ATTENTION-BASED ADVERSARIAL PARTIAL DOMAIN ADAPTATION

Mengzhu Wang[1], *Shan An*[2], *Xiao Luo*[3],*Xiong Peng*[1],*Wei Yu*[1], *Junyang Chen*[4*], *Zhigang Luo*[1]

[1]National University of Defense Technology
[2]Tech & Data Center JD.COM Inc
[3]Peking University, [4]Shenzhen University

ABSTRACT

With the rapid development of vision-based deep learning (DL), it is an effective method to generate large-scale synthetic data to supplement real data to train the DL models for domain adaptation. However, previous vanilla domain adaptation methods generally assume the same label space, and such an assumption is no longer valid for a more realistic scenario where it requires adaptation from a larger and more diverse source domain to a smaller target domain with less number of classes. To handle this problem, we propose an attention-based adversarial partial domain adaptation (AADA). Specifically, we leverage adversarial domain adaptation to augment the target domain by using source domain, then we can readily turn this task into a vanilla domain adaptation. Meanwhile, to accurately focus on the transferable features, we apply attention-based method to train the adversarial networks to obtain better transferable semantic features. Experiments on four benchmarks demonstrate that the proposed method outperforms existing methods by a large margin, especially on the tough domain adaptation tasks, e.g. VisDA-2017.

Index Terms— Deep learning, Adversarial domain adaptation, Attention-based method

1. INTRODUCTION

It is usually assumed that the training and testing data are derived from the same distribution in statistical learning theory [1]. However, collecting and labeling large-scale domain-specific training data for domain adaptation is time consuming and expensive [2]. The sample selection bias in the data collection constrains the unsupervised domain adaptation (UDA) [3] to be used in limited environments. It uses previously labeled source domain to facilitate tasks in target domain with little or no label. As the recent progress in UDA shows that more transferable and domain-invariant features can be extracted through deep frameworks, UDA techniques have also shifted from shallow learning based to DL based.

However, most current DL related methods still assume that the label space between the source domain and target domain are the same. This paper focus on another more practical environment where the target domain has only a subset of source domain or even no label is available. Intuitively, when the target domain contains only a subset of the source domain, it may be challenging to reduce the shift by directly comparing the source domain distribution and the target domain distribution since the label distribution of the two domains are different. In other words, reducing the distribution shift will not benefit the target domain. In this case, a natural and possible way to transfer from the source domain to the target domain is to weight the source samples whose classes may appear in the target domain during the training process.

To this end, we propose an attention-based adversarial partial domain adaptation (AADA). Specifically, we study the UDA problem from a new perspective that target domain is a part of source domain, and propose to expand the target label space to make it the same as the source label space. Besides, to accurately focus on the transferable features, we leverage coordinate attention [4] to capture long-ranged dependencies along one spatial direction and achieve precise positional information along the other spatial direction. Extensive experiments validate the promising performance of the proposed novel UDA scheme.

To sum up, we make the following contributions.

- We leverage adversarial learning to augment the target domain to effectively overcome the practical problem of smaller label space or no label for target domain. So far as we know, it is the first attempt for partial UDA by borrowing from the source domain to target domain and transforming it into a vanilla UDA problem.

- we leverage coordinate attention to capture transferable features and it can bring significant performance improvements.

- We theoretically analyze the transfer error bound of the proposed new scheme, and show that aligning the distributions of the target domain and the source domain can be bounded by adversarial alignment error in novel scenario of partial UDA through Ben-David [5] theory.

*This is the corresponding author. This work was supported by National Natural Science Foundation of China under Grant 62102265.

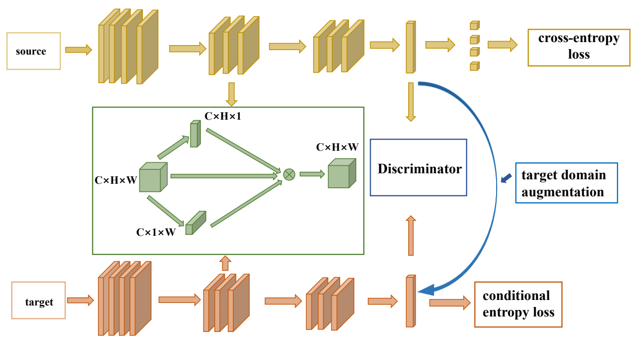

Fig. 1. The overall network of our AADA method. It includes two modules: a vanilla unsupervised adversarial module and a transferable attention module. The blue arrow denotes that we borrow from the source domain to target domain.

2. PROBLEM STATEMENT

We follow the protocol of partial UDA, which has a labeled source domain $\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$, $x_i^s \in \mathbb{R}^d$ of n_s labeled examples associated with $|\mathcal{C}_s|$ classes and an unlabeled target domain $\mathcal{D}_t = \{(x_i^t)\}_{i=1}^{n_t}$, $x_i^t \in \mathbb{R}^d$ of n_t unlabeled examples associated with $|\mathcal{C}_t|$ classes during the training process. The distribution of the source domain and target domain are not the same. Different from vanilla UDA, the target label space is a subset of the source label space for partial UDA.

As shown in Fig. 1, our AADA consists of two modules, the first one is an adversarial learning based on target domain augmentation, which leverages domain adversarial neural networks to develop a two-player game, including a domain discriminator and a feature extractor. Based on adversarial networks, we consider to borrow some samples from source domain to target domain which transform partial UDA to vanilla UDA. The second one is a transferable attention module, which improves adversarial methods by learning fine-grained feature information and transferable feature.

2.1. Adversarial Learning Based on Target Domain Augmentation

Inspired by the seminal idea of GAN [6], unsupervised transfer learning by backpropagation (DANN) [7] proposes a two player game for unsupervised domain adaptation, where one is discriminator D tries to distinguish the source domain and the target domain, another one is feature extractor F which try to confuse the discriminator D. Then the adversarial network DANN [7] can be written as:

$$\begin{aligned}
&\min_{\theta_f,\theta_g} \max_{\theta_d} \mathcal{L}_{cls}\left(\theta_f,\theta_g\right) + \lambda_1 \mathcal{L}_{adv}\left(\theta_f,\theta_d\right) \\
&\mathcal{L}_{adv}\left(\theta_f,\theta_d\right) = \mathbb{E}_{x_s\sim p(x_s)} \log\left[D\left(F\left(x_s\right)\right)\right] \\
&\qquad + \mathbb{E}_{x_t\sim p(x_t)} \log\left[1 - D\left(F\left(x_t\right)\right)\right], \\
&\mathcal{L}_{cls}\left(\theta_f,\theta_g\right) = \mathbb{E}_{x_s\sim p(x_s)} l_{ce}\left(C\left(F\left(x_s\right)\right), y_s\right)
\end{aligned} \tag{1}$$

where $l_{ce}(\cdot,\cdot)$ indicates the cross-entropy loss, $C(\cdot)$ indicates the source domain classifier, $F(\cdot)$ indicates the feature extractor, $D(\cdot)$ indicates the discriminator, λ_1 is a hyper-parameter to trade-off the classifier loss and adversarial loss. θ is learnable parameters corresponding to networks weight that need to train in the adversarial learning process. Due to the simplicity and applicability, DANN [7] and its variants appear in many previous works [8, 9]. Meanwhile, inspired by the sample selection strategy of CDAN [10], we expect that those hard and easy samples will have higher and lower weights during the adversarial alignment. Specifically, towards safe transfer, we quantify the difficulty of classifier prediction by the entropy criterion $H(g) = -\sum_{k=1}^{K} g_k \log g_k$, where K is the number of classes and g_k is the probability of predicting an example to class k. By weighting each training exemplar of the discriminator using entropy-aware weights, we prioritize those exemplars with certain predictions that are easily transferable by an entropy criterion weight $w(H(g)) = 1 + e^{-H(g)}$, then Eq. (1) can be rewritten as:

$$\begin{aligned}
\mathcal{L}_{adv}^{e}\left(\theta_f,\theta_d\right) = &\mathbb{E}_{x_s\sim p(x_s)} w(x_s) \log\left[D\left(F\left(x_s\right)\right)\right] \\
&+ \mathbb{E}_{x_t\sim p(x_t)} w(x_t) \log\left[1 - D\left(F\left(x_t\right)\right)\right]
\end{aligned} \tag{2}$$

Meanwhile, it is desirable that all the unlabeled target samples have highly-confident predictions. Then the conditional entropy term [11] that is widely used in unlabeled real samples can be formulated as:

$$\mathcal{L}_{ent}\left(\theta_f,\theta_y\right) = \mathbb{E}_{x_t\sim p(x_t)} H\left(G\left(F\left(x_t\right)\right)\right) \tag{3}$$

But the above formulation only considers that source domain and target domain have the same class while ignores the uncommon classes of the source domain. The previous methods [3, 12] rely on real prediction to generate class-level weights, and to some extent effectively avoid negative transfer. But all these methods only consider weighting on the common classes that are shared by source domain and target domain, and reducing the weight of uncommon classes that only appear in source domain. In contrast, we propose to augment target domain by borrowing original source domain. Then we can readily turn this task into a large vanilla UDA task. To handle this problem, we propose to augment target domain by copying of source domain to target domain. Specifically, we also use DANN [7] as our backbone, and augment the target domain by using source domain instead of weighting the source domain. Then our adversarial alignment formulation can be written as:

$$\begin{aligned}
\mathcal{L}_{adv}^{new}\left(\theta_f,\theta_d\right) = &\mathbb{E}_{x_s\sim p(x_s)} w\left(x_s\right) \log\left[D\left(F\left(x_s\right)\right)\right] \\
&+ \mathbb{E}_{x_t\sim p(x_t)} w\left(x_t\right) \log\left[1 - D\left(F\left(x_t\right)\right)\right] \\
&+ \mathbb{E}_{x_s\sim p(x_s)} w\left(x_s\right) \log\left[1 - D\left(F\left(x_s\right)\right)\right]
\end{aligned} \tag{4}$$

We can find that in the third term, we copy original samples from the source domain rather than using weight strategy in the adversarial alignment. By such a novel adversarial method, we can transform the partial UDA method in the new scenario into the vanilla UDA scenario.

2.2. Attention-Based Method

With the help of the improved adversarial learning mentioned above, we can perform transfer learning by optimizing Eq (4). However, the assumption that Eq (4) achieves significant performance is based on the assumption that all extracted features can be transferred for domain adaptation. Unfortunately, this assumption does not always hold. As we know, adversarial domain adaptation ultimately aims to learn transferable representations across source domain and target domain. However, insufficient extracted features may deceive the domain discriminator, but cannot learn truly both transferable and discriminative of features. One way to address this problem is to introduce attention mechanism to focus transferable features. Recently, coordinate attention [4] propose to capture long-range dependencies along one spatial direction, while retaining accurate position information along another spatial direction. Then, the generated feature maps are respectively encoded into a pair of direction-aware and position-sensitive attention maps that corresponding to image high-level semantics, which can be complementarily applied to the input feature maps to enhance the representation of the object of interest. The framework of transferable attention is the green color based blocks in the Fig 1. We use two spatial extents of pooling kernels $(1, W)$ and $(H, 1)$ to encode each channel along the vertical coordinate and the horizontal coordinate, respectively. Thus we can keep both spatial information globally and positional information.

2.3. Overall Networks and Generalization Bound Analysis

Finally, we integrate all the terms mentioned above to analyze the source and target samples, and derive a unified framework for effective partial UDA. The overall min-max objective is formulated as:

$$\min_{\theta_f,\theta_g} \max_{\theta_d} \mathcal{L}_{cls}(\theta_f,\theta_g) + \lambda_1 \mathcal{L}_{adv}^{new}(\theta_f,\theta_d) + \lambda_2 \mathcal{L}_{ent}(\theta_f,\theta_g) \tag{5}$$

where λ_1 and λ_2 are trade-off hyper-parameters in the training process.

To better understand our work, we present a short theoretical analysis on the widely used Ben-David [5] domain adaptation theory. The distribution of source domain and target domain are denoted as P and Q, respectively. In addition, we denote the distribution of augmented target domain as J. Given the fixed representation space $f = F(x)$ and a family of classifiers R for the source domain in hypothesis space $\mathcal{H}$, we denote $\epsilon_P(R) = \mathbb{E}_{(f,y)\sim P}[R(f) \neq y]$ as the risk of a hypothesis $R \in \mathcal{H}$ with respect to distribution P and the disagreement between hypotheses $R_1, R_2 \in \mathcal{H}$ as $\epsilon_P(R_1, R_2) = \mathbb{E}_{(f,y)\sim P}[R_1(f) \neq R_2(f)]$. In this work, we suppose that the ideal hypothesis is $R^* = \operatorname{argmin}_R \epsilon_P(R) + \epsilon_Q(R)$.

Generalization Error Analysis: According to [5], the real data risk $\epsilon_Q(R)$ of hypothesis R can be bounded by the simulation data risk $\epsilon_P(R)$ and the distribution discrepancy

$$\epsilon_Q(R) \leq \epsilon_P(R) + |\epsilon_P(R, R^*) - \epsilon_Q(R, R^*)| + C_0 \tag{6}$$

where $C_0 = \epsilon_P(R^*) + \epsilon_Q(R^*)$ is a constant. The goal of domain adaptation is to reduce the distribution discrepancy $|\epsilon_P(R, R^*) - \epsilon_Q(R, R^*)|$. Followed by Ben-David theory [5], we can draw the conclusion that discrepancy between source domain and target domain can be upper bounded by discriminator D:

$$\begin{aligned} &|\epsilon_P(R, R^*) - \epsilon_Q(R, R^*)| \\ &\leq \sup_{D\in\mathcal{H}_D} |\mathbb{E}_{h\sim Q_R}[D(h) \neq 0] - \mathbb{E}_{h\sim P_R}[D(h) \neq 0]| \\ &\leq \sup_{D\in\mathcal{H}_D} (|\ \mathbb{E}_{h\sim P_R}[D(h) \neq 0] - \mathbb{E}_{h\sim J_R}[D(h) \neq 0]\ | \\ &+ |\mathbb{E}_{h\sim Q_R}[D(h) \neq 0] - \mathbb{E}_{h\sim J_R}[D(h) \neq 0]|) \end{aligned} \tag{7}$$

The supremum of Eq. 7 consists of two parts. Then following DANN [7] and the analysis in [7], we have the supremum in the following equations:

$$\begin{aligned} |\mathbb{E}_{h\sim P_R}[D(h) \neq 0] - \mathbb{E}_{h\sim J_R}[D(h) \neq 0]| \leq \\ 2 \sup_{D\in\mathcal{H}_D} |D(\mathrm{h}) - 1| \end{aligned} \tag{8}$$

$$\begin{aligned} |\mathbb{E}_{h\sim Q_R}[D(h) \neq 0] - \mathbb{E}_{h\sim J_R}[D(h) \neq 0]| \leq \\ 2 \sup_{D\in\mathcal{H}_D} |D(\mathrm{h}) - 1| \end{aligned} \tag{9}$$

where D is the discriminator, $D(h)$ is maximized by the optimal D. Please note that this conclusion is based on the following assumption: we are free to choose $\mathcal{H}_D$. Fortunately, multilayer neural networks can adapt to any function. As long as the feature representation networks learn domain-invariant feature, the $|\epsilon_P(R, R^*) - \epsilon_Q(R, R^*)|$ tends to decrease during the adversarial learning.

3. EXPERIMENTAL RESULTS

3.1. Setup

Datasets. Office-31 [18] contains images of 31 object classes from three different domains (namely Amazon, DSLR and Webcam). Office-Home [19] is a more challenging transfer learning dataset, which consists of 4 different domains: Art (Ar), Clipart (Cl), Product (Pr), Real-World (Rw). VisDA-2017 [20] is a large-scale dataset for cross-domain object classification which has first present in 2017 Visual Domain Adaptation (VisDA) Challenge. We train adversarial layer and classifier layer by using back-propagation. We adopt min-batch SGD with momentum of 0.9 and the learning rate is adjusted by $\eta_p = \eta_0(1 + \alpha p)^{-\beta}$, where $\eta = 0.01$, α=10, β=0.75, and p changes with the training process from 0 to 1 [10]. For all the tasks, we set λ_1 = 1 for fair comparison.

Table 1. Accuracy of effective domain adaptation tasks on Office-Home (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-50 [13]	46.33	67.51	75.87	59.14	59.94	62.73	58.22	41.79	74.88	67.40	48.18	74.17	61.35
ADDA [14]	45.23	68.79	79.21	64.56	60.01	68.29	57.56	38.89	77.45	70.28	45.23	78.32	62.82
CDAN [10]	47.52	65.91	75.65	57.07	54.12	63.42	59.60	44.30	72.39	66.02	49.91	72.80	60.73
SAN [15]	44.42	68.68	74.60	67.49	64.90	77.80	59.78	44.72	80.07	72.18	50.21	78.66	65.30
MWPDA [16]	55.39	77.53	81.27	57.08	61.03	62.33	68.74	56.42	86.67	76.70	56.67	80.06	68.41
PADA [15]	51.95	67.00	78.74	52.16	53.78	59.03	52.61	43.22	78.79	73.73	56.60	77.09	62.06
ETN [17]	59.24	77.03	79.54	62.92	65.73	75.01	68.29	55.37	84.37	75.72	57.66	**84.54**	70.45
AADA	**60.27**	**79.53**	**86.41**	**64.10**	**73.78**	**82.17**	**72.09**	**54.25**	**86.24**	**76.18**	**61.04**	84.28	**73.36**

Table 2. Accuracy of effective domain adaptation tasks on Office-31 (ResNet-50)

Method	A → D	A → W	D → A	D → W	W → A	W → D	Avg
ResNet-50 [13]	83.44	75.59	83.92	96.27	84.97	98.09	87.05
ADDA [14]	83.41	75.67	83.62	95.38	84.25	99.85	87.03
CDAN [10]	77.07	80.51	93.58	98.98	91.65	98.09	89.98
SAN [15]	94.27	93.90	94.15	99.32	88.73	99.36	94.96
PADA [15]	82.17	86.54	92.69	99.32	95.41	**100.0**	92.69
MWPDA [16]	95.12	96.61	95.02	**100.0**	**95.51**	**100.0**	97.05
ETN [17]	95.03	94.52	96.21	**100.0**	94.64	**100.0**	96.73
AADA	**98.79**	**98.65**	**99.00**	**100.0**	92.34	**100.0**	**98.13**

Table 3. Classification accuracy on ImageNet-Caltech (ResNet-50) and VisDA-2017 (ResNet-101)

Method	I → C	C → I	Avg	R → S	S → R	Avg
ResNet-50 [13]	69.7	71.3	70.5	64.3	45.3	54.8
DAN [21]	71.3	60.1	65.7	68.4	47.6	58.0
DANN [7]	70.8	67.7	69.3	73.8	51.0	62.4
RTN [22]	75.5	66.2	70.9	72.9	50.0	61.5
PADA [15]	75.0	70.5	72.8	76.5	53.5	65.0
AADA	**75.9**	**74.1**	**75.00**	**77.8**	**54.5**	**66.1**

3.2. Empirical Analysis

Ablation study. We show the ablation study in Table 4. AADA (w/ A) denotes that we only use adversarial learning based on target domain augmentation for effective domain adaptation. From the result of Table 4, we find that AADA (w/ A) always performs well for effective domain adaptation task since it transforms the partial UDA problem into the vanilla UDA problem. Secondly, results on all the tasks of the Office-31 demonstrate that our method is better than than DANN, which show the effectiveness of our AADA method.

Convergence Performance. As shown in Fig. 2 (a), we study the convergence performance of the proposed methods for A → W. We demonstrate the different convergence of AADA (w/A), AADA and DANN. It is easy to see that loss curve of AADA is smoother, which indicates that the coordinate attention mechanism is effective to the proposed method.

Table 4. Ablation study on Office-31 (ResNet-50)

Method	A → D	A → W	D → A	D → W	W → A	W → D	Avg
DANN	85.61	78.63	83.60	97.28	85.07	99.37	88.26
AADA (w/ A)	93.63	95.91	94.89	**100.00**	**94.78**	**100.0**	96.54
AADA	**98.79**	**98.13**	**99.00**	**100.00**	92.34	**100.0**	**98.30**

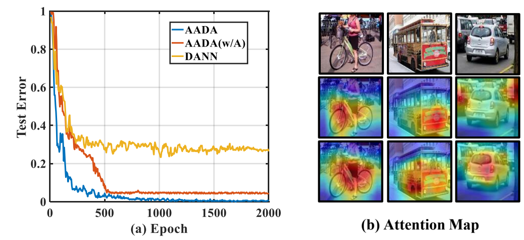

Fig. 2. (a) Convergence analysis on task A→W. (b) Attention Visualization of the last convolutional layer of different models on the VisDA-2017

Attention Visualization. It is believed that the general feature representation should be able to minimize the difference in cross-domain distribution while retaining specific attributes in the domain. In order to clarify our point intuitively, we use Grad-CAM visualization [23] to generate attention maps of DANN (the second row) and AADA (the third row) in Fig. 2 (b). Although DANN can focus on the desirable regions for some difficult samples, such as cars in the third column and plants in the sixth column, DANN cannot accurately focus on the desirable regions. Compared with DANN, AADA effectively solves the problem that DANN cannot focus on some difficult samples.

4. CONCLUSION

In this paper, we present an effective attention-based adversarial partial domain adaptation(AADA), which represents the effort in this more practical situation for UDA. AADA addresses two key problems. The first one is to handle the problem that the label space of target domain is much smaller than that of the source domain. We leverage adversarial learning to augment target domain by using original source domain to make it as the source label space. The second one is that existing adversarial learning mainly concentrates on global feature representation, while we take coordinate attention to focus on the local region features of interest to capture vision semantics. Extensive experimental results on several standard cross-domain datasets have demonstrated that AADA outperforms some competitive domain methods by a large margin. In the future, we plan to further explore domain invariance for domain adaptation under multiple environments.

5. REFERENCES

[1] Vladimir N Vapnik, “An overview of statistical learning theory,” *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[2] Mengzhu Wang, Wei Wang, Baopu Li, Xiang Zhang, Long Lan, Huibin Tan, Tianyi Liang, Wei Yu, and Zhigang Luo, “Interbn: Channel fusion for adversarial unsupervised domain adaptation,” in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 3691–3700.

[3] Mengzhu Wang, Xiang Zhang, Long Lan, Wei Wang, Huibin Tan, and Zhigang Luo, “Improving unsupervised domain adaptation by reducing bi-level feature redundancy,” *arXiv preprint arXiv:2012.15732*, 2020.

[4] Qibin Hou, Daquan Zhou, and Jiashi Feng, “Coordinate attentcoordinateion for efficient mobile network design,” *arXiv preprint arXiv:2103.02907*, 2021.

[5] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan, “A theory of learning from different domains,” *Machine Learning*, vol. 79, no. 1, pp. 151–175, 2010.

[6] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, “Generative adversarial networks,” *arXiv preprint arXiv:1406.2661*, 2014.

[7] Yaroslav Ganin and Victor Lempitsky, “Unsupervised domain adaptation by backpropagation,” in *International Conference on Machine Learning*, 2015, pp. 1180–1189.

[8] Ajay Kumar Tanwani, “Domain-invariant representation learning for sim-to-real transfer,” *CoRR*, vol. abs/2011.07589, 2020.

[9] Jing Zhang, Zewei Ding, Wanqing Li, and Philip Ogunbona, “Importance weighted adversarial nets for partial domain adaptation,” in *Computer Vision and Pattern Recognition*, 2018, pp. 8156–8164.

[10] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan, “Conditional adversarial domain adaptation,” *Neural Information Processing Systems*, pp. 1647–1657, 2017.

[11] Yves Grandvalet, Yoshua Bengio, et al., “Semi-supervised learning by entropy minimization.,” in *CAP*, 2005, pp. 281–296.

[12] Dongting Sun, Mengzhu Wang, Xurui Ma, Tianming Zhang, Nan Yin, Wei Yu, and Zhigang Luo, “A focally discriminative loss for unsupervised domain adaptation,” in *International Conference on Neural Information Processing*. Springer, 2021, pp. 54–64.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[14] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell, “Adversarial discriminative domain adaptation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7167–7176.

[15] Zhangjie Cao, Lijia Ma, Mingsheng Long, and Jianmin Wang, “Partial adversarial domain adaptation,” in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 135–150.

[16] Jian Hu, Hongya Tuo, Chao Wang, Lingfeng Qiao, Haowen Zhong, and Zhongliang Jing, “Multi-weight partial domain adaptation.,” in *British Machine Vision Conference*, 2019, p. 5.

[17] Zhangjie Cao, Kaichao You, Mingsheng Long, Jianmin Wang, and Qiang Yang, “Learning to transfer examples for partial domain adaptation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2985–2994.

[18] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell, “Adapting visual category models to new domains,” in *European Conference on Computer Vision*, 2010, pp. 213–226.

[19] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan, “Deep hashing network for unsupervised domain adaptation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5018–5027.

[20] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko, “Visda: The visual domain adaptation challenge,” *CoRR*, vol. abs/1710.06924, 2017.

[21] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan, “Learning transferable features with deep adaptation networks,” in *International Conference on Machine Learning*, 2015, pp. 97–105.

[22] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan, “Unsupervised domain adaptation with residual transfer networks,” *Advances in Neural Information Processing Systems*, pp. 136–144, 2016.

[23] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *IEEE International Conference on Computer Vision*, 2017, pp. 618–626.