Burn After Reading: Online Adaptation for Cross-domain Streaming Data

Luyu Yang[1], Mingfei Gao[2], Zeyuan Chen[2], Ran Xu[2], Abhinav Shrivastava[1], Chetan Ramaiah[3*]

[1]University of Maryland
[2]Salesforce Research
[3]Google
{loyo, abhinav}@cs.umd.edu, {mingfei.gao, zeyuan.chen, ran.xu}@salesforce.com, cramaiah@google.com

Abstract. In the context of online privacy, many methods propose complex security preserving measures to protect sensitive data. In this paper we note that: not storing any sensitive data is the best form of security. We propose an online framework called "Burn After Reading", *i.e.* each online sample is permanently deleted after it is processed. Our framework utilizes the labels from the public data and predicts on the unlabeled sensitive private data. To tackle the inevitable distribution shift from the public data to the private data, we propose a novel unsupervised domain adaptation algorithm that aims at the fundamental challenge of this online setting–the lack of diverse source-target data pairs. We design a **Cro**ss-**Do**main **Bo**otstrapping approach, named **CroDoBo**, to increase the combined data diversity across domains. To fully exploit the valuable discrepancies among the diverse combinations, we employ the training strategy of multiple learners with co-supervision. CroDoBo achieves state-of-the-art online performance on four domain adaptation benchmarks. *Code is available here*

Keywords: domain adaptation, online learning, privacy preserving

1 Introduction

With the onslaught of the pandemic, the internet has become an even more ubiquitous presence in all of our lives. Living in an enormous web connecting us to each other, we now face a new reality: it is very hard to escape one's past on the Internet since every photo, status update, and tweet lives forever in the cloud [54,13]. Moreover, recommender systems that actively explore the user data [16,82] for data-driven algorithms have brought controversy that the right to privacy is more important than the convenience. Fortunately, we have *the Right to Be Forgotten* (RTBF), which gives individuals the right to ask organizations to delete their personal data. Recently, many solutions [83,88] have been proposed that try to preserve privacy in the context of deep learning, mostly focused on

* Work was done at Salesforce.

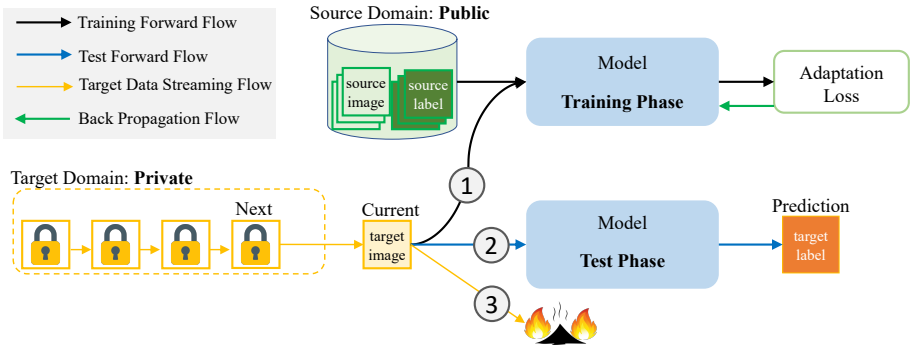

Fig. 1. The data-flow of the proposed **Burn After Reading** framework at one iteration. The iteration contains a training and a test phase. In training phase, the model takes labeled data from the public source domain, and the current unlabeled target data from the private target domain. The model updates based on the adaptation loss and then moves to test phase. After prediction, the current target data is permanently **deleted** from the target domain. Each target data is (1) trained (2) tested (3) deleted. *Best viewed in color.*

the Federated Learning [75,24]. Federated Learning allows asynchronous update of multiple nodes, in which sensitive data is stored only on a few specific nodes. However, recent studies [87,23,79] show that private training data can be leaked through the gradients sharing mechanism deployed in distributed models. In this paper, we argue that: *not storing any sensitive data is the best form of security.*

The best form of security requires us to delete the user data after use, which necessitates an online framework. However, existing online learning frameworks [56,34] cannot meet this need without addressing the distribution shift from public data, *i.e.* source domain, to the private user data, *i.e.* target domain. Therefore, in this paper we propose an online domain adaptation framework in which the target domain streaming data is deleted immediately after adapted. We name the framework "Burn After Reading", as illustrated in Figure 1. The task that is seemingly an extended setting of unsupervised domain adaptation (UDA), however, cannot simply be solved by the online implementation of the offline UDA methods. We explain the reason with a comprehensive analysis of the existing domain adaptation methods. To begin with, existing offline UDA methods rely heavily on the rich combinations of cross-domain mini-batches that gradually adjust the model for adaptation [59,29,67,80,48,85,70,62,77,73,76], which the online streaming setting cannot afford to provide. In particular, many domain adversarial-based methods [72,27,19,3] depend on a slowly annealing adversarial mechanism that requires discriminating large number of source-target pairs to achieve the adaptation. Recently, state-of-the-art offline methods [25,32,33] show promising results by exploiting target-oriented clustering, which requires an offline access to the entire target domain. Therefore, the online UDA task needs new solutions to succeed at scarcity of the data from target domain.

We aim straight at the most fundamental challenge of the online task—the lack of diverse cross-domain data pairs—and propose a novel algorithm based on cross-domain bootstrapping for online domain adaptation. At each online query, we increase the data diversity across domains by bootstrapping the source domain to form diverse combinations with the current target query. To fully exploit the valuable discrepancies among the diverse combinations, we train a set of independent learners to preserve the differences. Inspired by [81], we later integrate the knowledge of learners by exchanging their predicted pseudo-labels on the current target query to co-supervise the learning on the target domain, but without sharing the weights to maintain the learners' divergence. We obtain more accurate prediction on the current target query by an average ensemble of the diverse expertise of all the learners. We call it **CroDoBo**: **Cr**oss-**Do**main **Bo**otstrapping for online domain adaptation, an overview of CroDoBo pipeline is shown in Figure 3.

We conduct extensive evaluations on our method, including the classic UDA benchmark *VisDA-C* [49], a practical medical imaging benchmark *COVID-DA* [84] and the large-scale distribution shift benchmark *WILDS* [26] subset *Camelyon*. Moreover, we propose a new adaptation scenario in this paper from *Fashion-MNIST* [78] to *DeepFashion* [35]. On all the benchmarks, our method outperforms the state-of-the-art UDA methods that are eligible for the online setting. Further, without the reuse of any target sample, our method achieves comparable performance to the offline setting. We summarize the contributions as follows.

- To our best knowledge, we are the first to propose an online domain adaptation framework to implement *the right to be forgotten*.
- We study the fundamental drawback of the online setting compared to offline–the lack of data diversity, and designed a novel online domain adaptation method that improves, and exploits the data diversity.
- Our proposed algorithm achieves new state-of-the-art online results on four challenging benchmarks.
- Although designed for online setting, our method yields comparable performance to the offline setting, suggesting that it is a superior choice even just for time efficiency.

2 Related Work

The Right to Be Forgotten [69,13,46,15], also referred to as *right to vanish*, *right to erasure* and *courtesy vanishing*, is the right given to each individual to ask organizations to delete their personal data. RTBF is part of the General Data Protection Regulation (GDPR). As a legal document, the GDPR outlines the specific circumstances under which the right applies in *Article 17* GDPR [1]. The first item is: *The personal data is no longer necessary for the purpose an organization originally collected or processed it.* Yet, the exercise of this right

[1] Article 17 GDPR - Right to be forgotten
https://gdpr.eu/article-17-right-to-be-forgotten/

has become a thorny issue in applications. Politou *et al.* [50] discussed that the technical challenges of aligning modern systems and processes with the GDPR provisions are numerous and in most cases insurmountable. In the context of machine learning, Villaronga *et al.* [69] addressed that the core issue of the AI and Right to Be Forgotten problem is the dearth of interdisciplinary scholarship supporting privacy law and regulation. Graves *et al.* [15] proposed three defense mechanisms against a general threat model to enable deep neural networks to forget sensitive data while maintaining model efficacy. In this paper, we focus on how to obtain model efficacy while erasing data online to protect the user's right to be forgotten.

Online Adaptation to Shifting Domains was first investigated in Signal Processing [9] and later studied in Natural Language Processing [8] and Vision tasks [52,22,40,42,6]. Jain *et al.* [22] assumed the original classifier output a continuous number of which a threshold gives the class, and reclassify points near the original boundary using a Gaussian process regression scheme. The procedure is presented as a Viola-Jones cascade of classifiers. Moon *et al.* [42] proposed a four-stage method by assuming a transformation matrix between the source subspace and the mean-target subspace embedded in the Grassmann manifold. The method is designed for handcrafted features. In the context of deep neural network, one transformation matrix might not be sufficient to describe the correlation between source and target deep representations [44]. Taufique *et al.* [66] approached the task by selectively mixing the online target samples with those that were saved in a buffer. Since in [66] the approach relies on saved target samples, it is not applicable to the "Burn After Reading" framework.

Active Domain Adaptation [53,39,4,51] also benefits the online learning of shifting domains. But it has a different setting: the target domain can actively acquire labeled data online. Rai *et al.* [53] presented an algorithm that harnessed the source domain data to learn a initializer hypothesis, which is later used for active learning on the target domain. Ma *et al.* [39] allowed a small budget of target data for the categories that appeared only in target domain and presented an algorithm that jointly trains two sub-networks of different learning strategies. Chen *et al.* [4] proposed an algorithm that can adaptively deal with interleaving spans of inputs from different domains by a tight trade-off that depends on the duration and dimensionality of the hidden domains. The active acquisition of target labels is not feasible for the unsupervised domain adaptation, thus is beyond the scope of this paper.

Test-Time Domain Adaptation [68,71,65] is another related task. Similar to the "burn after reading", test-time DA also aims at a fast adaptation to the target samples. Differently, test-time DA is motivated by the unavailability of the source domain [71], which is a variant of source-free domain adaptation [32]. Thus, it is based on a continual setting. Meanwhile, test-time domain adaptation does not require target samples being deleted after training, although Wang *et al.* [71] and Sun *et al.* [65] both discussed the extension to an online setting in the experiments. Without the access to source samples, Varsavsky *et al.* [68] leverages a combination of adversarial learning and consistency under augmentation.

Sun *et al.* [65] exploits the self-supervision with auxiliary rotation prediction. In this paper, we compare with test-time DA with a devised continual version of our method in the supplementary.

Ensemble Methods for Online Learning [1,41] such as bagging and boosting have shown advantages handling *concept drift* [38] and class imbalance, which are common challenges in the online learning task. MinKu *et al.* [41] addressed the importance of ensemble diversity to improve accuracy in changing environments and proposed the measurement of ensemble diversity. Han *et al.* [17] proposed a regularization for online tracking with a subset of branches in the neural network that are randomly selected. Although online learning and online domain adaptation share similar streaming form of data input, the two tasks face fundamentally different challenges. For online learning, the challenge is to select the most trustworthy supervisions from the streaming data by differentiating the informative vs. misleading data points, also known as the *stability-plasticity dilemma* [21]. However, for online domain adaptation (our task), the streaming data of target domain naturally comes unlabeled, and the challenge is the scarcity of supervision. Thus the goal is how to maximize the utilization of the supervision from a different but related labeled source domain.

3 Approach

In this section, we introduce the proposed method for "Burn After Reading" framework, in which the samples from the public source domain are fully accessible, while only one/a batch of the target samples is available at each iteration. The model "reads" the current target data, updates, then predicts, after which the target data is deleted permanently from the target domain. In Sec 3.1 we describe the difference between online and the offline setting. In Sec 3.2, we first introduce the cross-domain bootstrapping strategy and the theoretical insights behind. Then we describe the details of the co-supervision.

3.1 Offline vs. Online

Given the labeled source data $D_{\mathcal{S}} = \{(s_i, y_i)\}_{i=1}^{N_S}$ drawn from the source distribution $p_s(x, y)$, and the unlabeled target data $D_{\mathcal{T}} = \{t_i\}_{i=1}^{N_T}$ drawn from the target distribution $p_t(x, y)$, where N_S and N_T represent the number of source and target samples, both offline and online adaptation aim at learning a classifier that make accurate predictions on $D_{\mathcal{T}}$. The *offline adaptation* assumes access to every data point in both $D_{\mathcal{S}}$ and $D_{\mathcal{T}}$, synchronous [12,59,62,67] or domain-wise asynchronous [32]. The inference on $D_{\mathcal{T}}$ happens after the model is trained on both $D_{\mathcal{S}}$ and $D_{\mathcal{T}}$ entirely. Differently for *online adaptation*, we assume the access to the entire $D_{\mathcal{S}}$, while the data from $D_{\mathcal{T}}$ arrives in a streaming data of random mini-batches $\{T_j = \{t_b\}_{b=1}^{B}\}_{j=1}^{M_T}$. B is the batch size and M_T is the total number of target batches. Each mini-batch T is first adapted, tested and then erased from $D_{\mathcal{T}}$ without replacement, as shown in Figure 1 and 3. We refer each online batch of target data as a target *query*.

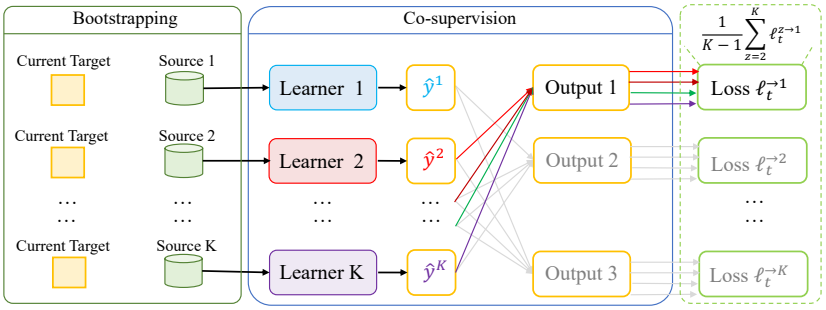

Fig. 2. Illustration of computing co-supervision loss ($\ell_t^{z\to k}$ in Eq. 4), taking $\ell_t^{\to 1}$ for example. The co-supervision for learner 1 is from the *other* K-1 learners. The current target data is repeatedly paired with each bootstrapped source data to improve data diversity. Each learner takes a unique data combination and generates pseudo-label $\hat{y}^k$ of the current target data. Then $\ell_t^{\to 1}$ receives co-supervision averagely from the pseudo-labels $\{\hat{y}^2, \hat{y}^3, ..., \hat{y}^K\}$.

The fundamental challenge of our online task is the limited access to the training data at each inference query, compared to the offline task. For generality, we can assume there are 10^3 source and target batches, respectively. In an offline setting, the model is tested after training on at most 10^6 combinations of source-target data pairs, while in an online setting, an one-stream model can see at most $10^3 + 500$ combinations at the 500-th query. Undoubtedly, the online adaptation faces a significantly compromised data diversity. The training process of our task suffers from two major drawbacks: (I) The model is prone to underfitting on target domain due to the lack of seen target samples, especially at the early stage of training. (II) Due to the deletion of previous data, the model lacks the diverse combinations of source-target data pairs that enable the deep network to find the optimal cross-domain classifier [30].

The goal of the proposed method is to minimize the two drawbacks of the online setting. We first propose to increase the data diversity by cross-domain bootstrapping, and we preserve the discrepancy in independently trained learners. Then we fully exploit the valuable discrepancies of these learners by exchanging their expertise on the current target query to co-supervise each other.

3.2 Proposed Method

Cross-domain Bootstrapping for Data Diversity The diversity of cross-domain data pairs is crucial for most prior offline methods [12,48,59] to succeed. Since the target samples cannot be reused in the online setting, we propose to increase the data diversity across domains by bootstrapping the source domain to form diverse combinations with the current target domain query, as shown in Figure 2. Specifically, for each target query T_j, we randomly select a set of K mini-batches $\{S_j^k = \{(s_b)_{b=1}^B\}\}_{k=1}^K$ of the same size from the source domain with replacement. Correspondingly, we define a set of K base learners $\{\boldsymbol{w}^k\}_{k=1}^K$.

At each iteration, a learner $\boldsymbol{w}^k$ makes prediction for query T_j after trained on $\{T_j, S_j^k\}$, and updates via

$$\begin{aligned} \boldsymbol{w}^k \leftarrow\ & \boldsymbol{w}^k - \eta \left(\nabla \mathcal{L}(\boldsymbol{w}^k, \{T_j, S_j^k\})\right), \\ p_j^k =\ & p\left(c|T_j; \boldsymbol{w}^k\right), \end{aligned} \tag{1}$$

where η is the learning rate, c is the number of classes, p_j^k is the predicted probability by the k-th learner, and $\mathcal{L}(,)$ is the objective function. The predicted class for T_j is the average of K predictions of the base learners. We justify our design choice from the perspective of uncertainty estimation in the following discussion.

Theoretical Insights As mentioned in Sec. 3.1, we aim at the best estimation of the current target query. We first consider a single learner situation. At the j-th query, the learner faces a fundamental trade-off: by minimizing the uncertainty of the j-th query, the learner can attain the best current estimation. Yet the risk of fully exploring the uncertainty is to spoil the existing knowledge from the previous j-1 target domain queries. However, if we don't treat the uncertainty, the single observation on j-th query is less informative for current query estimation. Confronting the dilemma, we should not ignore that the uncertainty captures the variability of a learner's posterior belief which *can* be resolved through statistical analysis of the appropriate data [45]. This gives us hope for a more accurate model via uncertainty estimation. One popular suggestion for resolving uncertainty is to use *Dropout* [10,11,58] sampling, where individual neurons are independently set to zero with a probability. As a sampling method on the neurons, *Dropout* works in a similar form of *bagging* [74,60] of multiple decision trees. It might equally reduce the overall noise of the network regardless of domain shift but it does not address the problem of our task, which is the lack of diverse cross-domain combinations.

Alternatively, we employ another pragmatic approach *Bootstrap* for uncertainty estimation on the target domain that offsets the source dominance. With the scarcity of target samples, we propose to bootstrap source-target data pairs for a more balanced cross-domain simulation. At high-level, the bootstrap simulates multiple realizations of a specific target query given the diversity of source samples. Specifically, the bootstrapped source approximate a distribution over the current query T_j via the bootstrap.

The bootstrapping brings multi-view observations on a single target query by two means. First, given K sampling subsets from $D_{\mathcal{S}}$, let $\mathcal{F}$ be the ideal estimate of T_j, $\hat{\mathcal{F}}$ be the practical estimate of the dataset, and $\hat{\mathcal{F}}^*$ be the estimate from a bootstrapped source paired with the target query, $\hat{\mathcal{F}}^* = K^{-1}\sum_{k=1}^{K}\hat{\mathcal{F}}_k^*$ will be the average of the multi-view K estimates. Second, besides the learnable parameters, the *Batch-Normalization* layers of K learners generate result in a set of different means and variances $\{\mu_k, \sigma_k\}_{k=1}^{K}$ that serve as K different initializations that affects the learning of $\hat{\mathcal{F}}^*$.

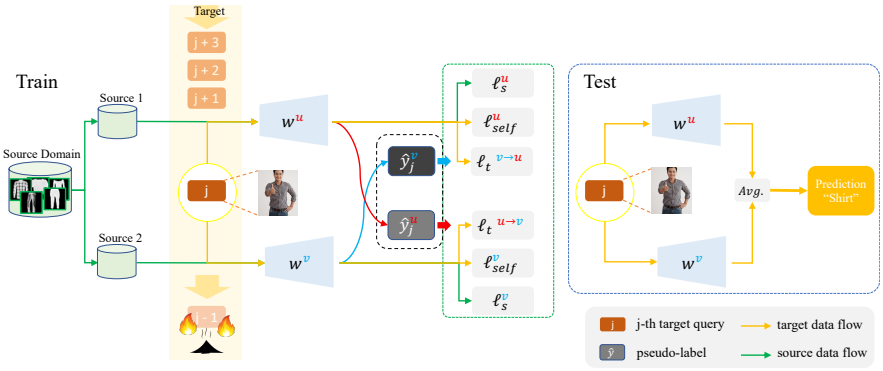

Fig. 3. The full pipeline of the proposed CroDoBo K=2 method at j-th iteration. Only one target query j is currently available from target domain in this iteration. We bootstrap the source domain and combine with the current j-th query. The learners w^u (k=1) and w^v (k=2) exchange the generated pseudo-labels $\hat{y}_j^u$ and $\hat{y}_j^v$ as co-supervision. Each learner is updated by a supervised loss ℓ_s on source data, a self-supervised loss ℓ_{self} on the target data and a co-supervised loss ℓ_t. The test result is recorded by averaging the predictions of both learners. Once tested, query j is immediately deleted.

Exploit the Discrepancies via Co-supervision After the independent learners have preserved the valuable discrepancies of cross-domain pairs, the question now is how to fully exploit the discrepancies to improve the online predictions on the target queries. On one hand, we want to integrate the learners' expertise into one better prediction on the current target query, on the other we hope to maintain their differences. Inspired by [81], we train the K learners jointly by exchanging their knowledge on the target domain as a form of co-supervision. Specifically, the K learners are trained independently with bootstrapped source supervision, but they exchange the pseudo-labels generated for target queries. We followed the *FixMatch* [63] to compute pseudo-labels on the target domain. We first consider K=2 for simplicity, we denote the learners as $\boldsymbol{w}^u$ for $k = 1$ and $\boldsymbol{w}^v$ for $k = 2$, respectively.

Given the current target query T_j, the loss function $\mathcal{L}$ consists a supervised loss term ℓ_s from the source domain with the bootstrapped samples, and a self-supervised loss term ℓ_t from the target domain with pseudo-labels $\hat{y}_b$ from the peer learner, as illustrated in Figure 3. We denote the cross-entropy between two probability distributions as $\mathcal{H}(;)$. Thus, the co-supervision objective ℓ_t is obtained via:

$$\ell_t^{v\to u} = B^{-1}\sum_{b=1}^{B} \mathbb{1}\left(p_b^v \geq \tau\right) \mathcal{H}\left(\hat{y}_b^v; p(c|\tilde{t}_b; \boldsymbol{w}^u)\right),$$
$$\ell_t^{u\to v} = B^{-1}\sum_{b=1}^{B} \mathbb{1}\left(p_b^u \geq \tau\right) \mathcal{H}\left(\hat{y}_b^u; p(c|\tilde{t}_b; \boldsymbol{w}^v)\right), \tag{2}$$

p_b^u and p_b^v are the predicted probabilities of t_b by $\boldsymbol{w}^u$ and $\boldsymbol{w}^v$, respectively. τ is the threshold for pseudo-label selection, and $\tilde{t}_b$ is a strongly-augmented version of t_b using *Randaugment* [5]. However, we note that *RandAug* is a technique only employed to increase data diversity, but is **not** required for CroDoBo. We denote the version without any augmentation as CroDoBo, and we denote the version with *RandAug* as CroDoBo+.

To further exploit the supervision from the limited target query, from p_b^u and p_b^v we compute a self-supervised loss $\ell_{\text{self}} = \ell_{\text{ent}} + \lambda \ell_{\text{div}}$, in which ℓ_{ent} is standard entropy and ℓ_{div} is a balancing term for class-diversity, λ is a weighting factor. The ℓ_{self} is widely used in prior domain adaptation works [70,57,32]. Finally, we update the learners by

$$\begin{aligned} \boldsymbol{w}^u &\leftarrow \boldsymbol{w}^u - \eta(\ \nabla \ell_s(\boldsymbol{w}^u, S_j^u) + \nabla \ell_t^{v \to u} + \nabla \ell_{\text{self}}(\boldsymbol{w}^u, T_j)), \\ \boldsymbol{w}^v &\leftarrow \boldsymbol{w}^v - \eta(\ \nabla \ell_s(\boldsymbol{w}^v, S_j^v) + \nabla \ell_t^{u \to v} + \nabla \ell_{\text{self}}(\boldsymbol{w}^v, T_j)). \end{aligned} \tag{3}$$

For $K > 2$, each learner $\boldsymbol{w}^k$ is updated with the co-supervision from the other $K-1$ learners (Figure 2), weighted by $1/(K-1)$ for each $\ell_t^{z \to k}$ (z is the learner's index other than k). We update $\boldsymbol{w}^k$ by

$$\boldsymbol{w}^k \leftarrow \boldsymbol{w}^k - \eta(\nabla \ell_s(\boldsymbol{w}^k, S_j^u) + \frac{1}{K-1} \sum_{z=1}^{K-1} \nabla \ell_t^{z \to k} + \nabla \ell_{\text{self}}(\boldsymbol{w}^k, T_j)). \tag{4}$$

4 Experiments

We consider two metrics for evaluating online domain adaptation methods: *online average accuracy* and *one-pass accuracy*. We provide formulations for our metrics. Given a target sequence $D_{\mathcal{T}} = \{T_0, T_1, ..., T_j, ..., T_{N_T}\}$, the online model at time j is $\boldsymbol{w}_j$. The test accuracy on the current T_j is acc_j, then the *online accuracy* $\text{ACC}_{\text{online}} = \frac{1}{N_T} \sum_{j=1}^{N_T} \text{acc}_j(T_j; \boldsymbol{w}_j)$ is the average of the entire streaming accuracies. Once the model finishes online update, we freeze the weights of $\boldsymbol{w}_{N_T}$ and compute *one-pass accuracy* $\text{ACC}_{\text{one-pass}} = \frac{1}{N_T} \sum_{j=1}^{N_T} \text{acc}(T_j; \boldsymbol{w}_{N_T})$. *One-pass* measures the model's generalizability to the entire target domain. We keep track of this metric in case the model keeps overfitting to the new target data only to achieve high *online accuracy* [43]. A *one-pass accuracy* much lower than *online average* indicates that the model might have overfitted to the fresh queries, but compromised its generalization ability to the early queries.

Dataset. We use **VisDA-C** [49], a classic benchmark adapting from synthetic images to real. We followed the data split used in prior offline settings [49,32,59]. We also use **COVID-DA** [84], adapting the CT images diagnosis from common pneumonia to the novel disease. This is a typical scenario where online domain adaptation is valuable in practice. When a novel disease breaks out, without any prior knowledge, one has to exploit a different but correlated domain to assist the diagnosis of the new pandemic in a time-sensitive manner. We also evaluate on a large-scale medical dataset *Camelyon17* from the **WILDS** [26], a histopathology image datasets with patient population shift from source to the

target. Camelyon17 has 455k samples of breast cancer patients from 5 hospitals. Another practical scenario is the online fashion where the user-generated content (UGC) might be time-sensitive and cannot be saved for training purposes. Due to the lack of cross-domain fashion prediction dataset, we propose to evaluate adapting from **Fashion-MNIST** [78]-to-**DeepFashion** [35] category prediction branch. We select 6 fashion categories shared between the two datasets, and design the task as adapting from $36,000$ grayscale samples of Fashion-MNIST to $200,486$ real-world commercial samples from DeepFashion.

Implementation details. We implement using Pytorch [47]. We follow [33,32] to use ResNet-101 [18] on VisDA-C pretrained on ImageNet [7,55]. We follow [84] to use pretrained ResNet-18 [18] on COVID-DA. We follow the leader-board on WILDS challenge [26] [2] to use DenseNet-121 [20] on Camelyon17 with random initialization, we use the official WILDS codebase (v1.1.0) for data split and evaluation. We use pretrained ResNet-101 [18] on Fashion-MNIST-to-DeepFashion. Our target query batch-size and bootstrapped source batch-size are both set as 64. The confidence threshold $\tau = 0.95$ and diversity weight $\lambda = 0.4$ are fixed throughout the experiments. Our method is not sensitive to hyperparameters, the results are reported in supplementary.

Baselines. We compare **CroDoBo** without data augmentation and **CroDoBo**$^+$ with *RandAug* with eight state-of-the-art domain adaptation approaches, including **DAN** [36], **CORAL** [64], **DANN** [12], **ENT** [14,57], **MDD** [85], **CDAN** [37], **SHOT** [32] and **ATDOC** [33]. ATDOC has multiple variants of the auxiliary regularizer, we compared with the *Neighborhood Aggregation* (ATDOC-NA) with the best performance in [33]. Among the compared approaches, SHOT and ATDOC-NA require a memory module that collects and stores information of all the target samples, thus only apply the offline setting. For the other six approaches, we compare both offline and online results. Each offline model is trained for 10 epochs. Each online model is trained batch-by-batch for 1 epoch, during which the online test results are recorded after each model update. All the online baselines take the same randomly-perturbed target queries to make a fair comparison. The results of CroDoBo and CroDoBo+ reported in Table 1-4 have 2 learners (*i.e.* K=2), the results with $K \geq 3$ are reported in Table 7.

Main results. We summarize the results on VisDA-C [49] in Table 1, and plot the online results in Figure 4 We follow [49,25,32,33] to provide the VisDA-C one-pass accuracy in class average. In Table 1: Online, the proposed CroDoBo largely outperforms other baselines. Without augmentation, our method outperforms the second by 11.5%. Our online result is on par with the state-of-the-art offline performance ATDOC-NA [32], outperforming many other offline baselines.

Comparing across the offline and online setting, the Source-Only baseline drops 2.4% in the online average and 7.2% in the one-pass accuracy, which indicates that the data diversity is also important in domain generalization. We observe that ENT [57], which is an entropy regularizer on the posterior probabilities of the unlabeled target samples, has a noticeable performance drop in the online setting, and illustrates more obvious imbalanced results over the categories

[2] https://wilds.stanford.edu/leaderboard/

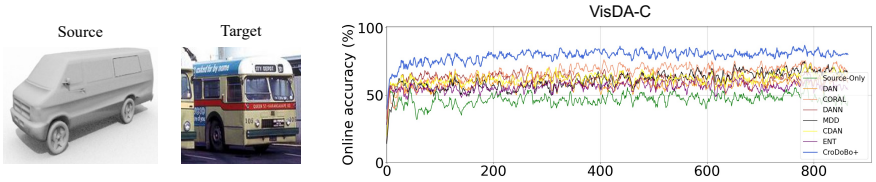

Fig. 4. Results of online adaptation from synthetic source domain to real target domain on VisDA-C [49] with "Burn After Reading". The x-axis is the online streaming timestep. Each query contains 64 samples. Each approach takes the same randomly perturbed sequence of target queries. Source-Only is in green, the proposed CroDoBo is in blue. Smoothed with 1-D uniform filter with length=5. *Best viewed in color.*

(superior at class "knife" but poor at "person" and "truck"). We consider it a typical example of bad objective choice for the online setting when the dataset is imbalanced. Without sufficient rounds to provide data diversity, entropy minimization might easily overfit the current target query. The 2.5% drop in one-pass from online further confirmed the model has deviated from the beginning.

Results on two medical imaging datasets COVID-DA [84] and *WILDS*-Camelyon17 [26] are respectively summarized in Table 2 and Table 3.The online streaming accuracy is presented in Figure 5. COVID-DA* is the method proposed along with the dataset in [84], which is a domain adversarial-based multi-classifier approach with focal loss regularization. Our method outperforms the other approaches on COVID-DA regarding the online and one-pass metric, and achieves competitive performance against the best offline accuracy. On the large-scale benchmark *WILDS*-Camelyon17, our CroDoBo is on par the the best offline result, and CroDoBo+ outperforms the offline results by 1.7%,

Table 1. Accuracy on VisDA-C (%) using ResNet-101. In the online setting, individual class reports accuracy after one-pass, *one-pass* is the class average. Best offline (***italic bold***), best online (**bold**).

	Methods (Syn → Real)	plane	bike	bus	car	horse	knife	motor	person	plant	skate	train	truck	Online	One-pass	Per-Class Acc.
Offline	Source-Only	67.7	27.4	50.0	61.7	69.5	13.7	85.9	11.5	64.4	34.4	84.2	19.2	-	-	49.1
	DAN [36]	84.4	50.9	68.4	66.8	82.0	17.0	82.3	22.0	73.3	47.4	81.2	18.3	-	-	57.8
	CORAL [64]	94.7	46.8	78.0	62.4	86.5	70.1	90.4	73.5	84.2	34.9	87.7	24.9	-	-	69.5
	DANN [12]	81.9	77.7	***82.8***	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	-	-	57.4
	ENT [57]	88.6	29.5	82.5	***75.8***	88.7	16.0	***93.2***	63.4	***94.2***	40.1	87.3	12.1	-	-	64.3
	MDD [85]	89.2	58.9	70.5	54.5	71.1	42.9	78.8	22.5	68.6	54.7	88.6	15.4	-	-	59.6
	CDAN [37]	89.4	40.3	74.6	65.2	81.5	62.2	90.1	69.3	73.3	58.6	84.8	19.1	-	-	67.4
	SHOT [32]	94.3	***88.5***	80.1	57.3	93.1	94.9	80.7	***80.3***	91.5	89.1	86.3	58.2	-	-	82.9
	ATDOC-NA [33]	***95.3***	84.7	82.4	75.6	***95.8***	***97.7***	88.7	76.6	94.0	***91.7***	***91.5***	***61.9***	-	-	***86.3***
Online	Source-Only	73.3	6.5	44.9	67.8	58.6	5.7	67.2	18.3	47.7	19.2	84.1	9.3	46.7	41.9	-
	DAN [36]	87.7	45.9	69.9	70.9	77.4	17.7	80.7	18.6	79.9	29.9	82.7	16.6	57.8	56.5	-
	CORAL [64]	94.7	51.0	79.6	63.2	88.2	69.4	**91.1**	73.1	87.7	41.8	88.4	24.2	66.7	71.0	-
	DANN [12]	84.5	39.2	70.2	60.4	77.1	28.6	90.9	20.5	67.7	39.9	**89.8**	10.5	49.0	56.6	-
	ENT [57]	87.1	14.8	87.9	71.9	87.8	**98.9**	90.3	0.0	5.2	15.0	80.4	0.2	55.8	53.3	-
	MDD [85]	**95.1**	52.2	87.9	57.9	90.3	94.8	88.4	45.7	76.2	50.5	77.7	25.7	60.4	70.1	-
	CDAN [37]	88.5	44.3	74.3	68.4	80.3	60.2	89.9	69.9	74.3	57.1	84.8	13.9	62.3	67.1	-
	CroDoBo (ours)	93.7	76.4	86.3	77.4	92.5	94.0	90.8	77.6	90.1	88.4	85.4	**37.7**	77.9	82.5	-
	CroDoBo^{+}(ours)	94.8	**87.5**	**90.5**	**76.0**	**94.9**	93.7	88.7	**80.1**	**94.8**	**89.4**	84.6	30.7	**79.4**	**84.0**	-

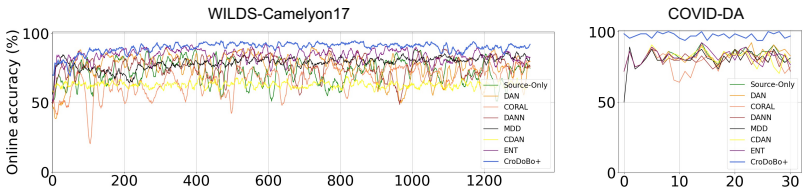

Fig. 5. Results of online accuracy on *WILDS*-Camelyon17 [26] with hospital patient population shift, and COVID-DA [84] adapting from common pneumonia to COVID-19 medical images with "Burn After Reading". Source-Only is in green, the proposed CroDoBo is the solid blue line. Smoothed with 1-D uniform filter with length=5 for *WILDS*-Camelyon17.

which validates the effectiveness of the approach. The good performance on larger number of target queries indicates that CroDoBo can well exploit the underlying information from the target domain. Similar observations are made on the large-scale Fashion benchmark [78,35]. Meanwhile, we reprint *Domain Generalization* results from the *WILDS* leaderboard for reference.

Results on large-scale Fashion dataset, from Fashion-MNIST [78] to DeepFashion [35] category prediction branch, is summarized in Table 4. We provide the online results in Figure 6. To the best of our knowledge, we are the first to report results on this meaningful adaptation scenario. The offline Source-Only merely achieves 23.1% accuracy, only 6.5% gain on the basis of the probability of guessing, which indicates the benchmark is challenging. The sharp drop of performance from Source-Only online accuracy to one-pass accuracy (-6.8%) indicates the large domain gap, and how easy the model is dominated by the source domain supervision. Similar observation is made on *WILDS*-Camelyon17

Table 2. Offline and online accuracy (%) on COVID-DA [84], adaptation from pneumonia to Covid. All the baselines use ResNet-18 as the backbone. COVID-DA* is the method proposed in [84] along with dataset.

Methods (Pneumonia → Covid)		Online	One-pass	Offline
Offline & Online	Source-Only	83.6	82.0	88.9
	DAN [36]	84.4	85.7	87.7
	CORAL [64]	67.6	45.4	65.4
	DANN [12]	83.0	87.1	87.7
	ENT [57]	84.3	87.3	89.8
	MDD [85]	83.2	86.2	81.0
	CDAN [37]	83.0	86.4	86.3
	SHOT [32]	-	-	93.2
	ATDOC-NA [33]	-	-	***98.1***
	COVID-DA* [84]	-	-	***98.1***
	CroDoBo (ours)	95.0	**97.1**	-
	CroDoBo$^+$(ours)	**96.5**	**97.1**	-

Table 3. Accuracy on *WILDS*-Camelyon17 [26] (%) using DenseNet-121. *Domain Generalization* results are reprinted from *WILDS* leaderboard (see Footnote 2).

Methods (Hospital 1,2,3 → Hospital 5)		Online	One-pass	Offline
Domain Generalization	ERM [26]	-	-	70.3
	Group DRO [26]	-	-	68.4
	IRM [26]	-	-	64.2
	FISH [61]	-	-	74.7
Offline & Online	Source-Only	71.7	60.1	63.6
	DAN [36]	76.3	78.0	69.0
	CORAL [64]	66.0	87.1	85.0
	DANN [12]	76.4	81.4	86.7
	ENT [57]	83.1	82.3	***87.5***
	MDD [85]	77.8	52.5	63.7
	CDAN [37]	62.7	60.1	58.5
	SHOT [32]	-	-	73.8
	ATDOC-NA [33]	-	-	86.3
	CroDoBo (ours)	87.5	89.2	-
	CroDoBo$^+$(ours)	**89.2**	**91.9**	-

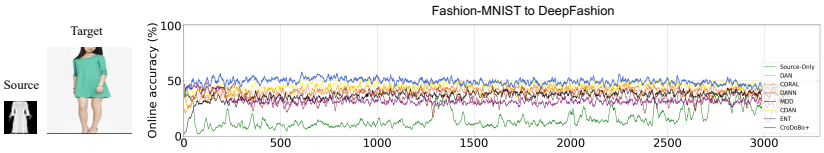

Fig. 6. Results of online adaptation from Fashion-MNIST [78]to DeepFashion [35] with "Burn After Reading". Smoothed with 1-D uniform filter with length=10.

Source-Only results(-11.6% from online to one-pass), this usually happens when the source domain is less challenging than the target domain, and the distribution of the two domains are far from each other. Faced with this challenging benchmark, CroDoBo improves the online performance to a remarkable 49.1%,

Table 4. Results on Fashion-MNIST [78] to DeepFashion [35] (%) using ResNet-101.

Methods (F-MNIST → DeepFashion)		Online	One-pass	Offline
Offline & Online	Source-Only	22.7	15.8	23.1
	DAN [36]	40.7	42.0	32.7
	CORAL [64]	40.4	40.7	39.6
	DANN [12]	35.6	26.5	40.5
	ENT [57]	31.9	31.2	31.1
	MDD [85]	36.5	38.0	39.0
	CDAN [37]	45.4	**47.6**	47.2
	SHOT [32]	-	-	42.3
	ATDOC-NA [33]	-	-	***47.4***
	CroDoBo (ours)	47.6	**47.6**	-
	CroDoBo$^+$ (ours)	**49.1**	46.3	-

Table 5. Ablation study of cross-domain bootstrapping on four datasets (%). VisDA-C one-pass accuracy is in per-class. Number of learners $K = 2$ in both w/ CRODOBO and w/o CRODOBO.

Method/Dataset		VisDA-C	COVID-DA	Camelyon17	Fashion
Online	w/o CroDoBo	78.5	94.4	86.2	42.3
	w/ CroDoBo	79.4	96.5	89.2	49.1
One-pass	w/o CroDoBo	84.0	97.1	89.4	39.9
	w/ CroDoBo	84.0	97.1	91.9	46.3

Table 6. Ablation study on the objectives on target domain on VisDA-C (%). T is the sharpening temperature in the MixMatch [2].

Method	Online	One-pass
default (w/o CroDoBo, τ=0.95, λ=0.4)	78.5 (-)	84.0 (-)
w/o ℓ_{ent}	63.7(↓)	53.1(↓)
w/o ℓ_{div}	72.6(↓)	73.0(↓)
replace $\ell_{\text{ent}} + \ell_{\text{div}}$ w/ Pseudo-labeling [28] (τ=0.95)	70.2(↓)	70.0(↓)
replace $\ell_{\text{ent}} + \ell_{\text{div}}$ w/ MixMatch [2] (T=0.5)	73.0(↓)	75.3(↓)
replace ℓ_t w/ MixMatch [2] (T=0.5)	76.3(↓)	81.5(↓)
use *Randaug* [5] on $\ell_{\text{ent}}, \ell_{\text{div}}$	77.6(↓)	83.7(↓)

Table 7. Accuracy on VisDA-C (%) using ResNet-101 with different number of learners K, and comparing the computation speed reported using 2 NVIDIA-P6000 GPUs.

CroDoBo$^+$	plane	bike	bus	car	horse	knife	motor	person	plant	skate	train	truck	Online	One-pass	samples/sec
$K = 2$	94.8	87.5	90.5	76.0	94.9	93.7	88.7	80.1	94.8	89.4	84.6	30.7	79.4	84.0	25
$K = 3$	95.0	85.6	84.2	73.3	94.4	95.7	88.5	82.2	94.4	83.4	89.3	36.6	79.2	83.5	16
$K = 4$	95.5	85.0	85.0	76.1	95.3	96.0	92.7	81.8	92.7	88.9	86.8	37.3	81.3	84.4	12
$K = 5$	96.3	82.3	86.7	83.0	93.7	95.6	91.6	83.2	96.3	87.0	85.2	43.0	82.0	85.3	10

outperforming the best result in the offline setting. Our one-pass accuracy is slightly shy compared to CDAN [37], but is better in online metric.

Ablation study. We conduct ablation study on the impact of cross-domain bootstrapping in Table 5. Following Table 1, we provide the VisDA-C one-pass accuracy in class average. This study is to evaluate whether the improvement is introduced by cross-domain bootstrapping or simply the strong baseline with the objectives on the target domain (see Sec. 3.2). Thus, we devise a baseline by removing only the cross-domain bootstrapping, called w/o CroDoBo. The baseline model has one learner that is optimized by minimizing the objective $\ell_s + \ell_t + \ell_{\text{ent}} + \lambda \ell_{\text{div}}$, where $\ell_t = B^{-1} \sum_{b=1}^{B} \mathbb{1}\left(p_b \geq \tau\right) \mathcal{H}\left(\hat{y}_b; p(c|\tilde{t}_b; \boldsymbol{w})\right)$, which is Eq. (2) without exchanging the pseudo-labels. In Table 5, we observe that w/ CroDoBo is consistently better than w/o in the online average accuracy on all the datasets. Regarding one-pass accuracy, the effectiveness of cross-domain bootstrapping is unapparent on smaller datasets VisDA-C and COVID-DA, yet clearly outperforms w/o on large-scale *WILDS*-Camelyon17 and Fashion-MNIST-to-DeepFashion.

We further conduct ablation study on the objective terms (see Sec. 3.2) and report the results in Table 6. To eliminate the benefit of cross-domain boosting, our default setting is the model w/o CRODOBO. We leave out ℓ_{ent} and observe significant performance drop. Without ℓ_{div}, the performance decrease slight in the online metric, but far more sharply on the one-pass metric (which is calculated per-class). We analyze that the diversity term is important for imbalanced dataset like VisDA-C to achieve high class-average accuracy. We also report the results by replacing ℓ_{ent} and ℓ_{div} with Pseudo-labeling [28]. We replace either $\{\ell_{\text{ent}}, \ell_{\text{div}}\}$ or ℓ_t with MixMatch, and observe decent performance when employed together with $\{\ell_{\text{ent}}, \ell_{\text{div}}\}$ (see Table 6 row6). The RandAugment [5] on the entropy and diversity terms does not enhance the performance.

Number of Learners $K \geq 3$. We report the results of CroDoBo with varying number of learners $K \in \{2, 3, 4, 5\}$ on VisDA-C in Table 7. We observe that when K=3 the performance is consistent with K=2. However, from K=4 the performance is improved with more learners with discrepancies. This observation reflects the effectiveness to exploit the discrepant learners via bootstrapping and co-supervision. The choice of K is a trade-off between computation cost and performance. We find that K=2 is sufficient to yield state-of-the-art performance in most times, thus is a better choice considering its computation efficiency.

5 Conclusion

In the context of the *the right to be forgotten*, we propose an online domain adaptation framework in which the target data is erased immediately after prediction. A novel online UDA algorithm is proposed to tackle the lack of data diversity, which is a fundamental drawback of the online setting. The proposed method achieves state-of-the-art online results and comparable results to the offline domain adaptation approaches. We would like to extend CroDoBo to more tasks like semantic segmentation [31,86].

References

1. de Barros, R.S.M., de Carvalho Santos, S.G.T., Júnior, P.M.G.: A boosting-like online learning ensemble. In: 2016 International Joint Conference on Neural Networks (IJCNN). pp. 1871–1878. IEEE (2016) 5
2. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.: Mixmatch: A holistic approach to semi-supervised learning. arXiv preprint arXiv:1905.02249 (2019) 13
3. Chen, J., Wu, X., Duan, L., Gao, S.: Domain adversarial reinforcement learning for partial domain adaptation. IEEE Transactions on Neural Networks and Learning Systems (2020) 2
4. Chen, Y., Luo, H., Ma, T., Zhang, C.: Active online learning with hidden shifting domains. In: International Conference on Artificial Intelligence and Statistics. pp. 2053–2061. PMLR (2021) 4
5. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 702–703 (2020) 9, 13, 14
6. Delussu, R., Putzu, L., Fumera, G., Roli, F.: Online domain adaptation for person re-identification with a human in the loop. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 3829–3836. IEEE (2021) 4
7. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009) 10
8. Dredze, M., Crammer, K.: Online methods for multi-domain learning and adaptation. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing. pp. 689–697 (2008) 4
9. Elliott, S.J., Rafaely, B.: Frequency-domain adaptation of causal digital filters. IEEE Transactions on Signal processing **48**(5), 1354–1364 (2000) 4
10. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: international conference on machine learning. pp. 1050–1059. PMLR (2016) 7
11. Gal, Y., Hron, J., Kendall, A.: Concrete dropout. arXiv preprint arXiv:1705.07832 (2017) 7
12. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. JMLR **17**(1), 2096–2030 (2016) 5, 6, 10, 11, 12, 13
13. Garg, S., Goldwasser, S., Vasudevan, P.N.: Formalizing data deletion in the context of the right to be forgotten. Advances in Cryptology–EUROCRYPT 2020 **12106**, 373 (2020) 1, 3
14. Grandvalet, Y., Bengio, Y., et al.: Semi-supervised learning by entropy minimization. CAP **367**, 281–296 (2005) 10
15. Graves, L., Nagisetty, V., Ganesh, V.: Does ai remember? neural networks and the right to be forgotten (2020) 3, 4
16. Guo, H., Chen, B., Tang, R., Zhang, W., Li, Z., He, X.: An embedding learning framework for numerical features in ctr prediction. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 2910–2918 (2021) 1
17. Han, B., Sim, J., Adam, H.: Branchout: Regularization for online ensemble tracking with convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3356–3365 (2017) 5

18. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) 10
19. Hu, J., Tuo, H., Wang, C., Qiao, L., Zhong, H., Yan, J., Jing, Z., Leung, H.: Discriminative partial domain adversarial network. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16. pp. 632–648. Springer (2020) 2
20. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017) 10
21. Jaber, G., Cornuéjols, A., Tarroux, P.: Online learning: Searching for the best forgetting strategy under concept drift. In: International Conference on Neural Information Processing. pp. 400–408. Springer (2013) 5
22. Jain, V., Learned-Miller, E.: Online domain adaptation of a pre-trained cascade of classifiers. In: CVPR 2011. pp. 577–584. IEEE (2011) 4
23. Jin, X., Chen, P.Y., Hsu, C.Y., Yu, C.M., Chen, T.: Cafe: Catastrophic data leakage in vertical federated learning. arXiv preprint arXiv:2110.15122 (2021) 2
24. Kaissis, G.A., Makowski, M.R., Rückert, D., Braren, R.F.: Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence **2**(6), 305–311 (2020) 2
25. Kang, G., Jiang, L., Yang, Y., Hauptmann, A.G.: Contrastive adaptation network for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4893–4902 (2019) 2, 10
26. Koh, P.W., Sagawa, S., Xie, S.M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R.L., Gao, I., Lee, T., et al.: Wilds: A benchmark of in-the-wild distribution shifts. In: International Conference on Machine Learning. pp. 5637–5664. PMLR (2021) 3, 9, 10, 11, 12
27. Lafarge, M.W., Pluim, J.P., Eppenhof, K.A., Moeskops, P., Veta, M.: Domain-adversarial neural networks to address the appearance variability of histopathology images. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 83–91. Springer (2017) 2
28. Lee, D.H., et al.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML. vol. 3, p. 896 (2013) 13, 14
29. Lee, S., Kim, D., Kim, N., Jeong, S.G.: Drop to adapt: Learning discriminative features for unsupervised domain adaptation. In: ICCV (2019) 2
30. Li, Y., Liang, Y.: Learning overparameterized neural networks via stochastic gradient descent on structured data. arXiv preprint arXiv:1808.01204 (2018) 6
31. Li, Y., Yuan, L., Vasconcelos, N.: Bidirectional learning for domain adaptation of semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6936–6945 (2019) 14
32. Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: International Conference on Machine Learning. pp. 6028–6039. PMLR (2020) 2, 4, 5, 9, 10, 11, 12, 13
33. Liang, J., Hu, D., Feng, J.: Domain adaptation with auxiliary target domain-oriented classifier. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16632–16642 (2021) 2, 10, 11, 12, 13
34. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E.: A survey of deep neural network architectures and their applications. Neurocomputing **234**, 11–26 (2017) 2

35. Liu, Z., Luo, P., Qiu, S., Wang, X., Tang, X.: Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1096–1104 (2016) 3, 10, 12, 13
36. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: International conference on machine learning. pp. 97–105. PMLR (2015) 10, 11, 12, 13
37. Long, M., Cao, Z., Wang, J., Jordan, M.I.: Conditional adversarial domain adaptation. arXiv preprint arXiv:1705.10667 (2017) 10, 11, 12, 13, 14
38. Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., Zhang, G.: Learning under concept drift: A review. IEEE Transactions on Knowledge and Data Engineering **31**(12), 2346–2363 (2018) 5
39. Ma, X., Gao, J., Xu, C.: Active universal domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8968–8977 (2021) 4
40. Mancini, M., Karaoguz, H., Ricci, E., Jensfelt, P., Caputo, B.: Kitting in the wild through online domain adaptation. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 1103–1109. IEEE (2018) 4
41. Minku, L.L., White, A.P., Yao, X.: The impact of diversity on online ensemble learning in the presence of concept drift. IEEE Transactions on knowledge and Data Engineering **22**(5), 730–742 (2009) 5
42. Moon, J., Das, D., Lee, C.G.: Multi-step online unsupervised domain adaptation. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 41172–41576. IEEE (2020) 4
43. Nakkiran, P., Neyshabur, B., Sedghi, H.: The deep bootstrap framework: Good online learners are good offline generalizers. arXiv preprint arXiv:2010.08127 (2020) 9
44. Nanni, L., Ghidoni, S., Brahnam, S.: Handcrafted vs. non-handcrafted features for computer vision classification. Pattern Recognition **71**, 158–172 (2017) 4
45. Osband, I.: Risk versus uncertainty in deep learning: Bayes, bootstrap and the dangers of dropout. In: NIPS workshop on bayesian deep learning. vol. 192 (2016) 7
46. Pagallo, U., Durante, M.: Human rights and the right to be forgotten. In: Human Rights, Digital Society and the Law, pp. 197–208. Routledge (2019) 3
47. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017) 10
48. Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B.: Moment matching for multi-source domain adaptation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1406–1415 (2019) 2, 6
49. Peng, X., Usman, B., Kaushik, N., Wang, D., Hoffman, J., Saenko, K.: Visda: A synthetic-to-real benchmark for visual domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 2021–2026 (2018) 3, 9, 10, 11
50. Politou, E., Alepis, E., Virvou, M., Patsakis, C.: The "right to be forgotten" in the gdpr: Implementation challenges and potential solutions. In: Privacy and Data Protection Challenges in the Distributed Era, pp. 41–68. Springer (2022) 4
51. Prabhu, V., Chandrasekaran, A., Saenko, K., Hoffman, J.: Active domain adaptation via clustering uncertainty-weighted embeddings. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8505–8514 (2021) 4

52. Qi, G.J., Hua, X.S., Rui, Y., Tang, J., Zhang, H.J.: Two-dimensional multilabel active learning with an efficient online adaptation model for image classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(10), 1880–1897 (2008) 4
53. Rai, P., Saha, A., Daumé III, H., Venkatasubramanian, S.: Domain adaptation meets active learning. In: Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing. pp. 27–32 (2010) 4
54. Rosen, J.: The right to be forgotten. Stan. L. Rev. Online **64**, 88 (2011) 1
55. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International journal of computer vision **115**(3), 211–252 (2015) 10
56. Sahoo, D., Pham, Q., Lu, J., Hoi, S.C.: Online deep learning: Learning deep neural networks on the fly. arXiv preprint arXiv:1711.03705 (2017) 2
57. Saito, K., Kim, D., Sclaroff, S., Darrell, T., Saenko, K.: Semi-supervised domain adaptation via minimax entropy. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8050–8058 (2019) 9, 10, 11, 12, 13
58. Saito, K., Ushiku, Y., Harada, T., Saenko, K.: Adversarial dropout regularization. arXiv preprint arXiv:1711.01575 (2017) 7
59. Saito, K., Watanabe, K., Ushiku, Y., Harada, T.: Maximum classifier discrepancy for unsupervised domain adaptation. In: CVPR (2018) 2, 5, 6, 9
60. Schmitz, A., Bansho, Y., Noda, K., Iwata, H., Ogata, T., Sugano, S.: Tactile object recognition using deep learning and dropout. In: 2014 IEEE-RAS International Conference on Humanoid Robots. pp. 1044–1050. IEEE (2014) 7
61. Shi, Y., Seely, J., Torr, P.H., Siddharth, N., Hannun, A., Usunier, N., Synnaeve, G.: Gradient matching for domain generalization. arXiv preprint arXiv:2104.09937 (2021) 12
62. Shu, R., Bui, H.H., Narui, H., Ermon, S.: A dirt-t approach to unsupervised domain adaptation. In: ICLR (2018) 2, 5
63. Sohn, K., Berthelot, D., Li, C.L., Zhang, Z., Carlini, N., Cubuk, E.D., Kurakin, A., Zhang, H., Raffel, C.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. arXiv preprint arXiv:2001.07685 (2020) 8
64. Sun, B., Feng, J., Saenko, K.: Return of frustratingly easy domain adaptation. In: AAAI (2016) 10, 11, 12, 13
65. Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A., Hardt, M.: Test-time training with self-supervision for generalization under distribution shifts. In: International Conference on Machine Learning. pp. 9229–9248. PMLR (2020) 4, 5
66. Taufique, A.M.N., Jahan, C.S., Savakis, A.: Conda: Continual unsupervised domain adaptation. arXiv preprint arXiv:2103.11056 (2021) 4
67. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: CVPR (2017) 2, 5
68. Varsavsky, T., Orbes-Arteaga, M., Sudre, C.H., Graham, M.S., Nachev, P., Cardoso, M.J.: Test-time unsupervised domain adaptation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 428–436. Springer (2020) 4
69. Villaronga, E.F., Kieseberg, P., Li, T.: Humans forget, machines remember: Artificial intelligence and the right to be forgotten. Computer Law & Security Review **34**(2), 304–313 (2018) 3, 4
70. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proceedings

of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2517–2526 (2019) 2, 9
71. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. In: International Conference on Learning Representations (2020) 4
72. Wang, Q., Rao, W., Sun, S., Xie, L., Chng, E.S., Li, H.: Unsupervised domain adaptation via domain adversarial training for speaker recognition. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 4889–4893. IEEE (2018) 2
73. Wang, R., Wu, Z., Weng, Z., Chen, J., Qi, G.J., Jiang, Y.G.: Cross-domain contrastive learning for unsupervised domain adaptation. IEEE Transactions on Multimedia (2022) 2
74. Warde-Farley, D., Goodfellow, I.J., Courville, A., Bengio, Y.: An empirical analysis of dropout in piecewise linear networks. arXiv preprint arXiv:1312.6197 (2013) 7
75. Wei, K., Li, J., Ding, M., Ma, C., Yang, H.H., Farokhi, F., Jin, S., Quek, T.Q., Poor, H.V.: Federated learning with differential privacy: Algorithms and performance analysis. IEEE Transactions on Information Forensics and Security **15**, 3454–3469 (2020) 2
76. Wei, Z., Chen, J., Goldblum, M., Wu, Z., Goldstein, T., Jiang, Y.G.: Towards transferable adversarial attacks on vision transformers. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 2668–2676 (2022) 2
77. Wu, Z., Han, X., Lin, Y.L., Uzunbas, M.G., Goldstein, T., Lim, S.N., Davis, L.S.: Dcan: Dual channel-wise alignment networks for unsupervised scene adaptation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 518–534 (2018) 2
78. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747 (2017) 3, 10, 12, 13
79. Xu, X., Wu, J., Yang, M., Luo, T., Duan, X., Li, W., Wu, Y., Wu, B.: Information leakage by model weights on federated learning. In: Proceedings of the 2020 Workshop on Privacy-Preserving Machine Learning in Practice. pp. 31–36 (2020) 2
80. Yang, L., Balaji, Y., Lim, S.N., Shrivastava, A.: Curriculum manager for source selection in multi-source domain adaptation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16. pp. 608–624. Springer (2020) 2
81. Yang, L., Wang, Y., Gao, M., Shrivastava, A., Weinberger, K.Q., Chao, W.L., Lim, S.N.: Deep co-training with task decomposition for semi-supervised domain adaptation. arXiv preprint arXiv:2007.12684 (2020) 3, 8
82. Zhang, N., Jia, Q., Deng, S., Chen, X., Ye, H., Chen, H., Tou, H., Huang, G., Wang, Z., Hua, N., et al.: Alicg: Fine-grained and evolvable conceptual graph construction for semantic search at alibaba. arXiv preprint arXiv:2106.01686 (2021) 1
83. Zhang, X., Chen, X., Liu, J.K., Xiang, Y.: Deeppar and deepdpa: privacy preserving and asynchronous deep learning for industrial iot. IEEE Transactions on Industrial Informatics **16**(3), 2081–2090 (2019) 1
84. Zhang, Y., Niu, S., Qiu, Z., Wei, Y., Zhao, P., Yao, J., Huang, J., Wu, Q., Tan, M.: Covid-da: Deep domain adaptation from typical pneumonia to covid-19. arXiv preprint arXiv:2005.01577 (2020) 3, 9, 10, 11, 12
85. Zhang, Y., Liu, T., Long, M., Jordan, M.: Bridging theory and algorithm for domain adaptation. In: International Conference on Machine Learning. pp. 7404–7413. PMLR (2019) 2, 10, 11, 12, 13

86. Zhao, S., Li, B., Yue, X., Gu, Y., Xu, P., Hu, R., Chai, H., Keutzer, K.: Multi-source domain adaptation for semantic segmentation. arXiv preprint arXiv:1910.12181 (2019) 14
87. Zhu, L., Han, S.: Deep leakage from gradients. In: Federated learning, pp. 17–31. Springer (2020) 2
88. Ziller, A., Usynin, D., Braren, R., Makowski, M., Rueckert, D., Kaissis, G.: Medical imaging deep learning with differential privacy. Scientific Reports **11**(1), 1–8 (2021) 1