M(g_t, h_t^k)$  is different from the entropy criterion  $M(g_s^k, h_s^k)$  in Eq. 3 since 1) it is calculated by adapted model instead of the source models to obtain a more accurate approximation of uncertainty, 2) it is computed on-the-fly in inference, harvesting importance weights that are adaptive to the input data.

## 4 EXPERIMENTS

### 4.1 EXPERIMENTAL SETUP

**Datasets** We perform extensive evaluation of SepRep-Net on five mainstream benchmark datasets, *i.e.*, **Office-31** Saenko et al. (2010), **Office-Home** Venkateswara et al. (2017), **Digit5** Peng et al. (2019), **Office-Caltech** Gong Boqing & Grauman (2012), and **DomainNet** Peng et al. (2019). Due to the page limit, the results on **Office-Caltech** and detailed descriptions of the datasets are included in Appendix.

### 4.2 EXPERIMENTAL RESULTS

We build SepRep-Net based on various baseline frameworks and evaluate the effectiveness, efficiency, and generalizability of SepRep-Net on different benchmarks. Specifically, we plug SepRep-Net into DECISION Ahmed et al. (2021), SHOT Liang et al. (2020) and CAiDA Dong et al. (2021), respectively. DECISION and CAiDA are designed for MSFDA tasks. SHOT is originally designed for SFDA, and it can be readily applied to MSFDA via model ensemble (SHOT-ens). We compare it with the vanilla results with multiple models and results with knowledge distillation Hinton et al. (2015) (KD). Following Yang et al. (2021), we report the average accuracy on source domains  $Acc_s$  (%), the accuracy on target domain  $Acc_t$  (%) and the H-score  $\frac{2 \cdot Acc_s \cdot Acc_t}{Acc_s + Acc_t}$  (%) which jointly evaluates effectiveness and generalizability. On efficiency, we report the FLOPs of each method. The experiment results are reported in Table 2, Table 3, Table 4 and Table 5.

Table 2: Source Accuracy (S) (%), Target Accuracy (T) (%), H-Score (H), and Model Efficiency on Office-31 dataset. ResNet-50 is adopted in experiments. A, D, W indicate different domains (A: Amazon, D: DSLR, W: Webcam). SHOT-ens indicates the performance of model ensemble with all models that are adapted via SHOT. KD indicates knowledge distillation.

METHOD	FLOPS	D, W $\rightarrow$ A			A, W $\rightarrow$ D			A, D $\rightarrow$ W			Average		
		S	T	H	S	T	H	S	T	H	S	T	H
DECISION + KD Hinton et al. (2015)	4.1	73.4	75.4	74.4	72.5	99.6	83.9	76.7	98.1	86.1	74.2	91.0	81.4
DECISION + <b>SepRep-Net</b>	4.1	87.5	75.9	81.3	90.3	99.7	94.8	88.6	98.5	93.3	<b>88.8</b>	<b>91.4</b>	<b>90.1</b>
DECISION Ahmed et al. (2021)	8.2	92.8	75.4	83.2	81.9	99.6	89.8	83.1	98.4	90.1	85.9	91.1	88.4
CAiDA + KD Hinton et al. (2015)	4.1	84.2	75.7	79.7	85.1	99.4	91.7	84.7	98.7	91.2	84.5	91.3	87.8
CAiDA + <b>SepRep-Net</b>	4.1	92.0	76.1	83.3	89.3	99.7	94.2	89.2	99.0	93.8	<b>90.2</b>	<b>91.6</b>	<b>90.9</b>
CAiDA Dong et al. (2021)	8.2	88.9	75.8	81.8	89.1	99.8	94.1	88.6	98.9	93.5	88.9	91.6	90.2
SHOT-ens + KD Hinton et al. (2015)	4.1	85.3	74.9	79.8	86.4	97.8	91.7	85.8	94.8	90.1	85.8	89.2	87.5
SHOT + <b>SepRep-Net</b>	4.1	92.8	75.7	83.4	93.9	98.8	96.3	91.4	96.6	93.9	<b>92.7</b>	<b>90.4</b>	<b>91.5</b>
SHOT-ens Liang et al. (2020)	8.2	90.0	75.0	81.8	90.2	97.8	93.8	90.5	94.9	92.6	90.2	89.3	89.7

Table 3: Source Accuracy (S) (%), Target Accuracy (T) (%), H-Score (H), and Model Efficiency on Office-Home dataset. ResNet-50 is adopted in experiments. Ar, Cl, Pr, Rw indicate different domains (Ar: Art, Cl: Clipart, Pr: Product, Rw: Real World). The abbreviations in METHOD are the same as Table 2.

METHOD	FLOPS	Cl,Pr,Rw $\rightarrow$ Ar			Ar,Pr,Rw $\rightarrow$ Cl			Ar,Cl,Rw $\rightarrow$ Pr			Ar,Cl,Pr $\rightarrow$ Rw			Average		
		S	T	H	S	T	H	S	T	H	S	T	H	S	T	H
DECISION + KD Hinton et al. (2015)	4.1	57.2	74.4	64.7	51.4	59.1	55.0	52.7	84.4	64.9	59.6	83.7	69.6	55.2	75.4	63.5
DECISION + <b>SepRep-Net</b>	4.1	73.2	74.6	73.9	69.5	60.2	64.5	68.8	84.0	75.6	71.8	83.8	77.3	<b>70.8</b>	<b>75.7</b>	<b>73.2</b>
DECISION Ahmed et al. (2021)	12.3	68.7	74.5	71.5	57.3	59.4	58.3	67.7	84.4	75.1	66.8	83.6	74.3	65.1	75.4	69.9
CAiDA + KD Hinton et al. (2015)	4.1	68.8	75.0	71.8	63.4	60.4	61.9	66.6	84.5	74.5	66.0	84.0	73.9	66.2	76.0	70.8
CAiDA + <b>SepRep-Net</b>	4.1	75.2	75.3	75.2	73.6	60.7	66.5	75.0	85.3	79.8	74.8	84.8	79.5	<b>74.7</b>	<b>76.5</b>	<b>75.6</b>
CAiDA Dong et al. (2021)	12.3	75.0	75.2	75.1	73.8	60.5	66.5	74.5	84.7	79.3	73.9	84.2	78.7	74.3	76.2	75.2
SHOT-ens + KD Hinton et al. (2015)	4.1	69.0	72.1	70.5	65.6	59.0	62.1	67.9	82.7	74.6	66.5	82.9	73.8	67.3	74.2	70.6
SHOT + <b>SepRep-Net</b>	4.1	82.6	72.3	77.1	75.6	59.2	66.4	75.1	83.3	79.0	77.8	83.3	80.5	<b>77.8</b>	<b>74.5</b>	<b>76.1</b>
SHOT-ens Liang et al. (2020)	12.3	76.5	72.2	74.3	75.5	59.3	66.4	75.4	82.8	78.9	75.3	82.9	78.9	75.7	74.3	75.0

In these tables, for SFDA methods such as SHOT, to tackle the multi-source free scenario, we first adapt multiple models to the target domain via SHOT one by one. SHOT-ens indicates the performance of model ensemble with all source models that are adapted via SHOT.

**Office31** Table 2 shows that knowledge distillation (KD) is effective in enhancing efficiency and maintaining target performance, but it harms generalizability (obviously low source accuracy). On the contrary, reducing computational costs by half, SepRep-Net preserves significantly more generalizability than KD, as well as achieves better results in both source and target domains.

**Office-Home** Table 3 demonstrates that when tackling three source models, our framework shows consistently higher performance than other methods in both source and target domains. We note that when compared with the vanilla methods, our method is still superior to them both in effectiveness and generalizability, with only 33% computational costs of them.

**Digit5** Moreover, we further evaluate our method on Digit5, a harder dataset with more domains and larger domain gaps. The results in Table 4 prove that our method only requires less than 25% of the computational costs, while consistently improving the original method with 100% costs in both source and target accuracy. Here, we also consider previous works that have access to source data during being adapted to the target domain. It is unfair to evaluate the source domain accuracy of these methods. Therefore, we only consider the accuracy of these methods in the target domain. Our method achieves a stronger target domain accuracy over these methods.

**DomainNet** Finally, we take DomainNet, a larger dataset in domain adaptation, with obvious domain gaps across the six domains. Table 5 shows that when compared with knowledge distillation, our method shows obvious improvements in effectiveness and generalizability under equal compu-

Table 4: Source Accuracy (S) (%), Target Accuracy (T) (%), H-Score (H), and Model Efficiency on Digit5. The backbone network in Peng et al. (2019) is adopted to recognize digits. MM, MT, UP, SV, and SY indicate different domains (MM: MNIST-M, MT: MNIST, UP: USPS, SV: SVHN, SY: Synthetic Digits).  $\rightarrow$  points to the target domain while the remaining domains serve as source domains. The abbreviations in METHOD are the same as Table 2, and traditional domain adaptation methods that require source data (from DAN to M3SDA) are also compared.

METHOD	FLOPS	$\rightarrow$ MM			$\rightarrow$ MT			$\rightarrow$ UP			$\rightarrow$ SV			$\rightarrow$ SY			Avg		
		S	T	H	S	T	H	S	T	H	S	T	H	S	T	H	S	T	H
DAN Long et al. (2015)	0.116	-	63.7	-	-	96.3	-	-	94.2	-	-	62.5	-	-	85.4	-	-	80.4	-
DANN Gamin et al. (2016)	0.116	-	71.3	-	-	97.6	-	-	92.3	-	-	63.5	-	-	85.3	-	-	82.0	-
MCD Saito et al. (2018)	0.116	-	72.5	-	-	96.2	-	-	95.3	-	-	78.9	-	-	87.5	-	-	86.1	-
CORAL Sun & Saenko (2016)	0.116	-	62.5	-	-	97.2	-	-	93.4	-	-	64.4	-	-	82.7	-	-	80.1	-
ADDA Tzeng et al. (2017)	0.116	-	71.6	-	-	97.9	-	-	92.8	-	-	75.5	-	-	86.5	-	-	84.8	-
M3SDA Peng et al. (2019)	0.116	-	72.8	-	-	98.4	-	-	96.1	-	-	81.3	-	-	89.6	-	-	87.6	-
DECISION + KD Hinton et al. (2015)	0.029	55.3	92.8	69.3	52.3	99.2	68.5	46.6	97.8	63.1	52.4	82.6	64.1	70.9	97.2	82.0	55.5	93.9	69.4
DECISION + SepRep-Net	0.032	61.8	93.1	74.3	57.0	99.1	72.4	65.8	98.0	78.7	62.1	83.9	71.4	73.4	97.4	83.7	<b>64.0</b>	<b>94.3</b>	<b>76.3</b>
DECISION Ahmed et al. (2021)	0.116	58.4	93.0	71.7	55.1	99.2	70.8	55.8	97.8	71.1	54.0	82.6	65.3	71.0	97.5	82.2	58.9	94.0	72.4
CAiDA + KD Hinton et al. (2015)	0.029	53.6	93.6	68.2	48.9	98.9	65.4	49.8	98.2	66.1	50.1	83.1	62.5	65.9	97.6	78.7	53.7	94.3	68.4
CAiDA + SepRep-Net	0.032	57.1	94.0	71.0	56.4	99.2	71.9	55.9	98.8	71.4	56.8	85.7	68.3	74.2	98.3	84.6	<b>60.1</b>	<b>95.2</b>	<b>73.7</b>
CAiDA Dong et al. (2021)	0.116	55.6	93.7	69.8	52.3	99.1	68.5	53.1	98.6	69.0	52.8	83.3	64.6	68.0	98.1	80.3	56.4	94.6	70.6
SHOT-ens + KD Hinton et al. (2015)	0.029	57.1	90.2	69.9	53.0	98.8	69.0	50.9	97.8	67.0	53.2	58.1	55.5	67.9	93.9	75.1	56.4	85.8	68.1
SHOT + SepRep-Net	0.032	67.6	95.8	79.3	63.1	98.6	77.0	70.7	97.6	82.0	71.2	82.9	76.6	75.9	93.1	83.6	<b>69.7</b>	<b>93.6</b>	<b>79.9</b>
SHOT-ens Liang et al. (2020)	0.116	60.3	90.4	72.3	56.7	98.9	72.1	60.8	97.7	75.0	54.9	58.3	56.5	68.9	93.9	75.6	60.3	85.8	70.8

Table 5: Source Accuracy (S) (%), Target Accuracy (T) (%), H-Score (H), and Model Efficiency on DomainNet dataset. ResNet-101 is adopted in experiments. C, I, P, Q, R, S indicate different domains. (C: Clipart, I: Infograph, P: Painting, Q: Quickdraw, R: Real World, S: Sketch). The abbreviations in METHOD are the same as Table 4.

METHOD	FLOPS	I.P.Q.R.S $\rightarrow$ C			C.P.Q.R.S $\rightarrow$ I			C.I.Q.R.S $\rightarrow$ P			C.I.P.R.S $\rightarrow$ Q			C.I.P.Q.S $\rightarrow$ R			C.I.P.Q.R $\rightarrow$ S			Avg		
		S	T	H	S	T	H	S	T	H	S	T	H	S	T	H	S	T	H	S	T	H
DAN Long et al. (2015)	39.2	-	39.1	-	-	11.4	-	-	33.3	-	-	16.2	-	-	42.1	-	-	29.7	-	-	28.6	-
DCTN Xu et al. (2018)	39.2	-	48.6	-	-	23.4	-	-	48.8	-	-	7.2	-	-	53.5	-	-	47.3	-	-	38.1	-
MCD Saito et al. (2018)	39.2	-	54.3	-	-	22.2	-	-	45.7	-	-	7.6	-	-	58.4	-	-	43.5	-	-	38.6	-
M3SDA Peng et al. (2019)	39.2	-	58.6	-	-	26.0	-	-	52.3	-	-	6.3	-	-	62.7	-	-	49.5	-	-	42.5	-
DECISION + KD Hinton et al. (2015)	7.8	25.3	61.1	35.8	31.5	21.0	25.2	32.0	53.4	40.0	2.4	18.2	4.2	29.8	67.1	41.3	29.0	50.4	36.8	25.0	45.2	30.6
DECISION + SepRep-Net	7.8	33.1	62.0	43.2	34.9	22.3	27.2	37.4	55.0	44.5	3.3	18.5	5.6	33.4	68.1	44.8	33.6	51.8	40.8	<b>29.3</b>	<b>46.3</b>	<b>35.9</b>
DECISION Ahmed et al. (2021)	39.2	31.4	61.5	41.6	34.3	21.6	26.5	36.2	54.6	43.5	3.1	18.9	5.3	32.4	67.5	43.8	32.8	51.0	39.9	28.4	45.9	35.1
CAiDA + KD Hinton et al. (2015)	7.8	25.1	61.3	35.6	29.8	21.8	25.2	32.2	54.6	40.5	2.5	19.1	4.4	27.6	67.8	39.2	30.0	50.9	37.8	24.5	45.9	32.0
CAiDA + SepRep-Net	7.8	32.5	62.5	42.8	33.8	22.8	27.2	36.5	55.6	44.1	3.1	20.1	5.4	32.5	68.5	44.1	34.0	51.7	41.0	<b>28.7</b>	<b>46.9</b>	<b>35.6</b>
CAiDA Dong et al. (2021)	39.2	29.8	61.9	40.2	32.8	22.2	26.5	34.6	55.0	42.5	3.0	19.3	5.2	29.3	68.1	41.0	31.3	51.2	38.8	26.8	46.3	33.9
SHOT-ens + KD Hinton et al. (2015)	7.8	35.3	58.2	43.9	34.6	24.9	29.0	37.4	54.8	44.5	2.2	15.4	3.9	28.9	70.0	40.9	31.7	52.1	39.4	28.4	45.9	35.1
SHOT + SepRep-Net	7.8	38.6	59.3	46.8	40.0	26.3	31.7	40.2	56.9	47.1	3.7	14.9	5.9	38.6	71.9	50.2	39.0	53.6	45.1	<b>33.3</b>	<b>47.2</b>	<b>39.0</b>
SHOT-ens Liang et al. (2020)	39.2	40.6	58.6	48.0	37.2	25.2	30.0	41.1	55.3	47.2	3.0	15.3	5.0	31.6	70.5	43.6	35.3	52.4	42.2	31.5	46.2	37.5

tational costs. For the vanilla methods, our method still surpasses them consistently, with only 20% of the original computation costs.

### 4.3 ANALYSES

**Ablation Study** We investigate the components in SepRep-Net: Separation (Sep), Reparameterization (Rep), and Importance Reweighting (ReW). We mention that when taking separate pathways alone, we adopt traditional knowledge distillation on the model with multiple pathways, which makes the model size equal to our method for a fair comparison. H-score results are reported in Table 6(a), justifying that the effect of each part is indispensable.

**Comparison of Different Uncertainty Criteria** In our method, we adopt entropy as the uncertainty-based criterion in our re-weighting strategy. There also exist various uncertainty metrics, such as confidence or margin. Here we give the comparison results of utilizing these different metrics to re-weight the multiple classifiers. Table 6(b) presents that our method performs stably with different uncertainty metrics, while the entropy metric enjoys the highest performance.

**Importance Reweighting** We analyze the importance weights for ensembling outputs when the model is evaluated on different domains respectively. As shown in Figure 3, our importance weight has an obviously positive correlation with the accuracy of each classifier on the unlabeled target domain. Moreover, our reweighting strategy is actually adapted to the input data during inference, leading to improved generalizability.



Table 6: **Ablation Study** of different designs in SepRep-Net and different uncertainty metrics for importance reweighting on Office-Home with ResNet-50 backbone. We take SHOT as the base method. H-score is used as the evaluation metric here. The meaning of  $\rightarrow$  is the same as Table 4. (Sep: Separation, Rep: Reparameterization, ReW: Importance Reweighting)

	FLOPs	$\rightarrow$ Ar	$\rightarrow$ Cl	$\rightarrow$ Pr	$\rightarrow$ Rw	Avg
+ Sep	12.3	73.4	65.0	77.5	77.9	73.5
+ Sep (+ KD)	4.1	70.2	62.3	74.1	76.5	70.8
+ Sep + Rep	4.1	73.4	65.0	77.5	77.9	73.5
+ Sep + Rep + ReW (ours)	4.1	<b>77.1</b>	<b>66.4</b>	<b>79.0</b>	<b>80.5</b>	<b>76.1</b>

	$\rightarrow$ Ar	$\rightarrow$ Cl	$\rightarrow$ Pr	$\rightarrow$ Rw	Avg
Confidence	75.4	66.0	78.3	78.9	74.7
Margin	76.5	65.8	78.5	79.9	75.2
<b>Entropy (ours)</b>	<b>77.1</b>	<b>66.4</b>	<b>79.0</b>	<b>80.5</b>	<b>76.1</b>

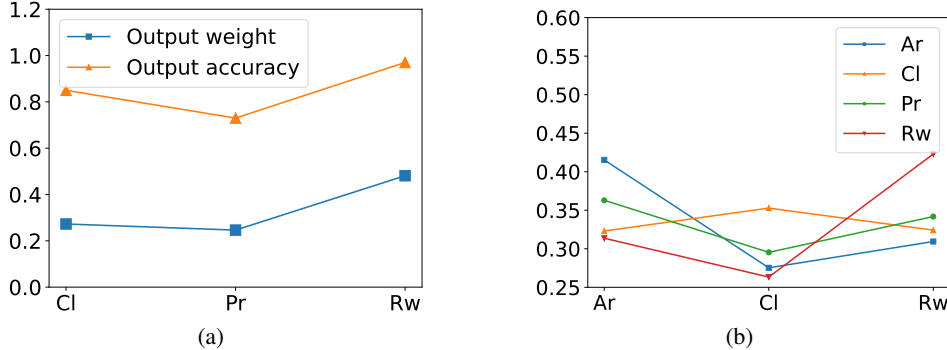


Figure 3: **Output Weights** for different domains in Office-Home with ResNet-50 backbone when applying SepRep-Net to SHOT. (a) The output weights calculated in our framework are positively correlated to the classifier accuracy. (b) Our importance weights are adaptive to the input data. Specifically, the importance weight of the input domain adaptively becomes larger than others, enhancing generalizability.

**More Analysis on Source Accuracy** The model accuracy on source domains is often taken as the metric to evaluate the model generalizability. For source accuracy, the ensemble results of multiple source models no doubt enjoys a strong performance. Therefore, it is interesting to compare our method with it. From Table 7, we can observe that ensembling multiple source models (Source-ens) shows strong performance in source domains, but poor accuracy in the target domain. The source accuracy gap between our method and Source-ens is much smaller than the traditional knowledge distillation, with similar computational costs. On the other hand, H-score validates that our method keeps the best trade-off between effectiveness and generalizability among these methods.

Table 7: Source Accuracy (S) (%), Target Accuracy (T) (%), H-Score (H), and Model Efficiency on Office-Home dataset. ResNet-50 is adopted in experiments. Ar, Cl, Pr, Rw indicate different domains (Ar: Art, Cl: Clipart, Pr: Product, Rw: Real World). Source-ens indicates the performance of model ensemble with all source models. The abbreviations in METHOD are the same as Table 2.

METHOD	FLOPS	Cl,Pr,Rw $\rightarrow$ Ar			Ar,Pr,Rw $\rightarrow$ Cl			Ar,Cl,Rw $\rightarrow$ Pr			Ar,Cl,Pr $\rightarrow$ Rw			S	Avg T	H
		S	T	H	S	T	H	S	T	H	S	T	H			
Source-ens He et al. (2016)	12.3	90.6	58.4	71.0	91.2	43.0	58.4	88.5	67.7	76.7	89.1	70.8	78.9	<b>89.9</b>	60.0	72.0
SHOF-ens + KD Hinton et al. (2015)	4.1	69.0	72.1	70.5	65.6	59.0	62.1	67.9	82.7	74.6	66.5	82.9	73.8	67.3	74.2	70.6
SHOT + SepRep-Net	4.1	82.6	72.3	77.1	75.6	59.2	66.4	75.1	83.3	79.0	77.8	83.3	80.5	77.8	<b>74.5</b>	<b>76.1</b>

## 5 CONCLUSION

In this paper, we target on reassembling multiple existing models into a single model and adapting it to a novel target domain without accessing source data. Towards this problem, we propose SepRep-Net, a framework that forms multiple separate pathways during training and further merges them via reparameterization to facilitate inference. As a general approach, SepRep-Net is easy to be plugged into various methods. Extensive experiments prove that SepRep-Net consistently improves existing methods in effectiveness, efficiency, and generalizability.

## 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 *IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
- Fabio Maria Carlucci, Lorenzo Porzi, Barbara Caputo, Elisa Ricci, and Samuel Rota Buló. Autodial: Automatic domain alignment layers. In *IEEE International Conference on Computer Vision*, pp. 5077–5085. IEEE, 2017.
- Jiahua Dong, Zhen Fang, Anjin Liu, Gan Sun, and Tongliang Liu. Confident anchor-induced multi-source free domain adaptation. In *Advances in Neural Information Processing Systems*, volume 34, 2021.
- Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(1):2096–2030, 2016.
- Sha Fei Gong Boqing, Shi Yuan and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, 2014.
- Arthur Gretton, Dino Sejdinovic, Heiko Strathmann, Sivaraman Balakrishnan, Massimiliano Pontil, Kenji Fukumizu, and Bharath K. Sriperumbudur. Optimal kernel choice for large-scale two-sample tests. In *Advances in Neural Information Processing Systems*, 2012.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.
- Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei A. Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International Conference on Machine Learning*, 2018.
- Ying Jin, Ximei Wang, Mingsheng Long, and Jianmin Wang. Minimum class confusion for versatile domain adaptation. In *European Conference on Computer Vision*, pp. 464–480. Springer, 2020.
- Guoliang Kang, Lu Jiang, Yi Yang, and Alexander Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- Jogendra Nath Kundu, Naveen Venkat, R Venkatesh Babu, et al. Universal source-free domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4544–4553, 2020a.
- Jogendra Nath Kundu, Naveen Venkat, Ambareesh Revanur, R Venkatesh Babu, et al. Towards inheritable models for open-set domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12376–12385, 2020b.
- Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9641–9650, 2020.
- Yanhao Li, Naiyan Wang, Jianping Shi, Xiaodi Hou, and Jiaying Liu. Adaptive batch normalization for practical domain adaptation. *Pattern Recognition*, 80:109–117, 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, 2020.
- Jian Liang, Dapeng Hu, Yunbo Wang, Ran He, and Jiashi Feng. Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I. Jordan. Learning transferable features with deep adaptation networks. In *International Conference on Machine Learning*, 2015.
- Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I. Jordan. Deep transfer learning with joint adaptation networks. In *International Conference on Machine Learning*, 2017.
- Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I. Jordan. Conditional adversarial domain adaptation. In *Advances in Neural Information Processing Systems*, 2018.
- Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
- Van-Anh Nguyen, Tuan Nguyen, Trung Le, Quan Hung Tran, and Dinh Phung. Stem: An approach to multi-source domain adaptation with guarantees. In *IEEE International Conference on Computer Vision*, pp. 9352–9363, October 2021.
- Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.
- Yingwei Pan, Ting Yao, Yehao Li, Yu Wang, Chong-Wah Ngo, and Tao Mei. Transferrable prototypical networks for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. In *AAAI Conference on Artificial Intelligence*, 2018.
- Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *IEEE International Conference on Computer Vision*, pp. 1406–1415, 2019.
- Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *European Conference on Computer Vision*, 2010.
- Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. Universal domain adaptation through self supervision. In *Advances in Neural Information Processing Systems*, 2020.
- Rui Shu, Hung H. Bui, Hirokazu Narui, and Stefano Ermon. A DIRT-T Approach to Unsupervised Domain Adaptation. In *International Conference on Learning Representations*, 2018.
- Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European Conference on Computer Vision*, 2016.
- Shiliang Sun, Honglei Shi, and Yuanbin Wu. A survey of multi-source domain adaptation. *Information Fusion*, 24:84–92, 2015.
- Hui Tang, Ke Chen, and Kui Jia. Unsupervised domain adaptation via structurally regularized deep clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8725–8735, 2020.
- Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. Deep domain confusion: Maximizing for domain invariance. *CoRR*, abs/1412.3474, 2014.

- Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep Hashing Network for Unsupervised Domain Adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- Hang Wang, Minghao Xu, Bingbing Ni, and Wenjun Zhang. Learning to combine: Knowledge aggregation for multi-source domain adaptation. In *European Conference on Computer Vision*, pp. 727–744. Springer, 2020a.
- Xiaofang Wang, Dan Kondratyuk, Kris M. Kitani, Yair Movshovitz-Attias, and Elad Eban. Multiple networks are more efficient than one: Fast and accurate models via ensembles and cascades. *CoRR*, abs/2012.01988, 2020b. URL <https://arxiv.org/abs/2012.01988>.
- Ximei Wang, Ying Jin, Mingsheng Long, Jianmin Wang, and Michael I Jordan. Transferable normalization: Towards improving transferability of deep neural networks. In *Advances in Neural Information Processing Systems*, pp. 1953–1963, 2019.
- Ruijia Xu, Ziliang Chen, Wangmeng Zuo, Junjie Yan, and Liang Lin. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- Ruijia Xu, Guanbin Li, Jihan Yang, and Liang Lin. Larger norm more transferable: An adaptive feature norm approach. In *IEEE International Conference on Computer Vision*, 2019.
- 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, 2020a.
- Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Unsupervised domain adaptation without source data by casting a bait. *arXiv preprint arXiv:2010.12427*, 2020b.
- Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Generalized source-free domain adaptation. In *IEEE International Conference on Computer Vision*, pp. 8978–8987, 2021.
- Yabin Zhang, Hui Tang, Kui Jia, and Mingkui Tan. Domain-symmetric networks for adversarial domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- Yang Zou, Zhiding Yu, B. Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *European Conference on Computer Vision*, 2018.
- Yang Zou, Zhiding Yu, Xiaofeng Liu, B.V.K. Vijaya Kumar, and Jinsong Wang. Confidence regularized self-training. In *IEEE International Conference on Computer Vision*, 2019.