Geometric Anchor Correspondence Mining with Uncertainty Modeling for Universal Domain Adaptation

Liang Chen [1], Yihang Lou [2], Jianzhong He [2*], Tao Bai [2], Minghua Deng [1†]
[1] School of Mathematical Sciences, Peking University, [2] Intelligent Vision Dept, Huawei Technologies

Abstract

*Universal domain adaptation (UniDA) aims to transfer the knowledge learned from a label-rich source domain to a label-scarce target domain without any constraints on the label space. However, domain shift and category shift make UniDA extremely challenging, which mainly lies in how to recognize both shared "known" samples and private "unknown" samples. Previous works rarely explore the intrinsic geometrical relationship between the two domains, and they manually set a threshold for the overconfident closed-world classifier to reject "unknown" samples. Therefore, in this paper, we propose a **G**eometric anchor-guided **A**dversarial and con**T**rastive learning framework with unc**E**rtainty modeling called GATE to alleviate these issues. Specifically, we first develop a random walk-based anchor mining strategy together with a high-order attention mechanism to build correspondence across domains. Then a global joint local domain alignment paradigm is designed, i.e., geometric adversarial learning for global distribution calibration and subgraph-level contrastive learning for local region aggregation. Toward accurate target private samples detection, GATE introduces a universal incremental classifier by modeling the energy uncertainty. We further efficiently generate novel categories by manifold mixup, and minimize the open-set entropy to learn the "unknown" threshold adaptively. Extensive experiments on three benchmarks demonstrate that GATE significantly outperforms previous state-of-the-art UniDA methods.*

1. Introduction

Deep neural networks have achieved impressive progress in image recognition tasks given abundant labeled data, however, they do not generalize well to novel unlabeled domain [28]. Annotating massive data from various domains is an expensive and time-consuming task, and these domains often sample from different distributions against each other. Such so-called domain shift would degenerate the model performance heavily [15]. To alleviate this issue, domain adaptation (DA) transfers knowledge from the label-rich source domain to the label-scarce target domain by diminishing the feature discrepancy across two domains [10]. Suppose that the label spaces of two domains are C_s and C_t, respectively, traditional unsupervised DA usually assumes $C_s = C_t$, i.e., close-set DA (CDA) [38]. In complex real-world scenarios, however, this assumption may not be easily satisfied. There are several situations we may encounter: $C_t \subset C_s$, i.e., partial DA (PDA) [2]; $C_s \subset C_t$, i.e., open-set DA (ODA) [29]; $C_s \cap C_t \neq \emptyset, C_s \cup C_t \neq C_s \, or \, C_t$, i.e., open-partial DA (OPDA) [16]. Although these variants have recently been resolved independently, the downside is that a method that is applicable to one variant may not be applicable to another. More difficultly, we cannot know which of these variants will occur in advance.

To better tackle these general situations, universal DA (UniDA) is raised to account for both domain shift and category shift, allowing two domains to own their private categories [44]. It assumes that we have no prior knowledge about their label space differences, thus being very practical in the real world. In UniDA, we aim to classify the target samples into either one of the "known" labels or the "unknown" label. However, there exist two main technical challenges in UniDA. First, we should constrain the domain bias removal into the common categories between two domains, and need to separate their respective private categories simultaneously. Second, we need to detect the potential "unknown" samples in the target domain without any target label supervision or other prior information. These two challenges are complementary and mutually constraining, since better removal of feature discrepancy can help identify the "unknown" sample, whereas, in turn, the model can help separate the common and private categories once it learns the concept of "unknown".

Recently, several efforts have been devoted to settling the UniDA task. For the first challenge, UAN [44] exploits both the domain similarity and the prediction uncertainty of each sample to incorporate a weighting mechanism into the adversarial network for promoting common-class adaptation. DANCE [32] moves each target sample either to a

*This work is completed in Huawei Technologies.
†Corresponding Author.

source "known" class prototype or to its target neighbors, and uses an entropy separation loss to encourage domain alignment. These methods rarely explore the intrinsic structural relationship between the two domains in the geometric manifold space, making it difficult to achieve fine-grained class-specific alignment and separation. For the second challenge, existing methods like CMU [9] manually set a threshold for confidence or entropy score outputted by the closed-world classifier to reject "unknown" samples. However, the class competition nature may cause the neural network to generate overconfident predictions for "unknown" instances, making the threshold hard to tune [43]. Besides, since the semantic information of "known" categories varies in different domains and tasks, it is hard to acquire an optimal threshold that suits all open-set recognition tasks [47].

To address these two challenges, here we propose a **G**eometric anchor-guided **A**dversarial and con**T**rastive learning framework with unc**E**rtainty modeling called GATE for UniDA task. First, we retrieve the mutual nearest neighbors (MNN) intra- and inter-domains as the geometric anchors to build the intrinsic structure correspondence between two domains. To better uncover the complete distribution of samples from the common categories, we develop a novel within-domain random walk-based strategy to expand the MNN anchor pair list. We also introduce a shared nearest neighbor (SNN) scoring mechanism to minimize the influence of incorrectly constructed anchors. Based on this fine-grained affinity knowledge across domains, we design a multi-anchor-constrained geometric adversarial learning paradigm for global distribution alignment, and propose a subgraph-level cross-domain contrastive learning objective to facilitate local region aggregation. Such a global joint local learning strategy better realizes the matching of common categories and the separation of private categories.

Second, to enable the model known "unknown", we transform the closed-world classifier into a universal formulation by adding extra dimension to model the energy-based uncertainty. Such an augmented classifier component also stands for a learnable threshold between "known" and "unknown". To efficiently anticipate the distribution of novel categories in training, we consider generating synthesized instances by manifold mixup technique, which mimics the target private categories with a limited computation cost. In this way, the open-set classification has in turn changed back to a generalized closed-set classification, thus adaptively estimating the "unknown" threshold. We also propose to apply open-set entropy minimization for unlabeled target samples, allowing the model to align them to either "known" or "unknown" categories. Given this insight, the idea of our method is simple yet powerful.

Our contribution can be summarized as follows:

- We propose a novel UniDA framework called GATE based on geometric anchor mining and uncertainty modeling. GATE designs a global joint local domain alignment strategy, i.e., geometric adversarial learning for global distribution calibration and subgraph-level contrastive learning for local region aggregation.
- We propose an energy-based universal classification paradigm that learns the "unknown" threshold adaptively. The open-set entropy minimization allows us to label the "known" samples and detect "unknown" samples in the target domain.
- We conduct extensive experiments on various UniDA benchmarks, and the empirical results show that GATE outperforms other state-of-the-art UniDA methods. Deeper analyses validate the effectiveness of individual components proposed in GATE.

2. Related work

2.1. Universal domain adaptation

Universal domain adaptation is a realistic but challenging DA scenario which allows both domains to have their own private categories. UAN [44] measures the sample-level transferability to distinguish the common and private categories. CMU [9] detects the target "unknown" samples by aggregating multiple complementary uncertainty measures. DANCE [32] designs two loss functions, neighborhood clustering and entropy separation, for category shift-agnostic adaptation. DCC [18] draws the domain consensus knowledge to facilitate the target clustering and the private category discovery. OVANet [33] trains a One-vs-All classifier for each source class and decides the "known" or "unknown" by using the output. However, these methods do not consider the intrinsic manifold structure relationship between two domains, thus making them suboptimal for domain alignment. In this paper, we design a global joint local learning framework to remove domain bias based on the geometric anchor correspondence across domains.

2.2. Contrastive learning

Recently, contrastive learning, a kind of self-supervised learning paradigm [23], has achieved impressively superior performance in many computer vision tasks [6, 13]. It aims to achieve instance-level discrimination and invariance, by pushing semantically distinct samples away while pulling semantically consistent samples closer in the feature space [7,42]. The existing contrastive learning methods can be mainly divided into three types: instance-based [11], cluster-based [4], and neighborhood-based [46]. In this paper, we first construct the random walk-based MNN pairs as positive anchors intra- and inter-domains, and then propose a cross-domain subgraph-level contrastive learning objective to aggregate local similar samples and separate dissimilar samples. By incorporating the high-order association into anchor scoring, the soften contrastive alignment process becomes more robust to avoid negative label transfer.

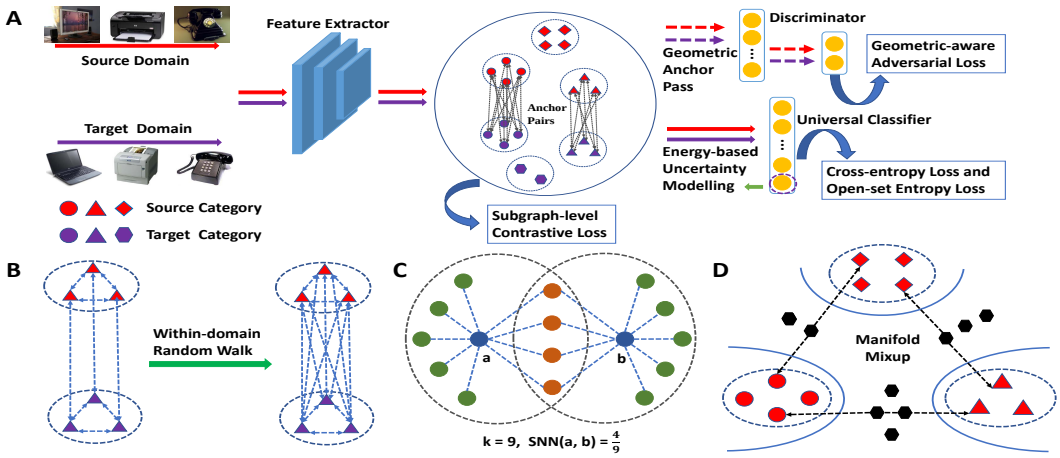


Figure 1. Schematics of GATE. (A) In the feature space, we use the subgraph-level contrastive loss to achieve local aggregation across domains. The geometric-aware adversarial loss aims to remove the global distribution-level domain shift. For the universal classifier, the open-set entropy minimization aligns the target samples to either "known" or "unknown" categories. (B) Within-domain random walk for anchor expansion. (C) The shared nearest neighbor score for robust correspondence. (D) Manifold mixup for generating novel samples.

3. Methods

In UniDA, we have access to a labeled source domain $\mathcal{D}_s = (x_i^s, y_i^s)_{i=1}^{n_s}$ and an unlabeled target domain $\mathcal{D}_t = (x_i^t)_{i=1}^{n_t}$ during training, which are sampled from different distributions p and q, respectively. Assume that the label sets of source domain and target domain are C_s and C_t, respectively, we denote $C = C_s \cap C_t$ as the common label set shared by both domains. $\bar{C}_s = C_s \setminus C$ and $\bar{C}_t = C_t \setminus C$ represent the label sets private to the source and target domain, respectively. Our goal is to train a model on $\mathcal{D}_s \cup \mathcal{D}_t$ to classify the target samples with either one of the "known" $|C_s|$ labels or the "unknown" label.

As shown in Figure 1(A), our model consists of three modules: (1) a feature extractor $\mathcal{F}$ that maps the input x into embedding feature $z = \mathcal{F}(x)$, and (2) an adversarial discriminator $\mathcal{D}$ that remove the global feature discrepancy of cross-domain geometric anchors, and (3) a universal classifier $\mathcal{G}$ that assigns the feature z into one of $|C_s|+1$ classes, with the $|C_s|+1$-th element representing the "unknown" uncertainty. The neural network should be able to produce high uncertainty scores for "unknown" samples.

3.1. Intrinsic geometric anchor correspondence mining intra- and inter-domains

The main challenge of UniDA is to align the common categories between the two domains and to separate the respective private categories. Naturally, removing feature discrepancy should be limited to the integration of common categories. Since the target domain does not have any prior structure knowledge of semantic categories, a natural question is how to identify the target sample is associated with the common category of the source domain? In manifold learning, a basic inductive bias is that samples that are neighbors in the feature space are more likely to come from the same category [5]. Correspondingly, we introduce a novel inductive bias for UniDA: samples from common categories (defined as anchors) are geometrically close to each other in the feature space. Therefore, a key step for domain alignment is the identification of anchors between two domains. These anchors represent two instances (with one instance from each domain), that we predict to originate from a common semantic category.

Specifically, we identify mutual nearest neighbor (MNN) pairs across domains as anchors. Assume that the features $\{z_m\}_{m=1}^{n_s+n_t}$ are l_2 normalized, for each sample i in source domain, we find the k samples in target domain with the smallest cosine distance to i, i.e., its k nearest neighbors $\mathcal{T}_k^i$ in target domain. We do the same for each sample j in target domain to find its k nearest neighbors $\mathcal{S}_k^j$ in source domain. If a pair of instances $(i \in \mathcal{D}_s, j \in \mathcal{D}_t)$ is contained in each other's set of k nearest neighbors, namely

$$i \in \mathcal{S}_k^j \wedge j \in \mathcal{T}_k^i, \tag{1}$$

those samples are considered to be mutual nearest neighbors. We interpret these anchor pairs as containing samples that belong to the same semantic category being generated in different domains. Thus, any systematic differences in feature representation between samples in MNN pairs should represent the domain shift. For convenience, we use $\mathcal{A}_{st} = \mathcal{A}_{ts}^T$ to denote the MNN affinity correspondence graph matrix between the two domains, and $\mathcal{A}_{st}(i,j) = 1$ when $(i \in \mathcal{D}_s, j \in \mathcal{D}_t)$ is a cross-domain MNN anchor pair, otherwise $\mathcal{A}_{st}(i,j) = 0$. Further, the semantic information represented by the MNN pair is also applicable to associate the clusters in target domains. Similarly, define $\mathcal{A}_{tt}$ as the MNN adjacency graph matrix in the target domain, we have $\mathcal{A}_{tt} = 1$ if and only if $(i \in \mathcal{D}_t, j \in \mathcal{D}_t)$ meets the following condition:

$$i \in \mathcal{T}_k^j \wedge j \in \mathcal{T}_k^i. \tag{2}$$

Considering that the source domain is labeled, its intrinsic

affinity adjacency matrix $\mathcal{A}_{ss}$ can be obtained according to whether the sample pairs have the same label.

3.2. Random walk-based MNN expansion and anchor pair scoring mechanism

One major potential drawback of the MNN pairs is that it is hard to assure these pairs could cover the complete distributions of samples from common categories. Therefore, we alternatively develop a within-domain random walk (rw) strategy to expand the MNN pair list (see Figure 1(B)). Specifically, suppose instance i_1 from source domain and instance j_1 from target domain is selected as an MNN pair. Among the k nearest neighbors $\mathcal{T}_k^{j_1}$ of j_1 within target domain, we arbitrarily pick one instance j_2. Conversely, the same procedure would give i_2 from source domain. Then, (i_2, j_2) is regarded as an extended MNN pair. For all MNN pairs, we could generate these kinds of new pairs. We call such pairs obtained from this procedure rwMNN pairs. The generated rwMNN pairs can better cover the distributions of matched common categories, which could strengthen the correspondence across domains. For convenience, we denote the anchor sets that can form cross-domain rwMNN pairs in the two domains as $\mathcal{M}_s$ and $\mathcal{M}_t$.

Robust identification of anchor correspondence is key for effective domain alignment. Incorrect anchor pairs representing instances from distinct categories can lead to error classification. In particular, instances that come from private categories should theoretically not participate in cross-domain anchor pairs, yet in practice, they will do so with low frequency. To minimize the influence of incorrectly identified anchors, we propose to weight anchors by shared nearest neighbor (SNN) score (see Figure 1(C)). Suppose $(i \in \mathcal{D}_s, j \in \mathcal{D}_t)$ is a rwMNN pair and their k nearest neighbor sets in two domains are $(\mathcal{S}_k^i, \mathcal{T}_k^i)$ and $(\mathcal{S}_k^j, \mathcal{T}_k^j)$, respectively, then their SNN score can be calculated by

$$\tilde{\mathcal{A}}_{st}(i,j) = SNN(i,j) = \frac{|\mathcal{S}_k^i \cap \mathcal{S}_k^j| + |\mathcal{T}_k^i \cap \mathcal{T}_k^j|}{2k}. \quad (3)$$

By examining the neighbor consistency in the local region, the SNN score introduces the high-order association to improve anchor reliability. The same procedure can also be able to apply on the target domain. For an anchor pair $(j_1 \in \mathcal{D}_t, j_2 \in \mathcal{D}_t)$, the SNN score is defined as

$$\tilde{\mathcal{A}}_{tt}(j_1,j_2) = SNN(j_1,j_2) = \frac{|\mathcal{T}_k^{j_1} \cap \mathcal{T}_k^{j_2}|}{k}. \quad (4)$$

After that, we replace the original binarized graph matrices with the weighted ones $\tilde{\mathcal{A}}_{st}$, $\tilde{\mathcal{A}}_{ts}$, $\tilde{\mathcal{A}}_{tt}$ and $\tilde{\mathcal{A}}_{ss}(=\mathcal{A}_{ss})$.

3.3. Feature alignment via geometric adversarial and contrastive learning

To achieve domain alignment in UniDA, we develop a global calibration joint local aggregation strategy from the constructed anchor correspondence within and across domains. Specifically, classical adversarial learning methods for CDA employ an adversarial domain discriminator $\mathcal{D}$ to distinguish the source samples from the target samples, while training the feature extractor $\mathcal{F}$ to generate embedding features to confuse the discriminator [36]:

$$\mathcal{L}_{adv} = -\mathbb{E}_{x\sim\mathcal{D}_t}[\log \mathcal{D}(\mathcal{F}(x))] \quad (5)$$

$$\mathcal{L}_{\mathcal{D}} = -\mathbb{E}_{x\sim\mathcal{D}_s}[\log \mathcal{D}(\mathcal{F}(x))] - \mathbb{E}_{x\sim\mathcal{D}_t}[\log(1-\mathcal{D}(\mathcal{F}(x)))]. \quad (6)$$

Min-max training between $\mathcal{L}_{\mathcal{D}}$ and $\mathcal{L}_{adv}$ yields the domain invariant feature representation. However, when the source domain and target domain have their own private classes, such an adversarial learning strategy will cause the distribution of private classes to be aligned to the common classes, resulting in the overcorrection of domain bias. To resolve this issue, we propose to perform adversarial learning on cross-domain rwMNN anchor pairs, namely,

$$\mathcal{L}_{adv}^{rw} = -\mathbb{E}_{x\sim\mathcal{M}_t}[\log \mathcal{D}(\mathcal{F}(x))] \quad (7)$$

$$\mathcal{L}_{\mathcal{D}}^{rw} = -\mathbb{E}_{x\sim\mathcal{M}_s}[\log \mathcal{D}(\mathcal{F}(x))] - \mathbb{E}_{x\sim\mathcal{M}_t}[\log(1-\mathcal{D}(\mathcal{F}(x)))]. \quad (8)$$

Our motivation here is that the difference in embedding features between samples in an rwMNN pair provides an estimate of the domain bias, which is made more precise by averaging across many such pairs. We use the adversarial network to obtain the estimation of domain shift implicitly and apply it to other samples to perform bias correction from the global feature distribution level. Using only the overlapping anchor subsets between two domains for feature discrepancy correction naturally avoids the assumption of equal category composition in unsupervised DA.

However, the global distribution-level adversarial alignment strategy may not be enough to guarantee local common class-specific alignment and private class separation, which might lead to negative transfer. To enforce the fine-grained class-level alignment, we design an rwMNN-guided contrastive learning framework to aggregate similar samples across domains and push away dissimilar samples. Specifically, let $S(z_i, z_j) = \exp(z_i z_j/\tau)$ represents the similarity of embedding features between two samples (x_i, x_j), where τ is a temperature parameter. Then we define two statistics for each sample based on the weighted adjacency graph matrices, i.e., total intra-subgraph similarity and total inter-subgraph similarity, as below

$$\mathcal{S}_{intra}(x_i) = \begin{cases} \text{if } x_i \in \mathcal{D}_s, \\ \sum_{\tilde{\mathcal{A}}_{ss}(i,j)>0} \tilde{\mathcal{A}}_{ss}(i,j) S(z_i,z_j) + \\ \sum_{\tilde{\mathcal{A}}_{st}(i,k)>0} \tilde{\mathcal{A}}_{st}(i,k) S(z_i,z_k); \\ \text{if } x_i \in \mathcal{D}_t, \\ \sum_{\tilde{\mathcal{A}}_{ts}(i,j)>0} \tilde{\mathcal{A}}_{ts}(i,j) S(z_i,z_j) + \\ \sum_{\tilde{\mathcal{A}}_{tt}(i,k)>0} \tilde{\mathcal{A}}_{tt}(i,k) S(z_i,z_k). \end{cases} \quad (9)$$

$$\mathcal{S}_{inter}(x_i) = \begin{cases} \text{if } x_i \in \mathcal{D}_s, \\ \sum_{\tilde{\mathcal{A}}_{ss}(i,j)=0} S(z_i,z_j) + \sum_{\tilde{\mathcal{A}}_{st}(i,k)=0} S(z_i,z_k); \\ \text{if } x_i \in \mathcal{D}_t, \\ \sum_{\tilde{\mathcal{A}}_{ts}(i,j)=0} S(z_i,z_j) + \sum_{\tilde{\mathcal{A}}_{tt}(i,k)=0} S(z_i,z_k). \end{cases} \quad (10)$$

Assume that the unified graph can be partitioned into several subgraphs, the intuition of category alignment and compactness implies that the similarities of feature representation in the same local subgraph should be larger than that between distinct subgraphs. To utilize this relation knowledge of intra- and inter-domains, we propose a novel subgraph-level multi-sample contrastive loss below,

$$\mathcal{L}_{Contra}^{rw} = -\frac{1}{n_s+n_t}\sum_{i=1}^{n_s+n_t}\log(\frac{\mathcal{S}_{intra}(x_i)}{\mathcal{S}_{inter}(x_i)}). \quad (11)$$

Minimizing $\mathcal{L}_{Contra}^{rw}$ can simultaneously increase total intra-subgraph similarity and decrease total inter-subgraph similarity, which improves the common categories matching across domains and the respective private categories separation. Note that the loss $\mathcal{L}_{Contra}^{rw}$ covers computations involving all the embedded features of two domains, so we employ a hybrid memory bank $\mathcal{Z} = \{z_1^s, ..., z_{n_s}^s, z_1^t, ..., z_{n_t}^t\}$ to store these features. Before every mini-batch training, we simply update the old items in the $\mathcal{Z}$ corresponding to the current mini-batch.

3.4. Energy-based uncertainty modeling for universal classification

Most existing UniDA methods use a softmax-based closed-world classifier to discriminate all samples, and they manually set a threshold for the confidence or entropy score output by the classifier to identify "unknown" samples in the target domain [33]. This strategy would face two thorny troubles: 1) since the model is asked to classify the input into that pre-defined source "known" categories with high probability during training, the class competition nature in softmax causes neural network classifiers to generate overconfident predictions, thus failing to model uncertainty in output probabilities; 2) in the absence of some prior "unknown" samples as supervision, it is difficult to tune an optimal threshold to apply it to all open-set tasks. To address these issues, we introduce an energy-based $|C_s|+1$-dimensional softmax formulation for universal classification, which incorporates the open-world uncertainty modeling as the augmented dimension.

Assume that $\mathcal{G}(\mathcal{F})(x)_k (1 \le k \le |C_s|)$ represents the k-th logit for the k-th category produced by closed classification head layer, then the discrimination probabilities $\mathcal{P}(x) = (\mathcal{P}_1(x), \mathcal{P}_2(x), ..., \mathcal{P}_{|C_s|}(x))$ can be obtained after putting a softmax normalization function on the $|C_s|$ logits,

$$\mathcal{P}_k(x) = \frac{\exp(\mathcal{G}(\mathcal{F})(x)_k)}{\sum_{i=1}^{|C_s|}\exp(\mathcal{G}(\mathcal{F})(x)_i)}. \quad (12)$$

According to [22], $\sum_{i=1}^{|C_s|}\exp(\mathcal{G}(\mathcal{F})(x)_i)$ can be regarded as a kind of negative energy function which encodes the classification uncertainty for out-of-distribution sample detection. The smaller the value of $\sum_{i=1}^{|C_s|}\exp(\mathcal{G}(\mathcal{F})(x)_i)$, the larger the classification uncertainty and the more the sample tends to be the "unknown" sample. To incorporate this characteristic into the classification layer, we extend the closed-world classifier into open-world one by adding an extra dimension representing uncertainty estimation, namely

$$\mathcal{P}_{|C_s|+1}(x) = \frac{\exp(\mathcal{G}(\mathcal{F})(x)_{|C_s|+1})}{\sum_{j=1}^{|C_s|+1}\exp(\mathcal{G}(\mathcal{F})(x)_j)}. \quad (13)$$

Now the next question is how to design a training principle that equips the $\mathcal{P}_{|C_s|+1}(x)$ with the ability for identifying "unknown" samples. Our motivation is to generate synthesized instances by manifold mixup [40] to mimic the emergence of novel categories (see Figure 1(D)). Specifically, we arbitrarily take two samples (x_i^s, x_j^s) from different categories in the source domain, and then mix their embedding features $(\mathcal{F}(x_i^s), \mathcal{F}(x_j^s))$ by linear interpolation,

$$\tilde{z}_\lambda = \lambda\mathcal{F}(x_i^s) + (1-\lambda)\mathcal{F}(x_j^s), y_i \ne y_j, \quad (14)$$

where $\lambda \in [0,1]$ is randomly sampled from Beta distribution (Beta(2, 2)). Since the interpolation regions between two source "known" categories are often places of low-confidence yet high-uncertainty predictions, we can treat $\tilde{z}_\lambda$ as the embedding of novel classes and give it label $|C_s|+1$. The set of these synthesized samples are defined as $\mathcal{D}_s^{mix}$. Naturally, the source samples together with their augmented novel instances are trained with a $|C_s|+1$-way cross-entropy loss, given by

$$\mathcal{L}_{sup} = -\mathbb{E}_{(x,y)\in\mathcal{D}_s\cup\mathcal{D}_s^{mix}}[log(\mathcal{P}_y(x))]. \quad (15)$$

Since the constructed universal classifier has the concept of "unknown", we apply open-set entropy minimization training for the unlabeled target samples to improve the discrimination reliability,

$$\mathcal{L}_{ent} = -\frac{1}{n_t}\sum_{i=1}^{n_t}\sum_{k=1}^{|C_s|+1}\mathcal{P}_k(x_i^t)\log\mathcal{P}_k(x_i^t). \quad (16)$$

3.5. Overall training objective

The overall training loss of our model can be written as a min-max game:

$$\max_{\mathcal{D}}\min_{\mathcal{F},\mathcal{G}}\mathcal{L}_{sup} + \mathcal{L}_{ent} + \alpha\mathcal{L}_{Contra}^{rw} - \alpha\mathcal{L}_{\mathcal{D}}^{rw}, \quad (17)$$

where α is a weight parameter to trade-off between alignment and discrimination. Here we set it as 0.1 to balance each loss component. During training, we split source and target samples into different mini-batches and forward them separately. Note that the synthesized samples are produced within source mini-batches, i.e., once we get a source mini-batch, another order of instances can be derived by shuffling this mini-batch. Then we mask the pairs from the same category and conduct manifold mixup with the pairs from different categories, thus free of extra time cost.

4. Results

4.1. Setup

Dataset. We conduct experiments on three benchmark datasets. **Office** [31] consists of about 4700 images in 31 categories from three domains: Amazon (A), DSLR (D), and Webcam (W). **OfficeHome** [39] is a larger dataset with

Table 1. H-score comparison in the OPDA setting. Some results for previous methods are cited from DCC [18] and OVANet [33].

		Office (10/10/11)							OfficeHome (10/5/50)												VisDA (6/3/3)
Methods	**Type**	A2W	D2W	W2D	A2D	D2A	W2A	Avg	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg	S2R
RTN	C	50.2	54.7	55.2	50.2	47.7	49.3	51.2	38.4	44.7	45.7	42.6	44.1	45.5	42.6	36.8	45.5	44.6	39.8	44.5	42.9	26.0
IWAN	P	50.1	54.1	55.4	50.6	49.7	49.8	51.6	40.5	47.0	47.8	45.0	45.1	47.6	45.8	41.4	47.6	46.3	42.5	46.5	45.3	27.6
OSBP	O	50.2	55.5	57.2	51.1	49.8	50.2	52.3	39.6	45.1	46.2	45.7	45.2	46.8	45.3	40.5	45.8	45.1	41.6	46.9	44.5	27.3
UAN	U	58.6	70.6	71.4	59.7	60.1	60.3	63.5	51.6	51.7	54.3	61.7	57.6	61.9	50.4	47.6	61.5	62.9	52.6	65.2	56.6	30.5
CMU	U	67.3	79.3	80.4	68.1	71.4	72.2	73.1	56.0	56.9	59.2	67.0	64.3	67.8	54.7	51.1	66.4	68.2	57.9	69.7	61.6	34.6
DANCE	U	75.8	90.9	87.1	79.6	82.9	77.6	82.3	61.0	60.4	64.9	65.7	58.8	61.8	73.1	61.2	66.6	67.7	62.4	63.7	63.9	42.8
DCC	U	78.5	79.3	88.6	**88.5**	70.2	75.9	80.2	58.0	54.1	58.0	**74.6**	70.6	77.5	64.3	**73.6**	74.9	**81.0**	**75.1**	80.4	70.2	43.0
OVANet	U	78.4	**95.9**	**95.5**	83.8	80.7	82.7	86.2	63.4	**77.8**	79.7	69.5	70.6	76.4	73.5	61.4	80.6	76.5	64.3	78.9	72.7	53.1
GATE	U	**81.6**	94.8	94.1	87.7	**84.2**	**83.4**	**87.6**	**63.8**	75.9	**81.4**	74.0	**72.1**	**79.8**	**74.7**	70.3	**82.7**	79.1	71.5	**81.7**	**75.6**	**56.4**

Table 2. H-score comparison in the ODA setting. Some results for previous methods are cited from DCC [18] and OVANet [33].

		Office (10/0/11)							OfficeHome (25/0/40)												VisDA (6/0/6)
Methods	**Type**	A2W	A2D	D2W	W2D	D2A	W2A	Avg	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg	S2R
STA	O	75.9	75.0	69.8	75.2	73.2	66.1	72.5	55.8	54.0	68.3	57.4	60.4	66.8	61.9	53.2	69.5	67.1	54.5	64.5	61.1	64.1
OSBP	O	82.7	82.4	97.2	91.1	75.1	73.7	83.7	55.1	65.2	72.9	64.3	64.7	70.6	63.2	53.2	73.9	66.7	54.5	72.3	64.7	52.3
ROS	O	82.1	82.4	96.0	**99.7**	77.9	77.2	85.9	60.1	69.3	**76.5**	58.9	65.2	68.6	60.6	56.3	74.4	68.8	60.4	**75.7**	66.2	66.5
UAN	U	46.8	38.9	68.8	53.0	68.0	54.9	55.1	40.3	41.5	46.1	53.2	48.0	53.7	40.6	39.8	52.5	53.6	43.7	56.9	47.5	51.9
CMU	U	55.7	52.6	75.9	64.7	76.5	65.8	65.2	45.1	48.3	51.7	58.9	55.4	61.2	46.5	43.8	58.0	58.6	50.1	61.8	53.3	54.2
DANCE	U	78.8	84.9	78.8	88.9	79.1	68.3	79.8	61.9	61.3	63.7	64.2	58.6	62.6	**67.4**	61.0	65.5	65.9	61.3	64.2	63.0	67.5
DCC	U	54.8	58.3	89.4	80.9	67.2	85.3	72.6	56.1	67.5	66.7	49.6	66.5	64.0	55.8	53.0	70.5	61.6	57.2	71.9	61.7	59.6
OVANet	U	**88.3**	**90.5**	**98.2**	98.4	**86.7**	**88.3**	**91.7**	58.9	66.0	70.4	62.2	65.7	67.8	60.0	52.6	69.7	68.2	59.1	67.6	64.0	66.1
GATE	U	86.5	88.4	95.0	96.7	84.2	86.1	89.5	**63.8**	**70.5**	75.8	**66.4**	**67.9**	**71.7**	67.3	**61.5**	**76.0**	**70.4**	**61.8**	75.1	**69.1**	**70.8**

15500 images from 65 categories in four domains: Artistic images (A), Clip-Art images (C), Product images (P), and Real-World images (R). **VisDA** [30] is a large-scale challenging dataset with 12 categories, with source domain containing about 150K synthetic images (S) and target domain containing 50K real world images (R). Let $|C|$, $|\bar{C}_s|$ and $|\bar{C}_t|$ denote the number of common categories, source private categories and target private categories, respectively. Following existing studies [32, 33], we show the category split ($|C|/|\bar{C}_s|/|\bar{C}_t|$) of each experimental setting in a corresponding result table. The split details can be seen in the supplementary.

Evaluation protocols. We use the same evaluation metrics as those in the previous studies [9, 18]. In CDA and PDA settings, we calculate the classification accuracy over all target samples. In ODA and OPDA settings, target private samples are grouped into a single "unknown" class. As such, the trade-off between the accuracy of "known" and "unknown" classes is important in evaluating performance. Thus, we use the H-score, i.e. the harmonic mean of the accuracy on common classes and accuracy on the "unknown" class, to evaluate each method. The H-score metric is high only when both the "known" and "unknown" accuracies are high. For all experiments, the averaged results of three runs are reported. Additionally, we assume no prior information about category shift in any of the above DA settings.

Implementation details. Our implementation is based on PyTorch and we conduct all experiments on 8 Tesla V100 GPUs. Following previous UniDA works, the feature extractor is ResNet50 [14] pre-trained on ImageNet [8]. A bottleneck layer with 256 units followed by a classifier and a domain discriminator, is added after the feature extractor. The batch size is set to 36 and the temperature parameter τ is set as 0.05. The nearest neighbor number k is set to 5 for Office and OfficeHome and 10 for VisDA as default. We train our model for 10000 iterations with Nesterov momentum SGD. The initial learning rate is set to 0.001, which is decayed with the same schedule as in previous study [25]. More details are illustrated in the supplementary.

4.2. Comparative results

Comparison baselines. We compare GATE with previous state-of-the-arts in four possible scenarios of UniDA, i.e., CDA (RTN [26], CDAN [25], MDD [19], SRDC [37]), PDA (PADA [2], IWAN [45], ETN [3], BA^{3}US [20]), ODA (OSBP [34], STA [21], ROS [1]) and OPDA (UAN [44], CMU [9], DANCE [32], DCC [18], OVANet [33]). For all cases, each UniDA method is tested without knowing the prior of category shift, and those baselines tailed for each setting are conducted by taking this prior into consideration. We use "C", "P", "O" and "U" to denote the methods designed for CDA, PDA, ODA and UniDA accordingly. Due to a limited space, we put some results in supplementary.

ODA and OPDA settings. From the results in Table 1, GATE achieves a new state-of-the-art on three datasets in the most challenging OPDA setting. With respect to H-score, GATE outperforms the previous state-of-the-art UniDA method OVANet on Office by 2% and on OfficeHome by 3%. On large-scale VisDA dataset, GATE also gives more than 3% improvement compared to all other methods in terms of H-score. Collectively, this evidence shows that GATE gains a better trade-off between common categories classification and private samples identification. For the ODA setting, the H-score comparison results are presented in Table 2. Although slightly inferior to OVANet on Office, our method consistently performs better than all the UniDA baselines on OfficeHome and VisDA datasets, with +3% H-score improvement. Even compared with ROS, a previous state-of-the-art method tailed for the ODA setting, GATE is also superior on all three datasets. Under these two scenarios with "unknown" samples, GATE shows a stronger capability on the separation of common and private categories, which benefits from the global joint local feature alignment paradigm and adaptive energy uncertainty calibration strategy.

Table 3. Accuracy comparison in the PDA setting. Some results for previous methods are cited from DANCE [32] and DCC [18].

		Office (10/21/0)							OfficeHome (25/40/0)													VisDA (6/6/0)
Methods	**Type**	A2W	A2D	D2W	W2D	D2A	W2A	Avg	A2C	A2P	A2R	C2A	C2P	C2R	P2A	P2C	P2R	R2A	R2C	R2P	Avg	S2R
PADA	P	82.2	86.5	92.7	99.3	95.4	**100.0**	92.7	52.0	67.0	78.7	52.2	53.8	59.1	52.6	43.2	78.8	73.7	56.6	77.1	62.1	53.5
ETN	P	94.5	95.0	**100.0**	**100.0**	**96.2**	94.6	96.7	59.2	77.0	79.5	62.9	65.7	75.0	68.3	55.4	84.4	75.7	57.7	84.5	70.5	59.8
BA^{3}US	P	**98.9**	**99.4**	**100.0**	98.7	94.8	95.0	**97.8**	**60.6**	**83.2**	**88.4**	71.8	**72.8**	83.4	**75.5**	61.6	**86.5**	79.3	62.8	**86.1**	**76.0**	54.9
UAN	U	76.8	79.7	93.4	98.3	82.7	83.7	85.8	24.5	35.0	41.5	34.7	32.3	32.7	32.7	21.1	43.0	39.7	26.6	46.0	34.2	39.7
CMU	U	84.2	84.1	97.2	98.8	69.2	66.8	83.4	50.9	74.2	78.4	62.2	64.1	72.5	63.5	47.9	78.3	72.4	54.7	78.9	66.5	65.5
DANCE	U	71.2	77.1	94.6	96.8	83.7	92.6	86.0	53.6	73.2	84.9	70.8	67.3	82.6	70.0	50.9	84.8	77.0	55.9	81.8	71.1	73.7
DCC	U	81.3	87.3	**100.0**	**100.0**	95.4	95.5	93.3	54.2	47.5	57.5	**83.8**	71.6	**86.2**	63.7	**65.0**	75.2	**85.5**	**78.2**	82.6	70.9	72.4
OVANet	U	61.7	69.4	90.2	98.7	61.4	66.4	74.6	34.1	54.6	72.1	42.4	47.3	55.9	38.2	26.2	61.7	56.7	35.8	68.9	49.5	34.3
GATE	U	86.2	89.5	**100.0**	98.6	93.5	94.4	93.7	55.8	75.9	85.3	73.6	70.2	83.0	72.1	59.5	84.7	79.6	63.9	83.8	73.9	**75.6**

Table 4. H-score comparison on the new low-light OPDA benchmark. Source domain: ImageCLEF [27]; Target domain: ExDark [24].

		ExDark without ZeroDCE (8/4/4)					ExDark with ZeroDCE (8/4/4)				
Methods	**Type**	B2E	C2E	I2E	P2E	Avg	B2E	C2E	I2E	P2E	Avg
UAN	U	41.4	50.7	52.3	42.9	46.8	42.8	52.5	54.9	41.9	48.1
CMU	U	52.1	50.4	54.3	46.7	50.9	54.4	54.2	58.6	55.8	55.7
DANCE	U	53.2	60.8	56.6	56.9	56.9	55.4	62.1	58.9	58.4	58.7
DCC	U	50.7	61.3	55.8	54.1	55.5	53.4	61.6	59.5	58.0	58.1
OVANet	U	49.0	50.1	57.5	56.6	53.3	51.4	53.1	60.0	58.7	55.8
GATE	U	**55.3**	**62.8**	**59.4**	**57.9**	**58.8**	**57.2**	**64.7**	**61.5**	**60.6**	**61.0**

CDA and PDA settings. In the PDA setting, the results in Table 3 tell us that GATE outperforms all other baselines including those tailored for PDA on VisDA. For Office and OfficeHome datasets, GATE also gives comparable results to BA^{3}US, which is the state-of-the-art method in PDA. The results in supplementary for CDA setting show that GATE outperforms other state-of-the-art UniDA methods on all three datasets. Even compared to those methods specialized for the CDA setting, GATE also achieves comparable performance to some of them, such as only inferior to SRDC on Office, OfficeHome and VisDA datasets. However, such methods customized for these two settings cannot adapt to situations where "unknown" samples exist, thereby limiting their application in real-world scenarios. It can also be seen that under these two settings OVANet does not perform as well as in ODA and OPDA settings, mainly because it pays more attention to the discovery of private categories and ignores the importance of domain alignment.

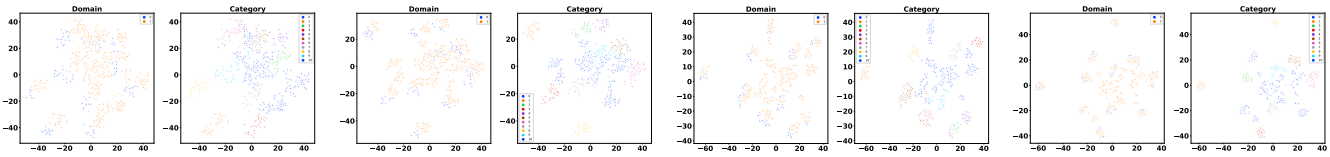
New low-light DA benchmark. High-level vision tasks, such as classification, detection and segmentation, would suffer severe performance degradation under low-light environments [35]. Recently, the low-light DA has attracted much attention of community [41]. Here, we integrate the existing datasets, ImageCLEF [27] and Exclusively Dark (ExDark) [24], to construct a new OPDA benchmark from normal-light to low-light. ImageCLEF consists of 4 domains, i.e., Bing (B), Caltech (C), ImageNet ILSVRC (I) and Pascal VOC (P), and each domain has 600 images with 12 categories. ExDark (E) is a collection of 7363 low-light images in 12 object classes. There are 8 common categories between these two datasets and 4 private categories for each dataset. We use ImageCLEF as the source domain and ExDark as the target domain for low-light transfer learning experiments. The parameter settings are the same as those on Office. To better illustrate the difficulty of low-light DA, we use ZeroDCE [12] to perform low-light enhancement on ExDark, and the enhanced dataset repeats the above experiment. From Table 4, the H-score of each method is improved when ExDark is enhanced. However, GATE consistently outperforms all other baseline methods, whether with or without low-light enhancement. This evidence shows that GATE is better able to cope with low-light conditions.

Feature visualization. We use t-SNE [17] to visualize the learned source and target features with corresponding domain labels and category labels. In this analysis, we use "D2W" in Office to conduct experiments under the ODA setting. As shown in Figure 2a, before adaptation, the common categories do not mix together and most target private samples are attached near the common samples. Using only geometric adversarial learning can reduce the margin between common categories, but they still don't mix well (see Figure 2b). Using only subgraph-level contrastive learning can mix some common categories and separate some target private samples, but there are still category mismatches and negative transfers (see Figure 2c). After applying both of them, we achieve the common categories mixing well and separate most target private samples from the common samples (see Figure 2d), demonstrating that our global joint local strategy is very effective for feature alignment.

4.3. Ablation studies

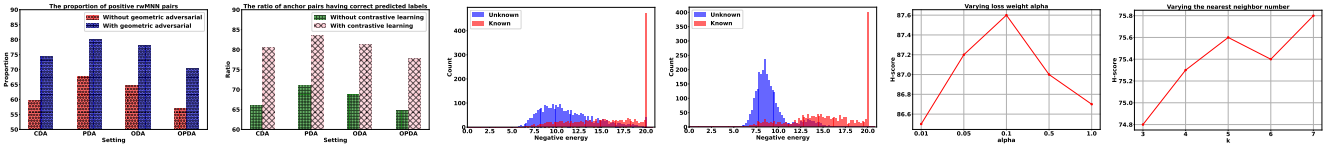
Effect of geometric adversarial learning. To verify the necessity of the geometric adversarial learning, we remove $\mathcal{L}_{\mathcal{D}}^{rw}$ in the overall loss and conduct control experiments in four settings on OfficeHome and VisDA datasets. From the results in Table 5, the performance would be degraded in all settings without geometric adversarial learning. This is mainly because, in the early stages of training, global feature discrepancy would cause the accuracy of cross-domain anchor pairs to be unguaranteed. Specifically, for all settings on VisDA, we observe that the proportion of true positive rwMNN pairs is sharply decreased after removing the geometric adversarial learning (see Figure 3a).

Effect of subgraph-level contrastive learning. To evaluate the contribution of $\mathcal{L}_{Contra}^{rw}$, we train the model without it and present the results in Table 5. Removing contrastive learning on the rwMNN anchor pairs would



(a) Before adaptation (b) Only adversarial (c) Only contrastive (d) After adaptation

Figure 2. Feature visualization on "D2W" in Office under the ODA setting. For domain, blue represents the source domain and orange refers to the target domain. For category, blue plots are "unknown" samples, others are "known" samples.



(a) Geometric adversarial (b) Contrastive learning (c) Without calibration (d) With calibration (e) Loss weight alpha (f) Neighbor number k

Figure 3. Case study. (A) The proportion of true positive rwMNN pairs in VisDA; (B) The ratio of anchor pairs having correct predicted labels in VisDA; (C & D) Histogram of energy scores on "R2P" of OfficeHome in OPDA setting; (E) Varying the loss weight α; (F) Varying the nearest neighbor number k.

severely hurt performance in all settings. Further tested results in Figure 3b tell us that without it, the ratio of anchor pairs having correct predicted labels decreases more than 15% in all settings on VisDA. It is reasonable since geometric adversarial learning cannot achieve fine-grained class-specific alignment, while subgraph-level contrastive learning can pull those rwMNN pairs closer to each other and push away dissimilar samples. Therefore, global adversarial learning and local contrast learning complement each other and jointly improve the effect of domain alignment.

Effect of energy-based universal classifier. To show the effectiveness of energy-based universal classifier trained by manifold mixup, we first train the model without $\mathcal{L}_{ent}$ and then use the energy score $\sum_{i=1}^{|C_s|} \exp(\mathcal{G}(\mathcal{F})(x))$ with a threshold δ to identify "unknown" samples. Suppose μ_s and σ_s are the mean and standard deviation values of set $\{\sum_{i=1}^{|C_s|} \exp(\mathcal{G}(\mathcal{F})(x_j^s))\}_{j=1}^{n_s}$, we decide the threshold as $\delta = \mu_s - 2 \times \sigma_s$ and divide those target samples with energy scores lower than δ as "unknown" samples. Table 5 shows that our universal classifier can obtain higher H-score in the ODA and OPDA settings. This is mainly due to the fact that the manifold mixup can tighten the classification area of the "known" class, while adaptively calibrating the distribution gap of the energy score between the "known" and the "unknown" (see an example in Figure 3c & 3d).

Table 5. Various ablation studies on the OfficeHome and VisDA dataset in all four settings.

Methods	OfficeHome				VisDA			
	CDA	PDA	ODA	OPDA	CDA	PDA	ODA	OPDA
GATE w/o $\mathcal{L}_{\mathcal{D}}^{rw}$	67.4	70.6	63.8	68.1	70.3	72.6	63.3	47.7
GATE w/o $\mathcal{L}_{Contra}^{rw}$	65.9	67.3	57.4	64.3	67.7	68.5	57.4	41.6
GATE w/o Universal classifier	70.0	73.5	65.2	70.5	74.5	75.1	66.7	52.9
GATE w/o Anchor expansion	68.8	72.4	66.5	73.0	72.1	73.9	67.1	53.3
GATE with larger k in MNN	69.5	72.8	67.0	73.8	73.4	74.8	68.2	54.4
GATE w/o SNN score	69.7	73.2	67.9	74.7	73.6	74.4	69.1	54.9
GATE (full)	70.2	73.9	69.1	75.6	74.8	75.6	70.8	56.4

Effect of random walk-based anchor expansion. To validate the effectiveness of anchor pair expansion by within-domain random walk, we replace rwMNN with MNN in adversarial and contrastive learning. The results in Table 5 show that the performance of rwMNN is better than that of MNN. Since the anchor pair number in rwMNN is nearly the square of that in MNN, we also show that a larger k in MNN would lead to worse performance. We chose k in MNN as 25 on OfficeHome and 100 on VisDA, and their results in Table 5 are really lower than those in rwMNN. This evidence demonstrates that rwMNN is more suitable to uncover the shared distribution of common categories.

Effect of SNN-based anchor score mechanism. To better understand the importance of the SNN-based anchor score, we set the affinity value of all rwMNN pairs as equal to 1. The results in Table 5 tell us that using expanded anchors pairs with consistent affinity values cannot obtain the same satisfactory performance as with the SNN-based score mechanism. This evidence illustrates that simply expanding MNN pairs would bring the risk of more noisy supervision, while SNN-based scores could reduce this negative impact and enhance the algorithm's robustness.

Hyperparameter sensitivity. To show the sensitivity of GATE to the loss weight α, we conducted control experiments on Office under the OPDA setting and presented the results in Figure 3e. Within a wide range of $\alpha \in [0.01, 1.0]$, the performance changes very little, showing that GATE is robust to the selection of α. We also analyze the behavior of GATE when changing the nearest neighbor number k on OfficeHome in the OPDA setting. As shown in Figure 3f, the H-score only varies slightly with $k \in [3, 7]$, validating that GATE is stable to the choices of k.

5. Conclusion

In this paper, we propose a novel UniDA framework called GATE from the perspective of anchor mining and uncertainty modeling. It performs feature alignment via global distribution-level geometric adversarial learning joint local subgraph-level contrastive learning. We also introduce an energy-based incremental classifier for "unknown" sample detection. A thorough evaluation shows the superior performance of GATE, compared to previous state-of-the-arts.

References

[1] Silvia Bucci, Mohammad Reza Loghmani, and Tatiana Tommasi. On the effectiveness of image rotation for open set domain adaptation. *In European Conference on Computer Vision*, pages 422–438, 2020. 6

[2] Zhangjie Cao, Lijia Ma, Mingsheng Long, and Jianmin Wang. Partial adversarial domain adaptation. *In Proceedings of the European Conference on Computer Vision*, pages 135–150, 2018. 1, 6

[3] Zhangjie Cao, Kaichao You, Mingsheng Long, Jianmin Wang, and Qiang Yang. Learning to transfer examples for partial domain adaptation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2985–2994, 2019. 6

[4] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *In Thirty-fourth Conference on Neural Information Processing Systems*, 2020. 2

[5] Lawrence. Cayton. Algorithms for manifold learning. *Univ. of California at San Diego Tech*, 1:1–17, 2005. 3

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *In International conference on machine learning*, pages 1597–1607, 2020. 2

[7] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 15750–15758, 2021. 2

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. *In 2009 IEEE conference on computer vision and pattern recognition*, pages 248–255, 2009. 6

[9] Bo Fu, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Learning to detect open classes for universal domain adaptation. *In European Conference on Computer Vision*, pages 567–583, 2020. 2, 6

[10] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. *In International conference on machine learning*, pages 1180–1189, 2015. 1

[11] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. *arXiv preprint arXiv:2006.07733*, 2020. 2

[12] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin. Cong. Zero-reference deep curve estimation for low-light image enhancement. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020. 7

[13] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020. 2

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6

[15] Taotao Jing, Hongfu Liu, and Zhengming Ding. Towards novel target discovery through open-set domain adaptation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9322–9331, 2021. 1

[16] Jogendra Nath Kundu, Naveen Venkat, and R. Venkatesh Babu. Universal source-free domain adaptation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4544–4553, 2020. 1

[17] Van der Maaten Laurens and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 7

[18] Guangrui Li, Guoliang Kang, Yi Zhu, Yunchao Wei, and Yi Yang. Domain consensus clustering for universal domain adaptation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9757–9766, 2021. 2, 6, 7

[19] Jingjing Li, Erpeng Chen, Zhengming Ding, Lei Zhu, Ke Lu, and Hengtao Shen. Maximum density divergence for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 2020. 6

[20] Jian Liang, Yunbo Wang, Dapeng Hu, Ran He, and Jiashi Feng. A balanced and uncertainty-aware approach for partial domain adaptation. *In Computer Vision–ECCV 2020: 16th European Conference*, pages 123–140, 2020. 6

[21] Hong Liu, Zhangjie Cao, Mingsheng Long, Jianmin Wang Wang, and Qiang Yang. Separate to adapt: Open set domain adaptation via progressive separation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2927–2936, 2019. 6

[22] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems*, 33, 2020. 5

[23] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Zhaoyu Wang, Li Mian, Jing Zhang, and Jie Tang. Self-supervised learning: Generative or contrastive. *arXiv preprint arXiv:2006.08218*, 2020. 2

[24] Yuen Peng Loh and Chee Seng. Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019. 7

[25] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I. Jordan. Conditional adversarial domain adaptation. *In Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 1647–1657, 2018. 6

[26] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I. Jordan. Unsupervised domain adaptation with residual transfer networks. *Advances in Neural Information Processing Systems*, pages 136–144, 2016. 6

[27] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I. Jordan. Deep transfer learning with joint adaptation networks. *In International conference on machine learning*, pages 2208–2217, 2017. 7

[28] Jaemin Na, Heechul Jung, Hyung Jin Chang, and Wonjun Hwang. Fixbi: Bridging domain spaces for unsupervised

domain adaptation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1094–1103, 2021. 1

[29] Pau Panareda and Juergen Gall. Open set domain adaptation. *In Proceedings of the IEEE International Conference on Computer Vision*, pages 754–763, 2017. 1

[30] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge. *arXiv preprint arXiv:1710.06924*, 2017. 6

[31] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. *In European conference on computer vision*, pages 213–226, 2010. 5

[32] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. Universal domain adaptation through self supervision. *Advances in Neural Information Processing Systems*, 33, 2020. 1, 2, 6, 7

[33] Kuniaki Saito and Kate Saenko. Ovanet: One-vs-all network for universal domain adaptation. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9000–9009, 2021. 2, 5, 6

[34] Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. *In Proceedings of the European Conference on Computer Vision*, pages 153–168, 2018. 6

[35] Yukihiro Sasagawa and Hajime. Nagahara. Yolo in the dark-domain adaptation method for merging multiple models. *In European Conference on Computer Vision*, pages 345–359, 2020. 7

[36] Astuti Sharma, Tarun Kalluri, and Chandraker Manmohan. Instance level affinity-based transfer for unsupervised domain adaptation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5361–5371, 2021. 4

[37] Hui Tang, Ke Chen, and Kui Jia. Unsupervised domain adaptation via structurally regularized deep clustering. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8725–8735, 2020. 6

[38] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017. 1

[39] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5018–5027, 2017. 5

[40] Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Yoshua . Bengio. Manifold mixup: Better representations by interpolating hidden states. *In International Conference on Machine Learning*, pages 6438–6447, 2019. 5

[41] Wenjing Wang, Wenhan Yang Yang, and Jiaying. Liu. Hlaface: Joint high-low adaptation for low light face detection. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 16195–16204, 2021. 7

[42] Xudong Wang, Ziwei Liu, and Stella X. Yu. Unsupervised feature learning by cross-level instance-group discrimination. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12586–12595, 2021. 2

[43] Yezhen Wang, Bo Li, Tong Che, Kaiyang Zhou, Ziwei Liu, and Li Dongsheng. Energy-based open-world uncertainty modeling for confidence calibration. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9302–9311, 2021. 2

[44] Kaichao You, Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I. Jordan. Universal domain adaptation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2720–2729, 2019. 1, 2, 6

[45] Jing Zhang, Zewei Ding, Wanqing Li, and Philip Ogunbona. Importance weighted adversarial nets for partial domain adaptation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8156–8164, 2018. 6

[46] Zhun Zhong, Enrico Fini, Subhankar Roy, Zhiming Luo, Elisa Ricci, and Nicu Sebe. Neighborhood contrastive learning for novel class discovery. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10867–10875, 2021. 2

[47] Da-Wei Zhou, Han-Jia Ye, and De-Chuan Zhan. Learning placeholders for open-set recognition. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2021. 2