Optimal Graph Learning and Nuclear Norm Maximization for Deep Cross-Domain Robust Label Propagation

Wei Wang[1], **Hanyang Li**[1], **Ke Shi**[2,3], **Chao Huang**[1],
Yang Cao[4], **Cong Wang**[1,5] and **Xiaochun Cao**[1*]
[1]School of Cyber Science and Technology, Shenzhen Campus of Sun Yat-sen University
[2]Institute of Information Engineering, Chinese Academy of Sciences
[3]School of Cyber Security, University of Chinese Academy of Sciences
[4]Department of Computer Science, Tokyo Institute of Technology
[5]Department of Computing, The Hong Kong Polytechnic University
{wangwei29, huangch253, caoxiaochun}@mail.sysu.edu.cn, lihy377@mail2.sysu.edu.cn,
shike@iie.ac.cn, cao@c.titech.ac.jp, supercong94@gmail.com

Abstract

Domain adaptation aims to achieve label transfer from a labeled source domain to an unlabeled target domain, where the two domains exhibit different distributions. Existing methods primarily concentrate on designing a feature extractor to learn better domain-invariant features, along with developing an effective classifier for reliable predictions. In this paper, we introduce optimal graph learning to generate a cross-domain graph that effectively connects the two domains, and two domain-specific graphs to capture domain-specific structures. On the one hand, we incorporate the three graphs into the label propagation (LP) classifier to enhance its robustness to distribution difference. On the other hand, we leverage the three graphs to introduce graph embedding losses, promoting the learning of locally discriminative and domain-invariant features. Furthermore, we maximize the nuclear norm of predictions in LP to enhance class diversity, thereby improving its robustness to class imbalance problem. Correspondingly, we develop an efficient algorithm to solve the associated optimization problem. Finally, we integrate the proposed LP and graph embedding losses into a deep neural network, resulting in our proposed deep cross-domain robust LP. Extensive experiments conducted on three cross-domain benchmark datasets demonstrate that our proposed approach could outperform existing state-of-the-art domain adaptation methods.

1 Introduction

Deep neural networks have advanced numerous computer vision tasks [Chen *et al.*, 2024; Huang *et al.*, 2021; Huang *et al.*, 2022a; Huang *et al.*, 2022b; Huang *et al.*, 2023a; Huang *et al.*, 2024]. However, they typically rely on large amounts of annotated data to ensure the model's generalization ability in a target domain of interest, making the process time-consuming and resource-intensive [Wu *et al.*, 2023b; Wu *et al.*, 2023c]. It is intuitive to leverage a well-labeled source domain to train a model for the target domain, assuming they follow the same distributions. However, in real-life scenarios, the assumption often becomes challenging to hold due to factors such as environmental variations, differences in equipment, and human-related factors. Domain adaptation (DA), in this regard, aims to reduce manual annotation costs by improving the model's generalization ability trained on a source domain to a target domain, where the two domains could be allowed to follow different distributions [Wang *et al.*, 2021; Wang *et al.*, 2023b; Wang *et al.*, 2020; Wu *et al.*, 2024].

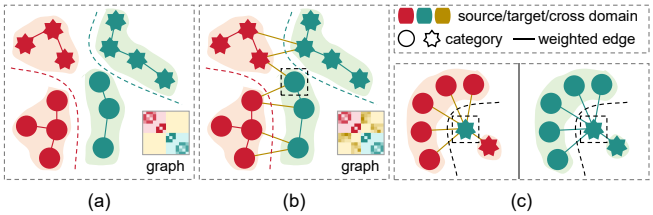

Figure 1: The motivation of this work. (a) LP is sensitive to distribution shift; (b) Optimal graph learning for LP; (c) LP is sensitive to class imbalance problem in the source (left)/target (right) domain.

Existing DA approaches enhance the model's generalization ability mainly from two aspects: feature extraction and classifier design. Feature extraction typically involves designing a distribution distance [Long *et al.*, 2015] or a network module [Long *et al.*, 2018] to discover domain-invariant features from both the source and target domains. As DA often employs the pseudo-labeling technique for label-induced losses related to the target domain, the classifier design primarily aims to improve the quality of pseudo-labels [Wang *et al.*, 2022; Wang and Breckon, 2020]. These two aspects are mutually reinforcing, as better domain-invariant features lead to improved pseudo-labels and vice versa. As an effec-

*Corresponding author: Xiaochun Cao.

tive classifier, label propagation (LP) constructs a similarity graph by exploring the structural relationships among different samples. It then designs a smooth function over this similarity graph to propagate labels from the source domain to the target domain based on the assumption that close samples in feature space should have similar labels as possible [Iscen *et al.*, 2019; Song *et al.*, 2023].

Although LP has been applied in many shallow DA methods [Ding *et al.*, 2018; Li *et al.*, 2020], it has not been fully explored in the context of deep DA as the graph construction has to consider all samples globally, which may not be well-suited for the batch sampling strategy in deep neural networks. Moreover, LP is sensitive to the distribution shift problem and fails to deal with the class imbalance problem. As shown in Figure 1 (a), the two domains follow very different distributions, and LP treats the source and target domains as a union to construct a similarity graph. However, the similarity weights between sample points that are separately from the two domains become close to 0 (i.e., no weighted edges), directly resulting in inaccurate LP from the source domain to the target domain. It can be observed that the colors are faint in the upper right corner and lower left corner of the similarity graph. As shown in Figure 1 (c), the number of class **heptagon** is smaller than that of class **circle** in the source (left) and target (right) domains. For example, we predict a target sample whose ground-truth is class **heptagon**. Although it is closer to the **heptagon** sample, the LP loss is smaller than the loss formed by all **circle** samples. Therefore, to minimize the overall LP loss, there is a tendency to wrongly classify this sample as the class **circle**.

This paper proposes a deep cross-domain robust label propagation (DCDRLP) approach to address the aforementioned issues. Specifically, we address the distribution shift and class imbalance issues in LP from two perspectives: graph construction and regularization construction. As for graph construction, we introduce optimal graph learning (OGL) to optimize three graphs. Specifically, instead of treating the source and target domains as a union, we construct a cross-domain graph across the two domains, ensuring that the similarity weight between two samples from different domains is not excessively small. As illustrated in Figure 1 (b), compared to Figure 1 (a), the colors in the upper right corner and lower left corner of the similarity graph have become darker. Moreover, we also construct two domain-specific graphs for the source and target domains to respect their own structures, respectively. As illustrated in Figure 1 (b), the sample in the black dashed box may be wrongly classified to the class **heptagon** with only the cross-domain graph. When we consider its own structure, the other samples from the class **circle** will correct it based on the assumption that close samples should have similar labels.

As for regularization construction, inspired by [Cui *et al.*, 2020], we introduce the nuclear norm maximization (NNM) into LP, to enhance the class diversity of prediction results, effectively addressing the class imbalance problem. However, as optimizing the LP model with NNM is nontrivial, we propose an efficient optimization algorithm. Finally, we incorporate the proposed LP into a deep neural network by supervising the classifier layer with the cross-entropy loss. Moreover, we utilize both features and labels to conduct OGL and generate three discriminative graphs. Then, inspired by the graph embedding (GE) technique, we construct three GE losses for the network to learn locally discriminative and domain-invariant features. Specifically, we enable two samples to close if they are close in feature space from the same class, where the two samples come from the same domain or two different domains. The main contributions of this paper are summarized as follows:

- We introduce OGL and NNM into LP so that it is robust to the problems of distribution shift and class imbalance. Moreover, we devise an efficient algorithm to solve the corresponding optimization problem.
- We incorporate the proposed LP into a deep neural network to generate reliable pseudo-labels. Moreover, we construct three GE losses with OGL to learn locally discriminative and domain-invariant features.
- We validate the effectiveness of our proposed approach on three cross-domain benchmark datasets, and it could outperform existing state-of-the-art methods.

2 Related Work

Existing DA approaches typically adopt the following strategies to bridge the gap between the source and target domains: statistical distribution distance metrics [Zhu *et al.*, 2021], optimal transport [Damodaran *et al.*, 2018], adversarial learning [Long *et al.*, 2018], self-training based on pseudo-labels [Zou *et al.*, 2019; Liu *et al.*, 2021], feature disentanglement [Bousmalis *et al.*, 2016], batch normalization [Chang *et al.*, 2019], and data augmentation [Li *et al.*, 2021b], among others. However, since DA assumes that target domain data can participate in the training process, many methods based on the aforementioned strategies utilize pseudo-labels of the target domain to enhance feature discriminability.

As discussed above, the pseudo-labeling technique has been widely applied in DA, making the quality of pseudo-labels crucial for DA performance. Existing methods mainly address this issue from three aspects: 1) using a certain threshold criterion to select more reliable pseudo-labels [Wang and Breckon, 2020]; 2) designing label-induced losses to make them less sensitive to the quality of pseudo-labels [Pan *et al.*, 2019; Pei *et al.*, 2018; Ding *et al.*, 2018; Wang *et al.*, 2022; Zhu *et al.*, 2021; Wang *et al.*, 2023c]; 3) devising classifiers capable of generating more reliable pseudo-labels [Wang and Breckon, 2020; Liang *et al.*, 2019; Wang *et al.*, 2022].

Considering the effectiveness of LP, recent work utilizes LP to generate pseudo-labels for DA [Ding *et al.*, 2018; Li *et al.*, 2019; Zhang *et al.*, 2020; Li *et al.*, 2020; Luo *et al.*, 2020; Wang *et al.*, 2022; Wang *et al.*, 2023a]. However, they usually utilize the most basic form of LP and apply LP to a shallow DA model. In contrast, this paper proposes a novel LP robust to the problems of distribution shift and class imbalance. Moreover, inspired by [Iscen *et al.*, 2019], we incorporate the proposed LP into a deep DA model.

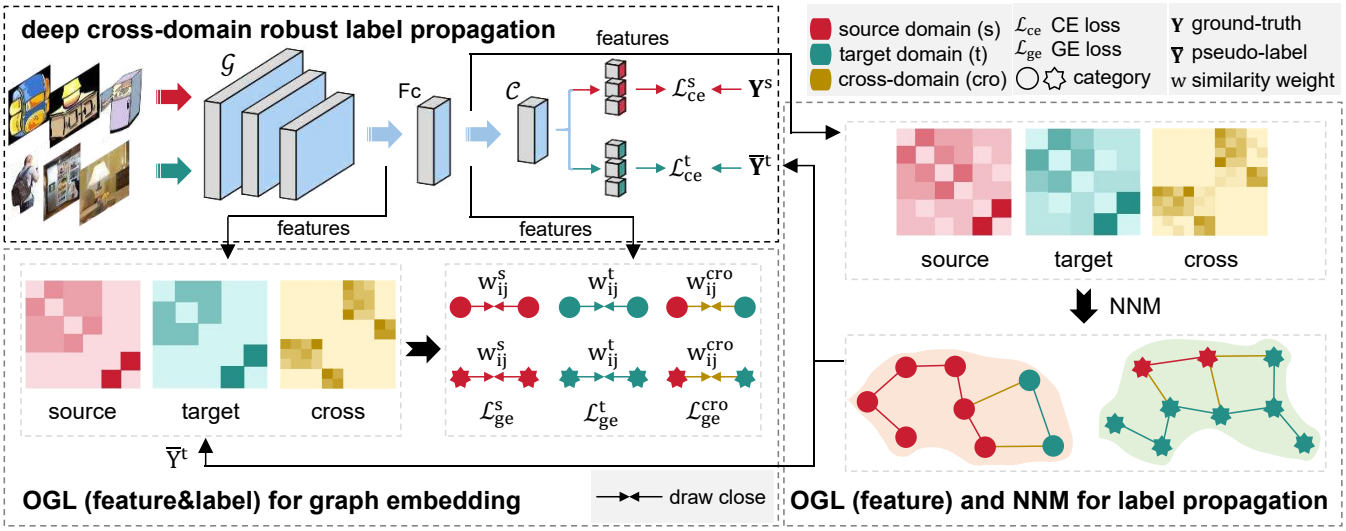

Figure 2: The whole pipeline of our proposed DCDRLP. (1) We utilize a backbone network $\mathcal{G}$ to extract features for the source and target domains, respectively. Moreover, we utilize a classifier layer $\mathcal{C}$ after a fully connected layer 'Fc' to realize label prediction for the two domains. (2) We conduct OGL to generate three graphs for LP and GE, respectively. We optimize the three graphs with features and labels for GE to prompt feature discriminability. Moreover, we introduce the constraint of NNM into LP. (3) We utilize the GE losses and pseudo-labels of the target domain generated by LP to supervise the network.

3 Proposed Approach

In DA, there exist a labeled source domain $\mathcal{D}^{\mathrm{s}} := \{\mathbf{X}^{\mathrm{s}} \in \mathbb{R}^{\mathrm{m}\times \mathrm{n}^{\mathrm{s}}}, \mathbf{Y}^{\mathrm{s}} \in \mathbb{R}^{\mathrm{C}\times \mathrm{n}^{\mathrm{s}}}\}$ and an unlabeled target domain $\mathcal{D}^{\mathrm{t}} := \{\mathbf{X}^{\mathrm{t}} \in \mathbb{R}^{\mathrm{m}\times \mathrm{n}^{\mathrm{t}}}\}$ with two different distributions. 'm' is the feature dimension, 'n^{s}' and 'n^{t}' are the sample numbers of source and target domains, and 'C' is the category number. We aim to utilize labels of the source to realize label prediction for the target as accurately as possible. The whole pipeline of our proposed DCDRLP is briefly depicted in Figure 2.

3.1 Label Propagation

As existing DA approaches usually utilize the pseudo-labeling technique to optimize some label-induced losses related to the target domain, it is crucial to devise a promising classifier for more reliable pseudo-labels. As an effective classifier, LP could propagate labels of the source domain to the target domain by representing the structural relationships among different samples as a similarity graph and devising a smooth function over the graph based on the assumption that close samples in feature space should have similar labels. We formulate LP as the following model,

$$\arg\min_{\bar{\mathbf{Y}}^{\mathrm{t}}} \sum_{\mathrm{i,j}=1}^{\mathrm{n}^{\mathrm{s}}+\mathrm{n}^{\mathrm{t}}} \left(\mathrm{w}_{\mathrm{ij}}||\bar{\mathbf{y}}_{\mathrm{i}} - \bar{\mathbf{y}}_{\mathrm{j}}||_2^2\right) \ \text{s.t.}\ \bar{\mathbf{Y}} = [\bar{\mathbf{Y}}^{\mathrm{s}}, \bar{\mathbf{Y}}^{\mathrm{t}}],\ \bar{\mathbf{Y}}^{\mathrm{s}} = \mathbf{Y}^{\mathrm{s}}, \tag{1}$$

where $\mathbf{Y}^{\mathrm{s}}$ is the ground-truth and $\bar{\mathbf{Y}}^{\mathrm{s}}$, $\bar{\mathbf{Y}}^{\mathrm{t}}$ are prediction results. $\mathrm{w}_{\mathrm{ij}} \in \mathbf{W}$ denotes the similarity weight between i-th and j-th samples in feature space, and w_{ij} is large when they are close in feature space. Optimizing model (1) enables two sample labels with large similarity weight close, and the model has a closed-form solution involving an inverse matrix [Iscen *et al.*, 2019]. Next, we will address the sensitivity of LP to distribution shift and class imbalance issues, proposing corresponding solutions from the perspectives of graph construction and regularization construction.

3.2 Optimal Graph Learning

As discussed, LP constructs the similarity graph $\mathbf{W}$ on the union of the source domain and target domain. As such, the similarity weights connecting the two domains are close to 0 when they follow very different distributions. To this end, we construct one cross-domain and two domain-specific graphs separately. We define the construction of these two types of graphs as the following optimization processes,

$$\begin{aligned}\arg\min_{\mathbf{W}} &\sum_{\mathrm{i,j}=1}^{\mathrm{n,n}^{\star}} \left(\mathrm{w}_{\mathrm{ij}}\left(||\mathbf{z}_{\mathrm{i}} - \mathbf{z}_{\mathrm{j}}^{\star}||_2^2 + ||\mathbf{y}_{\mathrm{i}} - \mathbf{y}_{\mathrm{j}}^{\star}||_2^2\right)\right) \\ &+ \alpha_1||\mathbf{W}||_{\mathrm{F}}^2,\ \text{s.t.}\ \mathbf{W} \succeq \mathbf{0},\ \mathbf{W1} = \mathbf{1},\end{aligned} \tag{2}$$

$$\begin{aligned}\arg\min_{\mathbf{W}} &\sum_{\mathrm{i,j}=1}^{\mathrm{n,n}} \left(\mathrm{w}_{\mathrm{ij}}\left(||\mathbf{z}_{\mathrm{i}} - \mathbf{z}_{\mathrm{j}}||_2^2 + ||\mathbf{y}_{\mathrm{i}} - \mathbf{y}_{\mathrm{j}}||_2^2\right)\right) \\ &+ \alpha_2||\mathbf{W}||_{\mathrm{F}}^2,\ \text{s.t.}\ \mathbf{W} \succeq \mathbf{0},\ \mathbf{W1} = \mathbf{1},\end{aligned} \tag{3}$$

where $\mathbf{z}$ is the learned feature vector from $\mathcal{G}$ or 'Fc'. To simplify the notation, n, $\mathbf{z}$, and $\mathbf{y}$ are for a specific domain, while $\mathrm{n}^{\star}$, $\mathbf{x}^{\star}$, and $\mathbf{y}^{\star}$ are for the other different domain.

The first term of the two models enables two close samples in feature space with similar labels to have large similarity weights. The second term avoids a trivial solution where only the nearest data point can be the neighbor of $\mathbf{z}_{\mathrm{i}}$ with a similarity weight of 1. Notably, α_1 and α_2 could be determined by the number of nearest neighbors k [Nie *et al.*, 2014].

We optimize model (2) between source and target domains to obtain two graphs $\mathbf{W}^{\text{st}}$ and $\mathbf{W}^{\text{ts}}$, and denote their average graph as $\mathbb{W}^{\text{cro}} = (\mathbf{W}^{\text{st}} + \mathbf{W}^{\text{ts}\top})/2$. Then, we construct our cross-domain graph as $\mathbf{W}^{\text{cro}} = [\mathbf{0}, \mathbb{W}^{\text{cro}}; \mathbb{W}^{\text{cro}\top}, \mathbf{0}]$. Moreover, we optimize model (3) on source and target domains to obtain our domain-specific graphs $\mathbf{W}^{\text{ss}}$ and $\mathbf{W}^{\text{tt}}$, respectively. Notably, we only utilize features to conduct OGL for LP, while both features and labels are adopted for graph embedding losses, which will be introduced later.

3.3 Deep Cross-Domain Robust Label Propagation

Given the data features of the source and target domains, we obtain three similarity graphs through the aforementioned OGL. Then, we utilize these graphs to enable LP to be robust to distribution shift. Additionally, inspired by [Cui *et al.*, 2020], we further introduce a nuclear norm maximization loss into LP to enhance its robustness to class imbalance issue. The formal definition is expressed in the following model,

$$\begin{aligned}&\arg\min_{\bar{\mathbf{Y}}^{\text{t}}} \sum_{\text{i,j=1}}^{\text{n}^{\text{s}}+\text{n}^{\text{t}}} \left(\text{w}_{\text{ij}}^{\text{cro}} ||\bar{\mathbf{y}}_{\text{i}} - \bar{\mathbf{y}}_{\text{j}}||_2^2\right) + \delta \sum_{\text{i,j=1}}^{\text{n}^{\text{t}},\text{n}^{\text{t}}} \left(\text{w}_{\text{ij}}^{\text{tt}} ||\bar{\mathbf{y}}_{\text{i}}^{\text{t}} - \bar{\mathbf{y}}_{\text{j}}^{\text{t}}||_2^2\right) \\ &- \sigma||\bar{\mathbf{Y}}^{\text{t}}||_\star, \text{ s.t. } \bar{\mathbf{Y}}^{\text{t}} \succeq \mathbf{0}, \ \bar{\mathbf{Y}}^{\text{t}\top}\mathbf{1} = \mathbf{1}, \ \bar{\mathbf{Y}}^{\text{s}} = \mathbf{Y}^{\text{s}},\end{aligned} \tag{4}$$

Optimizing the first term enables two samples close in feature space from different domains to have similar labels. Optimizing the second term makes two samples close in feature space from the target domain have similar labels. Optimizing the third term will increase the class diversity of predictions. Intuitively, the first term aims to propagate labels from source to target. In contrast, the second term utilizes the structure of the target domain itself to correct some label propagation errors in the first term. Moreover, we introduce a non-negative constraint, a normalization constraint, and a constraint that predicted source labels should be equal to the ground-truth. δ and σ are the hyper-parameters that balance the corresponding terms' relative importance.

Different from LP, 1) We split LP into two parts with two types of graphs so that it is robust to distribution shift; 2) We introduce a constraint of nuclear norm maximization to deal with the class imbalance problem; 3) We introduce non-negative and normalization constraints to ensure the probability meaning of predicted labels; 4) We solve the corresponding optimization problem with an efficient algorithm.

Although LP has been applied to many shallow DA methods, it has not been fully explored in deep DA as the graph construction requires all samples globally, which is unsuited for batch sampling strategy in deep neural networks. Inspired by [Iscen *et al.*, 2019], we utilize the predictions of the target domain by model (4) to supervise the classifier layer in a deep neural network with the cross-entropy loss, enabling it to obtain better pseudo-labels from model (4).

3.4 Graph Embedding

Besides an effective classifier, existing DA methods also focus on designing a feature extractor to obtain discriminative and domain-invariant features. Graph embedding (GE), as an effective feature learning technique, aims to incorporate the similarity graph constructed in the original feature space into a dimensionality reduction framework, preserving the local manifold structure in the original feature space [Zheng *et al.*, 2011; Cai *et al.*, 2011]. Inspired by this, we conduct OGL in the feature space extracted by $\mathcal{G}$, using ground-truth labels from the source domain and pseudo-labels from the target domain. This results in three similarity graphs with discriminative structures. Subsequently, the three graphs are applied into the feature space extracted by 'Fc', yielding the following three graph embedding losses,

$$\mathcal{L}_{\text{ge}}^{\text{s}} := \frac{1}{(\text{n}^{\text{s}})^2} \sum_{\text{i,j=1}}^{\text{n}^{\text{s}},\text{n}^{\text{s}}} \left(\text{w}_{\text{ij}}^{\text{ss}} ||\mathbf{z}_{\text{i}}^{\text{s}} - \mathbf{z}_{\text{j}}^{\text{s}}||_2^2\right) \tag{5}$$

$$\mathcal{L}_{\text{ge}}^{\text{t}} := \frac{1}{(\text{n}^{\text{t}})^2} \sum_{\text{i,j=1}}^{\text{n}^{\text{t}},\text{n}^{\text{t}}} \left(\text{w}_{\text{ij}}^{\text{tt}} ||\mathbf{z}_{\text{i}}^{\text{t}} - \mathbf{z}_{\text{j}}^{\text{t}}||_2^2\right) \tag{6}$$

$$\mathcal{L}_{\text{ge}}^{\text{cro}} := \frac{1}{2\text{n}^{\text{s}}\text{n}^{\text{t}}} \sum_{\text{i,j=1}}^{\text{n}^{\text{s}}+\text{n}^{\text{t}}} \left(\text{w}_{\text{ij}}^{\text{cro}} ||\mathbf{z}_{\text{i}} - \mathbf{z}_{\text{j}}||_2^2\right) \tag{7}$$

where the first two losses aim to prompt sample features from the same domain close if they are close in feature space of $\mathcal{G}$ and have similar labels. The third loss aims to prompt sample features from two domains to close if they are close in feature space $\mathcal{G}$ and have similar labels. As such, we could learn locally discriminative and domain-invariant features by minimizing the above losses for better DA effects.

After introducing the various modules above, we obtain the overall loss function for the proposed model as below,

$$\mathcal{L}_{\text{ce}}^{\text{s}} + \alpha\mathcal{L}_{\text{ce}}^{\text{t}} + \beta\Big(\lambda(\mathcal{L}_{\text{ge}}^{\text{s}} + \mathcal{L}_{\text{ge}}^{\text{t}}) + (1-\lambda)\mathcal{L}_{\text{ge}}^{\text{cro}}\Big) \tag{8}$$

where $\mathcal{L}_{\text{ce}}$ is the cross-entropy loss, $\mathcal{L}_{\text{ge}}$ is the loss for locally discriminative and domain-invariant feature learning. Moreover, α, β, and λ are the hyper-parameters that balance different losses' importance.

4 Experiments

We compare our proposed DCDRLP with 16 state-of-the-art DA methods, including 3 LP-based shallow DA methods: DTLC [Li *et al.*, 2020], CRLP [Wang *et al.*, 2022], RMMD [Wang *et al.*, 2023a], and 13 deep DA methods: CDAN [Long *et al.*, 2018], BSP [Chen *et al.*, 2019], BNM [Cui *et al.*, 2020], DSAN [Zhu *et al.*, 2021], ATM [Li *et al.*, 2021a], TSA [Li *et al.*, 2021b], GDCAN [Li *et al.*, 2022], DMP [Luo *et al.*, 2022], DALN [Chen *et al.*, 2022], SDAT [Rangwani *et al.*, 2022], DRDA [Huang *et al.*, 2023b], MEDM [Wu *et al.*, 2023a], MetaReg [Li *et al.*, 2023]. Besides, we adopt the variant of the highest results from their papers. As shallow DA methods fail to deal with large-scale datasets, we do not record their results on VisDA-2017.

4.1 Datasets

We conduct comprehensive experiments on three cross-domain image classification datasets. **Office-31** consists of 31 categories shared from three distinct domains: Amazon

Source	Venue	Amazon		Dslr	Webcam	Avg.
Target		D	W	A	A	
DTLC	TNNLS'20	91.0	85.3	73.5	73.5	80.8
CRLP	TCSVT'22	96.2	93.2	77.4	76.8	85.9
RMMD	TNNLS'23	90.8	88.9	75.4	75.2	82.6
CDAN	NeurIPS'18	92.9	94.1	71.0	69.3	81.8
BSP	ICML'19	93.0	93.3	73.6	72.6	83.1
BNM	CVPR'20	92.9	92.8	73.5	73.8	83.3
DSAN	TNNLS'21	90.2	93.6	73.5	74.8	83.0
ATM	TPAMI'21	96.4	95.7	74.1	73.5	84.9
TSA	CVPR'21	95.4	96.0	76.7	76.8	<u>86.2</u>
GDCAN	TPAMI'22	93.6	94.8	76.9	74.4	84.9
DMP	TPAMI'22	91.0	93.0	71.4	70.2	81.4
DALN	CVPR'22	95.4	95.5	75.0	75.1	85.3
DRDA	TIP'23	94.5	95.8	75.6	76.6	85.6
MEDM	TNNLS'23	93.2	93.4	75.1	75.4	84.3
MetaReg	TKDE'23	96.2	95.2	76.8	74.6	85.7
Ours	-	96.2	95.8	77.9	77.9	**87.0**

Table 1: Accuracy results (%) on Office-31

(A), Dslr (D), and Webcam (W), and they have 2,817, 498, and 795 images, respectively. **Office-Home** includes 65 categories shared from four different domains: Art (A), Clipart (C), Product (P) and Real-World (R). The four domains have 2,427, 4,365, 4,439, and 4,357 images. **VisDA-2017** is a large-scale synthetic-to-real dataset which have 12 shared categories. The synthetic domain and real domain have 152,397 and 55,388 images, respectively. Following [Li *et al.*, 2021b], we utilize one domain as the source domain and another as the target domain to construct 4, 12, and 1 DA tasks for the three datasets, respectively. Notably, the easy tasks of D→W and W→D in Office-31 are not included as most existing DA methods obtain nearly 100% accuracy results. Moreover, in VisDA-2017, we regard the synthetic as the source domain and the real as the target domain.

4.2 Implementation Details

We use ResNet-50 as the backbone network for Office-31 and Office-Home, while ResNet-101 is for VisDA-2017. Moreover, we use a mini-batch SGD optimizer with a momentum of 0.9 and a weight decay of 0.001. We set the batch size for the source and target domains to 32, with a learning rate of 0.01. The entire network training process spans 10 epochs, each comprising 500 iterations. For comparison results, we set hyper-parameters of α as 0.5. As for k, the numbers of nearest neighbors for both cross-domain graph and domain-specific graph are set as 10. For a fair comparison, we adopt 20 nearest neighbors for the graph construction of standard LP in the ablation study. As for β, λ, δ, and σ, we select the optimal values for each DA task through trials.

4.3 Results

We record the accuracy results on the three datatsets in Tables 1, 2, 3. The best average results are in bold font, and the second-best ones are underlined. From these tables, we can observe that our proposed approach outperforms all compared methods on average. The average results of our proposed approach are 87.0%, 73.6%, and 85.6%, which have 0.8%, 1.0%, and 1.3% improvements over the best baselines TSA, DALN, and SDAT/MetaReg, respectively. Furthermore, we can observe a greater improvement on Office-Home than on Office-31, attributed to the larger distribution shift within Office-Home. The most significant enhancement is obtained on VisDA-2017, as there exists a severe class imbalance issue, and we report accuracy results for each class. On the one hand, shallow DA methods also use LP to obtain reliable pseudo-labels. Still, they do not consider LP sensitive to the distribution shift problem and fail to deal with the class imbalance problem. Additionally, most of them suffer from the limitation of handling large-scale datasets. On the other hand, deep DA methods enhance the quality of pseudo-labels or promote discriminative feature learning through different strategies. However, they do not take advantage of LP in label prediction. In contrast, our proposed approach applies LP to a deep DA model and addresses the distribution shift and class imbalance problems in LP. As a result, the proposed approach could produce more reliable pseudo-labels and learn more effective discriminative features for better DA effects.

4.4 Empirical Analysis

To validate the effectiveness of our proposed approach, we conduct ablation experiments on various key components of the model, visualize the feature distributions, and perform sensitivity analysis on some key hyper-parameters.

Ablation Study. We select four relatively challenging DA tasks on Office-Home and test the model's performance with the following variations. **w/o OGL** means that the graph construction strategy in our model is kept consistent with LP; **w/o NNM** indicates that we do not consider nuclear norm maximization; **w/o LP** indicates that we do not use the proposed LP to supervise the classifier layer in the network; **w/o $\mathbf{GE^s}$, $\mathbf{GE^t}$** refers to removing the domain-specific graph embedding losses, while **w/o $\mathbf{GE^{cro}}$** refers to removing the cross-domain graph embedding loss. We keep all parameters consistent across these variations to ensure a fair comparison.

From Table 4, we observe that removing or replacing key components in our proposed model leads to a certain performance decrease. The performance degradation with the variants of w/o OGL (2.4%) and w/o NNM (4.1%) reflects that OGL and NNM could make LP robust to distribution difference and class imbalance, respectively. The performance drop with the variant of w/o LP (4.1%) indicates that leveraging the advantage of our proposed LP in label prediction could enhance the performance of deep DA models. The performance decline with the variants of w/o GE^s, GE^t (0.4%) and w/o GE^{cro} (1.2%) validates that GE losses could prompt locally discriminative and domain-invariant features for better DA effects. By observing the impact of the abovementioned components on our model, we can verify the significance of our proposed LP with OGL and NNM.

To further validate the effectiveness of NNM in terms of class imbalance issue, we directly use features extracted by the backbone on VisDA-2017 as inputs for LP, LP with OGL, and LP with OGL and NNM. As shown in Figure 3, VisDA-2017 exhibits a severe class imbalance problem in the target domain. It can be observed that the sample number predicted as the majority class using LP (i.e., Figure 3(a)) is

Source	Venue	Artistic			Clipart			Product			Real-World			Avg.
Target		C	P	R	A	P	R	A	C	R	A	C	P	
DTLC	TNNLS'20	56.0	74.9	76.5	57.4	71.3	70.9	61.0	52.4	76.3	68.3	56.7	80.4	66.8
CRLP	TCSVT'22	56.6	78.5	79.9	66.3	78.7	79.4	67.2	55.1	80.8	70.7	59.6	83.7	71.4
RMMD	TNNLS'23	58.4	77.8	79.3	61.6	72.8	73.0	62.7	55.3	78.9	70.4	60.1	83.2	69.5
CDAN	NeurIPS'18	50.7	70.6	76.0	57.6	70.0	70.0	57.4	50.9	77.3	70.9	56.7	81.6	65.8
BSP	ICML'19	52.0	68.6	76.1	58.0	70.3	70.2	58.6	50.2	77.6	72.2	59.3	81.9	66.3
BNM	CVPR'20	56.2	73.7	79.0	63.1	73.6	74.0	62.4	54.8	80.7	72.4	58.9	83.5	69.4
DSAN	TNNLS'21	54.4	70.8	75.4	60.4	67.8	68.0	62.6	55.9	78.5	73.8	60.6	83.1	67.6
ATM	TPAMI'21	52.4	72.6	78.0	61.1	72.0	72.6	59.5	52.0	79.1	73.3	58.9	83.4	67.9
TSA	CVPR'21	57.6	75.8	80.7	64.3	76.3	75.1	66.7	55.7	81.2	75.7	61.9	83.8	71.2
GDCAN	TPAMI'22	57.3	75.7	83.1	68.6	73.2	77.3	66.7	56.4	82.2	74.1	60.7	83.0	71.5
DMP	TPAMI'22	52.3	73.0	77.3	64.3	72.0	71.8	63.6	52.7	78.5	72.0	57.7	81.6	68.1
DALN	CVPR'22	58.1	79.6	83.7	67.7	77.9	78.7	66.8	56.0	81.9	73.9	60.9	86.1	72.6
SDAT	ICML'22	58.2	77.1	82.2	66.3	77.6	76.8	63.3	57.0	82.2	74.9	64.7	86.0	72.2
DRDA	TIP'23	58.2	74.2	81.2	65.6	75.1	73.3	65.8	57.1	80.4	75.6	63.2	85.1	71.2
MEDM	TNNLS'23	57.5	77.5	83.2	69.1	78.9	80.7	66.6	54.9	83.4	74.9	59.8	85.4	72.5
MetaReg	TKDE'23	58.0	75.5	78.9	65.0	74.7	75.0	67.9	57.2	81.8	74.4	63.5	83.8	71.3
Ours	-	60.4	80.1	81.7	68.1	82.3	81.8	66.3	58.2	83.4	71.4	63.3	86.0	**73.6**

Table 2: Accuracy results (%) on Office-Home

Methods	Venue	airplane	bicycle	bus	car	horse	knife	motorcycle	person	plant	skateboard	train	truck	Avg.
CDAN	NeurIPS'18	-	-	-	-	-	-	-	-	-	-	-	-	70.0
BSP	ICML'19	92.4	61.0	81.0	57.5	89.0	80.6	90.1	77.0	84.2	77.9	82.1	38.4	75.9
DSAN	TNNLS'21	90.9	66.9	75.7	62.4	88.9	77.0	93.7	75.1	92.8	67.6	89.1	39.4	75.1
TSA	CVPR'21	-	-	-	-	-	-	-	-	-	-	-	-	82.0
DMP	TPAMI'22	92.1	75.0	78.9	75.5	91.2	81.9	89.0	77.2	93.3	77.4	84.8	35.1	79.3
DALN	CVPR'22	-	-	-	-	-	-	-	-	-	-	-	-	83.7
SDAT	ICML'22	95.8	85.5	76.9	69.0	93.5	97.4	88.5	78.2	93.1	91.6	86.3	55.3	84.3
MEDM	TNNLS'23	93.1	74.2	86.0	68.7	93.9	87.2	91.8	80.4	92.9	83.1	88.0	49.8	82.4
MetaReg	TKDE'23	96.1	86.2	81.3	62.9	94.5	97.4	89.7	79.1	88.2	93.4	88.5	53.9	84.3
Ours	-	97.0	85.9	85.0	64.0	96.7	98.2	88.5	82.5	85.7	94.9	91.6	56.8	**85.6**

Table 3: Average accuracy results (%) on VisDA-2017

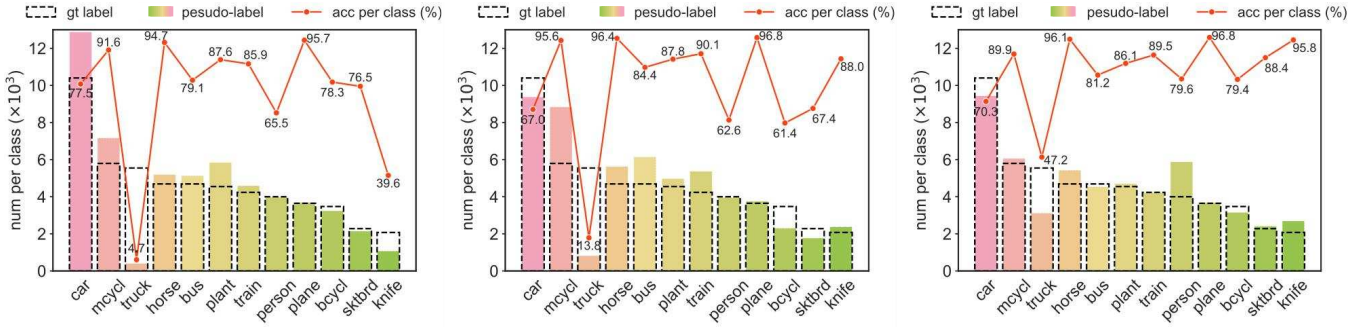

Figure 3: Empirical analysis for class imbalance problem on VisDA-2017. The sample numbers predicted with three different model variants and the accuracy results on each class, and the average accuracy results for all classes are 73.1%, 75.9%, and 83.4%, respectively. The black dashed boxes indicate the ground-truth sample numbers in each class, while the solid bars represent the predicted sample numbers.

much higher than the ground-truth. In contrast, the sample number predicted as the minority class is much smaller than the ground-truth. This leads to poor results for the minority classes, degrading the average accuracy for all classes accordingly. As graph construction with OGL is robust to distribution shift, including source domain samples into the target domain will somewhat mitigate the class imbalance problem.

From Figure 3(b), it can be observed that the sample numbers assigned to the minority classes increase, and the accuracy results of the minority classes begin to rise. When we further introduce NNM, the sample numbers of minority classes and the accuracy results further increase. Additionally, we observe that, although "truck" is a majority class, it is relatively smaller than "car". As there are similarities between

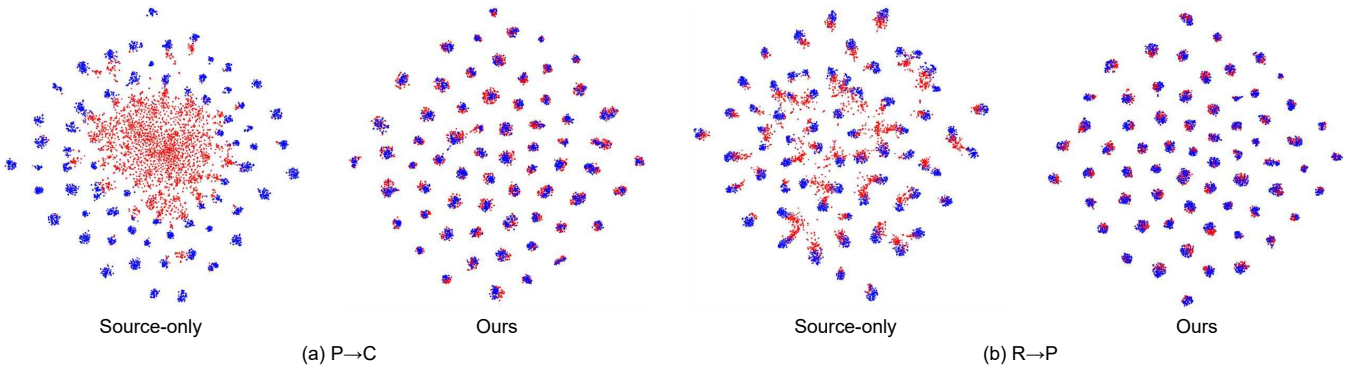

Figure 4: Feature visualization using tSNE on Office-Home.

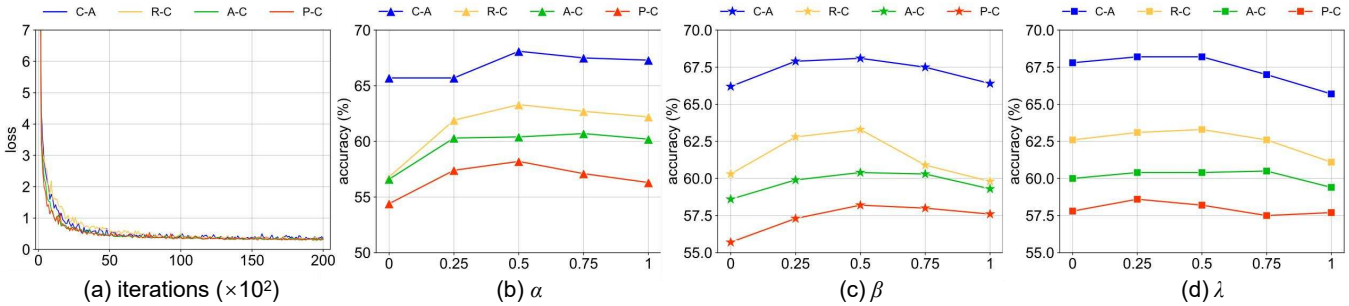

Figure 5: Empirical analysis for model convergence and hyper-parameter sensitivity on Office-Home.

Task	A→C	C→A	P→C	R→C	Avg.
w/o OGL	57.3	66.2	55.1	61.9	60.1
w/o NNM	58.7	58.6	55.4	60.7	58.4
w/o LP	56.6	65.7	54.4	56.8	58.4
w/o GE^s, GE^t	60.0	67.8	57.8	62.6	62.1
w/o GE^{cro}	59.4	65.7	57.5	62.6	61.3
Ours	**60.4**	**68.1**	**58.2**	**63.3**	**62.5**

Table 4: Ablation study (%) on Office-Home

these two categories, many samples belonging to the "truck" are misclassified as "car". Correspondingly, both Figure 3(b) and Figure 3(c) show significant improvements in the sample number and accuracy result on the class "truck".

Feature Visualization. We visualize the distribution of features extracted by our model on Office-Home. As shown in Figure 4, the good alignment of different color dots indicates better domain-invariant features. Moreover, better discriminative features are suggested when different sub-clusters are clustered more compactly. Our model's feature distribution exhibits better discriminative and domain-invariant characteristics than the source-only model (a model directly trained on the source for the target). Since our proposed LP can generate more reliable pseudo-labels, it facilitates the graph embedding process to extract more accurately locally discriminative and domain-invariant features.

Hyper-Parameter Sensitivity. We conduct sensitivity analysis on the crucial parameters of the proposed model. As shown in Figure 5(a), we conduct convergence analysis on our model and observe that the model achieves convergence at a relatively fast rate with increasing iterations. In Figure 5(b), α represents the impact of the proposed LP on the network's performance, and this curve shows a trend of first increasing and then decreasing, confirming the effectiveness of our proposed LP. In Figure 5(c), β demonstrates the significance of locally discriminative and domain-invariant feature learning, showing a trend of first increasing and then decreasing. In Figure 5(d), λ characterizes the importance between cross-domain GE and domain-specific GE. The former emphasizes learning domain-invariant features, while the latter focuses on preserving domain-specific structures. In addition, a larger λ will place more emphasis on domain-specific GE, and vice versa for cross-domain GE. It can be observed that better results are achieved when λ has a smaller value, indicating that cross-domain GE has a greater impact on model performance than domain-specific GE. Moreover, it can be observed that the proposed model is not particularly sensitive to these three parameters and can achieve good performance within a relatively wide range.

5 Conclusion

This paper proposes a novel classifier for DA, where we introduce OGL and NNM to be robust to distribution shift and class imbalance problems. Moreover, we incorporate the proposed classifier and GE losses into a deep neural network to obtain reliable pseudo-labels and learn locally discriminative and domain-invariant features for better DA effects. Extensive experiments on multiple cross-domain benchmarks could verify the effectiveness of our proposed approach.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.62306343, No.62025604, No.62301621), the Shenzhen Science and Technology Program (No.KQTD20221101093559018), and the Shenzhen Science and Technology Program (No.20231121172359002).

Contribution Statement

We state that Wei Wang, Hanyang Li, and Ke Shi made equal contributions to this work.

References

[Bousmalis *et al.*, 2016] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *NeurIPS*, pages 343–351, 2016.

[Cai *et al.*, 2011] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S. Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE TPAMI*, 33(8):1548–1560, 2011.

[Chang *et al.*, 2019] Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *IEEE CVPR*, pages 7354–7362, 2019.

[Chen *et al.*, 2019] Xinyang Chen, Sinan Wang, Mingsheng Long, and Jianmin Wang. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In *ICML*, volume 97, pages 1081–1090, 2019.

[Chen *et al.*, 2022] Lin Chen, Huaian Chen, Zhixiang Wei, Xin Jin, Xiao Tan, Yi Jin, and Enhong Chen. Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation. In *IEEE CVPR*, pages 7171–7180, 2022.

[Chen *et al.*, 2024] Ruoyu Chen, Hua Zhang, Siyuan Liang, Jingzhi Li, and Xiaochun Cao. Less is more: Fewer interpretable region via submodular subset selection. In *ICLR*, 2024.

[Cui *et al.*, 2020] Shuhao Cui, Shuhui Wang, Junbao Zhuo, Liang Li, Qingming Huang, and Qi Tian. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations. In *IEEE CVPR*, pages 3940–3949, 2020.

[Damodaran *et al.*, 2018] Bharath Bhushan Damodaran, Benjamin Kellenberger, Rémi Flamary, Devis Tuia, and Nicolas Courty. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. In *ECCV*, volume 11208, pages 467–483, 2018.

[Ding *et al.*, 2018] Zhengming Ding, Sheng Li, Ming Shao, and Yun Fu. Graph adaptive knowledge transfer for unsupervised domain adaptation. In *ECCV*, volume 11206, pages 36–52, 2018.

[Huang *et al.*, 2021] Chao Huang, Zongju Peng, Yong Xu, Fen Chen, Qiuping Jiang, Yun Zhang, Gangyi Jiang, and Yo-Sung Ho. Online learning-based multi-stage complexity control for live video coding. *IEEE TIP*, 30:641–656, 2021.

[Huang *et al.*, 2022a] Chao Huang, Zhihao Wu, Jie Wen, Yong Xu, Qiuping Jiang, and Yaowei Wang. Abnormal event detection using deep contrastive learning for intelligent video surveillance system. *IEEE TII*, 18(8):5171–5179, 2022.

[Huang *et al.*, 2022b] Chao Huang, Zehua Yang, Jie Wen, Yong Xu, Qiuping Jiang, Jian Yang, and Yaowei Wang. Self-supervision-augmented deep autoencoder for unsupervised visual anomaly detection. *IEEE TCYB*, 52(12):13834–13847, 2022.

[Huang *et al.*, 2023a] Chao Huang, Jie Wen, Yong Xu, Qiuping Jiang, Jian Yang, Yaowei Wang, and David Zhang. Self-supervised attentive generative adversarial networks for video anomaly detection. *IEEE TNNLS*, 34(11):9389–9403, 2023.

[Huang *et al.*, 2023b] Zenan Huang, Jun Wen, Siheng Chen, Linchao Zhu, and Nenggan Zheng. Discriminative radial domain adaptation. *IEEE TIP*, 32:1419–1431, 2023.

[Huang *et al.*, 2024] Chao Huang, Chengliang Liu, Jie Wen, Lian Wu, Yong Xu, Qiuping Jiang, and Yaowei Wang. Weakly supervised video anomaly detection via self-guided temporal discriminative transformer. *IEEE TCYB*, 54(5):3197–3210, 2024.

[Iscen *et al.*, 2019] Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. Label propagation for deep semi-supervised learning. In *IEEE CVPR*, pages 5070–5079, 2019.

[Li *et al.*, 2019] Jingjing Li, Mengmeng Jing, Ke Lu, Lei Zhu, and Heng Tao Shen. Locality preserving joint transfer for domain adaptation. *IEEE TIP*, 28(12):6103–6115, 2019.

[Li *et al.*, 2020] Shuang Li, Chi Harold Liu, Limin Su, Binhui Xie, Zhengming Ding, C. L. Philip Chen, and Dapeng Wu. Discriminative transfer feature and label consistency for cross-domain image classification. *IEEE TNNLS*, 31(11):4842–4856, 2020.

[Li *et al.*, 2021a] Jingjing Li, Erpeng Chen, Zhengming Ding, Lei Zhu, Ke Lu, and Heng Tao Shen. Maximum density divergence for domain adaptation. *IEEE TPAMI*, 43(11):3918–3930, 2021.

[Li *et al.*, 2021b] Shuang Li, Mixue Xie, Kaixiong Gong, Chi Harold Liu, Yulin Wang, and Wei Li. Transferable semantic augmentation for domain adaptation. In *IEEE CVPR*, pages 11516–11525, 2021.

[Li *et al.*, 2022] Shuang Li, Binhui Xie, Qiuxia Lin, Chi Harold Liu, Gao Huang, and Guoren Wang. Generalized domain conditioned adaptation network. *IEEE TPAMI*, 44(8):4093–4109, 2022.

[Li *et al.*, 2023] Shuang Li, Wenxuan Ma, Jinming Zhang, Chi Harold Liu, Jian Liang, and Guoren Wang. Meta-reweighted regularization for unsupervised domain adaptation. *IEEE TKDE*, 35(3):2781–2795, 2023.

[Liang *et al.*, 2019] Jian Liang, Ran He, Zhenan Sun, and Tieniu Tan. Aggregating randomized clustering-promoting invariant projections for domain adaptation. *IEEE TPAMI*, 41(5):1027–1042, 2019.

[Liu *et al.*, 2021] Hong Liu, Jianmin Wang, and Mingsheng Long. Cycle self-training for domain adaptation. In *NeurIPS*, pages 22968–22981, 2021.

[Long *et al.*, 2015] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I. Jordan. Learning transferable features with deep adaptation networks. In *ICML*, volume 37, pages 97–105, 2015.

[Long *et al.*, 2018] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I. Jordan. Conditional adversarial domain adaptation. In *NeurIPS*, pages 1647–1657, 2018.

[Luo *et al.*, 2020] Lingkun Luo, Liming Chen, Shiqiang Hu, Ying Lu, and Xiaofang Wang. Discriminative and geometry-aware unsupervised domain adaptation. *IEEE TCYB*, 50(9):3914–3927, 2020.

[Luo *et al.*, 2022] You-Wei Luo, Chuan-Xian Ren, Dao-Qing Dai, and Hong Yan. Unsupervised domain adaptation via discriminative manifold propagation. *IEEE TPAMI*, 44(3):1653–1669, 2022.

[Nie *et al.*, 2014] Feiping Nie, Xiaoqian Wang, and Heng Huang. Clustering and projected clustering with adaptive neighbors. In *KDD*, pages 977–986, 2014.

[Pan *et al.*, 2019] Yingwei Pan, Ting Yao, Yehao Li, Yu Wang, Chong-Wah Ngo, and Tao Mei. Transferrable prototypical networks for unsupervised domain adaptation. In *IEEE CVPR*, pages 2239–2247, 2019.

[Pei *et al.*, 2018] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. In *AAAI*, pages 3934–3941, 2018.

[Rangwani *et al.*, 2022] Harsh Rangwani, Sumukh K. Aithal, Mayank Mishra, Arihant Jain, and Venkatesh Babu Radhakrishnan. A closer look at smoothness in domain adversarial training. In *ICML*, volume 162, pages 18378–18399, 2022.

[Song *et al.*, 2023] Zixing Song, Xiangli Yang, Zenglin Xu, and Irwin King. Graph-based semi-supervised learning: A comprehensive review. *IEEE TNNLS*, 34(11):8174–8194, 2023.

[Wang and Breckon, 2020] Qian Wang and Toby P. Breckon. Unsupervised domain adaptation via structured prediction based selective pseudo-labeling. In *AAAI*, pages 6243–6250, 2020.

[Wang *et al.*, 2020] Wei Wang, Zhihui Wang, Haojie Li, Juan Zhou, and Zhengming Ding. Adaptive local neighbors for transfer discriminative feature learning. In *ECAI*, volume 325, pages 1595–1602, 2020.

[Wang *et al.*, 2021] Wei Wang, Shenglun Chen, Yuankai Xiang, Jing Sun, Haojie Li, Zhihui Wang, Fuming Sun, Zhengming Ding, and Baopu Li. Sparsely-labeled source assisted domain adaptation. *Pattern Recognit.*, 112:107803, 2021.

[Wang *et al.*, 2022] Wei Wang, Baopu Li, Mengzhu Wang, Feiping Nie, Zhihui Wang, and Haojie Li. Confidence regularized label propagation based domain adaptation. *IEEE TCSVT*, 32(6):3319–3333, 2022.

[Wang *et al.*, 2023a] Wei Wang, Haojie Li, Zhengming Ding, Feiping Nie, Junyang Chen, Xiao Dong, and Zhihui Wang. Rethinking maximum mean discrepancy for visual domain adaptation. *IEEE TNNLS*, 34(1):264–277, 2023.

[Wang *et al.*, 2023b] Wei Wang, Mengzhu Wang, Xiao Dong, Long Lan, Quannan Zu, Xiang Zhang, and Cong Wang. Class-specific and self-learning local manifold structure for domain adaptation. *Pattern Recognit.*, 142:109654, 2023.

[Wang *et al.*, 2023c] Wei Wang, Ziyi Wang, Mengzhu Wang, Haojie Li, and Zhihui Wang. Importance filtered soft label-based deep adaptation network. *Knowl. Based Syst.*, 265:110397, 2023.

[Wu *et al.*, 2023a] Xiaofu Wu, Suofei Zhang, Quan Zhou, Zhen Yang, Chunming Zhao, and Longin Jan Latecki. Entropy minimization versus diversity maximization for domain adaptation. *IEEE TNNLS*, 34(6):2896–2907, 2023.

[Wu *et al.*, 2023b] Yanan Wu, Zhixiang Chi, Yang Wang, and Songhe Feng. Metagcd: Learning to continually learn in generalized category discovery. In *IEEE ICCV*, pages 1655–1665, 2023.

[Wu *et al.*, 2023c] Yanan Wu, Tengfei Liang, Songhe Feng, Yi Jin, Gengyu Lyu, Haojun Fei, and Yang Wang. Metazscil: A meta-learning approach for generalized zero-shot class incremental learning. In *AAAI*, pages 10408–10416, 2023.

[Wu *et al.*, 2024] Yanan Wu, Zhixiang Chi, Yang Wang, Konstantinos N. Plataniotis, and Songhe Feng. Test-time domain adaptation by learning domain-aware batch normalization. In *AAAI*, pages 15961–15969, 2024.

[Zhang *et al.*, 2020] Yabin Zhang, Bin Deng, Kui Jia, and Lei Zhang. Label propagation with augmented anchors: A simple semi-supervised learning baseline for unsupervised domain adaptation. In *ECCV*, volume 12349, pages 781–797, 2020.

[Zheng *et al.*, 2011] Miao Zheng, Jiajun Bu, Chun Chen, Can Wang, Lijun Zhang, Guang Qiu, and Deng Cai. Graph regularized sparse coding for image representation. *IEEE TIP*, 20(5):1327–1336, 2011.

[Zhu *et al.*, 2021] Yongchun Zhu, Fuzhen Zhuang, Jindong Wang, Guolin Ke, Jingwu Chen, Jiang Bian, Hui Xiong, and Qing He. Deep subdomain adaptation network for image classification. *IEEE TNNLS*, 32(4):1713–1722, 2021.

[Zou *et al.*, 2019] Yang Zou, Zhiding Yu, Xiaofeng Liu, B. V. K. Vijaya Kumar, and Jinsong Wang. Confidence regularized self-training. In *IEEE ICCV*, pages 5981–5990, 2019.