LEAD: Learning Decomposition for Source-free Universal Domain Adaptation — Supplementary Material

Sanqing Qu[1] , Tianpei Zou[1], Lianghua He[1], Florian Röhrbein[2],
Alois Knoll[3], Guang Chen[1*], Changjun Jiang[1]
[1]Tongji University, [2]Chemnitz University of Technology, [3] Technical University of Munich

1. More Related Work

Source-free Domain Adaptation (SFDA). SFDA aims to adapt a pre-trained source model to a new domain without needing access to the original source data. Existing methods in this field generally fall into two categories: source distribution estimation methods [4, 9, 17, 24] and self-training methods [5, 16, 19, 23]. The former often utilizes generative networks to create synthetic-labeled data or to transfer target data into the style of the source domain, thereby bridging the domain gap. The latter, drawing inspiration from semi-supervised learning, typically adopts various pseudo-labeling techniques for model adaptation. However, these methods mainly focus on the vanilla closed-set scenarios where the source and target domains share identical label spaces, significantly limiting their applicability. In contrast, this paper focuses on Source-free Universal Domain Adaptation (SF-UniDA), targeting the more challenging setting that encompasses both covariate and label shifts.

Out-of-distribution (OOD) Detection. The primary objective of OOD detection is to identify test (target) samples that are distinct from the training (source) distribution. Typically, in the literature of OOD, the distribution is referred to as 'label distribution', with OOD samples being those unrecognizable or exclusive to the source label space. As such, OOD detection can be considered a component of UniDA/SF-UniDA, both of which aim to reject these target-private unknown data from the target domain. In addressing OOD, prevailing techniques [1, 2, 20] often use metrics such as the maximum of softmax outputs or confidence scores. Recent studies [3, 8, 21] have begun incorporating a collection of OOD samples for outlier exposure during model pre-training, fostering a clearer distinction between in-distribution (ID) and OOD samples. However, these methods typically depend on manually set thresholds for identifying target-private data, which can be both tedious and sub-optimal. In contrast, we start from the perspective of feature decomposition and utilize the feature projection onto source-unknown space as the indicator. Then, we consider the distance to both target prototypes and source anchors to establish adaptive instance-level decision boundaries. Our solution is effective under conditions of both covariate and label shift, offering a more flexible and robust approach for UniDA/SF-UniDA.

*Corresponding author: guangchen@tongji.edu.cn

2. More Details about Methodology

2.1. Target Prototype Construction

In the preceding discussion, we outlined the construction of instance-level pseudo-labeling decision boundaries. These boundaries are defined based on the distance of samples to source anchors, denoted as $\{\mathbf{c}_c^s \in \mathbb{R}^D | c = 1, \dots, C\}$, and target prototypes, denoted as $\{\mathbf{c}_c^t \in \mathbb{R}^D | c = 1, \dots, C\}$. Source anchors are derived directly from the classifier weight W_{cls}, where $W_{cls} \in \mathbb{R}^{C \times D}$. Regarding the target prototypes, we employ a top-K sampling strategy for their construction. Specifically, for each category c, we identify the top-K instances in the target domain with the highest $\delta_c(f_\theta^t(\mathbf{x}_i^t))$ scores. These instances are then averaged to form the target prototype $\mathbf{c}_c^t$ for category c. Formally,

$$\begin{aligned} \mathcal{M}_c &= \underset{\mathbf{x}_i^t \in \mathcal{X}_t;\ |\mathcal{M}_c|=K}{\arg\max} \ \delta_c(f_\theta^t(\mathbf{x}_i^t)), \\ \mathbf{c}_c^t &= \frac{1}{K} \sum_{i \in \mathcal{M}_c} g_\theta^t(\mathbf{x}_i^t). \end{aligned} \tag{1}$$

where $\delta_c(f_\theta^t(\mathbf{x}_i^t))$ denotes the c-th soft-max probability for the instance $\mathbf{x}_i^t$. As previously mentioned, f_θ^t and g_θ^t represent the entire target model and its feature extractor, respectively. A critical challenge is determining the appropriate value for K. Following previous work [7, 12], we adopt a straightforward yet effective method, setting $K = N_t/\hat{C}_t$. Here, N_t signifies the amount of target data, while $\hat{C}_t$ estimates the number of target categories. For the estimation of $\hat{C}_t$, we utilize the Silhouette metric[13]. This process involves initially enumerating possible values for C_t, followed by partitioning the target domain into respective clusters using an algorithm such as K-means. Subsequently, the Silhouette metric aids in selecting the most suitable value

for $\hat{C}_t$. Formally, the Silhouette score for $\mathbf{x}_i^t$ is defined as:

$$\begin{aligned} a(\mathbf{x}_i^t) &= \frac{1}{|\mathcal{C}_{\mathcal{I}}|-1} \sum_{\mathbf{x}_i^t \in \mathcal{C}_{\mathcal{I}}, i \neq j} d(g_\theta^t(\mathbf{x}_i^t), g_\theta^t(\mathbf{x}_j^t)), \\ b(\mathbf{x}_i^t) &= \min_{\mathcal{J} \neq \mathcal{I}} \frac{1}{|\mathcal{C}_{\mathcal{J}}|} \sum_{\mathbf{x}_j^t \in \mathcal{C}_{\mathcal{J}}} d(g_\theta^t(\mathbf{x}_i^t), g_\theta^t(\mathbf{x}_j^t)), \\ s(\mathbf{x}_i^t) &= \frac{b(\mathbf{x}_i^t) - a(\mathbf{x}_i^t)}{\max\{a(\mathbf{x}_i^t), b(\mathbf{x}_i^t)\}}. \end{aligned} \tag{2}$$

where $a(\mathbf{x}_i^t)$ and $b(\mathbf{x}_i^t)$ are functions used to calculate the distance of a data sample $\mathbf{x}_i^t$ to its own cluster $\mathcal{C}_{\mathcal{I}}$ and to other clusters $\mathcal{C}_{\mathcal{J}, \mathcal{J} \neq \mathcal{I}}$, respectively. The term $|\mathcal{C}_{\mathcal{I}}|$ denotes the size of the cluster $\mathcal{C}_{\mathcal{I}}$. $d(,)$ is used to measure the distance between two data samples. The appropriateness of the clustering configuration is assessed by the Silhouette values of the data samples. A majority of high Silhouette values indicate a suitable clustering configuration. Conversely, predominantly low Silhouette values suggest that the clustering configuration may have an inappropriate number of clusters, either too many or too few.

2.2. Feature Consensus Regularization

Existing literature [12, 22, 23] in SFDA and SF-UniDA have observed that the integration of consensus regularization with nearest neighbors in the feature space significantly contributes to stable performance. Building upon these findings, we incorporate the feature consensus learning objective $\mathcal{L}_{con}$ into our proposed LEAD framework. Specifically,

$$\begin{aligned} l_{i,c}^t &= \frac{1}{|L_i^t|} \sum_{\mathbf{x}_j^t \in L_i^t} \delta_c(f_\theta^t(\mathbf{x}_j^t)), \\ \mathcal{L}_{con} &= -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} l_{i,c}^t \log \delta_c(f_\theta^t(\mathbf{x}_i^t)). \end{aligned} \tag{3}$$

where L_i^t represents the set of nearest neighbors for a given target domain data $\mathbf{x}_i^t$ within the feature space. We utilize the cosine similarity function to identify these nearest neighbors. Consistent with [12], we empirically determine the size of the nearest neighbor $|L_i^t| = 4$.

2.3. Integration into Existing Methods

As elaborated in the main text of our study, the LEAD framework introduces an innovative approach to SF-UniDA through feature decomposition. This unique concept establishes LEAD as a complementary addition to existing methodologies. To validate this merit, we have integrated LEAD with representative methods, specifically UMAD [6] and GLC [12]. Given the unsupervised nature of the SF-UniDA task, this integration is achieved by combining the optimization objectives of LEAD with those of the baseline methods. As a result, the integrated optimization objective is presented as follows:

$$\mathcal{L}_{overall} = \gamma \cdot \mathcal{L}_{LEAD} + (1-\gamma) \cdot \mathcal{L}_{baseline} \tag{4}$$

where γ is a trade-off hyper-parameter, generally set to 0.7.

3. More Details about Experiments

3.1. Datasets

In this study, we evaluate the performance and adaptability of our proposed LEAD framework using four standard datasets, each offering distinct challenges. First, Office-31 [14], a prevalent domain adaptation benchmark, encompasses 31 object classes (comprising 4,652 images) within an office setting, spread across three distinct domains: DSLR (D), Amazon (A), and Webcam (W). Next, Office-Home [18] presents a more extensive challenge with 65 object classes totaling 15,500 images, categorized into four domains: Artistic images (Ar), Clip-Art images (Cl), Product images (Pr), and Real-world images (Rw). The VisDA [10] dataset, a more challenging benchmark, involves 12 object classes; its source domain includes 152,397 synthetic images created from 3D models, and the target domain comprises 55,388 real-world images from Microsoft COCO. Lastly, DomainNet [11], the most extensive DA benchmark, contains approximately 0.6 million images across 345 object classes. In line with existing research, our experiments on DomainNet focus on three subsets: Painting (P), Real (R), and Sketch (S). Figure 1 illustrates some representative examples of these benchmarks.

3.2. Source Model Pre-training

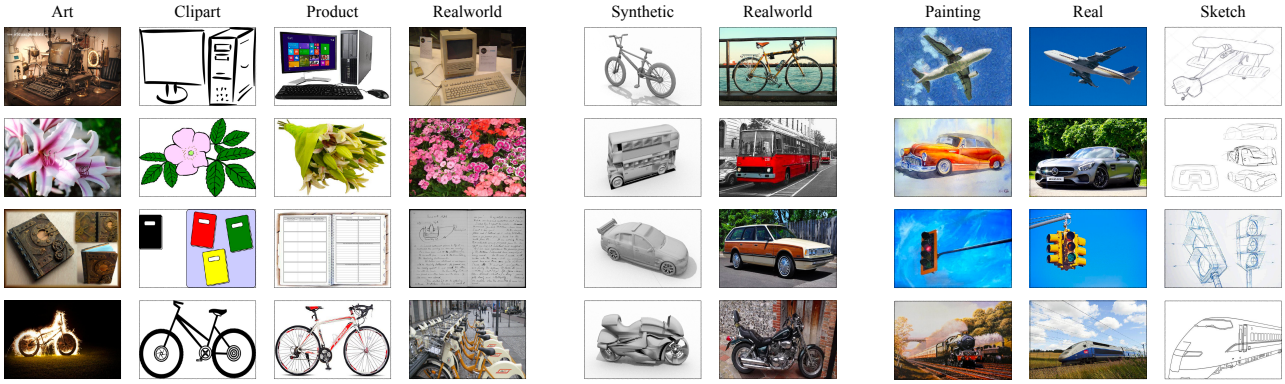
To prepare the source model, we adopt the training recipe [5, 12, 23] widely used in SFDA and SF-UniDA tasks. For a given labeled source domain $\mathcal{D}^s = \{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^{N_s}$ where $\mathbf{x}_i^s \in \mathcal{X}^s \subset \mathbb{R}^X, \mathbf{y}_i^s \in \mathcal{Y}^s \subset \mathbb{R}^C$. We train the source model f_θ^s using a smooth cross-entropy loss function, which is detailed as follows:

$$\mathcal{L}_{src} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} q_{i,c}^s \log \delta_c(f_\theta^s(\mathbf{x}_i^s)) \tag{5}$$

where $\delta_c(f_\theta^s(\mathbf{x}_i^s))$ denotes the c-th soft-max probability for the instance $\mathbf{x}_i^s$. $q_{i,c}^s$ corresponds to c-th smooth one-hot encoded label for $\mathbf{y}_i^s$, i.e., $q_{i,c}^s = (1-\beta) \cdot \mathbb{1}(\mathbf{y}_i^s) + \beta/C$. Here, $\mathbb{1}$ denotes the one-hot encoding operator, and β is the smoothing parameter which is set to 0.1 for all benchmarks.

3.3. Target Inference Details

In the inference phase, we leverage the strategy employed in existing methods [7, 12] to distinguish between common and private data. Specifically, we apply the normalized



(a) Office-Home (b) VisDA (c) DomainNet

Figure 1. Representative examples from the benchmark datasets used in our study, illustrating various types of domain shift. The selected samples highlight the distinct characteristics and environments of the domains within each dataset.

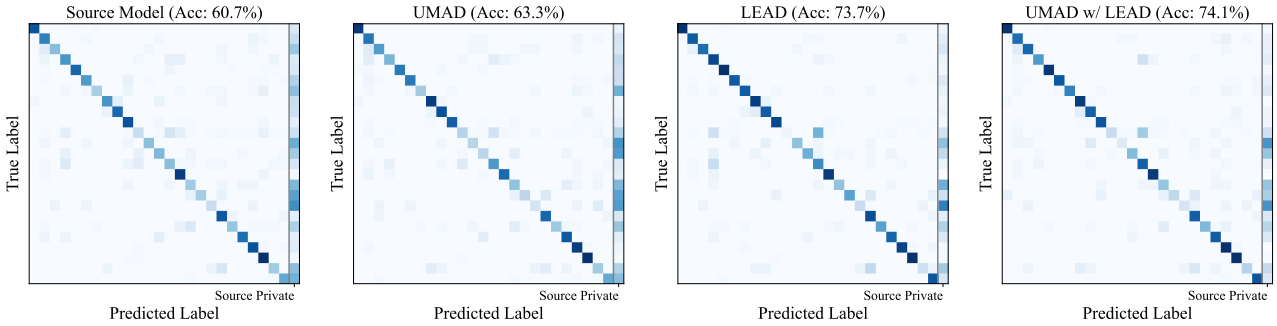


Figure 2. The confusion matrix for Source Model, UMAD, LEAD, and UMAD w/ LEAD on Pr→Ar (Office-Home) PDA task.

Shannon Entropy [15] as the metric for this differentiation. The implementation details of this strategy are as follows:

$$I(\mathbf{x}_i^t) = -\frac{1}{\log C}\sum_{c=1}^{C}\delta_c(f_\theta^t(\mathbf{x}_i^t))\log\delta_c(f_\theta^t(\mathbf{x}_i^t)) \quad (6)$$

$$p(\mathbf{x}_i^t) = \begin{cases}\text{unknown}, & \text{if } I(\mathbf{x}_i^t) \geq \omega \\ \text{argmax}(f_\theta^t(\mathbf{x}_i^t)), & \text{if } I(\mathbf{x}_i^t) < \omega\end{cases} \quad (7)$$

where the inference result $p(\mathbf{x}_i^t)$ hinges on the normalized entropy value $I(\mathbf{x}_i^t)$. A higher value of $I(\mathbf{x}_i^t)$ indicates a greater likelihood of the model f_θ^t classifying the data sample $\mathbf{x}_i^t$ as unknown. In particular, when $I(\mathbf{x}_i^t)$ exceeds a pre-defined threshold ω, the data sample $\mathbf{x}_i^t$ is categorized as target-private. Conversely, if $I(\mathbf{x}_i^t)$ falls below this threshold, the sample is recognized as common. In our experiments across all datasets, we set $\omega = 0.55$, aligning with the thresholds used in existing methods [7, 12].

3.4. More Experimental Analysis

Confusion Matrix Visualization. Figure 2 provides a visual representation of the confusion matrices for four different models: the model trained solely on source data, UMAD, LEAD, and a combined UMAD with LEAD. This comparison is conducted in the Pr→Ar (Office-Home) PDA task. The source-only model exhibits a tendency towards inaccurate predictions due to distributional covariate shifts, notably misclassifying target common data as belonging to source private categories. The implementation of model adaptation, as observed in the matrices, evidently mitigates this confusion. However, due to that UMAD is designed primarily for OPDA and OSDA scenarios, its effectiveness is somewhat constrained. In contrast, our LEAD framework demonstrates significant versatility and substantial performance enhancement. Notably, it also proves to be a complementary approach when integrated with UMAD.

t-SNE Feature Visualization. In Figure 3, we provide a t-SNE visualization of the features extracted by the source model, UMAD, LEAD, and UMAD integrated with LEAD. This analysis is conducted within the Cl→Re (Office-Home) OPDA task. An obvious finding from this visualization is the initially ambiguous boundaries between target private and common data in the feature space, where these data appear intermingled. As expected, performing model adaptation significantly contributes to the separation between common and private data. Taking a closer look at the visualization, it becomes evident that both LEAD and UMAD w/ LEAD are particularly effective in achieving a more distinct separation between common and private data.

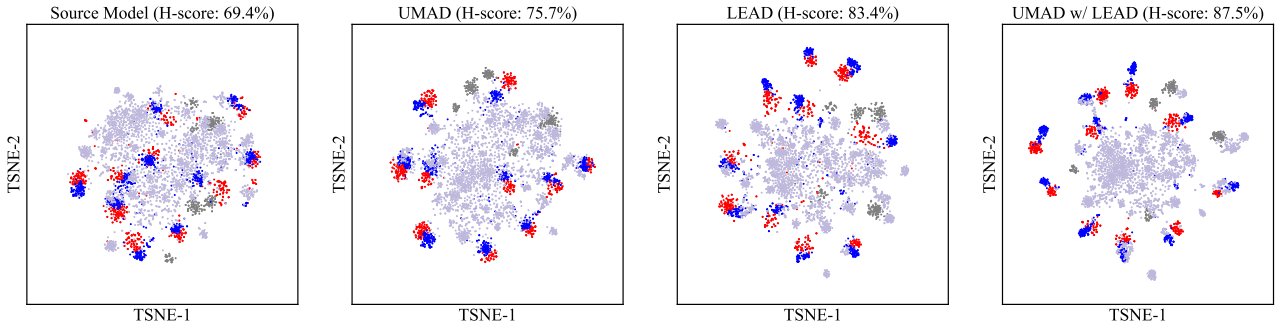
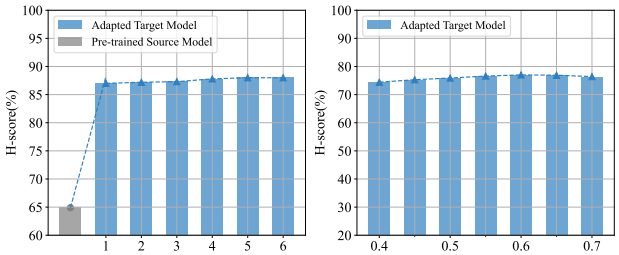


Figure 3. The t-SNE feature visualization for Source Model, UMAD, LEAD, and UMAD w/ LEAD on Cl→Re (Office-Home) OPDA task. Points in red denote unavailable source-common data, points in gray represent unavailable source-private data, points in blue illustrate target-common data, points in lavender illustrate target-private data, individually. It is easy to conclude that both LEAD and UMAD w/ LEAD are effective in achieving a clear separation of common and private data.



(a) Sensitivity to $|L_i^t|$ on Office-31 (b) H-score with respect to ω on VisDA

Figure 4. **More Robustness Analysis**. (a) shows the sensitivity to $|L_i^t|$ on Office-31 in the OPDA scenario. (b) illustrates the H-score with respect to ω on VisDA in the OPDA scenario.

More Robustness Analysis. In addition to the robustness analysis presented in the main paper, we extend our investigation to the parameter sensitivity of the nearest neighbor set size $|L_i^t|$ in the context of the feature consensus learning objective $\mathcal{L}_{con}$. This analysis is conducted on the Office-31 dataset within the OPDA scenario, where the range of λ is set to $[1, 2, 3, 4, 5, 6]$. As depicted in Figure 4 (a), the results indicate that our LEAD framework demonstrates stability around the chosen parameter value of $|L_i^t| = 4$. Additionally, in Figure 4 (b), we further analyze the H-score in relation to the threshold ω on the VisDA dataset, also in the OPDA scenario. It demonstrates a relative stability of the H-score around $\omega = 0.55$. Through an oracle validation, we could even achieve better performance when $\omega = 0.60$.

References

[1] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *ICLR*, 2017. 1

[2] Yen-Chang Hsu, Yilin Shen, Hongxia Jin, and Zsolt Kira. Generalized odin: Detecting out-of-distribution image without learning from out-of-distribution data. In *CVPR*, 2020. 1

[3] Julian Katz-Samuels, Julia B Nakhleh, Robert Nowak, and Yixuan Li. Training ood detectors in their natural habitats. In *ICML*, 2022. 1

[4] Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. In *CVPR*, 2020. 1

[5] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *ICML*. PMLR, 2020. 1, 2

[6] Jian Liang, Dapeng Hu, Jiashi Feng, and Ran He. Umad: Universal model adaptation under domain and category shift. *arXiv preprint arXiv:2112.08553*, 2021. 2

[7] Xinghong Liu, Yi Zhou, Tao Zhou, Chun-Mei Feng, and Ling Shao. Coca: Classifier-oriented calibration for source-free universal domain adaptation via textual prototype. *arXiv preprint arXiv:2308.10450*, 2023. 1, 2, 3

[8] Yifei Ming, Ying Fan, and Yixuan Li. Poem: Out-of-distribution detection with posterior sampling. In *ICML*, 2022. 1

[9] Gaurav Kumar Nayak, Konda Reddy Mopuri, Saksham Jain, and Anirban Chakraborty. Mining data impressions from deep models as substitute for the unavailable training data. *IEEE TPAMI*, 2021. 1

[10] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge. *arXiv preprint arXiv:1710.06924*, 2017. 2

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

[12] Sanqing Qu, Tianpei Zou, Florian Röhrbein, Cewu Lu, Guang Chen, Dacheng Tao, and Changjun Jiang. Upcycling models under domain and category shift. In *CVPR*, 2023. 1, 2, 3

[13] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987. 1

[14] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *ECCV*, 2010. 2

[15] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948. 3

[16] Tao Sun, Cheng Lu, and Haibin Ling. Prior knowledge guided unsupervised domain adaptation. In *ECCV*, 2022. 1

[17] Jiayi Tian, Jing Zhang, Wen Li, and Dong Xu. Vdm-da: Virtual domain modeling for source data-free domain adaptation. *IEEE TCSVT*, 2021. 1

[18] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *CVPR*, 2017. 2

[19] Yuxi Wang, Jian Liang, and Zhaoxiang Zhang. Source data-free cross-domain semantic segmentation: Align, teach and propagate. *arXiv preprint arXiv:2106.11653*, 2021. 1

[20] Hongxin Wei, Renchunzi Xie, Hao Cheng, Lei Feng, Bo An, and Yixuan Li. Mitigating neural network overconfidence with logit normalization. In *ICML*, 2022. 1

[21] Jingkang Yang, Haoqi Wang, Litong Feng, Xiaopeng Yan, Huabin Zheng, Wayne Zhang, and Ziwei Liu. Semantically coherent out-of-distribution detection. In *ICCV*, 2021. 1

[22] Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Generalized source-free domain adaptation. In *ICCV*, 2021. 2

[23] Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Exploiting the intrinsic neighborhood structure for source-free domain adaptation. In *NeurIPS*, 2021. 1, 2

[24] Ziyi Zhang, Weikai Chen, Hui Cheng, Zhen Li, Siyuan Li, Liang Lin, and Guanbin Li. Divide and contrast: Source-free domain adaptation via adaptive contrastive learning. In *NeurIPS*, 2022. 1