

Gradually Vanishing Bridge for Adversarial Domain Adaptation

Shuhao Cui^{1,2} Shuhui Wang^{1*} Junbao Zhuo^{1,2} Chi Su³ Qingming Huang^{1,2,4} Qi Tian⁵

¹Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS, Beijing, China

²University of Chinese Academy of Sciences, Beijing, China ³Kingsoft Cloud, Beijing, China

⁴Peng Cheng Laboratory, Shenzhen, China ⁵Noah's Ark Lab, Huawei Technologies

{cuishuhao18s, wangshuhui}@ict.ac.cn, junbao.zhuo@vipl.ict.ac.cn, suchi@kingsoft.com, qmhuang@ucas.ac.cn, tian.qil@huawei.com

Abstract

In unsupervised domain adaptation, rich domain-specific characteristics bring great challenge to learn domain-invariant representations. However, domain discrepancy is considered to be directly minimized in existing solutions, which is difficult to achieve in practice. Some methods alleviate the difficulty by explicitly modeling domain-invariant and domain-specific parts in the representations, but the adverse influence of the explicit construction lies in the residual domain-specific characteristics in the constructed domain-invariant representations. In this paper, we equip adversarial domain adaptation with Gradually Vanishing Bridge (GVB) mechanism on both generator and discriminator. On the generator, GVB could not only reduce the overall transfer difficulty, but also reduce the influence of the residual domain-specific characteristics in domain-invariant representations. On the discriminator, GVB contributes to enhance the discriminating ability, and balance the adversarial training process. Experiments on three challenging datasets show that our GVB methods outperform strong competitors, and cooperate well with other adversarial methods. The code is available at <https://github.com/cuishuhao/GVB>.

1. Introduction

Diverse vision applications have gained significant improvement with the cutting-edge technologies such as deep convolutional neural networks (CNN). Despite of success already achieved, even a subtle departure from training data may still cause existing shallow or deep models to make spurious predictions. Directly applying a deep CNN well-trained on millions of images to a new domain encounters performance degradation, while collecting labeled samples for the new domain is expensive and time-consuming. To alleviate such notorious domain discrepancy, researchers re-

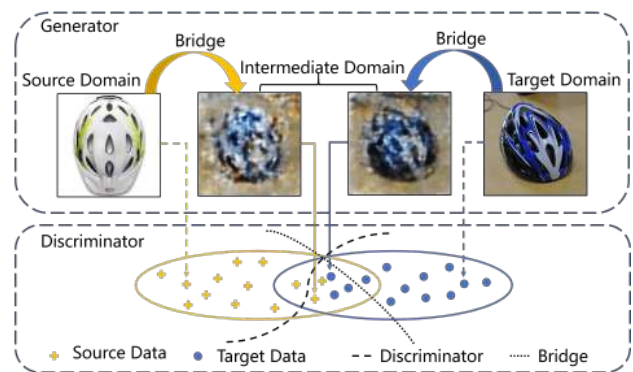


Figure 1. Illustration of the bridge for adversarial domain adaptation. On the generator, the bridge models domain-specific representations and connects either source or target domain to intermediate domain. On the discriminator, the bridge balances the adversarial game by providing additive adjustable discriminating ability. In these processes, the method to construct bridge is the key issue in our study.

sort to unsupervised domain adaptation (UDA) [5, 13, 37, 39, 46, 50] that transfers knowledge from labeled source data to related unlabeled target data.

In traditional UDA methods, domain discrepancy is reduced by learning transferable representations [10, 14, 28], or reweighing the importance of the samples [29]. Deep methods carry forward the thought of learning transferable representations by distribution alignment on multiple layers [8, 9, 23, 44, 51]. Recently, many domain adaptation methods are inspired by Generative Adversarial Networks (GAN) [1, 12, 27, 49]. Based on GAN, the transferable representations are constructed by the generator that tries to fool the discriminator in the *minmax* game [9, 17, 35, 36, 40]. In these studies, the discrepancy between source and target domains is considered to be *directly* minimized. Towards more robust UDA, we reinvestigate the intrinsic weakness possessed by existing domain discrepancy minimization paradigms.

*Corresponding author.

To reduce domain discrepancy, the great challenge lies in the rich domain-specific characteristics, which are deemed to be mitigated by distribution alignment [8, 23] or adversarial training [9, 35]. Towards more comprehensive reduction of domain-specific characteristics, some methods [2, 3, 11] explicitly model domain-specific and domain-invariant representations by input image reconstruction. The time-consuming reconstruction function demands rich domain characteristics in domain-specific representations. Such characteristics inevitably result in more domain properties in the intermediate domain, which is expected to be domain-invariant. From another perspective of adversarial learning, the adversarial minmax game could achieve better results by balancing the learning ability between the generator and discriminator. Strengthening the discriminating ability by multiple discriminators [30] may facilitate better collaboration with a strong generator in some situations, while the overly critical multiple discriminators may also break the brittle balance of adversarial training.

In this paper, we define the bridge as a basic concept, which is a measurement modeling the discrepancy between the existing and ideal representations. As shown by example in Figure 1, the bridge is applied to the generator and discriminator. The bridge on generator models domain-specific parts and connects both source and target domains to the intermediate domain, thus it could mitigate the overall transfer difficulty and facilitate more comprehensive domain alignment. The bridge on discriminator measures the discrepancy between the discriminating function hyperplane and the ideal domain decision boundary.

The key constraint on our constructed bridge is that the range of the bridge is progressively reduced. With the constraint, the bridge mechanism is called Gradually Vanishing Bridge (GVB) for adversarial domain adaptation. We denote GVB applied to generator and discriminator as GVB-G and GVB-D, respectively. Among the bridges, the bridge on generator models domain-specific parts, thus GVB-G could explicitly reduce domain-specific characteristics. Therefore, GVB-G plays the main role in learning domain-invariant representations. It could also reduce the adverse influence of domain characteristics in the representations, and avoid the influence of hard examples with excessive domain characteristics.

Both GVB-G and GVB-D are integrated into the whole framework, denoted as GVB-GD, to ensure the balance and robustness of the two-player minmax game. Experiments show that GVB-GD outperforms competitors on three challenging datasets and achieves state-of-the-art on Office-Home. We also apply GVB to other UDA methods including CDAN [24] and Symnets [48]. The improvements on CDAN and Symnets show the general applicability of GVB on adversarial domain adaptation. For better explanation, we visualize the bridge in GVB-GD to validate that the

bridge could measure the domain-specific characteristics. We further verify the functionality and necessity of GVB, by observation that larger range of GVB-G output tends to result in higher misclassification probability.

In summary, the key contributions are as follows:

- We propose a new framework which constructs GVB on both generator and discriminator to achieve more balanced adversarial training for domain adaptation.
- As the key role in the framework, GVB on generator could alleviate the transfer difficulty and explicitly reduce the domain characteristics in the representations.
- GVB outperforms competitors in most cases, and visualization also show the positive effect of GVB.

2. Related Work

Visual domain adaptation, first proposed by [34], receives increasing attention and fruitful results [45, 43], especially with deep learning techniques [8, 9, 15, 18, 19, 6]. Techniques developed on deep networks could be mainly categorized into moment alignment and adversarial training. Moment-alignment-based methods [14, 23, 25, 38, 44] are proposed to measure and minimize the domain discrepancy with maximum mean discrepancy (MMD) [23], correlation distance [38, 51] or other distance metric [20, 25] calculated on task-specific features.

Motivated by the Generative adversarial networks (GAN) [12], adversarial learning has been successfully explored in various research fields including domain adaptation [4, 36, 42]. In Domain Adversarial Neural Network (DANN) [9], the authors propose a gradient reversal layer to confuse the domain classifier. Many methods such as ADDA [40], CyCADA [17], SBADA [33], and MCD [35] are proposed to build adversarial frameworks to directly minimize the domain discrepancy. Typically, CDAN [24] is a principled framework that conditions the adversarial adaptation models on discriminative information conveyed in the classifier predictions. Symnets [48] makes a symmetric design of source and target task classifiers, in which both domain classifiers produce classification outputs.

Some methods try to explicitly model the bridge on generator, but they still face many problems. SGF [28] and GFK [10] build the bridge between domains on Grassmannian manifolds, but could not be easily applied to deep networks. In deep methods, DSN [2], DISE [3] and DLOW [11] formulate the bridge between source and target by reconstructing input images on pixel level. However, the time-consuming reconstruction function does not guarantee less residual domain characteristics in the domain-invariant representations. We avoid using image reconstruction function and reduce the range of bridge on generator to get

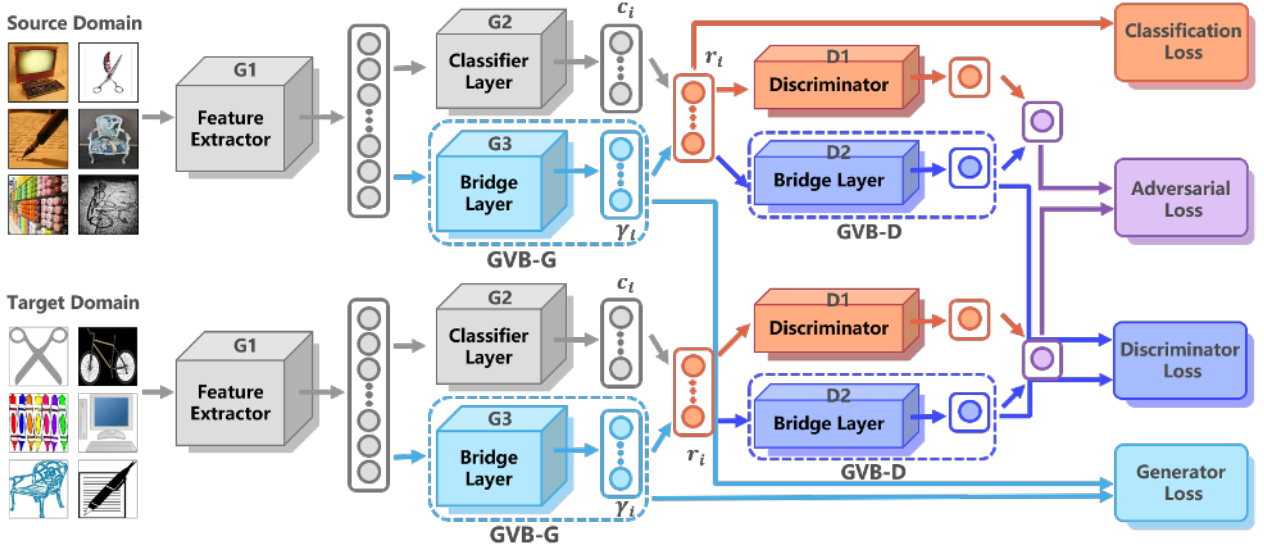


Figure 2. The proposed GVB-GD framework, which is an end-to-end network shared by source domain and target domain. In GVB-G, the bridge layer G_3 outputs γ_i that captures domain-specific properties. The intermediate representation r_i is the classifier response c_i subtracted by bridge γ_i . Similarly, in GVB-D, the bridge layer D_2 provides additive discriminating ability to the base domain classifier in discriminator. The network is trained by minimizing classification loss, adversarial loss, discriminator loss and generator loss.

rid of the negative influence of domain-specific characteristics. Besides, multiple discriminators in adversarial networks [30] may break the balance in the adversarial training process with overly critical discriminating ability. In contrast, we pursue a more balanced minmax adversarial game by building the bridge on both generator and discriminator.

3. Method

We propose GVB for adversarial domain adaptation as shown in Figure 2. In Sec. 3.1, we introduce the baseline for adversarial domain adaptation. The gradually vanishing bridge on generator (GVB-G) is introduced in Sec. 3.2. In Sec. 3.3, we describe the gradually vanishing bridge on discriminator (GVB-D) and the whole framework GVB-GD based on GVB-G and GVB-D.

3.1. Adversarial Domain Adaptation

We are given source domain $\mathcal{D}_S = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ with N_s labeled examples covering C classes where $y_i^s \in \{1, \dots, C\}$ and target domain $\mathcal{D}_T = \{x_i^t\}_{i=1}^{N_t}$ with N_t unlabeled examples that belong to the same C classes.

It is widely accepted that methods for domain adaptation can be achieved by minimizing a classification loss and an additional transfer loss. The classification loss in \mathcal{D}_S can be calculated as:

$$\mathcal{L}_{cls} = \frac{1}{N_s} \sum_{i=1}^{N_s} L_{ce}(G_*(x_i^s), y_i^s) \quad (1)$$

where L_{ce} is the cross-entropy loss function and G_* is the

network structure to obtain class responses. To construct transfer loss, inspired by [23, 35], we directly minimize the domain discrepancy on classifier responses. Besides, we adopt popular methods of adversarial training, the same with other deep domain adaptation methods [8, 17, 24]. Thus the transfer loss could be formulated as:

$$\begin{aligned} \mathcal{L}_{trans}^{adv} = & -\frac{1}{N_s} \sum_{i=1}^{N_s} \log D_*(G_*(x_i^s)) \\ & -\frac{1}{N_t} \sum_{j=1}^{N_t} \log(1 - D_*(G_*(x_j^t))), \end{aligned} \quad (2)$$

where D_* is the constructed discriminator. Some methods utilize extra loss, such as reconstruction loss, which can be denoted by \mathcal{L}_{ext} . The overall objective function is:

$$\begin{aligned} \min_{G_*} \quad & \mathcal{L}_{cls} + \mathcal{L}_{trans}^{adv} + \mathcal{L}_{ext} \\ \max_{D_*} \quad & \mathcal{L}_{trans}^{adv}. \end{aligned} \quad (3)$$

This formulation is regarded as the general framework for adversarial domain adaptation. In particular, suppose that the classification network consists of a feature extractor G_1 and a successive classifier layer G_2 . For the input data x_i , the feature output and the classifier response can be computed as $f_i = G_1(x_i)$ and $c_i = G_2(f_i)$, respectively. With the discriminator D_1 , the baseline of our method could be expressed as:

$$\begin{cases} G_*(x_i) = G_2(G_1(x_i)) \\ D_*(G_*) = D_1(G_*), \quad \mathcal{L}_{ext} = 0, \end{cases} \quad (4)$$

3.2. Gradually Vanishing Bridge on Generator

Actually, divergence between source and target domains can hardly be minimized to zero, thus it is quite challenging to seek direct knowledge transfer from source domain \mathcal{D}_S to target domain \mathcal{D}_T . Alternatively, to alleviate the transfer difficulty, we divide the transfer process into two separate processes from either source or target domain to intermediate domain \mathcal{D}_I . \mathcal{D}_I is supposed to lie in the middle of \mathcal{D}_S and \mathcal{D}_T , with much less domain-specific characteristics than \mathcal{D}_S and \mathcal{D}_T . To model domain-specific representations, we construct a bridge layer G_3 in light blue boxes in Figure 2. The output of G_3 is the bridge denoted as γ_i , which captures the domain-specific representation from the input x_i in either \mathcal{D}_S or \mathcal{D}_T , i.e., $\gamma_i = G_3(G_1(x_i))$. Subtracting bridge γ_i from the classifier response c_i , results in the constructed domain-invariant representation, denoted as intermediate representation r_i , calculated as follows:

$$r_i = c_i - \gamma_i. \quad (5)$$

With bridge γ_i , the framework is the same as Eqn. 4 except the generator. And the overall generator becomes:

$$G_*(x_i) = G_2(G_1(x_i)) - G_3(G_1(x_i)). \quad (6)$$

The adversarial training process enforces that the distributions of r_i across domains are as similar as possible. Therefore, γ_i tends to be dominated by domain-specific representations according to Eqn. 5.

In existing study, some bridge-based methods on the generator have been proposed [2, 3, 11], but the ways of constructing the bridge of existing study remain to be discussed. The bridge is built by image reconstruction in [2, 3, 11]. However, the image reconstruction function inevitably results in too much residual domain-specific characteristics and large overall range in the constructed bridge. In connection to our framework, if the range of γ_i , measured in vector-norm, is overly large, it has the following interpretations. First, it implies that data x_i is a hard example which contains heavy domain-specific properties. Second, the domain-specific and domain-invariant parts for x_i can hardly be separated. Consequently, the large range of the bridge means rich domain characteristics in the γ_i , and inevitably impacts c_i and r_i . The influence on the representations leads to a higher misclassification probability in \mathcal{D}_T , which is validated by experiment results in Figure 5(a).

Towards more effective constraint on the bridge, we seek to reduce the influence of domain characteristics in γ_i . Thus we gradually reduce the influence of γ_i by progressively minimizing the overall range of γ_i as follows:

$$\mathcal{L}_G = \frac{1}{(N_s + N_t) * C} \sum_{i=1}^{N_s+N_t} \sum_{j=1}^C |\gamma_{i,j}|, \quad (7)$$

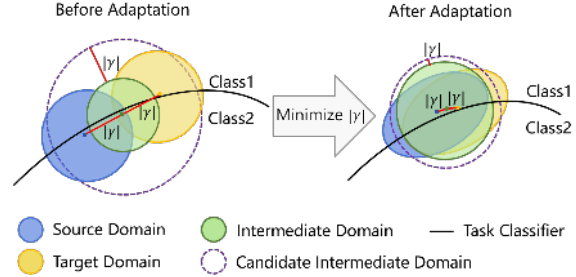


Figure 3. A toy example of the training process with bridge γ . With gradually minimized range of γ , the discrepancy between source and target domains is reduced. After adaptation, more points are covered in intermediate domain. But the points still outside of intermediate domain are regarded as hard examples or noisy data.

where each γ_i is encouraged to be near zeroes. With the reduced range of γ_i , the characteristics in γ_i is suppressed, which results in minor negative effects on c_i and r_i . The mechanism of minimizing the range of γ_i is denoted as GVB. We denote GVB applied to generator as GVB-G. As shown in the experiment in Figure 4, visualization of the data with and without γ_i also validates that γ_i still contains rich domain characteristics without input reconstruction.

For better comprehension of the gradually vanishing bridge mechanism, we show the example in two-dimension space in Figure 3 working on generator. We denote \mathcal{D}_S , \mathcal{D}_T and \mathcal{D}_I with circles in different colors. For simplicity, we assume that the range of bridge γ is the single parameter to be optimized, which can measure the distance between the centers of $\mathcal{D}_S(\mathcal{D}_T)$ and \mathcal{D}_I . The distance shows the discrepancy between $\mathcal{D}_S(\mathcal{D}_T)$ and \mathcal{D}_I , which could approximately measure the quantity of domain-specific characteristics in $\mathcal{D}_S(\mathcal{D}_T)$. $|\gamma|$ appears to be large before domain adaptation, which demonstrates the heavy domain characteristics in $\mathcal{D}_S(\mathcal{D}_T)$. Based on Eqn. 7, $|\gamma|$ is minimized to reduce domain discrepancy. In this way, \mathcal{D}_S and \mathcal{D}_T will be pushed to be closer and overlapped with \mathcal{D}_I as much as possible to obtain domain-invariant characteristics. In this case, there is less domain-specific characteristics in $\mathcal{D}_S(\mathcal{D}_T)$, resulting more reliable domain-invariant representations in \mathcal{D}_I .

Furthermore, all the data points, from either \mathcal{D}_S or \mathcal{D}_T that are possibly dragged into \mathcal{D}_I (i.e., translated with a distance of less than $|\gamma|$), are within the region bounded by dashed circle, which is referred to as the candidate intermediate domain. In the training process, the points in the candidate intermediate domain will be dragged into \mathcal{D}_I . After domain adaptation, $|\gamma|$ becomes small, and the candidate intermediate domain will shrink to near the sphere of \mathcal{D}_I . In this case, most of the examples will be included into \mathcal{D}_I (i.e., $|\gamma_i|$ is near zero), while the ones in \mathcal{D}_S or \mathcal{D}_T but outside the candidate intermediate domain are recognized as

the hard examples that contain too much domain-specific characteristics.

3.3. Gradually Vanishing Bridge on Discriminator

In adversarial domain adaptation, less attention has been paid to the discriminator. Besides, unlike the generator, which could be pretrained on large-scale datasets such as ImageNet [7], the discriminator is usually randomly initialized. Due to the large computation cost and searching space, the discriminator is prone to run into local minimum. Towards more effective discriminator, we propose to build a bridge layer D_2 on the discriminator as shown in dark blue boxes in Figure 2. D_2 models the distance between the basic discriminator D_1 function hyperplane and ideal decision boundary, and provides additive discriminating power to D_1 . Thus the overall discriminator becomes:

$$D_*(G_*(x_i)) = D_1(G_*(x_i)) + D_2(G_*(x_i)). \quad (8)$$

In the two-player game of adversarial learning, it is a common sense that balanced players could stimulate and improve each other more significantly. For example, multiple discriminators [30] may provide domain discrimination from different aspects, but they should be combined properly for more balanced adversarial learning, *e.g.*, combined with a more capable generator. For GVB, the ability of generator is enhanced with the introduction of GVB-G, thus the discriminating ability of the discriminator side should also be adjusted accordingly. Therefore, similar to GVB-G, we progressively minimize the overall range of the bridge on discriminator side as follows:

$$\mathcal{L}_D = \frac{1}{(N_s + N_t)} \sum_{i=1}^{N_s+N_t} |\sigma_i|, \quad (9)$$

where σ_i is the constructed bridge, *i.e.*, $\sigma_i = D_2(G_*(x_i))$. We denote GVB applied to the discriminator as GVB-D.

Towards more balanced minmax game in the whole training process, we combine GVB-G and GVB-D, and formulate the gradually vanishing bridge on generator and discriminator (GVB-GD). The whole framework could be described as follows:

$$\begin{cases} G_*(x_i) = G_2(G_1(x_i)) - G_3(G_1(x_i)) \\ D_*(G_*) = D_1(G_*) - D_2(G_*) \\ \mathcal{L}_{ext} = \lambda \mathcal{L}_G + \mu \mathcal{L}_D, \end{cases} \quad (10)$$

where the extra loss combines both Eqn. 7 and Eqn. 9 with hyper-parameters λ and μ . Since our proposed GVB mechanism focuses on optimization of minmax game, it is general applicable to adversarial domain adaptation framework. Thus we also apply GVB to existing methods including CDAN [24] and Symnets [48]. CDAN multiplies

the feature representations and class responses on the generator, to play against the discriminator. Both GVB-G and GVB-D could be applied to CDAN. Symnets builds symmetric generators for domains, but the discriminator is not modeled explicitly. Therefore, only GVB-G could be applied to SymNets.

4. Experiment

We evaluate the GVB mechanism on three challenging standard benchmarks and compare GVB with state-of-the-art domain adaptation methods.

4.1. Datasets and Settings

Office-31 [34] is a standard benchmark for visual domain adaptation which contains 4,652 images in 31 categories. It consists of three domains: Amazon (A), Webcam (W) and DSLR (D). We evaluate the methods across the three domains, *i.e.*, six transfer tasks in total.

Office-Home [41] is a challenging dataset for visual domain adaptation with 15,500 images in 65 categories. It has four significantly different domains: Artistic images (Ar), Clip Art (Cl), Product images (Pr), and Real-World images (Rw). Among four domains, there are totally 12 challenging domain adaptation tasks.

VisDA-2017 [31] is a simulation-to-real dataset for domain adaptation with over 280,000 images across 12 categories in the training, validation and testing domains. The training images are generated from the same object under different circumstances, while the validation images are collected from MSCOCO. [21].

For GVB-GD, we adopt ResNet-50 [16] pre-trained on ImageNet [7] as our backbone network. We build the bridge layer on both the generator and discriminator with several fully-connected layers in experiments. All the experiments are implemented with PyTorch. For each method, we run four random experiments and report the average classification accuracy and we report the standard deviation in Office-31. We compare our method with state-of-the-art domain adaptation methods. For fair comparison, results on Office-31, Office-Home and VisDA-2017 are directly reported from their original papers if available. Similar to our methods, for CDAN and Symnet, when applying GVB on the generator, ‘-G’ is added to the method name, and ‘-D’ is added when GVB is applied on the discriminator. We compare our methods with state-of-art domain adaptation methods: DAN [23], DANN [9], ADDA [40], GTA [36], CDAN [24], Simnet [32], MCD [35] and MDD [47].

In the adversarial training, we apply gradient reversal layer (GRL) [8] to the network. In GRL, instead of fixing the adaptation factor, progressive training strategy is adopted to suppress noisy signal from the discriminator at the early stages of the training procedure. To optimize the transfer process, we utilize entropy conditioning proposed

Table 1. Accuracies (%) on Office-31 for ResNet50-based unsupervised domain adaptation methods.

Method	A → D	A → W	D → W	W → D	D → A	W → A	Avg
ResNet-50 [16]	68.9±0.2	68.4±0.2	96.7±0.1	99.3±0.1	62.5±0.3	60.7±0.3	76.1
GFK [10]	74.5±0.0	72.8±0.0	95.0±0.0	98.2±0.0	63.4±0.0	61.0±0.0	77.5
DAN [23]	78.6±0.2	80.5±0.4	97.1±0.2	99.6±0.1	63.6±0.3	62.8±0.2	80.4
DANN [9]	79.7±0.4	82.0±0.4	96.9±0.2	99.1±0.1	68.2±0.4	67.4±0.5	82.2
ADDA [40]	77.8±0.3	86.2±0.5	96.2±0.3	98.4±0.3	69.5±0.4	68.9±0.5	82.9
Simnet [32]	85.3±0.3	88.6±0.5	98.2±0.2	99.7±0.2	73.4±0.8	71.8±0.6	86.2
GTA [36]	87.7±0.5	89.5±0.5	97.9±0.3	99.8±0.4	72.8±0.3	71.4±0.4	86.5
MCD [35]	92.2±0.2	88.6±0.2	98.5±0.1	100.0±0.0	69.5±0.1	69.7±0.3	86.5
MDD [47]	93.5±0.2	94.5±0.3	98.4±0.1	100.0±0.0	74.6±0.3	72.2±0.1	88.9
Baseline	89.9±0.3	92.5±0.5	98.5±0.3	99.9±0.1	70.0±0.4	71.1±0.3	87.0
GVB-G	93.9±0.4	94.2±0.4	98.6±0.2	100.0±0.0	71.8±0.3	73.5±0.4	88.7
GVB-D	92.8±0.5	93.9±0.4	98.4±0.2	100.0±0.0	72.0±0.3	72.6±0.3	88.3
GVB-GD	95.0±0.4	94.8±0.5	98.7±0.3	100.0±0.0	73.4±0.3	73.7±0.4	89.3
CDAN [24]	92.9±0.2	93.1±0.1	98.6±0.1	100.0±0.0	71.0±0.3	69.3±0.3	87.5
CDAN-G	92.1±0.3	92.9±0.2	98.2±0.1	100.0±0.0	73.5±0.3	72.8±0.2	88.2
CDAN-D	93.5±0.4	92.9±0.2	98.6±0.1	100.0±0.0	73.0±0.3	73.1±0.3	88.5
CDAN-GD	93.7±0.2	94.0±0.2	98.6±0.1	100.0±0.0	73.4±0.3	73.0±0.2	88.8
Symnets [48]	93.9±0.5	90.8±0.1	98.8±0.3	100.0±0.0	74.6±0.6	72.5±0.5	88.4
Symnets-G	96.1±0.3	93.8±0.4	98.8±0.2	100.0±0.0	74.9±0.4	72.8±0.3	89.4

Table 2. Accuracies (%) on Office-Home for ResNet50-based unsupervised domain adaptation methods.

Method	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-50 [16]	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
DAN [23]	43.6	57.0	67.9	45.8	56.5	60.4	44.0	43.6	67.7	63.1	51.5	74.3	56.3
DANN [9]	45.6	59.3	70.1	47.0	58.5	60.9	46.1	43.7	68.5	63.2	51.8	76.8	57.6
MCD [35]	48.9	68.3	74.6	61.3	67.6	68.8	57	47.1	75.1	69.1	52.2	79.6	64.1
BNM [6]	52.3	73.9	80.0	63.3	72.9	74.9	61.7	49.5	79.7	70.5	53.6	82.2	67.9
MDD [47]	54.9	73.7	77.8	60.0	71.4	71.8	61.2	53.6	78.1	72.5	60.2	82.3	68.1
Symnets [48]	47.7	72.9	78.5	64.2	71.3	74.2	63.6	47.6	79.4	73.8	50.8	82.6	67.2
Symnets-G	48.0	74.3	78.5	65.1	72.2	74.4	65.1	49.4	79.7	73.8	51.7	82.5	67.8
CDAN [24]	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
CDAN-G	53.6	72.1	77.9	61.9	71.8	71.7	63.8	52.9	79.9	71.5	58.0	83.3	68.2
CDAN-D	54.4	72.9	77.7	62.6	71.9	71.8	62.4	52.1	79.5	72.1	58.1	83.3	68.2
CDAN-GD	55.3	74.1	78.2	62.4	72.6	71.8	63.8	54.1	80.1	73.1	58.7	83.6	69.0
Baseline	54.7	72.8	78.5	62.3	71.1	73.1	61.0	53.0	80.0	72.8	56.5	83.4	68.3
GVB-G	56.5	74.0	79.2	64.2	73.0	74.1	65.2	55.9	81.2	74.2	58.2	84.0	70.0
GVB-D	55.0	73.8	79.0	64.3	72.9	75.0	63.4	54.2	80.9	73.1	58.0	83.6	69.4
GVB-GD	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

in [24] that pays more attention to easy-to-classify data. We choose the baseline model as described in Sec. 3.1. For the training of the network, we employ mini-batch stochastic gradient descent (SGD) with momentum of 0.9 to train our model. We keep fixing $\lambda = 1$ and $\mu = 1$ to balance the loss, since our method of GVB is stable under different values of parameters λ and μ .

4.2. Results

The results on Office-31, Office-Home and VisDA-2017 datasets are separately shown in Table 1, Table 2 and Table 3. Both GVB-G and GVB-D outperform the baseline and most of other competitors significantly, which means the bridge provides prominent improvement for domain adaptation. GVB-GD further outperforms GVB-G and

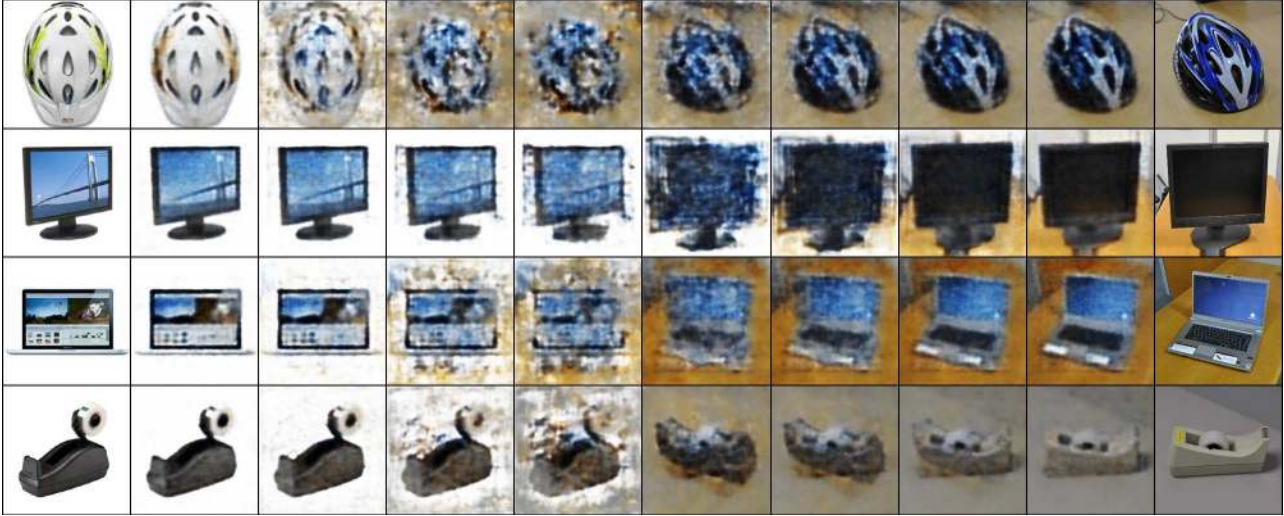


Figure 4. Visualization of data with and without the bridge γ_i on generator on $A \rightarrow D$ for GVB-GD. The images shown in the first and last columns are respectively sampled from source domain (Amazon) and target domain (DSLR). The 2nd~5th columns show the generated source images corresponding to different representations by subtracting γ_i with $0\times$, $5\times$, $10\times$ and $15\times$ amplifications from classifier responses c_i . Similarly, the 6th~9th columns show the generated target images by subtracting γ_i with $15\times$, $10\times$, $5\times$ and $0\times$ amplifications from c_i .

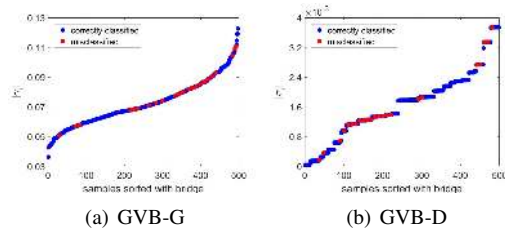


Figure 5. Relationship between the range of bridge on target domain and classification result on GVB-GD.

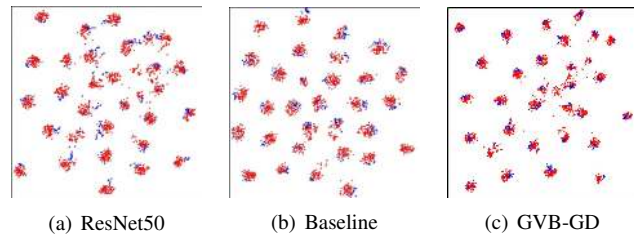


Figure 6. t -SNE of classifier responses by ResNet50, Baseline and GVB-GD (red: Amazon, blue: Webcam).

Table 3. Accuracies (%) of Synthetic \rightarrow Real on VisDA-2017 for unsupervised domain adaptation methods using ResNet-50.

Method	Acc	Method	Acc	Method	Acc
DAN [23]	61.6	CDAN [24]	70.0	Baseline	71.3
DANN [9]	57.4	CDAN-G	73.4	GVB-G	73.1
GTA [36]	69.5	CDAN-D	73.8	GVB-D	72.8
MDD [47]	74.6	CDAN-GD	74.9	GVB-GD	75.3

GVB-D for most tasks, which means the bridge on the generator and discriminator can boost each other to learn better domain-invariant representation. Besides, we apply GVB-G to CDAN and Symnets, and GVB-D to CDAN. The well-performed results of CDAN-G, CDAN-D, CDAN-GD and Symnets-G mean the GVB mechanism is generally applicable for mainstream adversarial domain adaptation methods.

For different datasets on Office-31, Office-Home and VisDA-2017, methods with GVB could achieve well-performed results. Among the datasets, there exists large domain discrepancy in Office-Home and VisDA-2017, while less domain discrepancy in Office. Thus our methods

with GVB are suitable for both the easy and difficult tasks. Besides, VisDA-2017 is a large-scale dataset, while Office-Home and Office are relative small datasets. This means our GVB methods are applicable to both small and large datasets. Meanwhile, GVB could cooperate well with methods such as CDAN and Symnets for different tasks. This shows that GVB is generally applicable for a wide range of domain adaptation tasks.

4.3. Qualitative Validation

Representation visualization. Since we avoid using image reconstruction in the bridge γ on generator, there remain concerns on whether there are rich domain-specific characteristics in γ . We visualize the representations with and without γ in Figure 4 on Office-31. 500 random images from Amazon and 498 images from DSLR are selected to train the network. Similar to FCN [22], we use the encoder-decoder network to achieve the input images reconstruction. The encoder is the same as GVB-GD and the decoder is a deconvolutional neural network that maps classifier responses into images. We utilize the weights of

Table 4. Accuracies (%) on Office-Home with ResNet50. Methods with notion of BG (D) are the bridge without minimization.

Method	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Avg
Baseline	54.7	72.8	78.5	62.3	67.1
BG	54.1	72.3	78.7	62.9	67.0
GVB-G	56.5	74.0	79.2	64.2	68.5
BD	55.3	73.1	79.0	63.9	67.8
GVB-D	55.0	73.8	79.0	64.3	68.0
GVB-G+BD	56.2	74.4	79.6	64.2	68.6
GVB-GD	57.0	74.7	80.0	64.8	69.1

encoder with pre-trained GVB-GD on $A \rightarrow D$ and fix the encoder parameters in the training process. We train the encoder-decoder network to reconstruct images from both domains by minimizing L_2 loss between input image and reconstructed image. As the range of γ is small compared with classification responses c , it is difficult to directly visualize γ . Thus we amplify the γ and construct several representations $r = c - \sigma\gamma$ with $\sigma \in \{0, 5, 10, 15\}$ as the inputs of decoder for better visualization.

From the left to right, we can see the gradual progress from source domain to target domain in Figure 4. Take the first row for example, the body of the bike helmet is white in source domain and blue in target domain. With increasing amplification factor σ , the generated helmets from the 2nd to the 5th column become darker and more blue, while those from the 9th to the 6th column becomes whiter. Besides, the helmet is vertical in source domain and slant in target domain, while the object region of the generated images with amplified γ is in a circle shape which covers the overlapping region of the original images. Furthermore, the background is white in source domain and brown in target domain, and it appears as half white and half brown in the images with amplified γ . All these results validate that γ captures rich domain-specific representations.

Statistical analysis on the bridge. To analyze the relationship between the bridge on target domain and classification results, we calculate the range of the bridge as $|\gamma_i|$ on GVB-GD. In Figure 5, we show the results of $A \rightarrow D$ on DSLR and sort samples according to the range of bridge. In Figure 5(a), with increasing range of the bridge γ_i on generator, the red dots become more intensive and the classification error goes higher. This validates that higher misclassification probability appears with larger $|\gamma_i|$, thus minimizing the range of γ_i tends to result in lower misclassification probability.

On the discriminator side, as observed in Figure 5(b), σ_i tends to be large when the input representations could be easily transformed into intermediate domain, *i.e.*, they tend to be more domain-invariant. These “easy” examples need more discriminative ability from discriminator to distinguish which domain they come from, and they need less

efforts on generator side to produce domain-invariant representations, resulting smaller $|\gamma_i|$. More importantly, these “easy” samples tend to be easily classified with higher probability since they show more domain-invariance. By contrast, samples with larger $|\gamma_i|$ deliver more domain-specific properties, which results in smaller $|\sigma_i|$ and higher misclassification probability.

Ablation study. To validate the effect of GVB compared with the single bridge, we conduct ablation experiments as shown in Table 4. Methods with notion of B-G and G-D are the bridge without minimization, *i.e.*, $\lambda = 0$ and $\mu = 0$. (GVB-G)+(B-D) means to add a bridge on discriminator without minimization to replace the GVB-D, *i.e.*, $\mu = 0$. GVB-G outperforms B-G by a large margin, this means the gradually vanishing mechanism is indispensable on the generator. Results on GVB-D and B-D are similar, but GVB-GD outperforms (GVB)-G+(B-D). This means GVB-D could cooperate with GVB-G better than B-D for a more balanced adversarial training.

Feature visualization. We visualize the features of ResNet50 and our methods on $A \rightarrow D$ by t -SNE [26] in Figure 6. It is shown in Figure 6(a) that ResNet50 works well in source domain (Amazon) but poorly in target domain (Dslr). The Baseline achieves well-performed global alignment by transferring on classifier responses as shown in Figure 6(b), but the direct transfer leads to a large average distance between target data and its nearest source data. For GVB-GD, target samples are located closer to the source examples. The clusters in Figure 6(c) are more compact than the other two cases, which further validates the learning ability of our GVB mechanism.

5. Conclusion

In this paper, we propose the gradually vanishing bridge mechanism that can work on the generator and discriminator in adversarial domain adaptation. On the generator, GVB could result in more domain-invariant representation and better reduction of the negative influence of the rich domain-specific characteristics. To achieve a more balanced adversarial training process, GVB is also built on the discriminator side to provide additive discrimination power. We also apply GVB to existing methods and achieve remarkable improvement over their original counterparts.

Acknowledgement. This work was supported in part by the National Key R&D Program of China under Grant 2018AAA0102003, in part by National Natural Science Foundation of China: 61672497, 61620106009, 61836002, 61931008 and U1636214, and in part by Key Research Program of Frontier Sciences, CAS: QYZDJ-SSW-SYS013. Authors would like to thank Kingsoft Cloud for support of free GPU cloud computing resource and helpful discussion.

References

- [1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223, 2017.
- [2] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *Advances in Neural Information Processing Systems*, pages 343–351, 2016.
- [3] Wei-Lun Chang, Hui-Po Wang, Wen-Hsiao Peng, and Wei-Chen Chiu. All about structure: Adapting structural information across domains for boosting semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1900–1909, 2019.
- [4] Qingchao Chen, Yang Liu, Zhaowen Wang, Ian Wassell, and Kevin Chetty. Re-weighted adversarial adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7976–7985, 2018.
- [5] Yi-Hsin Chen, Wei-Yu Chen, Yu-Ting Chen, Bo-Cheng Tsai, Yu-Chiang Frank Wang, and Min Sun. No more discrimination: Cross city adaptation of road scene segmenters. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1992–2001, 2017.
- [6] Shuhao Cui, Shuhui Wang, Junbao Zhuo, Liang Li, Qingming Huang, and Tian Qi. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations. *arXiv preprint arXiv:2003.12237*, 2020.
- [7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. 2009.
- [8] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. *arXiv preprint arXiv:1409.7495*, 2014.
- [9] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030, 2016.
- [10] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2066–2073. IEEE, 2012.
- [11] Rui Gong, Wen Li, Yuhua Chen, and Luc Van Gool. Dlow: Domain flow for adaptation and generalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2477–2486, 2019.
- [12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [13] Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Domain adaptation for object recognition: An unsupervised approach. In *2011 international conference on computer vision*, pages 999–1006. IEEE, 2011.
- [14] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(Mar):723–773, 2012.
- [15] Philip Haeusser, Thomas Frerix, Alexander Mordvintsev, and Daniel Cremers. Associative domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2765–2773, 2017.
- [16] 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.
- [17] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei A Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. *arXiv preprint arXiv:1711.03213*, 2017.
- [18] Lanqing Hu, Meina Kan, Shiguang Shan, and Xilin Chen. Duplex generative adversarial network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1498–1507, 2018.
- [19] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4893–4902, 2019.
- [20] Guoliang Kang, Liang Zheng, Yan Yan, and Yi Yang. Deep adversarial attention alignment for unsupervised domain adaptation: the benefit of target expectation maximization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 401–416, 2018.
- [21] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [22] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [23] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I Jordan. Learning transferable features with deep adaptation networks. *arXiv preprint arXiv:1502.02791*, 2015.
- [24] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *Advances in Neural Information Processing Systems*, pages 1647–1657, 2018.
- [25] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2208–2217. JMLR.org, 2017.
- [26] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.
- [27] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

- [28] Jie Ni, Qiang Qiu, and Rama Chellappa. Subspace interpolation via dictionary learning for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 692–699, 2013.
- [29] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, 2011.
- [30] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [31] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge, 2017.
- [32] Pedro O Pinheiro. Unsupervised domain adaptation with similarity learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8004–8013, 2018.
- [33] Paolo Russo, Fabio M Carlucci, Tatiana Tommasi, and Barbara Caputo. From source to target and back: symmetric bi-directional adaptive gan. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8099–8108, 2018.
- [34] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *European conference on computer vision*, pages 213–226. Springer, 2010.
- [35] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3723–3732, 2018.
- [36] Swami Sankaranarayanan, Yogesh Balaji, Carlos D Castillo, and Rama Chellappa. Generate to adapt: Aligning domains using generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8503–8512, 2018.
- [37] Kihyuk Sohn, Sifei Liu, Guangyu Zhong, Xiang Yu, Ming-Hsuan Yang, and Manmohan Chandraker. Unsupervised domain adaptation for face recognition in unlabeled videos. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3210–3218, 2017.
- [38] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European Conference on Computer Vision*, pages 443–450. Springer, 2016.
- [39] Eric Tzeng, Judy Hoffman, Trevor Darrell, and Kate Saenko. Simultaneous deep transfer across domains and tasks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4068–4076, 2015.
- [40] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7167–7176, 2017.
- [41] Hemant 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.
- [42] Riccardo Volpi, Pietro Morerio, Silvio Savarese, and Vittorio Murino. Adversarial feature augmentation for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5495–5504, 2018.
- [43] Shuhui Wang, Shuqiang Jiang, Qingming Huang, and Qi Tian. Multi-feature metric learning with knowledge transfer among semantics and social tagging. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2240–2247. IEEE, 2012.
- [44] Hongliang Yan, Yukang Ding, Peihua Li, Qilong Wang, Yong Xu, and Wangmeng Zuo. Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2272–2281, 2017.
- [45] Jun Yang, Rong Yan, and Alexander G Hauptmann. Cross-domain video concept detection using adaptive svms. In *Proceedings of the 15th ACM international conference on Multimedia*, pages 188–197, 2007.
- [46] Yang Zhang, Philip David, and Boqing Gong. Curriculum domain adaptation for semantic segmentation of urban scenes. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2020–2030, 2017.
- [47] Yuchen Zhang, Tianle Liu, Mingsheng Long, and Michael Jordan. Bridging theory and algorithm for domain adaptation. In *International Conference on Machine Learning*, pages 7404–7413, 2019.
- [48] Yabin Zhang, Hui Tang, Kui Jia, and Mingkui Tan. Domain-symmetric networks for adversarial domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5031–5040, 2019.
- [49] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232, 2017.
- [50] Junbao Zhuo, Shuhui Wang, Shuhao Cui, and Qingming Huang. Unsupervised open domain recognition by semantic discrepancy minimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 750–759, 2019.
- [51] Junbao Zhuo, Shuhui Wang, Weigang Zhang, and Qingming Huang. Deep unsupervised convolutional domain adaptation. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 261–269. ACM, 2017.