Distilling and Refining Domain-Specific Knowledge for Semi-Supervised Domain Adaptation

Ju Hyun Kim[1]
kjhyun18@dgu.ac.kr
Ba Hung Ngo[1]
ngohung@dgu.ac.kr
Jae Hyeon Park[1]
pjh0011@dongguk.edu
Jung Eun Kwon[1]
kje_9912@dgu.ac.kr
Ho Sub Lee[2]
hslee34@daegu.ac.kr
Sung In Cho[1]
csi2267@dongguk.edu

[1] Department of Multimedia Engineering
Dongguk University
Seoul, Korea

[2] Department of Electronic Engineering
Daegu University
Gyeongsan, Korea

Abstract

We propose a novel framework, ***D****istilling* ***A****nd* ***R****efining domain-specific* ***K****nowledge* (DARK), for Semi-supervised Domain Adaptation (SSDA) tasks. The proposed method consists of three strategies: *Multi-view Learning*, *Distilling*, and *Refining*. In *Multi-view Learning*, to acquire domain-specific knowledge, DARK trains a shared generator and two domain-specific classifiers using the labeled source and target data. Then, in *Distilling*, two classifiers exchange the domain-specific knowledge with each other to exploit a cross-view consistency regularization using soft labels between differently augmented unlabeled target samples. During this, DARK leverages information from low-confidence unlabeled target samples in addition to the high-confidence unlabeled target samples. To prevent a trivial collapse problem caused by the low-confidence samples, we propose the utilization of a sample-wise dynamic weight based on prediction reliability (SDWR). Finally, in *Refining*, for class alignment, class confusion of the unlabeled target data is minimized considering the model maturity. Simultaneously, to maintain model consistency between the predictions of differently augmented unlabeled target samples, a bridging loss with SDWR is used. Consequently, the experimental results on the SSDA datasets demonstrate that DARK outperforms the state-of-the-art benchmark methods for SSDA tasks. The code can be found at https://github.com/Juh-yun/DARK.

1 Introduction

Deep neural networks (DNNs) are widely used in various real-world applications of computer vision tasks such as image classification [7, 17], semantic segmentation [3, 4], and

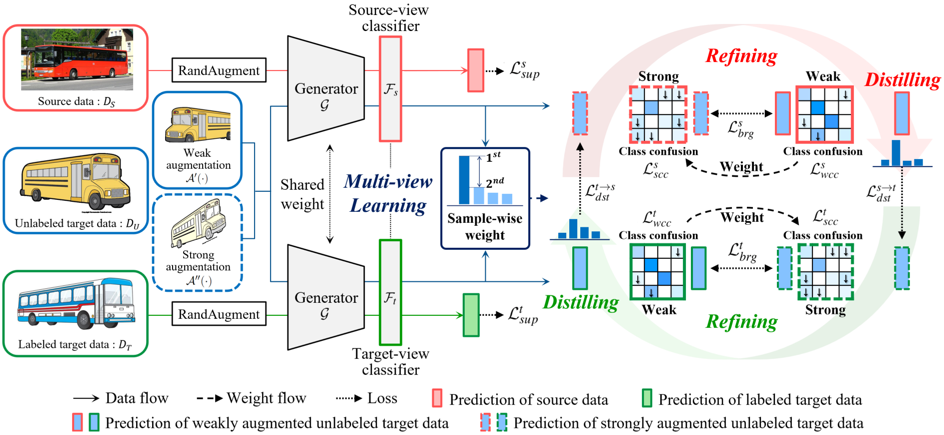

Figure 1: Illustration of the proposed method.

object detection [16, 29] with outstanding performance. However, DNN often suffers from a generalization ability problem that causes a performance degradation when there is a large discrepancy (domain shift) between the distributions of training and testing sets. Domain Adaptation (DA) is widely used to solve this problem. Previously, Unsupervised Domain Adaptation (UDA) methods [6, 8, 14, 22, 34] that assume the labeled source domain and unlabeled target domain are more actively developed. However, in recent works, Semi-supervised Domain Adaptation (SSDA) approaches [9, 13, 20, 28, 30, 32, 38] using partially labeled target samples have been actively researched. Specifically, in industrial applications, there are many cases where the performance could be boosted by utilizing a small set of label information of the target domain.

The most important factors that determine the performance of SSDA for image classification can be described as follows: 1) Efficiency of using partially labeled target samples and 2) quality of domain alignment between the source and target domains.

Concerning the first factor, when the model is trained with the mixed labeled samples from source and target domains, a data imbalance problem in which the labeled source samples dominate the training can occur. Recently, the multi-view approaches [23, 28, 38] make different training pipelines of source and target data to solve the above problem. They divide training scenarios for the labeled source and target data to capture the unique information of each domain. Yang *et al.* [38] proposes a representative method of the multi-view approach. It is composed of two models, each consisting of a generator and a classifier. It unifies Semi-supervised Learning (SSL) and UDA models into a framework using co-training [2], respectively. However, it requires a high computational complexity, and the pre-training step is inevitable for the training to converge.

Concerning the second factor, the quality of domain alignment is important for performance improvement in SSDA as well as DA. Specifically, inter-domain alignment to reduce divergence between two domains and intra-domain alignment to create class-wise separable distribution should be considered together. For the inter-domain alignment, DA methods mainly use cross-domain alignment approaches [6, 18, 34]. However, if only the inter-domain alignment is considered without intra-domain alignment, the performance cannot increase successfully since low-density distributions for each class in the target domain can be generated. Therefore, various DA methods focus on the intra-domain alignment by using a categorical prototype matching [9, 24] and a pair-wise class prediction alignment [8, 15] in addition to the inter-domain alignment. Kim and Kim [9] performs the intra-domain alignment by reducing the distance between class-wise prototypes and their nearby unlabeled target samples. However, since only unlabeled target samples adjacent to the prototypes from the source domain are aligned, the class-wise alignment cannot be achieved for all unlabeled samples.

Consequently, 1) a simple but efficient framework for easy training convergence should be considered when designing the multi-view setting with a few labeled target samples, and 2) effective inter-and intra-class alignment should be considered together.

In this paper, we propose a novel approach, called ***D**istilling **A**nd **R**efining domain-specific **K**nowledge* (DARK), to address the problems mentioned above for SSDA tasks as shown in Figure 1. To summarize, our contributions are as follows:

- We propose DARK, a novel approach to SSDA tasks. In our method, unique domain specific-knowledge is distilled reciprocally through cross-view consistency regularization using a soft pseudo-labeling and is refined to enhance the class alignment. When using the soft pseudo-labeling, to maximally utilize the available information, low-confidence samples are actively used, unlike existing techniques.

- We design a sample-wise dynamic weight based on prediction reliability (SDWR) to reduce the negative effect of the information from the unlabeled samples having low confidence scores when using the pseudo labelling.

- We evaluate DARK over DA datasets and compare it with SSDA benchmark methods. To the best of our knowledge, DARK achieves state-of-the-art results for all datasets.

2 Related Work

2.1 Semi-supervised Domain Adaptation

In Semi-supervised Domain Adaptation (SSDA), Yao *et al.* [39] and Ao *et al.* [1] that are pioneer SSDA methods are derived from Unsupervised Domain Adaptation (UDA). They only focus on reducing the domain shift like the UDA methods, so labeled target data is not effectively utilized. Saito *et al.* [30] addresses the problem of abovementioned methods, and subsequent SSDA approaches [9, 13, 20, 23, 28, 32, 38] design the model considering the intra-domain adaptation as well as the inter-domain adaptation using the labeled target data effectively.

Saito *et al.* [30] aligns the prototypes of labeled data and the unlabeled target samples through an adversarial entropy minimax strategy to reduce the domain divergence between the two domains. Kim and Kim [9] utilizes maximum mean discrepancy (MMD) [18] to minimize the inter-domain discrepancy and reduces the distance between the categorical prototypes and the nearby unlabeled target samples for the intra-domain alignment. Li *et al.* [13] proposes adversarial adaptive clustering considering the top 5 class predictions of samples to reduce intra-class variance and increase inter-class variance for the intra-domain alignment. Singh [32] uses class-wise contrastive learning across domains and target sample-level contrastive alignment strategy. Qin *et al.* [28] performs source expansion and clustering of target samples to make the target distribution being fitted within the source distribution.

Mishra *et al.* [20] proposes a pre-training step that aligns features to make class cluster easier on a domain adaptation step, where target features are generalized through consistency regularization [33] using the pseudo label [11]. Ngo *et al.* [23] composes the inter-and intra-domain-view classifiers and employs collaborative learning with hard pseudo labels for the domain alignment. Yang *et al.* [38] decomposes SSDA with SSL and UDA to leverage the information of labeled data in each domain and finds an optimal decision boundary of the classification for unlabeled target data using the co-training [2] and Mixup [40].

2.2 Low-confidence Unlabeled Samples

There are various strategies [12, 19, 36, 37] that are used for low-confidence unlabeled samples in tasks using unlabeled samples. They apply a separate strategy for the low-confidence samples or adopt a novel pseudo-labeling method and class-wise dynamic weight.

Wang *et al.* [36] classifies reliable and unreliable pixels by the predefined ratio and employs different strategies to successfully utilize many unreliable pixels for semi-supervised semantic segmentation. Li *et al.* [12] divides groups using the adaptive confidence margin proposed in the semi-supervised deep facial expression recognition task. It applies contrastive learning for the low-confidence groups and the pseudo labeling for the high-confidence group. Mei *et al.* [19] is one of the unsupervised domain adaptation methods for semantic segmentation applying the instance adaptive threshold. This approach removes the noise of pseudo labels by dynamically reducing the proportion of hard class samples with low confidence. Xu *et al.* [37] focuses on domain adaptive object detection with severe class imbalance problem. It assigns large weight to unreliable class samples to solve the class imbalance problem. All of these methods show performance improvements by dealing with low-confidence samples in their specific tasks. Thus, we design the model to focus on exchanging domain-specific knowledge between two classifiers without information loss by using not only high-confidence but also low-confidence unlabeled samples.

2.3 Consistency Regularization

Consistency regularization is an effective technique that is widely used in semi-supervised learning [12, 31, 33]. This approach keeps the consistency of predictions between different views of the same data to encourage a perturbation-invariant model. Sohn *et al.* [33] proposes consistency regularization method with differently augmented images in semi-supervised learning and achieves prominent performance. Since then, various DA methods [20, 23, 25, 28] utilizing unlabeled data employ this consistency regularization technique.

3 Proposed Method

In SSDA, we are given a set of labeled source samples $D_S = \{\mathbf{x}_S^i, \mathbf{y}_S^i\}_{i=1}^{N_S}$, a set of labeled target samples $D_T = \{\mathbf{x}_T^i, \mathbf{y}_T^i\}_{i=1}^{N_T}$, and a set of unlabeled target $D_U = \{\mathbf{x}_U^i\}_{i=1}^{N_U}$, where $\mathbf{x}^i$ and $\mathbf{y}^i$ are an image sample of each dataset and its corresponding one-hot label vector, respectively. For the label vector $\mathbf{y}^i$ of each set, $(\mathbf{y}^i)_k$ is a k-th element of the label vector, where $k \in \{1, 2, \ldots, K\}$ is an index of K classes. N_S, N_T, and N_U are the number of samples in D_S, D_T, and D_U, respectively. The goal of the proposed method is to design a novel model that maximizes prediction accuracy on the D_U using D_S, D_T, and D_U.

The proposed method mainly consists of *Multi-view Learning* for supervision, *Distilling* for inter-and intra-domain alignment, and *Refining* for evolutional intra-domain alignment as shown in Figure 1. First, in *Multi-view Learning*, we train a generator ($\mathcal{G}$) and two domain-specific classifiers ($\mathcal{F}_s$, $\mathcal{F}_t$) with standard cross-entropy ($\mathcal{L}^s_{sup}$, $\mathcal{L}^t_{sup}$) using D_S and D_T. Then, in *Distilling*, we use a cross-view consistency regularization ($\mathcal{L}_{dst}$) using D_U with different augmentations to distill the knowledge between the two classifiers. Finally, in *Refining*, we minimize class confusion ($\mathcal{L}_{wcc}, \mathcal{L}_{scc}$) with the bridging loss ($\mathcal{L}_{brg}$). We try to relieve a negative effect of uncertainty from large perturbations when the class confusion of strongly augmented samples is minimized during *Refining*. To cope with the negative effect, we set weights derived from the class confusion level of the weakly perturbated samples to reduce the class confusion of large perturbated samples. In addition, we employ sample-wise dynamic weights based on prediction reliability (SDWR) to alleviate the negative effect of the low-confidence unlabeled target samples.

3.1 Multi-view Learning for Supervision

To prevent a bias problem from imbalanced data in the training and to extract the domain-specific knowledge from the labeled data, we separate the training pipeline of D_S and D_T into the source-and target-view classifiers, respectively, as shown in Figure 1. We apply the data augmentation techniques of RandAugment [5] to the labeled source and target data. In training, the shared generator $\mathcal{G}$, the source-view classifier $\mathcal{F}_s$, and the target-view classifier $\mathcal{F}_t$ are trained using the standard cross-entropy loss:

$$\begin{aligned} \mathcal{L}^s_{sup}(\mathbf{x}^i_S, \mathbf{y}^i_S) &= -\sum_{k=1}^{K} \left(\mathbf{y}^i_S\right)_k \log\left(\sigma(\mathcal{F}_s(\mathcal{G}(\mathbf{x}^i_S)))_k\right), \\ \mathcal{L}^t_{sup}(\mathbf{x}^i_T, \mathbf{y}^i_T) &= -\sum_{k=1}^{K} \left(\mathbf{y}^i_T\right)_k \log\left(\sigma(\mathcal{F}_t(\mathcal{G}(\mathbf{x}^i_T)))_k\right), \end{aligned} \tag{1}$$

where σ is a SoftMax function. By this, the two classifiers can obtain the domain-specific class knowledge.

3.2 Distilling Strategy for Inter-and Intra-Domain Alignment

After *Multi-view Learning*, $\mathcal{F}_s$ has the class information of labeled source data while $\mathcal{F}_t$ has the poor class information from partially labeled target data. To compensate for the shortage of $\mathcal{F}_t$ and take advantage of their strengths, the classifiers exchange each domain-specific knowledge using a collaborative learning approach [23]. For weak augmentation $\mathcal{A}'(\cdot)$, horizontal flipping and cropping are used randomly. For the strong augmentation $\mathcal{A}''(\cdot)$, we employ the RandAugment [5]. Weakly and strongly augmented unlabeled target samples are predicted through the source-view ($\mathcal{F}_s$) and target-view ($\mathcal{F}_t$) classifiers, and the corresponding prediction vectors are as follows:

$$\mathbf{p}'_{s,t}(\mathbf{x}^i_U) = \sigma(\mathcal{F}_{s,t}(\mathcal{G}(\mathcal{A}'(\mathbf{x}^i_U)))),\ \mathbf{p}''_{s,t}(\mathbf{x}^i_U) = \sigma(\mathcal{F}_{s,t}(\mathcal{G}(\mathcal{A}''(\mathbf{x}^i_U)))). \tag{2}$$

Not only to transfer the knowledge but also to encourage model invariance and consistency, we minimize the cross-entropy loss between the pseudo label generated over the $\mathcal{A}'(\mathbf{x}^i_U)$ from each classifier and the $\mathbf{p}''(\mathbf{x}^i_U)$ from the other classifier for cross-view consistency regularization. For the pseudo labeling, we use $\mathbf{p}'_s(\mathbf{x}^i_U)$ and $\mathbf{p}'_t(\mathbf{x}^i_U)$ as soft label vectors. If a one-hot hard pseudo label is used as in the previous works [13, 23, 38], the influence of incorrect class information from samples having low confidence could be minimized. However, this hard label could lead to confirmation bias because of the missing inter-class information and information from the low-confidence samples. To deal with these, we use all of the $\mathbf{p}'(\mathbf{x}_U)$ as the soft labels to deliver the inter-and intra-class information. In this approach, however, low-confidence samples can contaminate the model. To alleviate this trivial collapse problem, we use the sample-wise dynamic weight that will be explained in Section 3.4. Then, we apply label smoothing (LS) [21] to the soft label for regularization by adding a constant prediction value (α/K) and rescaling the result. LS mitigates negative effects caused by uncertain soft pseudo-labels of an overconfident model in the early training phase. Consequently, cross-view consistency regularization loss of *Distilling* can employ domain-balanced knowledge by exchanging domain-specific knowledge as follows:

$$\begin{aligned}
\mathcal{L}_{dst}^{s\rightarrow t}\left(\mathbf{x}_U^i\right) &= -\sum_{k=1}^{K}\frac{1}{1+\alpha}\left(\mathbf{p}_s'(\mathbf{x}_U^i)_k+\frac{\alpha}{K}\right)\log\left(\mathbf{p}_t''(\mathbf{x}_U^i)_k\right),\\
\mathcal{L}_{dst}^{t\rightarrow s}\left(\mathbf{x}_U^i\right) &= -\sum_{k=1}^{K}\frac{1}{1+\alpha}\left(\mathbf{p}_t'(\mathbf{x}_U^i)_k+\frac{\alpha}{K}\right)\log\left(\mathbf{p}_s''(\mathbf{x}_U^i)_k\right),
\end{aligned}\tag{3}$$

where α is the label smoothing parameter, and empirically set to 0.1.

3.3 Refining Strategy for Intra-Domain Alignment

Through *Distilling*, the model can have high class confusion by large intra-class variance and small inter-class variance due to the soft label-based knowledge exchange. So, we propose *Refining* to perform evolutional intra-domain alignment through a class-wise arrangement. Jin *et al.* [8] find distinguishable features between confused classes by minimizing a pair-wise class confusion. Inspired by this, we use the class confusion loss to minimize categorical confusion for unlabeled target samples as follows:

$$\mathcal{L}_{wcc}=\frac{1}{K}\sum_{k=1}^{K}\sum_{\acute{k}\neq k}^{K}C'_{k\acute{k}},\ \mathcal{L}_{scc}=\frac{1}{K}\sum_{k=1}^{K}\sum_{\acute{k}\neq k}^{K}C''_{k\acute{k}},\tag{4}$$

where $\mathcal{L}_{wcc}$ and $\mathcal{L}_{scc}$ are the class confusion losses of a batch of the $\mathcal{A}'(\mathbf{x}_U)$ and $\mathcal{A}''(\mathbf{x}_U)$, respectively. $C'_{k\acute{k}}$ and $C''_{k\acute{k}}$ are normalized class correlations of softened probabilities using temperature scaling between two classes k and $\acute{k}$ in each batch of the $\mathcal{A}'(\mathbf{x}_U)$ and the $\mathcal{A}''(\mathbf{x}_U)$, respectively. The temperature scaling factor for softened probability is set to 2.5 as in [8]. Since the original class confusion loss is designed for a vanilla unlabeled sample, there is a following problem to apply it on the $\mathcal{A}''(\mathbf{x}_U)$. If the model tries to minimize the class confusion of $\mathcal{A}''(\mathbf{x}_U)$, the negative effect of dark knowledge can occur because strong augmentation may cause noise in the prediction of the model. To alleviate this, we make a dynamic weight for the $\mathcal{L}_{scc}$ using the $\mathcal{L}_{wcc}$ as follows:

$$\lambda_{scc}^{s,t}=\begin{cases}\exp\left(-3\cdot\mathcal{L}_{wcc}^{s,t}\right) & \mathcal{L}_{wcc}^{s,t}\leq 1,\\ 0 & \mathcal{L}_{wcc}^{s,t}>1.\end{cases}\tag{5}$$

As the $\mathcal{L}_{wcc}$ decreases, that is, when the class confusion of the $\mathcal{A}'(\mathbf{x}_U)$ is sufficiently lowered, the $\mathcal{L}_{scc}$ is progressively activated by the $\lambda_{scc}^{s,t}$. By this loss, $\mathbf{p}'_{s,t}(\mathbf{x}_U)$ and $\mathbf{p}''_{s,t}(\mathbf{x}_U)$ can

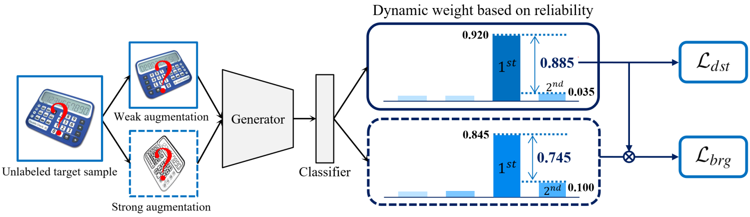

Figure 2: Illustration of SDWR for $\mathcal{L}_{dst}$ and $\mathcal{L}_{brg}$.

be varied during training independently. To maintain the consistency of these, we use the bridging loss as follows:

$$\mathcal{L}^{s,t}_{brg}(\mathbf{x}^i_U) = \left\|\mathbf{p}'_{s,t}(\mathbf{x}^i_U) - \mathbf{p}''_{s,t}(\mathbf{x}^i_U)\right\|^2. \tag{6}$$

Finally, we organize the overall loss of the *Refining* strategy $\mathcal{L}^{s,t}_{ref}$ as follows:

$$\mathcal{L}^{s,t}_{ref} = \mathcal{L}^{s,t}_{wcc} + \lambda^{s,t}_{scc} \cdot \mathcal{L}^{s,t}_{scc} + \mathcal{L}^{s,t}_{brg}. \tag{7}$$

3.4 Sample-wise Dynamic Weights Determination

As mentioned, we propose a novel sample-wise dynamic weight based on prediction reliability (SDWR) for the loss functions of unlabeled samples.

Information from low-confidence samples. In *Distilling*, we focus on the utilization of the information from low-confidence unlabeled target samples mentioned in Section 3.2. To reduce an effect of the noise caused by uncertain (low-confidence) samples, we set the sample-wise weights using the first and second largest prediction values of classes as follows:

$$\begin{aligned} w'^{i}_{s,t} &= \mathrm{topk}_{[\mathrm{k}=1]}\left(\mathbf{p}'_{s,t}(\mathbf{x}^i_U)\right) - \mathrm{topk}_{[\mathrm{k}=2]}\left(\mathbf{p}'_{s,t}(\mathbf{x}^i_U)\right), \\ w''^{i}_{s,t} &= \mathrm{topk}_{[\mathrm{k}=1]}\left(\mathbf{p}''_{s,t}(\mathbf{x}^i_U)\right) - \mathrm{topk}_{[\mathrm{k}=2]}\left(\mathbf{p}''_{s,t}(\mathbf{x}^i_U)\right), \end{aligned} \tag{8}$$

where $\mathrm{topk}(\cdot)$ is the top k-th of confidence score of the input sample. The model finds the prediction reliability of the samples by itself and generates the weights during the training. When using only the $\mathrm{topk}_{[\mathrm{k}=1]}(\cdot)$ as a confidence level for SDWR, the model becomes overconfidence in the early training phase because it cannot consider the probability of other classes. So, we also consider the $\mathrm{topk}_{[\mathrm{k}=2]}(\cdot)$ to design the quantitative sample-wise dynamic weight. The $w'^{i}_{s,t}$ and $w''^{i}_{s,t}$ are multiplied to the $\mathcal{L}^{s,t}_{dst}$ in *Distilling* to consider the reliability of the soft label for the knowledge exchange, which will be described in Section 3.5.

Progressive weight based on the model performance. To apply the bridging loss considering the convergence status of the model accurately, a product of SDWRs of the $\mathcal{A}'(\mathbf{x}_U)$ and $\mathcal{A}''(\mathbf{x}_U)$ is used to maintain consistency when each prediction is reliable as shown in Figure 2. We rewrite the $\mathcal{L}^{s,t}_{brg}$ with SDWR as follows:

$$\mathcal{L}^{s,t}_{brg}(\mathbf{x}^i_U, w'^{i}_{s,t}, w''^{i}_{s,t}) = w'^{i}_{s,t} \cdot w''^{i}_{s,t} \cdot \left\|\mathbf{p}'_{s,t}(\mathbf{x}^i_U) - \mathbf{p}''_{s,t}(\mathbf{x}^i_U)\right\|^2. \tag{9}$$

3.5 Overall Loss and Inference

The overall loss for training the proposed method can be formulated as follows:

$$\begin{aligned}\mathcal{L}_s((\mathbf{x}_S^i,\mathbf{y}_S^i),\mathbf{x}_U^i) &= \mathcal{L}_{sup}^s(\mathbf{x}_S^i,\mathbf{y}_S^i) + w'^i_s \cdot \mathcal{L}_{dst}^{s\to t}(\mathbf{x}_U^i) + \mathcal{L}_{ref}^s(\mathbf{x}_U^i, w'^i_s, w''^i_s),\\ \mathcal{L}_t((\mathbf{x}_T^i,\mathbf{y}_T^i),\mathbf{x}_U^i) &= \mathcal{L}_{sup}^t(\mathbf{x}_T^i,\mathbf{y}_T^i) + w'^i_t \cdot \mathcal{L}_{dst}^{t\to s}(\mathbf{x}_U^i) + \mathcal{L}_{ref}^t(\mathbf{x}_U^i, w'^i_t, w''^i_t).\end{aligned} \tag{10}$$

As mentioned in Section 3.4, SDWR is applied to $\mathcal{L}_{dst}$ and $\mathcal{L}_{brg}$ for each sample. We sequentially update the gradients of the source-and target-views using $\mathcal{L}_s$ and $\mathcal{L}_t$, respectively. In the inference stage, the probability of the unlabeled target sample is predicted by taking an averaged output prediction from the two classifiers, $\mathcal{F}_s$ and $\mathcal{F}_t$.

4 Experimental Results

In this section, we evaluate the performance of the proposed method on SSDA tasks using well-known DA datasets [27, 35]. As in the previous study [30], we conduct experiments using one or three labeled target samples for each class.

4.1 Setups

Datasets. We evaluate the performance on DomainNet [27] and Office-Home [35] datasets. DomainNet includes 345 classes and six domains. For comparison with the benchmark methods, we use 126 classes and four domains of *Real* (R), *Painting* (P), *Clipart* (C), and *Sketch* (S) following Saito *et al.* [30], using seven DA scenarios. Office-Home includes 65 classes and four domains: *Real* (R), *Product* (P), *Clipart* (C), and *Art* (A) for 12 scenarios.

Implementation details. We use Pytorch [26] for implementation. The backbone network is ResNet-34 [7] which is pre-trained on ImageNet dataset [10] following Saito *et al.* [30]. For the model optimization, we use SGD with momentum of 0.9, an initial learning rate of 0.001, and the same learning rate scheduler with Saito *et al.* [30]. We train our model for 50K/10K iterations on DomainNet/Office-Home as in Yang *et al.* [38].

Benchmarks. We compare DARK to state-of-the-art (SOTA) SSDA methods, MME [30], APE [9], PAC [20], CDAC [13], ASDA [28], CLDA [32], and DECOTA [38].

4.2 Comparisons with State-of-the-art Methods

DomainNet. Table 1 shows classification accuracies of DARK and benchmark methods for DomainNet. DARK obtains SOTA performance by showing 3.1% and 2.4% higher classification accuracies than ASDA [28] that provides the best performance in benchmark methods, in average accuracy in one-and three-shot settings. For the performance of each scenario, DARK shows outstanding classification accuracies compared to all benchmark methods except for the R-C scenario of DECOTA [38].

Office-Home. Table 2 shows classification accuracies of DARK and benchmark methods for Office-Home. DARK outperforms all benchmark methods on the average accuracy. In the one-shot setting, the performance is improved by 0.9% compared to CDAC [13], which achieves the highest average accuracy among benchmarks, and by 0.2% compared to DECOTA [38], which is the current SOTA in the three-shot setting.

Method	R → C		R → P		P → C		C → S		S → P		R → S		P → R		Mean	
	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot
MME [30]	70.0	72.1	67.7	69.7	69.0	71.7	56.3	61.8	64.8	66.8	61.0	61.9	76.1	78.5	66.4	68.9
APE [9]	70.4	76.6	70.8	72.1	72.9	76.7	56.7	63.1	64.5	66.1	63.0	67.8	76.6	79.4	67.8	71.7
PAC [20]	74.9	78.6	73.0	74.3	72.6	76.0	65.8	69.6	67.9	69.4	68.7	70.2	76.7	79.3	71.4	73.9
CDAC [13]	77.4	79.6	74.2	75.1	75.5	79.3	67.6	69.9	71.0	73.4	69.2	72.5	80.4	81.9	73.6	76.0
ASDA [28]	77.0	79.4	75.4	76.7	75.5	78.3	66.5	70.2	72.1	74.2	70.9	72.1	79.7	82.3	73.9	76.2
CLDA [32]	76.1	77.7	75.1	75.7	71.0	76.4	63.7	69.7	70.2	73.7	67.1	71.1	80.1	82.9	71.9	75.3
DECOTA [38]	**79.1**	**80.4**	74.9	75.2	76.9	78.7	65.1	68.6	72.0	72.7	69.7	71.9	79.6	81.5	73.9	75.6
DARK (ours)	78.3	79.4	**77.9**	**78.6**	**79.1**	**81.0**	**71.8**	**74.8**	**75.1**	**77.4**	**72.5**	**73.8**	**84.4**	**85.4**	**77.0**	**78.6**

Table 1: Quantitative results (%) on DomainNet. The best accuracy is indicated in bold.

# Shot	Method	R → C	R → P	R → A	P → R	P → C	P → A	A → P	A → C	A → R	C → R	C → A	C → P	Mean
1-shot	MME [30]	61.9	82.8	71.2	79.2	57.4	64.7	75.5	59.6	77.8	74.8	65.7	74.5	70.4
	APE [9]	60.7	81.6	72.5	78.6	58.3	63.6	76.1	53.9	75.2	72.3	63.6	69.8	68.9
	CLDA [32]	60.2	**83.2**	72.6	81.0	55.9	66.2	76.1	56.3	79.3	76.3	66.3	73.9	70.6
	CDAC [13]	61.9	83.1	72.7	80.0	**59.3**	64.6	75.9	**61.2**	78.5	75.3	64.5	75.1	71.0
	DARK (ours)	**62.0**	83.1	**74.8**	**81.5**	57.0	**67.4**	**77.1**	57.3	**80.5**	**77.1**	**68.2**	**76.5**	**71.9**
3-shot	MME[30]	64.6	85.5	71.3	80.1	64.6	65.5	79.0	63.6	79.7	76.6	67.2	79.3	73.1
	APE [9]	66.4	86.2	73.4	82.0	65.2	66.1	81.1	63.9	80.2	76.8	66.6	79.9	74.0
	CLDA [32]	66.0	87.6	**76.7**	82.2	63.9	**72.4**	81.4	63.4	81.3	80.3	70.5	80.9	75.5
	CDAC [13]	67.8	85.6	72.2	81.9	67.0	67.5	80.3	**65.9**	80.6	80.2	67.4	81.4	74.2
	DECOTA [38]	**70.4**	**87.7**	74.0	82.1	**68.0**	69.9	81.8	64.0	80.5	79.0	68.0	**83.2**	75.7
	DARK (ours)	66.0	87.4	76.0	**82.9**	65.1	71.2	**82.5**	64.0	**82.0**	**81.5**	**70.8**	82.0	**75.9**

Table 2: Quantitative results (%) on Office-Home. The best accuracy is indicated in bold.

4.3 Ablation Studies

We perform ablation studies and evaluate DARK on DomainNet using ResNet-34 under the three-shot setting. We evaluate the contribution to the performance depending on each proposed strategy. In addition, we provide detailed analysis to evaluate the effectiveness of the components.

Effectiveness of *Multi-view Learning*. We compare the performance of *Multi-view Learning* and a single-view strategy where a single classifier is trained on mixed labeled data. To focus on the evaluation of the domain-specific knowledge exchange of *Multi-view Learning*, we use only *Distilling* and the bridging loss of *Refining* except for the class confusion loss. For the single-view, *Distilling* of unlabeled target data is replaced by the consistency regularization loss between the soft labels of weakly augmented samples and the predictions of strongly augmented samples. As shown in Table 3, *Multi-view Learning* outperforms the performance by up to 5.5% in an average of four scenarios compared with the single-view learning. This result shows that *Multi-view Learning* captures domain-balanced class knowledge for domain alignment and alleviates the data imbalance problem.

Synergy of *Distilling* and *Refining*. Table 4 implies the effectiveness of *Distilling* and *Refining* for unlabeled target data. In the case of using only *Distilling*, we use the bridging loss together to ensure consistent predictions in each view. First, without *Refining*, *Distilling* shows satisfactory performance in DomainNet for the three-shot setting, with an average performance difference of 0.5% compared to the current SOTA, ASDA [28]. When both *Refining* and *Distilling* are conducted simultaneously, the performance is greatly improved for all scenarios, and classification accuracy is increased by 2.9% on average.

Method	Classifier	P → C	C → S	S → P	R → S	Mean
Single-view	Single-view	71.9	65.8	70.4	62.5	67.6
Multi-view	Source-view	**77.4**	**71.2**	73.8	**70.0**	**73.1**
	Target-view	77.4	71.0	73.8	69.7	73.0
	Ensemble	**77.4**	**71.2**	**73.9**	**70.0**	**73.1**

Table 3: Ablation study for *Multi-view Learning* of the proposed method. We report the classification accuracy (%) on DomainNet of four scenarios for the three-shot setting.

Distilling	*Refining*	R → C	R → P	P → C	C → S	S → P	R → S	P → R	Mean
✗	✗	61.9	65.1	62.0	58.1	62.4	56.0	75.2	63.0
✓	✗	76.5	76.8	77.4	71.2	73.9	70.0	83.8	75.7
✓	✓	**79.4**	**78.6**	**81.0**	**74.8**	**77.4**	**73.8**	**85.4**	**78.6**

Table 4: Ablation study for *Distilling* and *Refining*. We report the classification accuracy (%) on DomainNet of all scenarios for the three-shot setting.

$\mathcal{L}_{brg}$	$\mathcal{L}_{wcc}$	λ_{scc}	$\mathcal{L}_{scc}$	P → C	C → S	S → P
✗	✓	✗	✗	79.1	72.1	75.8
✗	✓	✗	✓	79.2	72.1	75.9
✓	✓	✗	✗	80.5	74.0	77.0
✓	✓	✗	✓	80.7	74.4	77.1
✓	✓	✓	✓	**81.0**	**74.8**	**77.4**

Table 5: Ablation study of components in *Refining*. We report the classification accuracy (%) on DomainNet of three scenarios for the three-shot setting.

Ablation study for the *Refining* strategy. We conduct experiments to prove the contribution of each of the four components in *Refining* for the performance: the bridging loss $\mathcal{L}_{brg}$, the class confusion loss of the weak augmentation $\mathcal{L}_{wcc}$, the dynamic weight λ_{scc}, and the class confusion loss of the strong augmentation $\mathcal{L}_{scc}$ in Table 5. When λ_{scc} is not used, a constant weight (0.3) is assigned to $\mathcal{L}_{scc}$. When all components are used, it provides the highest performances because the negative effect of $\mathcal{L}_{scc}$ is alleviated. The classification accuracies of this case are 1.9%, 2.7%, and 1.6% higher than when only $\mathcal{L}_{wcc}$ is used in the three scenarios. When the model is trained using $\mathcal{L}_{scc}$ and $\mathcal{L}_{wcc}$, the performance is slightly increased. Also, combining $\mathcal{L}_{wcc}$ and $\mathcal{L}_{brg}$ can boost the performance because weakly augmented samples help strongly augmented samples with aligning class distribution.

5 Conclusion

We introduce DARK for SSDA tasks with distilling and refining domain-specific knowledge strategy. Our method leverages the domain-dominant class knowledge from multi-view to acquire the domain-balanced features by the knowledge transfer. In this process, we utilize the inter-and intra-class information and the information from low-confidence samples using the soft label and the prediction reliability-aware weight to maximize usable information of unlabeled target information. Then, the domain-balanced knowledge is refined to minimize the class confusion of unlabeled target data. We prove that the components of our method are necessary to improve the performance of each other and that our approach is more effective compared with other benchmark methods. Although we design the dynamic weight to minimize the noise of the low-confidence samples to exploit the information, it could not completely reduce the negative effect on the low-confidence samples in some cases. In future work, we will study to minimize the negative effect by using different strategies depending on their reliability so that a model can exploit the information from all unlabeled target samples.

6 Acknowledgments

This research was supported by Culture, Sports and Tourism RD Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2022 (Project Name: System Enhancement and Core Module Development for Implementing Position-based Metaverse Service Platform, Project Number: R2022010122, Contribution Rate: 50%) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1C1C1009662)

References

[1] Shuang Ao, Xiang Li, and Charles Ling. Fast generalized distillation for semi-supervised domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[2] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 92–100, 1998.

[3] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[4] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2613–2622, 2021.

[5] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020.

[6] 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.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. cvpr. 2016. *arXiv preprint arXiv:1512.03385*, 2016.

[8] Ying Jin, Ximei Wang, Mingsheng Long, and Jianmin Wang. Minimum class confusion for versatile domain adaptation. In *European Conference on Computer Vision*, pages 464–480. Springer, 2020.

[9] Taekyung Kim and Changick Kim. Attract, perturb, and explore: Learning a feature alignment network for semi-supervised domain adaptation. In *European conference on computer vision*, pages 591–607. Springer, 2020.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[11] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896, 2013.

[12] Hangyu Li, Nannan Wang, Xi Yang, Xiaoyu Wang, and Xinbo Gao. Towards semi-supervised deep facial expression recognition with an adaptive confidence margin. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4166–4175, 2022.

[13] Jichang Li, Guanbin Li, Yemin Shi, and Yizhou Yu. Cross-domain adaptive clustering for semi-supervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2505–2514, 2021.

[14] Jingjing Li, Erpeng Chen, Zhengming Ding, Lei Zhu, Ke Lu, and Heng Tao Shen. Maximum density divergence for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 43(11):3918–3930, 2020.

[15] Shuang Li, Mixue Xie, Fangrui Lv, Chi Harold Liu, Jian Liang, Chen Qin, and Wei Li. Semantic concentration for domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9102–9111, 2021.

[16] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[17] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986, 2022.

[18] Mingsheng Long, Jianmin Wang, Guiguang Ding, Jiaguang Sun, and Philip S Yu. Transfer feature learning with joint distribution adaptation. In *Proceedings of the IEEE international conference on computer vision*, pages 2200–2207, 2013.

[19] Ke Mei, Chuang Zhu, Jiaqi Zou, and Shanghang Zhang. Instance adaptive self-training for unsupervised domain adaptation. In *European conference on computer vision*, pages 415–430. Springer, 2020.

[20] Samarth Mishra, Kate Saenko, and Venkatesh Saligrama. Surprisingly simple semi-supervised domain adaptation with pretraining and consistency. *arXiv preprint arXiv:2101.12727*, 2021.

[21] Rafael Müller, Simon Kornblith, and Geoffrey E Hinton. When does label smoothing help? *Advances in neural information processing systems*, 32, 2019.

[22] Jaemin Na, Heechul Jung, Hyung Jin Chang, and Wonjun Hwang. Fixbi: Bridging domain spaces for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1094–1103, 2021.

[23] Ba Hung Ngo, Ju Hyun Kim, Yeon Jeong Chae, and Sung In Cho. Multi-view collaborative learning for semi-supervised domain adaptation. *IEEE Access*, 9:166488–166501, 2021.

[24] Ba Hung Ngo, Jae Hyeon Park, So Jeong Park, and Sung In Cho. Semi-supervised domain adaptation using explicit class-wise matching for domain-invariant and class-discriminative feature learning. *IEEE Access*, 9:128467–128480, 2021.

[25] Ba Hung Ngo, Ju Hyun Kim, So Jeong Park, and Sung In Cho. Collaboration between multiple experts for knowledge adaptation on multiple remote sensing sources. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022.

[26] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[27] 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*, pages 1406–1415, 2019.

[28] Can Qin, Lichen Wang, Qianqian Ma, Yu Yin, Huan Wang, and Yun Fu. Semi-supervised domain adaptive structure learning. *arXiv preprint arXiv:2112.06161*, 2021.

[29] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[30] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell, and Kate Saenko. Semi-supervised domain adaptation via minimax entropy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8050–8058, 2019.

[31] Kuniaki Saito, Donghyun Kim, and Kate Saenko. Openmatch: Open-set semi-supervised learning with open-set consistency regularization. *Advances in Neural Information Processing Systems*, 34:25956–25967, 2021.

[32] Ankit Singh. Clda: Contrastive learning for semi-supervised domain adaptation. *Advances in Neural Information Processing Systems*, 34:5089–5101, 2021.

[33] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020.

[34] 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*, pages 7167–7176, 2017.

[35] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5018–5027, 2017.

[36] Yuchao Wang, Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Guoqiang Jin, Liwei Wu, Rui Zhao, and Xinyi Le. Semi-supervised semantic segmentation using unreliable pseudo-labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4248–4257, 2022.

[37] Minghao Xu, Hang Wang, Bingbing Ni, Qi Tian, and Wenjun Zhang. Cross-domain detection via graph-induced prototype alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12355–12364, 2020.

[38] Luyu Yang, Yan Wang, Mingfei Gao, Abhinav Shrivastava, Kilian Q Weinberger, Wei-Lun Chao, and Ser-Nam Lim. Deep co-training with task decomposition for semi-supervised domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8906–8916, 2021.

[39] Ting Yao, Yingwei Pan, Chong-Wah Ngo, Houqiang Li, and Tao Mei. Semi-supervised domain adaptation with subspace learning for visual recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2142–2150, 2015.

[40] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.