Deep Visual Unsupervised Domain Adaptation for Classification Tasks

A Survey

Madadi, Yeganeh; Seydi, Vahid; Nasrollahi, Kamal; Hossieni, Reshad; Moeslund, Thomas B.

Published in:
IET Image Processing

DOI (link to publication from Publisher):
10.1049/iet-ipr.2020.0087

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

IET Research Journals

Submission Template for IET Research Journal Papers

ISSN 1751-8644
doi: 0000000000
www.ietdl.org

Deep Visual Unsupervised Domain Adaptation for Classification Tasks: A Survey

Yeganeh Madadi[1,2,3], Vahid Seydi[1*], Kamal Nasrollahi[3,4], Reshad Hosseini[2], Thomas B. Moeslund[3]

[1] Department of Computer Engineering, Faculty of Technical and Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran
[2] University of Tehran, Tehran, Iran
[3] Aalborg University, Aalborg, Denmark
[4] Milestone Systems A/S, Copenhagen, Denmark
* E-mail: vahidseydi@gmail.com

Abstract: Learning methods are challenged when there is not enough labeled data. It gets worse when the existing learning data have different distributions in different domains. To deal with such situations, deep unsupervised domain adaptation techniques have newly been widely used. This paper surveys such domain adaptation methods that have been used for classification tasks in computer vision. The survey includes the very recent papers on this topic that have not been included in the previous surveys and introduces a taxonomy by grouping methods published on unsupervised domain adaptation into five groups of: discrepancy-, adversarial-, reconstruction-, representation-, and attention-based methods.

1 Introduction

By exploiting massive labeled data, deep NNs have shown improved performance in many applications, like image classification, object detection, semantic segmentation, text recognition, person re-identification, to name a few. The performance of these systems highly depends on the qualification of the labeled training data. The major assumption here is that the training and testing data have independent and identical distributions. This assumption can, however, be easily challenged on differences of illumination, pose, quality, background, etc, between the domains.

If the training (labeled) data is not sufficient, one could use domain adaptation techniques to transfer the knowledge a model has gained on a domain with enough labeled data to a domain with limited labeled data, even when the source and target domains are of different distributions. Labeling however is also a time and resource consuming process. This survey therefore focuses on deep unsupervised domain adaptation methods that have been utilized for classification purposes in computer vision.

There are few survey papers already published on transfer learning and domain adaptation [1–6]. However, they are not as comprehensive as this survey paper. Here we review the focus areas of these already published survey papers. Pan and Yang [1] introduced the first survey on transfer learning. They compared traditional machine learning with different kinds of transfer learning techniques then they categorized transfer learning techniques into three groups: inductive, transductive, and unsupervised. Shao et al. [2] reviewed transfer learning for visual classification tasks and categorized these techniques into instance-based transfer learning and parameter-based transfer learning. Patel et al. [3] focus was on reviewing a representative subset of the computer vision. The survey presented by Csurka [4] was about domain adaptation methods in visual applications covering both non-deep and deep domain adaptation. She investigated the state-of-the-art non-deep domain adaptation approaches, and then briefly expressed the deep domain adaptation ones, and categorized them into three loss-models: classification, discrepancy, and adversarial. Tan et al. [5] focused more on deep methods in transfer learning. Wang and Deng [6] extended Csurka's work. they added concepts of one-step and multi-step domain adaptation methods and categorized them into hand-crafted based, feature-based, and finally representation-based approaches. In other words, surveys [1, 2] covered only techniques on non-deep transfer learning and [5] extended this topic to deep ones. [3] focused only on non-deep domain adaptation on visualization tasks and [4][6] tried to extend their work on domain adaptation into deep methods.

The main contributions of this survey in comparison to other related surveys are as follows:

- There are many recent papers on deep visual unsupervised domain adaptation approaches, that are not mentioned in any of the previous surveys but are included in our paper.
- This survey paper represents a comprehensive coverage of deep methods for domain adaptation, while previous surveys were mostly focused on non-deep methods and have mentioned deep methods only briefly.
- This survey paper presents a new taxonomy for deep visual UDA for classification tasks. This taxonomy is useful because, it covers almost all existing techniques to solve UDA problem, which are categorized into five main groups based on the technology of adopted for domain adaptation. The first group is discrepancy-based which consists of used techniques for decreasing the difference between the domains and making more similarity between data distributions by utilizing statistical techniques (i.e. *maximum mean discrepancy*, *correlation alignment*, *entropy minimization*, *batch normalization*, *moment matching*, and *wasserstein discrepancy*). The second group is adversarial-based which consists of used techniques for minimizing the distribution difference across domains by using an adversarial objective with a domain discriminator through assuming that the source labels are equivalent to the target labels or not (i.e. *partial adversarial networks*, and *non-partial adversarial networks* with three subsetting: *discriminative adversarial networks*, *generative adversarial networks*, and *feature matching adversarial networks*). The third group is reconstruction-based which consists of used techniques for decreasing the difference between the domains by mapping the source and target, or both domain samples into a shared representation domain (i.e. *encoder-decoder models*, *dictionary and sparse coding models*, and *graph-based models*). The fourth group is representation-based which consists of used techniques for decreasing the difference between the domains by utilizing the trained network as input to use intermediate representations to a new network (i.e. *domain confusion representation*, *domain invariant representation*, and *representation disentangling*).

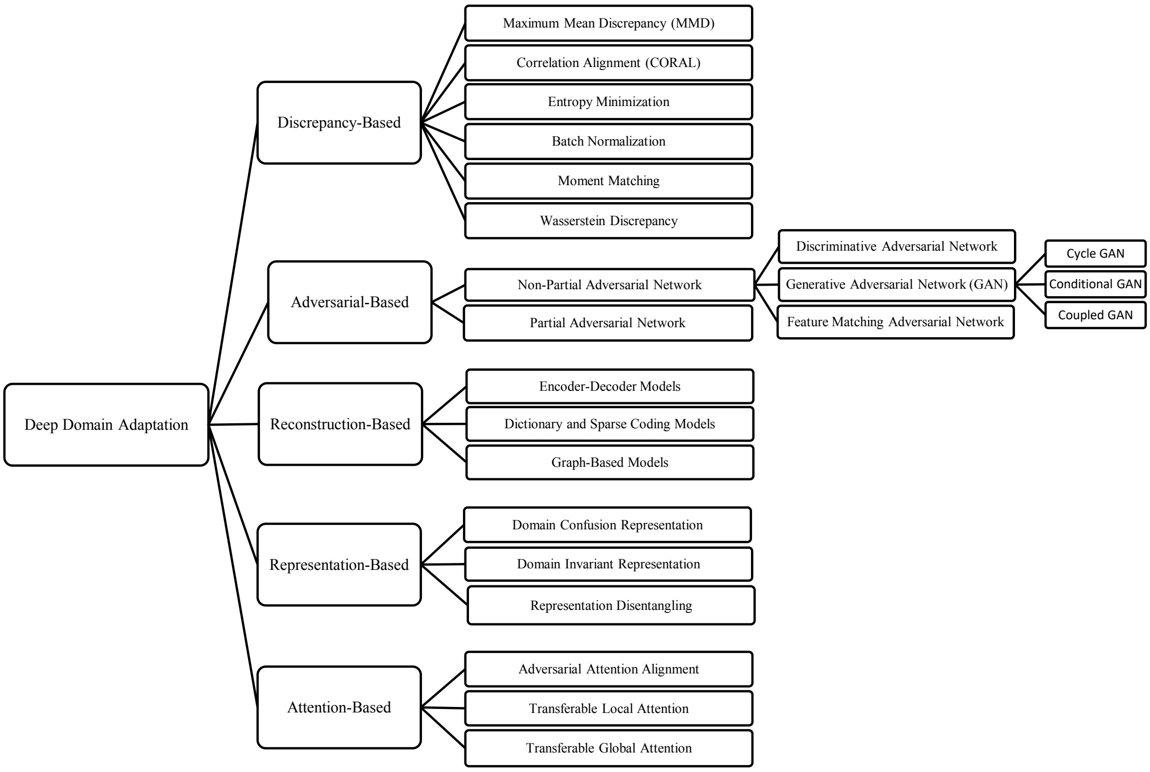


Fig. 1: Our taxonomy of deep visual unsupervised domain adaptation for classification tasks.

The fifth group is attention-based which consists of used techniques for decreasing the difference between the domains by focusing on some transferable attention regions or images from source data and relating them to the target data (i.e. *adversarial attention alignment*, *transferable local attention*, *transferable global attention*).
• We investigate and analyze some important methods in each category of our taxonomy, based on the results reported by different methods on well-known public databases. To the best of our knowledge, this is the first survey paper in deep visual unsupervised domain adaptation for classification tasks that quantitatively compares the performance of different deep UDA methods. It can help to provide proper insight for designing accurate and robust deep UDA methods.

The rest of this paper is organized as follows: In Section 2, our new taxonomy of deep domain adaptation methods is introduced. In the next five sections (Section 3-Section 7), different methods of our taxonomy including discrepancy-based, adversarial-based, reconstruction-based, representation-based, attention-based, are discussed, respectively. In Section 8, the most important benchmark datasets for deep visual domain adaptation are presented. Then, some unsupervised domain adaptation methods for other applications during the recent two years are mentioned in Section 9. Finally, the discussion and summary of this paper are presented in Sections 10 and 11, respectively.

2 Taxonomy

In this section, we introduce our taxonomy of deep visual unsupervised domain adaptation for classification tasks. As it is shown in the Fig. 1, we categorize deep domain adaptation methods into five main groups based on the technology of these methods adopted:

1. Discrepancy-based: These methods measure the distance between the source and target domains on the corresponding activation layers of the two networks and apply statistical techniques to diminish discrepancy between domains.
2. Adversarial-based: These methods are one of the novel research topics in machine learning approaches. These techniques contain deep NN architectures comprising of two competing networks.
3. Reconstruction-based: These methods map source/target/or both domain samples into a shared representation domain(s), while preserving individual characteristics of each domain.
4. Representation-based: These methods utilize the trained network as input to use intermediate representations to a new network.
5. Attention-based: These methods focus on some transferable attention regions or images from source data and relate them to the target data. In other words, it is introduced as the mechanism of guiding the network to focus on the particular parts of target images that contain related information to the source images.

In the following, we elaborate on the difference between the subgroups.

Regarding the taxonomy of discrepancy-based methods, we categorize this group into six subgroups: maximum mean discrepancy (MMD), correlation alignment (CORAL), entropy minimization (EM), batch normalization (BN), moment matching, and Wasserstein discrepancy. The MMD is a kernel method which measures the difference of two-sample in the RKHS, while the CORAL matches the second-order statistics (covariances) between the data distributions by a linear transformation. The entropy minimization (EM) calculates the difference between two probability distributions where

IET Research Journals, pp. 1–18

encourages both low entropy and consistency on the domain predictions for perturbations of the same input features in order to perturbation to be both consistent. The batch normalization (BN) in the form of the BN layer(s) tries to decrease the covariance shift. first, each feature is standardized in a mini-batch, and then a common bias and slope are learned for each mini-batch. The moment matching employs one or more moments to compute of difference between domains feature distribution. The Wasserstein metric is applied as a discrepancy measure for measuring the distance among the different domains samples.

Regarding the taxonomy of adversarial-based methods, we first categorize this group into two subgroups: non-partial adversarial network, and partial adversarial network. The non-partial adversarial methods usually assume that the source labels are equivalent to the target labels, so there is a restriction of identical labels between domains, but the partial adversarial methods relax the same label space assumption between source and target spaces. In other word, the target labels are supposed to be a subset of the source labels. Then, we categorize non-partial adversarial network into three subgroups: discriminative adversarial network, generative adversarial network (GAN), and feature matching adversarial network. Discriminative adversarial network refers to some methods which usually utilize an adversarial scheme for obtaining domain confusion concerning a domain discriminator. However, the GANs are not very optimal on discriminative purposes and maybe tangled in smaller domain shifts while discriminative adversarial methods with applying the shared weights can handle the larger domain shifts. Feature matching adversarial networks are adversarial methods where the source classifier is trained with the labeled data from the source domain, and the target classifier is regularized by minimizing a distance metric between the source classifier and the target one by using all the data. The main difference between feature matching adversarial networks with the discriminative adversarial networks is utilizing distance metrics to improve more efficiency for domain adaptation.

Regarding the taxonomy of reconstruction-based methods, we categorize this group into three subgroups: encoder-decoder models, dictionary and sparse coding models, and graph-based models. Encoder-decoder models map source/target/or both domain samples into a shared representation by using the encoder, while dictionary-based models do by updating or adapting the dictionary, and finally, graph-based models do by building the connected graph.

Regarding the taxonomy of representation-based methods, we categorize this group into three subgroups: domain confusion representation, domain invariant representation, and representation disentangling. The key difference between these groups is on the learning process and loss function. The domain confusion representation method utilizes the domain confusion loss to learn a domain invariant representation. Domain confusion loss seeks to learn domain invariance via finding a representation in which the best domain classifier performs poorly. So this method uses the maximization strategy on the loss function. While domain invariant representation method applies the source and the target data to learn a common representation and utilizes the embedding loss for enforcing prediction and structural consistency on the target data. So, this method uses the minimization strategy for loss function. Representation disentangling is a learning method for concluding a hidden feature space that decomposes the derived representation so that the visual attributes can be recognized and described.

Regarding the taxonomy of attention-based methods, we categorize this group into three subgroups: adversarial attention alignment, transferable local attention, and transferable global attention. Adversarial attention alignment method reviews transferable attention methods which uses adversarially learning. Transferable local attention method specifies which regions of images are better to transfer from the source domain to the target domain. While transferable global attention specifies which images are better to transfer from the source domain to the target domain that leads to better performance in domain adaptation.

The various methods for above-mentioned groups are discussed in the following subsections.

3 Discrepancy Based Methods

Discrepancy based methods are usually used to decrease the difference between the domains and to make more similarity between data distributions. These methods measure the difference between the source and the target domains on corresponding activation layers of the two networks. Discrepancy based methods can be further divided into subgroups shown in Table 1. We review these methods in the following subsections.

Table 1 Our Taxonomy of Discrepancy Based Methods

Maximum Mean Discrepancy	DDC[7], DAN[8], RTN[9], JAN[10], WMMD[11], [12], [13], MRAN[14], CAN[15], SCA[16]
Correlation Alignment	Deep CORAL[17], DGCAN[18], MCA[19], JDDA[20], CORAL&MMD[21]
Entropy Minimization	FTN[22], [23], [24]
Batch Normalization	[25], AdaBN[26]
Moment Matching	[27], M^3SDA[28], CV-CMD[29]
Wasserstein Discrepancy	WDGRL[30], JDOT[31], [32], SWD[33]

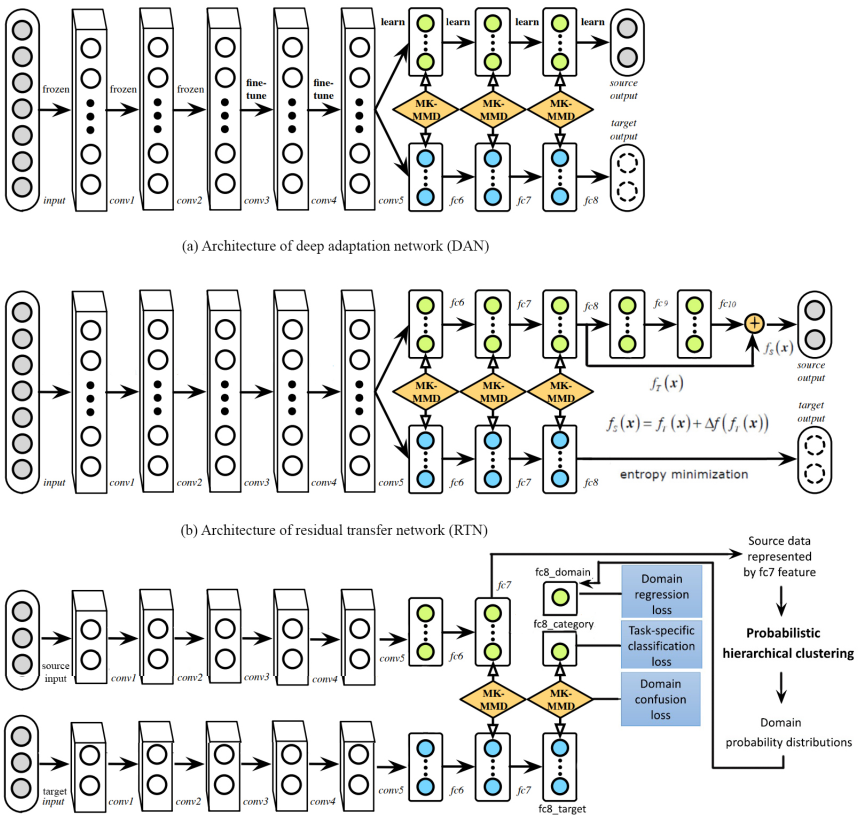
3.1 Maximum Mean Discrepancy (MMD)

Maximum Mean Discrepancy (MMD) [34] is a kernel method which measures the difference of two-sample in the RKHS. The MMD assumes that if generating distributions are the same then all their statistics are equal. MMD can be defined as the following difference discrepancy measure in Equation (1):

$$D(P(X_s), P(X_t)) \triangleq \sup_{f \in H} ||E_{X_s}[f(X_s)] - E_{X_t}[f(X_t)]||_H^2 \quad (1)$$

where D mentions distance between source $P(X_s)$ and target $P(X_t)$ distributions, and f represents the kernel function.

Based on the above, Tzeng et al. [7] proposed a CNN architecture, DDC, which contains a confusion loss function based on MMD, and a new layer for adaptation. Long et al. [8] proposed a Deep Adaptation Network (DAN) model, which generalizes deep CNN in the domain adaptation approaches. Different from previous work, DAN uses of two-sample test statistics, multi-kernel MMD and Mean Embedding Test (ME), for matching the domain distributions [35, 36]. Fig. 2(a) shows the DAN architecture. Long et al. introduced both additional works based on DAN, Residual Transfer Network (RTN) [9] and Joint Adaptation Networks (JAN) [10]. RTN simultaneously learns classifiers and features, and relaxes the common classifier criteria used in DAN [8], and assumes a residual layer for classifier adaptation (see Fig. 2(b)). JAN learns the network by aligning the joint distributions by using the joint MMD (JMMD) criteria. A weighted MMD (WMMD) is proposed by Yan et al. [11], which constructs a reference source distribution relying on target one to reduce the effect of class weight bias. Luo et al. [12] in their work to reduce the discrepancy between domains and to ensure the generalization ability of their model, constrained the MMD metric between domains. Wu et al. [13]'s work pointed out that changing deep representations in CNNs from generic to specific makes it more difficult to transfer their knowledge, specifically in higher fc layers. To deal with this, they introduced a domain confusion loss represented by MMD and placed it on fc layers (see Fig. 2(c)). Zhu et al. [14] proposed Multi Representation Adaptation Network (MRAN) to match the distributions of multiple representations by utilizing a hybrid structure for visual classification tasks. They presented a multi representation alignment and extended the marginal distribution discrepancy measure MMD to conditional MMD (CMMD). Kang et al. [15] presented the Contrastive Adaptation Network (CAN) for optimizing a discrepancy metric, Contrastive Domain Discrepancy (CDD) metric based on MMD. They jointly optimized the intra-class distance and inter-class distance for improving the adaptation performance. They stated that CDD compared to MMD is more robust to the noise especially when working on a large amount of data. Deng et al. [16] introduced Similarity-Constrained



(a) Architecture of deep adaptation network (DAN)

(b) Architecture of residual transfer network (RTN)

(c) Architecture of the joint learning of multiple latent domains and deep representation for domain adaptation

Fig. 2: Different approaches with the MMD metric. (a) Architecture of deep adaptation network (DAN) [8], (b) Architecture of residual transfer network (RTN) [9], and (c) Architecture of the joint learning of multiple latent domains and deep representation for domain adaptation [13]

Alignment (SCA) model which aligns data distributions at both domain-level and class-level. For aligning domain-level, they utilized the JMMD metric [10] which reduces the discrepancy in the joint distributions in the activation layers. For aligning class-level, they utilized Similarity Guided Constraint (SGC) [37, 38] on both source and target domains to achieve intra-class compactness and inter-class separability.

3.2 *Correlation Alignment (CORAL)*

In this section, we represent methods based on CORAL [39]. CORAL is an unsupervised domain adaptation approach via linear transformation for matching the second-order statistics (covariances) between the data distributions.

Sun and Saenko [17] extended CORAL for learning a nonlinear projection that utilizes deep NNs (Deep CORAL) to align correlations in the activation layer. In this model, the CORAL loss is applied to the last layer of AlexNet to use as a distance metric between the covariances of different domain data features. Coral loss is shown in Equation (2).

$$L_{CORAL} = \frac{1}{4d^2}||C_s - C_t||_F^2 \tag{2}$$

where $||.||_F^2$ is the Frobenius norm. C_s and C_t are the covariance matrices for the source and target data, respectively.

Peng and Saenko [18] proposed Deep Generative Correlation Alignment Network (DGCAN), which utilizes shape-preserving loss function for combining both fake and real images data, and utilizes a CORAL loss function for minimizing the domain discrepancy in deep features. Zhang et al. [19] introduced a model, Mapped Correlation Alignment (MCA), which projects covariances of different domains from Riemannian manifold to RKHS. In order to align the distributions, a non-linear transformation is learned by augmenting MCA loss to the classification loss. Chen et al. [20] introduced Joint Discriminative Domain Alignment (JDDA), a domain discrepancy loss which is measured by CORAL and applies a discriminative loss on the bottleneck layer. They jointly learned both instance-based and center-based discriminative learning scheme for deep domain adaptation. Other work on this topic [21] presents an unsupervised deep domain adaptation method based on CORAL and MMD. This method by jointly utilizing MMD and CORAL loss layers in the last two layers of source network and target network, aligns the second-order statistics and higher-order statistics, respectively.

3.3 *Entropy Minimization*

Entropy minimization [40] is a popular training objective in unsupervised domain adaptation. Also, it can be used as a distance metric adaptation between different domains.

Sohn et al. [22] proposed Feature Transfer Network (FTN) to divide the transformed source domain and target domain using a entropy minimization loss function to enhance the discriminative ability of FTNs in the target domain. Roy et al. [23] presented Min-Entropy Consensus (MEC), where the objective function jointly merges consistency loss and entropy loss to improve the domain adaptation, as in (3):

$$L^t(B_1^t, B_2^t) = \frac{1}{m}\sum_{i=1}^{m} l^t(x_i^{t_1}, x_i^{t_2}) \quad (3)$$

$$l^t(x_i^{t_1}, x_i^{t_2}) = -\frac{1}{2}\max_{y\in Y}(\log p(y|x_i^{t_1}) + \log p(y|x_i^{t_2})) \quad (4)$$

where $x_i^{t_1} \in B_t^1$ and $x_i^{t_2} \in B_t^2$ and also B_t^1 and B_t^2 are two different target batches that contain images with duplicate pairs in various image perturbations.

Based on previous work in the domain discovery [25], Mancini et al. [24] produced multiple domain predictions on perturbations of the features of given samples using MEC loss of [23]. This loss performs both consistency and low entropy for the perturbed domain predictions of the same input features.

3.4 Batch Normalization

Batch Normalization (BN) [41] in form of BN layer(s) was originally introduced to decrease the covariance shift.

Mancini et al. [25] introduced an CNN architecture with a new batch normalization layer (mDA-layer) for domain adaptation. The Multi-domain DA layer (mDA-layer) re-normalizes the multi-modal feature distributions. This layer exploit domain membership information to match the distributions. The mDA-layer can be normalized according to (5):

$$mDA(x_i, w_i, \hat{\mu}, \hat{\sigma}) = \sum_{d\in D} w_{i,d}\frac{x_i - \hat{\mu}_d}{\sqrt{\hat{\sigma}_d^2 + \epsilon}} \quad (5)$$

where $w_i = [w_{i,d}]_{d\in D}$, $\hat{\mu} = [\hat{\mu}_d]_{d\in D}$, and $\hat{\sigma} = [\hat{\sigma}_d^2]_{d\in D}$.

Adaptive Batch Normalization (AdaBN) [26], was introduced to improve the generalization power of a DNN. The AdaBN modifies the statistics of BN layers in the target domain by statistics of each BN layer in the source domain to update the weights in CNN for domain adaptation purposes.

3.5 Moment Matching

Moment matching is another approach to decrease the discrepancy between different domains. this technique employs one or more moments to compute of difference between domains feature distribution.

Li et al. [27] introduced Generative Moment Matching Networks (GMMNs). Training the GMMN for minimizing the distribution discrepancy is done via aligning all distribution moments of the model. The GMMN uses MMD as a loss function that is the main part of training objective for this method. Zellinger et al. [29] introduces a metric, Central Moment Discrepancy (CMD), which is the sum of discrepancies of higher-order central moments of the domain distributions. Peng et al. [28] proposed a moment matching approach, M^3SDA, for multi-source domain adaptation models which not only aligns the source domains with target domain but also source domain with each other simultaneously.

3.6 Wasserstein Discrepancy

The Wasserstein metric is applied as a discrepancy measure between the task-specific classifiers for measuring the distance among the different domains samples [42]. Equation (6) expresses Wasserstein Metric.

$$W_r(P(X_s), P(X_t)) = [\inf_{\mu\in\Gamma(P(X_s),P(X_t))} \int \rho(x,y)^r d\mu(x,y)]^{\frac{1}{r}} \quad (6)$$

where $P(X_s), P(X_t)$ are probability distributions source and target on *X*, respectively, and W_r is the Wasserstein distance of order *r* between $P(X_s)$ and $P(X_t)$. Also $\rho(x, y)$ is a distance function for two samples x and y in the set *X*.

Damodaran et al. [32] improved JDOT [31] by jointly matching feature and label space distributions in a DNN layer. Lee et al. [33] presented Sliced Wasserstein discrepancy (SWD), which utilizes the geometrically 1-Wasserstein as the discrepancy measure for obtaining the dissimilarity probability of source and target domains.

4 Adversarial Based Methods

Recently, adversarial based adaptation methods have become an expanding important type of domain adaptation method to deal with visual domain adaptation problems, which intends to minimize distribution difference across domains by using an adversarial objective with a domain discriminator. Our taxonomy of adversarial based models is summarized in Table 2, and the recent works on this topic are reviewed in the following subsections.

Table 2 Our Taxonomy of ADVERSARIAL BASED METHODS

Non-partial Adversarial Based Methods	
Generative Adversarial Network (GAN)[43]	[44], S+U[45], AdvKin[46], AdvNet[47], DupGAN[48], GAADGAN[49], GAGL[50], DR-GAN[51], GADM[52] **CycleGAN**[53] : DiscoGAN[54], ACAL[55], CyCADA[56] **CGAN**[57] : DualGAN[58], DR-GAN[59], AC-GAN[60], FF-GAN[61], PixelDA[62], DAGAN[63], [49], [64], CAPG-GAN[65], SketchGAN [66] **CoGAN**[67] : UNIT [68], TarGAN[69]
Discriminative Adversarial Network	DANN[70], ADDA[71], MADA[72], MDANN[73]
Feature Matching Adversarial Network	RevGrad[74], DSN[75], [76], WDGRL[30], CDANs[77], [78, 79], CDAN[77], CAN[80], DCTN[81], [82], RAAN[83], SimNet[84], [85], H+L[86], ARTN[87], M-ADDA [38], HAN [88]
Partial Adversarial Based Methods	
Partial Adversarial Networks	SAN[89], PADA[90], IWAN[91], ETN[92]

4.1 Non-partial Adversarial Based Methods

Non-partial adversarial methods usually assume that the source labels are equivalent to the target labels, So there is a restriction of identical labels between domains. In this section, we categorize non-partial adversarial based methods as follows:

4.1.1 Generative Adversarial Network (GAN): Generative Adversarial Networks (GANs) provide an approach to learn deep representations without extensive training data. This approach is executed by deriving backpropagation through a competition between the networks. A GAN model contains two parts: a discriminator and a generator. The generator learns to generate artificial samples that are hard to distinguish from the real ones, while the discriminator learns to perform this distinction. The GAN architecture is illustrated in Fig. 3(a).

GAN was introduced by Goodfellow et al. [43], by defining a pair of networks that are competing against each other. GAN contains two parts: a generator *g*, which obtains the distribution of data, and a discriminator *f*, which computes the probability of a sample is a real training data or is a fake one generated by *g*. Both the generator and discriminator have fully connected NNs architecture. GANs can be seen as a minimax two-player game by solving the following optimization Equation (7):

$$\min_g \max_f V(f,g) = E_{x\sim p(x)}[-\log f(x)] + E_{z\sim p_z(z)} [\log(1 - f(g(z)))] \quad (7)$$

where *z* is a random vector as input to the generative model.

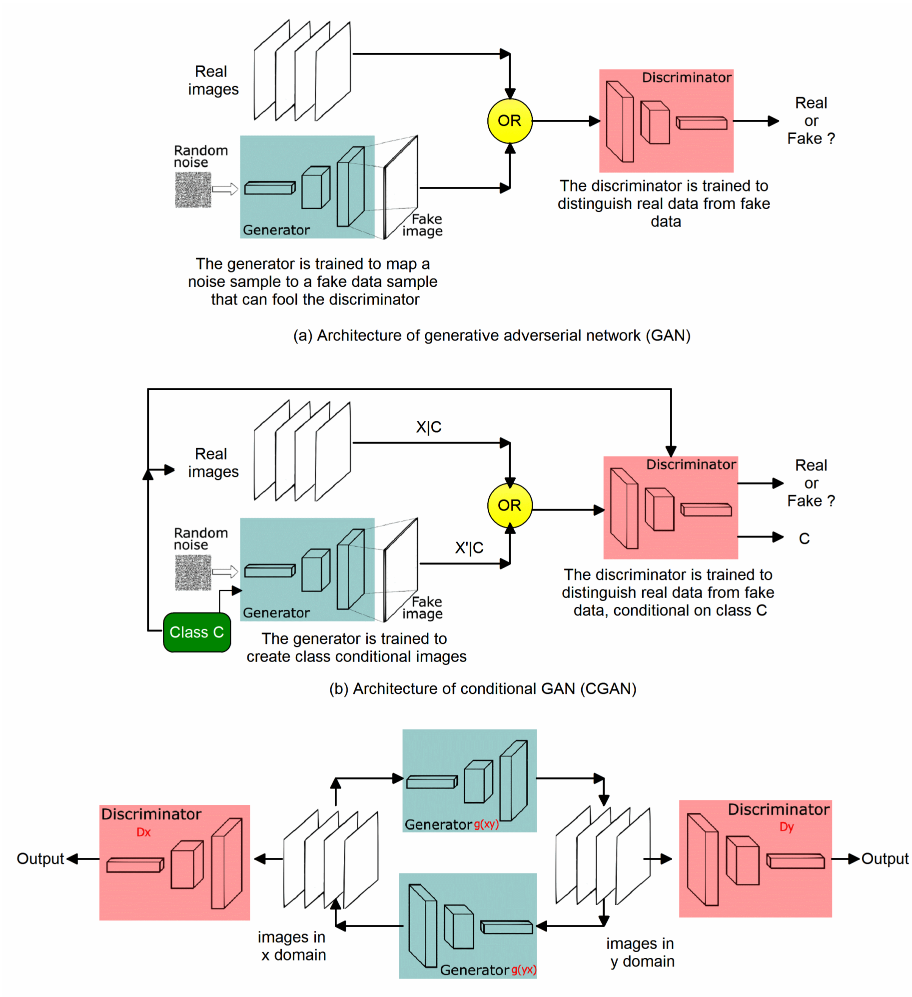
Yoo et al. [44] used GANs to transfer information from the source domain to the pixel-level target domain. They measured a pixel-level similarity via the color version of structural similarity. Shrivastava et al. [45] developed a model for simulated + unsupervised (S + U) learning which combines two losses: self-regularization and adversarial. Also, they trained a refiner network for adding realism to synthetic images. In contrast to GANs which use random vectors as input, this model uses synthetic images for input data. Duan et al. presented two adversarial contrastive methods, AdvKin [46] and AdvNet [47], which use GAN for kinship verification issue. Advkin utilizes MMD loss to decrease distributions discrepancy and an adversarial loss based on GAN to further improve the discrimination and robustness in deep features. Similar to Advkin, AdvNet not only utilizes MMD and GAN, but also applies a contrastive loss for increasing the inter-class distance and minimizing the intra-class distance in the second fc layer. Tran et al. [51] proposed a Disentangled Representation GAN (DR-GAN) by modeling the face rotation process. DR-GAN is the first work that is able to frontalize an extreme pose in the wild face and also constructs the generator in GAN for feature learning with an encoder-decoder structure. Wei et al. [50] proposed a Generative Adversarial Guided Learning (GAGL) model, which is a classification model to learn the decision boundary through the generator. Firstly, They learned a domain-invariant classification model via adversarial training, and then, they introduced an additional generative model to further enlarge the decision boundary of the classification model from the target data. Hu et al. [48] proposed a GAN architecture with two adversarial discriminators, DupGAN. It utilizes a generator, an encoder, and duplex adversarial discriminators for domain-invariant feature extraction and domain transformation. Zhang et al. [52] introduced Generative Adversarial Distribution Matching (GADM), which consists of three stages: source domain pre-training by using the labeled source image data, adversarial distribution matching across domains via augmenting the MMD to the objective function of the generator to reduce distribution discrepancy across domains, and target domain classification by using target mapping and target classification model.

Cycle Generative Adversarial Network (CycleGAN): CycleGAN [53] is a framework based on GANs, which utilizes cycle-consistency constraint along with adversarial training manner for mapping inputs data from source domain to target. The advantage of CycleGAN is that it does not need the paired input-output instances in the domains to match. It can learn the relations between the domains [93]. The CycleGAN architecture is illustrated in Fig. 3(c). Hoffman et al. [56] proposed a Cycle-Consistent Adversarial Domain Adaptation model, CyCADA, which is an extension of CycleGAN that discriminatively trains cycle-consistency loss and aligns representations at both the feature-level and pixel-level. Hosseini-Asl et al. [55] proposed two models, Relaxed Cycle-Consistent model (RCAL) and extended of RCAL, Augmented Cycle-Consistent Model (ACAL). They obtained better accuracy compared to CycleGAN, by relaxing the cycle-consistency constraint and integrating the discriminator in the training phase. Kim et al. [54] proposed DiscoGAN based on cycleGAN, which preserves major properties between the input image and the translated image by applying the cycle consistency loss function. The core of DiscoGAN is based on two different GANs. A key intuition of DiscoGAN is to put all data in one of the domains to be representable through data in another domain. DualGAN [58], CycleGAN [53], and DiscoGAN[54] apply two generators to generate photo-sketch conversion and viceversa.

Conditional Generative Adversarial Network (CGAN): Conditional Generative Adversarial Net (CGAN) [57] is based on GANs which extends to a conditional model. In CGAN, the generative model and discriminative model are conditioned on data and class labels. The CGAN architecture is illustrated in Fig. 3(b). DualGAN[58] adopts deep domain adaptation with using of two GANs to develop dual learning by adversarial reconstruction. The mean of absolute difference between the reconstructed and original data within each domain data is a reconstruction error in the DualGAN model. DR-GAN [59] gains domain-invariant feature extraction and image translation based on CGAN with one encoder in their architectures for all domains. Bousmalis et al. [62] presented pixel-level domain adaptation (PixelDA), for adopting source data to the target data. Previous works usually perform both image classification and domain adaptation in a single proceeding network, but PixelDA separates the proceeding of classification from the proceeding of domain adaptation. Odena et al. [60] proposed Auxiliary Classifier GAN (AC-GAN) with class label conditioning to learn a suitable adaptation. AC-GAN introduces a metric function of spatial resolution for image classification. Sankaranarayanan et al. [49] improved AC-GAN [60] where there are two parallel streams in training phase: classification stream, and adversarial stream which is an AC-GAN framework in the adversarial branch. Antoniou et al. [63] proposed a model based on CGANs, Data Augmentation Generative Adversarial Network (DAGAN), which augments standard vanilla classifiers and improves a kind of few-shot learning approaches. Yin et al. [61] introduced Face Frontalization GAN, FF-GAN, for generating 3D shapes face images. FF-GAN framework differs from CGAN based on modeling. Incorporating FF-GAN into the GAN structure gains appearance and shape priors for less training data and fast convergence. The FF-GAN utilizes not only the generator and discriminator loss similar GANs but also applies a masked symmetry loss for retaining the visual quality and an identity loss for recovering high-frequency information. Volpi et al. [64] utilized domain-invariance in a single feature extractor which is trained by GANs, and augmented the features by preparing a feature generator trained with a CGAN. Hu et al. [65] presented a Couple Agent Pose Guided Generative Adversarial Network (CAPG-GAN), which utilizes an identity preserving loss to keep identity knowledge and also applies a total variation regularization for refining local textures, besides of generator and conditional adversarial loss. SketchGAN [66] is another model based on CGANs with a cascade encode-decoder architecture. The SketchGAN has an incomplete sketch as input and a completed sketch with its classification label as output.

Coupled Generative Adversarial Network (CoGAN): Coupled Generative Adversarial Networks (CoGAN) [67] has two GANs. In CoGAN architecture, each source and target have a unique generative adversarial objective, and also source and target representations are jointly learned via weight sharing in specific layers. The main advantage of CoGAN is the effective learning of joint distribution from the two domains by separately drawing the samples from the marginal distributions [93]. Liu et al. [68] introduced a model based on CoGANs, UNsupervised Image-to-image Translation (UNIT), which combines variational autoencoders with CoGAN. Lv et al. [69] proposed an unsupervised domain adaptation model, TarGAN, which generates target data along with class labels are obtained by GANs, to improve the classification accuracy. The TarGAN baseline is CoGAN [67] and DDC [7].

4.1.2 *Discriminative Adversarial Network:* This section implies to some methods which usually utilize an adversarial manner for obtaining domain confusion concerning a domain discriminator. GANs are not very optimal on discriminative purposes and maybe tangled in smaller domain shifts while discriminative methods with applying the shared weights can handle the larger ones. Ganin et al. [70] proposed Domain Adversarial NN (DANN), which shares weights across the source and target domains and transfers both domains data to a common feature space. DANN by adding a new layer, gradient reversal, to the standard feed-forward NN with a backpropagation training manner solves the domain adaptation problems. Adversarial Discriminative Domain Adaptation (ADDA) [71] combines discriminative modeling with a GAN loss for domain adaptation purposes. First, ADDA utilizes the source domain labels to learn a discriminative representation and then learns a discriminative mapping of encoded target domain images to the source domain feature space by a domain adversarial loss. ADDA has the following unconstrained optimization (8):



(a) Architecture of generative adverserial network (GAN)

(b) Architecture of conditional GAN (CGAN)

(c) Architecture of cycle GAN

Fig. 3: (a) Architecture of GAN [43], (b) Architecture of CGAN [57] and (c) Architecture of CycleGAN [53].

$$
\begin{aligned}
\min_{M_s,C} L_{cls}(X_s,Y_s) &= -E_{(x_s,y_s)\sim(X_s,Y_s)} \sum_{k=1}^{K} 1_{[k=y_s]} \\
&\log C(M_s(x_s)) \\
\min_{D} L_{advD}(X_s,X_s,M_s,M_t) &= -E_{(x_s)\sim(X_s)}[\log D(M_s(x_s))] \\
&-E_{(x_t)\sim(X_t)}[\log(1-D(M_t(x_t)))] \\
\min_{M_t} L_{advM}(X_s,X_t,D) &= -E_{(x_t)\sim(X_t)}[\log D(M_t(x_t))]
\end{aligned} \tag{8}
$$

where M_s and M_t are source mapping and target mapping respectively. L_{cls} is classification loss function for source data, L_{advD} is adversarial loss function for domain discriminator, and L_{advM} is adversarial loss function for source and target mappings.

Pei et al. [72] presented a Multi Adversarial Domain Adaptation (MADA) approach, which addresses domain adaptation problems by applying multiple domain discriminators. A key advance about previous approaches (i.e. DAAN and ADDA) is the ability to simultaneously enhance positive transfer on relevant data and reduce negative transfer on irrelevant data. Another work was introduced by Qi et al. [73], which considers a problem of Multi-modal Domain Adaptation NN (MDANN) for domain adaptation by attending and fusing in an adversarial manner. Hybrid domain constraints are presented to jointly learn discriminative and domain-adaptive multi-modal features. Zhang et. al. [94] presented Domain Symmetric Networks (SymNets), which is based on a symmetric plan of target and source task classifiers. To train the SymNet, a new adversarial learning objective is introduced that the key plan is based on a two-level domain confusion manner, where the class-level confusion loss boosts over the domain-level via driving the learning of intermediate network features to be invariant at the corresponding classes of the two domains.

4.1.3 Feature Matching Adversarial Network: In this section, some adversarial methods are reviewed where the source classifier is trained with the labeled data from source domain, and the target classifier is regularized by minimizing a distance metric between the source classifier and the target one by using all the data.

RevGrad [74] proposed a shared encoder and two discriminator streams for predicting domain and class where a classification loss is used for class prediction while multiplying the gradient by a negative value during the backpropagation training is used for domain prediction. Bousmalis et al. [75] introduced Domain Separation Networks (DSN), which encodes representations by using domain adversarial similarity loss and maximizes confusion by utilizing Gradient Reversal Layer. Shen et al. [76] used Wasserstein distance as a discrepancy loss in adversarial network. Shen et al. [30] introduced WDGRL, which applied an adversarial scheme to obtain domain-invariant representations by iteratively learning features with lower Wasserstein distance. [78, 79], use the same idea as RevGrad [74], but the optimization approach for discriminator and generator losses is different from multiplying the gradient with a negative constant as in RevGrad. Conditional domain adversarial networks (CDANs) [77] was introduced to consider multi-linear condition and entropy condition framework, where the first condition enhances the discriminative ability by obtaining the covariance between representations and classifier predictions, and the second condition engages the transferring ability by controlling the uncertainty of classifier predictions. Zhang et al. [80] introduced Collaborative and Adversarial Network (CAN), which learns discriminative representations, unlike GANs don't generate new images, but similar to DANN learns the network by back-propagating the gradients from domain loss. Xu et al. [81] proposed a deep cocktail network (DCTN), which utilizes multi-source multi-process adversarial learning to reduce the difference between the target domain and each of the source domains. DCTN relaxes the assumption of having the shared class among any source domains. Saito et al. [82] proposed an adversarial learning method for domain adaptation by applying the decision boundaries that are specific for each task to increase the distance between the classifiers. In other words, this method utilizes the task-specific classifiers as a discriminator for the relationship between boundaries of classes and samples of target domain. Chen et al. [83] introduced the Re-weighted Adversarial Adaptation Network (RAAN), which applies EM distance to matching the feature distributions in an adversarial scheme for domain adaptation. RAAN is the first approach to learn domain invariant representations in UDA which utilizes the optimal transport [95] based on EM distance. Pinheiro [84] presented similarity adversarial learning method base on DANN [70] with a different approach to do classification where each image from source or target domains is evaluated by some of prototypes with a similarity-based classifier. Cicek and Soatto [85] proposed a conditional domain adaptation method where a cross-entropy loss is utilized for training a class predictor with the labeled source samples, and an adversarial regularization is applied to progress the performance of the classifier on the target domain. The work of [77] is similar to [85], which conditions the domain alignment loss to labels. Wen et al. [86] utilized an additional conditional domain adversarial loss to learn domain-invariant local feature, and to jointly align global and local feature statistics. Cai et al. [87] proposed Adversarial Residual Transform Network (ARTN), which is a feature-shared model directly transforming the source features into the target feature space. ARTN utilizes residual connections between the feature extractor and transform network to relax the learning of distribution mapping by sharing features. Laradji et.al. [38] introduced Metric-based Adversarial Domain Adaptation (M-ADDA), which utilizes similar architecture with ADDA [71], but M-ADDA applies a metric learning method to train the source classifier on the source data via optimizing a triplet loss function. Then, it applies the adversarial scheme to extract featured from both source and target data indistinguishable. They optimized a new loss function, which encourages the target data's embeddings to form clusters with large margins between them. Zhang et. al. [88] proposed Hybrid Adversarial Network (HAN). HAN minimizes the source data classifier loss, conditional adversarial loss similar to [77], and the correlation alignment loss. They introduced a new adaptation layer for further promoting the performance in the HAN model.

4.2 Partial Adversarial Based Methods

Partial adversarial based is a new scheme, which relaxes the same label space assumption between source and target spaces. So, the target labels are supposed to be subset of the source labels. In other words, the number of class labels in the target domain is less than the number of class labels in the source domain. Fig. 4 shows this concept. As mentioned before in subsection 4.1, non-partial approaches typically align all of the source space with the target space, which might result in the negative transfer for domain adaptation problems.

Cao et al. [89] proposed Selective Adversarial Network (SAN), which improves positive transfer by considering a weighting mechanism via multiple adversarial networks, and tries to prevent negative transfer by ignoring the outlier source classes. A key progress over related models is the ability to simultaneously boost positive transfer on relevant data and reduce negative transfer on irrelevant data. Cao et al. [90] presented Partial Adversarial Domain Adaptation (PADA), which improves SAN [89] by utilizing only one adversarial network and giving more importance to class-level in source classifier. Zhang et al. [91] introduced an unsupervised partial domain adaptation model, Importance Weighted partial Adversarial Network (IWAN), with two domain classifiers where the first classifier obtains the source instance importance weights, and the second classifier by utilizing the weighted source instances and the target instances executes the minimax game. They showed that the minimax method between the second domain classifier and the feature extractor is equivalent to minimizing the Jensen-Shannon distance between the weighted source density and the target density. Cao et al. [92] proposed another approach to partial domain adaptation, Example Transfer Network (ETN), which computes the transferring power of source examples by combining the discriminative information and down-weighting the outlier source examples. The key contribution of ETN is that the weights are put in the source classifier loss, which significantly enhances the ability to decrease the irrelevant source examples that damage the final model. Based on the evaluation, the ETN model is strongly performed for partial domain adaptation in comparison with previous models.

5 Reconstruction Based Methods

Reconstruction based methods reconstruct all domains samples to make the same representation of the domains along with preserving the special properties of each domain. Our taxonomy of reconstruction based methods is summarized in Table 3, and the recent works on this topic are reviewed in the following subsections.

Table 3 Our Taxonomy of RECONSTRUCTION BASED METHODS

Encoder-Decoder Models	[96], mSDA[97], MTAE[98], [99], DRCN[100], DTL[101], [102], [103]
Dictionary and Sparse Coding Models	[104], [105], [106], CDMDA[107]
Graph-Based Models	PUnDA[108], OBTL[109], ABLR[110], GM+PL[111], GAKT[112], GCAN [113], AdaGraph[114], GCAN[113]

5.1 Encoder-Decoder Models

Encoder-decoder models are the approaches that first encode an input to some hidden representation by the encoder, then decode this representation back for a reconstructed version by the decoder. The domain-invariant features are learned by a common encoder while domain-special features are preserved by reconstruction loss[115].

Glorot et al. [96] extracted a high-level feature space by using Stacked Denoising Autoencoders (SDA) [116]. SDAs have high computational costs along with a lack of scalability in high-dimensional feature spaces. So, the marginalized SDA (mSDA) method [97] was introduced to solve these problems, which utilizes linear denoisers to marginalize noise where parameters are computed in a closed-form solution without using a stochastic gradient descent approach. Ghifary et al. [98] introduced an autoencoder model, Multi-Task Autoencoder (MTAE), which learns self-domain and between-domain reconstruction. The most important contribution of MTAE is the training approach. It constructs the generalized denoising autoencoder, which learns invariances for naturally happening

IET Research Journals, pp. 1–18

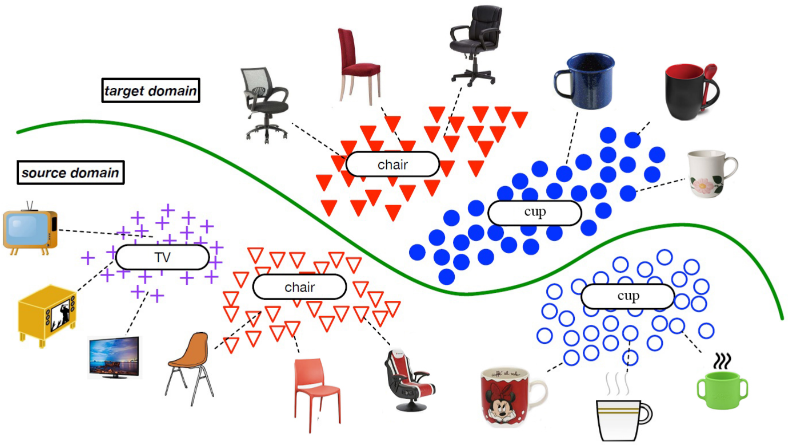


Fig. 4: The partial scenario. The target label space ('chair', 'cup') is contained subsetting of the source label space ('TV', 'chair', 'cup') [90].

transformations. Mittal et al. [99] utilized the deep belief network with stacked autoencoder, which jointly learns the representation for matching composite sketches. Ghifary et al. [100] proposed Deep Reconstruction Classification Network (DRCN), which jointly learns a common encoder for encoding source features in visual classification and a decoder for reconstructing the unlabeled target data. Wen et al. [101] proposed a deep transfer learning (DTL) method, which applies a sparse autoencoder with three layers for feature extraction, and utilizes the MMD metric to minimize the feature discrepancy. Murez et al. [102] proposed a UDA method, which utilizes an encoder network to apply constraint in the features extraction scheme. Then, decoding the features back to the source and target domains is done similar to [100]. Finally, the cycle consistency is used for both domains to certify that the mappings are learned correctly. Jiang et al. [103] presented a Cross Domain Minimization with Deep autoencoder (CDMDA) for unsupervised domain adaptation, which simultaneously learns the classifier by predicting of labels in the source domain and input reconstruction in the target domain using shared features aligned with CORAL as a regularizer in a unified scheme. The process of encoding and decoding of the basic autoencoder can be summarized as Equation (9):

$$\begin{aligned} z &= g_{enc}(x;\theta) \\ \hat{x} &= g_{dec}(z;\theta') \end{aligned} \tag{9}$$

where encoder g_{enc} and a decoder g_{dec} are both multilayer neural networks. The encoder first maps input x to latent feature codings z, then the decoder reconstructs the input from z. Corresponding parameters θ and θ' can be optimized by minimizing mean square reconstruction error Equation (10):

$$\min_{\theta,\theta'} ||x - \hat{x}||^2 = \min_{\theta,\theta'} ||x - g_{dec}(g_{enc}(x))||^2 \tag{10}$$

Bousmalis et al. [117] presented an encoder-decoder model with a common encoder for shared representations and a private encoder for domain-specific representation in each domain. Peng et al. [118] extended previous works based on SDA [96, 97, 101, 116], Stacked Local Constraint Auto-Encoder (SLC-AE), by proposing a variant of SDA for domain adaptation, which learns domain-invariant features by utilizing SDA and the low-dimensional manifold.

5.2 Dictionary and Sparse Coding Models

The idea of learning a dictionary was first introduced by Olshausen and Field [119]. These methods perform the process of updating or adapting the dictionary to better match the data between different domains. Dictionary-based models can make robust discriminant representations by adapting to particular data samples.

DLRC [104] is a dictionary-based CNN in which convolutional layers are general and shared by two domains while fully connected layers are task-specific and adapted with multilayer low-rank coding. DLRC applies the transformed source domain as a dictionary and uses it to reconstruct the transformed all data from the both domains. Ding et al. [107] enhanced the feature adaptation performance of DLRC [104] over general deep representations and investigated multilayer low-rank coding at the top task-specific layers. They extended the previous low-rank coding with one shared dictionary to the multilayer dictionaries. Yang et al. [105, 106] introduced a dictionary learning model with shared domains and sparse groups, DsGsDL. They introduced domain-shared group-sparsity criteria which is an equivalent condition on conditional distribution matching. A domain-shared group-sparse dictionary learning approach is developed through the joint alignment on marginal and conditional distributions.

5.3 Graph Based Models

In this section, we review the topic of UDA on graphs. These methods try to build the connected graph from the training samples for label deployment in graph modeling. It is usually considered a source graph and a target graph with samples drawn from data manifolds. These models solve the problem of estimating the unknown class labels of the target graph utilizing the label information of the source graph and the similarity among the two graphs though the weights of graph edges.

Gholami et al. [108] proposed graphical Probabilistic Unsupervised Domain Adaptation (PUnDA) model, which learns the classifier in a common space by using the MMD metric with utilizing a graphical framework. Karbalayghareh et al. [109] presented Optimal Bayesian Transfer Learning (OBTL) model, which combines graph model concept with a Bayesian method for domain adaptation. Perrone et al. [110] introduced Adaptive Bayesian Linear Regression (ABLR) model for multi-task applications, which is a graph based method for Bayesian optimization. In the ABLR, each task is modeled by a Bayesian linear regression layer on top of common feature space. Mancini et al. [114] introduced a deep architecture, AdaGraph, to get information from the auxiliary domains by utilizing a graph. AdaGraph is similar to graph based domain adaptation methods GM-PL[111] and GAKT[112]. Different from these works, in the AdaGraph model, a node represents a whole domain not a sample, and edges link domains with related metadata, while in GM-PL and GAKT, links are drawn between related samples. GCAN [113] utilizes a dense connected graph to solve unsupervised domain adaptation problem where the data structure is jointly integrated with domain and class labels in a deep NN. GCAN is designed via three alignment schemes: structure-aware alignment, domain alignment, and class centroid alignment.

6 Representation Based Methods

Representation based methods apply the trained network to use intermediate representation as input to the new NN. Our taxonomy of representation based methods is summarized in Table 4 and the recent works on this topic are reviewed in the following subsections.

Table 4 Our Taxonomy of REPRESENTATION BASED METHODS

Domain Confusion Representation	[120], BSW [121], [122]
Domain Invariant Representation	[123], [124], [125]
Representation Disentangling	[70], [75], [62], DR-GAN[59], CDRD[126], [127], UFDN[128], MTDA-ITA[129]

6.1 Domain Confusion Representation

This section introduces methods that use the confusion loss to learn the features that provide domain invariance.

Tzeng et al. [120] presented a CNN architecture with an additional fc layer for aligning domains via domain confusion loss, and transfers classification information between domains via a cross-entropy soft label loss function. Rozantsev et al. [121] introduced a two-stream adaptation model where in comparison with other methods, the corresponding layers weights are related but they are not shared. So the loss functions are presented to learn corresponding weights. This method learns source and target streams parameters through the overall loss functions, and maximizes the confusion between the domains by using exactly the same classification layer for both source and target data. Rozantsev et al. [122] improved their previous model [121] through the residual transfer network.

6.2 Domain Invariant Representation

This section contains methods that use the source and the target data to learn a common representation, which can enhance the domain adaptation.

Chen et al. [123] proposed Transfer Neural Trees (TNT) on heterogeneous domain adaptation. TNT contains two-stream networks for learning invariant features where the TNT prediction layer utilizes Transfer-NDF [58] for adapting the neurons in TNT by stochastic pruning. TNT utilizes the embedding loss for performing prediction and preserving structural consistency among the target-domain data. In [124, 125], pseudo-labels are utilized for invariant representation that are proven effective for improving domain adaptation models. While in [124], first, some of the labeled source and unlabelled target samples are taken as input, and these input samples are mapped into a deep representation. Then, the two-stage optimization on the input loss is computed by the model. This method learns a metric to minimize the loss function where the triplet loss is defined among the labeled source data with their nearest positive/negative neighbors between the unsupervised target data. The advantage of this method is that it can jointly learn the optimal feature representation and the optimal cross-domain transformation parameter, and target label inference for UDA. Saito et al. [125] applied four networks in their domain adaptation model where one network works as a shared feature extractor, and two networks learn from labeled source instances and pseudo-labeled target instances, and one of the other networks is trained by the pseudo-labeled target instances to attain target discriminative features.

6.3 Representation Disentangling

Representation disentangling is a learning method for concluding a hidden feature space that decomposes the derived representation so that the visual attributes can be recognized and described [130, 131].

Tran et al. [59] proposed Disentangled Representation learning-GAN (DR-GAN), which utilizes an encoder-decoder to learn the discriminative and generative representation. This representation relies on the learned coefficients, which is disentangled from face variations and face estimations. Liu et al. [126] proposed Cross Domain Representation Disentangler (CDRD), which aligns labeled source data with unlabeled target data where representation disentanglement and adaptation are jointly performed for visual domain adaptation. In [62, 70, 75, 126], are proposed disentanglement representation methods based on class labels to gain invariant feature representation. Gonzalez et al. [127] proposed an image-to-image translation for representation disentangling based on GANs and autoencoders. In this model, the internal representation has three parts where a shared part has information for different domains, and two exclusive parts have only factors of particular domain variations. Liu et al. [128] presented a Unified Feature Disentanglement Network (UFDN), which learns deep disentangled features for image translation and manipulates image outputs in the multi-domain scheme. Gholami et al. [129] proposed a Multi Target Domain Adaptation Information Theoretic Approach (MTDA-ITA), which makes a solid relationship between the hidden feature spaces and the source data, they utilized a unified approach for disentangling the shared and private knowledge.

7 Attention Based Methods

Attention based methods focus more on some transferable attention regions or images of source data and connect them to the target data. In other words, it is introduced as the mechanism of guiding the network to focus on the spatial parts of target images that contain related information to the source images.

Our taxonomy of attention based methods is summarized in Table 5 and the recent works on this topic are reviewed in the following subsections.

Table 5 Our Taxonomy of ATTENTION BASED METHODS

Adversarial Attention Alignment	TADA[132], DAAA [133]
Transferable Local Attention	DUCDA[134], [135], [132]
Transferable Global Attention	CHTL[136]

7.1 Adversarial Attention Alignment

This section reviews transferable attention methods which uses adversarially learning.

Wang et al. [132] proposed a domain adaptation model for transferable regions, Transferable Attention for Domain Adaptation (TADA) where there are two types of transferable adversarial attention network: local attention network obtains the features of the higher transferability regions generated by discriminators in multi-region level, and global attention network obtains more similar images by discriminator in the single-image level. The TADA architecture is shown in Fig. 5.

Kang et al. [133] proposed an attention alignment model based on CycleGAN, which transfers information in all convolutional layers by attention alignment. Also, they maximized the likelihood of target data, which enables the target network to apply more training data for better domain adaptation.

7.2 Transferable Local Attention

This section reviews transferable local attention methods that specify which regions of images are better to transfer from the source

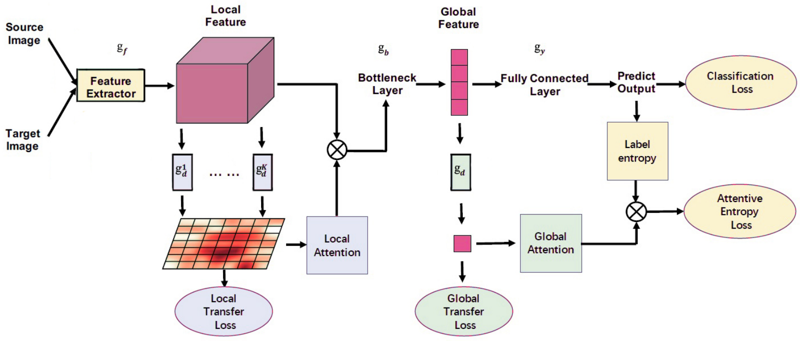


Fig. 5: Architecture of TADA [132].

domain to the target domain that leads to better performance in domain adaptation.

Zhuo et al. [134] presented Deep Unsupervised Convolutional Domain Adaptation (DUCDA) model with two loss functions consisting of source classification loss and correlation alignment loss. In this model, correlation alignment loss is utilized on both convolutional layers and fully connected layers. Wang et al. [135] introduced a residual approach for local attention, which is robust to negative local attention. Negative local attention is disadvantageous in the domain adaptation task. Wang et al. [132] added a residual connection in extending the previous idea to reduce the negative effects [135].

7.3 Transferable Global Attention

This section reviews transferable global attention methods that specify which images are better to transfer from the source domain to the target domain that leads to better performance in domain adaptation.

Moon et al. [136] proposed Completely Heterogeneous Transfer Learning (CHTL), where a transferred loss determines the source related images for transferring to the target domain. In the introduced model's Wang et al. [132], in addition of local attention, transferable global attention is implied, which utilizes a single image-level domain discriminator to specify the best related images to transfer.

8 Benchmark Datasets for Deep Visual Domain Adaptation

In this section, some important benchmark datasets for deep visual domain adaptation are reviewed. The most common ones are summarized in Table 6.

1. Office-31 Dataset :
(https://people.eecs.berkeley.edu/ jhoffman/domainadapt)
This dataset includes 3 domains, Amazon, Webcam, and DSLR. Each domain contains images from amazon.com or office facility images, which consists of different lighting and poses. This dataset has 2817, 498, and 795 images in each domain, respectively, and 31 categories. For deep learning applications, the dataset has been extracted from pre-trained AlexNet with 4096 dimensional feature vectors.

2. Office+Caltech Dataset :
(https://people.eecs.berkeley.edu/ jhoffman/domainadapt)
This dataset contains 10 categories for every 4 domains, which are Amazon, Caltech, DSLR, and Webcam. There are 958, 1123, 157, and 295 images in each domain, respectively. Amazon domain has SURF features with vector quantized to 800 dimensions, Caltech domain has DeCAF features with 4096 dimensions, and 4096 dimensional feature vectors are extracted from pre-trained VGG-Net for DSLR and Webcam domains.

3. Office+Home Dataset :
(http://hemanthdv.org/OfficeHome-Dataset)
This dataset consists of 4 different domains, Art, Clipart, Product, and Real-World, where each domain contains 65 categories with Office and Home settings images. For deep learning applications, the dataset has been extracted from pre-trained ResNet50 with 2048 dimensional feature vectors.

4. PIE Dataset :
(https://www.ri.cmu.edu/project/pie-database)
This database contains 41368 images of 68 people, where each person is represented under 13, 43, and 4, different poses, illuminations, and expressions, respectively. It has 5 subsets containing left pose, up pose, down pose, front pose, right pose.

5. MNIST+USPS Dataset :
(http://yann.lecun.com/exdb/mnist)
The MNIST dataset and the US Postal (USPS) dataset are two famous handwritten digit datasets. Each dataset has 10 categories. The MNIST dataset is derived from the NIST dataset. The MNIST dataset has 60000 training and 10000 test samples. The USPS dataset obtains recognizing handwritten digits, too. The training set and the test set have 7291 and 2007 samples, respectively.

6. COIL20 Dataset :
(http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php)
Columbia Object Image Library (COIL20) is a dataset of 1440 normalized images with 20 object categories. The images are at pose intervals of 5 degrees. Also, There is COIL100 database with 100 object categories,
(http://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php).

7. CIFAR10 Dataset :
(http://www.cs.utoronto.ca/ kriz/cifar.html)
The CIFAR10 dataset includes 60000 color images with 10 different classes. There are 6000 images per each class, which are 50000 images for training and 10000 images for testing. CIFAR100 is another dataset, which is similar to the CIFAR10, but it has 100 classes. These classes are categorized into 20 superclasses, containing 600 images in each of them. There are 500 and 100 training and testing images in each class, respectively.

8. SVHN Dataset :
(http://ufldl.stanford.edu/housenumbers)
The Street View House Numbers (SVHN) Dataset is a dataset with real-world images from Google Street View images, which have been obtained of house numbers with a minimum need on data preprocessing. It has over 600000 digit images with 10 classes, where digit '0' indicates label 10, '1' indicates label 1, '9' indicates label 9,

Table 6 Benchmark Datasets for Deep Visual Domain Adaptation

Dataset	Subsets	Abbreviation	#Images	Feature(size)	#Classes
Office-31	Amazon	A	2,817	$AlexNet - FC_7$(4,096)	31
	DSLR	D	498		
	Webcam	W	795		
Office+Caltech	Amazon	A	958	$SURF$(800)	10
	Caltech	C	1,123	$DeCAF_6$(4,096)	
	DSLR	D	157	$VGG - FC_6$(4,096)	
	Webcam	W	295	$VGG - FC_7$(4,096)	
Office+Home	Art	Ar	2,421	$ResNet50 - P_5$(2,048)	65
	Clipart	Cl	4,379	$ResNet50 - P_5$(2,048)	
	Product	Pr	4,428		
	Real-World	Rw	4,357		
PIE	C05 (left pose)	P1	3,332	Pixel(1,024)	68
	C07 (up pose)	P2	1,629		
	C09 (down pose)	P3	1,632		
	C27 (front pose)	P4	3,329		
	C29 (right pose)	P5	1,632		

and so on. It has 73257, 26032, 531131, digits for training, testing, and additional (somewhat less difficult samples), respectively.

9. NORB Dataset :
(https://cs.nyu.edu/ ylclab/data/norb-v1.0)
This database is suitable for 3D object recognition. It consists of 50 different toys images from 5 general groups: four-legged animals, human figures, airplanes, trucks, and cars. The images were taken by two cameras under 6 lighting conditions, 9 heights from 30 to 70 degrees for every 5 degrees, and 18 directions from 0 to 340 for every 20 degrees.

10. ImageCLEF Dataset :
(https://www.imageclef.org/2014/adaptation)
It consists of three domains which are Caltech-256, ImageNet ILSVRC2012, and PASCAL VOC2012 with a total of 600 images for each domain and 50 images for each class. This dataset contains 12 classes which are common to all domains: aereoplane, bike, bird, boat, bottle, bus, car, dog, horse, monitor, motorbike, and people.

11. ImageNet Dataset :
(http://www.image-net.org)
This huge image database is created to the WordNet hierarchy where the whole number of non-empty subcategories is 21841, the whole number of images is 14197122, the number of bounding box annotations images is 1034908, the number of subcategories with SIFT features is 1000, and the number of SIFT features images is 1.2 million.

9 Unsupervised Domain Adaptation for Other Applications

This survey paper focuses on the domain adaptation methods used in image classification tasks, but domain adaptation techniques are applied to many other applications. These are semantic segmentation, object detection, text recognition, person re-identification, which are reviewed briefly in this section.

Semantic segmentation applications recognize the relation between each image pixel and a suitable class label. Zhao et al. [137] proposed the semantic segmentation algorithm under classification and regression methods for domain adaptation, whereas, Tsai et al. [138] learned discriminative feature representations under space clustering. In [139–141], domain adaption for semantic segmentation are structured by learning the autoencoder. In [142, 143], domain adaptation frameworks are introduced that solve semantic segmentation problems by re-creating pseudo labels in the target domain and re-training the network with these labels. The learning semantic representation method which was proposed by Huang et al. [144] aligns the distributions of intermediate layers activations, whereas, Xie et al.'s method [145] aligns labeled source centroid and pseudo-labeled target centroid. In [146–150], GAN based methods are applied for semantic segmentation problems. Hong et al. [151] proposed a modular model consisting of two modules: a perception and a control policy, where semantic image segmentation is utilized for relating these modules. In [152, 153], adversarial domain adaptation methods are utilized for semantic segmentation. Vu et al. [154] addressed this problem with entropy loss for pixel-wise predictions. Chen et al. [155] proposed a cross-domain semantic segmentation model by utilizing auxiliary geometric information, whereas, Mousavian et al. [156] learned their model by using semantic texture and capturing spatial layout.

Object detection is another application in computer vision tasks to detect instances of objects in the target domain with the source certain class. In [157, 158], proposed domain adaptation models solve object detection problems by utilizing autoencoder. Abdullah-Jamal et al. [159] proposed a domain adaptation method, which can be used in the supervised and unsupervised scheme for face detection. Chen et al. [160] built their approach relied on Faster R-CNN model [161], which designs two domain adaptation modules based on H-divergence theory with the adversarial learning approach. Hsu et al. [162] to solve easier adaptation tasks utilized an intermediate domain for object detection problems. Yu et al. [163] trained a CNN by using the refined pseudo labels and a weighted loss function. Zhu et al. [164] proposed a domain adaption model for object detection to answer, "where to look" and "how to align". They answered them with mining the discriminative regions.

Chen et al. [165] proposed a Multinomial Adversarial Network (MAN) to address the text recognition problem by using adversarial approach. In [166, 167], the encoder-decoder models are introduced for text recognition problem. Zhan et al. [168] presented Geometry-Aware Domain Adaptation Network (GA-DAN), which models the shift between domains in both geometry and appearance spaces, and converts images with different characteristics across domains.

Li et al. [169] introduced Adaptation and Re-identification NN (ARN), which utilizes domain-invariant feature representations for person re-identification (Re-ID). In [170, 171], domain adaptation methods are proposed that apply GAN based scheme for Re-ID. Bak et al. [172] presented a domain adaptation model, which performs fine-tuning in an unsupervised way with using synthetic data. One of the challenging problems in person re-identification models is model generalization. Deng et al. [173] presented a "learning via translation" framework to translate the source labeled images to the target domain for addressing this problem. Wang et al. [174] proposed Transferable Joint Attribute-Identity Deep Learning (TJ-AIDL), which simultaneously learns feature representations with semantic and identity determinative for Re-ID problems. Li et al. [175] presented an unsupervised Re-ID deep learning model, which discovers and exploits the information from person tracklet data, which is generated from an end-to-end model.

Table 7 Comparison of Various Methods in Deep Visual UDA for Classification Tasks on the Office31 Dataset (Classification Accuracy %). The ResNet50 is as Base Network.

Category	Model	A → W	A → D	W → A	W → D	D → A	D → W	Avg.
Discrepancy-based	DDC [7]	75.6	76.5	61.5	98.2	62.2	96.0	78.3
	DAN [8]	80.5	78.6	62.8	99.6	63.6	97.1	80.4
	RTN [9]	84.5	77.5	64.8	99.4	66.2	96.8	81.6
	JAN [10]	85.4	84.7	70.0	99.8	68.6	97.4	84.4
	SCA [16]	93.6	89.5	72.4	100	72.6	98.0	87.6
	CAN [15]	94.5	95.0	77.0	99.8	78.0	99.1	90.6
Adversarial based	RevGrad [74]	82.0	79.7	67.4	99.1	68.2	96.9	82.2
	DANN [70]	82.0	79.7	67.4	99.1	68.2	96.9	82.2
	ADDA [71]	86.2	77.8	68.9	98.4	69.5	96.2	82.9
	MADA [72]	90.0	87.8	66.4	99.6	70.3	97.4	85.2
	SimNet [84]	88.6	85.3	71.8	99.7	73.4	98.2	86.2
	iCAN [80]	92.5	90.1	69.9	100	72.1	98.8	87.2
	CDAN [77]	93.1	93.4	70.3	100	71.0	98.6	87.7
	SCA-Rev [16]	93.6	89.5	72.7	100	72.5	98.5	87.8
	SymNets [94]	90.8	93.9	72.5	100	74.6	98.8	88.4
	HAN [88]	95.3	94.4	71.7	100	72.1	98.8	88.7
Representation-based	TCA [176]	72.7	74.1	60.9	99.6	61.7	96.7	77.6
	GFK [177]	72.8	74.5	61.0	98.2	63.4	95.0	77.5
Reconstruction-based	DLRC[104]	61.3	60.3	48.8	94.9	52.9	93.7	68.7
	DTLC[107]	70.4	68.2	53.9	99.3	54.9	96.9	73.9
Attention-based	DAAA [133]	86.8	88.8	73.9	100	74.3	99.3	87.2
	TADA [132]	94.3	91.6	73.0	99.8	72.9	98.7	88.4

Table 8 Comparison of Various Methods in Deep Visual UDA for Classification Tasks on the ImageCLEF Datasets (Classification Accuracy %). The ResNet50 is as Base Network.

Category	Model	I → P	P → I	I → C	C → I	C → P	P → C	Avg.
Discrepancy-based	DAN [8]	74.5	82.2	92.8	86.3	69.2	89.8	82.5
	RTN [9]	75.6	86.8	95.3	86.9	72.7	92.2	84.9
	JAN [10]	76.8	88.0	94.7	89.5	74.2	91.7	85.8
	SCA [16]	78.1	89.2	96.8	91.3	78.2	94.0	87.9
Adversarial based	DANN [70]	75.6	84.0	93.0	86.0	71.7	87.5	83.0
	RevGrad [74]	75.0	86.0	96.2	87.0	74.3	91.5	85.0
	MADA [72]	75.0	87.9	96.0	88.8	75.2	92.2	85.8
	iCAN [80]	79.5	89.7	94.7	89.9	78.5	92.0	87.4
	CDAN [77]	77.2	88.3	98.3	90.7	76.7	94.0	87.5
	HAN [88]	77.9	91.7	97.0	91.9	76.7	95.3	88.4
	SymNets [94]	80.2	93.6	97.0	93.4	78.7	96.4	89.9

10 Discussion

In this section, we investigate, analyze, and discuss a list of different experiments reported in reviewed deep visual unsupervised domain adaptation papers that are included in our taxonomy.

In Table 7 we compare various state-of-the-art methods in deep visual UDA for classification tasks on the Office31 dataset. As shown in Table 7, CAN [15] achieves 90.6% in average classification accuracy, which is the best performance between discrepancy-based methods, including DDC [7], DAN [8], RTN [9], JAN [10], and SCA [16] on the Office-31 dataset and gets +12.3, +10.2, +9, +6.2, and +3 higher classification accuracy on average, respectively. Also, the CAN accuracy is higher than some adversarial-based methods like RevGrad [74], DANN [70], ADDA [71], MADA [72], SimNet [84], iCAN [80], CDAN [77], SCA-rev [16], SymNets [94], and HAN [88] (90.6% v.s. 82.2%, 82.2%, 82.9%, 85.2%, 86.2%, 87.2%, 87.7%, 87.8%, 88.4% and 88.7, respectively). As we see in Table 7, CAN achieves the best result between listed discrepancy-based methods and obtains very competitive results compare to different methods in other groups. The reasons can be as follows: first, CAN applies two kinds of domain discrepancy metrics, i.e. the inter-class and the intra-class, while many of existing methods only consider intra-class domain difference in their models which causes different classes data may be matched incorrectly, i.e. some methods based on MMD or JMMD can be minimized even when the target-domain data are mismatch with the source-domain data of a different class, while CAN proposes an effective discrepancy metric, Contrastive Domain Discrepancy (CDD), to perform class-level alignment for UDA. Second, CAN performs training with both

Table 9 Comparison of Various Methods in Deep Visual UDA for Classification Tasks on the Digit Datasets (Classification Accuracy %). The ResNet50 is as Base Network.

Category	Model	SVHN → MNIST	MNIST → USPS	USPS → MNIST
Discrepancy-based	DDC [7]	68.1	-	66.5
	DAN [8]	71.1	-	-
	SCA [16]	92.0	96.1	95.5
	SWD [33]	98.9	98.1	97.1
Adversarial-based	DANN [70]	73.9	85.1	73.0
	RevGrad [74]	73.9	77.1	73.0
	ADDA [71]	76.0	89.4	90.1
	DSN [75]	82.7	-	-
	ADGAN [49]	-	92.50	90.80
	M-ADDA [38]	-	95.2	94.0
	UNIT [68]	90.5	96.0	93.6
	Cycada [56]	90.4	95.6	96.5
	PixelDA [62]	-	95.9	-
	CoGAN [67]	-	91.2	89.1
	GAGL [50]	96.7	74.6	-
	GADM [52]	78.0	93.8	95.1
	RAAN [83]	89.2	89.0	92.1
	TarGAN [69]	98.1	93.8	94.1
Representation-based	BSW [121]	82.8	60.7	67.3
	CDRD [126]	-	95.05	94.35
Reconstruction-based	DRCN [100]	-	91.80	73.67

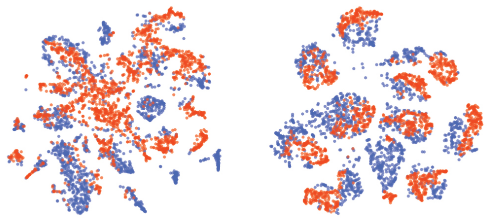
source labels and pseudo target labels which are achieved by clustering, So the learned decision boundary can generalize more strongly on the target domain.

In Table 7, among adversarial-based methods (i.e. RevGrad [74], DANN [70], ADDA [71], MADA [72], SimNet [84], iCAN [80], CDAN [77], SCA-Rev [16], SymNets [94], HAN [88]), HAN achieves higher accuracy result. HAN is jointly adopted by both correlation alignment and conditional adversarial learning. HAN considers both the class-level distribution matching and the correlation between domains. HAN incorporates a classification loss for learning a good classifier. A domain adversarial network is utilized to invariant feature representations of learning to domain differences, and a correlation alignment is utilized to reduce the discrepancy in the correlation between domains. Additionally, an adaptation layer is introduced to further boost the performance of the HAN model.

According to results in Table 7, we see that CAN and HAN achieved better results between all categories. These two methods utilize some similar techniques, which are Led to their success. First, both of them consider class-level alignment and domain-level alignment, which lead to better matching between domains. Second, both of them use discrepancy measurement in their models, which suggests that utilizing metric-learning for domain adaptation can lead to large developments in classification accuracy for domain adaptation.

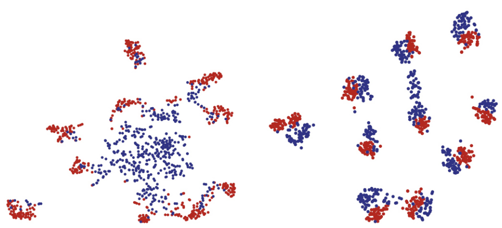
In Table 8, we compare various state-of-the-art methods in deep visual UDA for classification tasks on the ImageCLEF dataset. Results show that SymNets achieves competitive average classification accuracy on the ImageCLEF dataset. The accuracy of SymNets is 1.5% higher than the second-best method HAN. SymNets applies an adversarial learning method to overcome the limitation in matching the joint distributions of feature and class across domains via two-level domain confusion losses. Class-level confusion loss boosts over domain-level one via driving the learning of intermediate network representations to be invariant in the corresponding classes of two domains. So, the result obtained from Table 8 also confirms the first result in Table 7 about consideration on class-level and domain-level alignments.

In Table 9, we compare various state-of-the-art methods in deep visual UDA for classification tasks on the SVHN-MNIST-USPS dataset. Results on these three digits datasets show that SWD achieves the best accuracy between listed methods in 9 with 98.9%



(a) Source only (b) SWD adapted

Fig. 6: The t-SNE [178] visualization of features obtained from SVHN to MNIST adaptation by (a) source domain only, and (b) SWD adaptation. Blue and red points denote the source and target samples, respectively. [33].



(a) MMD (b) TarGAN

Fig. 7: The t-SNE [178] embeddings of 1000 test samples from SVHN (source, red) and MNIST (target, blue). (a) MMD metric, and (b) TarGAN method. [69].

IET Research Journals, pp. 1–18

on SVHN → MNIST, 98.1% on MNIST → USPS, 97.1 on USPS → MNIST. This represents the importance of using the task-specific decision boundaries (discrepancy) to guide the domain adaptation instead of simply matching the distributions among the source and target domains in pixel, feature, or output space in most of the other domain adaptation methods. We can see a visualization of features obtained from SVHN to MNIST adaptation by (a) source domain only, and (b) SWD adaptation in Fig 6. From Table 9, it is clear that TarGAN method achieves competitive results in comparison with the other methods in adversarial-based category. As we can see in Fig. 7, the intermediate feature obtained by TarGAN is more discriminative over the target domain compared with the baseline trained with MMD. This is consistent with the truth that domain invariance doesn't necessarily imply discriminative representations on the target data.

11 summary

In this paper, we provided a survey of different methods for deep visual unsupervised domain adaptation for classification tasks. We categorized the image classification methods into five main groups based on the technology of adopted for domain adaptation: discrepancy-, adversarial-, reconstruction-, representation-, and attention-based methods. Then, each of these groups were further categorized into several subgroups. Also, some benchmark datasets for deep visual domain adaptation were investigated. Then, we reviewed some recent papers for different applications in deep visual domain adaptation, for example, image classification, semantic segmentation, object detection, text recognition, person re-identification. Finally, some domain adaptation experiments reported in the reviewed papers included in our taxonomy were summarized and discussed. Experiments results indicate that using metric-learning in domain adaptation can lead to large advancements in classification accuracy for the domain adaptation task. Also, the results are highly affirming the effectiveness of utilizing class-level and domain-level in aligning the joint distributions of feature and category across domains.

Although deep visual UDA has achieved success recently, many issues still remain to be addressed. Some of future research challenges and directions can be as follows:

The traditional UDA algorithms assume that the training and test data have the same feature spaces, while this assumption doesn't necessarily hold in real applications. So it will be good to extend UDA models from traditional assumption to the new one.

Recently, some new methods to reduce negative transfer are proposed such areas as partial domain adaptation, attention-based domain adaptation, which focus on some transferable attention regions or images from source data and relating them to the target data. We believe these issues are worthy of more attention.

Finally, according to obtained experimental results, discrepancy-based and adversarial-based methods could obtain better performance. So, we suggest introducing new models with a combination of these methods.

12 References

[1] Pan, S.J., Yang, Q.: 'A survey on transfer learning', *IEEE Transactions on Knowledge and Data Engineering*, 2010, **22**, (10), pp. 1345–1359
[2] Shao, L., Zhu, F., Li, X.: 'Transfer learning for visual categorization: A survey', *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**, (5), pp. 1019–1034
[3] Patel, V.M., Gopalan, R., Li, R., Chellappa, R.: 'Visual domain adaptation: A survey of recent advances', *IEEE Signal Processing Magazine*, 2015, **32**, (3), pp. 53–69
[4] Csurka, G. 'A comprehensive survey on domain adaptation for visual applications'. In: Domain adaptation in computer vision applications. (Springer, 2017. pp. 1–35
[5] Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C. 'A survey on deep transfer learning'. In: Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I., editors. Artificial Neural Networks and Machine Learning – ICANN 2018. (Cham: Springer International Publishing, 2018. pp. 270–279
[6] Wang, M., Deng, W.: 'Deep visual domain adaptation: A survey', *Neurocomputing*, 2018, **312**, pp. 135–153
[7] Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: 'Deep domain confusion: Maximizing for domain invariance', *arXiv preprint arXiv:14123474*, 2014,
[8] Long, M., Cao, Y., Wang, J., Jordan, M. 'Learning transferable features with deep adaptation networks'. In: International Conference on Machine Learning. (, 2015. pp. 97–105
[9] Long, M., Zhu, H., Wang, J., Jordan, M.I. 'Unsupervised domain adaptation with residual transfer networks'. In: Advances in Neural Information Processing Systems. (, 2016. pp. 136–144
[10] Long, M., Zhu, H., Wang, J., Jordan, M.I. 'Deep transfer learning with joint adaptation networks'. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. (JMLR. org, 2017. pp. 2208–2217
[11] Yan, H., Ding, Y., Li, P., Wang, Q., Xu, Y., Zuo, W. 'Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation'. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (, 2017.
[12] Luo, Z., Hu, J., Deng, W., Shen, H. 'Deep unsupervised domain adaptation for face recognition'. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). (IEEE, 2018. pp. 453–457
[13] Wu, X., Chen, J., Yu, F., Yao, M., Luo, J.: 'Joint learning of multiple latent domains and deep representations for domain adaptation', *IEEE transactions on cybernetics*, 2019,
[14] Zhu, Y., Zhuang, F., Wang, J., Chen, J., Shi, Z., Wu, W., et al.: 'Multi-representation adaptation network for cross-domain image classification', *Neural Networks*, 2019, **119**, pp. 214–221
[15] Kang, G., Jiang, L., Yang, Y., Hauptmann, A.G. 'Contrastive adaptation network for unsupervised domain adaptation'. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (, 2019.
[16] Deng, W., Zheng, L., Sun, Y., Jiao, J.: 'Rethinking triplet loss for domain adaptation', *IEEE Transactions on Circuits and Systems for Video Technology*, 2020,
[17] Sun, B., Saenko, K. 'Deep coral: Correlation alignment for deep domain adaptation'. In: European Conference on Computer Vision. (Springer, 2016. pp. 443–450
[18] Peng, X., Saenko, K. 'Synthetic to real adaptation with generative correlation alignment networks'. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). (IEEE, 2018. pp. 1982–1991
[19] Zhang, Y., Wang, N., Cai, S., Song, L.: 'Unsupervised domain adaptation by mapped correlation alignment', *IEEE Access*, 2018, **6**, pp. 44698–44706
[20] Chen, C., Chen, Z., Jiang, B., Jin, X. 'Joint domain alignment and discriminative feature learning for unsupervised deep domain adaptation'. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33. (, 2019. pp. 3296–3303
[21] Rahman, M.M., Fookes, C., Baktashmotlagh, M., Sridharan, S. 'On minimum discrepancy estimation for deep domain adaptation'. In: Domain Adaptation for Visual Understanding. (Springer, 2020. pp. 81–94
[22] Sohn, K., Shang, W., Yu, X., Chandraker, M.: 'Unsupervised domain adaptation for distance metric learning', , 2018,
[23] Roy, S., Siarohin, A., Sangineto, E., Bulo, S.R., Sebe, N., Ricci, E. 'Unsupervised domain adaptation using feature-whitening and consensus loss'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 9471–9480
[24] Mancini, M., Porzi, L., Cermelli, F., Caputo, B. 'Discovering latent domains for unsupervised domain adaptation through consistency'. In: International Conference on Image Analysis and Processing. (Springer, 2019. pp. 390–401
[25] Mancini, M., Porzi, L., Rota.Bulò, S., Caputo, B., Ricci, E. 'Boosting domain adaptation by discovering latent domains'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 3771–3780
[26] Li, Y., Wang, N., Shi, J., Hou, X., Liu, J.: 'Adaptive batch normalization for practical domain adaptation', *Pattern Recognition*, 2018, **80**, pp. 109–117
[27] Li, Y., Swersky, K., Zemel, R. 'Generative moment matching networks'. In: International Conference on Machine Learning. (, 2015. pp. 1718–1727
[28] Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B. 'Moment matching for multi-source domain adaptation'. In: The IEEE International Conference on Computer Vision (ICCV). (, 2019.
[29] Zellinger, W., Moser, B.A., Grubinger, T., Lughofer, E., Natschläger, T., Saminger.Platz, S.: 'Robust unsupervised domain adaptation for neural networks via moment alignment', *Information Sciences*, 2019, **483**, pp. 174–191
[30] Shen, J., Qu, Y., Zhang, W., Yu, Y.. 'Wasserstein distance guided representation learning for domain adaptation'. (, 2017
[31] Courty, N., Flamary, R., Habrard, A., Rakotomamonjy, A. 'Joint distribution optimal transportation for domain adaptation'. In: Advances in Neural Information Processing Systems. (, 2017. pp. 3730–3739
[32] Damodaran, B.B., Kellenberger, B., Flamary, R., Tuia, D., Courty, N. 'Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation'. In: European Conference on Computer Vision. (Springer, 2018. pp. 467–483
[33] Lee, C.Y., Batra, T., Baig, M.H., Ulbricht, D. 'Sliced wasserstein discrepancy for unsupervised domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 10285–10295
[34] Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: 'A kernel two-sample test', *Journal of Machine Learning Research*, 2012, **13**, (Mar), pp. 723–773
[35] Chwialkowski, K.P., Ramdas, A., Sejdinovic, D., Gretton, A. 'Fast two-sample testing with analytic representations of probability measures'. In: Advances in Neural Information Processing Systems. (, 2015. pp. 1981–1989
[36] Jitkrittum, W., Szabó, Z., Chwialkowski, K.P., Gretton, A. 'Interpretable distribution features with maximum testing power'. In: Advances in Neural Information Processing Systems. (, 2016. pp. 181–189

[37] Schroff, F., Kalenichenko, D., Philbin, J. 'Facenet: A unified embedding for face recognition and clustering'. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (, 2015. pp. 815–823
[38] Laradji, I.H., Babanezhad, R. 'M-adda: Unsupervised domain adaptation with deep metric learning'. In: Domain Adaptation for Visual Understanding. (Springer, 2020. pp. 17–31
[39] Sun, B., Feng, J., Saenko, K. 'Return of frustratingly easy domain adaptation'. In: Thirtieth AAAI Conference on Artificial Intelligence. (, 2016.
[40] Grandvalet, Y., Bengio, Y. 'Semi-supervised learning by entropy minimization'. In: Advances in neural information processing systems. (, 2005. pp. 529–536
[41] Ioffe, S., Szegedy, C. 'Batch normalization: Accelerating deep network training by reducing internal covariate shift'. In: International Conference on Machine Learning. (, 2015. pp. 448–456
[42] Cédric, V.. 'Optimal transport: old and new'. (Springer, 2009
[43] Goodfellow, I., Pouget.Abadie, J., Mirza, M., Xu, B., Warde.Farley, D., Ozair, S., et al. 'Generative adversarial nets'. In: Advances in neural information processing systems. (, 2014. pp. 2672–2680
[44] Yoo, D., Kim, N., Park, S., Paek, A.S., Kweon, I.S. 'Pixel-level domain transfer'. In: European Conference on Computer Vision. (Springer, 2016. pp. 517–532
[45] Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., Webb, R. 'Learning from simulated and unsupervised images through adversarial training'. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (, 2017. pp. 2107–2116
[46] Duan, Q., Zhang, L., Jia, W. 'Adv-kin: An adversarial convolutional network for kinship verification'. In: Chinese Conference on Biometric Recognition. (Springer, 2017. pp. 48–57
[47] Duan, Q., Zhang, L. 'Advnet: Adversarial contrastive residual net for 1 million kinship recognition'. In: Proceedings of the 2017 Workshop on Recognizing Families In the Wild. (ACM, 2017. pp. 21–29
[48] Hu, L., Kan, M., Shan, S., Chen, X. 'Duplex generative adversarial network for unsupervised domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 1498–1507
[49] Sankaranarayanan, S., Balaji, Y., Castillo, C.D., Chellappa, R. 'Generate to adapt: Aligning domains using generative adversarial networks'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 8503–8512
[50] Wei, K.Y., Hsu, C.T. 'Generative adversarial guided learning for domain adaptation.'. In: BMVC. (, 2018. p. 100
[51] Tran, L.Q., Yin, X., Liu, X.: 'Representation learning by rotating your faces', *IEEE transactions on pattern analysis and machine intelligence*, 2018,
[52] Zhang, K., Kang, Q., Pan, L., Wang, X., Cui, C. 'A generative adversarial distribution matching framework for visual domain adaptation'. In: 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC). (IEEE, 2019. pp. 252–256
[53] Zhu, J.Y., Park, T., Isola, P., Efros, A.A. 'Unpaired image-to-image translation using cycle-consistent adversarial networks'. In: Proceedings of the IEEE international conference on computer vision. (, 2017. pp. 2223–2232
[54] Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J. 'Learning to discover cross-domain relations with generative adversarial networks'. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. (JMLR. org, 2017. pp. 1857–1865
[55] Hosseini.Asl, E., Zhou, Y., Xiong, C., Socher, R.: 'Augmented cyclic adversarial learning for low resource domain adaptation', , 2018,
[56] Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., et al. 'Cycada: Cycle-consistent adversarial domain adaptation'. In: International Conference on Machine Learning. (, 2018. pp. 1989–1998
[57] Mirza, M., Osindero, S.: 'Conditional generative adversarial nets', *arXiv preprint arXiv:14111784*, 2014,
[58] Kontschieder, P., Fiterau, M., Criminisi, A., Rota.Bulo, S. 'Deep neural decision forests'. In: Proceedings of the IEEE international conference on computer vision. (, 2015. pp. 1467–1475
[59] Tran, L., Yin, X., Liu, X. 'Disentangled representation learning gan for pose-invariant face recognition'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2017. pp. 1415–1424
[60] Odena, A., Olah, C., Shlens, J. 'Conditional image synthesis with auxiliary classifier gans'. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. (JMLR. org, 2017. pp. 2642–2651
[61] Yin, X., Yu, X., Sohn, K., Liu, X., Chandraker, M. 'Towards large-pose face frontalization in the wild'. In: Proceedings of the IEEE International Conference on Computer Vision. (, 2017. pp. 3990–3999
[62] Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., Krishnan, D. 'Unsupervised pixel-level domain adaptation with generative adversarial networks'. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (, 2017. pp. 3722–3731
[63] Antoniou, A., Storkey, A., Edwards, H. 'Augmenting image classifiers using data augmentation generative adversarial networks'. In: International Conference on Artificial Neural Networks. (Springer, 2018. pp. 594–603
[64] Volpi, R., Morerio, P., Savarese, S., Murino, V. 'Adversarial feature augmentation for unsupervised domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 5495–5504
[65] Hu, Y., Wu, X., Yu, B., He, R., Sun, Z. 'Pose-guided photorealistic face rotation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 8398–8406
[66] Liu, F., Deng, X., Lai, Y.K., Liu, Y.J., Ma, C., Wang, H. 'Sketchgan: Joint sketch completion and recognition with generative adversarial network'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 5830–5839
[67] Liu, M.Y., Tuzel, O. 'Coupled generative adversarial networks'. In: Advances in neural information processing systems. (, 2016. pp. 469–477
[68] Liu, M.Y., Breuel, T., Kautz, J. 'Unsupervised image-to-image translation networks'. In: Advances in neural information processing systems. (, 2017. pp. 700–708
[69] Lv, F., Zhu, J., Yang, G., Duan, L.: 'Targan: Generating target data with class labels for unsupervised domain adaptation', *Knowledge-Based Systems*, 2019, **172**, pp. 123–129
[70] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., et al.: 'Domain-adversarial training of neural networks', *The Journal of Machine Learning Research*, 2016, **17**, (1), pp. 2096–2030
[71] Tzeng, E., Hoffman, J., Saenko, K., Darrell, T. 'Adversarial discriminative domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2017. pp. 7167–7176
[72] Pei, Z., Cao, Z., Long, M., Wang, J. 'Multi-adversarial domain adaptation'. In: Thirty-Second AAAI Conference on Artificial Intelligence. (, 2018.
[73] Qi, F., Yang, X., Xu, C. 'A unified framework for multimodal domain adaptation'. In: 2018 ACM Multimedia Conference on Multimedia Conference. (ACM, 2018. pp. 429–437
[74] Ganin, Y., Lempitsky, V. 'Unsupervised domain adaptation by backpropagation'. In: International Conference on Machine Learning. (, 2015. pp. 1180–1189
[75] Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., Erhan, D. 'Domain separation networks'. In: Advances in neural information processing systems. (, 2016. pp. 343–351
[76] Shen, J., Qu, Y., Zhang, W., Yu, Y.: 'Adversarial representation learning for domain adaptation', *stat*, 2017, **1050**, pp. 5
[77] Long, M., Cao, Z., Wang, J., Jordan, M.I. 'Conditional adversarial domain adaptation'. In: Advances in Neural Information Processing Systems. (, 2018. pp. 1640–1650
[78] Kumar, A., Sattigeri, P., Wadhawan, K., Karlinsky, L., Feris, R., Freeman, B., et al. 'Co-regularized alignment for unsupervised domain adaptation'. In: Advances in Neural Information Processing Systems. (, 2018. pp. 9345–9356
[79] Shu, R., Bui, H.H., Narui, H., Ermon, S. 'A dirt-t approach to unsupervised domain adaptation'. In: Proc. 6th International Conference on Learning Representations. (, 2018.
[80] Zhang, W., Ouyang, W., Li, W., Xu, D. 'Collaborative and adversarial network for unsupervised domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 3801–3809
[81] Xu, R., Chen, Z., Zuo, W., Yan, J., Lin, L. 'Deep cocktail network: Multi-source unsupervised domain adaptation with category shift'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 3964–3973
[82] Saito, K., Watanabe, K., Ushiku, Y., Harada, T. 'Maximum classifier discrepancy for unsupervised domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 3723–3732
[83] Chen, Q., Liu, Y., Wang, Z., Wassell, I., Chetty, K. 'Re-weighted adversarial adaptation network for unsupervised domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 7976–7985
[84] Pinheiro, P.O. 'Unsupervised domain adaptation with similarity learning'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 8004–8013
[85] Cicek, S., Soatto, S. 'Unsupervised domain adaptation via regularized conditional alignment'. In: The IEEE International Conference on Computer Vision (ICCV). (, 2019.
[86] Wen, J., Liu, R., Zheng, N., Zheng, Q., Gong, Z., Yuan, J. 'Exploiting local feature patterns for unsupervised domain adaptation'. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33. (, 2019. pp. 5401–5408
[87] Cai, G., Wang, Y., He, L., Zhou, M.: 'Unsupervised domain adaptation with adversarial residual transform networks', *IEEE Transactions on Neural Networks and Learning Systems*, 2019,
[88] Zhang, C., Zhao, Q., Wang, Y.: 'Hybrid adversarial network for unsupervised domain adaptation', *Information Sciences*, 2020, **514**, pp. 44–55
[89] Cao, Z., Long, M., Wang, J., Jordan, M.I. 'Partial transfer learning with selective adversarial networks'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 2724–2732
[90] Cao, Z., Ma, L., Long, M., Wang, J. 'Partial adversarial domain adaptation'. In: Proceedings of the European Conference on Computer Vision (ECCV). (, 2018. pp. 135–150
[91] Zhang, J., Ding, Z., Li, W., Ogunbona, P. 'Importance weighted adversarial nets for partial domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 8156–8164
[92] Cao, Z., You, K., Long, M., Wang, J., Yang, Q. 'Learning to transfer examples for partial domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 2985–2994
[93] Kamath, U., Liu, J., Whitaker, J. 'Transfer learning: Domain adaptation'. In: Deep Learning for NLP and Speech Recognition. (Springer, 2019. pp. 495–535
[94] Zhang, Y., Tang, H., Jia, K., Tan, M. 'Domain-symmetric networks for adversarial domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 5031–5040
[95] Courty, N., Flamary, R., Tuia, D., Rakotomamonjy, A.: 'Optimal transport for domain adaptation', *IEEE transactions on pattern analysis and machine intelligence*, 2016, **39**, (9), pp. 1853–1865
[96] Glorot, X., Bordes, A., Bengio, Y. 'Domain adaptation for large-scale sentiment classification: A deep learning approach'. In: Proceedings of the 28th international conference on machine learning (ICML-11). (, 2011. pp. 513–520

[97] Chen, M., Xu, Z., Weinberger, K.Q., Sha, F. 'Marginalized denoising autoencoders for domain adaptation'. In: Proceedings of the 29th International Coference on International Conference on Machine Learning. (, 2012. pp. 1627–1634
[98] Ghifary, M., Bastiaan.Kleijn, W., Zhang, M., Balduzzi, D. 'Domain generalization for object recognition with multi-task autoencoders'. In: Proceedings of the IEEE international conference on computer vision. (, 2015. pp. 2551–2559
[99] Mittal, P., Vatsa, M., Singh, R. 'Composite sketch recognition via deep network-a transfer learning approach'. In: 2015 International Conference on Biometrics (ICB). (IEEE, 2015. pp. 251–256
[100] Ghifary, M., Kleijn, W.B., Zhang, M., Balduzzi, D., Li, W. 'Deep reconstruction-classification networks for unsupervised domain adaptation'. In: European Conference on Computer Vision. (Springer, 2016. pp. 597–613
[101] Wen, L., Gao, L., Li, X.: 'A new deep transfer learning based on sparse auto-encoder for fault diagnosis', *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017, **49**, (1), pp. 136–144
[102] Murez, Z., Kolouri, S., Kriegman, D., Ramamoorthi, R., Kim, K. 'Image to image translation for domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 4500–4509
[103] Jiang, B., Chen, C., Jin, X.: 'Unsupervised domain adaptation with target reconstruction and label confusion in the common subspace', *Neural Computing and Applications*, 2018, pp. 1–14
[104] Ding, Z., Shao, M., Fu, Y. 'Deep low-rank coding for transfer learning'. In: Twenty-Fourth International Joint Conference on Artificial Intelligence. (, 2015.
[105] Yang, B., Ma, A.J., Yuen, P.C. 'Domain-shared group-sparse dictionary learning for unsupervised domain adaptation'. In: Thirty-Second AAAI Conference on Artificial Intelligence. (, 2018.
[106] Yang, B., Ma, A.J., Yuen, P.C.: 'Learning domain-shared group-sparse representation for unsupervised domain adaptation', *Pattern Recognition*, 2018, **81**, pp. 615–632
[107] Ding, Z., Fu, Y.: 'Deep transfer low-rank coding for cross-domain learning', *IEEE transactions on neural networks and learning systems*, 2018, **30**, (6), pp. 1768–1779
[108] Gholami, B., Pavlovic, V., et al. 'Punda: Probabilistic unsupervised domain adaptation for knowledge transfer across visual categories'. In: Proceedings of the IEEE International Conference on Computer Vision. (, 2017. pp. 3581–3590
[109] Karbalayghareh, A., Qian, X., Dougherty, E.R.: 'Optimal bayesian transfer learning', *IEEE Transactions on Signal Processing*, 2018, **66**, (14), pp. 3724–3739
[110] Perrone, V., Jenatton, R., Seeger, M.W., Archambeau, C. 'Scalable hyperparameter transfer learning'. In: Advances in Neural Information Processing Systems. (, 2018. pp. 6845–6855
[111] Das, D., Lee, C.G. 'Graph matching and pseudo-label guided deep unsupervised domain adaptation'. In: International Conference on Artificial Neural Networks. (Springer, 2018. pp. 342–352
[112] Ding, Z., Li, S., Shao, M., Fu, Y. 'Graph adaptive knowledge transfer for unsupervised domain adaptation'. In: Proceedings of the European Conference on Computer Vision (ECCV). (, 2018. pp. 37–52
[113] Ma, X., Zhang, T., Xu, C. 'Gcan: Graph convolutional adversarial network for unsupervised domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 8266–8276
[114] Mancini, M., Bulò, S.R., Caputo, B., Ricci, E. 'Adagraph: Unifying predictive and continuous domain adaptation through graphs'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 6568–6577
[115] Bengio, Y., et al.: 'Learning deep architectures for ai', *Foundations and trends® in Machine Learning*, 2009, **2**, (1), pp. 1–127
[116] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: 'Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion', *Journal of machine learning research*, 2010, **11**, (Dec), pp. 3371–3408
[117] Bousmalis, K., Silberman, N., Krishnan, D., Trigeorgis, G., Erhan, D.. 'Domain separation neural networks'. (Google Patents, 2019. uS Patent App. 16/321,189
[118] Peng, X., Li, Y., Murphey, Y.L., Luo, J.: 'Domain adaptation by stacked local constraint auto-encoder learning', *IEEE Access*, 2019, **7**, pp. 108248–108260
[119] Olshausen, B.A., Field, D.J.: 'Sparse coding with an overcomplete basis set: A strategy employed by v1?', *Vision research*, 1997, **37**, (23), pp. 3311–3325
[120] Tzeng, E., Hoffman, J., Darrell, T., Saenko, K. 'Simultaneous deep transfer across domains and tasks'. In: Proceedings of the IEEE International Conference on Computer Vision. (, 2015. pp. 4068–4076
[121] Rozantsev, A., Salzmann, M., Fua, P.: 'Beyond sharing weights for deep domain adaptation', *IEEE transactions on pattern analysis and machine intelligence*, 2018, **41**, (4), pp. 801–814
[122] Rozantsev, A., Salzmann, M., Fua, P. 'Residual parameter transfer for deep domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 4339–4348
[123] Chen, W.Y., Hsu, T.M.H., Tsai, Y.H.H., Wang, Y.C.F., Chen, M.S. 'Transfer neural trees for heterogeneous domain adaptation'. In: European Conference on Computer Vision. (Springer, 2016. pp. 399–414
[124] Sener, O., Song, H.O., Saxena, A., Savarese, S. 'Learning transferrable representations for unsupervised domain adaptation'. In: Advances in Neural Information Processing Systems. (, 2016. pp. 2110–2118
[125] Saito, K., Ushiku, Y., Harada, T. 'Asymmetric tri-training for unsupervised domain adaptation'. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70. (JMLR. org, 2017. pp. 2988–2997
[126] Liu, Y.C., Yeh, Y.Y., Fu, T.C., Wang, S.D., Chiu, W.C., Frank.Wang, Y.C. 'Detach and adapt: Learning cross-domain disentangled deep representation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 8867–8876
[127] Gonzalez.Garcia, A., van de Weijer, J., Bengio, Y. 'Image-to-image translation for cross-domain disentanglement'. In: Advances in Neural Information Processing Systems. (, 2018. pp. 1287–1298
[128] Liu, A.H., Liu, Y.C., Yeh, Y.Y., Wang, Y.C.F. 'A unified feature disentangler for multi-domain image translation and manipulation'. In: Advances in Neural Information Processing Systems. (, 2018. pp. 2590–2599
[129] Gholami, B., Sahu, P., Rudovic, O., Bousmalis, K., Pavlovic, V.: 'Unsupervised multi-target domain adaptation: An information theoretic approach', *IEEE Transactions on Image Processing*, 2020, **29**, pp. 3993–4002
[130] Tenenbaum, J.B., Freeman, W.T. 'Separating style and content'. In: Advances in neural information processing systems. (, 1997. pp. 662–668
[131] Hinton, G.E., Krizhevsky, A., Wang, S.D. 'Transforming auto-encoders'. In: International Conference on Artificial Neural Networks. (Springer, 2011. pp. 44–51
[132] Wang, X., Li, L., Ye, W., Long, M., Wang, J. 'Transferable attention for domain adaptation'. In: AAAI Conference on Artificial Intelligence (AAAI). (, 2019.
[133] Kang, G., Zheng, L., Yan, Y., Yang, Y. 'Deep adversarial attention alignment for unsupervised domain adaptation: the benefit of target expectation maximization'. In: Proceedings of the European Conference on Computer Vision (ECCV). (, 2018. pp. 401–416
[134] Zhuo, J., Wang, S., Zhang, W., Huang, Q. 'Deep unsupervised convolutional domain adaptation'. In: Proceedings of the 25th ACM international conference on Multimedia. (ACM, 2017. pp. 261–269
[135] Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., et al. 'Residual attention network for image classification'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2017. pp. 3156–3164
[136] Moon, S., Carbonell, J.G. 'Completely heterogeneous transfer learning with attention-what and what not to transfer.'. In: IJCAI. (, 2017. pp. 2508–2514
[137] Zhao, H., Zhang, S., Wu, G., Moura, J.M., Costeira, J.P., Gordon, G.J. 'Adversarial multiple source domain adaptation'. In: Advances in Neural Information Processing Systems. (, 2018. pp. 8559–8570
[138] Tsai, Y.H., Sohn, K., Schulter, S., Chandraker, M. 'Domain adaptation for structured output via discriminative patch representations'. In: Proceedings of the IEEE International Conference on Computer Vision. (, 2019. pp. 1456–1465
[139] Hong, W., Wang, Z., Yang, M., Yuan, J. 'Conditional generative adversarial network for structured domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 1335–1344
[140] Wu, Z., Han, X., Lin, Y.L., Gokhan.Uzunbas, M., Goldstein, T., Nam.Lim, S., et al. 'Dcan: Dual channel-wise alignment networks for unsupervised scene adaptation'. In: Proceedings of the European Conference on Computer Vision (ECCV). (, 2018. pp. 518–534
[141] Chen, Y., Li, W., Van.Gool, L. 'Road: Reality oriented adaptation for semantic segmentation of urban scenes'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 7892–7901
[142] Zou, Y., Yu, Z., Vijaya.Kumar, B., Wang, J. 'Unsupervised domain adaptation for semantic segmentation via class-balanced self-training'. In: Proceedings of the European Conference on Computer Vision (ECCV). (, 2018. pp. 289–305
[143] Zhang, J., Liang, C., Kuo, C.C.J. 'A fully convolutional tri-branch network (fctn) for domain adaptation'. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). (IEEE, 2018. pp. 3001–3005
[144] Huang, H., Huang, Q., Krahenbuhl, P. 'Domain transfer through deep activation matching'. In: Proceedings of the European Conference on Computer Vision (ECCV). (, 2018. pp. 590–605
[145] Xie, S., Zheng, Z., Chen, L., Chen, C. 'Learning semantic representations for unsupervised domain adaptation'. In: International Conference on Machine Learning. (, 2018. pp. 5419–5428
[146] Sankaranarayanan, S., Balaji, Y., Jain, A., Nam.Lim, S., Chellappa, R. 'Learning from synthetic data: Addressing domain shift for semantic segmentation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 3752–3761
[147] Mao, X., Wang, S., Zheng, L., Huang, Q.: 'Semantic invariant cross-domain image generation with generative adversarial networks', *Neurocomputing*, 2018, **293**, pp. 55–63
[148] Xu, W., Li, Y., Lu, C. 'Srda: Generating instance segmentation annotation via scanning, reasoning and domain adaptation'. In: Proceedings of the European Conference on Computer Vision (ECCV). (, 2018. pp. 120–136
[149] Chen, C., Dou, Q., Chen, H., Heng, P.A. 'Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation'. In: International Workshop on Machine Learning in Medical Imaging. (Springer, 2018. pp. 143–151
[150] Gong, R., Li, W., Chen, Y., Gool, L.V. 'Dlow: Domain flow for adaptation and generalization'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 2477–2486
[151] Hong, Z.W., Chen, Y.M., Yang, H.K., Su, S.Y., Shann, T.Y., Chang, Y.H., et al. 'Virtual-to-real: learning to control in visual semantic segmentation'. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. (, 2018. pp. 4912–4920
[152] Romijnders, R., Meletis, P., Dubbelman, G. 'A domain agnostic normalization layer for unsupervised adversarial domain adaptation'. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). (IEEE, 2019. pp. 1866–1875
[153] Luo, Y., Zheng, L., Guan, T., Yu, J., Yang, Y. 'Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 2507–2516
[154] Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P. 'Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019.

pp. 2517–2526
[155] Chen, Y., Li, W., Chen, X., Gool, L.V. 'Learning semantic segmentation from synthetic data: A geometrically guided input-output adaptation approach'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 1841–1850
[156] Mousavian, A., Toshev, A., Fišer, M., Košecká, J., Wahid, A., Davidson, J. 'Visual representations for semantic target driven navigation'. In: 2019 International Conference on Robotics and Automation (ICRA). (IEEE, 2019. pp. 8846–8852
[157] Inoue, N., Furuta, R., Yamasaki, T., Aizawa, K. 'Cross-domain weakly-supervised object detection through progressive domain adaptation'. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (, 2018. pp. 5001–5009
[158] Park, H., Ju, M., Moon, S., Yoo, C.D. 'Unsupervised domain adaptation for object detection using distribution matching in various feature level'. In: International Workshop on Digital Watermarking. (Springer, 2018. pp. 363–372
[159] Abdullah.Jamal, M., Li, H., Gong, B. 'Deep face detector adaptation without negative transfer or catastrophic forgetting'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 5608–5618
[160] Chen, Y., Li, W., Sakaridis, C., Dai, D., Van.Gool, L. 'Domain adaptive faster r-cnn for object detection in the wild'. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (, 2018. pp. 3339–3348
[161] Ren, S., He, K., Girshick, R., Sun, J. 'Faster r-cnn: Towards real-time object detection with region proposal networks'. In: Advances in neural information processing systems. (, 2015. pp. 91–99
[162] Hsu, H.K., Hung, W.C., Tseng, H.Y., Yao, C.H., Tsai, Y.H., Singh, M., et al. 'Progressive domain adaptation for object detection'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. (, 2019. pp. 1–5
[163] Yu, Y., Xu, X., Hu, X., Heng, P.A.: 'Dalocnet: Improving localization accuracy for domain adaptive object detection', *IEEE Access*, 2019, **7**, pp. 63155–63163
[164] Zhu, X., Pang, J., Yang, C., Shi, J., Lin, D. 'Adapting object detectors via selective cross-domain alignment'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 687–696
[165] Chen, X., Cardie, C. 'Multinomial adversarial networks for multi-domain text classification'. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). (, 2018. pp. 1226–1240
[166] Chen, F., Chen, Y.Y. 'Adversarial domain adaptation using artificial titles for abstractive title generation'. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. (, 2019. pp. 2197–2203
[167] Zhang, Y., Nie, S., Liu, W., Xu, X., Zhang, D., Shen, H.T. 'Sequence-to-sequence domain adaptation network for robust text image recognition'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2019. pp. 2740–2749
[168] Zhan, F., Xue, C., Lu, S. 'Ga-dan: Geometry-aware domain adaptation network for scene text detection and recognition'. In: Proceedings of the IEEE International Conference on Computer Vision. (, 2019. pp. 9105–9115
[169] Li, Y.J., Yang, F.E., Liu, Y.C., Yeh, Y.Y., Du, X., Frank.Wang, Y.C. 'Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. (, 2018. pp. 172–178
[170] Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y. 'Camera style adaptation for person re-identification'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 5157–5166
[171] Sun, W., Liu, F., Xu, W. 'Unlabeled samples generated by gan improve the person re-identification baseline'. In: Proceedings of the 2019 5th International Conference on Computer and Technology Applications. (ACM, 2019. pp. 117–123
[172] Bak, S., Carr, P., Lalonde, J.F. 'Domain adaptation through synthesis for unsupervised person re-identification'. In: Proceedings of the European Conference on Computer Vision (ECCV). (, 2018. pp. 189–205
[173] Deng, W., Zheng, L., Ye, Q., Kang, G., Yang, Y., Jiao, J. 'Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 994–1003
[174] Wang, J., Zhu, X., Gong, S., Li, W. 'Transferable joint attribute-identity deep learning for unsupervised person re-identification'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (, 2018. pp. 2275–2284
[175] Li, M., Zhu, X., Gong, S. 'Unsupervised person re-identification by deep learning tracklet association'. In: Proceedings of the European Conference on Computer Vision (ECCV). (, 2018. pp. 737–753
[176] Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q.: 'Domain adaptation via transfer component analysis', *IEEE Transactions on Neural Networks*, 2010, **22**, (2), pp. 199–210
[177] Gong, B., Shi, Y., Sha, F., Grauman, K. 'Geodesic flow kernel for unsupervised domain adaptation'. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. (IEEE, 2012. pp. 2066–2073
[178] Maaten, L.v.d., Hinton, G.: 'Visualizing data using t-sne', *Journal of machine learning research*, 2008, **9**, (Nov), pp. 2579–2605