RESEARCH-ARTICLE

Quantifying Samples with Invariance for Source-Free Class Incremental Domain Adaptation

ZHIYU YE, Sun Yat-Sen University, Guangzhou, Guangdong, China

GUOWEN LI, Sun Yat-Sen University, Guangzhou, Guangdong, China

HAOYUAN LIANG, Sun Yat-Sen University, Guangzhou, Guangdong, China

ZIXI WANG, Sun Yat-Sen University, Guangzhou, Guangdong, China

SHILEI CAO, Sun Yat-Sen University, Guangzhou, Guangdong, China

YUSHAN LAI, Sun Yat-Sen University, Guangzhou, Guangdong, China

View all

Open Access Support provided by:

Sun Yat-Sen University

Published: 27 October 2025

Citation in BibTeX format

MM '25: The 33rd ACM International Conference on Multimedia
October 27 - 31, 2025
Dublin, Ireland

Conference Sponsors:
SIGMM

MM '25: Proceedings of the 33rd ACM International Conference on Multimedia (October 2025)
https://doi.org/10.1145/3746027.3755053
ISBN: 9798400720352

Quantifying Samples with Invariance for Source-Free Class Incremental Domain Adaptation

Zhiyu Ye
Sun Yat-Sen University
Zhuhai, China
yezhy26@mail2.sysu.edu.cn

Guowen Li
Sun Yat-Sen University
Zhuhai, China
ligw8@mail2.sysu.edu.cn

Haoyuan Liang
Sun Yat-Sen University
Zhuhai, China
lianghy68@mail2.sysu.edu.cn

Zixi Wang
Sun Yat-Sen University
Zhuhai, China
wangzx87@mail2.sysu.edu.cn

Shilei Cao
Sun Yat-Sen University
Zhuhai, China
caoshlei@mail2.sysu.edu.cn

Yushan Lai
Sun Yat-Sen University
Zhuhai, China
laiysh6@mail2.sysu.edu.cn

Juepeng Zheng*
Sun Yat-Sen University
Zhuhai, China
National Supercomputing Center in Shenzhen
Shenzhen, China
zhengjp8@mail.sysu.edu.cn

Abstract

In response to the growing demands of real-world applications, models must be capable of learning continuously under inconsistent data distribution. However, existing Class-Incremental (CI) methods fail to alleviate domain shifts, while traditional Unsupervised Domain Adaptation (UDA) techniques suffer from catastrophic forgetting and privacy concerns. To address these limitations, we explore Source-Free Class Incremental Domain Adaptation (SF-CIDA) and propose a novel approach, Quantifying Samples with Invariance (QSI), for this scenario. Our proposed method involves two main strategies: (1) **Semantic Restructuring**. We identify confusing source category pairs and restructure images to create a negative dataset that is semantically similar to the source features, refining accurate decision boundary among source categories. (2) **Invariance Quantification**. The sample's confidence is then quantified by its spatial location under the special data distribution, reflecting the trade-off between invariant features and domain shifts. Guided by such strategy, samples' confidence is accumulated for the target model to prioritize reliable categories, not only mitigating the poor performance of experience replay in unsupervised scenarios, but alleviating distribution discrepancies simultaneously. Experiments demonstrate that our approach outperforms previous methods, establishing new state-of-the-art performance on the Office-31, Office-Home and DomainNet-126 datasets, with average accuracy improvements of over 7.3%, 4.9% and 10.2% respectively.

*Corresponding author.

MM '25, Dublin, Ireland

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3755053

CCS Concepts

• **Computing methodologies → Computer vision**.

Keywords

Source-free Domain Adaptation, Class Incremental Learning

ACM Reference Format:
Zhiyu Ye, Guowen Li, Haoyuan Liang, Zixi Wang, Shilei Cao, Yushan Lai, and Juepeng Zheng. 2025. Quantifying Samples with Invariance for Source-Free Class Incremental Domain Adaptation. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3746027.3755053

1 Introduction

Deep learning models have demonstrated excellent performance on a variety of tasks where the source and target domains are of the same distribution. However, in practical situations, target domains often experience domain shifts [18, 43], leading to inconsistencies in the data distribution. As a result, models trained solely on source data perform poorly when applied to out of distribution domains due to limited generalization capabilities. Therefore, Unsupervised Domain Adaptation (UDA) methods have been developed to address the challenge of distributional invariance.

However, for real-world applications, data often arrive as a continuous stream, conflicting with the assumption that all data is available in advance in UDA methods. For example, consider a medical robot equipped with disease diagnosis model trained on a large dataset of adult data. While some diseases are now developing in a younger age, such a model may perform poorly for children due to differences in manifestation of the same symptoms compared

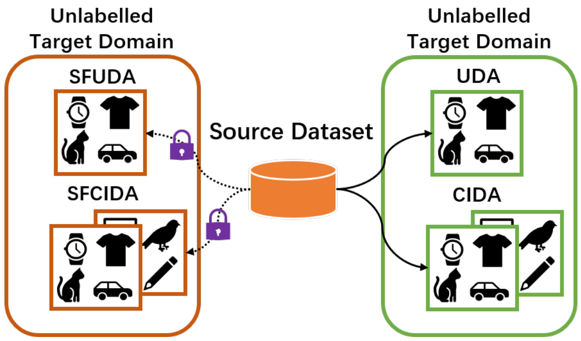

Figure 1: An illustration of different settings (UDA, SFUDA, CIDA, SFCIDA). Overlapping images indicate incremental tasks, while the locks on the dashed lines imply restricted access to the data.

to adults. Furthermore, given the growing demand for diagnosing a wider range of diseases and concerns over medical privacy, the model needs to preserve learned knowledge without requiring training data.

Nevertheless, UDA falls short in addressing the above medical scenario, as it not only necessitate access to source data but also assumes the source and target domains share the same label set [17, 22]. Recently, several studies have delved into Source-Free Unsupervised Domain Adaptation (SFUDA) to address the absence of source data during adaptation [27, 29, 41, 48, 49]. However, SFUDA still suffers from catastrophic forgetting when learning from continuous data-streams. In contrast, Class-Incremental Learning (CIL) [11, 37] focuses on preserving prior knowledge while acquiring new information, but cannot effectively handle domain shifts. Given the limitations of previous works, we propose a novel setting called Source-Free Class Incremental Domain Adaptation (SFCIDA) to meet the demands of real-world applications.

The SFCIDA paradigm introduces two main challenges for the models. First, the models are prone to forgetting previous knowledge during continuous learning. To address this, experience replay, which selectively stores a limited number of past training samples in a memory buffer, has been shown to be effective in preserving prior knowledge in supervised scenarios [31, 37]. However, the absence of source and unlabelled target data weaken the effectiveness of CI methods, as it becomes difficult to denoise samples and store accurate images. Secondly, to mitigate distribution discrepancies, models [17, 22] must be capable of performing domain alignment in the absence of source data. Although SFUDA approaches [27, 29], which rely only on target samples, are well-suited for this, they still face poor long-term performance.

Therefore, we propose a novel Source-Free Class Incremental Domain Adaptation approach named Quantifying Samples with Invariance (QSI), which leverages the confidence quantification for target samples estimated by a well-designed pre-trained model. To tackle the challenges encountered, our method employs a two-stage strategy: (1) **Semantic Restructuring**. We construct a negative dataset that is semantically similar to source domain and train the source model with both positive and negative datasets, which enables to configure decision boundaries precisely. Furthermore, with the confidence accumulation method, the model can effectively address the challenges posed by domain shift and the absence of source data. (2) **Invariance Quantification**. Target samples are weighted based on the degree of invariance between themselves and the source domain images of positive categories, ensuring the use of more reliable samples throughout the training process. Building upon the confidence we compute, we weight the transfer loss function and samples in experience replay. This stage strengthens the model's robustness and retains higher-quality images in the memory buffer to simultaneously address the problems of domain shift and catastrophic forgetting.

To summarize, the main contributions of this paper are as follows:

(1) To the best of our knowledge, we are the first to explore Source-Free Class Incremental Domain Adaptation (SFCIDA) and we propose a novel Quantifying Samples with Invariance (QSI) method to address two challenges in SFCIDA setting.
(2) We propose Semantic Restructuring and Invariance Quantification to sift reliable samples, which enhance the robustness of the training model and alleviates both catastrophic forgetting and distribution discrepancies.
(3) We evaluate our method on three datasets (**Office-31**, **Office-Home** and **DomainNet-126**), with average accuracy improvements of over 4.9%, 7.3% and 10.2% compared to SOTA methods, respectively.

2 Related Works

Domain Adaptation and Source-Free Domain Adaptation. To address the distribution discrepancies, Unsupervised Domain Adaptation (UDA) methods aim to mitigate domain-shift and improve performance on the target domain [8, 10, 21, 24, 25]. Numerous prior works have made great progress in reducing domain gap by minimizing transfer loss [7, 9, 15, 30, 44] and statistically aligning data distributions across different domains [5, 16, 36, 45], or by employing a generative models like Generative Adversarial Nets (GANs) to generate target-style samples from source domain [15, 39, 50]. A subfield of DA is Partial Domain Adaptation (PDA), which assumes that the target label set is a subset of the source domain. To transfer knowledge into a smaller label set, source samples are assigned by class-level or instance-level weights [2–4]. Other methods address negative-transfer problem by employing a discriminator [3] or by conducting alignment between source and target domains [14].

But for real-world applications, source data is often inaccessible due to private concerns or data corruption, rendering traditional UDA methods that require both source and target data inapplicable. Consequently, recent researches have focused on addressing the problem in the context of Source-free Unsupervised Domain Adaptation (SFUDA), where only source models and unlabelled target data are available [6, 26, 27, 35, 41, 47]. GPUE [29] introduced an approach to refine pseudo-labels by evaluating samples' uncertainty through their neighbours. SHOT [27] provided another refinement strategy to assign labels to unlabelled target samples based on cosine similarity to the category centroid. Additionally,

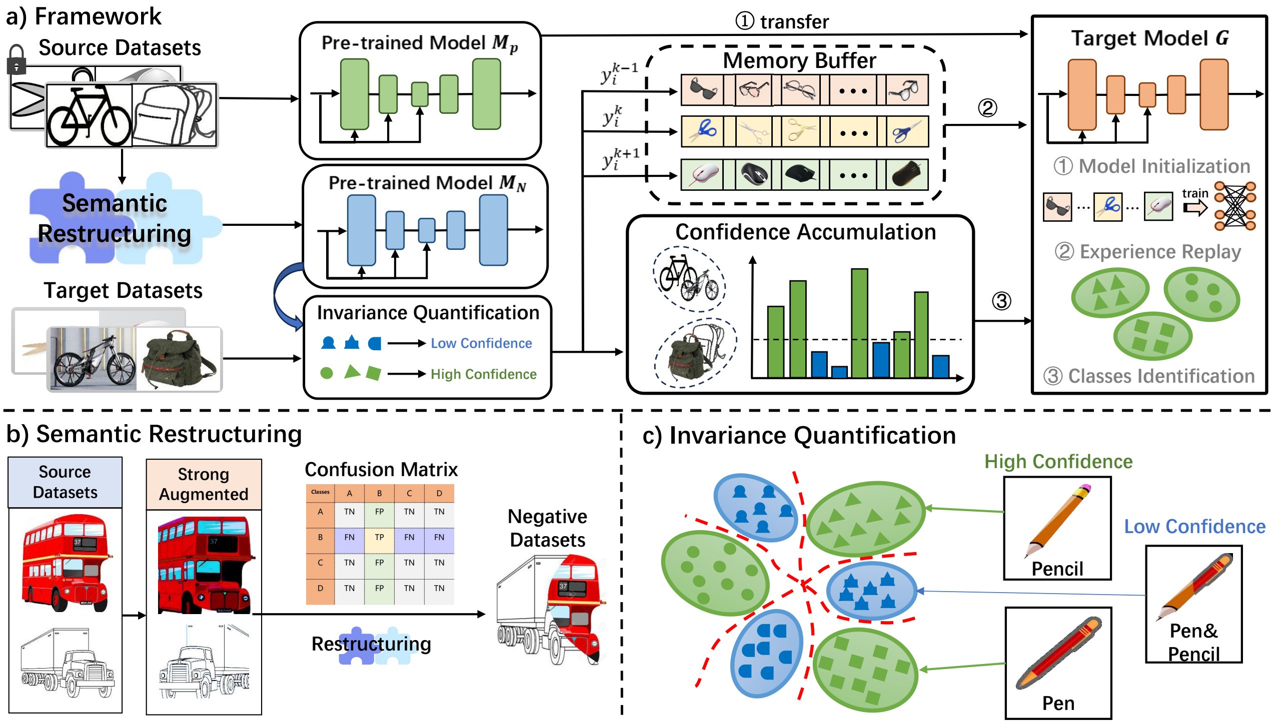

Figure 2: (a) Framework. The source models M_N and M_P are trained with and without negative datasets, respectively. M_N is utilized to implement our Invariance Quantification strategy for denoising target samples, while the target model G is initialized by M_P. To build the memory buffer, we identify prototypes of the target label set in each incremental task using the Confidence Accumulation method. Satisfying with the SFCIDA scenario requirements, we train model G and the classifier using only target and stored images. (b) Semantic Restructuring. We augment the source samples to generate a confusion matrix, from which we obtain negative pairs, and employ Semantic Restructuring to construct a negative dataset. (c) Invariance Quantification. Based on the constructed decision boundaries, we can quantify the quality of target samples. Samples falling within the green distribution exhibit high confidence. Conversely, those within the blue distribution indicate low confidence.

other methods have focused on generative models [26, 35], self-supervised strategies [6], and class-prototypes [27]. However, both UDA and SFUDA methods are prone to catastrophic forgetting when data arrives in a streamming manner.

Continual Learning. Motivated by the human learning process, where data streams continuously, Continual Learning has been studied, with the expectation that artificial intelligence can accumulate knowledge [12, 20]. Among the various branches of Continual Learning, Class-Incremental Learning (CIL) is one of the most widely studied. CIL methods aim to strike a balance between efficiently learning new knowledge and maintaining the ability to preserve old knowledge, which can be challenging due to the risk of catastrophic forgetting. To address this, CIL methods employ strategies like weight regularization [19], generative replay [40], gradient projection [32], self-supervised learning [33] and parameter allocation [34]. For example, GEM [32] and AQM [1] stores a few images that are close to the feature centroid in a small memory buffer during training and then replay them in subsequent tasks.However, most CIL methods are not well-suited in DA scenarios as they lack the ability to bridge the domain gap.

Class Incremental Domain Adaptation. Existing DA methods often suffer from catastrophic forgetting when learning new knowledge. Conversely, CI methods are not designed to mitigate domain shift in transfer learning. Therefore, a few works have introduced CI approaches to DA. For example, CIDA [23] learned a target-specific latent space to align target samples from shared classes and preserve the semantic granularity. ProCA [28] introduced a prototype-guided method to detect shared categories in PDA, which alleviates both domain discrepancies and catastrophic forgetting via prototype-based methods. However, both methods rely on data or meta-data from the source domain, which is not suitable for some realistic scenario (e.g. biometric data). Despite the inconsistent data distributions between the source and target domains, using source data can effectively preserve old knowledge, since the same categories across different domains often share common features. Consequently, the Source-Free (source model only) setting poses greater challenges for the CIDA scenario.

3 Proposed Approach

Our study focuses on the Source-Free Class Incremental Domain Adaptation scenario. In this setting, the source data is no longer accessible during adaptation, and all subsequent incremental tasks in the target domain rely solely on the pre-trained source model. We train the model M_p only by source dataset and the model M_n with well-defined decision boundaries using both the source and the negative dataset. We initialize target model G using M_p, and assign confidence to target samples through the output of model M_N by Invariance Quantification strategy. High-quality target images are stored in a memory buffer and the model G will be trained jointly on the target and the stored samples. With two-stage strategy and experience replay, the target model, initialized by the source model, effectively mitigates both catastrophic forgetting and domain shift. An overall framework is illustrated in Fig. 2.

3.1 Semantic Restructuring

In unsupervised settings, pseudo-labels inevitably contain noise due to the absence of ground truth labels. Domain shifts and unseen source data can further push image features toward the edges of the source distribution, degrading their quality. Although [35] proposes a method for configuring decision boundaries using randomly composed samples, the generated images do not contain sufficient information to mislead the model effectively, providing minimal assistance in configuring accurate decision boundaries. Therefore, we propose a novel method for restructuring images semantically to achieve precise division of decision boundaries.

Let $D_s = \{(x_s, y_s)|x_s \in X_s, y_s \in C_s)\}$ denote the source domain, where X_s is the positive sample space and y_s is the corresponding label in the source label set C_s for x_s. Then we construct samples that do not belong to the positive (source) categories but still share similarities with them. To distinguish these samples, we refer to them as the negative dataset, $D_n = \{(x_n, y_n)|x_n \in X_n, y_n \in C_n)\}$ represents the constructed negative datasets, where X_n denotes the negative sample space and y_n is a negative label in the negative source label set C_n. To configure reliable negative data sets D_n, the following properties must be satisfied. First, the center of the negative distribution should lie between the most similar positive categories, which may confine the source distribution within a limited area, thereby weakening the generalization ability to obtain samples with low uncertainty. Second, the pre-trained model must maintain a balance between positive and negative samples to avoid bias towards any dataset, with the decision boundary configured for each source category ($i.e.$ $|C_s| = |C_n|$).

The underlying hypothesis is that a high-quality target sample x_t will share some invariance and distinctiveness in features with source samples of the same category. Consequently, $G(x_t)$ is expected to be close to the centroid of the positive distribution, where $G()$ denotes a feature extractor in target model and $G(x_t)$ denotes the feature extraction of sample x_t. In contrast, ambiguous samples will fall outside the decision boundary and be classified as negative categories. The level of noise in such samples correlates with their proximity to the center of the negative distribution.

By compositing local parts from different source samples (as shown in Fig. 3), we can create a feature combination consisting of confusing categories, potentially falling in the latent space beyond the decision boundary. To identify similar categories, we generate a strongly-augmented sample x_{ag} for each source sample x_s. Generally, the model M_P trained only on normal source samples, will

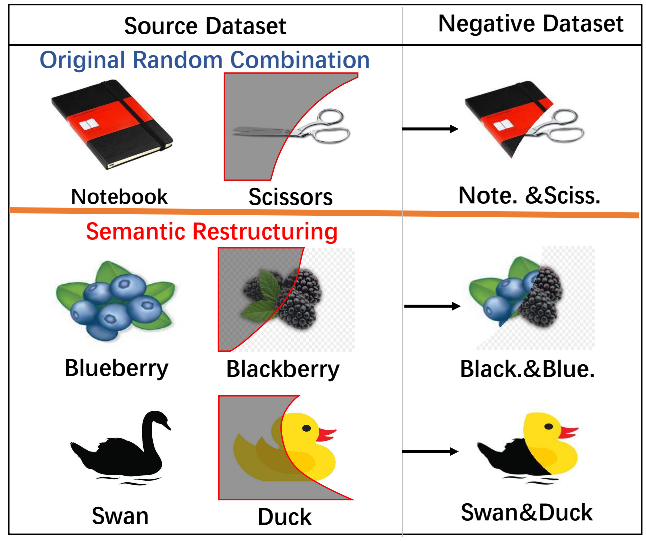

Figure 3: Random combinations of the source dataset [35] versus negative samples generated by highly similar category pairs in the confusion matrix. Clearly, the samples constructed using the latter strategy represent more meaningful image information

predict incorrect labels for some x_{ag}. We classify the augmented source samples using model M_P. For category y_s, the model may misclassify some images from other categories as y_s. Therefore, we construct a confusion matrix to identify the category $\tilde{y_s}$ that has the highest number of images misclassified as y_s and form the negative pair $(y_s, \tilde{y_s})$. These pairs allow us to configure a negative dataset between the most similar categories(see Fig. 3) and subsequently train another pre-trained M_N with both positive and negative samples to quantify the confidence of target samples.

3.2 Invariance Quantification

As described in Sec. 3.1, due to the domain change and the unsupervised nature, target samples often contain noise and deviate from the source centroid in the same categories. Since a reliable target sample will be semantically similar to the sample with the same label in the source domain and contain less noise, samples with low invariance to the source category are more likely to be misclassified and fall outside the decision boundary. Treating all samples with equal confidence may degrade the model's performance. Therefore, inspired by [29], we propose a novel method to estimate the confidence of unsupervised samples to overcome the mentioned challenges, which is based on the pre-trained model M_N and its probability outputs y_t.

To estimate confidence, we dig into the misclassified samples from the pre-trained model. We observe that a sample composed of features from multiple categories will have a large Euclidean distance from the source center, indicating great variance to the category it belongs to and making it challenging even for humans to distinguish. These images are more likely to be classified into the negative class dataset by the M_N, while reliable samples will be correctly classified. Accordingly, we assign weights to all target samples based on both the confidence of positive categories and the uncertainty in negative probability as follows:

$$W(x_t) = e^{-(W_n(x_t)+W_p(x_t))} \tag{1}$$

where $W_n(x_t)$ denotes the uncertainty of sample x_t computed by negative categories (variant part) and $W_p(x_t)$ denotes the entropy of positive categories (invariant part). To keep the weights stay within the range of 0 to 1, we use an exponential function to regularize the sum of the metrics. More specifically, we compute the $W_p(x_t)$ as:

$$W_p(x_t) = -\frac{\sum_{i=1}^{|C_s|} y_t^i \log y_t^i}{\log_2 |C_s|} \tag{2}$$

where $|C_s|$ denotes the number of classes in source domain and $\log_2 |Cs|$ denotes a normalization method for different $|C_s|$. Meanwhile, we evaluate $W_n(x_t)$ as:

$$W_n(x_t) = -y_t^n \log(1-y_t^n),\ y_t^n = \sum_{i=|C_s|+1}^{|C_s|+|C_n|} y_t^i \tag{3}$$

where $|C_n|$ denotes the number of classes in negative datasets, and y_n represents the cumulative probabilities of the negative output. As the value of y_n increases, it indicates that the image is characterized by more categories and exhibits greater variance relative to the source category. Thus, $W_n(x_t)$ penalizes samples that shift beyond the decision boundaries, ensuring that they contribute less to the training process. Now, we weight the standard cross-entropy loss L_{CE} as:

$$L_{CE} = \mathbb{E}_{x_t \in \mathcal{X}_t} \left[-W(x_t) \sum_{i}^{|C_s|} y_i \log y_i \right] \tag{4}$$

3.3 Buffer Construction and Replay

One of the key challenges in our setting is detecting shared categories in Partial Domain Adaptation and handling unsupervised target samples. This is crucial, as specific labels are essential for experience replay in Class-Incremental Learning. To address this challenge, we propose a detection approach based on the confidence accumulation evaluated in Eq. (1), which integrates the majority of the samples' information. Let $q_t = G_n(x_t)$ denote the extracted feature from the pre-trained model M_N and $\hat{y}_t = \sigma(q_t)$ denote the probability prediction.

We select the highest values from $\hat{y}_t$ to assign pseudo-labels to the sample and convert it into a weighted one-hot encoding $\delta(x_t) = \arg\max_{k \in |C_s|} \hat{y}_t^k$. By summing all the vectors, we can obtain a denoised estimate for each category as:

$$v_k = \sum_{x \in \mathcal{X}_t} W(x_t)\delta(x_t), k \in C_s \tag{5}$$

where v_k denotes the count for the k-th class. By comparing v_k with the threshold α, set by the buffer capacity, we can roughly estimate whether a class is included in the training task. If $v_k < \alpha$, it indicates that there are limited samples in class k, and it can be considered as noise. Otherwise, class k is a shared class in both source and target domains. Due to the presence of unlabeled samples, shared classes may contain noise and can lead to misclassification. Therefore, the buffer for each category should be updated once it has been constructed. The v_k^a computed in Eq. (5) will be recorded when the class is initially identified and compared with the v_k^b computed in the subsequent task b. If $v_k^a > v_k^b$, we update the buffer for the k-th class with target samples from the current task b. A higher value of v_k represents that there are more images classified into class k, which suggests less noise.

As shared classes are identified, we refine pseudo-labels with a self-supervised pseudo-labeling strategy [27]. First, we attain the feature centroid for each identified class:

$$c_k^{(0)} = \frac{\sum_{x \in \mathcal{X}_t} \hat{y}_t \cdot \hat{q}_t}{\sum_{x \in \mathcal{X}_t} \hat{y}_t} \tag{6}$$

where $\hat{q}_t = G_n(x_t)$ denotes the feature extracted from target model G. For each sample, we generate a pseudo-label via a nearest centroid classifier:

$$\hat{y}_t = \arg\min_k D(\hat{q}_t, c_k^{(0)}) \tag{7}$$

where D is a distance function, typically measuring cosine similarity. Then we repeat Eq. (6) and Eq. (7) for another iteration to obtain an improved pseudo-label.

$$c_k^{(1)} = \frac{\sum_{x \in \mathcal{X}_t} \mathbb{I}(\hat{y}_t = k) \cdot \hat{q}_t}{\sum_{x \in \mathcal{X}_t} \mathbb{I}(\hat{y}_t = k)}, \quad \hat{y}_t = \arg\min_k D(\hat{q}_t, c_k^{(1)}) \tag{8}$$

where $\mathbb{I}$ denotes the indicator function.

To mitigate catastrophic forgetting, we establish a memory buffer to store images from detected shared classes. Due to buffer constraints, we can only store a limited number of images. Inspired by iCaRL [37], let μ be the feature centroid of class k by: $\mu_k = \mathbb{E} \sum_{x_t \in D_k} G(x_t)$, and m as the maximum number of exemplars that can be stored for each class. The exemplar is chosen using the following criteria:

$$y_i^k = \arg\max_{x_i \in D_k} ||\mu - \frac{1}{i}[G(x_i) + \sum_{j=1}^{i-1} G(\overline{x}_j)]|| \tag{9}$$

where i iterates from 1 to m. During the iteration, an exemplar is stored if the current feature centroid of exemplars has the closest similarity to the centroid over all samples in the same class.

3.4 Overall Objective

We replay the stored images (in Sec. 3.3) to maintain the knowledge learned from previous tasks. To alleviate domain shift in PDA, we employ contrastive loss to facilitate domain alignment. To this end, we attain the feature centroid for categories in the buffer and current task. Then for all the sample in target domain, we optimize InfoNCELoss [46] by:

$$L_{con} = -\mathbb{E}_{x \in \mathcal{X}_t} \log \frac{\exp(G(x_t) \cdot c_+/\tau)}{\sum_{i \in D_N \cup D_t} \exp(G(x_t) \cdot c_i/\tau)} \tag{10}$$

where c_+ denotes the feature centroid of samples of same label, c_i denotes the centroid of different labels, and τ denotes a temperature factor. We use a weighted cross-entropy loss as:

$$L_{bank} = E_{x_t \in \mathcal{X}_b}[-W(x_t) \cdot \sum_{i=1}^{|C_b|} p_i \log p_i] \tag{11}$$

Table 1: Classification accuracy (%) of Office-Home. Our approach outperform prior methods in 12 experiments.

Method	Office-Home												
	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg.
Resnet	47.2	67.8	82.9	58.2	64.8	71.2	55.9	41.9	69.0	65.2	44.6	71.3	61.7
Source Only	49.8	65.8	75.8	53.5	62.3	65.7	53.2	43.8	74.6	64.6	48.8	78.9	61.4
DANN [10]	33.1	40.0	45.8	36.8	36.6	44.1	32.0	29.8	49.8	42.4	40.2	55.2	40.5
PADA [3]	24.8	41.4	55.1	18.3	35.0	36.3	25.9	26.2	53.7	46.8	31.4	50.0	37.1
iCaRL [37]	14.6	20.5	14.6	10.3	19.7	14.6	11.2	21.3	14.5	11.0	15.6	21.6	15.8
FeCAM [11]	14.8	19.8	15.5	16.8	16.8	21.6	33.4	21.5	23.1	33.9	23.0	35.2	23.0
GPUE [29]	43.4	58.9	66.1	41.9	54.0	55.8	45.7	34.1	62.4	50.7	41.3	63.0	51.4
CIDA [23]	32.2	45.9	49.1	36.5	48.6	46.6	51.6	33.5	59.0	64.0	38.0	65.1	47.5
ProCA [28]	52.6	72.1	79.1	58.0	64.0	65.5	60.1	47.9	73.9	69.4	52.6	79.1	64.5
DIFO [42]	43.2	66.9	73.7	54.7	62.0	65.0	52.3	39.5	73.0	64.8	45.2	77.3	59.8
QSI (Ours)	**59.1**	**81.1**	**85.6**	**63.8**	**75.5**	**75.1**	**63.6**	**50.4**	**85.6**	**74.8**	**59.0**	**88.2**	**71.8**

Therefore, the overall loss function used is as follows,where λ_1,λ_2,λ_3 are hyper-parameters.:

$$L = \lambda_1 L_{CE} + \lambda_2 L_{con} + \lambda_3 L_{bank} \tag{12}$$

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate our approach on three datasets, including **Office-31**, **Office-Home** and **DomainNet-126**. **Office-31** consists of 31 categories across 3 different domains including Amazon (**A**), DSLR (**D**) and Webcam (**W**). We divide each domain into six parts with 5 categories, each subset can be regarded as an incremental task in our experiments. **Office-Home** includes 65 classes of images from 4 distinct domains including Art (**Ar**), Clipart (**Cl**), Product (**Pr**), and Real World (**Rw**). We conduct experiments for 12 domain-shifts by 6 disjoint subsets, each containing 10 categories, within a single incremental setting. **DomainNet-126** is a challenging large-scale dataset with 7 invariant domains. Follow [38], we use 7 domain adaptation tasks built form a subset containing 126 categories from 4 domains including Clipart (**C**), Painting (**P**), Sketch (**S**), Real (**R**) to evaluate our method. Each adaptation consists of 6 incremental task, each with 20 disjoint categories.

Implementation details. We employ the same ResNet50 [13] models trained on the source domain as a feature extractor followed by a classifier. For training process, we initialize the model with pre-trained models' parameters, and utilize the SGD optimizer with a learning rate of 0.001. For hyper-parameters, we set $\lambda_1 = 1$, $\lambda_2 = 0.1$, $\lambda_3 = 1$, $\alpha = 15$.

Baselines. We compare our proposed QSI with other prior methods in different fields within the same SFCIDA scenario. Additionally, some existing methods are not applicable to the new scenario we introduced, resulting in suboptimal performance. (1) Source model only. The baseline models are evaluated directly on the target domain after training on the source domain; (2) Unsupervised Domain Adaptation: DANN [10], PADA [3]. These models are no longer evaluated in a standard DA scenario. Instead, we train the models on multiple incremental tasks in a continual learning setting and evaluate their performance on the overall target dataset after training; (3) Source-Free Unsupervised Domain Adaptation: GPUE [29], DIFO [42]. Similar to UDA methods, the performance is also evaluated after training on multiple incremental tasks; (4) Class-Incremental Learning: iCaRL [37], FeCAM [11]. Traditional CIL methods are not affected by domain shift during training. Therefore, we first train a baseline model on the labeled source dataset. Then, we divide the target domain into multiple incremental tasks and continue training the target model, initialized from the source model, using the aforementioned methods; (5) Class-Incremental Domain Adaptation: CIDA [23], ProCA [28]. These methods are not allowed to access source data during training.

4.2 Results

Office-31. In Tab. 2, we show results on 6 experiments by average accuracy. We surpass the SOTA in 6 tasks under SFCIDA setting with average accuracy increment over +4.9%. For some adaptations with little domain shift like ($D \rightarrow W$), UDA experiments can be seen as a supervised tasks. Thus, DANN [10], ProCA [28] reach a competitive accuracy. Although there are some easy tasks, we still achieve the best overall performance.

Office-Home. As shown in Tab. 1, we perform 12 domain shifts experiment under SFCIDA setting and report the accuracy of each tasks and compute the mean over all adaptations. Some DA methods, such as GPUE, lack of memory stability while CI methods like iCaRL and FeCAM are ineffective in mitigating domain discrepancies. Both of them perform poorly in the SFCIDA setting. We outperform the SOTA in every experiments and a notable margin of +7.3% on average, even though without source data during training.

DomainNet-126. We evaluate our approach and some recent works under SFCIDA setting on 7 adaptation experiments of DomainNet-126 in Tab. 3. Neither CI nor SFUDA methods are suitable for this scenario. We achieve the highest accuracy on the average performance and a improvement over +5.8%. For CIDA methods, we outperform the SOTA by +10.2%. Due to the challenging nature of DomainNet as a large-scale dataset, the results validates the reliability of our model and its applicability to real-world scenarios.

4.3 Discussions

Average Accuracy and Forgetting Measure. The performance of Class-Incremental Learning methods is typically estimated from two aspects, overall performance at the current moment evaluated by average accuracy (AA) and memory stability evaluated by forgetting measure (FM). We report the AA and FM metrics in Fig. 4. Our method outperforms other approaches in every task, and forgets least knowledge during the learning process. The FM metric of DA method have a high value in subsequent tasks, indicating the phenomenon of catastrophic forgetting. While CI method could

Table 2: Final accuracy (%) of Office-31, where SF, DA, CI indicate Source-Free, Domain Adaptation and Class-Incremental Learning respectively. The proposed approach achieve the highest accuracy compared to existing methods.

Method	DA	CI	SF	Office-31						
				A→D	A→W	D→A	D→W	W→A	W→D	Avg.
Source only	✗	✗	✗	65.6	63.6	58.9	95.3	65.3	99.0	74.6
DANN [10]	✓	✗	✗	74.9	72.5	55.7	96.6	51.4	97.7	74.8
PADA [3]	✓	✗	✗	56.9	61.5	12.5	82.4	46.7	84.3	57.4
iCaRL [37]	✗	✓	✗	65.2	59.5	30.9	60.1	29.9	46.9	48.8
FeCAM [11]	✗	✓	✗	72.3	74.1	30.9	77.3	36.9	69.1	60.1
GPUE [29]	✓	✗	✓	56.0	56.1	60.6	86.2	58.4	95.6	68.8
CIDA [23]	✓	✓	✗	70.4	64.5	48.1	95.1	52.7	98.8	71.6
ProCA [28]	✓	✓	✗	71.4	64.1	61.4	95.9	61.4	**99.4**	75.6
DIFO [42]	✓	✗	✓	58.6	64.2	53.7	71.1	51.5	84.9	64.0
QSI (Ours)	✓	✓	✓	**76.6**	**74.0**	**67.5**	**97.0**	**68.1**	**99.4**	**80.5**

Table 3: Classification accuracy (%) of DomainNet-126 for 7 adaptation. SF, DA, CI indicates Source-Free, Domain Adaptation and Class-Incremental Learning respectively. Our approach have the best performance over all methods in 7 experiments.

Method	DA	CI	SF	DomainNet-126							
				R→C	R→P	P→C	C→S	S→P	R→S	P→R	Avg.
Source Only	✗	✗	✗	56.6	61.9	58.2	50.4	49.3	46.2	73.6	56.6
FeCAM [11]	✗	✓	✗	15.6	14.3	9.56	13.2	14.6	12.7	29.4	15.6
GPUE [29]	✓	✗	✓	47.9	51.8	51.4	45.2	53.4	37.3	57.2	49.2
ProCA [28]	✓	✓	✗	56.9	56.8	58.1	44.1	33.6	41.7	73.9	52.2
QSI (Ours)	✓	✓	✓	**64.3**	**64.2**	**67.2**	**54.1**	**57.3**	**52.2**	**77.5**	**62.4**

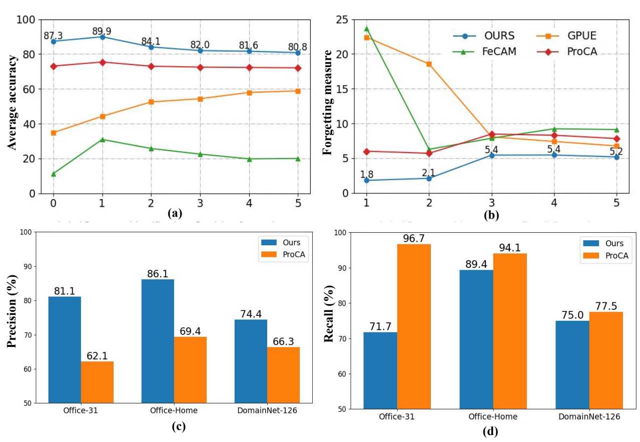

Figure 4: We evaluate CI metrics on Office-Home (Ar→ Pr) and identification metrics for prototype labels on 3 datasets. (a) Average accuracy (AA) of different methods in 6 incremental tasks. (b) Forgetting Measure (FM) of different methods in 6 incremental tasks. (c) Identification precision between our method and ProCA in the PDA scenario. (d) Identification recall on 3 datasets

not handle domain shift, resulting in a significant drop in classification accuracy compared to DA methods. In contrast, our method effectively preserves old knowledge and reduces the domain gap, achieving the best performance.

Confusion Matrix for Negative Datasets As described in Sec. 3.2, some strongly augmented samples x_{ag} are misclassified by the source model M_p, and confusion matrices are then constructed for each domain based on the predictions. These misclassifications (Fig.5) provide sufficient information to construct negative pairs $(y_s, \tilde{y_s})$ for each category y_s. For each row in the confusion matrix, corresponding to a different y_s, the column with the highest number of misclassified images is selected as $\tilde{y_s}$ to form the negative pair.

Confidence Accumulation. An essential challenge in Unsupervised Partial Domain Adaptation is the identification of shared classes, where the target label set is a subset of the source label set. We compare our identification strategy with the method in ProCA [28] on three datasets in Fig. 4. Two metrics are employed to assess the performance: 1) Recall: The percentage of correctly detected categories out of the total correct categories. 2) Precision: The percentage of correctly detected categories out of the total detected categories. Fig. 4 shows that our approach outperforms prior works with a notable margin in precision among all datasets. Although, ProCA reach a higher results in recall, the low precision indicates that the ability to filter out correct labels is relatively weak, and just simply selects labels widely without distinguishing them.

Semantic Restructuring. Fig. 6 visualizes the positioning of samples in the feature space after being constructed by Semantic Restructuring, with source dataset represented in blue and negative dataset depicted in red. Obviously, the constructed samples fill in the latent space (i.e. the feature regions that are challenging for image classification) among source categories(As shown in Fig. 6, a comparison between sub-figure (a) and sub-figure (b)). By training

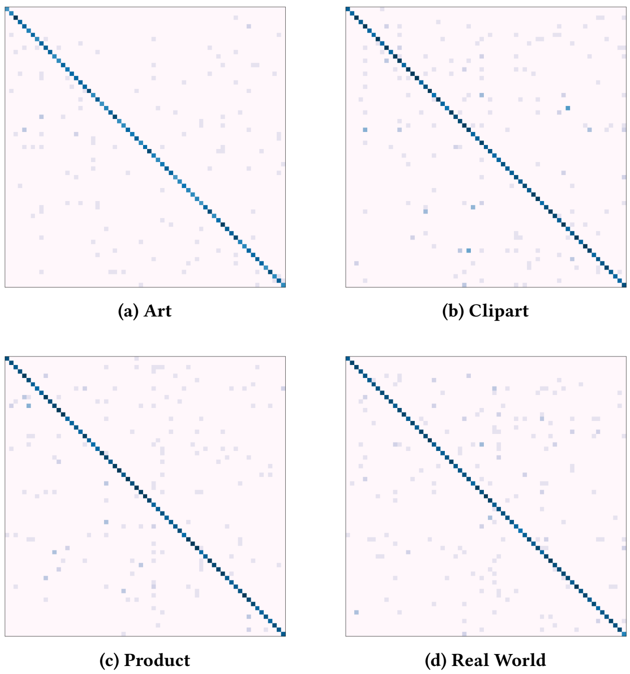

Figure 5: In the Office-Home dataset, the confusion matrices are obtained by feeding augmented samples from each domain into the corresponding model. Brighter colors outside the diagonal indicate higher misclassification counts.

Table 4: Ablation studies of our modules on Office-Home. The first row is backbone. The second row is label refinement in Sec. 3.3. The third row is Invariance Quantification in Sec. 3.2. The fourth and fifth rows are cross entropy in memory buffer and contrastive loss respectively, both in Sec. 3.3.

L_{CE}	Label Refinement	Invariance Quantification	L_{bank}	L_{con}	Avg. acc.
✓	✗	✗	✗	✗	61.7
✓	✓	✗	✗	✗	63.6
✓	✓	✓	✗	✗	66.4
✓	✓	✓	✓	✗	69.0
✓	✓	✓	✓	✓	**71.8**

with both source and negative datasets, the model is capable of clustering categories within the latent space, proposing a clear metric (Eq. (3)) to measure the positions beyond the decision boundaries.

Ablation Study. We show the results of our ablation study on the Office-Home in Tab. 4 to estimate the effectiveness of our modules. First, we utilize the ResNet50 model with pseudo-label cross-entropy as the backbone. By refining the pseudo-labels, we improve the accuracy by +1.9% . Next, we apply our weighting strategy to the classification loss, which further boosts the accuracy by +2.8%. To alleviate catastrophic forgetting, we modify the total loss function to include a weighted cross entropy for stored images, leading to an additional increase in accuracy of +2.6% . Finally, by employing a contrastive loss to amplify category gaps, we achieve the highest accuracy of 71.8%.

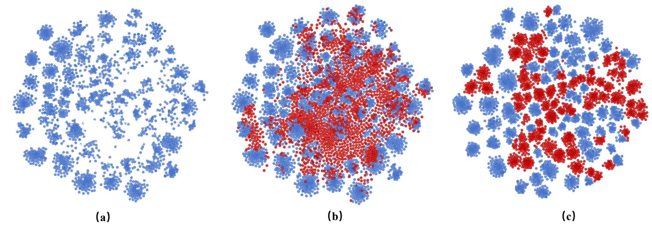

Figure 6: We present a visualization of the feature space in the source domain (Art) of Office-Home. (a) Only the source samples in the feature space of model M_p. (b) Both the source and negative samples in the feature space of model M_p. (c) Both the source and negative samples in the feature space of model M_n.

Table 5: Classification accuracy (%) of different methods with or without our weighting strategy on Office-Home

	without weight	with weight
ProCA [28]	64.5	68.2
QSI (Ours)	**70.4**	**71.8**

Weighting Strategy. Unsupervised scenarios often bring a significant amount of noise to the samples. To this end, we introduce a weighting strategy to denoise and sift high-quality samples. We integrated our strategies into different methods and compared the model's performance before and after the integration, as shown in Tab. 5. Compared to the original method's accuracy, our strategy significantly enhances the model's performance by +3.7% in ProCA [28] and by +1.4% in ours.

5 Conclusion

We have introduced a novel approach to address a practical scenario called Source-Free Class Incremental Domain Adaptation (SFCIDA). In our proposed method, we employ a novel Semantic Restruction strategy to build a negative dataset, which is essential for configuring decision boundaries. Based on the model trained on both positive and negative datasets, we conduct Invariance Quantification strategy to extract samples' confidence from the model's output and assign weights to the samples, which are then applied throughout the training process. Despite the absence of source data, we successfully mitigated catastrophic forgetting and domain gap simultaneously. Experiments on three datasets show that our method significantly outperforms prior approaches, demonstrating its robustness and effectiveness.

6 Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant T2125006 and Grant 42401415; in part by Shenzhen Science and Technology Program under Grant KCXFZ20240903093759004 and Grant KJZD20230923115106012; and in part by Jiangsu Innovation Capacity Building Program under Project BM2022028.

References

[1] Lucas Caccia, Eugene Belilovsky, Massimo Caccia, and Joelle Pineau. 2020. Online Learned Continual Compression with Adaptive Quantization Modules. In *Proceedings of the 37th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 119)*, Hal Daumé III and Aarti Singh (Eds.). PMLR, 1240–1250. https://proceedings.mlr.press/v119/caccia20a.html

[2] Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Michael I. Jordan. 2018. Partial Transfer Learning With Selective Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[3] Zhangjie Cao, Lijia Ma, Mingsheng Long, and Jianmin Wang. 2018. Partial Adversarial Domain Adaptation. In *Computer Vision – ECCV 2018*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer International Publishing, Cham, 139–155.

[4] Zhangjie Cao, Kaichao You, Mingsheng Long, Jianmin Wang, and Qiang Yang. 2019. Learning to Transfer Examples for Partial Domain Adaptation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2980–2989. doi:10.1109/CVPR.2019.00310

[5] Chao Chen, Zhihang Fu, Zhihong Chen, Sheng Jin, Zhaowei Cheng, Xinyu Jin, and Xiansheng Hua. 2019. HoMM: Higher-order Moment Matching for Unsupervised Domain Adaptation. In *AAAI Conference on Artificial Intelligence*. https://api.semanticscholar.org/CorpusID:209500519

[6] Dian Chen, Dequan Wang, Trevor Darrell, and Sayna Ebrahimi. 2022. Contrastive Test-time Adaptation. In *CVPR*.

[7] Wendong Chen and Haifeng Hu. 2020. Unsupervised Domain Adaptation via Discriminative Classes-Center Feature Learning in Adversarial Network. *Neural Processing Letters* 52, 1 (01 Aug 2020), 467–483. doi:10.1007/s11063-020-10266-z

[8] Koby Crammer, Michael Kearns, and Jennifer Wortman Vaughan. 2006. Learning from Multiple Sources. *J. Mach. Learn. Res.* 9 (2006), 1757–1774. https://api.semanticscholar.org/CorpusID:2709712

[9] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. 2016. Domain-Adversarial Training of Neural Networks. *Journal of Machine Learning Research* 17, 59 (2016), 1–35. http://jmlr.org/papers/v17/15-239.html

[10] Yaroslav Ganin, E. Ustinova, Hana Ajakan, Pascal Germain, H. Larochelle, François Laviolette, Mario Marchand, and Victor S. Lempitsky. 2015. Domain-Adversarial Training of Neural Networks. In *Journal of machine learning research*. https://api.semanticscholar.org/CorpusID:2871880

[11] Dipam Goswami, Yuyang Liu, Bartlomiej Twardowski, and Joost van de Weijer. 2023. FeCAM: Exploiting the Heterogeneity of Class Distributions in Exemplar-Free Continual Learning. *ArXiv* abs/2309.14062 (2023). https://api.semanticscholar.org/CorpusID:262464670

[12] Raia Hadsell, Dushyant Rao, Andrei A. Rusu, and Razvan Pascanu. 2020. Embracing Change: Continual Learning in Deep Neural Networks. *Trends in Cognitive Sciences* 24, 12 (2020), 1028–1040. doi:10.1016/j.tics.2020.09.004

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778. doi:10.1109/CVPR.2016.90

[14] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. 2019. Deep Anomaly Detection with Outlier Exposure. *Proceedings of the International Conference on Learning Representations* (2019).

[15] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei A. Efros, and Trevor Darrell. 2017. CyCADA: Cycle-Consistent Adversarial Domain Adaptation. *ArXiv* abs/1711.03213 (2017). https://api.semanticscholar.org/CorpusID:7646250

[16] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G. Hauptmann. 2019. Contrastive Adaptation Network for Unsupervised Domain Adaptation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 4888–4897. doi:10.1109/CVPR.2019.00503

[17] Myeonginn Kang and Seokho Kang. 2021. Data-free knowledge distillation in neural networks for regression. *Expert Systems with Applications* 175 (2021), 114813. doi:10.1016/j.eswa.2021.114813

[18] Aditya Khosla, Tinghui Zhou, Tomasz Malisiewicz, Alexei A. Efros, and Antonio Torralba. 2012. Undoing the Damage of Dataset Bias. In *European Conference on Computer Vision*. https://api.semanticscholar.org/CorpusID:9286850

[19] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences* 114, 13 (2017), 3521–3526. doi:10.1073/pnas.1611835114 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1611835114

[20] Dhireesha Kudithipudi, Mario Aguilar-Simon, Jonathan Babb, Maxim Bazhenov, Douglas Blackiston, Josh Bongard, Andrew P. Brna, Suraj Chakravarthi Raja, Nick Cheney, Jeff Clune, Anurag Daram, Stefano Fusi, Peter Helfer, Leslie Kay, Nicholas Ketz, Zsolt Kira, Soheil Kolouri, Jeffrey L. Krichmar, Sam Kriegman, Michael Levin, Sandeep Madireddy, Santosh Manicka, Ali Marjaninejad, Bruce McNaughton, Risto Miikkulainen, Zaneta Navratilova, Tej Pandit, Alice Parker, Praveen K. Pilly, Sebastian Risi, Terrence J. Sejnowski, Andrea Soltoggio, Nicholas Soures, Andreas S. Tolias, Darío Urbina-Meléndez, Francisco J. Valero-Cuevas, Gido M. van de Ven, Joshua T. Vogelstein, Felix Wang, Ron Weiss, Angel Yanguas-Gil, Xinyun Zou, and Hava Siegelmann. 2022. Biological underpinnings for lifelong learning machines. *Nature Machine Intelligence* 4, 3 (01 Mar 2022), 196–210. doi:10.1038/s42256-022-00452-0

[21] Jogendra Nath Kundu, Nishank Lakkakula, and R. Venkatesh Babu. 2019. UM-Adapt: Unsupervised Multi-Task Adaptation Using Adversarial Cross-Task Distillation. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 1436–1445. https://api.semanticscholar.org/CorpusID:199543275

[22] Jogendra Nath Kundu, Naveen Venkat, Revanur Ambareesh, V. RahulM., and R. Venkatesh Babu. 2020. Towards Inheritable Models for Open-Set Domain Adaptation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 12373–12382. https://api.semanticscholar.org/CorpusID:215548694

[23] Jogendra Nath Kundu, Rahul Venkatesh, Naveen Venkat, Ambareesh Revanur, and R. Venkatesh Babu. 2020. Class-Incremental Domain Adaptation. *ArXiv* abs/2008.01389 (2020). https://api.semanticscholar.org/CorpusID:220961583

[24] Seiichi Kuroki, Nontawat Charoenphakdee, Han Bao, Junya Honda, Issei Sato, and Masashi Sugiyama. 2019. Unsupervised domain adaptation based on source-guided discrepancy. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence* (Honolulu, Hawaii, USA) *(AAAI'19/IAAI'19/EAAI'19)*. AAAI Press, Article 506, 8 pages. doi:10.1609/aaai.v33i01.33014122

[25] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324. doi:10.1109/5.726791

[26] Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. 2020. Model Adaptation: Unsupervised Domain Adaptation Without Source Data. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 9638–9647. doi:10.1109/CVPR42600.2020.00966

[27] Jian Liang, Dapeng Hu, and Jiashi Feng. 2020. Do We Really Need to Access the Source Data? Source Hypothesis Transfer for Unsupervised Domain Adaptation. In *International Conference on Machine Learning*. https://api.semanticscholar.org/CorpusID:211205159

[28] Hongbin Lin, Yifan Zhang, Zhen Qiu, Shuaicheng Niu, Chuang Gan, Yanxia Liu, and Mingkui Tan. 2022. Prototype-Guided Continual Adaptation for Class-Incremental Unsupervised Domain Adaptation. In *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer Nature Switzerland, Cham, 351–368.

[29] Mattia Litrico, Alessio Del Bue, and Pietro Morerio. 2023. Guiding Pseudo-labels with Uncertainty Estimation for Source-free Unsupervised Domain Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[30] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. 2015. Learning Transferable Features with Deep Adaptation Networks. In *Proceedings of the 32nd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 37)*, Francis Bach and David Blei (Eds.). PMLR, Lille, France, 97–105. https://proceedings.mlr.press/v37/long15.html

[31] Noel Loo, Siddharth Swaroop, and Richard E. Turner. 2020. Generalized Variational Continual Learning. *ArXiv* abs/2011.12328 (2020). https://api.semanticscholar.org/CorpusID:227162606

[32] David Lopez-Paz and Marc' Aurelio Ranzato. 2017. Gradient Episodic Memory for Continual Learning. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/f87522788a2be2d171666752f97ddebb-Paper.pdf

[33] Divyam Madaan, Jaehong Yoon, Yuanchun Li, Yunxin Liu, and Sung Ju Hwang. 2022. Representational Continuity for Unsupervised Continual Learning. In *International Conference on Learning Representations*. https://openreview.net/forum?id=9Hrka5PA7LW

[34] Arun Mallya, Dillon Davis, and Svetlana Lazebnik. 2018. Piggyback: Adapting a Single Network to Multiple Tasks by Learning to Mask Weights. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

[35] Jogendra Nath Kundu, Naveen Venkat, M. V. Rahul, and R. Venkatesh Babu. 2020. Universal Source-Free Domain Adaptation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 4543–4552. doi:10.1109/CVPR42600.2020.00460

[36] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. 2019. Moment Matching for Multi-Source Domain Adaptation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 1406–1415. doi:10.1109/ICCV.2019.00149

[37] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, G. Sperl, and Christoph H. Lampert. 2016. iCaRL: Incremental Classifier and Representation Learning. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 5533–5542. https://api.semanticscholar.org/CorpusID:206596260

[38] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell, and Kate Saenko. 2019. Semi-supervised Domain Adaptation via Minimax Entropy. *ICCV* (2019).

[39] Swami Sankaranarayanan, Yogesh Balaji, Carlos D. Castillo, and Rama Chellappa. 2018. Generate to Adapt: Aligning Domains Using Generative Adversarial Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8503–8512. doi:10.1109/CVPR.2018.00887

[40] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. 2017. Continual Learning with Deep Generative Replay. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/0efbe98067c6c73dba1250d2beaa81f9-Paper.pdf

[41] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei A. Efros, and Moritz Hardt. 2019. Test-Time Training for Out-of-Distribution Generalization. *ArXiv* abs/1909.13231 (2019). https://api.semanticscholar.org/CorpusID:203593944

[42] Song Tang, Wenxin Su, Mao Ye, and Xiatian Zhu. 2024. Source-free domain adaptation with frozen multimodal foundation model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 23711–23720.

[43] Antonio Torralba and Alexei A. Efros. 2011. Unbiased look at dataset bias. In *Computer Vision & Pattern Recognition*.

[44] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. 2017. Adversarial Discriminative Domain Adaptation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2962–2971. doi:10.1109/CVPR.2017.316

[45] Eric Tzeng, Judy Hoffman, N. Zhang, Kate Saenko, and Trevor Darrell. 2014. Deep Domain Confusion: Maximizing for Domain Invariance. *ArXiv* abs/1412.3474 (2014). https://api.semanticscholar.org/CorpusID:17169365

[46] Aäron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation Learning with Contrastive Predictive Coding. *ArXiv* abs/1807.03748 (2018). https://api.semanticscholar.org/CorpusID:49670925

[47] Dequan Wang, Shaoteng Liu, Sayna Ebrahimi, Evan Shelhamer, and Trevor Darrell. 2021. On-target Adaptation. *ArXiv* abs/2109.01087 (2021). https://api.semanticscholar.org/CorpusID:237386019

[48] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. 2021. Tent: Fully Test-Time Adaptation by Entropy Minimization. In *International Conference on Learning Representations*. https://openreview.net/forum?id=uXl3bZLkr3c

[49] Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. 2021. Generalized Source-free Domain Adaptation. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), 8958–8967. https://api.semanticscholar.org/CorpusID:236881316

[50] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 2242–2251. doi:10.1109/ICCV.2017.244