GCA: Geometry-aware Conditional Alignment for Partial Domain Adaptation with Coding Rate Reduction

Xiaohui Chen, Chuan-Xian Ren*

School of Mathematics, Sun Yat-Sen University, China
chenxh278@mail2.sysu.edu.cn, rchuanx@mail.sysu.edu.cn

Abstract

Partial Domain Adaptation (PDA) aims to transfer knowledge from a labeled source domain to an unlabeled target domain, where the target label space is a subset of the source label space. In PDA scenario, existing methods typically achieve transferability through distribution alignment in a statistical framework, and discriminability through geometric modeling. These two aspects are often treated as separate frameworks, which severs the intrinsic connection between them. To bridge this gap, we propose a unified framework termed Geometry-aware Conditional Alignment (GCA), which is derived from theoretical insights of Maximum Coding Rate Reduction. GCA collaboratively achieves conditional alignment and orthogonal discriminability in a unified framework, making the learned features more interpretable in both statistical and geometric aspects. As a result, GCA effectively enhances both the transferability and discriminability of features. Extensive experiments on four benchmark datasets validate the effectiveness of GCA.

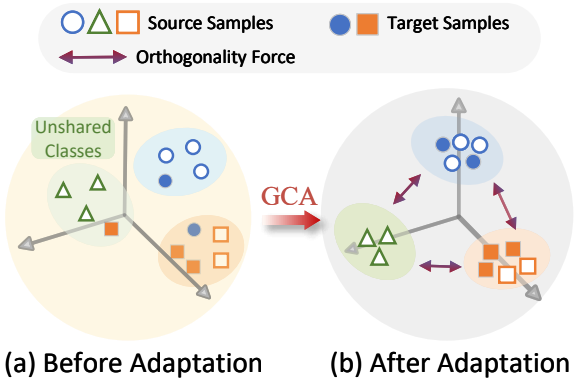

Figure 1: Illustration of feature subspaces before and after adaptation. (a) Before adaptation, class-specific feature subspaces are ambiguous. (b) After adaptation with GCA, features become more compact within classes and more discriminative between classes.

Introduction

Deep neural networks have achieved notable success in computer vision (Krizhevsky, Sutskever, and Hinton 2012; He et al. 2016; Simonyan and Zisserman 2014), mainly due to large-scale labeled datasets and strong representation capabilities. However, traditional deep learning methods typically assume that the training and testing data come from the same distribution. This assumption is often violated in real-world applications (Pan and Yang 2009). Such distribution discrepancies cause domain shift, significantly degrading the performance of source-trained models on new domains.

To address this practical challenge, domain adaptation (DA) has been proposed (Courty et al. 2016; Long et al. 2018a; Luo et al. 2020). PDA is a scenario of DA where the target label space is a subset of the source. Compared to traditional closed-set domain adaptation, PDA presents a significantly greater challenge (Li et al. 2020b). Source-private classes may induce negative transfer if the entire source domain is naively aligned with the target domain.

To mitigate this negative transfer, the paradigm of learning domain-invariant representations has garnered substantial research attention. Existing PDA methods primarily focus on two key properties of the representation model: transferability and discriminability. Intuitively, transferability implies that representations across domains should be indistinguishable, while discriminability requires that representations across different classes should be separable.

Within the statistical framework, where domains are defined as probability distributions, transferability is characterized by the discrepancy between distributions. Statistical learning theory (Kirchmeyer et al. 2021; Zhao et al. 2019) shows that minimizing this domain discrepancy allows the source risk estimate to approximate the target risk. Consequently, many methods learn invariant representations by minimizing statistical measures of domain divergence, such as the Maximum Mean Discrepancy (MMD) (Borgwardt et al. 2006). Alternatively, other methods employ reweighting schemes to reduce distribution discrepancy. These methods diminish the influence of source-private class samples by assigning lower weights to them during adaptation.

In contrast, the geometric framework views domains as manifolds or subspaces. Methods within this framework typically learn invariant representations based on geometric criteria, such as manifold metrics, matrix rank, matrix

*Corresponding Author.

norms (Cui et al. 2020), or subspace bases (Fernando et al. 2013). Compared to the statistical framework, geometric approaches often offer more intuitive geometric interpretations and clearer algebraic significance.

Statistical alignment and geometric modeling have made substantial contributions to the transferability of learned representations (Luo and Ren 2021; Stojanov et al. 2021), yet their discriminability is often neglected. Chen et al. demonstrate that during the pursuit of transferability, the local structure and label information of downstream tasks may be lost (Chen et al. 2019). Solely relying on global distribution alignment may cause feature collapse that undermines discriminability. Moreover, prior studies have shown that models overly focused on transferability often suffer from a decline in discriminability. Although recent works (Xu et al. 2019; Cui et al. 2020) introduce intuitive approaches to enhance both discriminability and transferability, these two properties are rarely addressed within a unified framework and lack solid theoretical grounding. Existing studies often characterize geometric relationships as linear dependence and independence, but the stronger property of feature orthogonality and its possible integration with conditional alignment in a unified framework remain unexplored.

To address these challenges and fully leverage the strengths of both statistical and geometric frameworks for learning transferable and discriminative features, we employ Maximum Coding Rate Reduction (MCR^2), which quantifies the compactness of overall features based on coding rate. Specifically, the orthogonality condition that achieves the MCR^2 upper bound provides a theoretical guarantee for learning discriminative features. By exploring the alignment properties implied by its lower bound, we retain the statistical characteristics that are critical for representation learning. Based on these theories, we unify statistical conditional alignment and geometric orthogonal discriminability into a single framework, forming our full learning criterion Geometry-aware Conditional Alignment(GCA). As illustrated in Fig. 1, GCA employs conditional alignment and orthogonality constraints to transform originally ambiguous class spaces into compact intra-class and mutually orthogonal inter-class subspaces, thereby enhancing the transferability and discriminability of the learned features. Overall, our contributions can be summarized as follow.

- We propose a unified statistical and geometric framework for PDA. It learns invariant representations that possess both statistical properties and geometric structure, ensuring the learned features are interpretable and effective.
- We thoroughly explore the implicit alignment property revealed by the lower bound of MCR^2. Based on the theory of MCR^2, we propose a novel algorithm that integrates geometric orthogonal discriminability with statistical conditional alignment, effectively enhancing both feature transferability and discriminability.
- Extensive experiments on four public PDA benchmarks demonstrate that our method attains top performance. It consistently outperforms recent state-of-the-art adaptation methods and substantially exceeds the baselines.

Related Work

Partial Domain Adaptation. PDA transfers knowledge from a labeled source domain to an unlabeled target domain whose label space is a subset, aiming to reduce negative transfer from irrelevant source classes. Existing methods mainly fall into three categories: sample reweighting, alignment strategies, and geometric structure modeling. Sample reweighting approaches, such as PADA (Cao et al. 2018b), SAN (Cao et al. 2018a) and IWAN (Zhang et al. 2018), estimate the probability of source samples belonging to shared classes and adjust their weights to suppress irrelevant classes. Alignment methods, such as DRCN, SLM (Sahoo et al. 2023), ISRA (Xiao, Ding, and Liu 2021) and IDSP (Li and Chen 2023), explicitly match source and target representations to improve transferability; Geometric structure modeling techniques, such as SAFN (Xu et al. 2019), ML-Net (Lu et al. 2024) and LEAD (Qu et al. 2024), enhance feature properties like norm and neighborhood relations for better generalization. Although conditional alignment and geometric modeling have shown effectiveness in PDA, they have limitations and remain largely unintegrated. Therefore, a unified framework combining both is still necessary.

Transferability and Discriminability. Recent invariant representation models demonstrate strong generalization in transfer learning, highlighting two core properties: transferability and discriminability. Many methods (Chen et al. 2019; Xu et al. 2019) study these from a norm perspective. These norm-based strategies reveal the link between feature generalization and feature structure. Some methods, such as CDAN (Long et al. 2018b), ETD (Li et al. 2020a) and DWL (Xiao and Zhang 2021) enhance discriminability by incorporating labels and weights. Additional approaches improve model performance using class-balance constraints (Zou et al. 2019) or consistency regularization (Liu, Wang, and Long 2021; Sohn et al. 2020; Prabhu et al. 2021). Despite these advances, transferability and discriminability are still modeled separately, and a unified theoretical foundation is lacking.

Geometry-aware Conditional Alignment

Preliminary

Problem Setting. In PDA scenarios, we are given a labeled source domain $\mathcal{D}^s = \{(x_i^s, y_i^s)\}_{i=1}^{n^s}$ and an unlabeled target domain $\mathcal{D}^t = \{x_i^t\}_{i=1}^{n^t}$. The target label space $\mathcal{Y}_t$ is a subset of the source label space $\mathcal{Y}_s$, i.e., $\mathcal{Y}_t \subset \mathcal{Y}_s$. We use $g : \mathcal{X} \to \mathcal{Z}$, $h : \mathcal{Z} \to \mathcal{Y}$ to denote feature extractor and classifier, respectively. Denote $X^s \in \mathbb{R}^{d\times n^s}$, $X^t \in \mathbb{R}^{d\times n^t}$ as data matrices, $X = [X^s, X^t] \in \mathbb{R}^{d\times n}$ as the concatenation of X^s and X^t. Let $X_i = [X_i^s, X_i^t] \in \mathbb{R}^{d\times n_i}$ denote the concatenation of X_i^s and X_i^t, where $X_i^s \in \mathbb{R}^{d\times n_i^s}$ is the data matrix of the class i in the source domain. Note that the total sample size satisfies $n = n^s + n^t = n_1 + n_2 + \cdots + n_k$, where each $n_i = n_i^s + n_i^t$.

Previous work on conditional alignment and orthogonal discriminability has not been addressed within a unified framework. To bridge this gap, we aim to integrate both objectives under a mathematically principled approach. Specif-

ically, guided by theoretical principles, we seek to learn representations that are compressed within the same class to achieve conditional alignment, while expanded between different classes to the extent of orthogonality. This orthogonal expansion enhances inter-class discriminability, providing the learned features with more precise geometric and statistical properties. To quantify the degree of feature compression or expansion, we leverage MCR², which effectively measures the compactness of both the entire feature set and its subsets from the perspective of coding rate. This provides a powerful tool for unifying conditional alignment and orthogonal discriminability within a single framework.

MCR^2. Given a set of feature representations $\boldsymbol{Z}$ obtained via a feature encoder g, the average coding rate—i.e., the number of bits required to encode each sample with a prescribed precision $\varepsilon > 0$ is defined as:

$$R(\boldsymbol{Z}, \varepsilon) = \frac{1}{2} \log\det(I + c\boldsymbol{Z}\boldsymbol{Z}^\top), \text{ where c} = \frac{d}{n\varepsilon^2} \quad (1)$$

To learn ideal representations, the Maximizing Coding Rate Reduction (MCR^2) principle optimizes the feature encoder g by maximizing the difference between the coding rate of the entire set $\boldsymbol{Z}$ and the weighted average coding rates of the class-wise subsets $\boldsymbol{Z}_i$:

$$\Delta R(\boldsymbol{Z}, \varepsilon) = R(\boldsymbol{Z}, \varepsilon) - \sum_{i=1}^{k} \frac{n_i}{n} R(\boldsymbol{Z}_i, \varepsilon) \quad (2)$$

In the following, we focus on its geometric properties and the implied alignment characteristics, without discussing its information-theoretic aspects. For a more detailed introduction to this principle, interested readers are referred to the original works (Ma et al. 2007; Yu et al. 2020).

Theoretical Analysis

Theorem 1. *(Yu et al. 2020) For any $\boldsymbol{Z} \in \mathbb{R}^{d\times n}$ and $\varepsilon > 0$. Assuming that the features of each class have a zero mean, we have*

$$\Delta R(\boldsymbol{Z}, \varepsilon) \leq \sum_{i=1}^{k} \frac{1}{2n} \log\left(\frac{\det^n(\boldsymbol{I} + \frac{d^2}{n\varepsilon^2}\boldsymbol{Z}_i\boldsymbol{Z}_i^\top)}{\det^{n_i}(\boldsymbol{I} + \frac{d^2}{n_i\varepsilon^2}\boldsymbol{Z}_i\boldsymbol{Z}_i^\top)} \right), \quad (3)$$

with equality holds if and only if $\boldsymbol{Z}_{i_1}^\top \boldsymbol{Z}_{i_2} = 0$ for all $1 \leq i_1 < i_2 \leq k$.

Theorem 1 provides an explicit upper bound for $\Delta R(\boldsymbol{Z}, \varepsilon)$, which is attained when different classes are orthogonal. By maximizing $\Delta R(\boldsymbol{Z}, \varepsilon)$, the angles between subspaces of different classes are increased, encouraging these subspaces to become mutually orthogonal and thereby enhancing inter-class discriminability.

To enhance the discriminability between subspaces of different classes, we learn the orthogonal discriminability criterion by partitioning the overall feature set $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_k]$ according to classes. Let $\hat{\boldsymbol{Z}} = \boldsymbol{Z} - \mu$ and $\hat{\boldsymbol{Z}}_i = \boldsymbol{Z}_i - \mu_i$, where $\mu_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Z_{ij}$, the orthogonal discriminability criterion based on MCR^2 is formulated as:

$$\arg\max_{g(\cdot)} \mathcal{L}_{\text{Ort}} = \Delta R(\hat{\boldsymbol{Z}}, \varepsilon) = R(\hat{\boldsymbol{Z}}, \varepsilon) - \sum_{i=1}^{k} R(\hat{\boldsymbol{Z}}_i, \varepsilon) \quad (4)$$

According to Theorem 1, the upper bound of $\mathcal{L}_{\text{Ort}}$ is achieved when the set $\{\hat{\boldsymbol{Z}}_i\}$ is mutually orthogonal. Such orthogonality between different clusters encourages them to lie in disjoint subspaces, thereby enhancing inter-class discriminability. This orthogonal structure effectively mitigates the negative transfer caused by source-private class samples and significantly improves the model's generalization performance on the target domain.

Proposition 1. *(Yu et al. 2020) For any $\{\boldsymbol{Z}_i \in \mathbb{R}^{d\times n_i}\}_{i=1}^k$ and any $\varepsilon > 0$, let $\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_k] \in \mathbb{R}^{d\times n}$. We have*

$$\frac{n}{2}\log\det(\boldsymbol{I} + \frac{d}{n\varepsilon^2}\boldsymbol{Z}\boldsymbol{Z}^\top) \geq \sum_{j=1}^{k} \frac{n_j}{2}\log\det(\boldsymbol{I} + \frac{d}{n_j\varepsilon^2}\boldsymbol{Z}_j\boldsymbol{Z}_j^\top)$$

with equality holds if and only if

$$\frac{\boldsymbol{Z}_1\boldsymbol{Z}_1^\top}{n_1} = \frac{\boldsymbol{Z}_2\boldsymbol{Z}_2^\top}{n_2} = \cdots = \frac{\boldsymbol{Z}_k\boldsymbol{Z}_k^\top}{n_k}$$

Proposition 1 states that the lower bound of $\Delta R(\boldsymbol{Z}, \varepsilon)$ is zero, and equality holds if and only if the uncentered second-order moments of each class are equal.

The original MCR^2 formulation partitions features only by class and only focuses on the orthogonality when $\Delta R(\boldsymbol{Z}, \varepsilon)$ attains its upper bound. However, this is insufficient for domain adaptation problems. It offers little explanation of the lower bound in Proposition 1, and due to the conflict of simultaneously optimizing the upper and lower bounds, the alignment properties implied by the lower bound remain unexplored and unused. By changing the data partitioning strategy from class-wise to domain-wise, we are able to better exploit the alignment properties implicit in the MCR^2 lower bound. This leads to Corollary 1, which provides a clearer interpretation of the lower bound, resolves the conflict and enables the coexistence of MCR^2-based conditional alignment and orthogonal discriminability.

Corollary 1. *Assuming Gaussian priors on each class distribution in source and target domains. Given the features of the i-th class $\boldsymbol{Z}_i = [\boldsymbol{Z}_i^s, \boldsymbol{Z}_i^t]$, let $\hat{\boldsymbol{Z}}_i^s$ and $\hat{\boldsymbol{Z}}_i^t$ denote the centered forms of $\boldsymbol{Z}_i^s$ and $\boldsymbol{Z}_i^t$, respectively. Define $\hat{\boldsymbol{Z}}_i = [\hat{\boldsymbol{Z}}_i^s, \hat{\boldsymbol{Z}}_i^t]$, then we have*

$$\Delta R(\hat{\boldsymbol{Z}}_i, \varepsilon) \geq 0, \quad (5)$$

with equality holds if and only if $P^s_{\hat{\boldsymbol{Z}}|Y=i} = P^t_{\hat{\boldsymbol{Z}}|Y=i}$.

This corollary provides a more detailed analysis of the lower bound and fully uncovers its inherent alignment properties. It shows the lower bound is a necessary and sufficient condition for cross-domain conditional distribution alignment, surpassing existing frameworks. Furthermore, it offers a theoretically supported method to implement MCR^2-based conditional alignment, achieving both conditional alignment and orthogonal discriminability within a unified framework. This leads to invariant representations with better geometric structure and clearer statistical meaning.

Based on this corollary, we learn the conditional alignment criterion by performing domain-wise partitioning of

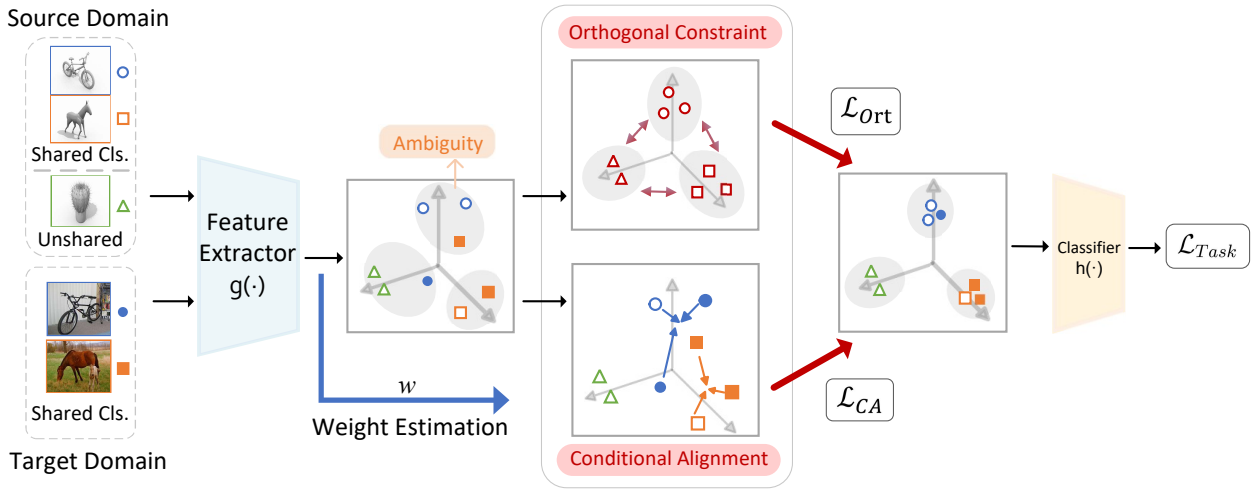

Figure 2: This is a pipeline of our GCA method. Before adaptation, class-specific feature subspaces are ambiguous. After applying GCA, the orthogonality constraint promotes inter-class orthogonality to enhance discriminability, and the weighted class-conditional alignment improves intra-class compactness. Consequently, the learned features possess better geometric structure and statistical properties, resulting in improved transferability and discriminability.

features within each class, i.e., $\boldsymbol{Z}_i = [\boldsymbol{Z}_i^s, \boldsymbol{Z}_i^t]$. Firstly, we estimate the weight $\boldsymbol{w}$ using the BBSE (Lipton, Wang, and Smola 2018) algorithm, whose convergence property is theoretically ensured. Then, our conditional alignment criterion is formulated as follows:

$$\begin{aligned} \arg\min_{g(\cdot)} \mathcal{L}_{\text{CA}} &= \sum_{i=1}^{k} w_i \Delta R(\hat{\boldsymbol{Z}}_i, \varepsilon) \\ &= \sum_{i=1}^{k} w_i [R(\hat{\boldsymbol{Z}}_i, \varepsilon) - R(\hat{\boldsymbol{Z}}_i^s, \varepsilon) - R(\hat{\boldsymbol{Z}}_i^t, \varepsilon)] \end{aligned} \tag{6}$$

By minimizing this loss, we achieve conditional distribution alignment between domains. Specifically, features of shared classes are aligned class-wisely across the source and target domains. This process encourages the features of shared classes across domains to become more compact, thereby enhancing the overall intra-class consistency. Consequently, the decision boundaries become more distinct, further improving inter-class discriminability. These improvements collectively enhance the model's classification performance on the target domain.

Theorem 2. $\log\det(I + c\boldsymbol{Z}\boldsymbol{Z}^\top) = ||\log(I + c\boldsymbol{Z}\boldsymbol{Z}^\top)||_* = \sum_{i=1}^{r} \log(c\sigma_i^2 + 1)$

This theorem provides valuable insight into the optimization objective of $\Delta R(\boldsymbol{Z}, \varepsilon)$. Maximizing $\Delta R(\boldsymbol{Z}, \varepsilon)$ encourages an increase in the rank of the overall representation matrix, which essentially expands the overall representation space. This helps prevent the model from learning a trivial solution with most subspaces collapsing to a point as zero element, which would cause clusters to lose the discriminative properties they are supposed to learn. Conversely, minimizing $\Delta R(\boldsymbol{Z}_i, \varepsilon)$ reduces the rank of each class-specific representation matrix, which essentially compresses the representation space of each class. This helps the model achieve better alignment between the source and target domains.

From the above, we propose a unified framework that jointly achieves conditional alignment and orthogonal discriminability. The optimization objective is formulated as:

$$\arg\min_{g(\cdot)} \mathcal{L}_{\text{GCA}} = \lambda_1 \mathcal{L}_{\text{CA}} - \lambda_2 \mathcal{L}_{\text{Ort}}, \tag{7}$$

where λ_1 and λ_2 are non-negative coefficients. $\mathcal{L}_{\text{GCA}}$ provides a unified perspective that integrates orthogonal constraints and conditional alignment to learn invariant representations. When $\lambda_1 = 0$, the objective degenerates to $\mathcal{L}_{\text{Ort}}$, in which case the model only focuses on orthogonal discriminability among classes. When $\lambda_2 = 0$, the objective degenerates to $\mathcal{L}_{\text{CA}}$, in which case the model only focuses on conditional alignment across domains. The later numerical experiments will show that combining both components yields better performance than using either alone.

Model and Algorithm

The overall GCA model consists of two parts, i.e., the risk objective for task learning and the geometry-aware conditional alignment constraint for invariant representation learning. We estimate the weight $\boldsymbol{w}$ using the BBSE algorithm, where $\sum_{i=1}^{k} w_i = 1$.

Task Learning. Given the weight vector $\boldsymbol{w}$, minimizing the weighted empirical risk on the source domain associated with the cross-entropy loss can be formulated as:

$$\arg\min_{g(\cdot), h(\cdot)} \mathcal{L}_{RCE} = \sum_{i=1}^{n^s} \sum_{j=1}^{k} -w_j y_{ij}^s \log \hat{y}_{ij}^s \tag{8}$$

where $\hat{\boldsymbol{y}}_i^s = h(g(x_i^s))$ satisfying $\sum_{j=1}^{k} \hat{y}_{ij}^s = 1$ and $\hat{y}_{ij}^s$ is the prediction probability of x_i^s belonging to the j-th class. y_{ij}^s is the ground truth label of x_i^s.

Algorithm 1: Geometry-aware Conditional Alignment

Input: Source $\mathcal{D}^s = \{(x_i^s, y_i^s)\}_{i=1}^{n^s}$, target $\mathcal{D}^t = \{x_i^t\}_{i=1}^{n^t}$.
Parameter: Loss parameter λ_1,λ_2, prescribed precision ε, maximum iteration N and pre-training iteration N_{pre}.
Output: Network parameters $\mathbf{W}_g$, $\mathbf{W}_h$, predictions of target domain samples $\{\hat{y}_j^t\}_{j=1}^{n^t}$.

for $step = 1, \cdots, N_{pre}$ **do**
  Update the parametere of g and h via cross-entropy loss on the source domain $\mathcal{D}^s$;
end for
for $step = 1, \cdots, N - N_{pre}$ **do**
  Estimate the weight w by BBSE algorithm.
  Compute per-class mean of source feature μ_i^s, per-class mean of target feature μ_i^t.
  Use $\{z_i^s, y_i^s\}_{i=1}^{n^s}$ to compute $\mathcal{L}_{RCE}$ via Eq.(8).
  Compute $\mathcal{L}_{Ent}$ via Eq.(9).
  Compute the class-wise centralized data $\hat{\boldsymbol{Z}}^s$, $\hat{\boldsymbol{Z}}^t$.
  Use $\hat{\boldsymbol{Z}}^s$,$\hat{\boldsymbol{Z}}^t$ to compute the loss $\mathcal{L}_{GCA}$ via Eq. (7).
  Update the parametere of g and h via Eq.(10)
end for

The entropy minimization criterion is commonly used to reduce the uncertainty of predictions in the target domain and is typically considered as a regularization. The target entropy objective is formulated as:

$$\underset{g(\cdot),h(\cdot)}{\arg\min}\ \mathcal{L}_{Ent} = \sum_{i=1}^{n^t}\sum_{j=1}^{k} -\hat{y}_{ij}^t \log \hat{y}_{ij}^t \tag{9}$$

Invariant Representation Learning. By integrating the task learning loss with the geometry-aware conditional alignment constraint, we obtain the overall objective:

$$\underset{g(\cdot),h(\cdot)}{\arg\min}\ \mathcal{L} = \mathcal{L}_{Task} + \mathcal{L}_{GCA}, \tag{10}$$

where $\mathcal{L}_{Task} = \mathcal{L}_{RCE} + \lambda_t \mathcal{L}_{Ent}$ is the total task learning objective. Although this loss helps improve classification performance, the intrinsic properties of the learned features are unclear. Fortunately, the features learned by $\mathcal{L}_{GCA}$ are more interpretable. They show clear geometric structure and statistical meaning. Specifically, it encourages the subspaces of different classes to be orthogonal and the features of the same class across domains to be compact. An intuitive illustration is shown in Fig. 2. We refer to our method as GCA, and its detailed procedure is given in Algorithm 1.

Experiments and Analysis

Set Up. We evaluate the performance of GCA on four commonly used datasets for PDA: Office-Home (Venkateswara et al. 2017), VisDA-2017 (Peng et al. 2017), Office-31 (Saenko et al. 2010), and ImageCLEF (Caputo et al. 2014). Details of implementations are provided in appendix.

Comparison with SOTA Methods. We compare our method with several recent SOTA PDA methods, including DANN (Ganin et al. 2016), PADA, SAFN, DMP, ETN (Cao et al. 2019), DRCN, AR (Gu et al. 2021), AGAN (Kim and Hong 2021), GLC, GATE (Chen et al. 2022), CSDN (Li et al. 2023), IDSP, SAN++ (Cao et al. 2023), SLM (Sahoo et al. 2023), MLNet (Lu et al. 2024), Ma et al. (Ma et al. 2024), CC-loss (Jin et al. 2024) and LEAD (Qu et al. 2024).

Method	Office-31							VisDA
	A→W	A→D	W→A	W→D	D→A	D→W	**Avg.**	S→R
Source-only	75.6	83.4	85.0	98.1	83.9	96.3	87.1	45.3
DANN	73.6	81.5	86.1	98.7	82.8	96.3	86.5	51.0
PADA	86.5	82.2	95.4	**100**	92.7	93.3	92.7	53.5
SAFN	87.5	89.8	92.7	99.4	92.6	96.6	93.1	67.7
ETN	94.5	95.0	94.6	**100**	**96.2**	**100**	96.7	59.8
DRCN	90.8	94.3	94.8	**100**	95.2	**100**	95.9	58.2
AR	93.5	96.8	96.0	99.7	95.5	**100**	96.9	88.7
AGAN	97.3	94.3	95.7	**100**	95.7	**100**	97.2	67.7
GATE	86.2	89.5	94.4	98.6	93.5	**100**	93.7	75.6
MLNet	92.4	84.6	94.7	99.4	94.9	**100**	94.3	80.4
Ma et al.	94.6	91.7	94.1	99.4	94.1	98.7	95.4	-
LEAD	93.9	89.8	96.0	99.4	95.6	98.6	95.6	75.3
GCA	**98.6**	**98.3**	**96.4**	**100**	96.0	99.9	**98.2**	**93.5**

Table 1: Accuracies (%) on Office-31 and VisDA-2017.

ImageCLEF	I→P	P→I	I→C	C→I	C→P	P→C	**Avg**
Source-only	78.3	86.9	91.0	84.3	72.5	91.5	84.1
DANN	78.1	86.3	91.3	84.0	72.1	90.3	83.7
PADA	81.7	92.1	94.6	89.8	77.7	94.1	88.3
SAFN	79.5	90.7	93.0	90.3	77.8	94.0	87.5
DMP	82.4	**94.5**	96.7	**94.3**	78.7	96.4	90.5
Ma et al.	86.7	92.0	**97.0**	89.3	83.3	97.0	90.9
GCA	**88.1**	91.0	96.8	92.5	**86.0**	**97.4**	**92.0**

Table 2: Accuracies (%) on ImageCLEF.

VisDA-2017. As shown in Table 1, compared to the other datasets, VisDA-2017 presents a significantly larger sample size compared to other datasets, making it a more challenging domain adaptation scenario. Due to the difficulty of the task, most methods achieve results below 80%. Only MLNet and AR reach 80.4% and 88.7%, respectively. In contrast, our method achieves 93.5%, outperforming AR by 4.8%. These results highlight the effectiveness of our approach on large-scale datasets with complex domain shifts.

Office-31. Table 1 shows the results on Office-31, where GCA achieves the highest average accuracy of 98.2%, outperforming the second-best method AGAN by 1%. It is worth noting that our approach achieves the best performance on the challenging tasks A→W and W→A. Although the results on D→W and D→A are slightly lower than the best, they still reach state-of-the-art levels.

Office-Home. Table 3 reports the results on Office-Home, which features more categories and a larger domain shift compared to other datasets, making the transfer task more challenging. GCA outperforms geometry-based methods like GATE and LEAD by approximately 3.5%, as it learns geometric properties more accurately under theoretical guidance. Compared with domain alignment methods such as

Office-Home	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→C	R→P	**Avg.**
Source-only	46.3	67.5	75.9	59.1	59.9	62.7	58.2	41.8	74.9	67.4	48.2	74.2	61.3
PADA	52.0	67.0	78.7	52.2	53.8	59.0	52.6	43.2	78.8	73.7	56.6	77.1	62.1
ETN	59.2	77.0	79.5	62.9	65.7	75.0	68.3	55.4	84.4	75.7	57.7	84.5	70.4
AGAN	56.4	77.3	85.1	74.2	73.8	81.1	70.8	51.5	84.5	79.0	56.8	83.4	72.8
GLC	55.9	79.0	87.5	72.5	71.8	82.7	74.9	41.7	82.4	77.3	60.4	84.3	72.5
CSDN	57.3	78.1	87.0	71.0	70.1	79.0	75.8	54.9	86.0	79.6	61.3	84.7	73.7
GATE	55.8	75.9	85.3	73.6	70.2	83.0	72.1	59.5	84.7	79.6	63.9	83.8	74.0
IDSP	60.8	80.8	87.3	69.3	76.0	80.2	74.7	59.2	85.3	77.8	61.3	85.7	74.9
Ma et al.	60.6	75.2	85.3	67.4	66.8	77.8	70.2	55.0	84.7	74.2	53.7	81.2	71.2
LEAD	58.2	83.1	87.0	70.5	75.4	**83.3**	73.7	50.4	83.7	78.3	58.7	83.2	73.8
SAN++	61.3	81.6	88.6	72.8	76.4	81.9	74.5	57.7	87.2	79.7	63.8	**86.1**	76.0
CC-loss	61.1	84.4	85.6	72.3	75.9	79.2	75.0	62.0	83.4	78.8	**68.4**	83.3	75.8
SLM	61.1	84.0	**91.4**	**76.5**	75.0	81.1	74.6	58.6	**87.8**	**82.3**	57.8	83.5	76.0
GCA	**63.0**	**85.9**	89.7	75.8	**78.5**	**83.3**	**77.4**	**63.2**	87.3	79.7	61.9	85.1	**77.6**

Table 3: Accuracies (%) on Office-Home.

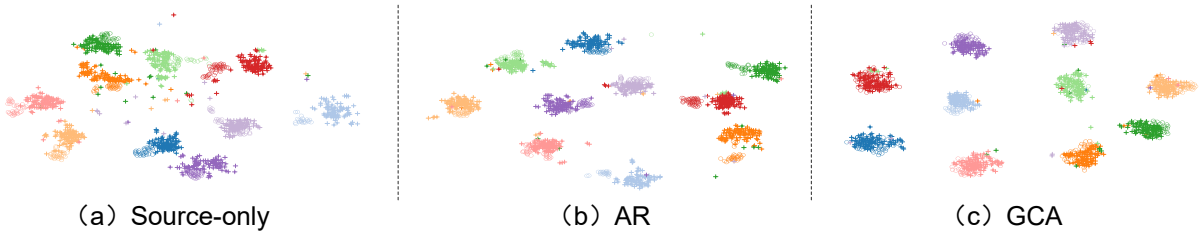

Figure 3: t-SNE visualization of features generated by Source-only, AR and GCA on Office-31 task W→A, respectively. Here, "o" denotes source domain and "+" denotes target domain. Each color denotes one class. Best viewed in color.

SAN++, IDSP, and Ma et al., our theoretically grounded method enables finer-grained conditional alignment, resulting in improvements of 1.6% to 6.4%. With an average accuracy of 77.6%, GCA surpasses all other methods, further demonstrating the effectiveness of our approach, which combines conditional alignment and orthogonal discriminability to address complex transfer scenarios.

ImageCLEF. Table 2 presents the results on ImageCLEF. Our method jointly learns orthogonal discriminability and conditional alignment, outperforming adversarial and metric-based methods with a top accuracy of 92%. It surpasses the second-best by 1.1% overall.

Feature Visualization. To intuitively evaluate the alignment of features in the low-dimensional subspace, we use t-SNE (Maaten and Hinton 2008) to visualize the features of the W→A task in Office-31. We compare the feature distributions of the Source-only (before adaptation), AR, and GCA, as shown in Fig. 3. Fig. 3(a) shows a significant discrepancy between the source and target domains before adaptation, with the source-only model struggling to classify target samples. In Fig. 3(b), although AR alleviates negative transfer to some extent, the intra-class compactness remains suboptimal for certain classes. In contrast, Fig. 3(c) shows that GCA achieves closer alignment of each class across domains. Compared to AR, our approach yields better inter-class separability and intra-class compactness. These results confirm GCA's effectiveness in learning a more discriminative and domain-invariant representation space.

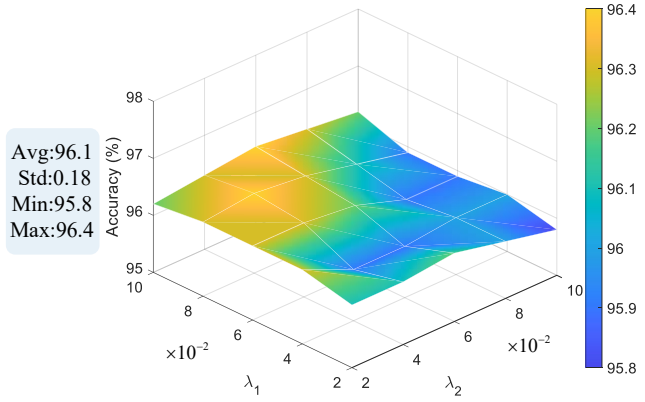

Figure 4: Classification accuracy under different settings of hyper-parameters. Some statistics are also provided.

Hyper-parameter. We investigate the sensitivity of hyper-parameters on Office-31 W→A. Parameters λ_1 and λ_2 act on the conditional alignment loss $\mathcal{L}_{CA}$ and orthogonal discriminative loss $\mathcal{L}_{Ort}$, respectively. The parameters λ_1 and λ_2 are searched from {2e-2,4e-2,6e-2,8e-2,1e-1}. Accuracies of the task by varying λ_1 and λ_2 are shown in Fig. 4. The results show that the performance of the model is robust for the different choice of parameters. The deviation is slight in a wide range of parameters variation. In fact, the accuracies of the model are stable in random experiments,

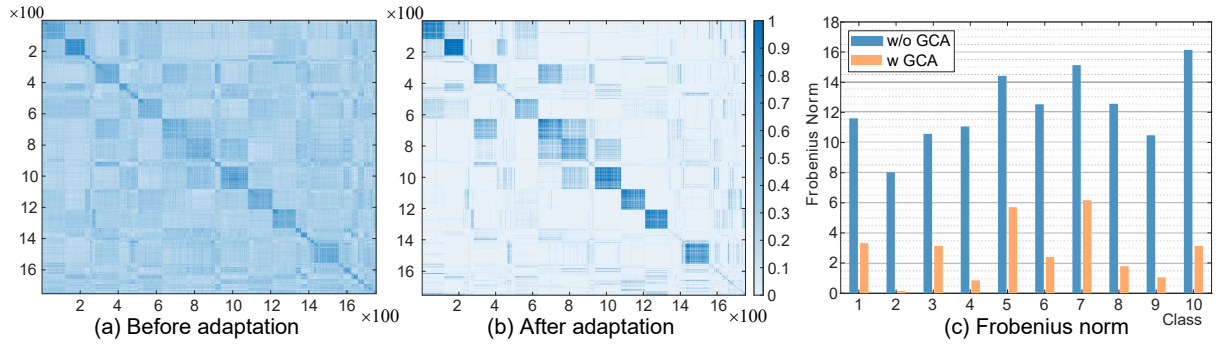

Figure 5: (a) (b) Cosine similarity between all features before and after adaptation. (c) The Frobenius norm of the difference between source and target covariance for each class before (blue) and after (orange) adaptation with GCA.

Objectives			ImageCLEF	VisDA-2017	Office-31
$\mathcal{L}_{task}$	$\mathcal{L}_{Ort}$	$\mathcal{L}_{CA}$	I→P	S→R	W→A
✓			78.3	45.3	85
✓	✓		80.3	92.0	89.2
✓		✓	86.3	82.7	95.3
✓	✓	✓	**88.1**	**93.5**	**96.4**

Table 4: Ablation study on major components of GCA.

with a standard deviation of 0.18. This further demonstrates that the proposed method is both robust from the view of randomness and parameter setting.

Ablation Study. To investigate the contributions of the conditional alignment objective $\mathcal{L}_{CA}$ and the orthogonal discriminative objective $\mathcal{L}_{Ort}$, an ablation study is conducted on three challenging tasks: W→A (Office-31), S→R (VisDA-2017) and I→P (ImageCLEF). The results are in Table 4. As shown in the second and third rows, both $\mathcal{L}_{CA}$ and $\mathcal{L}_{Ort}$ lead to notable performance improvements. Specifically, the domain gap is larger and the baseline accuracy is lower on VisDA-2017. The reliability of pseudo-labels for $\mathcal{L}_{CA}$ is limited, making $\mathcal{L}_{Ort}$ more critical. However, incorporating $\mathcal{L}_{CA}$ remains essential for achieving better performance. Lastly, the full model consistently outperforms each individual objective by at least 1.4% on average, demonstrating the complementary nature of $\mathcal{L}_{CA}$ and $\mathcal{L}_{Ort}$. Overall, these results confirm that both orthogonal discriminability and conditional alignment are vital for effective PDA.

Orthogonal Analysis. To quantitatively analyze and verify the orthogonal discriminability of the learned representations, We compute the cosine similarity between all learned features $\boldsymbol{Z}$ (mixed from both source and target domains), and visualize the results as a heatmap on the W→A task of Office-31. This provides a more intuitive explanation for the orthogonality between classes. As shown in Fig. 5(a) and Fig. 5(b), compared with the model before adaptation, the model trained with our method exhibits a more distinct diagonal structure in the heatmap. Moreover, the off-diagonal values are significantly reduced. Most off-diagonal values are close to zero, indicating that the subspaces of different classes are nearly orthogonal and that the learned features are discriminative. Meanwhile, a clearer block-wise structure appears at the shared classes, where the values within the diagonal blocks are close to 1, demonstrating that the proposed conditional alignment effectively learns transferable features. In summary, these results indicate that the proposed principles ensure both the geometric properties and the transferability of the learned representations.

Statistical Analysis. To verify the correctness of our theorem, we conduct experiments on the W→A task of Office-31. Specifically, we focus on the shared classes between the source and target domains (10 in total for Office-31). For each class, we compute the Frobenius norm of the difference between its source and target covariance matrices, i.e., $||\boldsymbol{Cov}_i^s - \boldsymbol{Cov}_i^t||_F$. The results in Fig. 5(c) show that w/o $\mathcal{L}_{GCA}$, the differences in covariance matrices for each shared class are large, indicating significant discrepancies between source and target distributions for each class. This leads to poor cross-domain generalization. In contrast, with $\mathcal{L}_{GCA}$ applied, the Frobenius norm of the covariance difference for each shared class is significantly reduced, demonstrating that the covariance matrices of shared classes from different domains become much more aligned. This further confirms the effectiveness of our proposed learning principle based on Gaussian prior for conditional alignment.

Conclusion

In this paper, we propose a novel framework for Partial Domain Adaptation, termed Geometry-aware Conditional Alignment (GCA). The main theoretical results demonstrate the feasibility of jointly learning both geometry-related and statistic-related properties of features. Based on the derived learning principle, GCA simultaneously accounts for geometric and statistical characteristics to enhance both transferability and discriminability of learned representations. This theoretically grounded model ensures the interpretability of the learned invariant features. Extensive experiments on four benchmark datasets validate the effectiveness of GCA. The theoretical insights are well supported by empirical results, which confirm that our model successfully captures the desired geometric and statistical properties. In future work, we plan to explore broader applications of GCA in other domain adaptation scenarios.

Acknowledgments

This work is supported in part by National Key R&D Program of China (2024YFA1011900), National Natural Science Foundation of China (Grant No. 62376291), Guangdong Basic and Applied Basic Research Foundation (2023B1515020004), Science and Technology Program of Guangzhou (2024A04J6413), and the Fundamental Research Funds for the Central Universities, Sun Yat-sen University (24xkjc013).

References

Borgwardt, K. M.; Gretton, A.; Rasch, M. J.; Kriegel, H.-P.; Schölkopf, B.; and Smola, A. J. 2006. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22: e49–e57.

Cao, Z.; Long, M.; Wang, J.; and Jordan, M. I. 2018a. Partial transfer learning with selective adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2724–2732.

Cao, Z.; Ma, L.; Long, M.; and Wang, J. 2018b. Partial Adversarial Domain Adaptation. In *Proceedings of the European Conference on Computer Vision*.

Cao, Z.; You, K.; Long, M.; Wang, J.; and Yang, Q. 2019. Learning to transfer examples for partial domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2985–2994.

Cao, Z.; You, K.; Zhang, Z.; Wang, J.; and Long, M. 2023. From Big to Small: Adaptive Learning to Partial-Set Domains. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45: 1766–1780.

Caputo, B.; Müller, H.; Martinez-Gomez, J.; Villegas, M.; Acar, B.; Patricia, N.; Marvasti, N.; Üsküdarlı, S.; Paredes, R.; Cazorla, M.; et al. 2014. Imageclef 2014: Overview and analysis of the results. In *International conference of the cross-language evaluation forum for European languages*, 192–211.

Chen, L.; Lou, Y.; He, J.; Bai, T.; and Deng, M. 2022. Geometric anchor correspondence mining with uncertainty modeling for universal domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 16134–16143.

Chen, X.; Wang, S.; Long, M.; and Wang, J. 2019. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In *International Conference on Machine Learning*, 1081–1090.

Courty, N.; Flamary, R.; Tuia, D.; and Rakotomamonjy, A. 2016. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39: 1853–1865.

Cui, S.; Wang, S.; Zhuo, J.; Li, L.; Huang, Q.; and Tian, Q. 2020. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3941–3950.

Fernando, B.; Habrard, A.; Sebban, M.; and Tuytelaars, T. 2013. Unsupervised visual domain adaptation using subspace alignment. In *Proceedings of the IEEE International Conference on Computer Vision*, 2960–2967.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; March, M.; and Lempitsky, V. 2016. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17: 1–35.

Gu, X.; Yu, X.; Sun, J.; Xu, Z.; et al. 2021. Adversarial reweighting for partial domain adaptation. *Advances in Neural Information Processing Systems*, 34: 14860–14872.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Jin, Y.; Cao, Z.; Wang, X.; Wang, J.; and Long, M. 2024. One fits many: Class confusion loss for versatile domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46: 7251–7266.

Kim, Y.; and Hong, S. 2021. Adaptive graph adversarial networks for partial domain adaptation. *IEEE Transactions on Circuits and Systems for Video Technology*, 32: 172–182.

Kirchmeyer, M.; Rakotomamonjy, A.; de Bezenac, E.; and Gallinari, P. 2021. Mapping conditional distributions for domain adaptation under generalized target shift. *arXiv preprint arXiv:2110.15057*.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.

Li, M.; Zhai, Y.-M.; Luo, Y.-W.; Ge, P.-F.; and Ren, C.-X. 2020a. Enhanced transport distance for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 13936–13944.

Li, S.; Gong, K.; Xie, B.; Liu, C. H.; Cao, W.; and Tian, S. 2023. Critical Classes and Samples Discovering for Partial Domain Adaptation. *IEEE Transactions on Cybernetics*, 53: 5641–5654.

Li, S.; Liu, C. H.; Lin, Q.; Wen, Q.; Su, L.; Huang, G.; and Ding, Z. 2020b. Deep residual correction network for partial domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43: 2329–2344.

Li, W.; and Chen, S. 2023. Partial Domain Adaptation Without Domain Alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45: 8787–8797.

Lipton, Z.; Wang, Y.-X.; and Smola, A. 2018. Detecting and correcting for label shift with black box predictors. In *International Conference on Machine Learning*, 3122–3130.

Liu, H.; Wang, J.; and Long, M. 2021. Cycle self-training for domain adaptation. *Advances in Neural Information Processing Systems*, 34: 22968–22981.

Long, M.; Cao, Y.; Cao, Z.; Wang, J.; and Jordan, M. I. 2018a. Transferable representation learning with deep adaptation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41: 3071–3085.

Long, M.; Cao, Z.; Wang, J.; and Jordan, M. I. 2018b. Conditional adversarial domain adaptation. *Advances in Neural Information Processing Systems*, 31.

Lu, Y.; Shen, M.; Ma, A. J.; Xie, X.; and Lai, J.-H. 2024. Mlnet: Mutual learning network with neighborhood invariance for universal domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 3900–3908.

Luo, Y.-W.; and Ren, C.-X. 2021. Conditional bures metric for domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 13989–13998.

Luo, Y.-W.; Ren, C.-X.; Dai, D.-Q.; and Yan, H. 2020. Unsupervised domain adaptation via discriminative manifold propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44: 1653–1669.

Ma, Y.; Derksen, H.; Hong, W.; and Wright, J. 2007. Segmentation of multivariate mixed data via lossy data coding and compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29: 1546–1562.

Ma, Y.; Yao, X.; Chen, R.; Li, R.; Shen, X.; and Yu, B. 2024. Small is Beautiful: Compressing Deep Neural Networks for Partial Domain Adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 35: 3575–3585.

Maaten, L. v. d.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9: 2579–2605.

Pan, S. J.; and Yang, Q. 2009. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22: 1345–1359.

Peng, X.; Usman, B.; Kaushik, N.; Hoffman, J.; Wang, D.; and Saenko, K. 2017. Visda: The visual domain adaptation challenge. *arXiv preprint arXiv:1710.06924*.

Prabhu, V.; Khare, S.; Kartik, D.; and Hoffman, J. 2021. Sentry: Selective entropy optimization via committee consistency for unsupervised domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, 8558–8567.

Qu, S.; Zou, T.; He, L.; Röhrbein, F.; Knoll, A.; Chen, G.; and Jiang, C. 2024. Lead: Learning decomposition for source-free universal domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 23334–23343.

Saenko, K.; Kulis, B.; Fritz, M.; and Darrell, T. 2010. Adapting visual category models to new domains. In *Proceedings of the European Conference on Computer Vision*, 213–226.

Sahoo, A.; Panda, R.; Feris, R.; Saenko, K.; and Das, A. 2023. Select, label, and mix: Learning discriminative invariant feature representations for partial domain adaptation. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 4210–4219.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sohn, K.; Berthelot, D.; Carlini, N.; Zhang, Z.; Zhang, H.; Raffel, C. A.; Cubuk, E. D.; Kurakin, A.; and Li, C.-L. 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in Neural Information Processing Systems*, 33: 596–608.

Stojanov, P.; Li, Z.; Gong, M.; Cai, R.; Carbonell, J.; and Zhang, K. 2021. Domain adaptation with invariant representation learning: What transformations to learn? *Advances in Neural Information Processing Systems*, 34: 24791–24803.

Venkateswara, H.; Eusebio, J.; Chakraborty, S.; and Panchanathan, S. 2017. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5018–5027.

Xiao, N.; and Zhang, L. 2021. Dynamic weighted learning for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 15242–15251.

Xiao, W.; Ding, Z.; and Liu, H. 2021. Implicit semantic response alignment for partial domain adaptation. *Advances in Neural Information Processing Systems*, 34: 13820–13833.

Xu, R.; Li, G.; Yang, J.; and Lin, L. 2019. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, 1426–1435.

Yu, Y.; Chan, K. H. R.; You, C.; Song, C.; and Ma, Y. 2020. Learning diverse and discriminative representations via the principle of maximal coding rate reduction. *Advances in Neural Information Processing Systems*, 33: 9422–9434.

Zhang, J.; Ding, Z.; Li, W.; and Ogunbona, P. 2018. Importance weighted adversarial nets for partial domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8156–8164.

Zhao, H.; Des Combes, R. T.; Zhang, K.; and Gordon, G. 2019. On learning invariant representations for domain adaptation. In *International Conference on Machine Learning*, 7523–7532.

Zou, Y.; Yu, Z.; Liu, X.; Kumar, B.; and Wang, J. 2019. Confidence regularized self-training. In *Proceedings of the IEEE International Conference on Computer Vision*, 5982–5991.