Discriminative Pattern Calibration Mechanism for Source-Free Domain Adaptation

Haifeng Xia[1], Siyu Xia[1]*, Zhengming Ding[2]
[1]School of Automation, Southeast University, [2]Department of Computer Science, Tulane University
{hfxia,xsy}@seu.edu.cn, zding1@tulane.edu

Abstract

Source-free domain adaptation (SFDA) assumes that model adaptation only accesses the well-learned source model and unlabeled target instances for knowledge transfer. However, cross-domain distribution shift easily triggers invalid discriminative semantics from source model on recognizing the target samples. Hence, understanding the specific content of discriminative pattern and adjusting their representation in target domain become the important key to overcome SFDA. To achieve such a vision, this paper proposes a novel explanation paradigm "Discriminative Pattern Calibration (DPC)" mechanism on solving SFDA issue. Concretely, DPC first utilizes learning network to infer the discriminative regions on the target images and specifically emphasizes them in feature space to enhance their representation. Moreover, DPC relies on the attention-reversed mixup mechanism to augment more samples and improve the robustness of the classifier. Considerable experimental results and studies suggest that the effectiveness of our DPC in enhancing the performance of existing SFDA baselines.

1. Introduction

Deep learning [1, 14] recently attracts considerable attentions due to its powerful representation capability, especially ResNet [9] and ViT [46]. Their appearance indeed provides a promising direction on solving the challenging tasks, i.e., image classification [31], object detection [35], semantic segmentation [27], from computer vision and machine learning community. However, deploying them into the real-world scenarios always suffers from obstruction. The primary reason results in that it is difficult to collect sufficient well-annotated instances for the optimization of abundant network parameters [7, 15, 43].

This demand naturally stimulates the exploration of unsupervised domain adaptation (UDA). It aims to reuse knowledge learned from the off-the-shelf source domain with supervisions to address similar task on one novel target domain [39]. The main challenge lies in that cross-domain shift easily induces performance degradation of source model [50]. Fortunately, UDA methods can simply observe source and target images to learn domain-invariant attributions by eliminating distribution divergence [37, 41]. However, real applications fail to satisfy such assumption when considering data privacy and storage. For example, as the improvement of intelligent computing power, multiple basic large models will be embedded into several specific industries such as medical system. Although large models achieve high-generalization ability, it still needs conduct essential model adaptation due to distribution shift across training set and samples of subdivided areas. It is worth noting that such knowledge transfer hardly accesses original training instances. Hence, UDA strategies become invalid under these conditions and this application scenario is further formulated as source-free domain adaptation (SFDA).

Formally, SFDA merely provides permission to access the well-trained source model and unlabeled target images for knowledge transfer [48]. To overcome this setting, the core is discovering meaningful and transferable information from source model and matching them with target visual signals to perform decision. Along with this direction, SHOT [21] freezes source classifier to preserve its recognition ability and fine-tunes feature extractor to adjust target representations closer to source distribution. In order to increase the flexibility of model, Co-learning [47] introduces one ImageNet pre-trained network with classifier to be updated and utilizes the collaboration of it and source model to reach domain adaptation. Similarly, A^{2}Net [38] constructs an auxiliary learnable target classifier and explores the adversarial relation between dual-classifier and generator to achieve cross-domain alignment. Moreover, the other branch expects to boost model performance via self-learning fashion on target domain. Specifically, NRC [44] adopts clustering constraint over target features to allocate samples from the identical category into one tight subspace. And, AaD [45] considers the consistency between paired features and their predictions and theoretically

*Corresponding Author.

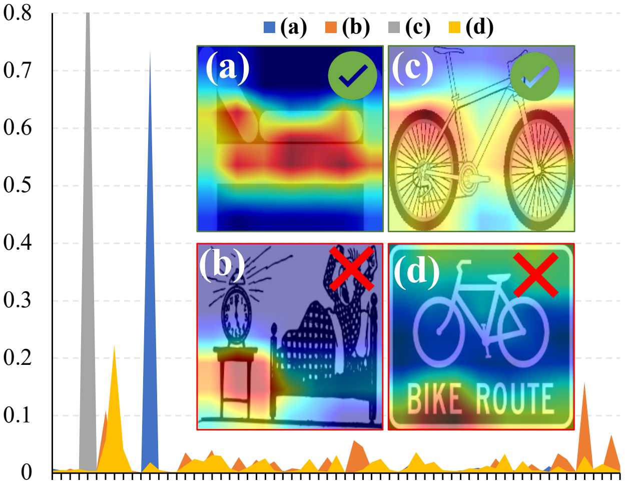


Figure 1. The relationship between prediction confidence (y-axis) and attention shift, where x-axis denotes the category index. When producing lower prediction confidence on samples, the model likely focuses on classification-irrelevant contents.

deduces the upper bound to realize this anticipation. Although existing approaches have obtained obvious progress on dealing with SFDA, they fail to provide deep insight on what discriminative and transferable semantics the source model can provide. Hence, it is difficult for them to comprehensively exploit well-learned source knowledge on adapting the property of target domain.

To surmount this drawback, this paper explores a novel "Discriminative Pattern Calibration (DPC)" mechanism to delicately interpret the influence of source knowledge on target visual signals and provides a feasible technical route to better solve SFDA. Concretely, we first explore the relation between attention shift and prediction confidence. As Figure 1 shows, when the attention map derived from the model is out of the region of interests, this image is very likely to be predicted incorrectly. In order to avoid such a situation, the intuitive strategy is adjusting distribution of attention matrix to involve more informative patterns for effective adaptation. This consideration evolves into our discriminative semantic enhancement (DSE) module in DPC mechanism. Specifically, given the attention map from Grad-CAM [29] per target image, our DSE deploys an autoencoder architecture to calibrate the attention distribution with the prediction confidence. The adjusted attention emphasizes the contribution of important regions in raw images and DSE integrates these discriminative patterns into high-level representations via semantic filling manner. In addition, DPC considers further to enhance the robustness of the classifier by expanding the diversity of target samples. This anticipation is formulated as the second module "Attention Induced Mixture" (AIM). Importantly, when augmenting the additional instances, AIM avoids the conflict of multiple objects in the same picture. In a nutshell, our main contributions are summarized as three folds:

- First, this paper rethinks SFDA from interpretable perspective to gain a deeper understanding of the source knowledge and rely on such target-relevant information to benefit model adaptation.
- Second, a novel discriminative pattern calibration mechanism is presented to adaptively discover and highlight discriminative semantics and to augment feasible target images via attention reverse manner.
- Third, extensive experimental results and empirical studies demonstrate that our proposed DPC effectively advances the existing SFDA methods, especially our DPC improves SHOT and AaD by 1.4% and 1.3% on average classification accuracy of Office-Home.

2. Related Works

Source-Free Domain Adaptation. In SFDA, the well-learned source model is adjusted to adapt unlabeled target domain with the absence of source data. Several methods [21, 38] freeze certain components of source model such as classifier to instruct model adaptation on target dataset. Moreover, [44, 45] explore structural information of target samples and adopt clustering manner to capture class-discriminative representations. And, [22] introduces additional tasks such as rotation prediction to enrich semantics. Similarly, [19] exploits generative adversarial network to produce target-style image and achieve more accurate prediction. In addition, [3, 17] consider boosting pseudo label quality by selecting target instances with low entropy or loss, and utilize these selected visual signals to fine-tune source model. Different from them, our proposed DPC starts from interpretable learning to discover discriminative and transferable knowledge from source model and target images, and achieve better model adaptation via the adjustment and usage of attention map.

Interpretable Machine Learning. Several interpretability techniques have been introduced to enhance the transparency and trustworthiness of machine learning models. The post-hoc interpretability [2, 8, 29, 33, 49] establishes fidelity interpretation methods using causal inference, visualization, and other strategies to elucidate the operational mechanism and decision criteria of trained models. The classic class activation mapping (CAM) [49] enables the models to locate important and discriminative regions. To generalize CAM to any deep CNN, Grad-CAM [29] utilizes the gradient backpropagation of any target concept to produce a coarse localization map, which highlights the important regions in the image for concept prediction. Thus, Grad-CAM enhances the interpretability of image classification models [12], object detection [42], image segmentation [40]. This paper mainly explores Grad-CAM to understand source knowledge during model adaptation and utilizes attention matrix to better assist model learning.

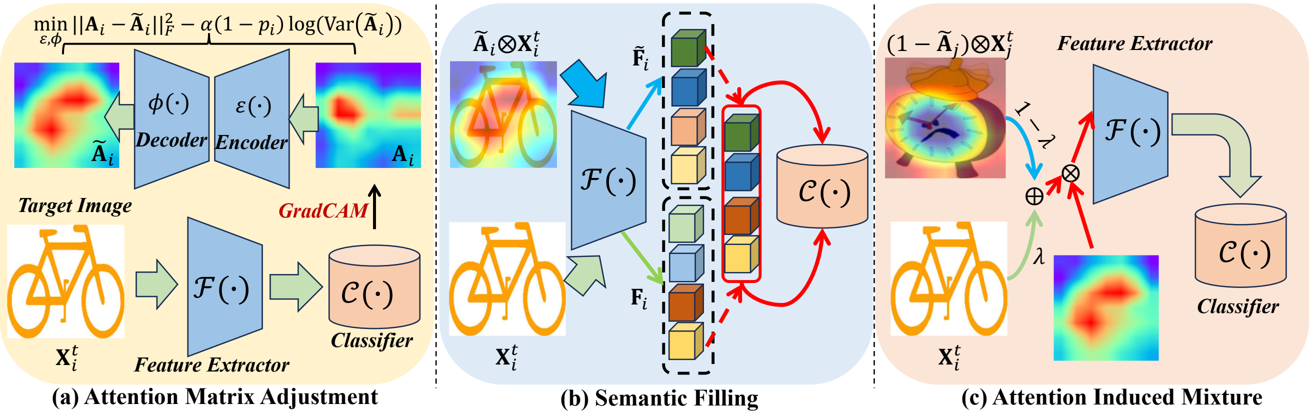


Figure 2. Overview of the proposed method. Our discriminative pattern calibration (DPC) includes three main procedures. (a) Given target image $\mathbf{X}_i$, model produces original attention matrix via Grad-CAM and $\mathbf{A}_i$ will be calibrated into $\widetilde{\mathbf{A}}_i$ by the introduced encoder and decoder. (b) The model takes the highlighted image $\widetilde{\mathbf{A}}_i \otimes \mathbf{X}_i$ and raw picture $\mathbf{X}_i$ as inputs to obtain their features $\mathbf{F}_i$ and $\widetilde{\mathbf{F}}_i$ and perform cross-feature semantic filling for learning more discriminative representations. (c) To avoid the conflict of multiple objects in the same image, DPC adopts attention reverse manner to blend $\mathbf{X}_i$ with $\mathbf{X}_j$ and increase sample diversity to promote the robustness of the classifier.

3. Proposed Method

3.1. Preliminary & Motivation

In source-free domain adaptation (SFDA), knowledge transfer procedure explicitly accesses well-trained source model $\{\mathcal{F}(\cdot), \mathcal{C}(\cdot)\}$ and unlabeled target domain $\mathcal{D}_t = \{\mathbf{X}_i^t | \mathbf{X}_i^t \in \mathbb{R}^{C_x \times H_x \times W_x}\}_{i=1}^{n_t}$ with n_t visual signals $\mathbf{X}_i^t$, where $\{\mathcal{F}(\cdot), \mathcal{C}(\cdot)\}$ denotes feature extractor and classifier, respectively, and C_x, H_x, W_x are channel number, height and width of the input image. It is noteworthy that the supervised learning over n_s source samples with cross-entropy loss, i.e., $\min_{\mathcal{F}(\cdot),\mathcal{C}(\cdot)} \sum_{i=1}^{n_s} \ell_{ce}(\mathcal{C}(\mathcal{F}(\mathbf{X}_i^s)), y_i^s)$, produces a data-free source model, where $\mathbf{X}_i^s$, y_i^s are source image and its corresponding annotation. Moreover, source and target inputs are collected from different distributions, but share the identical category space including c specific classes, i.e., $\mathcal{P}(\mathbf{X}^s) \neq \mathcal{P}(\mathbf{X}^t)$, $\mathcal{P}(y^s|\mathbf{X}^s) = \mathcal{P}(y^t|\mathbf{X}^t)$. This cross-domain shift likely results in that source model suffers from significant performance degradation when directly evaluated on target domain [47]. The solution to SFDA needs discover domain-invariant knowledge from available source model and adapt them into target distribution. Along with this direction, SHOT [21] explores the frozen source classifier to preserve discriminative information and adjusts target features to fit the source classifier boundary by optimizing feature generator. Moreover, AaD [45] focuses on the consistency between paired features and their predictions and deduces one upper bound to achieve this constraint.

Different from them, this paper proposes a novel method "Discriminative Pattern Calibration (DPC)" to surmount the bottleneck of SFDA from interpretable perspective and perform model adaptation in continuous learning manner. As Figure 2 shows, DPC first deploys post-hoc back-propagation-based tools such as Grad-CAM [30] and Score-CAM [34] to interpret which regions of the given image make more contributions to the decision, and then adjusts their heatmap distribution according to the prediction confidence. The calibrated attention map not only further emphasizes necessary contents to improve discriminability of representations but also expands sample diversity to enhance the robustness of the classifier.

3.2. Discriminative Semantics Enhancement

Given a target image $\mathbf{X}_i^t$, source model easily predicts its annotation via $\hat{y}_i^t = \arg\max_j \mathbf{p}_{ij} = \mathcal{C}(\mathcal{F}(\mathbf{X}_i^t))$, where $\mathbf{p}_{ij}$ is the j-th element of $\mathbf{p}_i \in \mathbb{R}^c$. Under this condition, Grad-CAM measures the difference between $\hat{y}_i^t$ and $\mathbf{p}_i$ to infer attention matrix $\mathbf{A}_i$ with the combination of feature maps from $\mathcal{F}(\cdot)$. When attaching $\mathbf{A}_i$ over its input image $\mathbf{X}_i^t$ in Figure 1, it is straightforward to observe that the activated regions in several images are out of our interested objects, causing the source classifier to struggle in accurately identifying their categories. In other words, the performance degradation of source model on target domain actually results from the **attention region shift** due to distribution divergence across source and target images. Moreover, while compared with probability distribution $\mathbf{p}_i$ of samples with correct predictions, that of the remaining instances have **lower probability**. It suggests that prediction confidence of target instance with wrong prediction is lower. Based on observations, we naturally post a question "*Can we adaptively adjust attention map with the guidance of predictive confidence and facilitate discriminative feature learning?*"

To answer this question, it is necessary to explore the relation between predictive confidence and attention map in order to clearly understand adjustment direction. In fact, the aforementioned phenomenons explicitly demonstrate that the attention maps with region shift likely correspond to these samples with lower confidence. Hence, whether this attention map needs to be adjusted depends on the predicted confidence score. Moreover, moving attention region into our interested object means searching these pixels in high-dimensional image space. This ideal operation will become difficult and time-consuming due to scarcity of realistic pixel-level annotation. To this end, we can gain a deep insight into the property of Gaussian distribution and seek for an optimal strategy to adjustment.

Figure 3 illustrates three Gaussian distributions with the same mean and various standard deviation settings. As standard deviation increases, the distribution will transform from a steep form to a flat shape. In other words, the peak corresponding to mean value is reduced and others are elevating. Thus, when assuming that distribution of the deduced attention map follows Gaussian distribution, the improvement of its standard deviation declines attention of highlighted region and intensifies focus of other areas. In this way, the interested object is easily involved into important regions in attention map $\mathbf{A}_i$. This simple mechanism suggests one reasonable adjustment direction. Hence, discriminative semantic enhancement (DSE) module is presented to implement it. Concretely, DSE consists of an encoder $\varepsilon(\cdot)$ mapping $\mathbf{A}_i$ into hidden space and a decoder $\phi(\cdot)$ recovering latent representation to $\widetilde{\mathbf{A}}_i$, i.e., $\mathbf{h}_i = \varepsilon(\mathbf{A}_i)$ and $\widetilde{\mathbf{A}}_i = \phi(\mathbf{h}_i)$. The above discussions reveal that $\widetilde{\mathbf{A}}_i$ not only preserves most of the basic information from $\mathbf{A}_i$ but also moderately expands significant regions to involve discriminative patterns. The expectation further evolves into the following objective function for the training of DSE:

$$\min_{\varepsilon,\phi} \|\mathbf{A}_i - \widetilde{\mathbf{A}}_i\|_{\mathbf{F}}^2 - \alpha(1-p_i)\log\big(\mathsf{var}(\widetilde{\mathbf{A}}_i)\big), \tag{1}$$

where $\|\cdot\|_{\mathbf{F}}$ is the Frobenius norm, $\mathsf{var}(\cdot)$ calculates the variance of the input, $p_i = \max_{\mathbf{p}_{ij}} \mathbf{p}_{ij} = \mathcal{C}(\mathcal{F}(\mathbf{X}_i^t))$ is the maximum probability of prediction, and α is by default set as 1e-2 to balance the recovery and expansion. The second term denotes that samples with lower predictive confidence tend to be more constrained on the variance of its attention matrix. Note that Eq. (1) only optimizes network parameters of encoder and decoder.

So far, the adjusted attention matrix has been in hand and provided more accurate emphasis on salient regions of raw visual signals. The next consideration attempts to gain support from $\widetilde{\mathbf{A}}_i$ to assist model in capturing more discriminative semantics. It naturally induces the second important operation "semantic filling" in DSE module. Specifically, attaching $\widetilde{\mathbf{A}}_i$ into image produces a novel input visual signal, i.e., $\widetilde{\mathbf{X}}_i^t = \widetilde{\mathbf{A}}_i \otimes \mathbf{X}_i^t$ where $\otimes$ denotes the element-wise multiplication of attention map to each channel of the inputs. Compared with $\mathbf{X}_i^t$, $\widetilde{\mathbf{X}}_i^t$ weights each pixel with the corresponding importance which is brought into high-level features via network forward propagation as $\widetilde{\mathbf{F}}_i = \mathcal{F}(\widetilde{\mathbf{X}}_i^t)$ where $\widetilde{\mathbf{F}}_i \in \mathbb{R}^{C_f \times H_f \times W_f}$ and C_f, H_f, W_f represent the channel number, height and width of this feature map. Similarly, raw image flowing into feature extractor also produces $\mathbf{F}_i$ with the same dimension. Under this condition, the identical location across $\widetilde{\mathbf{F}}_i$ and $\mathbf{F}_i$ should represent consistent semantic. And if the value of this point in $\widetilde{\mathbf{F}}_i$ is larger than that in $\mathbf{F}_i$, it implies that $\widetilde{\mathbf{F}}_i$ offers more meaningful information on this location due to the guidance of adjusted attention matrix. Such a useful content is supposed to be embedded into $\mathbf{F}_i$ via:

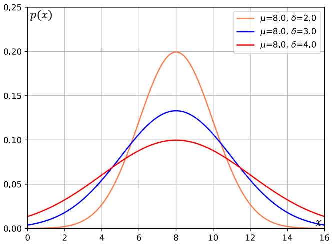


Figure 3. Three different Gaussian distribution with the same mean and various standard deviations.

$$\hat{\mathbf{F}}_i = \mathsf{ReLU}(\widetilde{\mathbf{F}}_i - \mathbf{F}_i) + \mathbf{F}_i, \tag{2}$$

where $\mathsf{ReLU}(\cdot)$ only reserves positive values by replacing negative ones with zero. The calibrated representation $\hat{\mathbf{F}}_i$ obtains additional knowledge related to object from $\widetilde{\mathbf{F}}_i$. And these activated contents likely belong to domain-invariant features and are simply recognized by source classifier due to their powerful transferability.

Remark: Our proposed Discriminative Semantics Enhancement (DSE) module is designed to extract transferable knowledge from input images using an adjusted attention matrix. Crucially, the DSE module highlights this information in the hidden feature space, enabling the source classifier to accurately identify target instances. The collaboration of these two procedures results in a feasible model adaptation even in the absence of source data.

3.3. Attention Induced Mixture

For the SFDA scenario, one task is to discover shared and transferable patterns across source and target images and multiplexs classification boundary learned from source domain. Our DSE module has reached this demand by boosting contribution of important pixels. From another viewpoint, mining more useful knowledge from the current available instances also has positive influence on performance improvement since the model pays more attention to target distribution. On this direction, NRC [44] considers structural information of target features and makes the distribution of subspace more compact via clustering fashion. Similarly, A^{2}Net [38] introduces contrastive learning mechanism to constrain distribution of hidden representations. Although they effectively endow the features with stronger discriminative ability, using target images to promote robustness of classifier is ignored. The intuitive solution to learn robust classifier is increasing the diversity of observed images. For existing UDA works [25, 36, 50], they typically adopt immovable or random ratio to mix source and target images to create more samples in the intermediate domain. However, it is difficult to conduct this data augmentation due to the absence of source instances in SFDA. Certainly, the combination of two arbitrary target images also achieves data augmentation. Moreover, the composite image generally involves two or more objects which can obviously confuse classifier and trigger incorrect classification boundary deformation. This simple mixup manner not only fails to promote robustness but also hurts its recognition ability.

To overcome such a dilemma, we consider utilizing attention matrix to remove discriminative region from one target image and integrate it with the other instance. Formally, given arbitrary two target images $\mathbf{X}_i$ and $\mathbf{X}_j$, utilizing the source model produces their attention matrix $\mathbf{A}_i$ and $\mathbf{A}_j$. Sequentially, the frozen DSE module reassigns importance for each pixel via:

$$\widetilde{\mathbf{A}}_i = \phi(\varepsilon(\mathbf{A}_i)), \quad \widetilde{\mathbf{A}}_j = \phi(\varepsilon(\mathbf{A}_j)). \tag{3}$$

When regarding $\mathbf{X}_i$ and $\mathbf{X}_j$ as the primary and auxiliary samples respectively, the expected situation is gradually mitigating the effect of object region and retaining the remaining pixels in $\mathbf{X}_j$. In fact, $(\mathbf{1} - \widetilde{\mathbf{A}}_j)$ has realized it by moving values in high activation region closer to zero and upgrading values of others closer to one. The image with reduced discriminability is formulated as $(\mathbf{1} - \widetilde{\mathbf{A}}_j) \otimes \mathbf{X}_j$, where $\mathbf{1}$ is a matrix of all ones. As a result, the linear combination between primary image and the weighted auxiliary sample generates one novel instance as:

$$\hat{\mathbf{X}}_i = \boldsymbol{\lambda}\mathbf{X}_i + (\hat{\mathbf{1}} - \boldsymbol{\lambda})(\mathbf{1} - \widetilde{\mathbf{A}}_j) \otimes \mathbf{X}_j, \tag{4}$$

where $\hat{\mathbf{1}}$ is a vector with all ones, $\boldsymbol{\lambda} \in \mathbb{R}^{C_x}$ is collected from the parameterized distribution, i.e., $\boldsymbol{\lambda} \sim \mathbf{Beta}(\beta, \gamma)$, and it adopts various combination coefficients $\{\boldsymbol{\lambda}_k\}_{k=1}^{C_x}$ for different channels by considering their respective attributions. Since DSE module explores semantic filling mechanism to activate more potential features, we take $\hat{\mathbf{X}}_i$ and $\widetilde{\mathbf{A}}_i \otimes \hat{\mathbf{X}}_i$ as the paired input and obtain features to rewrite Eq. (2) as:

$$\hat{\mathbf{F}}_i = \mathsf{ReLU}\Big(\mathcal{F}(\widetilde{\mathbf{A}}_i \otimes \hat{\mathbf{X}}_i) - \mathcal{F}(\hat{\mathbf{X}}_i)\Big) + \mathcal{F}(\hat{\mathbf{X}}_i). \tag{5}$$

And then classifier transforms $\hat{\mathbf{F}}_i$ into the label space as $\mathcal{C}(\hat{\mathbf{F}}_i)$. Moreover, when deriving $\mathbf{A}_{i/j}$, we simultaneously also obtain their pseudo labels as $\hat{y}^t_{i/j}$. In order to improve the model robustness, we introduce the following:

$$\mathcal{L}_c = \mathop{\arg\min}_{\mathcal{F}(\cdot),\mathcal{C}(\cdot)} \sum_{i=1}^{n_t} \lambda \ell_{ce}\Big(\mathcal{C}(\hat{\mathbf{F}}_i), \hat{y}_i^t\Big) + (1-\lambda)\ell_{ce}\Big(\mathcal{C}(\hat{\mathbf{F}}_i), \hat{y}_j^t\Big), \tag{6}$$

where $\lambda = \frac{1}{C_x}\sum_{i=1}^{C_x} \boldsymbol{\lambda}_k$ and $\ell_{ce}(\cdot)$ is the cross-entropy loss. It is worth noting that the auxiliary sample $\mathbf{X}_j$ is randomly selected from dataset for each given image $\mathbf{X}_i$.

3.4. Embedded Instruction

For SFDA issue, our "Discriminative Pattern Calibration" method mainly focuses on the design of input image to capture discriminative representation and improve the robustness of the classifier by utilizing attention matrix. First, DSE module in our DPC method adjusts attention maps to emphasize discriminative region in raw image and exploits its features to calibrate original representation via Eq. (2). Second, according to $\mathbf{A}_{i/j}$, DPC conducts linear combination of paired images via Eq. (4) and maintains the discriminating attributes of one object as much as possible to optimize the overall model.

In fact, since DPC only conducts operations on input layer and high-level feature, it is convenient to plug our DPC into the existing SFDA works to advance their performance. Take the popular SHOT [21] as one example. **First**, it utilizes $\min_{\mathcal{F}(\cdot),\mathcal{C}(\cdot)} \sum_i^{n_s} \ell_{ce}(\mathcal{F}(\mathcal{C}(\mathbf{X}_i^s)), y_i^s)$ to acquire source model. And then, DPC can take target images as input to pre-train encoder and decoder with the overall frozen source model. **Second**, given single image $\mathbf{X}_i$, we obtain its attention matrix $\widetilde{\mathbf{A}}_i$ from the previous model and rely on DSE module to get $\hat{\mathbf{F}}_i$ which classifier uses to calculate all objective functions mentioned in SHOT. Note that the clustering procedure also depends on $\hat{\mathbf{F}}_i$. During this stage, three sub-networks $\{\varepsilon,\ \phi,\ \mathcal{F}(\cdot)\}$ will be optimized. **Third**, with the paired images, $\hat{\mathbf{F}}_i$ in Eq. (5) will be explored to compute Eq. (6) updating parameters in $\mathcal{F}(\cdot)$ and $\mathcal{C}(\cdot)$. In the next iteration, the updated model will be utilized to repeatedly implement the above processes.

4. Experiments

4.1. Experimental Setup

Datasets. To evaluate the performance of our DPC, we conduct considerable experiments on three popular domain adaptation benchmarks. **Office-31** [28] includes three domains: Amazon (**A**), Webcam (**W**) and DSLR (**D**) and they share the identical 31 categories such as laptop and backpack. **Office-Home** [32] collects images from 65 common household items and these samples belong to four different domains: Art (**Ar**), Clipart (**Cl**), Product (**Pr**) and Real-World (**Rw**). Compared with Office-31, the difficulty of cross-domain knowledge transfer on Office-Home results from the increasing number of categories. **VisDA-C** [26] is a large-scale dataset and consists of one source domain with synthetic rendering images of 3D models and one target domain with Microsoft COCO real pictures. Source and target domains have the same 12 classes. This dataset is generally considered as one standard to assess transferability of model on dealing with synthetic-to-real shift.

Implementation Details. In SFDA setting, the first step is to obtain a well-trained model with the supervision of source samples and it typically includes one feature extractor $\mathcal{F}(\cdot)$ and one classifier $\mathcal{C}(\cdot)$. As for source model, we follow the protocol of [21, 45], and adopt ResNet-50 on Office-31/Office-Home and ResNet-101 on VisDA-C as the basic backbone $\mathcal{F}_s$ to extract high-level representation from image input. The network parameters are initialized with

Table 1. Classification accuracy of 12 domain adaptation tasks on Office-Home benchmark. The best performance of SFDA is highlighted in **bold**, while the best one achieved by UDA works is highlighted with underline. **AS** means the access of source and target images.

Method	AS	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg
Resnet[10]	✓	46.3	67.5	75.9	59.1	59.9	62.7	58.2	41.8	74.9	67.4	48.2	74.2	61.3
CDAN [23]	✓	50.7	70.6	76.0	57.6	70.0	70.0	57.4	50.9	77.3	70.9	56.7	81.6	65.8
GVB-GD [4]	✓	57.0	74.7	79.8	64.6	74.1	74.6	65.2	55.1	81.0	74.6	59.7	84.3	70.4
RSDA [6]	✓	53.2	77.7	81.3	66.4	74.0	76.5	67.9	53.0	82.0	75.8	57.8	85.4	70.9
TSA [20]	✓	57.6	75.8	80.7	64.3	76.3	75.1	66.7	55.7	81.2	75.7	61.9	83.8	71.2
SRDC [31]	✓	52.3	76.3	81.0	69.5	76.2	78.0	68.7	53.8	81.7	76.3	57.1	85.0	71.3
FixBi [25]	✓	58.1	77.4	80.4	67.7	79.5	78.1	65.8	57.9	81.7	76.4	62.9	86.7	72.7
SFDA [17]	✗	48.4	73.4	76.9	64.3	69.8	71.7	62.7	45.3	76.6	69.8	50.5	79	65.7
NRC [44]	✗	58.0	79.3	81.8	**70.1**	78.7	78.7	63.5	57.0	82.8	71.6	58.2	84.3	72.0
A^{2}Net [38]	✗	58.4	79.0	82.4	67.5	79.3	78.9	68.0	56.2	82.9	**74.1**	60.5	85.0	72.8
SHOT [21]	✗	57.1	78.1	81.5	68.0	78.2	78.1	67.4	54.9	82.2	73.3	58.8	84.3	71.8
AaD [45]	✗	58.7	79.8	81.4	67.5	79.4	78.7	64.7	56.8	82.5	70.3	58.0	83.3	71.8
SHOT+DPC	✗	59.2	79.8	82.6	68.9	**79.7**	79.5	**68.6**	56.5	82.9	73.9	**61.2**	**85.4**	**73.2**($\uparrow_{1.4}$)
AaD+DPC	✗	**59.5**	**80.6**	**82.9**	69.4	79.3	**80.1**	67.3	**57.2**	**83.7**	73.1	58.9	84.9	73.1($\uparrow_{1.3}$)

ImageNet-1k weights and the last fully-connected (FC) layer is replaced with a new bottleneck layer. Moreover, the classifier involves two FC layers with weight normalization. When conducting model adaptation, our method specially introduces two additional sub-networks: encoder $\varepsilon(\cdot)$ and decoder $\phi(\cdot)$. The former includes two convolutional layers with ReLU as activation function, while the latter involves two deconvolutional ones. For the overall training procedure, we adopt SGD as the optimizer with a momentum of 0.9 and a weight decay of 1e-3. During the model adaptation, we pre-train the new added encoder and decoder with learning rate 1e-2. And then, the learning rate for optimizing $\{\mathcal{F}(\cdot), \varepsilon, \phi\}$ is set as 1e-3 and 1e-4 on Office-31/Office-Home and VisDA-C, respectively. When using Eq. (6) to slightly finetune classifier, we deploy 1e-5 as the learning rate.

Baselines. To illustrate the effectiveness of our DPC on solving SFDA, we select the classical UDA methods and recent SFDA algorithms as competitors. These UDA strategies concurrently observe source and target samples to achieve cross-domain alignment. They include CDAN [23], MCC [13], CAN [16], GSDA [11], SRDC [31], GVB-GD [4], RSDA [6], TSA [20], FixBi [25], SFAN [41], STAR [24], and SE [5]. Differently, SFDA solutions only perform model adaptation by using source model and target instances, which are SDDA [18], SFDA [17], NRC [44], A^{2}Net [38], SHOT [21], AaD [45]. In our experiments, our proposed training mechanism is plugged into SHOT [21] and AaD [45] by adding operations on image input and high-level feature activation.

Table 2. Classification accuracy of six domain adaptation tasks on Office-31. The best performance for SFDA is emphasized in **bold**, while the best one achieved by UDA works is highlighted with underline. **AS** means the access of source and target images.

Method	AS	A→D	A→W	D→A	D→W	W→A	W→D	Avg
ResNet [10]	✓	68.9	68.4	62.5	96.7	60.7	99.3	76.1
CDAN [23]	✓	92.9	94.1	71.0	98.6	69.3	100.0	87.7
MCC [13]	✓	95.6	95.4	72.6	98.6	73.9	100.0	89.4
CAN [16]	✓	95.0	94.5	78.0	99.1	77.0	99.8	90.6
GSDA [11]	✓	94.8	95.7	73.5	99.1	74.9	100.0	89.7
SRDC [31]	✓	95.8	95.7	76.7	99.2	77.1	100.0	90.8
SDDA [18]	✗	85.3	82.5	66.4	**99.0**	67.7	99.8	83.5
SFDA [17]	✗	92.2	91.1	71.0	98.2	71.2	99.5	87.2
NRC [44]	✗	92.0	91.6	74.5	97.9	74.8	**100.0**	88.5
A^{2}Net [38]	✗	94.5	94.0	**76.7**	99.2	76.1	**100.0**	90.1
SHOT [21]	✗	94.0	90.1	74.7	98.4	74.3	99.9	88.6
AaD [45]	✗	94.4	93.3	75.9	98.4	76.3	99.8	89.7
SHOT+DPC	✗	**95.9**	92.6	75.4	98.6	76.2	**100.0**	89.8 ($\uparrow_{1.2}$)
AaD+DPC	✗	95.8	**94.5**	76.5	98.9	**76.8**	**100.0**	**90.5** ($\uparrow_{0.8}$)

4.2. Comparison Results

Table 1, Table 2 and Table 3 summarize the performance of recent UDA methods and SFDA solutions on dealing with domain adaptation issue. It is straightforward to observe several significant phenomenons and conduct the intuitive analysis. **First**, with respect to the average classification accuracy, the integration of our DPC to SHOT/AaD surpasses most SFDA methods by a large margin and achieves comparable results with UDA works. Specifically, when conducting knowledge transfer on VisDA-C dataset, the combination of our DPC and AaD exceeds FixBi by 1.6%, which suggests that our proposed method effectively facilitates model to adapt target distribution even with the absence of source images. **Second**, our DPC brings the additional benefits to the existing SFDA algorithms to boost their model transferability. For example, on task **Ar→Cl**, DPC mechanism assists SHOT boosting recognition accuracy about 2.1%. The success mainly results from the exploration and feasible usage of the derived attention matrix. It directly

Table 3. Classification accuracy of domain adaptation task on VisDA-C benchmark. The best performance for SFDA is emphasized in **bold**, while the best one achieved by UDA works is highlighted with underline. **AS** means the access of source and target images.

Methods	AS	plane	bike	bus	car	horse	knife	mcycle	person	plant	sktbrd	train	truck	Avg
Resnet [10]	✓	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
CDAN [23]	✓	85.2	66.9	83.0	50.8	84.2	74.9	88.1	74.5	83.4	76.0	81.9	28.0	73.9
SFAN [41]	✓	93.6	61.3	84.1	70.6	94.1	79.0	91.8	79.6	89.9	55.6	89.0	24.4	76.1
MCC [13]	✓	88.7	80.3	80.5	71.5	90.1	93.2	85.0	71.6	89.4	73.8	85.0	36.9	78.8
STAR [24]	✓	95.0	84.0	84.6	73	91.6	91.8	85.9	78.4	94.4	84.7	87.0	42.2	82.7
SE [5]	✓	95.9	87.4	85.2	58.6	96.2	95.7	90.6	80.0	94.8	90.8	88.4	47.9	84.3
CAN [16]	✓	97.0	87.2	82.5	74.3	97.8	96.2	90.8	80.7	96.6	96.3	87.5	59.9	87.2
FixBi [25]	✓	96.1	87.8	90.5	90.3	96.8	95.3	92.8	88.7	97.2	94.2	90.9	25.7	87.2
SFDA [41]	✗	86.9	81.7	84.6	63.9	93.1	91.4	86.6	71.9	84.5	58.2	74.5	42.7	76.7
A^{2}Net [38]	✗	94.0	87.8	85.6	66.8	93.7	95.1	85.8	81.2	91.6	88.2	86.5	56.0	84.3
NRC [44]	✗	96.8	**92.0**	83.8	57.2	96.6	95.3	84.2	79.6	94.3	93.9	90.0	59.8	85.3
SHOT [21]	✗	94.3	88.5	80.1	57.3	93.1	94.9	80.7	80.3	91.5	89.1	86.3	58.2	82.9
AaD [45]	✗	**96.9**	90.2	85.7	82.8	**97.4**	96.0	89.7	83.2	**96.8**	94.4	90.8	49.0	87.7
SHOT+DPC	✗	95.6	88.2	82.8	59.4	92.5	95.7	85.6	81.7	91.6	90.9	87.6	**60.1**	84.3($\uparrow_{1.4}$)
AaD+DPC	✗	96.5	89.3	**86.5**	**83.2**	**97.4**	**97.3**	**91.8**	**83.7**	96.4	**94.8**	**92.1**	56.2	**88.8**($\uparrow_{1.1}$)

helps model to discover domain-invariant semantics by emphasizing discriminative and transferable regions in raw images. **Third**, with our DPC, SHOT and AaD obtain 1.9% and 1.4% gain on task **A**→**D** of Office-31 benchmark. In fact, the number of samples in **D** is much less than that in **A**. Hence, it is difficult to capture more discriminative contents from target images. Under this challenging situation, our DPC also provides complementary information for them. The main reason lies in that our method depends on attention matrix to increase diversity of visual signals.

4.3. Empirical Analysis

Effect of Calibration. When solving SFDA task, our DPC focuses on the adjustment of input layer to advance the existing source-free methods by calibrating attention matrix and conducting semantic filling. To provide an in-depth analysis of how our approach achieves this outcome, we attempt to draw attention map derived by target models over the raw images. Concretely, with **Ar** as source domain, target images of **Pr** are fed into the learned models by SHOT and SHOT+DPC. In Figure 4, the first two rows list several samples which SHOT incorrectly identify yet SHOT+DPC correctly recognize. For example, given "clock" picture in the first column, SHOT fails to accurately capture the discriminative regions of object to make decision, while our DPC assists SHOT locating the specific time "12:03" to produce correct prediction. The success of this case illustrates that our DPC effectively adjusts the distribution of attention and intensifies domain-invariant contents in input images to enable model to learn more useful patterns. Moreover, with **Ar** as source domain, we take images of **Rw** as input for models of AaD and AaD+DPC. From the last two rows of Figure 4, the deployment of our DPC on AaD significantly promotes the robustness of classification model. Specifically, for "desk lamp" image in the third column, AaD pays more attention to background information instead of the interested object, leading to the incorrect prediction. However, our DPC mitigates the negative influence of irrelevant background and focuses on lamp holder.

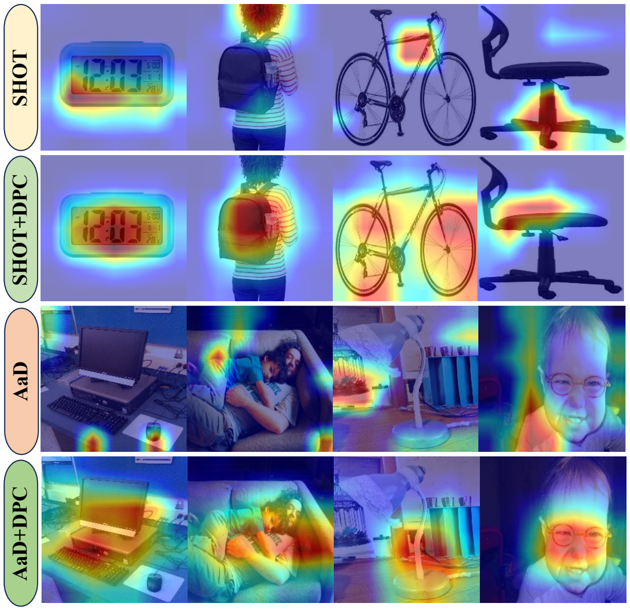
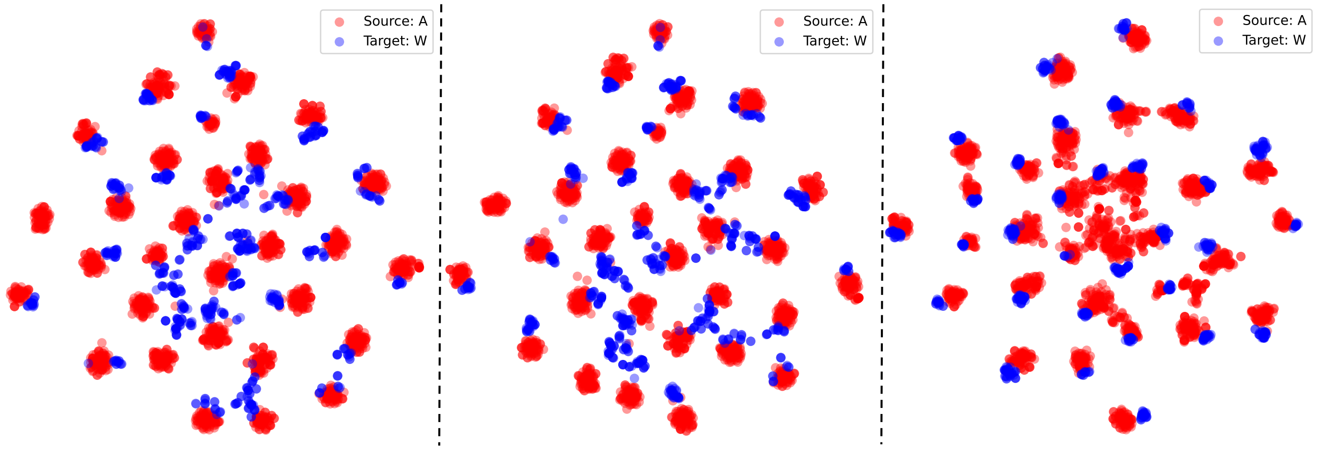


Figure 4. Comparison of attention matrix. The images of the first two rows come from **Pr** and describe clock, backpack, bike and chair. The images of the bottom two rows are sampled from **Rw** and describe computer, sofa, desk lamp and glasses.

Feature Visualization. In fact, the main challenge of model adaptation under SFDA scenario is how to learn discriminative and transferable representation by reusing source



(a) Source Model

(b) SHOT

(c) SHOT + DPC

Figure 5. Visualization of feature distribution. Source (**A**) and target (**W**) images are fed into the source model and the learned network by SHOT and SHOT+DPC to obtain their hidden representations.

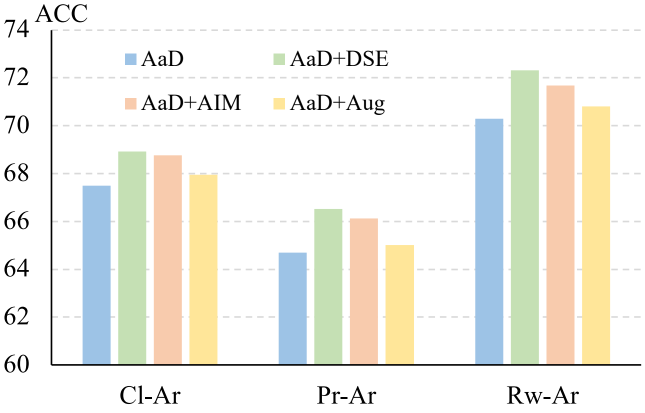


Figure 6. Ablation study. To analyse effect of each module, we gradually introduce the proposed components into AaD.

knowledge. To give insight on these properties of features, we integrate images from **A** and **W** into the learned target models of SHOT and SHOT+DPC to obtain their hidden representations and visualize them in 2D-plane with t-SNE tool in Figure 5 (b) and (c). Moreover, feature distribution derived from source model in Figure 5 (a) is considered as one basic reference. According to the comparison, target features per category mostly reside in a more compact subspace and distribute closer to the corresponding source features. It demonstrates that our DPC mechanism enables model to learn more discriminative representation and gradually eliminates distribution shift. In addition, we have one interesting observation. Concretely, the source classifier in SHOT is frozen, which guarantees the explicit classification boundary among source features from various categories. However, this boundary starts becoming vague due to the weak optimization of classifier via Eq. (6) by using DPC. This supplies model with more freedom to adapt target distribution and reach performance improvement.

Ablation Studies. Our DPC utilizes the combination of "discriminative semantic enhanment" (DSE) and "attention induced mixture" (AIM) to advance the existing SFDA methods. In order to analyse the contribution of each component, we design three ablation references. Concretely, AaD+DPC only integrates DSE module into the basic algorithm. AaD+AIM only uses attention matrix derived from normal GradCAM to do sample augmentation without attention adjustment and directly utilizes these mixed images to train AaD model without "semantic filling". Different from AaD+AIM, AaD+Aug simply blend one target image with the other one to conduct data augmentation without the guidance of attention matrix. As Figure 6 shows, DSE module can produce more positive effect on performance improvement when compared with AIM operation. This mainly results from that the adjustment of attention matrix detects more discriminative information from visual signals and embeds them into high-level representation. Moreover, the comparisons between AaD+AIM and AaD+Aug verify that AIM avoids the conflict of multiple objects in the augmented images and provides reasonable mixture.

5. Conclusion

In this paper, we propose a novel discriminative pattern calibration (DSE) mechanism to better solve SFDA issue. Specifically, it adjusts attention matrix to highlight discriminative patterns and conduct semantic filling on high-level features. Moreover, DPC increases sample diversity via attention reverse manner to promote the robustness of classifier. Experimental results and analysis on several popular benchmarks illustrate that DPC effectively advances the existing works to achieve better knowledge transfer.

Acknowledgment

This work is partially supported by the Louisiana Board of Regents Support Fund (LEQSF(2022-25)-RD-A-22) and NIH 2U19 AG055373-06A1.

References

[1] Yannis Assael, Thea Sommerschield, Brendan Shillingford, Mahyar Bordbar, John Pavlopoulos, Marita Chatzipanagiotou, Ion Androutsopoulos, Jonathan Prag, and Nando de Freitas. Restoring and attributing ancient texts using deep neural networks. *Nature*, 603(7900):280–283, 2022. 1

[2] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 839–847. IEEE, 2018. 2

[3] Weijie Chen, Luojun Lin, Shicai Yang, Di Xie, Shiliang Pu, and Yueting Zhuang. Self-supervised noisy label learning for source-free unsupervised domain adaptation. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10185–10192. IEEE, 2022. 2

[4] Shuhao Cui, Shuhui Wang, Junbao Zhuo, Chi Su, Qingming Huang, and Qi Tian. Gradually vanishing bridge for adversarial domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12455–12464, 2020. 6

[5] Geoffrey French, Michal Mackiewicz, and Mark Fisher. Self-ensembling for domain adaptation. *arXiv preprint arXiv:1706.05208*, 7, 2017. 6, 7

[6] Xiang Gu, Jian Sun, and Zongben Xu. Spherical space domain adaptation with robust pseudo-label loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9101–9110, 2020. 6

[7] Kai Han, Yunhe Wang, Hanting Chen, Xinghao Chen, Jianyuan Guo, Zhenhua Liu, Yehui Tang, An Xiao, Chunjing Xu, Yixing Xu, et al. A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence*, 45(1):87–110, 2022. 1

[8] Syed Nouman Hasany, Caroline Petitjean, and Fabrice Mériaudeau. Seg-xres-cam: Explaining spatially local regions in image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3732–3737, 2023. 2

[9] 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. 1

[10] 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, 7

[11] Lanqing Hu, Meina Kan, Shiguang Shan, and Xilin Chen. Unsupervised domain adaptation with hierarchical gradient synchronization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4043–4052, 2020. 6

[12] Hongyang Jiang, Jie Xu, Rongjie Shi, Kang Yang, Dongdong Zhang, Mengdi Gao, He Ma, and Wei Qian. A multi-label deep learning model with interpretable grad-cam for diabetic retinopathy classification. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1560–1563. IEEE, 2020. 2

[13] Ying Jin, Ximei Wang, Mingsheng Long, and Jianmin Wang. Minimum class confusion for versatile domain adaptation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16*, pages 464–480. Springer, 2020. 6, 7

[14] Taotao Jing, Haifeng Xia, Renran Tian, Haoran Ding, Xiao Luo, Joshua Domeyer, Rini Sherony, and Zhengming Ding. Inaction: Interpretable action decision making for autonomous driving. In *European Conference on Computer Vision*, pages 370–387. Springer, 2022. 1

[15] Taotao Jing, Haifeng Xia, Jihun Hamm, and Zhengming Ding. Marginalized augmented few-shot domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 1

[16] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4893–4902, 2019. 6, 7

[17] Youngeun Kim, Donghyeon Cho, Kyeongtak Han, Priyadarshini Panda, and Sungeun Hong. Domain adaptation without source data. *IEEE Transactions on Artificial Intelligence*, 2(6):508–518, 2021. 2, 6

[18] Vinod K Kurmi, Venkatesh K Subramanian, and Vinay P Namboodiri. Domain impression: A source data free domain adaptation method. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 615–625, 2021. 6

[19] Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9641–9650, 2020. 2

[20] Shuang Li, Mixue Xie, Kaixiong Gong, Chi Harold Liu, Yulin Wang, and Wei Li. Transferable semantic augmentation for domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11516–11525, 2021. 6

[21] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International conference on machine learning*, pages 6028–6039. PMLR, 2020. 1, 2, 3, 5, 6, 7

[22] Jian Liang, Dapeng Hu, Yunbo Wang, Ran He, and Jiashi Feng. Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8602–8617, 2021. 2

[23] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *Advances in neural information processing systems*, 31, 2018. 6, 7

[24] Zhihe Lu, Yongxin Yang, Xiatian Zhu, Cong Liu, Yi-Zhe Song, and Tao Xiang. Stochastic classifiers for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Con-*

ference on Computer Vision and Pattern Recognition, pages 9111–9120, 2020. 6, 7

[25] 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. 4, 6, 7

[26] 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. 5

[27] Lixiang Ru, Heliang Zheng, Yibing Zhan, and Bo Du. Token contrast for weakly-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3093–3102, 2023. 1

[28] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV 11*, pages 213–226. Springer, 2010. 5

[29] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 2

[30] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 3

[31] Hui Tang, Ke Chen, and Kui Jia. Unsupervised domain adaptation via structurally regularized deep clustering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8725–8735, 2020. 1, 6

[32] 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

[33] Haofan Wang, Zifan Wang, Mengnan Du, Fan Yang, Zijian Zhang, Sirui Ding, Piotr Mardziel, and Xia Hu. Score-cam: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 24–25, 2020. 2

[34] Haofan Wang, Zifan Wang, Mengnan Du, Fan Yang, Zijian Zhang, Sirui Ding, Piotr Mardziel, and Xia Hu. Score-cam: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 24–25, 2020. 3

[35] Wei Wu, Hao Chang, Yonghua Zheng, Zhu Li, Zhiwen Chen, and Ziheng Zhang. Contrastive learning-based robust object detection under smoky conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4295–4302, 2022. 1

[36] Yuan Wu, Diana Inkpen, and Ahmed El-Roby. Dual mixup regularized learning for adversarial domain adaptation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*, pages 540–555. Springer, 2020. 4

[37] Haifeng Xia and Zhengming Ding. Cross-domain collaborative normalization via structural knowledge. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2777–2785, 2022. 1

[38] Haifeng Xia, Handong Zhao, and Zhengming Ding. Adaptive adversarial network for source-free domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9010–9019, 2021. 1, 2, 4, 6, 7

[39] Haifeng Xia, Taotao Jing, and Zhengming Ding. Maximum structural generation discrepancy for unsupervised domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3434–3445, 2022. 1

[40] Mengying Xiao, Liyuan Zhang, Weili Shi, Jianhua Liu, Wei He, and Zhengang Jiang. A visualization method based on the grad-cam for medical image segmentation model. In *2021 International Conference on Electronic Information Engineering and Computer Science (EIECS)*, pages 242–247. IEEE, 2021. 2

[41] Ruijia Xu, Guanbin Li, Jihan Yang, and Liang Lin. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1426–1435, 2019. 1, 6, 7

[42] Toshinori Yamauchi and Masayoshi Ishikawa. Spatial sensitive grad-cam: Visual explanations for object detection by incorporating spatial sensitivity. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 256–260. IEEE, 2022. 2

[43] Guanglei Yang, Haifeng Xia, Mingli Ding, and Zhengming Ding. Bi-directional generation for unsupervised domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, pages 6615–6622, 2020. 1

[44] Shiqi Yang, Joost van de Weijer, Luis Herranz, Shangling Jui, et al. Exploiting the intrinsic neighborhood structure for source-free domain adaptation. *Advances in neural information processing systems*, 34:29393–29405, 2021. 1, 2, 4, 6, 7

[45] Shiqi Yang, Shangling Jui, Joost van de Weijer, et al. Attracting and dispersing: A simple approach for source-free domain adaptation. *Advances in Neural Information Processing Systems*, 35:5802–5815, 2022. 1, 2, 3, 5, 6, 7

[46] Li Yuan, Yunpeng Chen, Tao Wang, Weihao Yu, Yujun Shi, Zi-Hang Jiang, Francis EH Tay, Jiashi Feng, and Shuicheng Yan. Tokens-to-token vit: Training vision transformers from scratch on imagenet. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 558–567, 2021. 1

[47] Wenyu Zhang, Li Shen, and Chuan-Sheng Foo. Rethinking the role of pre-trained networks in source-free domain adaptation. In *Proceedings of the IEEE/CVF International*

Conference on Computer Vision, pages 18841–18851, 2023. 1, 3

[48] Yixin Zhang, Zilei Wang, and Weinan He. Class relationship embedded learning for source-free unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7619–7629, 2023. 1

[49] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016. 2

[50] Jinjing Zhu, Haotian Bai, and Lin Wang. Patch-mix transformer for unsupervised domain adaptation: A game perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3561–3571, 2023. 1, 4