

Multi-Domain Learning from Insufficient Annotations

Rui He^{a,b}, Shengcai Liu^{a,*}, Jiahao Wu^{a,c}, Shan He^b and Ke Tang^a

^aDepartment of Computer Science and Engineering, Southern University of Science and Technology

^bSchool of Computer Science, University of Birmingham

^cDepartment of Computing, The Hong Kong Polytechnic University

ORCID ID: Rui He <https://orcid.org/0000-0002-6338-1387>, Shengcai Liu <https://orcid.org/0000-0002-4223-2438>

Abstract. Multi-domain learning (MDL) refers to simultaneously constructing a model or a set of models on datasets collected from different domains. Conventional approaches emphasize domain-shared information extraction and domain-private information preservation, following the shared-private framework (SP models), which offers significant advantages over single-domain learning. However, the limited availability of annotated data in each domain considerably hinders the effectiveness of conventional supervised MDL approaches in real-world applications. In this paper, we introduce a novel method called multi-domain contrastive learning (MDCL) to alleviate the impact of insufficient annotations by capturing both semantic and structural information from both labeled and unlabeled data. Specifically, MDCL comprises two modules: inter-domain semantic alignment and intra-domain contrast. The former aims to align annotated instances of the same semantic category from distinct domains within a shared hidden space, while the latter focuses on learning a cluster structure of unlabeled instances in a private hidden space for each domain. MDCL is readily compatible with many SP models, requiring no additional model parameters and allowing for end-to-end training. Experimental results across five textual and image multi-domain datasets demonstrate that MDCL brings noticeable improvement over various SP models. Furthermore, MDCL can further be employed in multi-domain active learning (MDAL) to achieve a superior initialization, eventually leading to better overall performance.

1 Introduction

In many machine learning tasks, models are built on datasets collected from various data sources with different distributions, known as domains. For instance, texts from different sources like news articles, social media posts, and scientific papers constitute distinct domains in natural language processing. In computer vision, images of differing styles, such as sketches, cartoons, art paintings, and camera photos [1] are considered distinct domains. While each domain possesses unique information, they often share a significant amount of information with other domains. Naive solutions involve jointly building a single model across domains or independently constructing models for each domain, as is done in conventional single domain learning (SDL) approaches. However, joint training may neglect the unique information on each domain, while independent training disregards the correlations among domains [2]. To address

these shortcomings, multi-domain learning (MDL)[3] has been proposed to simultaneously capture domain-shared and domain-private information. Most existing MDL works concentrate on sharing information among domains while preserving domain-private information through models under the shared-private framework (SP models) [2, 4]. Typically, following the concept of domain adaptation (DA) [5], shared information can be captured through distribution alignment across domains, allowing several DA methods to be utilized in MDL. Besides, private information is usually managed by a private component of the model for each domain. Accounting for both types of information has led to significant performance improvements over joint and independent training in the past decade [2].

In real-world applications, obtaining a sufficiently labeled dataset can be costly, even within a single domain [6, 7, 8]. This issue exacerbated in MDL since constructing labeled multi-domain dataset is even more challenging due to the difficulty in accessing data from multiple domain experts [9, 10, 11]. For instance, in the case of multi-domain medical image datasets [9], the high cost of manual annotations from medical experts across various research fields is just one of the challenges. The varying privacy and legal concerns, quality assurance processes, and labeling tools across domains also entail additional costs. The aforementioned MDL approaches face challenges in this high-cost scenario as they heavily rely on fully supervised training from a relatively sufficiently annotated multi-domain dataset. Therefore, a natural question arises: **can we perform cost-efficient MDL with insufficient annotated data?**

To the best of our knowledge, only a few works address the issue of insufficient annotations from multiple domains. Some works utilize contrastive [12, 13] and semi-supervised [14] learning to alleviate the impact of insufficient annotations across domains. The key is to utilize unlabeled data to improve the performance. However, these works focus on the domain adaptation problem, where only the target domain performance is concerned and the domain-private information is usually neglected. Other works propose to utilize active learning (AL) [15] in MDL, which is referred to as multi-domain active learning (MDAL) [16]. Given a budget for annotation, MDAL begins with a small set of labeled instances and iteratively selects the new instances for model building. However, without a good initial model trained on insufficient labeled instances, the selection process is likely to be biased and unreliable in the subsequent MDAL iterations. In summary, no method is readily applicable for MDL with insufficient annotations so far.

In this paper, we propose a novel MDL approach, called multi-domain contrastive learning (MDCL), to construct neural network

* Corresponding Author. Email: liusc3@sustech.edu.cn

models on a limited number of labeled instances from each domain. Figure 1 presents an intuitive understanding of MDCL, where the semantic and structural information are respectively captured from labeled and unlabeled data. Specifically, MDCL comprises two components: a supervised contrastive loss to align instances of the same category from different domains within a shared hidden space and an unsupervised contrastive loss that focuses on learning the cluster structure of instances from the same domain in a private hidden space. By integrating both components, MDCL can learn a well-aligned representation from insufficient annotations. Importantly, MDCL is readily compatible with various share-private (SP) models for MDL [17, 2, 4], requiring no additional model parameters and allowing for end-to-end training. Experimental results across five textual and image multi-domain datasets demonstrate that MDCL brings noticeable improvement over various SP models.

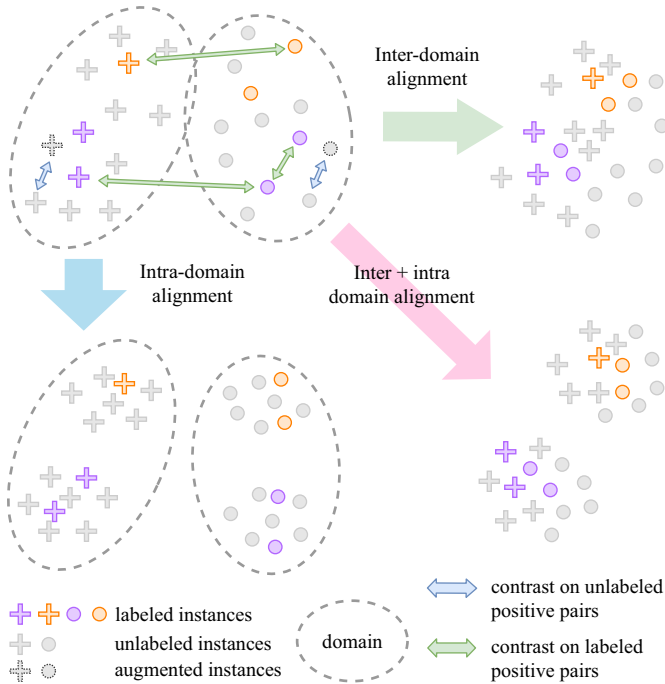


Figure 1: Intuitive understanding of MDCL. An illustrative example in the hidden space. (1) Inter-domain alignment aims to align items within the same category but from different domains closer to each other. (2) Intra-domain contrast aims to maintain a cluster structure in each domain and make instances more separable.

The main contributions of this paper are summarized as follows:

- We introduce a novel approach called MDCL to build neural network models on insufficient labels in MDL. MDCL is readily compatible with many SP models, requiring no additional model parameters and allowing for end-to-end training.
- Experimental results demonstrate that MDCL yields noticeable improvement over many SP models across five textual and image multi-domain datasets.
- Moreover, MDCL can be employed in MDAL to achieve a superior initialization and consequently result a better performance.

2 Related Work

This work is closely related to multi-domain learning (MDL) [3] and contrastive learning [18], which will be briefly reviewed in the following sections. The representative works and solutions would be included.

2.1 Multi-Domain Learning

MDL [3] differs from both multi-task learning (MTL) [19], domain adaptation (DA) [5] and domain generalization [20, 21, 22] in that MDL focuses on performing the same task across all domains simultaneously. By leveraging shared information from multiple related domains, MDL enhances the model’s performance on each individual domain.

Domain-shared information learning The domain-shared information learning is a crucial step in MDL, which aims to extract common features across different domains. This can be achieved by using a shared feature extractor that aligns the marginal distributions of different domains. The alignment is typically performed by minimizing the maximum mean discrepancy (MMD) between domains [23]. Adversarial training can also be used to handle distribution alignment for more than two domains. The extracted features are effective if a discriminator cannot distinguish the domain of the instances. Several models based on this intuition from domain adaptation (DA) can be adapted for use in MDL. Examples include Domain-adversarial neural networks (DANN) [24] and Adversarial discriminative domain adaptation (ADDA) [25]. Additionally, Feng *et al.* [26] proposed a method that directly learns representations for MDL using this distribution alignment approach.

Domain-private information imparted In MDL, the domain-invariant representations may not fully capture the unique information of each domain. To address this issue, several methods have been proposed to capture both domain-shared and domain-private information. One approach is concatenating the domain-private information with the domain-shared information for making predictions. For example, Domain separation networks (DSN) [17] use separate feature extractors for domain-private and domain-shared information in a DA setting. Adversarial shared-private model (ASP-MTL) [2] and Multinomial adversarial networks [4] apply this idea to MDL. This architecture that concatenates representations through domain-shared and domain-private feature extractors is commonly known as the **share-private framework** (SP models). Empirical study [16] has shown the superiority of the share-private framework.

Multi-domain active learning To address the labeling cost in MDL, MDAL [27] iteratively selects informative instances to reduce the labeling burden. A comprehensive comparative study [16] has shown that AL can improve performance compared to passive models. The MAN model under the share-private framework outperforms other methods in both passive and active scenarios. Moreover, taking into account the evaluation of multi-domain informativeness, perturbation-based two-stage multi-domain active learning (P2S-MDAL) [28] serves as the first ad hoc active learning strategy for MDAL. Some works have also applied AL to multiple domains but without considering information-sharing schemes. They either build classifiers on each domain [29] or use a single model for all domains [30].

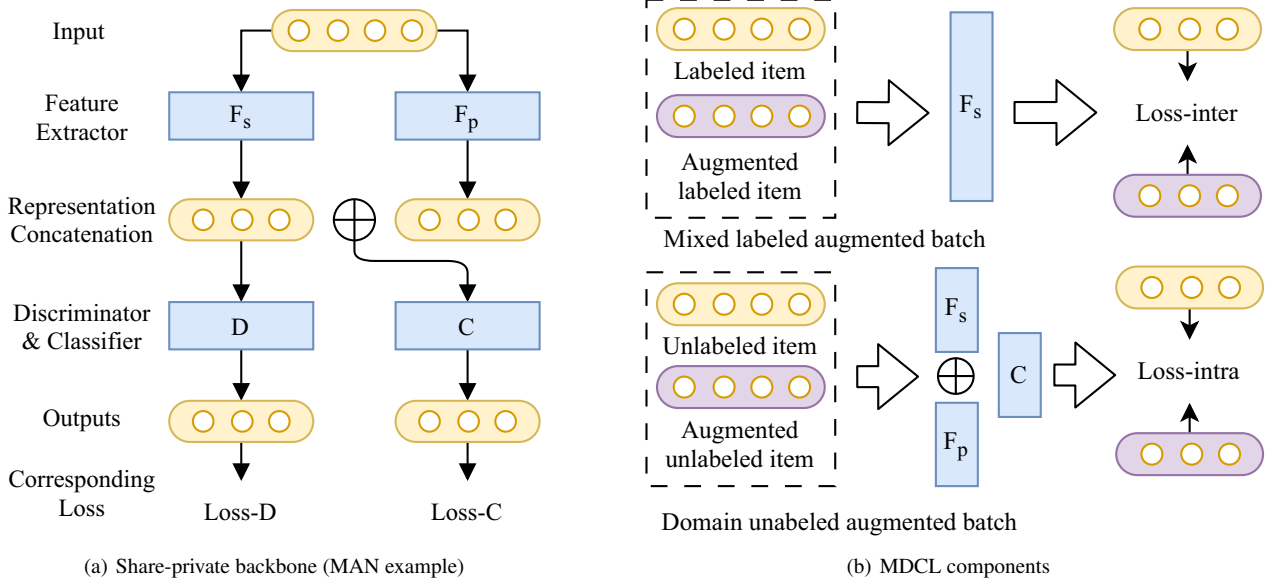


Figure 2: Overview of MDCL method. (a) As a representative model for MDL under the share-private framework, MAN is taken as an illustration. (b) MDCL comprises two components: an inter-domain semantic alignment and an intra-domain representation learning.

2.2 Contrastive Learning

Contrastive learning on single domain Contrastive learning (CL) is one of the most prevalent paradigms of self-supervised learning. The general learning paradigm of CL always performs a contrasting among different views augmented from the original data [18, 31, 32]. Due to the capability of extracting supervision signals from unlabeled data, CL is explored in a wide range of fields and achieves promising performance [18, 31, 33], including multi-domain learning.

Contrastive learning across domains Contrastive learning has been applied in DA, where the learning process is guided by a fully labeled source domain to improve performance on the unlabeled or sparsely labeled target domain. Self-supervised tasks can be directly used to jointly learn a shared feature extractor [34]. Besides, class prototypes [12, 13, 14] can be generated to align the categorical distributions between the source and target domains. Meanwhile, instance contrastive alignment [13, 35] can also be used to build a feature extractor without using labels on the target domain to learn the target domain representation.

These studies are most close to our work technically. However, they only focus on domain adaptation (DA) setting and only employ a single feature extractor for both domains. They do not consider the MDL problem, where both shared and private information should be taken into account and further be handled by certain model components. Besides, there is no fully labeled source domain in our setting to extract reliable class prototypes as in [12, 13, 14].

3 Problem Formulation

Following the definition given in [16], we re-formulate the problem of multi-domain learning, which further considers the insufficient annotations and utilizes both labeled and unlabeled instances. The formulation is written as follows:

In multi-domain learning, there are K different data sources (domains) $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_K\}$. A set of data pools $\mathcal{P} =$

$\{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_K\}$ containing both labeled and unlabeled data is collected from \mathcal{D} in advance. The labeled data from each pool constitute a labeled data set $\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_K\}$. Considering the scenario with limited labeled instances, the number of the labeled instances is much smaller than the collected data pool, i.e. $|\mathcal{L}| \ll |\mathcal{P}|$. In this scenario, MDL is to find a set of models $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_K\}$ for K domains by capturing common knowledge on the unlabeled data pools \mathcal{P} and limited labeled instances \mathcal{L} of different domains, which can be expressed as follows:

$$\min_{\mathcal{M}} Loss_{sup}(\mathcal{M}; \mathcal{L}) + \Omega(\mathcal{M}; \mathcal{P}, \mathcal{L}) \quad (1)$$

$Loss_{sup}(\mathcal{M}; \mathcal{L})$ denotes the supervised loss on the labeled set \mathcal{L} . $\Omega(\mathcal{M}; \mathcal{P}, \mathcal{L})$ denotes a designed loss on the set of data pools \mathcal{P} and limited labeled instances \mathcal{L} for capturing the common knowledge through model \mathcal{M} .

4 Methodology

This section introduces our novel multi-domain contrastive learning (MDCL) method for MDL with limited annotations. The key is to maximally utilize the unlabeled data pools and limited labeled instances to learn a set of models for K domains. The labeled data can be used to capture the semantic information and align the distributions across domains. The unlabeled data can further be used to learn the structural information of the data distribution and preserve the local information. MDCL introduces an inter-domain semantic alignment loss and an intra-domain representation learning loss to utilize the unlabeled and limited labeled data maximally. As we introduced in Section 2.1, the share-private framework is the commonly used architecture for MDL. MDCL takes the share-private framework as the backbone, where our designed loss can be easily integrated into the existing MDL methods as a plug-and-play component and trained in an end-to-end manner without introducing additional model parameters. The inter-domain semantic alignment and the intra-domain representation learning loss are based on the contrastive learning to guild the feature extraction. The overview of MDCL is depicted in Fig. 2.

The remainder of this section will first provide an overview of the share-private framework, which serves as the backbone of MDCL, and then describe our method and its components in detail.

4.1 Share-Private Framework

The popular solution for MDL is the share-private framework (SP structure), which uses two types of feature extractors to handle both domain-shared and domain-private information. It effectively combines shared and private feature extractors to capture domain-shared and domain-specific information in MDL. The shared-information across domains could improve overall performance by training a shared extractor, while domain-specific information may be lost during the joint training. Whereas, the private extractor can efficiently preserve domain-specific information. The representations from both extractors are then concatenated for inference. Consequently, the SP model outperforms both the single shared feature extractor model (e.g., DANN [24]) commonly used in DA, as well as private models. Some models have unique classifiers on each domain (e.g., ASP-MTL [2]), while others share a single classifier for all domains (e.g., MAN [4] and CAN [36]). As a representative and efficient SP model, the structure of MAN is illustrated in Fig. 2(a).

4.2 Inter-Domain Semantic Alignment

The conventional idea of MDL is to utilize a shared feature extractor [24] to align the marginal distributions of domains. However, this approach may map instances within the same category far apart. Therefore, it is crucial to consider the conditional distribution as well [37, 36], meaning that items with identical labels should be mapped closely in the latent space, which is referred to as inter-domain semantic alignment.

Different from the previous methods for DA [12, 13, 14], in the scenario with limited labeled instances in MDL, obtaining reliable class prototypes is not feasible. In this situation, we could employ a NT-Xent contrastive loss [18] in a supervised manner [32] directly on the limited labels to proceed semantic alignment and bypass the construction of prototypes. Pairs of items should be mapped together as long as they belong to the same category, regardless of the domain they originate from. As illustrated in Fig. 2(b), we introduce an inter-domain semantic alignment loss on the domain-shared representation space, which can be expressed as:

$$\mathcal{L}_{\text{inter}} = \sum_{i \in I} \mathcal{L}_i = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p / \tau)}{\sum_{a \in A(i)} \exp(\mathbf{z}_i \cdot \mathbf{z}_a / \tau)} \quad (2)$$

where i is the index of a sample in the augmented batch $I \equiv \{1 \dots 2N\}$, and N is the batch size. $A(i) \equiv I \setminus \{i\}$. $P(i)$ is the index set of positives $\{p \in A(i) : \tilde{\mathbf{y}}_p = \tilde{\mathbf{y}}_i\}$, and $|P(i)|$ is the cardinality. \mathbf{z}_l is the representation of the l -th item of the augmented batch extracted from the domain-shared feature extractor. τ is a scaler temperature factor. The original batch is chosen from a mixed labeled dataset that encompasses all domains, and then augmented to obtain I . This means that instances with the same label but from different domains can be included in the same batch. Using the semantic alignment loss, the shared feature extractor can align hidden representations of items within the same class to be close to each other.

However, solely using the proposed inter-domain semantic alignment loss may not preserve the local manifold structure and only affect on the few labeled instances. As illustrated in Fig. 1, solely using

inter-domain alignment might result in the unlabeled representations being mixed and losing the cluster structures.

4.3 Intra-Domain Representative Learning

Within each domain, unlabeled data can be utilized for intra-domain contrastive alignment. Unlike the previous inter-domain instance alignment, we use the classifier outputs instead of intermediate representations to align items within each domain. This approach goes beyond the intermediate representation space and ensures that the same item with different augmentations has a similar class assignment. The contrastive alignment ensures a low density (uniformity) and form robust clusters (alignment) in the representation space [38]. As a result, the domain representations are more robust to noise and augmentations. To achieve this unsupervised alignment, we also leverage the NT-Xent contrastive loss [18]. As illustrated in Fig. 2(b), we introduce an intra-domain contrastive loss on the domain-private output space, which can be expressed as:

$$\mathcal{L}_{\text{intra}} = \sum_{i \in I} \mathcal{L}_i = - \sum_{i \in I} \log \frac{\exp(\mathbf{o}_i \cdot \mathbf{o}_{j(i)} / \tau)}{\sum_{a \in A(i)} \exp(\mathbf{o}_i \cdot \mathbf{o}_a / \tau)} \quad (3)$$

where \mathbf{o}_i is the output of the classifier of the corresponding domain.

However, solely using the proposed intra-domain contrastive loss only affects the alignment in each domain. As illustrated in Fig. 1, the items from different domains would still be far from each other in the representation space after solely using intra-domain alignment.

4.4 Overall framework and Pseudocode

Our method serves as a plug-and-play solution for MDL and can be applied to different SP models. We provide the pseudocode for our MDAL in Algorithm 1. In the pseudocode, we also take MAN [4] as the backbone example for a better explanation.

5 Experiments

5.1 Research Questions

We evaluate our method on the following research questions:

1. As a plug-and-play method, can MDCL enhance the performance of various models under the SP structure with the limited number of labeled instances (around 5%-20%) or extremely few labeled instances (around 1%)? (Section 5.3)
2. How does each component of MDCL affect the performance? (Section 5.4)
3. Given a further labeling budget, MDAL could be utilized. Can MDCL improve the entire MDAL process with a relatively large number of unlabeled instances (5%-50%)? (Section 5.5)

5.2 Experimental Setup

5.2.1 Dataset

We evaluate our proposed MDCL method on several popular multi-domain textual and image datasets, namely Amazon [39], MNIST-USPS [25], Office-Home [40], FDUMTL [2], and PACS [1]. The Amazon dataset consists of four textual domains, each containing two categories, with instances encoded to a vector representation of length 5000. MNIST-USPS comprises two image domains with ten categories each, and instances are encoded to a vector representation

Table 1: The hyperparameters used for MDCL.

| Datasets | Optimizer | Learning Rate | Learning Rate Decay | Batch Size | Weight Decay | Early Stopping | Inter-Domain λ & τ | Intra-Domain λ & τ |
|-------------|-----------|---------------|---------------------|------------|--------------|----------------|---------------------------------|---------------------------------|
| Amazon | Adam | 3e-4 | False | 8 | 0.05 | 20 | 1/0.1 | 1/0.01 |
| MNIST-USPS | Adam | 3e-3 | 0.33 | 8 | 0.001 | 30 | 0.1/0.1 | 1/0.1 |
| Office-Home | Adam | 1e-2 | 0.33 | 8 | 0.001 | 15 | 1/0.01 | 1/0.01 |
| FDUMTL | Adam | 3e-4 | 0.1 | 8 | 0.001 | 30 | 0.1/0.01 | 1/0.01 |
| PACS | SGD | 1e-3 | 0.1 | 8 | 0.001 | 15 | 0.1/1 | 1/0.1 |

Algorithm 1 Multi-Domain Contrastive Learning.

Require: labeled dataset \mathbb{X} ; unlabeled dataset \mathbb{U} ; hyperparameter $\lambda_d > 0$, $\lambda_{inter} > 0$, $\lambda_{intra} > 0$, $k_{inter} \in \mathbb{N}$, $k_{adv} \in \mathbb{N}$

- 1: **repeat**
- 2: \triangleright Inter-domain alignment
- 3: **for** $iter = 1$ to k_{inter} **do**
- 4: $l_{inter} = 0$
- 5: Sample a mini-batch $(\mathbf{x}, \mathbf{y}) \sim \mathbb{X}$ \triangleright From mixed labeled dataset
- 6: $\mathbf{x}', \mathbf{y}' = \text{Aug}(\mathbf{x}, \mathbf{y})$ \triangleright Augmented batch
- 7: $\mathbf{z}_s = \mathcal{F}_s(\mathbf{x}); \mathbf{z}'_s = \mathcal{F}_s(\mathbf{x}')$ \triangleright Shared feature vector
- 8: $l_{inter} = \lambda_{inter} \cdot L_{inter}(\mathbf{z}_s, \mathbf{z}'_s; \mathbf{y}, \mathbf{y}')$ \triangleright **Inter-domain loss (Equ. 2)**
- 9: Update \mathcal{F}_s parameters using ∇l_{inter}
- 10: \triangleright Discriminator training
- 11: **for** $iter = 1$ to k_{adv} **do**
- 12: $l_{\mathcal{D}} = 0$
- 13: **for all** $d \in \Delta$ **do** \triangleright For all N domains
- 14: Sample a mini-batch $\mathbf{x} \sim \mathbb{U}_d$
- 15: $\mathbf{z}_s = \mathcal{F}_s(\mathbf{x})$
- 16: $l_{\mathcal{D}} += L_{\mathcal{D}}(\mathcal{D}(\mathbf{z}_s); d)$ \triangleright Accumulate \mathcal{D} loss
- 17: Update \mathcal{D} parameters using $\nabla l_{\mathcal{D}}$
- 18: \triangleright Main iteration
- 19: $loss = 0$
- 20: **for all** $d \in \Delta_L$ **do** \triangleright For all labeled domains
- 21: Sample a mini-batch $(\mathbf{x}, \mathbf{y}) \sim \mathbb{X}_d$
- 22: $\mathbf{z}_s = \mathcal{F}_s(\mathbf{x})$
- 23: $\mathbf{z}_d = \mathcal{F}_d(\mathbf{x})$ \triangleright Domain feature vector
- 24: $loss += L_{\mathcal{C}}(\mathcal{C}(\mathbf{z}_s, \mathbf{z}_d); \mathbf{y})$ \triangleright Compute \mathcal{C} loss
- 25: **for all** $d \in \Delta_U$ **do** \triangleright For all N domains
- 26: Sample a mini-batch $\mathbf{x} \sim \mathbb{U}_d$
- 27: \triangleright Discriminate
- 28: $\mathbf{z}_s = \mathcal{F}_s(\mathbf{x})$
- 29: $loss += \lambda_d \cdot L_{\mathcal{F}_s}^{\mathcal{D}}(\mathcal{D}(\mathbf{z}_s); d)$ \triangleright Domain loss of \mathcal{F}_s
- 30: \triangleright Intra-domain alignment
- 31: $\mathbf{x}' = \text{Aug}(\mathbf{x})$ \triangleright Augmented batch
- 32: $\mathbf{o} = \mathcal{C}(\mathcal{F}_s(\mathbf{x}), \mathcal{F}_d(\mathbf{x})); \mathbf{o}' = \mathcal{C}(\mathcal{F}_s(\mathbf{x}'), \mathcal{F}_d(\mathbf{x}'))$; \triangleright Output of the classifier
- 33: $loss += \lambda_{intra} \cdot L_{intra}(\mathbf{o}, \mathbf{o}')$ \triangleright **Intra-domain loss (Equ. 3)**
- 34: Update $\mathcal{F}_s, \mathcal{F}_d, \mathcal{C}$ parameters using $\nabla loss$
- 35: **until** convergence

of length 256. The Office-Home dataset contains four image domains with 65 categories each, with instances encoded to a vector representation of length 2048. The FDUMTL dataset comprises sixteen textual domains, and we use the first four of them in our experiment, each containing two categories, with raw texts utilizing word2vec embedding. Finally, the PACS dataset consists of four image domains with seven categories each, and raw images are preprocessed to tensor with shape (224,224,3).

5.2.2 Models

Our primary focus is evaluating the performance of MDCL on a single model. Therefore, we selected the most renowned and widely recognized models in MDL as baselines. In the majority of the ex-

periments, MAN [4] is utilized as the backbone of MDCL due to its simple structure and wide acceptance in the literature. ASP-MTL [2], as the most classic MDL model, is also used to verify the generation ability of MDCL. Compared to MAN, ASP-MTL has specific classifiers for each domain, while MAN has a shared classifier for all domains. While there are newer models available, such as CAN [36], the existing literature [16] suggests that these models do not outperform MAN. Some other recently proposed models in the MDL are designed for particular tasks [11, 9], which are beyond the scope of our paper.

The balancing parameter λ_d for $l_{\mathcal{D}}$ is set to 0.05 for all the datasets. For Amazon, MNIST-USPS and Office-Home datasets, we use one fully connected layer with a sigmoid activation function as the \mathcal{F}_s and \mathcal{F}_d feature extractor, both with the same output size of 64. For the FDUMTL dataset, we use a CNN as the feature extractor that takes 100d word embeddings as input from the sequence, and the output representation length is 128. A single convolution layer with 200 kernels of sizes 3, 4, and 5 is used. For the PACS dataset, we use a pre-trained Resnet-18 [41] as the feature extractor with an output size of 64. For all the datasets, the classifier and domain discriminator consist of a single fully connected layer. Table 1 provides all the hyper-parameters for optimizations in training details for each dataset.

5.2.3 MDCL Implementation Details

For MDCL, the inter-domain and intra-domain contrastive learning requires a balancing (λ) and a temperature (τ) hyperparameter, which are listed in Table 1. For the FDUMTL dataset, multiple outputs from dropout layers are utilized to obtain batch augmentation. For the rest datasets, we apply batch augmentation by using Gaussian noise with a standard deviation of 0.01. All the experiments were conducted five times, and the average performance and standard deviation were calculated to ensure the reliability of the results.

5.3 RQ1: Performance with insufficient labels

We assess the efficacy of our MDCL method with a limited number of labeled instances. First, two scenarios are included: 1) when the number of labeled instances is moderately limited (around 5%-20%), and 2) when the number of labeled instances is extremely low (1%). Both scenarios are evaluated on the MAN model. Then, ASPMTL is used to verify the generation ability of MDCL over SP models in the limited labeled case (around 5%).

Moderately insufficient labels We evaluate our method on datasets Amazon, MNIST-USPS, and Office-Home with 5% to 20% labeled instances, and on datasets FDUMTL and PACS with 5% to 15% labeled instances. The performance results are illustrated as learning curves in Figure 3. On all the datasets, MDCL brings obvious improvement on solely using MAN. The improvement is usually

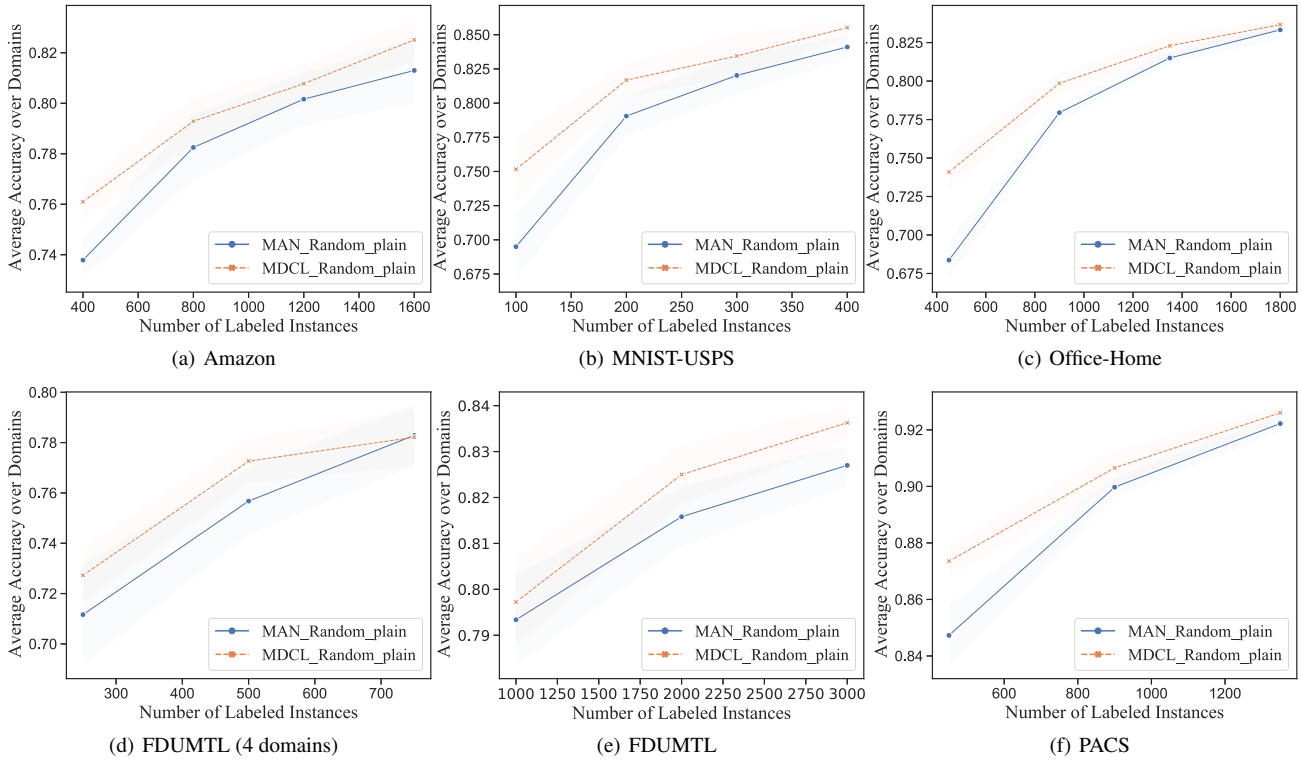


Figure 3: The results of MDCL with different number of labeled instances on different datasets.

more significant when the number of labeled instances is low and decreases with more labeled instances.

Table 2: MDCL on only 1% labeled instances. Average performance in 5 runs with the standard deviation in parentheses.

| Method | Amazon | MNIST-USPS | Office-Home | FDUMTL | PACS |
|--------|--------------------|--------------------|-------------------|--------------------|--------------------|
| MAN | 0.6374 (0.0180) | 0.3886 (0.0831) | 0.347 (0.0248) | 0.5204 (0.0163) | 0.6723 (0.0418) |
| MDCL | 0.6631 (0.0292) | 0.4750 (0.1205) | 0.421 (0.0315) | 0.5592 (0.0559) | 0.6911 (0.0290) |

Extremely insufficient labels We evaluated the performance of MDCL with an extremely small number of labeled instances on Amazon, FDUMTL, and PACS datasets. Only used 1% labeled instances are used to train the model. The results are presented in Table 2. The improvement is also obvious on all the datasets. The improvement is more significant on the Office-Home dataset, which has more categories than the other two datasets.

Table 3: MDCL with ASPMTL on 1% & 5% labeled instances. Average performance in 5 runs with the standard deviation in parentheses.

| Method | Amazon | | Office-Home | |
|--------|--------------------|--------------------|--------------------|--------------------|
| | 1% | 5% | 1% | 5% |
| ASPMTL | 0.5969 (0.0147) | 0.6874 (0.0128) | 0.1665 (0.0127) | 0.5816 (0.016) |
| MDCL | 0.617 (0.0164) | 0.7224 (0.0221) | 0.2348 (0.0242) | 0.6141 (0.0125) |

Compatibility for share-private framework As a plug-and-play method, MDCL should be compatible with other SP models. We compare the performance of our MDCL method with another popular model, ASPMTL [2], on the Amazon and Office-Home datasets with 5% labeled instances to evaluate its ability to be combined with other models. The results are shown in Table 3. MDCL outperforms ASPMTL on both datasets, which validates the generalization ability of MDCL on different SP models.

5.4 RQ2: Effectiveness of components

We explore the effectiveness of the two components of MDCL, i.e., the inter-domain and the intra-domain contrastive loss, by conducting ablation studies on the PACS and MNIST-USPS datasets with 5% labeled instances. The results are shown in Table 4. The inter-domain contrastive proves more effective than the intra-domain contrast on PACS, but the opposite is observed for MNIST-USPS. This discrepancy may be attributed to the fact that, on PACS, the pre-trained feature extractor already provides a reliable domain-invariant feature representation, making semantic information more critical. In contrast, on MNIST-USPS, the intra-domain contrastive loss is more effective, as it assists the feature extractor in learning features in each domain.

Table 4: Ablation study on PACS and MNIST-USPS datasets.

| Method | MNIST-USPS | PACS |
|--------------|------------|--------|
| MAN | 0.7007 | 0.8478 |
| MDCL (Inter) | 0.7024 | 0.8717 |
| MDCL (Intra) | 0.7505 | 0.852 |
| MDCL | 0.7575 | 0.8748 |

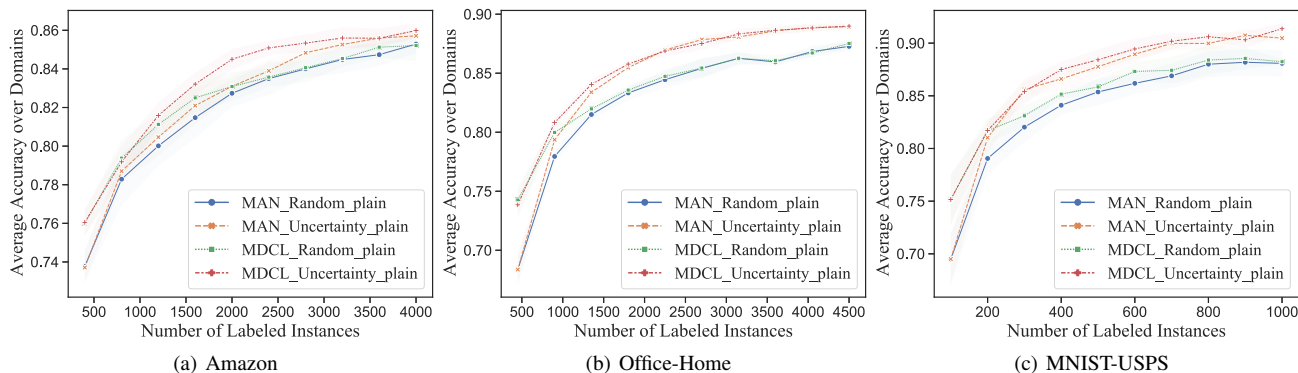


Figure 4: The results of MDCL combined with the Uncertainty strategy in a MDAL setting.

5.5 RQ3: Ability to integrate with MDAL

In section 5.3, the improvement decreases with more labeled instances. In real applications, a method that only performs well with fewer labeled instances but poorly with more labeled instances is not desirable. Thus, the proposed method should be evaluated with more labeled instances. We consider a setting where there is a larger budget for collecting labeled instances, which allows for more labeled instances to be included in the learning process. Specifically, we can use the MDAL method to iteratively collect informative labeled instances, with MDCL being used in each iteration.

We conduct experiments with 5% to 50% labeled instances on the Amazon, MNIST-USPS, and Office-Home datasets using the simplest but effective active learning (AL) query strategy, Best-vs-Second-Best [42] (BvSB). The learning curves on all the datasets are shown in Figure 4. Moreover, the area-under-learning-curves (AULC) results of all three datasets are presented in Table 5.

Table 5: AULC of MDCL. Average performance in 5 runs with the standard deviation in parentheses.

| Method | Amazon | MNIST-USPS | Office-Home |
|---------|--------|------------|-------------|
| MAN | 82.09 | 84.29 | 83.28 |
| +Random | (0.46) | (1.00) | (0.23) |
| MDCL | 82.67 | 85.46 | 83.96 |
| +Random | (0.37) | (0.68) | (0.15) |
| MAN | 82.63 | 86.74 | 85.25 |
| +BvSB | (0.31) | (0.50) | (0.09) |
| MDCL | 83.46 | 87.43 | 85.81 |
| +BvSB | (0.18) | (0.37) | (0.11) |

From the results, we can see that the proposed MDCL method performs well on fewer labeled instances and performs competitively to MAN with more labeled instances. Integrated with MDAL, MDCL further improves performance. With BvSB, MDCL starts from a better initialization and could provide more reliable selections for active learning. In Figure 4, MDCL dominates MAN in the whole MDAL learning process. In Table 5, MDCL obtains higher AULC scores on all the datasets in both passive and active learning settings.

6 Conclusion

In conclusion, we introduce a novel multi-domain contrastive learning (MDCL) approach for multi-domain learning. The primary objective of MDCL is to build neural network models on insufficient

labeled instances from multiple domains. MDCL comprises two components: a supervised contrastive loss for inter-domain semantic alignment and an unsupervised contrastive loss for intra-domain representation learning. MDCL is readily compatible with many SP models, requiring no additional model parameters and allowing for end-to-end training. Experimental results across five textual and image multi-domain datasets demonstrate that MDCL brings noticeable improvement over various SP models. Additionally, given a labeling budget, MDCL can be further employed in multi-domain active learning to enhance the performance of the entire learning process.

Acknowledgements

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFA1004102 and in part by the National Natural Science Foundation of China under Grant 62250710682.

References

- [1] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M. Hospedales. Deeper, broader and artier domain generalization. In *Proceedings of the 2017 IEEE International Conference on Computer Vision*, pages 5543–5551, Venice, Italy, October 2017.
- [2] Pengfei Liu, Xipeng Qiu, and Xuanjing Huang. Adversarial multi-task learning for text classification. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 1–10, Vancouver, Canada, July 2017.
- [3] Mark Dredze and Koby Crammer. Online methods for multi-domain learning and adaptation. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 689–697, Honolulu, HI, October 2008.
- [4] Xilun Chen and Claire Cardie. Multinomial adversarial networks for multi-domain text classification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1226–1240, New Orleans, LA, June 2018.
- [5] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.
- [6] Xueying Zhan, Huan Liu, Qing Li, and Antoni B. Chan. A comparative survey: Benchmarking for pool-based active learning. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence, Virtual Event / Montreal, Canada, August 2021*.
- [7] Ke Tang, Shengcai Liu, Peng Yang, and Xin Yao. Few-shots parallel algorithm portfolio construction via co-evolution. *IEEE Transactions on Evolutionary Computation*, 25(3):595–607, 2021.
- [8] Shengcai Liu, Ke Tang, and Xin Yao. Generative adversarial construction of parallel portfolios. *IEEE Transactions on Cybernetics*, 52(2):784–795, 2022.

- [9] Chao Huang, Hu Han, Qingsong Yao, Shankuan Zhu, and Shao-hua Kevin Zhou. 3d u²-net: A 3d universal u-net for multi-domain medical image segmentation. In *Proceedings of Medical Image Computing and Computer Assisted Intervention - MICCAI, Lecture Notes in Computer Science*, pages 291–299, Shenzhen, China, October 2019.
- [10] Hao Zheng, Jun Han, Hongxiao Wang, Lin Yang, Zhuo Zhao, Chaoli Wang, and Danny Z. Chen. Hierarchical self-supervised learning for medical image segmentation based on multi-domain data aggregation. In *Proceedings of Medical Image Computing and Computer Assisted Intervention - MICCAI, Lecture Notes in Computer Science*, pages 622–632, Strasbourg, France, September 2021.
- [11] Idriss Mghabbar and Pirashanth Ratnamogan. Building a multi-domain neural machine translation model using knowledge distillation. In *Proceedings of 24th European Conference on Artificial Intelligence*, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 2116–2123, Santiago de Compostela, Spain, August 2020.
- [12] Korawat Tanwisuth, Xinjie Fan, Huangjie Zheng, Shujian Zhang, Hao Zhang, Bo Chen, and Mingyuan Zhou. A prototype-oriented framework for unsupervised domain adaptation. In *Proceedings of the Annual Conference on Neural Information Processing Systems 2021*, pages 17194–17208, Virtual-only, December 2021.
- [13] Ankit Singh. Clda: Contrastive learning for semi-supervised domain adaptation. In *Proceedings of the Annual Conference on Neural Information Processing Systems 2021*, pages 5089–5101, Virtual-only, December 2021.
- [14] Kai Li, Chang Liu, Handong Zhao, Yulun Zhang, and Yun Fu. Ecacl: A holistic framework for semi-supervised domain adaptation. In *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops*, pages 8578–8587, Montreal, Canada, October 2021.
- [15] Burr Settles. Active learning literature survey. 2010.
- [16] Rui He, Shengcai Liu, Shan He, and Ke Tang. Multi-domain active learning: Literature review and comparative study. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2022.
- [17] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *Proceedings of the Annual Conference on Neural Information Processing Systems 2016*, pages 343–351, Barcelona, Spain, December 2016.
- [18] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, pages 1597–1607, Virtual Event, July 2020.
- [19] Yu Zhang and Qiang Yang. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [20] Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, and Tao Qin. Generalizing to unseen domains: A survey on domain generalization. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*, pages 4627–4635, Virtual Event / Montreal, Canada, August 2021.
- [21] Junkun Yuan, Xu Ma, Defang Chen, Kun Kuang, Fei Wu, and Lanfen Lin. Label-efficient domain generalization via collaborative exploration and generalization. In *Proceedings of The 30th ACM International Conference on Multimedia*, pages 2361–2370, Lisboa, Portugal, October 2022.
- [22] Ning Lu, Shengcai Liu, Rui He, and Ke Tang. Large language models can be guided to evade ai-generated text detection. *arXiv preprint arXiv:2305.10847*, 2023.
- [23] Sinno Jialin Pan, Ivor W. Tsang, James T. Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, 2011.
- [24] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor S. Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17:59:1–59:35, 2016.
- [25] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2962–2971, Honolulu, HI, July 2017.
- [26] Zeyu Feng, Chang Xu, and Dacheng Tao. Self-supervised representation learning from multi-domain data. In *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*, pages 3244–3254, Seoul, Korea (South), October 2019.
- [27] Lianghao Li, Xiaoming Jin, Sinno Jialin Pan, and Jian-Tao Sun. Multi-domain active learning for text classification. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1086–1094, Beijing, China, August 2012.
- [28] Rui He, Zeyu Dai, Shan He, and Ke Tang. Perturbation-based two-stage multi-domain active learning. *arXiv preprint arXiv:2306.10700*, 2023.
- [29] Vincent Vercauteren, Lorenzo Perini, Wannes Meert, and Jesse Davis. Multi-domain active learning for semi-supervised anomaly detection. *ECML/PKDD 2022 published proceedings*, 2022.
- [30] Ard Snijders, Douwe Kiela, and Katerina Margatina. Investigating multi-source active learning for natural language inference. *arXiv preprint arXiv:2302.06976*, 2023.
- [31] Jiahao Wu, Wenqi Fan, Jingfan Chen, Shengcai Liu, Qing Li, and Ke Tang. Disentangled contrastive learning for social recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 4570–4574, Atlanta, GA, USA, October 2022.
- [32] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschiot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *Proceedings of the Annual Conference on Neural Information Processing Systems 2020*, Virtual-only, December 2020.
- [33] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.
- [34] Yu Sun, Eric Tzeng, Trevor Darrell, and Alexei A Efros. Unsupervised domain adaptation through self-supervision. *arXiv preprint arXiv:1909.11825*, 2019.
- [35] Zizheng Yan, Yushuang Wu, Guanbin Li, Yipeng Qin, Xiaoguang Han, and Shuguang Cui. Multi-level consistency learning for semi-supervised domain adaptation. *arXiv preprint arXiv:2205.04066*, 2022.
- [36] Yuan Wu, Diana Inkpen, and Ahmed El-Roby. Conditional adversarial networks for multi-domain text classification. *arXiv preprint arXiv:2102.10176*, 2021.
- [37] Saeid Motiian, Marco Piccirilli, Donald A. Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *Proceedings of the 2017 IEEE International Conference on Computer Vision*, pages 5716–5726, Venice, Italy, October 2017.
- [38] Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *Proceedings of the 37th International Conference on Machine Learning*, pages 9929–9939, Virtual Event, July 2020.
- [39] Minmin Chen, Zhixiang Eddie Xu, Kilian Q. Weinberger, and Fei Sha. Marginalized denoising autoencoders for domain adaptation. In *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, UK, June 2012.
- [40] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pages 5385–5394, Honolulu, HI, July 2017.
- [41] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, Las Vegas, NV, June 2016.
- [42] Ajay J. Joshi, Fatih Porikli, and Nikolaos Papanikolopoulos. Multi-class active learning for image classification. In *Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2372–2379, Miami, FL, June 2009.