# Batch Singular Value Polarization and Weighted Semantic Augmentation for Universal Domain Adaptation

**Ziqi Wang** [1] **Wei Wang** [1 *] **Chao Huang** [1 *] **Jie Wen** [2] **Cong Wang** [1]

## Abstract

As a more challenging domain adaptation setting, universal domain adaptation (UniDA) introduces category shift on top of domain shift, which needs to identify unknown category in the target domain and avoid misclassifying target samples into source private categories. To this end, we propose a novel UniDA approach named **B**atch **S**ingular value **P**olarization and **W**eighted **S**emantic **A**ugmentation (BSP-WSA). Specifically, we adopt an adversarial classifier to identify the target unknown category and align feature distributions between the two domains. Then, we propose to perform SVD on the classifier's outputs to maximize larger singular values while minimizing those smaller ones, which could prevent target samples from being wrongly assigned to source private classes. To better bridge the domain gap, we propose a weighted semantic augmentation approach for UniDA to generate data on common categories between the two domains. Extensive experiments on three benchmarks demonstrate that BSP-WSA could outperform existing state-of-the-art UniDA approaches.

## 1. Introduction

The tremendous successes of deep learning are largely attributed to its reliance on diverse and massive annotated training data (Huang et al., 2021; 2022a;b; 2023; 2024). However, meeting such requirements is laborious and resource-intensive, especially in the military and medical fields (Chen et al., 2024). Intuitively, one can leverage a well-labeled source domain to assist in annotating a target domain of interest. However, in practice, the domain shift problem may exist between the two domains, which will degrade annotation performance in the target. To address this problem, domain adaptation (DA) as an effective technology allows for domain shift and aims to improve the generalization performance of a model trained on the source to the target (Wang et al., 2020; 2021; 2022a; 2023a;b).

However, DA assumes that the source and target domains share the same label space, which does not align with real-world scenarios. In recent years, researchers have relaxed the assumption of the same label space and proposed various new DA settings, such as partial DA (Cao et al., 2018; Zhang et al., 2018; Cao et al., 2019), open-set DA (Saito et al., 2018; Liu et al., 2019; Luo et al., 2020), universal DA (UniDA) (You et al., 2019b; Lifshitz & Wolf, 2021; Kundu et al., 2022). Among these scenarios, UniDA is the most challenging one, as it introduces category shift on top of domain shift. UniDA assumes that the two domains not only have common categories but also have their own private categories. Moreover, there is no prior knowledge about which classes are common and which are private.

For a UniDA task, it is crucial to classify target domain samples accurately. Samples that belong to private classes of the target domain should be identified as unknown, while those from classes common to both domains should be classified into the corresponding classes of the source domain. Since the source domain also contains private classes, it is essential to prevent target samples from being misclassified into source private classes, considering distribution differences between the two domains. Conversely, when we attempt to alleviate the domain shift using existing DA techniques, the lack of prior knowledge about the relationship between the label spaces of the two domains (category shift) makes it difficult to reduce the distribution discrepancy between common categories effectively.

To address the aforementioned problems in UniDA, we propose a **B**atch **S**ingular value **P**olarization and **W**eighted **S**emantic **A**ugmentation approach (BSP-WSA). Inspired by Saito et al., we employ an adversarial classifier to simultaneously detect samples of unknown category in the target domain and achieve distribution alignment between the two

*Corresponding Author. [1]School of Cyber Science and Technology, Shenzhen Campus of Sun Yat-sen University, Shenzhen, China. [2]School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China. Ziqi Wang <hhdxdwzq@gmail.com>. Correspondence to: Wei Wang <wangwei29@mail.sysu.edu.cn>, Chao Huang <huangch253@mail.sysu.edu.cn>.

*Proceedings of the 41st International Conference on Machine Learning*, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

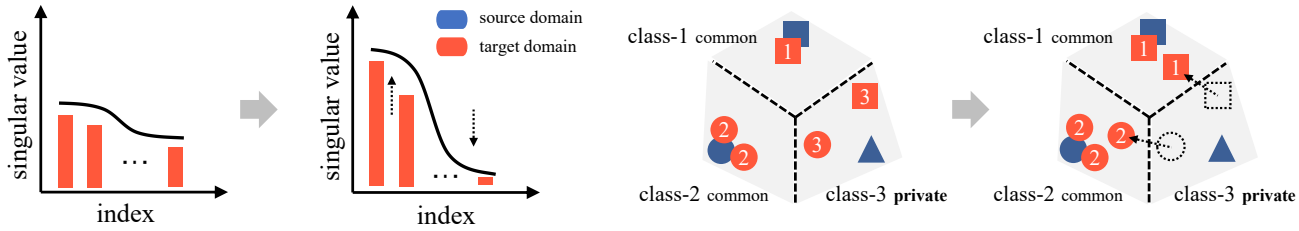



*Figure 1.* Effect of batch singular value polarization. The left part is the singular value fluctuation curve (the index is sorted in descending order of singular values), and the right part is the feature distribution of the source and target domains. Firstly, we perform SVD on the predicted label matrix of the target domain. Then, we maximize the larger singular values while minimizing the smaller ones to steep the singular value curve and reduce the diversity of predicted labels. As a result, we could prevent target domain samples from being wrongly assigned to private categories in the source domain.

domains. Specifically, the classifier aims to categorize as many target samples as possible into the unknown category. At the same time, the feature generator is employed to learn features that confuse the classifier, causing it to classify target samples into known categories. This approach unifies the sample discovery of unknown category and distribution alignment into a unified adversarial learning process. However, for UniDA, directly using this strategy would classify some target samples into the source private categories, compromising the correct distribution alignment process for the common categories.

In DA, designing loss functions based on SVD has gained widespread research due to its simplicity and effectiveness (Chen et al., 2019a;b; Cui et al., 2020). Cui et al. indicate that maximizing the sum of singular values of the label matrix could enhance the class diversity. In contrast, UniDA requires reducing the class diversity or preventing the possibility of target samples being classified into source private classes. Inspired by this, we observe that by maximizing the larger singular values but minimizing the smaller singular values (singular value polarization), the class diversity could be reduced, or the number of samples classified as source private classes could be reduced. As shown in the left part of Fig. 1, the singular value curve becomes steeper through the process of singular value polarization, indicating that the difference between larger and smaller values is increased. As depicted in the right part of Fig. 1, the target samples misclassified into source private classes are corrected to the common classes.

Li et al. and Xie et al. utilize semantic data augmentation (Wang et al., 2022b) to bridge the distribution gap between the two domains, which can be regarded as a supplement to feature distribution alignment strategies (Long et al., 2018; Chen et al., 2019b; Kang et al., 2019). Inspired by this, we introduce a weighted semantic augmentation approach specifically designed for UniDA, which utilizes the statistical information of the two domains to generate samples for each common category with larger weights. Our main contributions can be summarized as follows,

- We propose a batch singular value polarization method to prevent target samples from being assigned to source private classes, and achieve distribution alignment between common classes of the two domains.
- We propose a weighted semantic augmentation method to bridge the gap between two different domains further, producing samples for each common category with larger weights.
- Extensive experiments on three benchmarks with UniDA setting could verify the effectiveness of BSP-WSA compared to existing state-of-the-art approaches.

## 2. Related Work

**Universal Domain Adaptation.** Considering both the domain and category shifts, UniDA aims to classify target samples belonging to private classes as unknown, and those from common classes as corresponding classes in the source domain. Therefore, the most crucial problem in UniDA is distinguishing between common and private classes, thereby aligning distributions of common categories between the two domains and detecting unknown class in the target domain. To this end, You et al. construct a sample-level transferability criterion based on domain similarity and prediction uncertainty (entropy). Saito et al. only utilize the entropy of prediction results to characterize the probability of a sample belonging to known or unknown classes. Fu et al. introduce a novel transferability measure estimated by a mixture of uncertainty quantities.

As the methods mentioned above need to set a threshold manually, Saito & Saenko propose to learn the threshold using source samples and adapt it to the target domain. Chen et al. learn a category-aware heterogeneous threshold vector to reject diverse unknown samples. Chen et al. introduce a bimodality hypothesis for the maximum discriminative

probability distribution to detect the possible target private samples, and present a data-based statistical approach to separate the common and private categories. Chang et al. design an OT-based partial alignment with adaptive filling to detect common classes. Chen et al. try to minimize the open-set entropy to learn the unknown threshold adaptively.

In contrast, we employ singular value polarization to distinguish between shared and private classes more effectively. Additionally, we introduce weighted semantic augmentation to align the distributions of common classes better.

**SVD-based Domain Adaptation.** In recent years, designing regularization losses through SVD has received extensive attention in many computer vision tasks. During the SVD process on the feature matrix, Chen et al. observe that larger singular values correspond to transferability, while smaller ones correspond to discriminability. Therefore, they propose to penalize the larger singular values to improve feature discriminability in adversarial domain adaptation networks. Moreover, Chen et al. also propose to penalize the smaller singular values to prompt feature transferability. Shi et al. observe that minimizing the sum of singular values of the feature matrix could remove environmental features and prompt domain-invariant features accordingly. Cui et al. observe that maximizing the sum of singular values of the label matrix could enhance the certainty and diversity of predicted results. Similarly, this paper also performs SVD on the predicted label matrix. In contrast, we aim to maximize the larger singular values while minimizing those smaller ones to deal with UniDA problem.

## 3. Proposed Approach

### 3.1. Preliminaries

In UniDA, there exist a labeled source domain $\mathbf{D}^{s} = \{(\mathbf{x}_i^{s}, \mathbf{y}_i^{s})\}_{i=1}^{n^{s}}$ and an unlabeled target domain $\mathbf{D}^{t} = \{(\mathbf{x}_j^{t})\}_{j=1}^{n^{t}}$. Here, $\mathbf{y}_i^{s}$ represents the label of sample $\mathbf{x}_i^{s}$, and $n^{s}$, $n^{t}$ denote the sample numbers of source and target domains, respectively. UniDA aims to generalize the model trained on the source domain to the target domain with problems of domain shift ($\mathbf{P}^{s} \neq \mathbf{P}^{t}$) and category shift ($\mathbf{Y}^{s} \neq \mathbf{Y}^{t}$), where $\mathbf{y}^{s} \subset \mathbf{Y}^{s}$ and $\mathbf{y}^{t} \subset \mathbf{Y}^{t}$. We denote the common label space between the two domains as $\widetilde{\mathbf{Y}} = \mathbf{Y}^{s} \cap \mathbf{Y}^{t}$, while private ones as $\overline{\mathbf{Y}}^{s} = \mathbf{Y}^{s} \setminus \widetilde{\mathbf{Y}}$ and $\overline{\mathbf{Y}}^{t} = \mathbf{Y}^{t} \setminus \widetilde{\mathbf{Y}}$. Notably, we regard $\overline{\mathbf{Y}}^{t}$ as unknown class, while $\mathbf{Y}^{s}$ as known class. The category numbers in $\mathbf{Y}^{s}$, $\overline{\mathbf{Y}}^{s}$, and $\widetilde{\mathbf{Y}}$ are $C^{s} = |\mathbf{Y}^{s}|$, $\overline{C}^{s} = |\overline{\mathbf{Y}}^{s}|$, and $\widetilde{C} = |\widetilde{\mathbf{Y}}|$, respectively.

### 3.2. Proposed BSP-WSA

The whole pipeline of our proposed BSP-WSA is depicted in Fig. 2, and we will provide detailed explanations of various modules in the following subsections.

### 3.3. Adversarial Classifier

Inspired by an open-set DA study (Saito et al., 2018), we devise an adversarial classifier with an output dimensionality of C+1. Specifically, the classifier aims to categorize target samples into unknown class, while the feature generator endeavors to prevent the classifier from assigning target samples to unknown class. The corresponding loss for this adversarial learning process can be defined as follows,

$$\begin{aligned} &\arg\min_{\mathcal{C}} \mathcal{L}_{\text{adv}}\Big(\mathcal{C}\big(\mathcal{F}(\mathbf{D}^{t})\big)\Big), \\ &\arg\max_{\mathcal{F}} \mathcal{L}_{\text{adv}}\Big(\mathcal{C}\big(\mathcal{F}(\mathbf{D}^{t})\big)\Big), \end{aligned} \tag{1}$$

$$\begin{aligned} \mathcal{L}_{\text{adv}}(\hat{\mathbf{y}}_i^{t}) = &-\theta \log\Big(\hat{y}^{t}_{(i,C^{s}+1)}\Big) \\ &- (1-\theta)\log\Big(1-\hat{y}^{t}_{(i,C^{s}+1)}\Big). \end{aligned} \tag{2}$$

Optimizing the first term in Eq. (1) encourages the classifier to categorize target samples into unknown class, while optimizing the second term enables the feature generator to hinder the classifier from assigning target samples to unknown class. In Eq. (2), we calculate the adversarial loss using only the probability of predicted target samples belonging to the unknown class, denoted as $\hat{y}_{(i,C^{s}+1)}$. Notably, $\theta \in [0,1]$ controls the strength of the classifier in pushing target samples to the unknown class, and a larger $\theta$ will bring more target samples into the unknown class. Through the adversarial learning process, we are not only able to identify unknown class in the target domain, but also achieve alignment of feature distributions between the two domains.

### 3.4. Batch Singular Value Polarization

In the adversarial classifier, as the feature extractor reduces the distribution disparity between the two domains, more target samples will closely resemble the source feature distribution. However, this may cause target samples to be misclassified into source private categories, which is called error-t. Specifically, when the probability of a target sample being classified as unknown decreases, the probabilities of being classified into all source categories will increase. Since target samples cannot belong to any source private categories, we only want the probabilities to increase for those common categories. This could promote feature alignment for common classes during the adversarial classifier process, thereby enhancing the performance of UniDA. However, we do not know which categories are source private.

Cui et al. have shown that maximizing the sum of singular values on classifier outputs can enhance the class diversity of prediction results, addressing the long-tailed distribution problem. Inspired by this, in contrast, we aim to diminish the class diversity of prediction results for the given issue in UniDA discussed above. Specifically, we intend to prevent

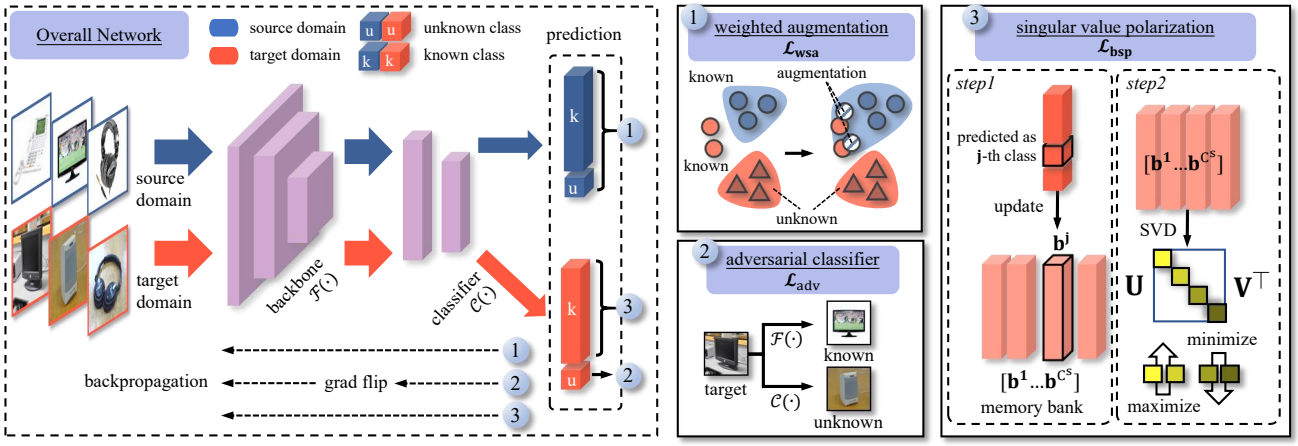


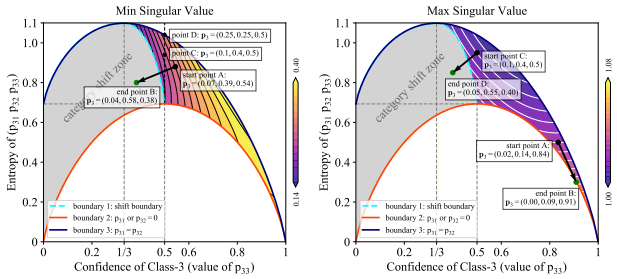

*Figure 2.* The whole pipeline of our proposed BSP-WSA, which could be decomposed into three parts: **i)** identifying unknown class using adversarial classifier; **ii)** preventing target samples from being mistakenly assigned to source private classes through singular value polarization; **iii)** closing the domain gap using weighted semantic augmentation. Specifically, an input sample $\mathbf{x}_i$ is first passed through $\mathcal{F}(\cdot)$ to extract feature of $\mathbf{z}_i = \mathcal{F}(\mathbf{x}_i) \in \mathbb{R}^d$. Afterward, $\mathbf{z}_i$ goes through $\mathcal{C}(\cdot)$ and obtains the predicted label $\hat{\mathbf{y}}_i = \mathcal{C}(\mathbf{z}_i) \in \mathbb{R}^{C^s+1}$, where an extra component denotes probability belonging to the unknown class. Finally, the predicted labels from one batch are fed into three losses: adversarial loss, batch singular value polarization loss, and weighted semantic augmentation loss.

the classifier from assigning target samples to source private classes. As we have observed that the likelihood of target samples being assigned to source private classes is lower compared to common classes, it corresponds to the "tail" of the long-tailed distribution. Building upon this observation, we propose a batch singular value polarization approach tailored for UniDA problem, which maximizes the larger singular values while minimizing the smaller ones.

To better understand our motivation, as depicted in Eq. 3, we consider a prediction matrix $\mathbf{P}$ with three samples and three classes (two shared classes, one private class), where the row represents a sample. In Fig. 3, we depict the changes in the entropy and confidence score for the predicted label of the third sample ($\mathbf{p}_3$), and their impact on the smallest and largest singular values of matrix $\mathbf{P}$. Notably, the confidence score is the probability of the third sample belonging to the third class, ranging between 0 and 1. Then, we have the following observations according to Fig. 3.

$$\mathbf{P} = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.2 \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{matrix} \leftarrow \mathbf{p}_1 \\ \leftarrow \mathbf{p}_2, \\ \leftarrow \mathbf{p}_3 \end{matrix} \quad \mathbf{P} = \mathbf{U}\mathbf{\Sigma}_\mathbf{P}\mathbf{V}^\top, \tag{3}$$
$$\text{entropy} = -\sum_{i=1,2,3} p_{3i} \log p_{3i}.$$

**Observation 1.** Fu et al. suggest that when the confidence corresponding to the predicted category of a target sample is lower, and the label entropy is higher, the sample is more likely to be an error-t. However, this intuition is not entirely accurate. As shown in Fig. 3 (a), for points C (0.1, 0.4, 0.5) and D (0.25, 0.25, 0.5), although both of them have a confidence score of 0.5 for being classified into the third




(a) The smallest singular value. (b) The largest singular value.

*Figure 3.* The changes of the smallest (largest) singular value with confidence&entropy of the sample $\mathbf{p}_3$. The sickle-like area between boundary 2 and boundary 3 represents the entire possible value range. This region consists of two parts: the right-colored area is the current region where the samples are labeled as 3. The left gray area is the category shift zone, where when a sample crosses boundary 1 and enters this region, it changes its category from the 3-th class to the other two classes.

category, due to the bimodal distribution in point C, the point C with lower entropy is more likely to be an error-t (Chen et al., 2022b). This paper suggests that when a point is closer to boundary 1, it is more likely to be an error-t. This is because the closer a point is to boundary 1, the sample label is more likely to undergo category shift. From Fig. 3 (a), it can be observed that point C is indeed closer to boundary 1 compared to point D. Therefore, if we need to determine whether a category in the source domain is a private class, we only need to calculate the average of labels for all target domain samples assigned to that class. Then,

we can use this average label to assess whether it is a private class based on how close it is to boundary 1.

**Observation 2.** According to the density of the contour lines in the colored region of Fig. 3 (a), it can be observed that when the confidence score $p_{33}$ for the third sample belonging to the third class is high, the change gradient of the smallest singular value ($\sigma^s$) is small. However, when the sample is close to boundary 1, the change gradient is large. Therefore, the impact of minimizing $\sigma^s$ on error-t samples is greater than on non-error-t samples. This is because smallest singular values often correspond to noisy patterns in a matrix (Gu et al., 2014; Xu et al., 2017), and error-t samples could be considered as noisy pattern due to their higher uncertainty and lower quantity.

**Observation 3.** When the third sample starts at point A (0.07, 0.39, 0.54) near boundary 1 and is moved towards the low-$\sigma^s$ region, the confidence of this sample belonging to the third class will decrease and the entropy will decrease. As such, it may eventually reach the endpoint B. At this point, the sample has been pushed out of the third class and reclassified into other categories. Notably, it will strive to push it towards higher-confidence category (in the direction of increasing confidence and decreasing entropy).

**Observation 4.** In Fig. 3 (b), our goal is to increase the largest singular value, thus moving towards the yellow region. Similarly, when maximizing the largest singular value, it will push error-t samples towards category shift from C to D. However, for non-error-t samples, it will push them towards the lower right corner (in the direction of increasing confidence and decreasing entropy). This is because for samples with a high probability of belonging to the common class, we encourage pushing them towards common class.

Based on the above analysis, we can prevent target samples from being misclassified into source private classes, by minimizing smaller singular values but maximizing larger singular values. Due to the lack of global context information in batch sampling, the computed singular values often struggle to capture information about common and private classes. In light of this, we employ a sliding window updated memory bank to smooth the prediction matrix of the current batch. We denote the predicted label matrix for a batch as $\hat{\mathbf{Y}} \in \mathbb{R}^{C^s \times n^b}$ where we discard the last component that is the probability belonging to unknown class. Subsequently, for each class, we utilize the mean of all label vectors belonging to that class within this batch as the label representation, as depicted by the following equation,

$$\bar{\mathbf{y}}^j = \frac{1}{|\mathbf{S}^j|} \sum \mathbf{S}^j, \tag{4}$$

where $\mathbf{S}^j$ is the set of sample labels that are predicted as j-th class. Notably, we ignore those classes with $|\mathbf{S}^j| = 0$. Then, we utilize $\{\bar{\mathbf{y}}^j\}_{j=1}^{C^s}$ to update memory bank $\{\mathbf{b}^j\}_{j=1}^{C^s}$.

$$\mathbf{b}^j \leftarrow \lambda \cdot \bar{\mathbf{y}}^j + \mathbf{b}^j, \ |\mathbf{S}^j| \neq 0, \tag{5}$$

where $\lambda$ represents the update ratio. Then, we conduct SVD on $\{\mathbf{b}^j\}_{j=1}^{C^s}$, and our proposed batch singular value polarization could be formulated as follows,

$$\begin{aligned} &\mathbf{B} = [\mathbf{b}^1, \ \ldots, \ \mathbf{b}^{C^s}], \ \mathbf{B} = \mathbf{U}\mathbf{\Sigma}_\mathbf{B}\mathbf{V}^\top, \\ &\arg\min_{\mathcal{C},\mathcal{F}} \mathcal{L}_{bsp}(\mathbf{B}) = \sum_{m=1}^{r} \sigma_m^s - \sigma_m^l, \end{aligned} \tag{6}$$

where $\sigma_m^s$ and $\sigma_m^l$ refer to the m-th smallest and m-th largest singular values in $\mathbf{\Sigma}_\mathbf{B}$, respectively. r controls the degree of polarization, with higher values leading to increased polarization and reduced diversity of predicted results.

### 3.5. Weighted Semantic Augmentation

Li et al. introduce a semantic augmentation strategy that implicitly generates augmented samples between the source and target domains, which could bridge the domain gap effectively and could be regarded as a supplement to existing feature distribution alignment strategies. Specifically, this involves estimating the mean and variance information for each class in both the source and target domains. Subsequently, using the calculated mean, a semantic direction for data augmentation from the source to the target domain is determined, and the amplitude of augmentation is determined based on the variance of the target domain. Consequently, for each sample in the source domain, an infinite number of augmented samples can be obtained using this augmentation approach, and these augmented samples maintain the same labels as the corresponding source samples. Based on theoretical analysis and derivation, Li et al. integrate these augmented samples along with their corresponding labels into the cross-entropy loss of the source domain. This integration leads to a concise loss function. However, this method equally generates samples for each class in the source domain including the private classes. As a result, employing this method directly in UniDA could lead to significant adverse effects. To address this problem, we introduce a weighted semantic augmentation tailored to UniDA, which is formally defined as below,

$$\begin{aligned} &\mathcal{L}_{wsa} = -\frac{1}{n^s} \sum_{i=1}^{n^s} \log \frac{e^{z_{i,g}^s}}{\sum_{j=1}^{C^s} e^{z_{i,j}^s}}, \\ &z_{i,j}^s = \hat{y}_{i,j}^s + \gamma_j (\mathbf{w}_j^\top - \mathbf{w}_g^\top) \mathbf{\Delta} \mu^g + \frac{\rho_{i,j}^s}{2}, \\ &\rho_{i,j}^s = \gamma_j (\mathbf{w}_j^\top - \mathbf{w}_g^\top) \mathbf{\Sigma}_g^t (\mathbf{w}_j - \mathbf{w}_g), \\ &\gamma_j = \frac{\hat{n}_j^t / n_j^s}{\sum_{k=1}^{C^s} (\hat{n}_k^t / n_k^s)}. \end{aligned} \tag{7}$$

where $\hat{y}^{s}_{i,j}$ represents the probability that the i-th sample in the source domain is predicted as the j-th class. 'g' represents the ground-truth. $\mathbf{W} = [\mathbf{w}_1^\top, \cdots, \mathbf{w}_{C^s}^\top] \in \mathbb{R}^{C^s \times K}$ represents the weight matrix in the last fully connected layer of the network, where K is the number of feature embedding dimension. $\mathbf{\Delta}\mu^{g} = \mu^{s}_{g} - \mu^{t}_{g}$ where $\mu^{s}_{g}$ and $\mu^{t}_{g}$ are the sample means of g-th category in the source domain and target domain, respectively. $\mathbf{\Sigma}^{t}_{g}$ represents the covariance matrix of class 'g' in the target domain. More details could be referred to (Li et al., 2021b).

The definition of Eq. (7) is essentially consistent with (Li et al., 2021b), but they use a fixed constant for $\gamma_j$, which is the same for each class. In contrast, this paper reformulates $\gamma_j$ in order to emphasize the semantic augmentation of common classes. Based on the adversarial classifier and batch singular value polarization introduced above, the feature distributions of common classes between the source domain and the target domain continuously align, and the number of target domain samples assigned to source domain private classes decreases. As shown in the third row of Eq. (7), when a class in the source domain has more target domain samples assigned to it, it is more likely to belong to the common classes, denoted as $\hat{n}^{t}_{j}$. Furthermore, to mitigate the impact of class imbalance problem, we consider dividing $\hat{n}^{t}_{j}$ by the number of samples from the j-th class in the source domain (i.e., $n^{s}_{j}$). As such, the strength of data augmentation for source private classes is reduced, and the negative effect are mitigated. The objective of weighted semantic augmentation is shown as below,

$$\min_{\mathcal{C},\mathcal{F}} \mathcal{L}_{\text{wsa}}\Big(\mathcal{C}\big(\mathcal{F}(\mathbf{D}^{s})\big)\Big). \tag{8}$$

# 4. Experiments

To validate the superiority of our BSP-WSA, we compare it with 7 state-of-the-art UniDA models, including UAN (You et al., 2019a), CMU (Fu et al., 2020), DANCE (Saito et al., 2020), DCC (Li et al., 2021a), OVANet (Saito & Saenko, 2021), GATE (Chen et al., 2022c), GLC (Qu et al., 2023).

## 4.1. Experimental Setups

**Datasets.** We conduct extensive experiments on 3 cross-domain object recognition datasets: Office-31 (Saenko et al., 2010), Office-Home (Venkateswara et al., 2017), and VisDA-2017 (Peng et al., 2018). **Office-31** consists of three domains: Amazon (A), DSLR (D), and Webcam (W), and they contain a total of 4,110 images from 31 common categories. **Office-Home** is a more challenging benchmark dataset, comprising 15,588 images from 65 common categories, which has four domains: Artistic (A), Clipart (C), Product (P), and Real-World (R). **VisDA-2017** is a large-scale dataset consisting of 207,785 images from 12 common categories. Following You et al., we choose the synthetic domain (S) as the source domain and the realistic domain (R) as the target domain. Following Saito et al., we conduct class splits for each dataset to construct UniDA scenario, which includes source and target private classes $\overline{\mathbf{Y}}^{s}$, $\overline{\mathbf{Y}}^{t}$, and the common classes between the two domains $\widetilde{\mathbf{Y}}$.

**Evaluation Details.** Following Fu et al., we adopt H-score as the evaluation metric as below,

$$\text{H}_{\text{score}} = \frac{2 \cdot \text{acc}_{\text{c}} \cdot \text{acc}_{\text{u}}}{\text{acc}_{\text{c}} + \text{acc}_{\text{u}}}, \tag{9}$$

where $\text{acc}_{\text{c}}$ denotes the average accuracy of target samples in common classes, and $\text{acc}_{\text{u}}$ denotes the accuracy of target samples in unknown class. This evaluation metric is the harmonic mean of the accuracy on common and unknown, which is higher only when both $\text{acc}_{\text{c}}$ and $\text{acc}_{\text{u}}$ are high.

**Implementation Details.** All experiments are implemented with Pytorch. Following Saito et al., we employ ResNet50 (He et al., 2016) pre-trained on ImageNet (Deng et al., 2009) as our backbone for feature extraction, and the classifier consists of two fully connected layers with batch normalization. Moreover, we train our model for 40K iterations using the mini-batch SGD optimizer with a momentum of 0.9 and a weight decay of 5e-4, and set batch size to 36. In the comparison results, we set the adversarial classifier parameter $\theta$ to 0.2 and the polarization strength r to 3.

## 4.2. Comparison Results

In Table 1, we set $\widetilde{\mathbf{Y}}$, $\overline{\mathbf{Y}}^{s}$, and $\overline{\mathbf{Y}}^{t}$ to be non-empty (OPDA), and their class numbers are 10,10,11; 10,5,50; 6,3,3 for the three datasets, respectively. It could be observed that our proposed BSP-WSA shows average improvements of 0.7%, 1.4%, and 0.7% compared to the second-best approach GATE (underlined) on the three datasets, respectively. In Table 2, we set $\widetilde{\mathbf{Y}}$ and $\overline{\mathbf{Y}}^{s}$ to be non-empty, while $\overline{\mathbf{Y}}^{t}$ is empty (PDA), and their class numbers are 10,21,0; 25,40,0; 6,6,0 for the three datasets, respectively. It could be seen that our proposed BSP-WSA achieves average increases of 1.3%, 0.7%, and 2.4% compared to the second-best approach GATE (underlined) on the three datasets, respectively. In Table 3, we set $\widetilde{\mathbf{Y}}$ and $\overline{\mathbf{Y}}^{t}$ to be non-empty, while $\overline{\mathbf{Y}}^{s}$ is empty (ODA), and their class numbers are 10,0,11; 25,0,40; 6,0,6 for the three datasets, respectively. As ODA does not include source private classes and our main contribution is to prevent the assignment of target samples to source private classes, the improvements are slightly lower compared to OPDA and PDA settings. Additionally, no improvement is observed on Office-31, while its performance is the closest to that of the best method, i.e., OVANet. These results could validate the effectiveness of our method for UniDA, especially when source private classes exist.

*Table 1.* H-score (%) on Office-31 (10/10/11), Office-Home (10/5/50) and VisDA-2017 (6/3/3) in OPDA setting.

| Methods | Venue | Office-31 | | | | | | | Office-Home | | | | | | | | | | | | | VisDA-2017 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A2W | D2W | W2D | A2D | D2A | W2A | Avg. | A2C | A2P | A2R | C2A | C2P | C2R | P2A | P2C | P2R | R2A | R2C | R2P | Avg. | S2R |
| UAN | CVPR'19 | 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 | ECCV'20 | 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 | NeurIPS'20 | 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 | CVPR'21 | 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 | ICCV'21 | 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 | CVPR'22 | **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 |
| BSP-WSA | - | 78.6 | **97.1** | **96.4** | 86.6 | **85.8** | **85.1** | **88.3** | **65.1** | **79.2** | **82.3** | **75.3** | **74.2** | **81.6** | 72.9 | 70.0 | 81.6 | **82.8** | **77.4** | **82.1** | **77.0** | **57.1** |

*Table 2.* H-score (%) on Office-31 (10/21/0), Office-Home (25/40/0) and VisDA-2017 (6/6/0) in PDA setting.

| Methods | Venue | Office-31 | | | | | | | Office-Home | | | | | | | | | | | | | VisDA-2017 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A2W | A2D | D2W | W2D | D2A | W2A | Avg. | A2C | A2P | A2R | C2A | C2P | C2R | P2A | P2C | P2R | R2A | R2C | R2P | Avg. | S2R |
| UAN | CVPR'19 | 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 | ECCV'20 | 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 | NeurIPS'20 | 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 | CVPR'21 | 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 | ICCV'21 | 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 | CVPR'22 | 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 |
| BSP-WSA | - | **86.8** | **89.6** | 99.2 | 99.6 | **97.7** | **97.0** | **95.0** | 49.8 | **77.2** | 80.5 | 81.3 | 69.9 | **87.2** | **75.3** | 60.1 | **83.6** | 83.5 | 60.8 | **85.8** | **74.6** | **78.0** |

*Table 3.* H-score (%) on Office-31 (10/0/11), Office-Home (25/0/40) and VisDA-2017 (6/0/6) in ODA setting.

| Methods | Venue | Office-31 | | | | | | | Office-Home | | | | | | | | | | | | | VisDA-2017 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A2W | A2D | D2W | W2D | D2A | W2A | Avg. | A2C | A2P | A2R | C2A | C2P | C2R | P2A | P2C | P2R | R2A | R2C | R2P | Avg. | S2R |
| UAN | CVPR'19 | 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 | ECCV'20 | 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 | NeurIPS'20 | 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 | CVPR'21 | 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 | ICCV'21 | 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 | CVPR'22 | 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 |
| BSP-WSA | - | **87.7** | 89.4 | 97.2 | 96.1 | **89.8** | 87.2 | 91.2 | 59.5 | **76.6** | **79.8** | **68.4** | 66.7 | **75.3** | 66.6 | **62.2** | **78.0** | 68.7 | **63.8** | **79.2** | **70.4** | **71.1** |

*Table 4.* Ablation study on Office-31 (10/10/11), Office-Home (10/5/50), and VisDA-2017 (6/3/3) in OPDA setting.

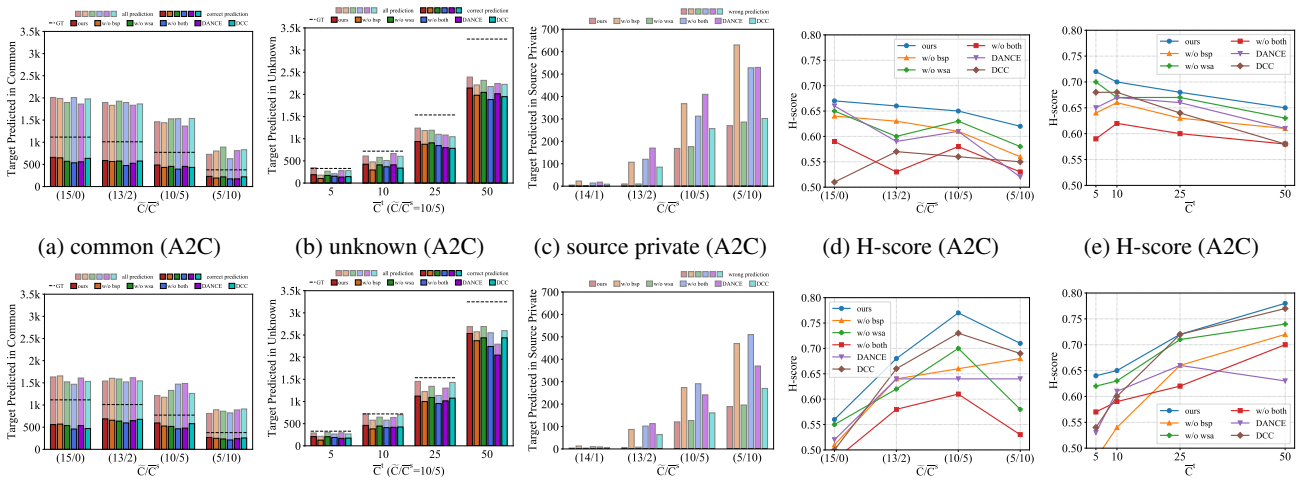
| Module | | Office-31 | Office-Home | VisDA-2017 |
|---|---|---|---|---|
| $\mathcal{L}_{\text{bsp}}$ | $\mathcal{L}_{\text{wsa}}$ | Avg. | Avg. | S2R |
| - | - | 70.5 | 63.2 | 44.9 |
| ✓ | - | 86.3 | 67.8 | 55.1 |
| - | ✓ | 80.1 | 66.3 | 50.5 |
| ✓ | ✓ | **88.3** | **77.0** | **57.1** |

*Table 5.* Ablation study on Office-31 (10/21/0), Office-Home (25/40/0), and VisDA-2017 (6/6/0) in PDA setting.

| Module | | Office-31 | Office-Home | VisDA-2017 |
|---|---|---|---|---|
| $\mathcal{L}_{\text{bsp}}$ | $\mathcal{L}_{\text{wsa}}$ | Avg. | Avg. | S2R |
| - | - | 90.7 | 69.8 | 64.8 |
| ✓ | - | 94.7 | 73.6 | 75.7 |
| - | ✓ | 91.3 | 70.2 | 70.0 |
| ✓ | ✓ | **95.0** | **74.6** | **78.0** |

### 4.3. Model Analysis

**Ablation Study.** To verify the effectiveness of two pivotal components, i.e., BSP and WSA, we conduct experiments by removing either one or both of them from BSP-WSA. Subsequently, we evaluate their performances on the three datasets and record the results in Tables 4∼ 6. We employ the average results of the three datasets in three different settings. From Tables 4 and 5, we observe that considering either of these two components individually enhances the performance of the baseline model, and simultaneously considering both of them results in a more substantial im-




(a) common (A2C) (b) unknown (A2C) (c) source private (A2C) (d) H-score (A2C) (e) H-score (A2C)

(f) common (R2C) (g) unknown (R2C) (h) source private (R2C) (i) H-score (R2C) (j) H-score (R2C)

*Figure 4.* With different UniDA setting, we utilize different methods to predict the numbers (**all prediction**) of target samples belonging to common classes **(a, f)**, unknown class **(b, g)**, and source private classes **(c, h)**. Additionally, we also report the ground-truth numbers (**GT**) of target samples belonging to common, unknown, and source private classes and the numbers (**correct prediction**) correctly predicted. We plot the H-score curves of different methods with varying numbers of source private classes **(d, i)** and target unknown classes **(e, j)**.

*Table 6.* Ablation study on Office-31 (10/0/11), Office-Home (25/0/40), and VisDA-2017 (6/0/6) in ODA setting.

| Module | | Office-31 | Office-Home | VisDA-2017 |
|---|---|---|---|---|
| $\mathcal{L}_{\text{bsp}}$ | $\mathcal{L}_{\text{wsa}}$ | Avg. | Avg. | S2R |
| - | - | 78.4 | 62.1 | 63.4 |
| ✓ | - | 77.2 | 61.9 | 59.8 |
| - | ✓ | **92.3** | **71.2** | **73.3** |
| ✓ | ✓ | 91.2 | 70.4 | 71.1 |

*Table 7.* Plug-and-Play of BSP on Office-31 (10/10/11), Office-Home (10/5/50), and VisDA-2017 (6/3/3) in OPDA setting.

| Methods | Venue | Office-31 | Office-Home | VisDA-2017 |
|---|---|---|---|---|
| - | - | Avg. | Avg. | S2R |
| GLC | CVPR'23 | 87.8 | 75.6 | 65.3$^\star$ |
| GLC+BSP | - | **88.7** | **76.9** | **68.5** |

provement. Furthermore, focusing on BSP proves to be more effective than WSA. However, since the source domain does not contain private classes under ODA setting, forcibly using BSP will degrade the model's performance, and only WSA is effective in this case (as shown in Table 6). This observation aligns with the conclusion from the subsection of 4.2, which shows that the improvements on ODA are not as significant as in the other two settings compared to state-of-the-art UniDA approaches. These results could verify the effectiveness of the two pivotal components within the proposed model.

*Table 8.* Plug-and-Play of BSP on Office-31 (10/21/0), Office-Home (25/40/0), and VisDA-2017 (6/6/0) in PDA setting.

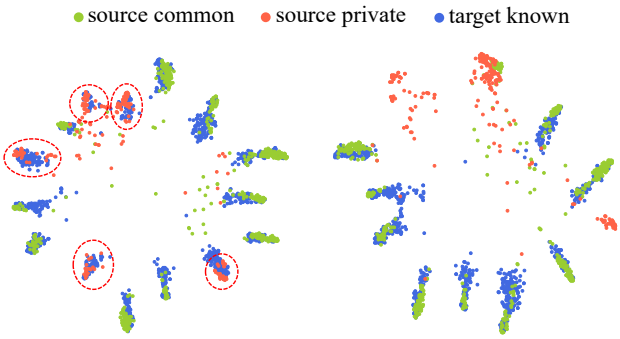
| Methods | Venue | Office-31 | Office-Home | VisDA-2017 |
|---|---|---|---|---|
| - | - | Avg. | Avg. | S2R |
| GLC | CVPR'23 | 94.1 | 72.5 | 76.2 |
| GLC+BSP | - | **94.8** | **73.7** | **76.8** |

In Fig. 4, we compare the performance of different methods under various UniDA settings by setting varying numbers of source private classes ($\overline{\mathbf{Y}}^{\text{s}}$), common classes ($\widetilde{\mathbf{Y}}$), and target unknown classes ($\overline{\mathbf{Y}}^{\text{t}}$). These methods include not only mainstream UniDA methods like DANCE and DCC but also variants of our model with different components removed. From Fig. 4 (a), (b), (c), (f), (g), (h), it can be observed that our method achieves higher accuracy while ensuring that the numbers of target domain predictions and ground-truth numbers are as close as possible. Please note that the dashed line for GT is not shown in (c, h), as target domain samples cannot belong to source domain private classes. Furthermore, from Fig. 4 (d), (e), (i), (j), it can be seen that our method consistently performs the best across different UniDA settings. These observations not only validate the effectiveness of our proposed approach compared to mainstream methods but also highlight the effectiveness of the two key proposed components.

**Plug-and-Play of BSP.** As a plug-and-play module, the proposed BSP module can be applied to many domain adaptation frameworks where the source domain has private

classes. To this end, we incorporate our BSP module into the latest method of GLC (Qu et al., 2023). As illustrated in Tables 7 and 8, we record the reproduction results of GLC and the results of GLC with our BSP module dubbed as GLC+BSP. Notably, we only record the results under OPDA and PDA settings, which could reflect the effectiveness of our BSP module in scenarios where the source domain contains private classes. We can observe From the two tables that employing BSP brings significant performance improvements across different datasets on both settings. Moreover, the result we reported on VisDA-2017 and OPDA setting is not the same as the original paper but rather our reproduced result, which we denote with a superscript '$\star$'.

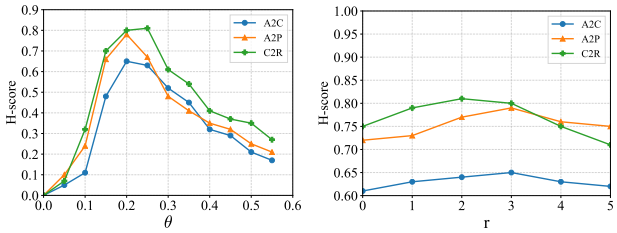
**Feature Visualization.** In Fig. 5, we present feature visualizations for all source samples as well as those target samples predicted as known classes. This visualization experiment is conducted on the A2C task of Office-Home. In Fig. 5 (a), the feature visualization is shown after removing $\mathcal{L}_{\text{bsp}}$. It can be observed that numerous blue and red points are intertwined, indicating that many target samples have been predicted as source private classes. In contrast, Fig. 5 (b) depicts the feature visualization with $\mathcal{L}_{\text{bsp}}$ considered. In this case, the overlap between blue and red points diminishes, and more blue points are observed to align with the green points. This is a consequence of the introduced $\mathcal{L}_{\text{bsp}}$ preventing the classifier from misclassifying target domain samples as source private classes and promoting the feature distribution alignment between common classes.




(a) BSP-WSA w/o $\mathcal{L}_{\text{bsp}}$ (b) BSP-WSA full

*Figure 5.* Feature visualization. Green dots are source common samples, red dots are source private samples, and blue dots are target samples predicted as known classes. For clarity, we remove target samples predicted as unknown class. The red circles indicate target samples are wrongly assigned into source private categories.

**Hyper-Parameter Sensitivity.** To demonstrate the performance influence of hyper-parameters in BSP-WSA, we conduct sensitivity analysis on three UniDA tasks of Office-Home, i.e., A2C, A2P, and C2R. As shown in Fig. 6 (a), the parameter $\theta$ corresponds to a trend in the curve of different




(a) sensitivity of $\theta$ (b) sensitivity of r

*Figure 6.* Sensitivity analysis for hyper-parameters on Office-Home (10/5/50) and OPDA setting.

H-score values, showing an initial increase followed by a decrease. The optimal $\theta$ range is approximately within [0.15, 0.25], indicating that this parameter exhibits relatively low sensitivity to the model within a certain range. As shown in Fig. 6 (b), the magnitude of r reflects the intensity of singular value polarization. It can be observed that the curve corresponding to different H-score values for this parameter shows a gradual rise followed by a gradual decline. The selection range for r is relatively robust, and it is generally found that using a value of 3 yields favorable results.

## 5. Conclusion and Future Work

**Conclusion.** This paper proposes a novel UniDA approach dubbed BSP-WSA. Specifically, BSP-WSA utilizes an adversarial classifier to detect the target unknown class and align feature distributions between the two domains simultaneously. To prevent target samples from being wrongly assigned to source private classes, we perform SVD on the classifier's outputs to maximize larger singular values while minimizing those smaller ones. To better bridge the domain gap, we introduce a weighted semantic augmentation tailored for UniDA to generate data on common classes. Extensive experiments on three cross-domain object recognition benchmarks could verify the superiority of BSP-WSA compared to some state-of-the-art UniDA approaches.

**Future Work.** The success of BSP relies on the assumption that the probability of target domain samples being assigned to private classes of the source domain is small, which requires a high discriminative ability among different classes, and the adversarial classifier can accurately distinguish unknown classes of the target domain. These factors may affect the performance of BSP. Additionally, the proposed approach treats the unknown classes in the target domain as a single category. It is a challenging problem to enhance the performance of UniDA task by exploiting the clustering structure within the unknown classes, as done in GLC (Qu et al., 2023) using clustering methods. The shortcomings in these aspects will be the focus of our future efforts.

## Acknowledgements

We thank all the reviewers for their valuable comments and suggestions during the review process, as well as the area chair for recognizing our work. Additionally, we appreciate all collaborators for their contributions to the ideas, writing, and experiments in this research. Finally, we acknowledge the following projects for funding this research: 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).

## Impact Statement

This work is the first to utilize singular values to analyze the relationship between common and private classes in universal domain adaptation. This work also introduces the concept of singular value polarization, which can be applied to classification tasks containing private classes in the training set.

This paper aims to advance the field of machine learning. Our research does not involve any ethical issues or concerns, so there are no specific ethical aspects to highlight. We ensure that the research process adheres to all relevant ethical and legal standards and poses no negative impact on society.

## References

Cao, Z., Ma, L., Long, M., and Wang, J. Partial adversarial domain adaptation. In *ECCV*, volume 11212, pp. 139–155, 2018.

Cao, Z., You, K., Long, M., Wang, J., and Yang, Q. Learning to transfer examples for partial domain adaptation. In *IEEE CVPR*, pp. 2985–2994, 2019.

Chang, W., Shi, Y., Tuan, H., and Wang, J. Unified optimal transport framework for universal domain adaptation. In *NeurIPS*, 2022.

Chen, L., Du, Q., Lou, Y., He, J., Bai, T., and Deng, M. Mutual nearest neighbor contrast and hybrid prototype self-training for universal domain adaptation. In *AAAI*, pp. 6248–6257, 2022a.

Chen, L., Du, Q., Lou, Y., He, J., Bai, T., and Deng, M. Mutual nearest neighbor contrast and hybrid prototype self-training for universal domain adaptation. In *AAAI*, pp. 6248–6257, 2022b.

Chen, L., Lou, Y., He, J., Bai, T., and Deng, M. Geometric anchor correspondence mining with uncertainty modeling for universal domain adaptation. In *IEEE CVPR*, pp. 16113–16122, 2022c.

Chen, L., Lou, Y., He, J., Bai, T., and Deng, M. Evidential neighborhood contrastive learning for universal domain adaptation. In *AAAI*, pp. 6258–6267, 2022d.

Chen, R., Zhang, H., Liang, S., Li, J., and Cao, X. Less is more: Fewer interpretable region via submodular subset selection. In *ICLR*, 2024.

Chen, X., Wang, S., Fu, B., Long, M., and Wang, J. Catastrophic forgetting meets negative transfer: Batch spectral shrinkage for safe transfer learning. In *NeurIPS*, pp. 1906–1916, 2019a.

Chen, X., Wang, S., Long, M., and Wang, J. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In *ICML*, volume 97, pp. 1081–1090, 2019b.

Cui, S., Wang, S., Zhuo, J., Li, L., Huang, Q., and Tian, Q. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations. In *IEEE CVPR*, pp. 3940–3949, 2020.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *IEEE CVPR*, pp. 248–255, 2009.

Fu, B., Cao, Z., Long, M., and Wang, J. Learning to detect open classes for universal domain adaptation. In *ECCV*, volume 12360, pp. 567–583, 2020.

Gu, S., Zhang, L., Zuo, W., and Feng, X. Weighted nuclear norm minimization with application to image denoising. In *IEEE CVPR*, pp. 2862–2869, 2014.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *IEEE CVPR*, pp. 770–778, 2016.

Huang, C., Peng, Z., Xu, Y., Chen, F., Jiang, Q., Zhang, Y., Jiang, G., and Ho, Y. Online learning-based multi-stage complexity control for live video coding. *IEEE TIP*, 30: 641–656, 2021.

Huang, C., Wu, Z., Wen, J., Xu, Y., Jiang, Q., and Wang, Y. Abnormal event detection using deep contrastive learning for intelligent video surveillance system. *IEEE TII*, 18 (8):5171–5179, 2022a.

Huang, C., Yang, Z., Wen, J., Xu, Y., Jiang, Q., Yang, J., and Wang, Y. Self-supervision-augmented deep autoencoder for unsupervised visual anomaly detection. *IEEE TCYB*, 52(12):13834–13847, 2022b.

Huang, C., Wen, J., Xu, Y., Jiang, Q., Yang, J., Wang, Y., and Zhang, D. Self-supervised attentive generative adversarial networks for video anomaly detection. *IEEE TNNLS*, 34(11):9389–9403, 2023.

Huang, C., Liu, C., Wen, J., Wu, L., Xu, Y., Jiang, Q., and Wang, Y. Weakly supervised video anomaly detection via self-guided temporal discriminative transformer. *IEEE TCYB*, 54(5):3197–3210, 2024.

Kang, G., Jiang, L., Yang, Y., and Hauptmann, A. G. Contrastive adaptation network for unsupervised domain adaptation. In *IEEE CVPR*, pp. 4893–4902, 2019.

Kundu, J. N., Bhambri, S., Kulkarni, A. R., Sarkar, H., Jampani, V., and R., V. B. Subsidiary prototype alignment for universal domain adaptation. In *NeurIPS*, 2022.

Li, G., Kang, G., Zhu, Y., Wei, Y., and Yang, Y. Domain consensus clustering for universal domain adaptation. In *IEEE CVPR*, pp. 9757–9766, 2021a.

Li, S., Xie, M., Gong, K., Liu, C. H., Wang, Y., and Li, W. Transferable semantic augmentation for domain adaptation. In *IEEE CVPR*, pp. 11516–11525, 2021b.

Lifshitz, O. and Wolf, L. Sample selection for universal domain adaptation. In *AAAI*, pp. 8592–8600, 2021.

Liu, H., Cao, Z., Long, M., Wang, J., and Yang, Q. Separate to adapt: Open set domain adaptation via progressive separation. In *IEEE CVPR*, pp. 2927–2936, 2019.

Long, M., Cao, Z., Wang, J., and Jordan, M. I. Conditional adversarial domain adaptation. In *NeurIPS*, pp. 1647–1657, 2018.

Luo, Y., Wang, Z., Huang, Z., and Baktashmotlagh, M. Progressive graph learning for open-set domain adaptation. In *ICML*, pp. 6468–6478, 2020.

Peng, X., Usman, B., Kaushik, N., Wang, D., Hoffman, J., and Saenko, K. Visda: A synthetic-to-real benchmark for visual domain adaptation. In *IEEE CVPR Workshops*, pp. 2021–2026, 2018.

Qu, S., Zou, T., Röhrbein, F., Lu, C., Chen, G., Tao, D., and Jiang, C. Upcycling models under domain and category shift. In *IEEE CVPR*, pp. 20019–20028, 2023.

Saenko, K., Kulis, B., Fritz, M., and Darrell, T. Adapting visual category models to new domains. In *ECCV*, volume 6314, pp. 213–226, 2010.

Saito, K. and Saenko, K. Ovanet: One-vs-all network for universal domain adaptation. In *IEEE ICCV*, pp. 8980–8989, 2021.

Saito, K., Yamamoto, S., Ushiku, Y., and Harada, T. Open set domain adaptation by backpropagation. In *ECCV*, volume 11209, pp. 156–171, 2018.

Saito, K., Kim, D., Sclaroff, S., and Saenko, K. Universal domain adaptation through self supervision. In *NeurIPS*, 2020.

Shi, Z., Ming, Y., Fan, Y., Sala, F., and Liang, Y. Domain generalization via nuclear norm regularization. *CoRR*, abs/2303.07527, 2023.

Venkateswara, H., Eusebio, J., Chakraborty, S., and Panchanathan, S. Deep hashing network for unsupervised domain adaptation. In *IEEE CVPR*, pp. 5385–5394, 2017.

Wang, W., Wang, Z., Li, H., Zhou, J., and Ding, Z. Adaptive local neighbors for transfer discriminative feature learning. In *ECAI*, volume 325, pp. 1595–1602, 2020.

Wang, W., Chen, S., Xiang, Y., Sun, J., Li, H., Wang, Z., Sun, F., Ding, Z., and Li, B. Sparsely-labeled source assisted domain adaptation. *Pattern Recognition*, 112: 107803, 2021.

Wang, W., Li, B., Wang, M., Nie, F., Wang, Z., and Li, H. Confidence regularized label propagation based domain adaptation. *IEEE TCSVT*, 32(6):3319–3333, 2022a.

Wang, W., Li, H., Ding, Z., Nie, F., Chen, J., Dong, X., and Wang, Z. Rethinking maximum mean discrepancy for visual domain adaptation. *IEEE TNNLS*, 34(1):264–277, 2023a.

Wang, W., Wang, M., Dong, X., Lan, L., Zu, Q., Zhang, X., and Wang, C. Class-specific and self-learning local manifold structure for domain adaptation. *Pattern Recognition*, 142:109654, 2023b.

Wang, Y., Huang, G., Song, S., Pan, X., Xia, Y., and Wu, C. Regularizing deep networks with semantic data augmentation. *IEEE TPAMI*, 44(7):3733–3748, 2022b.

Xie, B., Li, S., Li, M., Liu, C. H., Huang, G., and Wang, G. Sepico: Semantic-guided pixel contrast for domain adaptive semantic segmentation. *IEEE TPAMI*, 45(7): 9004–9021, 2023.

Xu, J., Zhang, L., Zhang, D., and Feng, X. Multi-channel weighted nuclear norm minimization for real color image denoising. In *IEEE ICCV*, pp. 1105–1113, 2017.

You, K., Long, M., Cao, Z., Wang, J., and Jordan, M. I. Universal domain adaptation. In *IEEE CVPR*, pp. 2720–2729, 2019a.

You, K., Long, M., Cao, Z., Wang, J., and Jordan, M. I. Universal domain adaptation. In *IEEE CVPR*, pp. 2720–2729, 2019b.

Zhang, J., Ding, Z., Li, W., and Ogunbona, P. Importance weighted adversarial nets for partial domain adaptation. In *IEEE CVPR*, pp. 8156–8164, 2018.