Class Subset Selection for Partial Domain Adaptation

Fariba Zohrizadeh Mohsen Kheirandishfard Farhad Kamangar
Department of Computer Science and Engineering,
University of Texas at Arlington
{fariba.zohrizadeh,mohsen.kheirandishfard,farhad.kamangar}@uta.edu

Abstract

Domain adaptation is the task of transferring knowledge from a labeled source dataset to an unlabeled target dataset. Partial domain adaptation (PDA) investigates the scenarios in which the target label space is a subset of the source label space. The main purpose of the PDA is to identify the shared classes between the domains and promote learning transferable knowledge from these classes. Inspired by the idea of subset selection, we propose an adversarial PDA approach which aims to not only automatically select the most relevant subset of source domain classes but also ignore the samples that are less transferable across the domains. In the absence of target labels, the proposed approach is able to effectively learn domain-invariant feature representations, which in turn can facilitate and enhance the classification performance in the target domain. Empirical results on Office-31 and Office-Home datasets demonstrate the high potential of the proposed approach in addressing different partial domain adaptation tasks.

1. Introduction

Deep neural networks have demonstrated superior performance in a variety of machine learning problems such as semantic image segmentation [5, 11, 17], object detection, and classification [16, 24, 30], etc. These impressive achievements heavily depend on the availability of large amounts of labeled training data. However, in many applications, the acquisition of sufficient labeled data is difficult and time-consuming. One potential solution to reduce the labeling consumption is to build an effective predictive model using richly-annotated datasets from different but related domains. However, this paradigm generally suffers from the domain shift between the distributions of the source and the target datasets. As a result, deep networks trained on labeled source datasets often exhibit unsatisfactory performance on the target domain classification task. In the absence of target labels, unsupervised domain adaptation (UDA) seeks to bridge different domains by learning feature representations that are discriminative and domain-invariant [1, 12, 21].

Recently, various approaches have been proposed to combine both domain adaptation and deep feature learning in a unified framework for exploiting more transferable knowledge across domains [6, 7, 15, 18, 31, 37] (see [34] for a comprehensive survey on deep domain adaptation methods). A class of deep domain adaptation methods aims to reduce the misfit between the distributions of the source and target domains through minimizing discrepancy measures such as maximum mean discrepancy [15, 18], correlation distance [27, 29], etc. In this way, they map the domains into the same latent space, which results in learning feature representations that are domain-invariant. A new line of research has recently emerged which uses the concept of generative adversarial networks [13] to align feature distributions across the domains and learn discriminators that are able to predict the domain labels of different samples [19, 23, 35]. Specifically, these methods try to generate feature representations that are difficult for the discriminators to differentiate.

Despite the advantages offered by the existing UDA methods, they mostly exhibit superior performance in scenarios in which the source and target domains share the same label space. With the goal of considering more realistic cases, [4] introduced partial domain adaptation (PDA) as a new adaptation scenario which assumes the target domain label space is a subset of the source domain label space. The primary challenge in PDA is to identify and reject the source domain classes that do not appear in the target domain, known as *outlier classes*, since they may have negative impacts on the transfer performance [3, 22]. Addressing this challenge enables the PDA methods to effectively transfer models learned on large labeled datasets (e.g. ImageNet) to small-scale datasets from different but re-

lated domains.

In this paper, we propose an adversarial approach for partial domain adaptation, which aims to not only automatically reject the outlier source classes, but also down-weight the relative importance of *irrelevant samples*, i.e. those samples that are highly dissimilar across different domains. Our method uses the same network architecture as partial adversarial domain adaptation (PADA) [4] and incorporates two additional regularization terms to boost the target domain classification performance. Inspired by the idea of subset selection, the first regularization is a row-sparsity term on the output of the classifier, which promotes the selection of a small subset of classes that are in common between the source and target domains. The second regularization is a minimum entropy term which utilizes the output of the discriminator to down-weight the relative importance of irrelevant samples from both domains. We empirically observe that our method can effectively enhance the target classification accuracy on different PDA tasks.

2. Related Work

To date, various deep unsupervised domain adaptation methods have been developed to extract domain-invariant feature representations from different domains. Some studies [9, 15, 18, 20, 36] have proposed to minimize the maximum mean discrepancy between the source and target distributions. In [28], a correlation alignment (CORAL) method is proposed that utilizes a linear transformation to match the second-order statistics between the domains. [29] presented an extension of the CORAL method that aligns correlations of layer activations in deep networks by learning a non-linear transformation. Despite the practical success of the aforementioned methods in aligning the domain distributions, it is shown that they are unable to completely eliminate the domain shift [7, 37].

Recently, adversarial learning has been widely employed to enhance the performance of UDA methods [2, 8, 10, 19, 32]. The basic idea behind the adversarial-based methods is to train a discriminator for predicting domain labels and a deep network for extracting features that are indistinguishable by the discriminator. By doing so, the discrepancy between the source and target domains can be efficiently eliminated, which results in significant improvement in the overall classification performance [8, 23, 32]. [39] developed an incremental adversarial scheme which gradually reduces the gap between the domain distributions by iteratively selecting the high confidence pseudo-labeled target samples to enlarge the training set.

Towards the task of PDA, great studies have been recently developed which simultaneously promote positive transfer from the common classes between the domains and alleviate the negative transfer from the outlier classes [3, 4, 38]. Selective adversarial network [3] trains separate domain discriminators for each source class to align the distributions of the source and target domains across the shared label space and to ignore the outlier classes. Partial adversarial domain adaptation (PADA) [4] proposed a new architecture which assigns a weight to each source domain class based on the target label prediction and automatically reduces the weights of the outlier classes. Importance weighted adversarial nets [38] develops a two-domain classifier strategy to estimate the relative importance of the source domain samples.

Closely related to our work, transferable attention for domain adaptation (TADA) [35] proposed an attention-based mechanism for UDA, which can highlight transferable regions or images. Unlike TADA, our method is focused on the PDA problem and utilizes a different network architecture with a novel loss function that efficiently assigns weights to both classes and samples. Our method differs from PADA [4] in the sense that we incorporate two novel regularization terms which not only able to discover and reject the outlier classes more effectively but also down-weight the relative importance of the irrelevant samples in the training procedure.

3. Problem Formulation

This section briefly reviews two well-established domain adaptation methods and then provides a detailed explanation on how our proposed method relates to them. Let $\{(\boldsymbol{x}_s^i, \boldsymbol{y}_s^i)\}_{i=1}^{n_s}$ be a set of n_s sample points drawn $i.i.d$ from the source domain $\mathcal{D}_s$, where $\boldsymbol{x}_s^i$ denotes the i^{th} source image with label $\boldsymbol{y}_s^i$. Similarly, let $\{\boldsymbol{x}_t^i\}_{i=1}^{n_t}$ be a set of n_t sample points collected $i.i.d$ from the target domain $\mathcal{D}_t$, where $\boldsymbol{x}_t^i$ indicates the i^{th} target image. To clarify notation, let $\mathcal{X} = \mathcal{X}_s \cup \mathcal{X}_t$ be the set of entire images from both domains, where $\mathcal{X}_s = \{\boldsymbol{x}_s^i\}_{i=1}^{n_s}$ and $\mathcal{X}_t = \{\boldsymbol{x}_t^i\}_{i=1}^{n_t}$. The UDA methods assume that the source and target domains possess the same label space, denoted as $\mathcal{C}_s$ and $\mathcal{C}_t$, respectively. In the absence of target labels, the primary goal of the UDA is to learn domain-invariant feature representations that can reduce the domain shift. One promising direction to achieve this goal is to train a domain adversarial neural network [8] which consists of a discriminator G_d for predicting the domain labels, a feature extractor G_f for confusing the discriminator by learning transferable feature representations, and a classifier G_y that classifies the source domain samples. Training the adversarial network is equivalent to solve the following

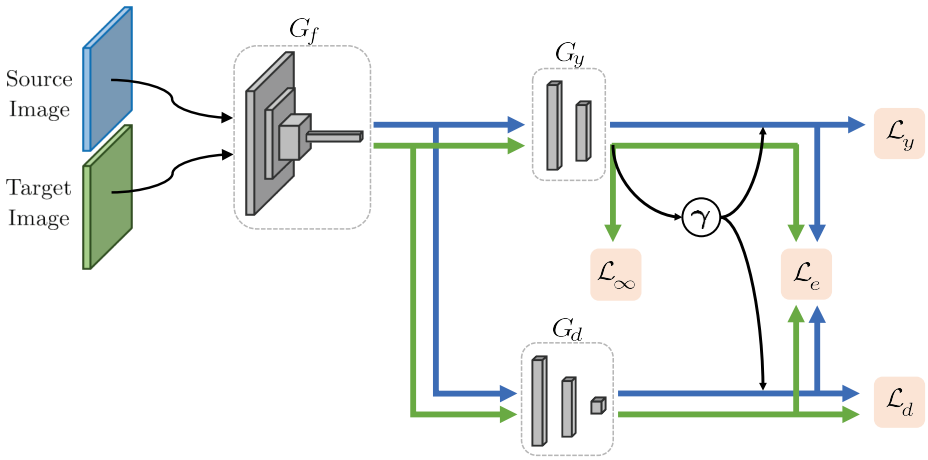

Figure 1: Overview of the proposed adversarial network for partial transfer learning. The network consists of a feature extractor, a classifier, and a domain discriminator, denoted by $\boldsymbol{G}_f$, $\boldsymbol{G}_y$, and $\boldsymbol{G}_d$, respectively. The blue and green arrows depict the source flow and target flow. Loss functions $\mathcal{L}_y$, $\mathcal{L}_d$, $\mathcal{L}_e$, and $\mathcal{L}_\infty$ denote the classification loss, the discriminative loss, the entropy loss, and the selection loss. *Best viewed in color.*

optimization problem

$$\max_{\boldsymbol{\theta}_d} \min_{\boldsymbol{\theta}_y,\boldsymbol{\theta}_f} \frac{\lambda}{n_s} \sum_{\boldsymbol{x}^i \in \mathcal{X}_s} L_y(G_y(G_f(\boldsymbol{x}^i;\boldsymbol{\theta}_f);\boldsymbol{\theta}_y), \boldsymbol{y}^i) - \frac{1}{n} \sum_{\boldsymbol{x}^i \in \mathcal{X}} L_d(G_d(G_f(\boldsymbol{x}^i;\boldsymbol{\theta}_f);\boldsymbol{\theta}_d), d^i),$$

where $n = n_s + n_t$, $\lambda > 0$ is a regularization parameter, $\boldsymbol{y}^i$ is the one-hot class label of image $\boldsymbol{x}^i$ and $d^i \in \{0, 1\}$ denotes its domain label; $d^i = 0$ if $\boldsymbol{x}^i$ belongs to the source domain and $d^i = 1$ otherwise. L_y and L_d are cross-entropy loss functions corresponding to the classifier G_y and the domain discriminator G_d, respectively. Moreover, variables $\boldsymbol{\theta}_f$, $\boldsymbol{\theta}_y$, and $\boldsymbol{\theta}_d$ are parameters associated with the networks G_f, G_y, and G_d, respectively. For the brevity of notation, we drop the reference to the parameters $\boldsymbol{\theta}_f$, $\boldsymbol{\theta}_y$, and $\boldsymbol{\theta}_d$ in the subsequent formulations.

As noted earlier, standard domain adaptation approaches assume that the source and target possess the same label space, i.e. $\mathcal{C}_s = \mathcal{C}_t$. This assumption may not be fulfilled in a wide range of practical applications in which $\mathcal{C}_s$ is large and diverse (e.g., ImageNet) and $\mathcal{C}_t$ only contains a small subset of source classes, i.e. $\mathcal{C}_t \subset \mathcal{C}_s$. Under this assumption, aligning the domain distributions may not necessarily facilitate the classification task in the target domain due to the adverse effect of transferring information from the outlier classes $\mathcal{C}_s \setminus \mathcal{C}_t$ [3, 4]. Hence, the primary goal in partial domain adaptation is to learn a feature extractor that can align the distributions of the source and target domains across the shared label space and simultaneously identify and reject the outlier classes. A classifier trained along such feature extractor can generalize well to the target domain. To this end, PADA [4] proposed the following weighting procedure to highlight the shared classes and reduce the importance of outlier classes

$$\boldsymbol{\gamma} = \frac{1}{n_t} \sum_{i=1}^{n_t} \hat{\boldsymbol{y}}_t^i$$

where $\hat{\boldsymbol{y}}_t^i = G_y(G_f(\boldsymbol{x}_t^i))$ denotes the output of G_y to the target sample $\boldsymbol{x}_t^i$. The weighting vector $\boldsymbol{\gamma}$ is further normalized as $\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma} \setminus \max(\boldsymbol{\gamma})$ to show the relative weights of the classes.

The weights associated with the outlier classes are expected to be much smaller than that of the shared classes, mainly because the target samples are significantly dissimilar to the samples belonging to the outlier classes. Ideally, $\boldsymbol{\gamma}$ is expected to be a vector whose elements are non-zero except those corresponding to the outlier classes. Given that PADA proposes to train the adversarial network through solving the following minimax optimization problem

$$\max_{\boldsymbol{\theta}_d} \min_{\boldsymbol{\theta}_y,\boldsymbol{\theta}_f} \frac{\lambda}{n_s} \sum_{\boldsymbol{x}^i \in \mathcal{X}_s} \gamma_{c_i} L_y(G_y(G_f(\boldsymbol{x}^i)), \boldsymbol{y}^i) - \frac{1}{n_s} \sum_{\boldsymbol{x}^i \in \mathcal{X}_s} \gamma_{c_i} L_d(G_d(G_f(\boldsymbol{x}^i)), d^i) - \frac{1}{n_t} \sum_{\boldsymbol{x}^i \in \mathcal{X}_t} L_d(G_d(G_f(\boldsymbol{x}^i)), d^i),$$

where $c_i = \operatorname{argmax}_j y_j^i$ denotes the index of the largest element in $\boldsymbol{y}^i$.

Besides the outlier classes, the irrelevant samples are inherently less transferable and they may significantly degrade the target classification performance in different PDA tasks. In the next section, we present a novel algorithm to simultaneously identify and reject the outlier classes and down-weight the relative importance of the irrelevant samples.

4. Proposed Method

We adopt the same network architecture as PADA and employ two novel regularization terms to better align the source and target distributions across the shared classes and learn more transferable features.

The first regularization is a row-sparsity term which promotes the selection of a small subset of source domain classes that appear in the target domain. This, in turn, encourages $\boldsymbol{\gamma}$ to be a vector of zeros except for the elements corresponding to the shared classes. This selection regularization can be defined as follows

$$\mathcal{L}_{\infty}(\mathcal{X}_t,\boldsymbol{\theta}_f,\boldsymbol{\theta}_y)=\frac{\mu}{|\mathcal{C}_s|}\left\|G_y(G_f(\boldsymbol{x}_t^1)),\ldots,G_y(G_f(\boldsymbol{x}_t^{|\mathcal{X}_t|}))\right\|_{1,\infty},$$

where $|.|$ denotes the cardinality of its input set, $\|.\|_{1,\infty}$ computes the sum of the infinity norms of the rows of an input matrix, and μ is a regularization parameter. Imposing the above term takes into account the relation between the entire target samples and encourages the classifier to generate a sparse output vector with its non-zero entries located at certain indices correspond to the shared classes.

The second regularization term seeks to reduce the importance of irrelevant samples in the training procedure by leveraging the following entropy minimization term

$$\begin{aligned}\mathcal{L}_e(\mathcal{X}_s,\mathcal{X}_t,\boldsymbol{\theta}_f,\boldsymbol{\theta}_d,\boldsymbol{\theta}_y)=&\\ &\frac{1}{n_s}\sum_{\boldsymbol{x}^i\in\mathcal{X}_s}\gamma_{c_i}(1+L_d^e(G_d(G_f(\boldsymbol{x}^i))))\,L_y^e(G_y(G_f(\boldsymbol{x}^i)))\\ &+\frac{1}{n_t}\sum_{\boldsymbol{x}^i\in\mathcal{X}_t}(1+L_d^e(G_d(G_f(\boldsymbol{x}^i))))\,L_y^e(G_y(G_f(\boldsymbol{x}^i))),\end{aligned}$$

where L_y^e and L_d^e are the entropy loss functions corresponding to the classifier G_y and the domain discriminator G_d, respectively. The above regularization encourages assigning higher weights to those samples whose domain labels are confidently predicted by the discriminator. This, in turn, reduces the relative importance of the irrelevant samples and helps to learn more transferable features for classification.

By integrating both regularization terms into the total loss function, our method can not only automatically identify and reject the outlier classes, but also down-weight the irrelevant samples that are inherently not transferable across domains. Figure 1 illustrates the architecture of our proposed network in details.

Figure 2: Example images of the Office-31 dataset.

5. Experiments

In this section, we conduct empirical experiments on two benchmark datasets to evaluate the efficacy of our approach, named SSPDA, for partial domain adaptation (PDA) across different tasks. The experiments are performed in an unsupervised setting, where the target labels are unknown. In what follows, we briefly explain the datasets, the PDA tasks, and the network hyperparameters used in the experiments.

5.1. Setup

Dataset: We evaluate the performance of SSPDA on two commonly used datasets for domain adaptation: Office-31 and Office-Home. The Office-31 object dataset [26] consists of 4,652 images from 31 classes, where the images are collected from three different domains: *Amazon* (**A**), *Webcam* (**W**), and *DSLR* (**D**). We follow the procedure presented in [4] to transfer knowledge from a source domain with 31 classes to a target domain with 10 classes. The results are provided as the target domain classification accuracy across six different PDA tasks: **A** → **W**, **W** → **A**, **D** → **W**, **W** → **D**, **A** → **D**, and **D** → **A**.

The Office-Home [33] is a more complex dataset consisting of around 15,500 images collected from four distinct domains: *Art* (**Ar**), *Clipart* (**Cl**), *Product* (**Pr**), and *Real-World* (**Rw**), where each domain has 65 classes. Following the procedure presented in [4], we aim to transfer information from a source domain containing 65 classes to a target domain with 25 classes. The results on this dataset are also reported as the target classification accuracy on twelve pairs of source-target domains: **Ar** → **Cl**, **Ar** → **Pr**, **Ar** → **Rw**, **Cl** → **Ar**, **Cl** → **Pr**, **Cl** → **Rw**, **Pr** → **Ar**, **Pr** → **Cl**, **Pr** → **Rw**, **Rw** → **Ar**, **Rw** → **Cl**, and **Rw** → **Pr**.

For each of the aforementioned tasks, we report the

Method	A → W	D → W	W → D	A → D	D → A	W → A	Avg
ResNet	54.52	94.57	94.27	65.61	73.17	71.71	75.64
DAN	46.44	53.56	58.60	42.68	65.66	65.34	55.38
DANN	41.35	46.78	38.85	41.36	41.34	44.68	42.39
ADDA	43.65	46.48	40.12	43.66	42.76	45.95	43.77
RTN	75.25	97.12	98.32	66.88	85.59	85.70	84.81
SAN	80.02	98.64	100	81.28	80.58	83.09	87.27
IWAN	76.27	98.98	100	78.98	89.46	81.73	87.57
PADA	86.54	**99.32**	**100**	82.17	92.69	95.41	92.69
SSPDA-selection	87.45	95.31	98.48	82.25	91.89	95.34	91.79
SSPDA-entropy	90.51	96.59	97.45	89.08	92.38	95.30	93.55
SSPDA	**93.42**	97.62	**100**	**90.43**	**93.45**	**95.53**	**95.07**

Table 1: Accuracy of partial domain adaptation tasks on *Office-31* (ResNet-50).

Method	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg
ResNet	38.57	60.78	75.21	39.94	48.12	52.90	49.68	30.91	70.79	65.38	41.79	70.42	53.71
DAN	44.36	61.79	74.49	41.78	45.21	54.11	46.92	38.14	68.42	64.37	45.37	68.85	54.48
DANN	44.89	54.06	68.97	36.27	34.34	45.22	44.08	38.03	68.69	52.98	34.68	46.50	47.39
RTN	49.37	64.33	76.19	47.56	51.74	57.67	50.38	41.45	75.53	70.17	51.82	74.78	59.25
PADA	51.95	67	78.74	**52.16**	**53.78**	59.03	52.61	**43.22**	78.79	**73.73**	**56.6**	77.09	62.06
SSPDA	**52.31**	**68.35**	**80.17**	50.79	51.29	**60.87**	**56.68**	42.53	**79.15**	70.94	56.43	**78.92**	**62.37**

Table 2: Accuracy of partial domain adaptation tasks on *Office-Home* (ResNet-50).

average target classification accuracy of five independent runs with different initialization as generated in [4]. We compare the performance of SSPDA against several deep transfer learning methods: Deep Adaptation Network (DAN) [18], Domain Adversarial Neural Network (DANN) [8], Residual Transfer Networks (RTN) [20], Adversarial Discriminative Domain Adaptation (ADDA) [32], Importance Weighted Adversarial Nets (IWAN) [38], Selective Adversarial Network (SAN) [3], and Partial Adversarial Domain Adaptation (PADA) [4].

Parameter: We adopt ResNet-50 [14] pre-trained on ImageNet [25] as the backbone for the network G_f. Also, we fine-tune the entire feature layers and apply back-propagation to train the domain discriminator G_d and the classifier G_y. Through the experiments, parameter λ is set to 1.0 and 2.0 for the Office-31 dataset and Office-Home dataset, respectively. Also, we set $\mu = 0.1$ for both datasets. Notice that since the classifier is not appropriately trained in the first few epochs, we gradually increase parameter μ from 0 to 0.1. To minimize the loss function, we use mini-batch stochastic gradient descent (SGD) with a momentum of 0.95 and the learning rate is adjusted during SGD by: $\eta = \frac{\eta_0}{(1+\alpha\times\rho)^\beta}$ where $\eta_0 = 10^{-2}$, $\alpha = 10$, $\beta = 0.75$, and ρ is the training progress linearly changing from 0 to 1 [4, 8]. We use a batch size of 72 with 36 samples for each domain.

5.2. Results

Tables 1 and 2 show the target domain classification accuracy of various methods on different PDA tasks including 6 tasks of Office-31 dataset and 12 tasks of Office-Home dataset. All the results are reported based on the ResNet-50 and the scores of the competitor methods are directly obtained from [3, 4, 38].

Observe that some deep domain adaptation methods such as DAN and DANN have exhibited worse performance than the standard ResNet-50 on few PDA tasks in both datasets. This can be attributed to the fact that these methods aim to align the marginal distributions across the domains and hence are prone to the negative transfer resulted from the outlier classes. On the other hand, the PDA methods, such as PADA, SAN, and IWAN, achieve promising results on most of the PDA tasks since they leverage weighting mechanisms to highlight a subset of samples that are more transferable. By doing so, these methods can effectively mitigate transferring knowledge from the outlier source classes and promote learning from the shared classes between the domains, which in turn enhance the classification accuracy in the target domain.

Notice that SSPDA uses the same network architecture as PADA, but introduce a novel loss function to identify and reject the outlier classes and irrelevant samples. The results in Table 1 indicate that SSPDA performs better than or close to the state-of-the-art methods at all PDA tasks on Office-31 dataset. In par-

ticular, it achieves considerable improvement on $\mathbf{A} \rightarrow \mathbf{W}$ and $\mathbf{A} \rightarrow \mathbf{D}$, and generally increases the average accuracy of all tasks by almost 2.4%. Moreover, Table 2 shows that SSPDA maintains the performance of PADA and exhibits slight improvement in the average classification accuracy over all partial domain adaptation tasks on Office-Home dataset.

The numerical results provided in the above tables imply that SSPDA has high potential in transferring semantic information and learning domain-invariant features in different tasks of partial domain adaptation.

Ablation Study: To demonstrate the improvements obtained by each of the proposed regularizations, in this part, we conduct an ablation study by discarding the selection regularization (SSPDA-selection) or the entropy minimization term (SSPDA-entropy). The results are reported in Table 1. It can be seen that both SSPDA-selection and SSPDA-entropy generally obtain better or close results than the baselines. In particular, SSPDA-entropy works better on some difficult tasks such as $\mathbf{A} \rightarrow \mathbf{W}$ and $\mathbf{A} \rightarrow \mathbf{D}$.

6. Conclusion

This work presented an adversarial approach for the task of partial domain adaptation. The proposed approach minimizes a novel loss function to reduce the effect of the outlier classes and the irrelevant samples, which results in learning more transferable feature representations for classification. The experiments conducted on standard benchmark datasets demonstrate the high potential of our approach for partial domain adaptation tasks and highlight the directions for future explorations and research. Future work will explore the effectiveness and the generalization power of our designed loss functions in different adversarial architectures.

References

[1] Mahsa Baktashmotlagh, Mehrtash T Harandi, Brian C Lovell, and Mathieu Salzmann. Unsupervised domain adaptation by domain invariant projection. In *ICCV*, 2013. 1

[2] Konstantinos Bousmalis, Nathan Silberman, David Dohan, Dumitru Erhan, and Dilip Krishnan. Unsupervised pixel-level domain adaptation with generative adversarial networks. In *CVPR*, 2017. 2

[3] Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Michael I Jordan. Partial transfer learning with selective adversarial networks. In *CVPR*, 2018. 1, 2, 3, 5

[4] Zhangjie Cao, Lijia Ma, Mingsheng Long, and Jianmin Wang. Partial adversarial domain adaptation. In *ECCV*, 2018. 1, 2, 3, 4, 5

[5] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE TPAMI*, 40(4):834–848, 2018. 1

[6] Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *CVPR*, 2018. 1

[7] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *ICML*, 2014. 1, 2

[8] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *J MACH LEARN RES*, 17(1):2096–2030, 2016. 2, 5

[9] Muhammad Ghifary, W Bastiaan Kleijn, and Mengjie Zhang. Domain adaptive neural networks for object recognition. In *PRICAI*, pages 898–904. Springer, 2014. 2

[10] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, David Balduzzi, and Wen Li. Deep reconstruction-classification networks for unsupervised domain adaptation. In *ECCV*, 2016. 2

[11] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014. 1

[12] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *CVPR*, 2012. 1

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, 2014. 1

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 5

[15] Hadi Kazemi, Sobhan Soleymani, Fariborz Taherkhani, Seyed Iranmanesh, and Nasser Nasrabadi. Unsupervised image-to-image translation using domain-specific variational information bound. In *NeurIPS*, 2018. 1, 2

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NeurIPS*, 2012. 1

[17] Ziwei Liu, Xiaoxiao Li, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Semantic image segmentation via deep parsing network. In *ICCV*, 2015. 1

[18] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *ICML*, 2015. 1, 2, 5

[19] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *NeurIPS*, 2018. 1, 2

[20] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Unsupervised domain adaptation with residual transfer networks. In *NeurIPS*, 2016. 2, 5

[21] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE T NEURAL NETWOR*, 22(2):199–210, 2011. 1

[22] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE T KNOWL DATA EN*, 22(10):1345–1359, 2010. 1

[23] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. In *AAAI*, 2018. 1, 2

[24] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NeurIPS*, 2015. 1

[25] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *IJCV*, 115(3):211–252, 2015. 5

[26] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *ECCV*, 2010. 4

[27] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. In *AAAI*, 2016. 1

[28] Baochen Sun, Jiashi Feng, and Kate Saenko. Correlation alignment for unsupervised domain adaptation. In *Domain Adaptation in Computer Vision Applications*, pages 153–171. Springer, 2017. 2

[29] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *EECV*, 2016. 1, 2

[30] Christian Szegedy, Alexander Toshev, and Dumitru Erhan. Deep neural networks for object detection. In *NeurIPS*, 2013. 1

[31] Eric Tzeng, Judy Hoffman, Trevor Darrell, and Kate Saenko. Simultaneous deep transfer across domains and tasks. In *ICCV*, 2015. 1

[32] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *CVPR*, 2017. 2, 5

[33] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *CVPR*, 2017. 4

[34] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018. 1

[35] Ximei Wang, Liang Li, Weirui Ye, Mingsheng Long, and Jianmin Wang. Transferable attention for domain adaptation. In *AAAI*, 2019. 1, 2

[36] Hongliang Yan, Yukang Ding, Peihua Li, Qilong Wang, Yong Xu, and Wangmeng Zuo. Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation. In *CVPR*, 2017. 2

[37] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *NeurIPS*, 2014. 1, 2

[38] Jing Zhang, Zewei Ding, Wanqing Li, and Philip Ogunbona. Importance weighted adversarial nets for partial domain adaptation. In *CVPR*, 2018. 2, 5

[39] Weichen Zhang, Wanli Ouyang, Wen Li, and Dong Xu. Collaborative and adversarial network for unsupervised domain adaptation. In *CVPR*, 2018. 2