WAMDA: Weighted Alignment of Sources for Multi-source Domain Adaptation

Surbhi Aggarwal surbhia@iisc.ac.in Jogendra Nath Kundu jogendrak@iisc.ac.in R. Venkatesh Babu venky@iisc.ac.in Anirban Chakraborty anirban@iisc.ac.in Department of Computational and Data Sciences, Indian Institute of Science Bangalore, India

Abstract

Unsupervised Domain Adaptation aims to learn a model for an unlabelled *target* domain, given access to a single labelled but differently distributed *source* domain. However, often multiple labelled sources which share complementary information are present, resulting in the more practical problem of *multi-source domain adaptation* (MSDA). Recent works in MSDA learn a domain-invariant space from the sources and target. However, they treat each source to be equally relevant to the target and are not sensitive towards the intrinsic statistical similarities amongst domains. In this work, we propose a novel method for MSDA, termed *WAMDA*, which utilizes the multiple sources based on their relative importance to the target. Our aim is to explore the relevance of each sourcetarget alignment and source-source alignment, and then perform weighted alignment of domains by using the relevance scores. We experimentally validate the performance of our proposed method on multiple datasets, and achieve either state-of-the-art results or competitive performances across all these datasets.

1 Introduction

Conventional supervised learning methods assume that the test data follows the same distribution as the training data. However, in practice, such an assumption is not valid as the models are often tested on related but differently distributed datasets. Due to such *domain shifts* [2], the performance of the pre-trained models suffer on unseen target distributions. Since data annotation is costly and time consuming, obtaining labelled target data is not always feasible. *Unsupervised Domain Adaptation* [1] addresses such scenarios by transferring the class-discriminative knowledge from labelled sources to the unlabelled target. Based upon the number of sources involved, unsupervised domain adaptation is classified into: *Single-Source Domain Adaptation* (SDA) [1, 12] and *Multi-Source Domain Adaptation* (MSDA) [16, 12].

In this work, we focus on the problem of MSDA. MSDA is a more realistic problem than SDA, as in practice, labelled data collected from multiple, diverse sources is often available.

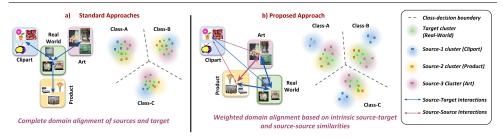


Figure 1: In the above example, the target domain, *Real-World*, is similar to sources *Art* and *Product*, while it is dissimilar from *Clipart*. **a**) Prior works [**1**, **2**] perform complete alignment of sources and target. **b**) In contrast, we propose to calculate *source-target* and *source-source* relevance (Section 3.1.2) and then use them to do weighted alignment of domains (Section 3.2). Here, the length of arrows indicate the degree of alignment between domains, such that shorter length implies better alignment.

However, MSDA is also more challenging because in single source setting, domain shift exists only between the source and the target, whereas in multi-source setting, there are varying levels of domain shifts between each source-target pair as well as between each source-source pair [1]. Over the past few years, deep learning based works have brought advances in MSDA [13, 26, 23]. However, most of these methods suffer from at least one of the following issues. a) Firstly, some works focus on learning a domain-invariant feature space from the sources and target (Fig. 1(a)) [\Box , \Box]. These methods treat all sources to be equally relevant and perform complete alignment of the target with each source. Such an alignment strategy does not regard the *intrinsic source-target correlations* and makes the target vulnerable to negative transfer from uncorrelated sources [1]. Some other methods explore source relevance [11, 12], but use these scores only to infer the target's class by aggregation of the outputs of source classifiers, and not to drive the source-target alignment. Hence, the first issue is that the source-target alignment is not sensitive to source relevance. b) Secondly, limited works probe the effect of source-source relevance [1]. Though M³SDA [D] performed source-source alignment, they did not use source-source relevance to guide it. c) Thirdly, although creating domain-invariant features enhances the transferability to the target, it can also make the feature space less discriminative [23]. Transferability refers to the ability of a feature space to be meaningful across domains, while discriminability denotes the degree of inter-class separation and within-class similarity [3]. Hence, the need is to preserve the discriminative information of sources, while encouraging domain-invariance.

Unlike prior methods that tackle either of these issues individually, we propose to address all drawbacks in a unified manner (Fig. 1(b)). First, we explore the *relevance of each source*, and harness upon the more relevant sources to transfer the discriminative information to the target. We also regard the *source-source relations* by gauging the importance of aligning each pair of sources. Such alignment of source pairs merges the important sources and minimizes interaction with any source that is uncorrelated to target. Finally, to make the source features both transferable and discriminative, we learn an *intermediate feature space* for each source. These spaces are designed such that the distributions of relevant sources align with each other. This allows positive aggregation of discriminative information into the intermediate spaces of the relevant sources. Thus, the separate intermediate spaces preserve the discriminability of sources, while the weighted alignment of the intermediate spaces enhances the transferability to target. Our contributions are, therefore, summarized as follows:

- We propose a novel method that performs weighted alignment of domains and enhances transfer of discriminative knowledge from the relevant sources to the target. We calculate *source relevance* and *source-source relevance* to guide the degree of alignment of each source-target pair and source-source pair respectively (Section 3.1.2).
- To enhance the discriminability and transferability of the relevant sources, a separate intermediate space is learnt per source (Section 3.2). The intermediate spaces of relevant sources interact to aggregate the discriminative information present across them.
- Our target inference method is much simpler than that of prior works, which often utilize complex rules to aggregate predictions from source classifiers (Section 3.2.3).
- Extensive experiments on benchmark datasets show the effectiveness of our approach, as we achieve either state-of-the-art results or competitive performances across all of them (Section 4).

2 Related Work

Single-Source Domain Adaptation. Several single-source DA methods aim at aligning source and target by optimizing various measures of domain divergence such as MMD [\square], KL-divergence [\square], and Wasserstein distance [\square]. Ganin *et al.* [\square] used adversarial-training to obtain discriminative but domain-invariant features. GAN-based algorithms [\square] are also shown to be effective for SDA.

Multi-Source Domain Adaptation. Early works on MSDA, such as [2, 12, 16], provided a theoretical analysis of the problem. Alongside the theoretical works, shallow learning based techniques for MSDA were also proposed [2, 23]. Recently, deep-learning based works have been proposed for MSDA [11, 12]. Deep Cocktail Networks (DCTN) [23] is based on the popular view that the target can be represented as a mixture of source distributions. DCTN proposes a two-stage optimization process, such that the first stage involves adversarial adaptation for achieving domain-invariance, and the second stage uses pseudo-labels to make the target features discriminative. Finally, the final target predictor is obtained by a weighted combination of source-specific predictors, such that the weight of the source is proportional to the domain-confusion loss for that source. In [11], Guo et al. propose to use Mahalanobis distance between the target instance and the mean source representation to obtain weight for that source. Recently, Peng et al. [1] proposed M³SDA, where a joint feature extractor is trained and the marginal features distributions are aligned by applying MMD [9] on each pair of domains. Two variants of the algorithm were proposed for imparting discriminative properties in the representations. In the first variant, M³SDA, a single classifier per source is learnt, while in the second variant, M^3 SDA- β , two classifiers are involved for each source and they are trained using the *maximum classifier discrepancy method* [1]. The predictions for the target is obtained through weighted average of classifiers per source, such that the weights are derived from the accuracy of the classifiers on labelled source data.

3 Proposed Approach

We present a two-stage algorithm (Fig. 2), such that in the first stage of *pre-adaptation training*, we obtain the relevance scores and learn the discriminative knowledge inherently present in each source. In the second stage of *multi-source adaptation training*, we perform the weighted alignment of domains and learn a classifier over this weighted aligned space.

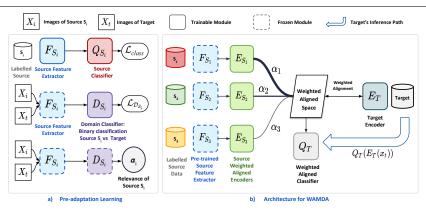


Figure 2: Overview of WAMDA. We propose a two-stage algorithm. a) In the first stage, we perform source-specific discriminative learning and obtain the *source relevance* and *source-source relevance* scores. b) In the second stage, we perform weighted alignment between domains and transfer discriminative information from the relevant sources to target.

3.1 Pre-Adaptation Training

The objective of the pre-adaptation training to learn the source-specific discriminative knowledge, and to set the ground for computing the *source relevance* and *source-source relevance*.

3.1.1 Source-Specific Discriminative Learning

During this step, the discriminative knowledge intrinsically present in each source is learnt. For each source S_i , a discriminative **source feature extractor** (F_{S_i}) and **source classifier** (Q_{S_i}) are trained using its labelled data through cross-entropy loss. We refer to the features space obtained by passing source data through F_{S_i} as *source feature space*. This space is rich in the source-specific discriminative information as it is obtained solely from the source data, without any interaction with other domains. Hence, the source feature extractor and classifier act as reservoirs of discriminative information, which shall be utilized during adaptation.

3.1.2 Learning Relevance Weightings

A source is termed *relevant*, if it has high statistical similarity with the target in the instance space, and hence the labelled information of the source is significantly transferable to the target. Inspired by the concept of \mathcal{H} -divergence and *domain confusion* [\square , \square], for each source S_i , we utilize a **domain classifier**, D_{S_i} to gauge *source relevance*. The task for domain classifier D_{S_i} is to discriminate between samples from S_i and target. The rationale for connecting domain classifier with source relevance is that high confusion of the domain classifier implies less domain shift between source and target, and consequently higher source relevance. When comparing target with S_i , the target features are obtained by passing target samples through the pre-trained F_{S_i} . When a large domain gap is present between source and target, the distribution induced by passing target through F_{S_i} shall be significantly different from source feature space. Therefore, the ease of a domain classifier D_{S_i} in discriminating between the source and target features is indicative of the source-target similarity and consequently, of source relevance. Each D_{S_i} is trained as follows by standard binary cross-entropy loss, where X_i denotes the set of images from S_i and X_t denotes the set of target images:

$$\mathcal{L}_{D_{S_i}} = -\frac{1}{|X_i|} \sum_{x_i \in X_i} \log D_{S_i}(F_{S_i}(x_i)) - \frac{1}{|X_t|} \sum_{x_t \in X_t} \log(1 - D_{S_i}(F_{S_i}(x_t)))$$
(1)

Algorithm 1 WAMDA: Proposed adaptation algorithm

Input: Unlabelled target images X_t ; Labelled source images $\{X_{S_i}, Y_{S_i}\}_{i=1}^{K}$; Pre-trained source feature extractors, category classifiers and domain classifiers $\{F_{S_i}, Q_{S_i}, D_{S_i}\}_{i=1}^{K}$ **Output:** Trained target encoder E_T , classifier Q_T

- 1: Create separate optimizer for each loss: { $\mathcal{L}_{qt}, \mathcal{L}_{align}, \mathcal{L}_{de}, \mathcal{L}_{T \rightarrow W}$ }
- 2: while current iteration $t \leq maxIter$ do
- 3: Sample mini-batch from $\{X_{S_i}, Y_{S_i}\}_{i=1}^K$ and X_t
- 4: **Calculate batch-wise relevance scores:** Calculate $\{\alpha_i\}_{i=1}^K, \{\beta_{i,j}\}_{i=1, j \neq i}^K$
- 5: Learn Weighted Aligned Space: Update $Q_T, E_T, \{E_{S_i}\}_{i=1}^K$ by alternate minimization of $\{\mathcal{L}_{qt}, \mathcal{L}_{align}\}$ using loss-specific optimizers.
- 6: **Target-Specific Learning:** Update E_T by alternate minimization of { $\mathcal{L}_{de}, \mathcal{L}_T \rightarrow W$ } using loss-specific optimizers.

7: end while

8: return E_T, Q_T

Once the domain classifiers are learnt, relevance scores are obtained as follows:

Source Relevance. For each source S_i , a source relevance score (α_i) is assigned, such that the score is proportional to the similarity of the target domain to S_i . To quantify the source relevance, we gauge the confusion of the trained domain classifier in discriminating between source and target features. If the domain classifier can confidently discriminate, then there would be high discrepancy between its predictions for source and target. Similarly, if the domain classifier is confused, its outputs would be similar across the two domains. Hence, for each S_i , we measure the confusion of the classifier by the difference in the predictions for source and target, $h_i = \left|\frac{\sum_{x_i \in X_i} D_{S_i}(F_{S_i}(x_i))}{|X_i|} - \frac{\sum_{x_t \in X_t} D_{S_i}(F_{S_i}(x_t))}{|X_i|}\right|$. These scores are then normalized to obtain the source relevance, $\alpha_i = \frac{\exp(-Kh_i)}{\sum_j \exp(-Kh_j)}$, where *K* is the number of sources. Source-Source Relevance. Having observed that the sources are of different relevance, a consequence of this is that alignment of each source-pair is not equally important. This is because, if all source-pairs are equally aligned, it effectively merges all the sources uniformly and no weighted alignment between source-target pairs can be done. Hence, the importance of aligning each pair of sources is gauged. We further observe that for any pair of sources, even if one of the sources involved in the pair is highly dissimilar to the target, then aligning that source pair is not crucial. Hence, we quantify the relevance of a source-pair (S_i, S_j) as $\beta_{ij} = \min(\alpha_i, \alpha_j)$. This permits knowledge transfer between sources which are mutually important for the target, and limits negative transfer from less relevant sources to the target.

3.2 Adaptation Algorithm

By virtue of our pre-adaptation training, the source-specific discriminative information is captured in the source feature extractors and classifiers. During adaptation, we learn *intermediate spaces* for each source, such that the spaces of relevant sources align well. This facilitates the aggregation of the discriminative knowledge into the intermediate spaces of the relevant sources, while limiting interactions with any uncorrelated source. Further, based on the relevance of S_i , the target is weighted aligned with the intermediate spaces of each S_i . Moreover, to efficiently perform the task of target classification, discriminative target features are learnt by transfer of knowledge from labelled sources. To achieve these objectives, for each S_i , **source weighted aligned encoder** (E_{S_i}) is introduced, which learns the intermediate space for S_i . The target features are learnt by a **target encoder** (E_T), such that its feature space is more aligned with the intermediate space of the relevant sources. The space so learnt, where the target is more aligned with the relevant sources and the sources which are mutually relevant are well aligned, is termed as **weighted aligned space**. To ensure that the discriminability of this space, a **weighted aligned classifier** (Q_T) is introduced, which is trained by interactions with the intermediate spaces of the relevant sources.

3.2.1 Learning Weighted Alignment of Domains

We firstly discuss the method used to ensure that the desired alignment of domains is achieved. **A. Learning Weighted Aligned Classifier.** The main goal of adaptation is to obtain a discriminative classifier which has good generalizability on the target. To achieve this, we build a weighted aligned classifier (Q_T) that is trained using the labelled sources which are highly similar to the target. This ensures that the decision boundaries of Q_T accommodate well the relevant sources, and thereby, allows Q_T to be more generalizable to the target. Furthermore, this classifier ensures that the intermediate space of each source remains discriminative, and the weighted interaction with a common classifier enhances the knowledge transfer amongst the relevant sources. Given labelled images for K sources $\{(X_i, Y_i)\}_{i=1}^K, Q_T$ is trained by weighting the per-source cross-entropy loss (\mathcal{L}_{ce}) by the source importance, as follows:

$$\mathcal{L}_{qt} = \sum_{i=1}^{K} \frac{\alpha_i}{|X_i|} \sum_{(x,y)\in(X_i,Y_i)} \mathcal{L}_{ce}(Q_T(E_{S_i}(F_{S_i}(x))), y)$$
(2)

B. Weighted Alignment of Sources and Target. To learn the weighted alignment of domains, domain matching is performed between source-target pairs and source-source pairs by using the source relevance and source-source relevance weights. Such an alignment strategy encourages the target to be aligned more with the relevant sources, and minimizes negative transfer from uncorrelated sources. For performing the domain alignment, we use the popular DA loss, Deep CORAL (\mathcal{L}_{coral}) [22]. For each S_i , given a batch of source images X_i , we obtain the source activations from E_{S_i} and denote the matrix of activations by $\mathbf{E_i}$. Similarly, given a batch of target images X_t , we encode them by E_T and denote the matrix of activations by $\mathbf{E_t}$. Using these notations, the weighted domain alignment loss is as follows:

$$\mathcal{L}_{align} = \sum_{i} \alpha_{i} \mathcal{L}_{coral}(\mathbf{E}_{i}, \mathbf{E}_{t}) + \sum_{i, j \neq i} \frac{\beta_{ij}}{(K-1)} \mathcal{L}_{coral}(\mathbf{E}_{i}, \mathbf{E}_{j})$$
(3)

3.2.2 Target-Specific Learning

Having performed the domain alignment of the target with sources, we introduce losses designed to make the target encoder E_T discriminative. In this part, for fine-grained knowledge transfer to target, for each target datapoint x_t , we use a per-instance weight $w_i(x_t)$ to capture the importance of S_i for x_t , defined as $w_i(x_t) = \frac{\alpha_i D_{S_i}(F_{S_i}(x_t))}{\sum_j \alpha_j D_{S_j}(F_{S_j}(x_t))}$. Basically, along with the prior of source relevance, the instance-wise score of $D_{S_i}(F_{S_i}(x_t))$, which denotes the probability that x_t belongs to S_i , is used to get these fine-grained weights.

A. Distillation and Entropy Minimization. During the initial stages of training, discriminative information is contained only in the pre-trained source-specific modules $\{F_{S_i}, Q_{S_i}\}_{i=1}^{K}$. Thus, an initial discriminative guidance can be obtained for each target image x_t by the weighted aggregation of the outputs of source-specific classifiers (Fig.3). Inspired by *distribution weighted combining rule* [IIG], for each x_t , we aggregate the source-wise class predictions to obtain a *pseudo-softmax* denoted by $\Phi(x_t) = \sum_i w_i(x_t) Q_{S_i}(F_{S_i}(x_t))$. To distill the

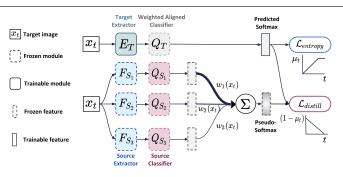


Figure 3: Weighted combination of source-classifiers is used to distill discriminative knowledge into E_T . $\mathcal{L}_{entropy}$ ensures that target falls in confident regions of Q_T . Contribution of $\mathcal{L}_{entropy}$ is increased over iterations, and that of $\mathcal{L}_{distill}$ is decreased.

discriminative signals from the relevant sources into E_T , we introduce a *distillation loss* (Eq. 4) which minimizes the difference between predictions of Q_T and the pseudo-softmax:

$$\mathcal{L}_{distill} = \|Q_T(E_T(x_t)) - \Phi(x_t)\|_1 \tag{4}$$

However, for successful adaptation to happen, Q_T should be a better predictor for target than an aggregation of pre-trained source classifiers. This means, as training progresses, the decision boundaries of Q_T should become better suited for the target. Therefore, inspired by works such as [**A**], an entropy minimization constraint ($\mathcal{L}_{entropy}$) is introduced (Eq. 5), to ensure that the target samples fall into the high-confidence regions of the classifier. Since in the beginning, Q_T is random, entropy minimization can hamper the learning. Hence, we gradually increase the weight for $\mathcal{L}_{entropy}$ as training progresses, and keep decreasing the weight of $\mathcal{L}_{distill}$, to ensure that $\Phi(x_t)$ does not act as a bottleneck. At any iteration *t*, the weight of $\mathcal{L}_{distill}$, denoted by $\mu_t \in [0,1]$, is formulated as $\mu_t = min(1,mt)$, where *m* is a hyperparameter. The combined loss, termed *distill-entropy loss* (\mathcal{L}_{de}), at any iteration *t* is formulated as follows, where H(.) denotes the entropy function:

$$\mathcal{L}_{entropv} = H(Q_T(E_T(x_t))) \tag{5}$$

$$\mathcal{L}_{de} = (1 - \mu_t) \,\mathcal{L}_{distill} + \mu_t \,\mathcal{L}_{entropy} \tag{6}$$

B. Guidance from Source Weighted Aligned Spaces. In this loss, we guide the output of E_T by referring to the source weighted aligned encoders $\{E_{S_i}\}_{i=1}^K$. This is because the encoders of the relevant sources are discriminative and have high transferability to target. Since the feature obtained from the relevant sources shall be more informative, for each target instance x_t , we minimize the distance between $E_{S_i}(x_t)$ and $E_T(x_t)$, weighted by the relevance of S_i . This loss, which guides the target's projection onto the weighted aligned space, is referred to as *target to weighted aligned space loss* $(\mathcal{L}_T \to W)$, and is defined as follows:

$$\mathcal{L}_{T \to W} = \sum_{i=1}^{K} w_i(x_t) \| E_T(x_t) - E_{S_i}(F_{S_i}(x_t)) \|_2^2$$
(7)

3.2.3 Target Predictions

Post training, final predictions for any target sample x_t are obtained by $Q_T(E_T(x_t))$. This prediction rule is much simpler than the aggregation rules in prior literature [12, 26].

Method	Acc. (%)
$\mathcal{L}_{qt} + \mathcal{L}_{T \to W}$	79.3
$\mathcal{L}_{qt} + \mathcal{L}_{T \to W} + \mathcal{L}_{align}$	80.0
$\mathcal{L}_{qt} + \mathcal{L}_{T \to W} + \mathcal{L}_{align} + \mathcal{L}_{distill}$	81.1
$\mathcal{L}_{at}^{T} + \mathcal{L}_{T \to W} + \mathcal{L}_{align} + \mathcal{L}_{de}$	82.3

Table 1: Loss-wise ablations on *Office-Home* on $ACP \rightarrow R$.

Alignment	$ACP \rightarrow R$	$ACR \rightarrow P$
Uniform	81.7	82.4
Weighted	82.3	84.1

Table 2: Effect of weighted

alignment on accuracy (%)

of Office-Home.

 Values
 ACR \rightarrow P

 K=1
 83.1

 K=3
 84.1

 K=5
 84.0

Table 3: Effect of K used in α_i on accuracy (%) of *Office-Home*.

4 Experiments and Analysis

4.1 Experimental Setup

A. Datasets. The experiments are conducted on standard unsupervised visual DA benchmarks: Office-31, Office-Caltech and Office-Home. *Office-31* [13] is a popular DA dataset, consisting of 31 object classes in 3 distinct domains: Amazon (A), DSLR (D) and Webcam (W). *Office-Caltech* [2] comprises of 10 classes captured in 4 domains: Amazon (A), Caltech (C), DSLR (D) and Webcam (W). The third benchmark, *Office-Home* [23], is a more challenging one as the domain shifts are much severe in it. It contains 65 categories and 4 diverse domains: Art (A), Clip Art (C), Product (P) and Real World (R). We have also presented the analysis of our method on the latest benchmark, *DomainNet* [13], in the supplementary.

B. Implementation Details. For each source S_i , F_{S_i} comprises of ImageNet pre-trained ResNet-50 backbones [III] followed by four ELU-activated linear layers. Similarly, E_{S_i} consists of four linear layers with ELU activations. The target encoder E_T comprises of a pre-trained ResNet-50 backbone followed by ELU-activated linear layers. For each adaptation loss $\{\mathcal{L}_{qt}, \mathcal{L}_{align}, \mathcal{L}_{de}, \mathcal{L}_T \rightarrow W\}$, we create a separate Adam optimizer with a learning rate of 1e-4 and optimize them as specified in Algorithm 1. In the pre-adaptation stage (Fig. 2), the validation set of source is used for the model selection of F_{S_i} and Q_{S_i} . In the adaptation stage, we follow the standard model selection technique [III, III] to obtain the final model of E_T and Q_T . More implementation details are presented in the supplementary.

C. Baselines. We follow the conventional baselines used to analyse the performance of MSDA methods, which are described as follows. (1) *No Adapt*: Source classifiers are directly used to get target predictions, and the best result is reported. We report the numbers obtained from ResNet models trained on source data. Since our source feature extractor has a few extra layers after ResNet, we also report the accuracy obtained by using our source-specific classifiers. (2) *Single-Source Best*: In this setting, conventional single-source methods [**b**, **c**], **c**] are used to perform single-source DA between each source-target pair, and the best result is reported. (3) *Source Combine*: All the sources are unified into a single domain, and then single-source DA methods are applied for obtaining target predictions. (4) *Multi-Source*: We compare our approach with state-of-the-art MSDA works, such as DCTN, M^3 SDA, and MFSAN [**c**], **c**], **c**]. Since DCTN had used AlexNet, to get a fair comparison, we follow the strategy of M^3 SDA, and report numbers by employing ResNet in their model.

4.2 Results and Ablations

A. Loss-wise ablations. We firstly analyse the effect of each proposed loss (Table 1). Our baseline of \mathcal{L}_{qt} (Eq. 2) + $\mathcal{L}_{T \to W}$ (Eq. 7) + \mathcal{L}_{align} (Eq. 3) yields 80.0%, which is better than our no-adaptation baseline of 73.2%. Further, we analyse the effect of \mathcal{L}_{de} (Eq. 6) by conducting two experiments: first where only $\mathcal{L}_{distill}$ is used, and second where both $\mathcal{L}_{distill}$ and $\mathcal{L}_{entropy}$ are used. The enhancement obtained in each ablation justifies our formulation of \mathcal{L}_{de} . These

8

AGGARWAL, KUNDU, BABU, CHAKRABORTY: WAMDA

Standard	Method	$ACD \rightarrow W$	$ACW \rightarrow D$	$ADW \rightarrow C$	$CDW \rightarrow A$	Avg
No Adapt	Ours	100.0	100.0	89.6	95.5	96.3
Single Best	DAN	99.5	99.1	89.2	91.6	94.8
	MCD	99.5	99.1	91.5	92.1	95.6
Source Combine	DAN	99.3	98.2	89.7	94.8	95.5
	DCTN	99.4	99.0	90.2	92.7	95.3
	M ³ SDA	99.5	99.2	92.2	94.5	96.4
Multi-Source	MFSAN	99.0	98.7	93.3	95.3	96.6
	Ours	100.0	100.0	94.7	96.2	97.7

Table 4: Classification accuracy (%) on Office-Caltech dataset.

Standard	Method	$\text{CPR} \to \text{A}$	$APR \rightarrow C$	$ACR \rightarrow P$	$ACP \to R$	Avg
No Adapt	ResNet	65.3	49.6	79.7	75.4	67.5
	Ours	65.6	53.8	78.6	73.2	67.8
Single Best	D-CORAL	67.0	53.6	80.3	76.3	69.3
	RevGrad	67.9	55.9	80.4	75.8	70.0
	DAN	68.2	56.5	80.3	75.9	70.2
Source Combine	RevGrad	68.4	59.1	79.5	82.7	72.4
	DAN	68.5	59.4	79.0	82.5	72.4
Multi-Source	MFSAN	72.1	62.0	80.3	81.8	74.1
	Ours	71.9	61.4	84.1	82.3	74.9

Table 5: Classification accuracy (%) on <i>Office-Home</i> datase	et.
---	-----

Standard	Method	$DW \rightarrow A$	$AD \!\rightarrow\! W$	$AW \rightarrow D$	Avg
No Adapt	ResNet	62.5	96.7	99.3	86.2
	Ours	64.7	96.8	98.8	86.7
Single Best	RevGrad	68.2	96.9	99.1	88.1
	DAN	66.7	96.8	99.5	87.7
	RTN	66.2	96.8	99.4	87.5
Source Combine	RevGrad	67.6	97.8	99.6	88.3
	D-CORAL	67.1	98.0	99.3	88.1
Multi-Source	DCTN	64.2	98.2	99.3	87.2
	MFSAN	72.7	98.5	99.5	90.2
	Ours	72.0	98.6	99.6	90.0

Table 6: Classification accuracy (%) on Office-31 dataset.

ablations, thus, show that each loss is contributing to the performance of our model.

B. Analysis of weighting scheme. Next, we study the effect of doing weighted alignment instead of uniform alignment between domains. To obtain the results for uniform alignment, each relevance weight is replaced by (1/K), where K is the number of sources. As shown in Table 2, weighted alignment gives better results compared to uniform alignment. In ACP \rightarrow R, R is similar to both A and P, and highly dissimilar to C. This ordering can be verified by checking the accuracy of target by source classifiers, more details of which are presented in the Supplementary. In this split, the gains in performance is due to the decreased interaction with C. In ACR \rightarrow P, R has high relevance while both A and C are less relevant. Here, we see that uniform weighting gave a bigger drop because it increases the interactions with A and C. Hence, these experiments validate the benefits of our weighted alignment method, in comparison to the strategy of learning a fully domain-invariant space. C. Results on Office-Caltech. The results on Office-Caltech are shown in Table 4. With an average accuracy of 97.7%, we obtain improved results over the state-of-the-art methods.

D. Results on Office-Home. We analyse the performance of our method on *Office-Home*, which has the more significant domain shifts. As shown in Table 5, for the target **P**, it can be observed that *source combine* setting underperforms compared to *single-source best* baseline. This is indicative that the target **P** is prone to negative transfer. In this split, we achieve a jump of 3.8% over MFSAN. Overall, we achieve state-of-art performance of 74.9%. We also checked the sensitivity of *K* used in calculating α_i on the accuracy. As shown in Table 3, the selected value of K = 3 is optimal for ACR \rightarrow **P**.

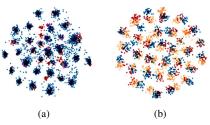


Figure 4: Visualization of weighted alignment: (a) $DW \rightarrow A$ of *Office-31*: D (pink) and W (red) are equally relevant to A (blue). Plot shows equal alignment of domains, and formation of class-wise clusters. (b) $AD \rightarrow W$ of *Office-31*: A (orange) is less relevant than D (blue) for the target W(red). We see that partial alignment of domains, such that W (red) is better aligned with D (blue) than with A (orange). Best viewed in color.

E. Results on Office-31. The results on *Office-31* are presented in Table 6. In **DW** \rightarrow **A**, both the sources are equally relevant to the target. In this split, MFSAN performed better than our method, but we significantly outperform DCTN. In the other splits of **AW** \rightarrow **D** and **AD** \rightarrow **W**, we obtain better results than MFSAN. On average, MFSAN performed slightly better. **F. Visualization.** To visualize the weighted alignment of domains, we study the t-SNE plots of the features from E_T and each E_{S_i} , shown in Fig. 4. We firstly visualize the features for **DW** \rightarrow **A** (Fig. 4(a)), which is a split where both sources are equally relevant to the target. As required, the target aligns well with both the sources and distinct clusters are formed for each class. We next visualize the split of **AD** \rightarrow **W**. In this split, **A** is less similar to **W** than **D**. The visualization (Fig. 4(b)) shows that weighted alignment of domains is performed, such that **W** is better aligned with **D**, and less aligned with **A**. We also observe that a cluster is formed per class. This shows that the intermediate spaces of the sources retain class information, and the target learns discriminative information by the weighted interaction.

5 Conclusion

In this work, we perform weighted alignment of domains, which leverages the source-target and source-source similarities for effective multi-source domain adaptation. We learn highly discriminative source-specific intermediate spaces and then, through the adaptation process, perform the weighted alignment of domains. This allows the learnt weighted aligned space to be optimally designed for the target. We motivate and formulate novel losses for the adaptation step and showcase the effectiveness of each contribution. In future, we shall extend our approach to MSDA in the presence of category shift.

Acknowledgments

This work is partially supported by Pratiksha Trust, Bangalore. The authors also thank Mr. Naveen Venkat and Ms. Siddhinita Wandekar for helpful discussions and technical support.

References

- [1] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. In *NeurIPS*, 2007.
- [2] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and

Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 2010.

- [3] Xinyang Chen, Sinan Wang, Mingsheng Long, and Jianmin Wang. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In *ICML*, 2019.
- [4] Lixin Duan, Ivor W Tsang, Dong Xu, and Tat-Seng Chua. Domain adaptation from multiple sources via auxiliary classifiers. In *ICML*, 2009.
- [5] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *ICML*, 2015.
- [6] 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*, 2016.
- [7] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In CVPR. IEEE, 2012.
- [8] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. In *NeurIPS*, 2005.
- [9] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 2012.
- [10] Jiang Guo, Darsh J Shah, and Regina Barzilay. Multi-source domain adaptation with mixture of experts. arXiv preprint arXiv:1809.02256, 2018.
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR. IEEE, 2016.
- [12] Judy Hoffman, Mehryar Mohri, and Ningshan Zhang. Algorithms and theory for multiple-source adaptation. In *NeurIPS*, 2018.
- [13] Yitong Li, David E Carlson, et al. Extracting relationships by multi-domain matching. In *NeurIPS*, 2018.
- [14] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning features with deep adaptation networks. In *ICML*, 2015.
- [15] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Unsupervised domain adaptation with residual transfer networks. In *NeurIPS*, 2016.
- [16] Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain adaptation with multiple sources. In *NeurIPS*, 2009.
- [17] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *ICCV*, 2018.
- [18] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In ECCV, 2010.

12	AGGARWAL, KUNDU, BABU, CHAKRABORTY: WAMDA
[19]	Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In <i>CVPR</i> . IEEE, 2018.
[20]	Swami Sankaranarayanan, Yogesh Balaji, Carlos D Castillo, and Rama Chellappa. Generate to adapt: Aligning domains using generative adversarial networks. In <i>CVPR</i> . IEEE, 2018.
[21]	Jian Shen, Yanru Qu, Weinan Zhang, and Yong Yu. Wasserstein distance guided repre- sentation learning for domain adaptation. In <i>AAAI Conference on Artificial Intelligence</i> , 2018.
[22]	Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In <i>ECCV</i> , 2016.
[23]	Qian Sun, Rita Chattopadhyay, Sethuraman Panchanathan, and Jieping Ye. A two-stage weighting framework for multi-source domain adaptation. In <i>NeurIPS</i> , 2011.
[24]	A Torralba and AA Efros. Unbiased look at dataset bias. In CVPR. IEEE, 2011.
[25]	Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Pan- chanathan. Deep hashing network for unsupervised domain adaptation. In <i>CVPR</i> . IEEE, 2017.
[26]	Ruijia Xu, Ziliang Chen, Wangmeng Zuo, Junjie Yan, and Liang Lin. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In <i>CVPR</i> . IEEE, 2018.
[27]	Han Zhao, Shanghang Zhang, Guanhang Wu, José MF Moura, Joao P Costeira, and Geoffrey J Gordon. Adversarial multiple source domain adaptation. In <i>NeurIPS</i> , 2018.
[28]	Yongchun Zhu, Fuzhen Zhuang, and Deqing Wang. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. In AAAI Conference on Artificial Intelligence, 2019.
[29]	Fuzhen Zhuang, Xiaohu Cheng, Ping Luo, Sinno Jialin Pan, and Qing He. Supervised representation learning: Transfer learning with deep autoencoders. In <i>IJCAI</i> , 2015.