CENet: Consolidation-and-Exploration Network for Continuous Domain Adaptation

Chi Zhang[1], Yalu Cheng[1,2], Pengxu Wei[3], Hongliang He[1,2], and Jie Chen[1,2,*]
[1]School of Electronic and Computer Engineering, Peking University, Shenzhen, China
[2]Peng Cheng Laboratory, Shenzhen, China
[3]Sun Yat-sen University, Guangzhou, China
{conor.zhangch,hehl}@pku.edu.cn, yalucheng@stu.pku.edu.cn
weipx3@mail.sysu.edu.cn, chenj@pcl.ac.cn

Abstract

Unsupervised Domain Adaptation (UDA) deals with transferring knowledge from labeled source domains to an unlabeled target domain under domain shift. However, this does not reflect the breadth of scenarios that arise in real-world applications since source domains could increase. A plausible conjecture is: can we train a lifelong learning model learned on continuous source domains given the target without the presence of labels? We formalize this task as the Continuous Domain Adaptation (CDA) and empirically show that conventional domain adaptation methods may suffer severe generalization deterioration due to the limited incremental transferability and negative transfer. To tackle this issue, we propose a novel sample-to-sample framework—Consolidation-and-Exploration Network (CENet) to facilitate incremental transferring. This method underscores both the qualitative and quantitative relationship between samples. Moreover, we conduct comprehensive experiments to evaluate the effectiveness of each component in our pair-based method. Extensive experiments show that our approach achieves significant improvement over related state-of-the-art methods. Our source code will be publicly available at https://github.com/GekFreeman/continuous_da.

1. Introduction

Unsupervised Domain Adaptation (UDA) has been widely explored to mitigate the domain shift between labeled source domains and the unlabeled target domain [1, 25, 30]. It specifically transfers the domain knowledge to the target domain from a single source domain (*i.e.*, Single-Source Domain Adaptation (SSDA)) or multiple domains (*i.e.*, Multi-Source Domain Adaptation (MSDA)). Nevertheless, existing UDA research works conventionally assume that all the data from source domains are pre-collected well for training UDA models. This inevitably fails to cope with a more practical scenario: the data from different source domains are sequentially collected, or it is realistically intractable to adapt the model for the target domain with all the available source data due to security issues and data privacy.

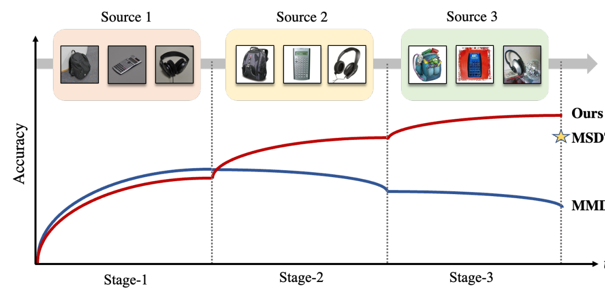


Figure 1. The adaptation performance on the unlabeled target domain of MMD [30] (*for UDA*), MSDTR [5] (*for MSDA*) and our CENet (*for CDA*) at different training stages. Given the same source domains, MMD and our CENet are trained sequentially with different labelled source domains at each training stage, respectively; MSDTR is trained in a MSDA manner with all the source domains together.

Recently, there have been a few works that attempt to transfer from continuous source domains [15, 22, 23]. However, their proposed problems are not scalable from the standard UDA and lack comparability with SSDA and MSDA methods. In this paper, we consider a practical continuous UDA problem (see Fig. 1): can we train a model to learn from the crescent source domains adapting to the unlabelled target domain? We refer to this realistic setting as Continuous Domain Adaptation (CDA). CDA sequentially receives different labeled source domains and is increasingly trained with one source domain at once. For instance, medical resources are unbalanced in the world, we can easily collect and train on the labeled domain from areas with rich medical resources continuously to adapt to the unlabeled target domain with resource-poor health facilities.

In the CDA task, the issue of *adaptation drift* arises remarkably due to the sensitivity to the learning source data. As shown in Fig. 1, MMD [30] (the UDA method) exhibits a declining adaptation trend when sequentially trained with different source domains. Due to the absence of annotations in the target domain, the supervision information of MMD is mainly determined by labels of the source domain and the domain discrepancy distance dominated by the source resulting in overfitting of the source domain. MS-DTR [5], trained in a MSDA manner, under-performs our CENet (trained in a CDA manner), although it has all the source domains in the whole training stage and costs more computation for training with so many data. The reason is that applying whole different source domains bring about the discrepancy of distinct modal information.

To mitigate the adaptation drift issue of CDA, we propose a Consolidation-and-Exploration Network (CENet). To prevent results in a tug-of-war dynamic, we propose a pair-based alternative termed Contrastive Pair (CP) rather than domain-based. CP aims to establish sample-to-sample connections in domain and category aspects based on target samples. However, due to the lack of annotation and domain shift, the noisy feature may lead to blunders in the relationship. Therefore, we first sieve the target derived from the sample-to-prototype distance to accrue the relational bank. Next, we use the embedding after the encoder and projector to establish an association map of the source domain, and relational samples based on the Hierarchical Navigable Small World (HNSW) [8]. We construct a contrastive pair of *convergence* (CP_c) that represents the correlation of semblable samples between different domains. We realize the exploration of the source domain by constraining the diagonal elements of the cross-correlation matrix of CP_c in the feature space to make their feature vectors as similar as possible. Besides, we propose another contrastive pair of *divergence* (CP_d) using the class discrepancy within the relational bank. The cross-correlation matrix of CP_d is constrained to make its eigenvectors orthogonal, thus preserving the reliable discriminative features learned from the historical source domain and avoiding negative transfer during exploration.

Instead of alignments between distributions or prototypes, we propose a sample-to-sample alignment to address the problem of incremental matching in continuous domain adaptation. We also propose a symmetric form of differentiation constraint to consolidate the reliable features of the historical source domain to avoid the negative transfer. We design class- and domain-based supervised relations for the task compared to classical contrastive learning methods. Ultimately, our method achieves desirable results on this new problem. Compared to MSDA methods, our method can be applied in more rigorous data usage scenarios, e.g., to learn the source domain sequentially without episodic memory. At the same time, our method only requires a light model to implement instead of designing domain-specific network for each domain. We outperform current state-of-the-art MSDA methods on multiple benchmarks.

2. Related Work

Domain Adaptation. Many works on UDA focus on the adaptation from single source domain, while Multi-Source Domain Adaptation (MSDA) attracts increasingly attentions to leverage abundant labelled data from multiple sources. MSDA originated from A-SVM [25] leverages the ensemble of source-specific classifiers to tune the target categorization model. Motivated by the distribution weighted combining rule, some methods [13, 24, 30] learn domain-specific classifier modules and obtain a weighted ensemble prediction for target samples. Besides, another MSDA strategy is prototype-based for sample-level domain alignment [12, 20, 26]. SRDC [20] propose to directly uncover the intrinsic target discrimination via discriminative clustering of the target data. Recently, self-training has emerged as a simple and effective technique for UDA, attaining state-of-the-art performance on many image recognition tasks [28]. However, when these UDA methods are applied to CDA, their network structures need to be expanded with the increase of source domains and may even lead to adaptation drift due to the sensitivity to the learning source data. Compared with [26], we further enrich the connections between samples and apply them to the exploration and consolidation representation learning tasks in CDA.

Multi-Domain Continual Learning. Multi-Domain Continual Learning (MDCL) is one of the most important scenarios in continual learning. MDCL is concerned with learning a task, such as image classification, sequentially over multiple visual domains with the different label spaces. Progressive Neural Networks [17], Dynamically Expandable Networks (DENs) [27], and Deep Adaptation Modules (DAMs) [16] are the earliest works in this field on the classification task. [9, 10] learn a domain-specific binary mask over a fixed backbone architecture to get a compact and memory-efficient solution. Other works [14, 15] mitigate forgetting by using parameter-isolated-based approaches to dedicate a domain-specific subset of parameters to each unique task. The learning objective in the CDA scenario is to enhance the adaptation to a given target domain with higher labeling costs. Since we are not concerned with catastrophic forgetting of the source domains, CDA is not required to set up any domain-specific modules to preserve the knowledge of the source domains. Besides, this setting

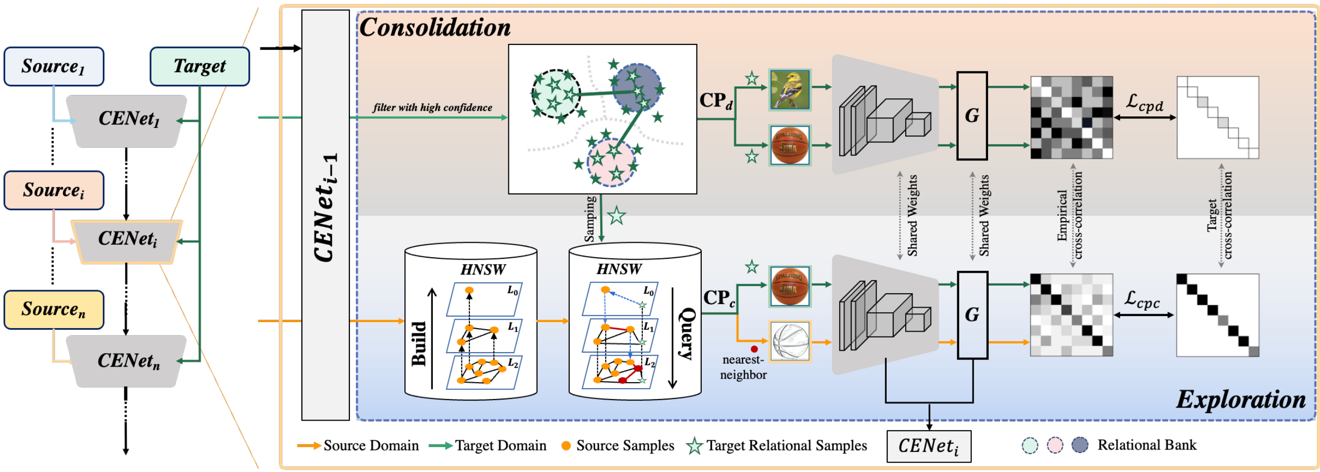


Figure 2. The architecture of the proposed CENet framework. (1) **Consolidation**: We use the margin and semantic association of the target domain in the prior model CENet$_{i-1}$ to accrue Relational Bank, which includes the information to be preserved. Furthermore, we construct contrastive pairs of divergence between banks of different objects and maintain the objects' variance by equating the diagonal elements of the cross-correlation matrix to zero. (2) **Exploration**: Based on the consolidation task, we propose to explore more generalizable feature representations by establishing a sample-to-sample alignment. We introduce the Hierarchical Navigable Small World (HNSW) to build the feature relation graph of the source domain samples. Then, we assemble contrastive pairs of convergence through the samples of the relational bank querying the HNSW and try to equate the diagonal elements of the cross-correlation matrix to 1, resulting in eliminating domain shift and exploring more adaptive feature representations.

is a natural extension of SSDA and MSDA, and one of the core problems is to deal with the domain shift between the labeled domain and the unlabeled domain, which does not need to be concerned in MDCL. Moreover, our setting can be directly compared with a broad range of UDA methods for a specific target domain, which more readily reflects the effectiveness of our method.

3. Method

3.1. Problem Setup

In CDA, n source domains are labeled and sequentially given: $D^{S_1}, D^{S_2}, ..., D^{S_n}$; the target domain D^T is unlabelled. They share the same class space: $\{l_0, l_1, ..., l_{V-1}\}$. In the n-th source domain D^{S_n}, it has N_n images X^{S_n} with labels Y^{S_n}, denoted as (X^{S_n}, Y^{S_n})=$\{(x_i^{S_n}, y_i^{S_n})\}_{i=1}^{N_n}$. Target domain D^T is X^T=$\{x_i^T\}_{i=1}^{N_T}$, where N_T is the number of samples in D^T. F is the feature extractor and G means the projector.

3.2. Consolidation-and-Exploration Network

Domain shift is also the core problem of continuous domain adaptation. However, methods based on distribution alignment may cause negative transfer of the target domain under continuous alignment. This paper proposes a continuous representation learning based on sample relations (*relational bank*) to endow the model with the ability of incremental transfer. In order to enhance the generalization performance of the model in the target domain without causing negative transfer, we correspondingly set two representation learning tasks of exploration and consolidation. Similar to contrastive learning, we take input pairs of related samples in some way, which is called a contrastive pair (CP). For exploring and consolidating the representation learning of the generalization of the target domain, we refine CPs to two kinds: cross-domain and cross-class, respectively.

Relational Bank. Unlike self-supervised representation learning, our CENet learns feature representations that are robust to specific domains through the association of samples between categories and domains. The target domain sample is the core of all sample relations. However, due to the inevitable noise generated by the lack of labeling, we filter unlabeled images with high confidence τ and select some samples with the closest distance to the class prototype to construct the *relational bank*, which is further used to construct differentiated sample relations for learning robust representation to continuous learning.

Consolidation. The objective of the consolidation task is to preserve reliable class-discriminative information in the current network to prevent negative transfer while exploring novel domain knowledge. Therefore, we propose the contrastive pair of divergence (CP$_d$) based on samples of diverse pseudo-classes in the relational bank. Moreover, learning the differential feature representation of CP$_d$ plays an essential role in disentangling the domain attribute out of the domain distribution and balancing the proportion of various classes. Defining a CP_d, i.e., $\langle x_i^T, x_j^T \rangle$ with pseudo labels $\langle \hat{y}_i \neq \hat{y}_j \rangle$. We set the symmetric representation learning objective as $\mathcal{L}_{cpd}$:

$$\mathcal{L}_{cpd} = \sum_i \mathbb{1}(\hat{y}_i^T \neq \hat{y}_j^T) C_{ii}(x_i^T, x_j^T)^2 \tag{1}$$

where C is the cross-correlation matrix computed between the outputs of the two identical networks along the batch dimension:

$$z = G(F(x)), \tag{2}$$

$$C_{ij}(z, z') = \frac{\sum_b z_{b,i} {z'}_{b,j}}{\sqrt{\sum_b (z_{b,i})^2} \sqrt{\sum_b ({z'}_{b,j})^2}} \tag{3}$$

where b indexes batch samples and i, j index the vector dimension of the networks' output. C is a square matrix with size same as the dimensionality of the network's output, and with values comprised between -1 (i.e. perfect anti-correlation) and 1 (i.e. perfect correlation). Intuitively, the $\mathcal{L}_{cpd}$ objective, by trying to equate the diagonal elements of the cross-correlation matrix to zero, makes the embedding orthogonal to the difference of labels.

Exploration. The exploration task aims to learn information from the source domain to improve the feature representation of hard samples. Recent work [2] in domain adaptation find that not all knowledge is transferable across domains, and indiscriminate transfer may be detrimental to the generalization. So it is necessary to pay attention to the relevance of samples cross domains: the adaptation of some samples in the target domain depends more on relevant samples of source domains than others. This paper proposes to construct CP_c to describe the partial connection between source and target domains, which is used to obtain class-unique feature cross domains.

CP_c consists of relational samples and their similar source domain samples. There are different domains and similar categories between samples of CP_c. To build the relational map of samples, we introduce the Hierarchical Navigable Small World (HNSW) [8], which is built by feature maps of the current source. HNSW builds a multi-layer structure incrementally consisting of a hierarchical set of proximity graphs for nested subsets of the source elements. The minimum layer in which an element is present is selected randomly with an exponentially decaying probability distribution. Starting search from the upper layer together with utilizing the scale separation boosts the performance allows a logarithmic complexity scaling. We query HNSW for its most similar top K points for each relational sample in the bank. Then we sample one point according to the probability with the relational sample to constitute the CP_c of this round. The probability of sampling is determined by the distance in the map as follows:

$$p(x_i^S | x^T) = \frac{\frac{1}{\exp^{d(x^T, x_i^S)}}}{\sum_j^K \frac{1}{\exp^{d(x^T, x_j^S)}}} \tag{4}$$

where $d(x^T, x_i^S)$ means the distance between the sample x^T and x_i^S in the built HNSW. When we get the CP_c as (x^T, x^S), we define the representation learning objective as $\mathcal{L}_{cpc}$:

$$\mathcal{L}_{cpc} = \sum_i p(x^S | x^T)(1 - C_{ii}(x^T, x^S))^2 \tag{5}$$

Unlike $\mathcal{L}_{cpd}$, the $\mathcal{L}_{cpc}$ objective, by trying to equate the diagonal elements of the cross-correlation matrix to 1, makes the embedding invariant to the domain shift, leading to the generalizable feature space.

Finally, we further simply use the $(x_i^T, \hat{y}_i^T)$ to calculate the cross-entropy loss as L_{cls}. Finally, our representational learning objective is as follows:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{cpc} + \lambda_2 \mathcal{L}_{cpd} + \mathcal{L}_{cls} \tag{6}$$

4. Experiment

To validate the effectiveness of our method, we compare our CENet against state-of-the-art domain adaptation methods on three datasets: *Office-31*, *Image-CLEF*, and *Office-Caltech*.

4.1. Experimental Settings

Dataset. *Office-31* [18] is widely used as a benchmark for domain adaptation, and it consists of three different domains with 31 categories: Amazon (A) with 2,817 images, Webcam (W) with 795 images and DSLR (D) with 498 images. *Image-CLEF* [7] is a benchmark dataset for ImageCLEF2014 domain adaptation challenge, which is organized by selecting 1800 images meanwhile as three domains: Caltech-256 (C), ImageNet ILSVRC 2012 (I), and Pascal VOC 2012 (P). The *Office-Caltech* [4] dataset consists of four different domains: Amazon (A), Caltech (C), DSLR (D), and Webcam (W).

Implementation Details. To realize a general and simple application, we adopt ResNet-50 as our backbone for all datasets. We set τ=0.95, K=80% $\times$ N_t, where N_t is the number of samples in the target domain. λ_1=λ_2=1.0 for all the experiments. In optimization trajectory, we use the SGD optimizer with the initial learning rate $\alpha = 0.01$ which is adjusted using the following formula: $lr = \frac{\alpha}{(1+u*p)^q}$, where p is the training progress linearly changing from 0 to 1. $u = 0.01$. $q = 0.70$ and $\alpha = 10$. As for HNSW, we set the dimension as 256. Moreover, the graph's maximum number of outgoing connections is set to 16. We choose the squared L_2 as the distance metric for querying.

We not only follow the same three evaluation protocols ((1)-(3)) as MFSAN [30] but also add the fourth (4): (1) Single Best (SB): we report the best single source transfer results among multiple source domains. (2) Source Combine (SC): all source domains are combined into a traditional single source. (3) Multi-Source (MS): training data from multiple sources are available and used simultaneously. (4) Continuous Source (CS): multiple sources are learned in continuous ways where the historical domains are not accessible and we report results on each domain and their average for different datasets in Tab. 1, Tab. 2, and Tab. 3.

4.2. Main Results

For all datasets, we take each domain as the target domain and remaining domains as source domains to train the model continuously. After all training trajectories, we compare the accuracy of the target domain. Since there is no published method for the CDA scenario, we extend A2LP [28] and MFSAN [30] to the CS protocol for comparison. We show that the existing UDA methods cannot maintain their excellent performance on SSDA and MSDA when applied to CDA. Using the same source domain set, MFSAN [30] drops from 90.2% (MS) to 89.5% (CS), for instance.

CENet surpasses these UDA methods in all benchmarks under the CS protocol. Counterintuitively, our method even outperforms the MS protocol under the CS protocol. We regarded the MS protocol as the upper limit of the CS due to the widespread domain-specific modules and looser data access constraints (all source domain data can be trained simultaneously). Nevertheless, our method employs a lighter and fully shared network to learn multiple source domains continuously without any data replay and achieves significant improvement over the MS protocol's methods. Taking the *Office-31* dataset as an example, the average target accuracy of CENet is 92.0%, which exceeds the 91.1% accuracy of the state-of-the-art method MSDTR [29] in the MS protocol. Moreover, our method dramatically outperforms other methods, using ResNet-50 as the backbone in the *Office-Caltech* dataset. Compared with MOST [11] which take the ResNet-101 as the backbone, our method also achieves very similar accuracy with a much more lightweight network.

Experiments show that our proposed CENet effectively solves the retrogressive adaptability of existing UDA methods in CS protocol. More importantly, exceeding the prior MS works, our method can provide novel insight into the multi-source domain adaptation. Recent work mainly approximates the target domain by mixing the distributions of multiple source domains, but is this the optimal paradigm for MSDA? The excellent performance of CENet on CDA may provide further understanding for future exploration.

Table 1. Classification accuracy (%) on *Office-31* dataset. The method with the highest accuracy on the given target is emphasized in bold. Our method achieves 92.0% average accuracy outperforming the others.

Protocols	Models	D	W	A	Avg
SB	Source Only [30]	99.3	96.7	62.5	86.2
	RevGrad [3]	99.1	96.9	68.2	88.1
	DAN [6]	99.5	96.8	66.7	87.7
	D-CORAL [19]	99.7	98.0	65.3	87.7
SC	DAN [6]	99.6	97.8	67.6	88.3
	D-CORAL [19]	99.3	98.0	67.1	88.1
	RevGrad [3]	99.7	98.1	67.6	88.5
MS	DCTN [24]	99.3	98.2	64.2	87.2
	SImpAl$_{50}$ [21]	99.2	97.4	70.6	89.0
	MFSAN [30]	99.5	98.5	72.7	90.2
	MSCLDA [5]	99.8	98.8	73.7	90.8
	MSDTR [29]	99.7	98.3	75.2	91.1
CS	MFSAN [30]	99.8	97.6	71.0	89.5
	CENet(Ours)	**100.0**	**99.8**	**76.1**	**92.0**

Table 2. Classification accuracy(%) on *Image-CLEF* dataset for multi-source unsupervised domain adaptation. The method with the highest accuracy on the given target is emphasized in bold. Our method achieves 90.7% accuracy.

Protocols	Models	P	C	I	Avg
SB	Source Only [30]	74.8	91.5	83.9	83.4
	RevGrad [3]	75.0	96.2	87.0	86.1
	DAN [6]	75.0	93.3	86.2	84.8
	D-CORAL [19]	76.9	93.6	88.5	86.3
	A2LP [28]	79.3	96.3	91.8	89.1
SC	DAN [6]	77.6	93.3	92.2	87.7
	D-CORAL [19]	77.1	93.6	91.7	87.5
	RevGrad [3]	77.9	93.7	91.8	87.8
MS	DCTN [24]	75.0	95.7	90.3	87.0
	SImpAl$_{50}$ [21]	77.5	93.3	91.0	87.3
	MFSAN [30]	79.1	95.4	93.6	89.4
	MSCLDA [5]	79.5	95.9	**94.3**	89.9
CS	MFSAN [30]	78.2	95.9	93.9	89.3
	A2LP [28]	73.3	94.3	90.2	85.9
	CENet(Ours)	**81.2**	**96.7**	**94.3**	**90.7**

4.3. Ablation Study

We conduct thorough ablation studies with *Office-31* dataset to demonstrate the effectiveness of each component in our model. We remove CP$_c$ or CP$_d$ individually with full CENet. As shown in Tab. 4, experiments have proved that both CP$_c$ and CP$_d$ can bring improvements no matter whether they work alone or together.

Based on the *Image-CLEF* dataset, we measure the average accuracy under all possible order settings, which is still close to the state-of-the-art performance in MSDA. In addition, we compare the performance of CENet and the classical DA method in preventing negative transfer during

Table 3. Classification accuracy (%) on *Office-Caltech* dataset. The method with the highest accuracy on the given target is emphasized in bold. Our method achieves 98.1% accuracy and surpasses all the methods with ResNet-50. MOST_{101} take the ResNet-101 as the backbone, and we gain a comparable performance with a lighter network that is easier to employ.

Protocols	Models	W	D	C	A	Avg
SC	Source Only [30]	99.0	98.3	87.8	86.1	92.8
	DAN [6]	99.3	98.2	89.7	94.8	95.5
MS	DAN [6]	99.5	99.1	89.2	91.6	94.8
	DCTN [24]	99.4	99.0	90.2	92.7	95.3
	SImpAl_{50} [21]	99.3	99.8	92.2	95.3	96.7
	MSCLDA [5]	99.1	98.5	94.1	95.3	96.8
	MOST_{101} [11]	**100.0**	**100.0**	**96.0**	**96.4**	**98.1**
CS	MFSAN [30]	**100.0**	97.8	92.5	94.2	96.1
	A2LP [28]	99.0	99.4	94.9	95.8	97.3
	CENet(Ours)	**100.0**	**100.0**	95.9	**96.4**	**98.1**

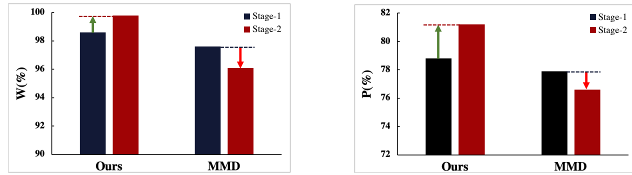
Table 4. Ablation study on *Office-31* dataset. "✓" means with this operation. $\rightarrow$ means the continuous learning sequence, $\Rightarrow$ means the adaptation to the target domain. We report the accuracy (%) on both one domain adaptation setting and two domain CDA settting.

CP_c	CP_d	$(\text{W}) \Rightarrow \text{A}$	$(\text{W} \rightarrow \text{D}) \Rightarrow \text{A}$
		73.1	74.8
✓		72.8	75.9
	✓	73.6	74.9
✓	✓	**73.8**	**76.1**

continuous domain adaptation. We take the domain closest to the real-world scenario as the target domain T. The learning sequence of the sources is set to the order of decreasing similarity with the target domain, which significantly increases the possibility of negative transfer. We use Maximum Mean Discrepancy (MMD), which made great success in many domain adaptation tasks. As shown in Fig. 3, the traditional method exhibits noticeable negative transfer under this challenging learning sequence. That is, the performance of the target domain degenerates with the increment of source domains. On the other hand, CENet achieves the desired goal of maintaining incremental transfer learning capabilities in the case of diminishing domain similarity.

5. Conclusion

In this paper, we propose a practical Continuous Domain Adaptation (CDA) task for real-world scenarios and devise a Consolidation-and-Exploration Network (CENet) to address the domain drift issue in CDA. To endow the model with incremental transferability, CENet utilizes the prior knowledge of the model to construct contrastive pairs for memory consolidation and adaptability exploration, respectively. Based on the differential connections of these samples, we design a representation learning objective based on the cross-correlation matrix, which can acquire further domain adaptability in the current source domain while preserving the reliable priors in the model. CENet effectively mitigates the problem of adaptation drift of existing UDA methods with lighter structure and higher computational efficiency. Under the strict constraints of model structure and data usage, our performance even exceeds state-of-the-art MSDA methods, which also provides new insights for domain adaptation.



(a) Office-31. (b) Image-CLEF.

Figure 3. We take the W domain as the target of the *Office-31* dataset, and the learning sequence of the sources is (D→W). The P domain is the target of *Image-CLEF*, and the corresponding order is (I→C). The red arrow indexes the negative transfer, which means target accuracy deteriorates after learning a new source. The green arrow means the incremental transfer. Our method performs powerful incremental transferability.

Acknowledgments This work is supported by the Nature Science Foundation of China (No.61972217, No.62081360152), Natural Science Foundation of Guangdong Province in China (No.2019B1515120049, 2020B1111340056).

References

[1] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *NIPS*, 2016. 1

[2] Xinyang Chen, Sinan Wang, Bo Fu, Mingsheng Long, and Jianmin Wang. Catastrophic forgetting meets negative transfer: Batch spectral shrinkage for safe transfer learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. 4

[3] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *ICML*, 2015. 5

[4] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *CVPR*, 2012. 4

[5] Keqiuyin Li, Jie Lu, Hua Zuo, and Guangquan Zhang. Multi-source contribution learning for domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 2021. 1, 2, 5, 6

[6] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *ICML*, 2015. 5, 6

[7] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *ICML*, 2017. 4

[8] Yu A Malkov and Dmitry A Yashunin. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE TPAMI*, 42(4):824–836, 2018. 2, 4

[9] Arun Mallya, Dillon Davis, and Svetlana Lazebnik. Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 67–82, 2018. 2

[10] Massimiliano Mancini, Elisa Ricci, Barbara Caputo, and Samuel Rota Bulo. Adding new tasks to a single network with weight transformations using binary masks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018. 2

[11] Tuan Nguyen, Trung Le, He Zhao, Quan Hung Tran, Truyen Nguyen, and Dinh Phung. Most: multi-source domain adaptation via optimal transport for student-teacher learning. In Cassio de Campos and Marloes H. Maathuis, editors, *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pages 225–235. PMLR, 27–30 Jul 2021. 5, 6

[12] Yingwei Pan, Ting Yao, Yehao Li, Yu Wang, Chong-Wah Ngo, and Tao Mei. Transferrable prototypical networks for unsupervised domain adaptation. In *CVPR*, 2019. 2

[13] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *ICCV*, 2019. 2

[14] Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Learning multiple visual domains with residual adapters. *Advances in neural information processing systems*, 30, 2017. 2

[15] Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Efficient parametrization of multi-domain deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8119–8127, 2018. 1, 2

[16] Amir Rosenfeld and John K. Tsotsos. Incremental learning through deep adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42:651–663, 2020. 2

[17] Andrei A Rusu, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. Progressive neural networks. *arXiv preprint arXiv:1606.04671*, 2016. 2

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

[19] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pages 443–450. Springer, 2016. 5

[20] Hui Tang, Ke Chen, and Kui Jia. Unsupervised domain adaptation via structurally regularized deep clustering. In *CVPR*, 2020. 2

[21] Naveen Venkat, Jogendra Nath Kundu, Durgesh Kumar Singh, Ambareesh Revanur, et al. Your classifier can secretly suffice multi-source domain adaptation. In *NeurIPS*, 2020. 5, 6

[22] Riccardo Volpi, Diane Larlus, and Grégory Rogez. Continual adaptation of visual representations via domain randomization and meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4443–4453, 2021. 1

[23] Hao Wang, Hao He, and Dina Katabi. Continuously indexed domain adaptation. *arXiv preprint arXiv:2007.01807*, 2020. 1

[24] Ruijia Xu, Ziliang Chen, Wangmeng Zuo, Junjie Yan, and Liang Lin. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *CVPR*, 2018. 2, 5, 6

[25] Jun Yang, Rong Yan, and Alexander G Hauptmann. Cross-domain video concept detection using adaptive svms. In *ACM MM*, 2007. 1, 2

[26] Jeongbeen Yoon, Dahyun Kang, and Minsu Cho. Semi-supervised domain adaptation via sample-to-sample self-distillation. *arXiv preprint arXiv:2111.14353*, 2021. 2

[27] Jaehong Yoon, Eunho Yang, Jeongtae Lee, and Sung Ju Hwang. Lifelong learning with dynamically expandable networks. *arXiv preprint arXiv:1708.01547*, 2017. 2

[28] Yabin Zhang, Bin Deng, Kui Jia, and Lei Zhang. Label propagation with augmented anchors: A simple semi-supervised learning baseline for unsupervised domain adaptation. In *European Conference on Computer Vision*, pages 781–797. Springer, 2020. 2, 5, 6

[29] Lihua Zhou, Mao Ye, Dan Zhang, Ce Zhu, and Luping Ji. Prototype-based multisource domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 2021. 5

[30] Yongchun Zhu, Fuzhen Zhuang, and Deqing Wang. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. *CoRR*, abs/2201.01003, 2022. 1, 2, 5, 6