2019

Recent advances in transfer learning for cross-dataset visual recognition: A problem-oriented perspective

Jing Zhang
University of Wollongong, jz960@uowmail.edu.au

Wanqing Li
University of Wollongong, wanqing@uow.edu.au

Philip O. Ogunbona
University of Wollongong, philipo@uow.edu.au

Dong Xu
University of Sydney

Recent advances in transfer learning for cross-dataset visual recognition: A problem-oriented perspective

Abstract

This article takes a problem-oriented perspective and presents a comprehensive review of transfer-learning methods, both shallow and deep, for cross-dataset visual recognition. Specifically, it categorises the cross-dataset recognition into 17 problems based on a set of carefully chosen data and label attributes. Such a problem-oriented taxonomy has allowed us to examine how different transfer-learning approaches tackle each problem and how well each problem has been researched to date. The comprehensive problem-oriented review of the advances in transfer learning with respect to the problem has not only revealed the challenges in transfer learning for visual recognition but also the problems (e.g., 8 of the 17 problems) that have been scarcely studied. This survey not only presents an up-to-date technical review for researchers but also a systematic approach and a reference for a machine-learning practitioner to categorise a real problem and to look up for a possible solution accordingly.

Disciplines

Engineering | Science and Technology Studies

Publication Details

Zhang, J., Li, W., Ogunbona, P. & Xu, D. (2019). Recent advances in transfer learning for cross-dataset visual recognition: A problem-oriented perspective. ACM Computing Surveys, 52 (1), 7-1-7-38.

Recent Advances in Transfer Learning for Cross-Dataset Visual Recognition: A Problem-Oriented Perspective

JING ZHANG, WANQING LI, and PHILIP OGUNBONA, University of Wollongong, Australia
DONG XU, University of Sydney, Australia

This paper takes a problem-oriented perspective and presents a comprehensive review of transfer learning methods, both shallow and deep, for cross-dataset visual recognition. Specifically, it categorises the cross-dataset recognition into seventeen problems based on a set of carefully chosen data and label attributes. Such a problem-oriented taxonomy has allowed us to examine how different transfer learning approaches tackle each problem and how well each problem has been researched to date. The comprehensive problem-oriented review of the advances in transfer learning with respect to the problem has not only revealed the challenges in transfer learning for visual recognition, but also the problems (e.g. eight of the seventeen problems) that have been scarcely studied. This survey not only presents an up-to-date technical review for researchers, but also a systematic approach and a reference for a machine learning practitioner to categorise a real problem and to look up for a possible solution accordingly.

1 INTRODUCTION

Humans have exceptional ability to transfer learning in one context to another context [178, 246]. Machine learning algorithms mostly inspired by human brains, however, usually require a huge amount of training examples to learn a new model from scratch and often fail to apply the learned model to test data acquired from the scenarios different from those of the training data mainly due to domain divergence and task divergence [170]. This is particularly true in visual recognition [223] where external factors such as environments, lighting, background, sensor types, view angles, and post-processing can cause the distribution shift or even feature space divergence of the same task in two datasets or even the tasks, i.e. categories of the objects, are different.

To use previously available data effectively for current tasks with scarce data, models or knowledge learned from one domain have to be transferred to a new domain for the current task. Transfer learning has been actively researched in the past decade and one of its topics, domain adaptation, has been especially extensively researched, where the previous and current tasks are the same. The extensive study has led to about a dozen of tutorial and survey papers published since 2009, from the analysis of the nature of dataset shift [184] to the formal definition and task-oriented categorization of transfer learning [170], and to the recent tutorial and survey on deep learning based domain adaptation [37, 228]. Most of these survey papers [12, 35, 154, 157, 160, 173, 195, 214, 228, 245] are method-driven and provide up to the time a review of the evolution of the technologies. Many of them are on particular topics, for instance, domain adaptation [12, 37, 157, 173, 214, 228], dataset shift [160], activity recognition [35], and speech and language processing [237]. While these review papers have provided researchers in the field valuable references and contributed significantly to the advances of the technologies, they have not examined the full landscape of transfer learning and maturity of technologies to serve as a reference for machine learning practitioners. Unlike these existing survey papers, this paper takes a new problem-oriented perspective and presents a comprehensive review of transfer learning methods for cross-dataset visual recognition. Specifically,

Authors' addresses: Jing Zhang; Wanqing Li; Philip Ogunbona, University of Wollongong, Northfields Ave, Wollongong, NSW, 2522, Australia, jz960@uowmail.edu.au, wanqing@uow.edu.au, philipo@uow.edu.au; Dong Xu, University of Sydney, Sydney, Australia, dong.xu@sydney.edu.au.

- It defines a set of data and label attributes, categorises in a fine-grained way the cross-dataset recognition into seventeen problems based on these attributes, and presents a comprehensive review of the transfer learning methods, both shallow and deep, developed to date for each problem.
- The paper has also provided an assessment of the suitability of widely used datasets for transfer learning in evaluating algorithms for each of the seventeen problems.
- The problem-oriented taxonomy has allowed us to examine how different transfer learning approaches tackle each problem, how well each problem has been studied to date and the available solutions to each problem.
- Through the problem-oriented analysis, challenges and future directions have been identified. Particularly, little studies have been reported on eight of the seventeen problems.
- This survey not only presents an up-to-date technical review for researchers, but also a systematic approach and a reference for a machine learning practitioner to categorise a real problem and to look up for a possible solution accordingly.

In addition, none of the previous survey papers covers all of the seventeen problems. For instance, Weiss et al. [245] focuses on nine (of the seventeen) problems on homogeneous and heterogeneous domain adaptation and transfer learning with heterogeneous label spaces; Venkateswara et al. [229] mainly reviewed the literature of two problems in homogeneous domain adaptation using deep-learning; and Csurka [37] focuses on seven problems in domain adaptation.

The rest of the paper is organised as follows. Section 2 explains the terminologies used in the paper, defines the problem-oriented taxonomy of cross-dataset recognition, and summarises the transfer learning approaches to cross-dataset recognition. The seventeen problems identified in the taxonomy are categorised into four scenarios: *homogeneous feature and label spaces*, *heterogeneous feature spaces*, *heterogeneous label spaces* and *heterogeneous feature and label spaces*. Sections 3 through 6 review and analyse respectively the advances of techniques in addressing the problems under the four scenarios. Section 7 discusses and examines the suitability of the most commonly used datasets for cross-dataset transfer learning for all the problems. Section 8 discusses the challenges and future research directions. Section 9 concludes the paper.

2 OVERVIEW

This section begins with the definitions of terminologies used throughout the paper and then provides a summary of the approaches that have been developed for transfer learning.

2.1 Terminologies and Definitions

In this paper, we follow the definitions of "domain" and "task" given by [170].

Definition 2.1. (**Domain** [170]) "A domain is defined as $\mathcal{D} = \{\mathcal{X}, P(\mathbf{x})\}$, which is composed of two components: a feature space $\mathcal{X}$ and a marginal probability distribution $P(\mathbf{x})$, where $\mathbf{x} \in \mathcal{X}$."

Definition 2.2. (**Task** [170]) "Given a specific domain, a task is defined as $\mathcal{T} = \{\mathcal{Y}, f(\mathbf{x})\}$, which is composed of two components: a label space $\mathcal{Y}$ and a predictive function $f(\mathbf{x})$, where $f(\mathbf{x})$ can be seen as a conditional distribution $P(y|\mathbf{x})$ and $y \in \mathcal{Y}$."

Definition 2.3. (**Dataset**) A dataset is defined as $\mathcal{S} = \{\mathcal{N}, \mathcal{X}, P(\mathbf{x}), \mathcal{Y}, f(\mathbf{x})\}$, which is a collection of $\mathcal{N}$ data that belong to a specific domain $\mathcal{D} = \{\mathcal{X}, P(\mathbf{x})\}$ with a specific task $\mathcal{T} = \{\mathcal{Y}, f(\mathbf{x})\}$.

Often $P(\mathbf{x})$ and $f(\mathbf{x})$ are unknown and need to be estimated and learned respectively. If for each sample in the dataset $\mathcal{N}$ its label $y \in \mathcal{Y}$ is given, S is labelled, Otherwise, S is unlabelled.

Definition 2.4. (**Transfer Learning** [170]) "In general, given a source domain $\mathcal{D}_S$ and learning task $\mathcal{T}_S$, a target domain $\mathcal{D}_T$ and learning task $\mathcal{T}_T$, transfer learning aims to help improve the learning of the target predictive function $f_T(\cdot)$ in $\mathcal{D}_T$ using the knowledge in $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$, or $\mathcal{T}_S \neq \mathcal{T}_T$." Note that a special topic where $\mathcal{T}_S = \mathcal{T}_T$ and $\mathcal{D}_S \neq \mathcal{D}_T$ is known as *Domain Adaptation*. Specifically, in the context of cross-dataset recognition, the aim of transfer learning is to learn a robust classifier $f(x)$ from a dataset (i.e. target dataset $\mathcal{S}_T$) by effectively utilising the knowledge offered through other datasets (i.e. source datasets $\mathcal{S}_S$).

2.2 Problem-oriented Taxonomy of Cross-dataset Recognition

In cross-dataset recognition, there are often two datasets. One, referred to as a source dataset, is used in training and the other, referred to as a target dataset, is to be recognized. Their domains and/or tasks are different and their characteristics determines what methods can or should be used. In this paper, we define a set of attributes to characterise the source or target datasets. These attributes have led to a comprehensive taxonomy of cross-dataset recognition problems that provides a unique perspective for this survey.

- **Attributes on data:**
  - *Feature space:* the consistency of feature spaces (i.e. different feature extraction methods or different data modalities) between the source and target datasets.
  - *Data availability:* the availability and sufficiency of target data in the training stage.
  - *Balanced data:* whether the numbers of data samples in each class are balanced.
  - *Sequential/Online data:* whether the data are sequential/online and evolving over time.
- **Attributes on label:**
  - *Label availability:* the availability of labels in source and target datasets.
  - *Label space:* whether the data categories of the two datasets are identical.

Based on these attributes, the following four scenarios are defined as the first layer of the problem taxonomy to guide the survey.

- *Homogeneous feature spaces and label spaces:* The feature spaces and label spaces of the source and target datasets are identical. But domain divergence (i.e. different data distributions) exists across the source and target datasets.
- *Heterogeneous feature spaces:* the feature spaces of the source and target datasets are different (i.e. domain divergence occurs), but their label spaces are the same.
- *Heterogeneous label spaces:* the label spaces of the source and target datasets are different (i.e. task divergence occurs), but their feature spaces are the same.
- *Heterogeneous feature spaces and label spaces:* both the feature spaces and the label spaces of the source and target datasets are different (i.e. both domain and task divergence occurs).

The problems corresponding to the four scenarios are further divided into sub-problems using other data attributes such as the data being balanced and/or sequential/online. Fig. 1 shows the problem-oriented taxonomy for cross-dataset recognition, which shows seventeen different problems.

2.3 Approaches

Many approaches have been developed for transfer learning across datasets [170] at instance level, i.e. re-weighting some source samples based on their divergence from the target domain, at the feature level, i.e. learning "good" feature representations that have minimum domain shift, and at the classifier level, i.e. learn an optimal target classification

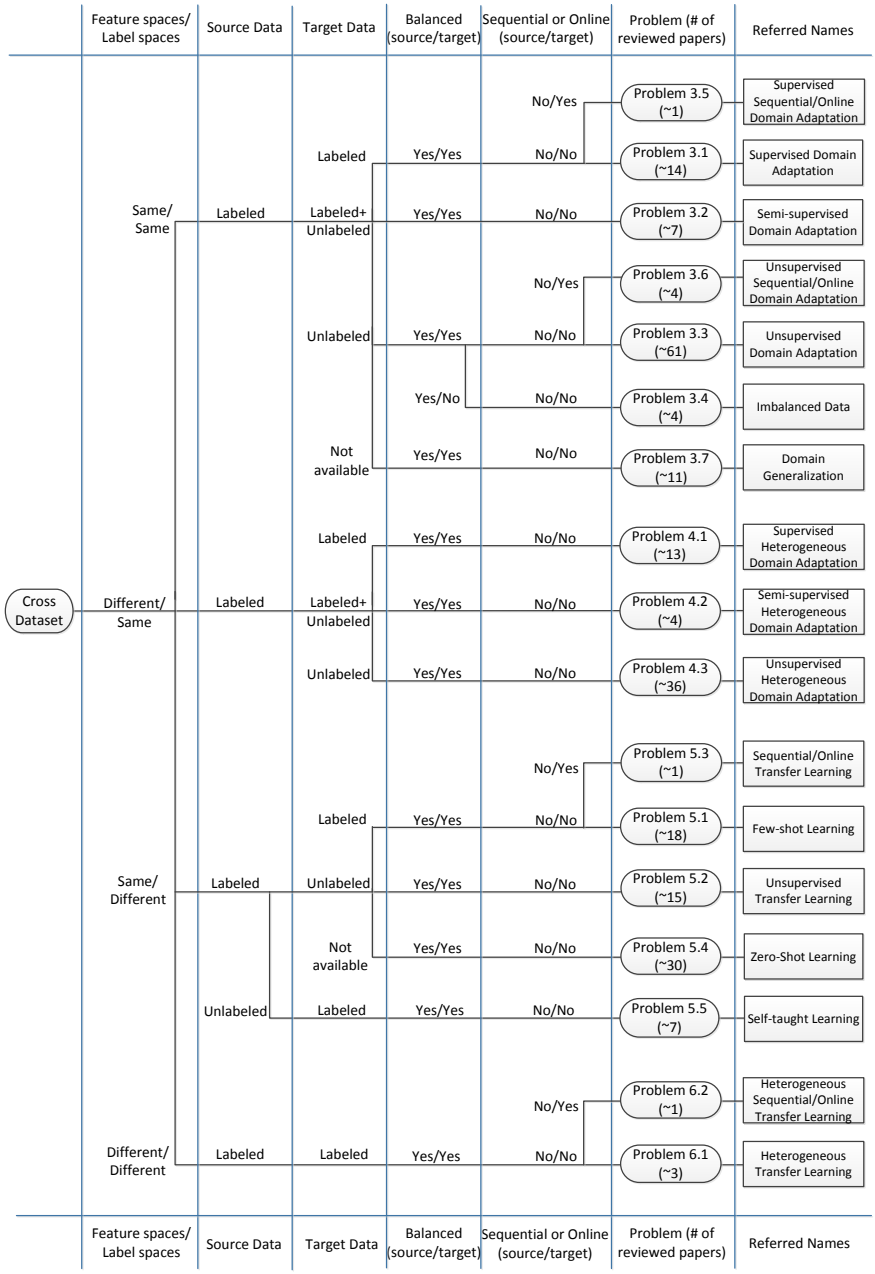

Fig. 1. A problem-oriented taxonomy for cross-dataset recognition including the number of papers that are found to address the problems.

model by using the data from both source and target domains as well as the source model. This section summarises several most typical approaches to transfer learning for cross-dataset recognition, including *Statistical approach*, *Geometric approach*, *Higher-level Representation*, *Correspondence approach*, *Class-based approach*, *Self Labelling*, and *Hybrid approach*. These approaches have been reported explicitly or implicitly in the literature. In particular, the basic assumptions of each approach are analysed and presented in this section. Moreover, several commonly used methods are illustrated under each approach. *Due to page limit, only brief description of each approach and its methods is presented. See the supplementary material for details.*

Statistical Approach: is employed in transferring the knowledge at the levels of instances, features and classifiers by measuring and minimizing the divergence of statistical distributions between the source and target datasets. This approach generally assumes sufficient data in each dataset to approximate the respective statistical distributions. The typical methods are Instance re-weighting [97], Feature space mapping [169] and Classifier parameter mapping [183].

Geometric Approach: bridges datasets according to their geometrical properties. It assumes domain shift can be reduced using the relationship of geometric structures between the source and target datasets. Typical methods include Subspace alignment [62], Intermediate subspaces [81, 85], and Manifold alignment (without correspondence) [39].

Higher-level Representation Approach: aims at finding higher-level representations that are representative, compact, and invariant between datasets. This approach does not require any labelled data, or the existence of correspondence set, but assumes that there exist the domain invariant higher-level representations between datasets. Note that this approach is commonly used together with other approaches for better transfer, but it is also used independently without any mechanism to reduce the domain divergence explicitly. Typical methods are Sparse coding [185], Low-rank representation [196], Deep Neural Networks [50, 189, 269], Stacked Denoising Auto-encoders (SDAs) [27, 77], and Attribute space [2, 124].

Correspondence Approach: uses paired correspondence samples from different domains to construct the relationship between domains. A set of corresponding samples (i.e. the same object captured from different view angles, or by different sensors) are required. The typical methods are Sparse coding with correspondence [285] and Manifold alignment (with correspondence) [271] .

Class-based Approach: uses label information as a guidance for connecting the source and target datasets. Hence, the labelled data from each dataset are assumed to be available, whether sufficient or not. The commonly used methods include Feature augmentation [44], Metric learning [193], Linear Discriminative Model [264], and Bayesian Model [60].

Self Labelling: uses the source domain samples to train an initial model to obtain the pseudo labels of target domain data. Then the target data and their pseudo labels are incorporated to retrain the model. The procedure continues iteratively until convergence. A typical example is Self-training [43, 216].

Hybrid Approach: combines two or more above approaches for better transferring of knowledge. Several example combinations are Correspondence and Higher-level representation [96], Higher-level representation and Statistic [147, 148, 243], Statistic and Geometric [273], Statistic and Self labelling [42], Correspondence and Class-based [46], Statistic and Class-based [52], and Higher-level representation and Class-based [288].

In the following sections, we present a comprehensive review on what approaches have been or can be used for the cross-dataset recognition problems shown in Figure 1.

3 HOMOGENEOUS FEATURE SPACES AND LABEL SPACES

In this scenario, $\mathcal{X}_S = \mathcal{X}_T$ and $\mathcal{Y}_S = \mathcal{Y}_T$. Hence, the $\mathcal{S}_S$ and $\mathcal{S}_T$ are generally different in their distributions ($P(X,Y)$). Sufficiently labelled source domain data are generally assumed available and different assumptions are made on the target domain, leading to different sub-problems.

3.1 Labelled Target Dataset

In this problem, a small number of labelled data in target domain are available. However, the labelled target data are generally insufficient for learning an effective classifier. This is also called *supervised domain adaptation* or *few-shot domain adaptation* in the literature.

Class-based Approach. The most commonly used approach in supervised domain adaptation is class-based since the labelled data from both domains are available in the training stage. For example, Daumé III [44] propose a feature augmentation based method where each feature is replicated into a high-dimensional space Φ containing the general and domain-specific version.

$$\Phi_s(x) = [x_s, x_s, 0]; \quad \Phi_t(x) = [x_t, 0, x_t]; \tag{1}$$

where $x_s \in \mathbb{R}^{f \times n_s}$ is the source domain data, $x_t \in \mathbb{R}^{f \times n_t}$ is the target domain data, f is the feature dimension, n_s and n_t are the total number of samples in the source and target domains, respectively.

The idea of supervised metric learning has also been used [179, 279]. The core idea is to exploit the task relationships between domains to boost the target task. Another group of methods [105, 254, 264] transfer the parameters of discriminative classifiers (e.g. SVM) across datasets. Recently, Motiian et al. [161] propose to create pairs of source and target instances to handle the scarce target labelled data. In addition, they extend adversarial learning [84] to align the semantic information of classes.

A more realistic setting is that samples from only a subset of classes are available in the target domain. Then the adapted features are generalized to unseen categories in the target dataset. While some categories are not available in the target dataset, we still assume the same label spaces between the two domains. So we discuss these methods under the problem of homogeneous label spaces. Generally, these methods assume the shift between domains is category-independent. For example, Saenko et al. [193] present a supervised metric learning-based method to learn a metric that minimizes the distribution shift by using target labelled data from a subset of categories:

$$\min Tr(W) - \log det(W) \quad s.t. \quad d_W(x_s, x_t) < u \quad \text{if} \quad y_s = y_t, \quad d_W(x_s, x_t) > l \quad \text{if} \quad y_s \neq y_t \tag{2}$$

where $u, l \in R$ are the threshold parameters, x_s and x_t represent the source domain sample and target domain sample, respectively and y_s and y_t represent their corresponding labels, $d_W = (\mathbf{x}_s - \mathbf{x}_t)^T W (\mathbf{x}_s - \mathbf{x}_t)$ is the distance between x_s and x_t, and W is the distance matrix that will be learned. Then the transformation is applied to unseen target test data that may come from different categories from the target training data. Similarly, some recent methods learn to recognize unseen target categories (but have been seen in the source domain) under the deep learning frameworks by exploiting the semantic structure either via soft labels (which is the averaged softmax activations over all source samples in each category) [225] or by the Siamese architecture [162]. For example, Figure 2 illustrates the network architecture of the domain and task transfer method proposed by Tzeng et. al. [225], which uses soft labels. In this work [225], the learned source semantic structure is transferred to the target domain by optimizing the network to produce activation distributions that match those learned for source data.

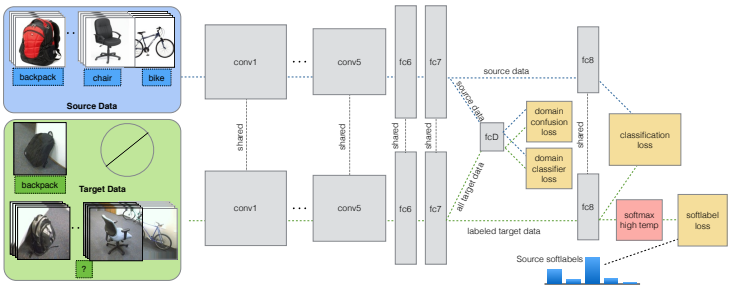

Fig. 2. The network architecture of the domain and task transfer method [225]. (Figure used courtesy of [225])

Self Labelling. Dai et al. [43] propose TrAdaBoost to extend boosting-based methods by decreasing the weights of the instances that are most dissimilar to the target distribution in order to weaken their impacts.

Hybrid Approach. The higher-level representation approach and class-based approach have been used together for better cross-dataset representation. For example, the discriminative dictionary can be learned such that the same class samples from different domains have similar sparse codes. [198, 290]. Except for the discriminative dictionary learning, the label information can also be used for guiding the deep neural networks to reduce domain shift. For example, Koniusz et al. [116] fuse the source and target CNN streams at the classifier level, where the scatters of the two network streams of the same class are aligned while the between-class are separated.

3.2 Labelled plus Unlabelled Target Dataset

Compared to the scenario where only limited labelled target data are presented, additional redundant unlabelled target data are also presented in training in this problem (often known as *semi-supervised domain adaptation* in the literature) to provide additional structural information. This setting is realistic in real-world applications because unlabelled data are easy to obtain.

Class-based Approach. Duan et al. [55] extend SVM-based supervised classifier transfer methods with unlabelled target data. They proposed a regularizer which enforces that the learned target classifiers and the pre-learned source classifiers should have the similar decision values on the unlabelled target instances:

$$\Omega_D(\mathbf{f}_u^T) = \frac{1}{2}\sum_{i=n_l+1}^{n_T}\sum_s \gamma_s (f_i^T - f_i^s)^2 = \frac{1}{2}\sum_s \gamma_s \|\mathbf{f}_u^T - \mathbf{f}_u^s\|^2 \tag{3}$$

where $\mathbf{f}_u^T = [f_{n_l+1}^T, ..., f_{n_T}^T]'$ and $\mathbf{f}_u^s = [f_{n_l+1}^s, ..., f_{n_T}^s]'$ represent the decision values of the unlabelled target samples from the target classifier and the s-th auxiliary classifier, n_l and n_T are the number of labelled target samples and the total number of target samples, γ_s is the weight for measuring the relevance between the s-th source domain and the target domain.

Self Labelling. Some researches extend distance-based classifiers, such as the k-Nearest Neighbour [219] and Nearest Class Mean [38] classifiers, to learn the domain invariant metric iteratively. Specifically, Tommasi and Caputo [219] present a method that learns a metric per class based on the NBNN algorithm. by progressively selecting target instances and combining it with a subset of the source data while imposing a large margin separation hyperplanes among classes. Similarly, Csurka et al. [38] extend the NCM classifier to a Domain Specific Class Means (DSCM) classifier and iteratively add high confidence unlabelled target samples to the training set. A co-training-based method is proposed by [26] to facilitate the gradual inclusion of target features and instances in training. This method iteratively learns feature views and a target predictor upon the views.

Hybrid Approach. A group of methods for semi-supervised domain adaptation combines class-based and statistical approach to make use of both labelled and unlabelled target data. The key idea is that the statistical criteria (e.g. MMD metric between source data and unlabelled target data) are used as an additional constraint in discriminative learning methods (e.g. multiple kernel learning (MKL) [52, 54], or least square method [266]).

Yamada et al. [261] generalize the EASYADAPT method [44] to semi-supervised setting. They proposed to project input features into a higher dimensional space as well as estimate weights for the training samples based on the ratio of test and training marginal distributions in that space using unlabelled target samples.

3.3 Unlabelled Target Dataset

In this problem, no labelled target domain data are available but sufficient unlabelled target domain data are observable for transfer learning. This problem is also named *unsupervised domain adaptation*. The unsupervised domain adaptation has attracted increasing attention nowadays, which is certainly more realistic and challenging.

Statistical Approach. The Maximum Mean Discrepancy (MMD) criterion is commonly used in unsupervised domain adaptation. Generally, the MMD distance between domains is reduced by re-weighting the samples [78, 97, 213], or mapping to another feature space [9, 150, 169, 273], or regularizing the source domain classifier using target domain unlabelled data [149, 183]. For example, Pan et al. [169] proposed to find a domain invariant feature mapping function ϕ such that the marginal distributions between the two domains P_s and P_t in the mapped feature space is small when using the MMD criterion:

$$D_{MMD}(P_s, P_t) = \| \frac{1}{n_s} \sum_{\mathbf{x}_i \in X_s} \phi(\mathbf{x}_i) - \frac{1}{n_t} \sum_{\mathbf{x}_j \in X_t} \phi(\mathbf{x}_j) \|_F^2 \tag{4}$$

Except for MMD, other statistical criteria, such as Kullback-Leibler divergence [208], Hellinger distance [10], Quadratic divergence [204], and mutual information [200], are also used for comparing two distributions. Sun et al. [210] propose the CORrelation ALignment (CORAL) to minimize distribution divergence by mapping the covariance of data.

Instead of learning a global transformation, Optimal Transport [36] learns a local transformation such that each source datum is mapped to target data and the marginal distribution is preserved.

Rather than assuming single domain in a dataset, some methods assume a dataset may contain several distinctive sub-domains due to the large variations in visual data. For example, Gong et al. [79] automatically discover latent domains from multi-source domains to characterize the inter-domain variations and, hence, to construct discriminative models.

Geometric Approach. Gopalan et al. [85] proposed a Sampling Geodesic Flow (SGF) method by sampling intermediate subspace representations between the source and target generative subspaces. The two generative subspaces are viewed as two points on a manifold. Then they sample the intermediate subspaces on the geodesic flow between the two subspaces. Lastly, all the data are mapped to the concatenation of all the subspaces to obtain the final representation. Figure 3 illustrates the SGF method. Gong et al. [81] extend SGF to a geodesic flow kernel (GFK) method by proposing a kernel method, such that an infinite number of subspaces are integrated to represent the incremental changes. The methods in [85, 86] and [80, 81] open the opportunity for researches to construct intermediate representations to characterize the domain changes. For example, Zhang et al. [283] bridge the source and target domains by inserting virtual views along a virtual path for cross-view recognition. Rather than manipulating on the subspaces, Cui et al. [40] represent source and target domains as covariance matrices and interpolate some intermediate covariance matrices to bridge the two domains. Some methods [165, 253] are proposed to generate several intermediate domains by learning

the domain-adaptive dictionaries between domains. The idea of intermediate domains is also employed in the deep learning framework [32].

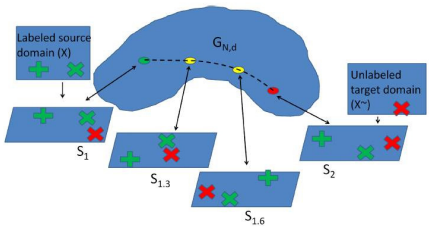

Fig. 3. Illustration of the SGF method (Figure used courtesy of [85])

Instead of modelling intermediate domains, some methods align the two domains directly [4, 39, 62, 153]. For instance, Fernando et al. [62] propose to align the source subspace to the target subspace directly by learning a linear transformation function.

Higher-level Representation. The low-rank criterion is commonly used to learn the domain invariant representations [47, 102, 197]. Generally, these methods assume that the data from different domains lie in a shared low-rank structure.

Bengio [14] argue that more transferable features can be learned by deep networks since they are able to extract the unknown factors of variation that are intrinsic to the data. Donahue et al. [50] propose the deep convolutional representations named DeCAF, where a deep CNN model is pre-trained using the source dataset (generally large-scale) in a fully supervised fashion. Then they transfer the features (defined by the pre-learned source convolutional network weights) to the target data. The deep auto-encoders are also used for the cross-dataset tasks by exploiting more transferable features by reconstruction [27, 75, 77, 104, 109]. For instance, Ghifary et al. [75] propose a Deep Reconstruction-Classification Network (DRCN) to learn a shared deep CNN model for both classification task of the source samples and reconstruction task of the target samples.

Self Labelling. Recently, Panareda Busto and Gall [171] propose an open set domain adaptation problem, where only some of the classes are shared between the source and target datasets. The task is to label all the target samples either by one of the classes shared between the two domains or as unknown. We discuss this setting under the homogeneous label space problem because the unknown classes are simply detected as unknown rather than recognized as certain classes. They solve this problem by first assigning some of the target data with the labels of the known classes and then reducing the shift between the shared classes in the source and target datasets by a subspace alignment method (similar to [62]). The two procedures are learned iteratively.

Hybrid Approach. Combining different approaches generally trigger better transferring of knowledge. Some methods [96, 285] learn two dictionaries on pairs of correspondence samples and encourage the sparse representation of each sample pair to be similar. Some methods use both geometric and statistical approach [211, 273]. For example, Zhang et al. [273] propose to learn two projections for the source and target domain respectively to reduce the geometrical shift and statistical shift. Differently, Gholami et al. [76] jointly learn a low dimensional subspace and a classifier through a Bayesian learning framework.

Though deep networks can generally learn more transferable features [14, 50], the higher level features computed by the last few layers are usually task-specific and are not transferable to new target tasks [269]. Hence, some recent work imposes statistical approach into the deep learning framework (high-level representation approach) to further reduce domain bias. For instance, the MMD loss is incorporated into the objective of the deep models to reduce the divergence

of marginal distributions [148, 152, 227, 229] (e.g. Figure 4 illustrates the Deep Adaptation Networks (DAN) proposed in [148]) or joint distributions [151] between domains.

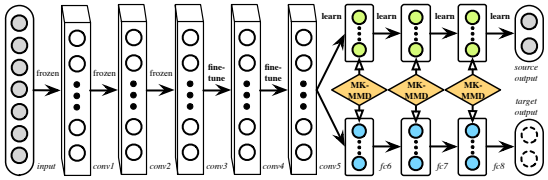

Fig. 4. Illustration of the DAN method (Figure used courtesy of [148])

Instead of using MMD metric, Sun and Saenko [212] extend the CORrelation ALignment (CORAL) method [210] that aligns the covariance of the source and target data to a deep learning-based method. Zellinger et al. [270] propose the Central Moment Discrepancy (CMD) method, which align the higher order central moments of distributions through order-wise moment differences. Instead of statistical approach, the self-labelling is also used in deep neural network-based method. Saito et al. [194] propose an asymmetric tri-training method, where feature extraction layers are used to drive three classifier sub-networks. The first two networks are used to label unlabelled target samples and the third network is to learn the final adapted classifier to operate on the target domain with the pseudo-labels obtained on the first two networks.

The statistical approaches (e.g. MMD distance [18, 243], and $\mathcal{H}$ divergence [18]) are also incorporated into deep autoencoders for learning more transferable features.

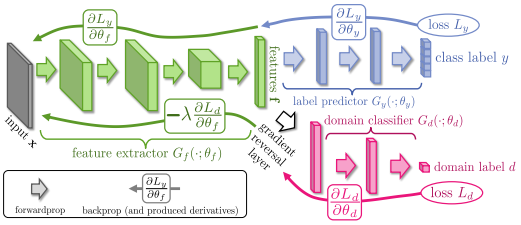

Fig. 5. Illustration of the ReverseGrad method (Figure used courtesy of [70])

Motivated by adversarial learning [84], the GAN-based domain adaptation methods are proposed with the key idea that the JS divergence between domains are reduced [17, 70, 71, 226]. For example, the gradient reversal algorithm (ReverseGrad) proposed by Ganin and Lempitsky [70] minimizes the $\mathcal{H}$-divergence by considering the domain invariance as a binary classification task and employing a gradient reversing strategy (as shown in Figure 5). Tzeng et al. [226] propose to learn separate feature extraction networks for different domains, and a domain classifier is incorporated such that the embeddings produced by the source or target CNN cannot be distinguished. Bousmalis et al. [17] propose a GAN-based method to adapt the source domain data from the pixel level, such that they are not distinguishable to the target domain data. Differently, Liu and Tuzel [145] propose a Coupled GAN (CoGAN) method that learns a joint distribution by jointly modelling two GANs, where the first one generates the source data while the second generates the target images. Instead of enforcing samples from different domains to be non-discriminant, the CoGAN enforce the layers that decode high-level features to share the weights so as to enforce the assumption that the images from different domains share the same high-level representations but have different low-level representations.

3.4 Imbalanced Unlabelled Target Dataset

This problem assumes the target domain is class imbalanced and only with unlabelled data. Thus, the statistical approach can be used. This problem is quite common in practice and known as *prior probability shift*, or *imbalanced data* in classification. For instance, the abnormal activities (e.g. kick, punch, fight, and fall down) are much less frequent than normal activities (e.g. walk, sit, eat, and drink) in the video surveillance but require higher recognition rate.

Statistical Approach. In the classification scenario, the prior probability ($P(Y)$) shift was often considered to be a class imbalance problem [100, 276]. Zhang et al. [276]tackle the prior probability shift by re-weighting the source samples using the similar idea as the Kernel Mean Matching method [97]. They also define the situation where both $P(Y)$ and $P(X|Y)$ are shifted across datasets and propose a kernel approach to reduce the distribution shift by re-weighting and transforming the source data. It is assumed that the source data are able to be transferred to the target domain by location-scale (LS) transformation (i.e. $P(X|Y)$ only differs in the location and scale). Instead of assuming that all the features can be transferred to the target domain by LS transformation, Gong et al. [82] propose to learn the conditional invariant components through a linear transformation, and then the source samples are re-weighted to reduce shift of $P(Y)$ and $P(Y|X)$ between domains.

Recently, Yan et al. [263] take both the domain shift and class weight bias across domains into account. To take the class prior probability into account, they introduce class-specific weights. Specifically, the domain adaptation is performed by iteratively generating the pseudo-labels to the target samples, learning the source class weights, and tuning the deep CNN model parameters.

3.5 Sequential/Online Labelled Target Data

In practice, the target data can be sequential video streams or continuous evolving data. The distribution of the target data may also change with time. Since the target data are labelled, this problem is named *supervised sequential/online* domain adaptation.

Self Labelling. Xu et al. [255] assume a weak-labelling setting and propose an incremental method for object detection across domains. Specifically, the adaptation model is a weighted ensemble of the source and target classifiers and the ensemble weights are updated with time.

3.6 Sequential/Online Unlabelled Target Data

Similar to the problem in 3.5, the target data are sequential in this problem, however, no labelled target data is available, which is named *unsupervised sequential/online domain adaptation* and related to but different from *concept drift*. The concept of drift [67] refers to changes in the conditional distribution ($P(Y|X)$), while the marginal distribution ($P(X)$) stays unchanged, whereas in sequential/online domain adaptation the changes between the two domains are caused by the changes of the input data distribution.

Geometric Approach. Hoffman et al. [92] extend the Subspace Alignment method [62] to handle continuous evolving target domain, as shown in Figure 6. Both the subspaces and subspace metrics that align the two subspaces are updated after each new target sample is received. Bitarafan et al. [15] tackle the continuously evolving target domain using the idea of GFK [81] to construct linear transformation. The linear transformation is updated after a new batch of unlabelled target domain data come. Each batch of arrived target data are classified after the transformation and included in the source domain for recognizing the next batch of data.

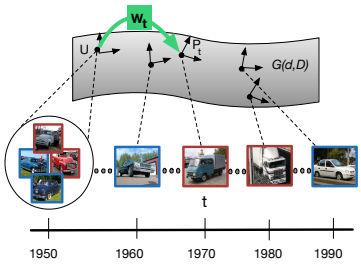

Fig. 6. Illustration of the continuous domain adaptation method [92] (Figure used courtesy of [92])

Self Labelling. Jain and Learned-Miller [99] address the online adaptation in the face detection task by adapting pre-trained classifiers using a Gaussian process regression scheme. The intuition is that the "easy-to-detect" faces can help the detection of "hard-to-detect" faces by normalizing the co-occurring "hard-to-detect" faces and thus reducing their difficulty of detection. Xu et al. [256] propose an online domain adaptation model for multiple object tracking using a two-level hierarchical tree framework, where the leaf nodes correspond to the object detectors while the root node corresponds to the class detector. The adaptation is executed in a progressive manner.

3.7 Unavailable Target Data

This problem is also named *domain generalization* in literature, where the target domain data are not presented for adaptation. Thus, multiple source datasets are generally required to learn the dataset invariant knowledge that can be generalized to a new dataset. Note that domain generalization is distinguished from multi-source domain adaptation (MSDA)[53, 55, 79, 93, 214, 257] since MSDA generally requires the access to the target data for adaptation. We will discuss transfer learning from multiple sources in details in Section 8.3.

Higher-level Representation. Most of the existing work tackle this problem by learning domain invariant and compact representation from multiple source domains [16, 58, 73, 74, 111, 132, 162, 163, 207]. For example, Khosla et al. [111] explicitly model the bias of each source domain and try to estimate the weights for the unbiased data by removing the source domain biases. Muandet et al. [163] propose the Domain-Invariant Component Analysis (DICA), a kernel-based method, to learn an invariant mapping that reduces the domain shift and preserve discriminative information at the same time. Fang et al. [58] propose an unbiased metric learning approach to learn unbiased metric from multiple biased datasets. Ghifary et al. [74] propose a Multi-Task Autoencoder (MTAE) method. It substitutes artificially induced corruption in standard denoising autoencoder with some specific variations of the objects (e.g. rotation) to form multiple views. Hence, MTAE learns representations that are invariant to multiple related domains.

Ensembling classifiers learned from multiple sources is also used for generalizing to unseen target domain [138, 166, 167, 260]. Xu et al. [260] propose to reduce the domain shift in an exemplar-SVMs framework by regularizing positive samples from the same latent domain to have similar likelihoods from each exemplar classifier. Similarly, Niu et al. [166] extend this idea to the source domain samples with multi-view features. Niu et al. [167] explicitly discover the multiple hidden domains [79], and then an ensemble of classifiers is formed by learning a single classifier for each individual category in each discovered hidden domain.

4 HETEROGENEOUS FEATURE SPACES

This section discusses the problems that $\mathcal{S}_S$ and $\mathcal{S}_T$ are different due to $\mathcal{X}_S \neq \mathcal{X}_T$, but $\mathcal{Y}_S = \mathcal{Y}_T$. The different feature spaces can be generated from different data modalities or different feature extraction methods. Similar to the scenario defined in Section 3, sufficient labelled source domain data are assumed to be available in the following sub-problems.

4.1 Labelled Target Dataset

This problem assumes limited labelled target data are presented for adaptation. This problem is named *supervised heterogeneous domain adaptation.*

Higher-level Representation. Some methods assume that only the feature spaces are different while the distributions are the same between source and target datasets. Since the labelled data in the target dataset are scarce, Zhu et al. [291] propose to use the auxiliary heterogeneous data that contain both modalities from Web to extract the semantic concept and find the shared latent semantic feature space between different modalities.

Class-based Approach. The class-based approach has also been used to connect heterogeneous feature spaces. Finding the relationship between different feature spaces can be seen as translating between different languages. Hence, Dai et al. [41] propose a translator using a language model to translate between different data modalities or feature spaces by borrowing the class label information. Kan et al. [110] propose a multi-view discriminant analysis method that learns view-specific linear mappings for each view to find a view-invariant space by using label information: $(w_1^*, w_2^*, ..., w_v^*) = \arg\max_{w_1,...,w_v} \frac{Tr(S_B^y)}{Tr(S_W^y)}$,where the between-class variation S_B^y from all views are maximized while the within-class variation S_W^y from all views are minimized, $w_1^*, w_2^*, ..., w_v^*$ are the optimized transformations for different views. Manifold alignment method [233] is also used for heterogeneous domain adaptation with the class-based approach.

Inspired by [44], the feature augmentation based method has also been proposed [56, 135] for heterogeneous domain adaptation, which transforms the data from two domains into a shared subspace, and then two transformations are proposed such that the transformed features in the subspace are augmented with the original data as well as zeros (as shown in Figure 7).

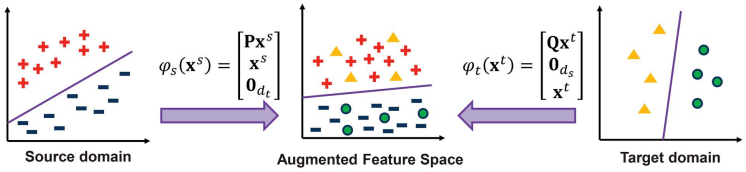

Fig. 7. Illustration of a feature augmentation method for heterogeneous domain adaptation. (Figure used courtesy of [135])

Kulis et al. [118] extend [193] to learn an asymmetric mapping that transforms samples between domains using labelled data from both domains, with the similar assumption as [193] that the label spaces of target training set and target test set are non-overlapping subsets of source label space. Different from previous metric learning based domain adaptation that learns the asymmetric feature transformation between heterogeneous features [118], the asymmetric metric of classifiers can also be learned to bridge source and target classifiers on heterogeneous features [287].

Hybrid Approach. The first group of work focuses on cross-modal representation learning by combing class-based and higher level representation approaches. Gong et al. [83] propose a three-view Canonical Correlation Analysis (CCA) model that explicitly incorporates the high-level semantic information (i.e. high-level labels or topics) as a third view. A recent work[232] incorporates the adversarial learning to the supervised representation learning for cross-modal retrieval.

Another line of research assumes that both the feature spaces and the data distributions are different. Shekhar et al. [199] extend [198] to heterogeneous feature spaces, where the two projections and a latent dictionary are jointly learned to simultaneously find a common discriminative low-dimensional space and reduce the distribution shift. Similarly, Sukhija et al. [209] assume the label distributions between domains are shared. Then the shared label distributions are used as pivots to derive a sparse projection between the two domains.

4.2 Labelled plus Unlabelled Target Dataset

In this problem, both limited labelled and sufficient unlabelled target data are presented, which is named *semi-supervised heterogeneous domain adaptation*.

Statistical Approach. Tsai et al. [224] propose the Cross-Domain Landmark Selection (CDLS) method for heterogeneous domain adaptation (HDA) using the statistical approach (MMD). Specifically, the CDLS method derives a heterogeneous feature transformation which results in a domain-invariant subspace for associating the heterogeneous domains. and assigns the weight to each instance according to their adaptation ability using both labelled and unlabelled target samples.

Correspondence Approach. Zhai et al. [271] assume in addition to a set of labelled correspondence pairs between the source and target datasets, some unlabelled data from both datasets are also available. Specifically, given a set of correspondence samples C between the two domains, one can learn the mapping matrices A_s and A_t for the source and target sets respectively in order to preserve the correspondence relationships after mapping:

$$< A_s, A_t >= \arg\min_{A_s, A_t} \sum_{(i,j)\in C} \|A_s^T x_i^s - A_t^T x_j^t\|^2 + J(A_s, X_s) + J(A_t, X_t) \tag{5}$$

where x_i^s and x_j^t represent the ith source domain sample and the jth target domain sample, respectively, $J(A_s, X_s)$ and $J(A_t, X_t)$ are the manifold regularization terms which are used to preserve the intrinsic manifold structures of the source and target domains.

Class-based Approach. Xiao and Guo [251] propose a kernel matching method, where a kernel matrix of the target domain is matched to a source domain sub-matrix by exploiting the label information such that the target samples are mapped to similar source samples. The unlabelled target samples are expected to be aligned with the source samples from the same class with the guides of labelled target samples via the function of kernel affinity measures between samples.

Hybrid Approach. Wu and Ji [247] introduce a constrained deep transfer feature learning method by incorporating the correspondence into the high-level representation approach. Specifically, several pairs of source and target samples are used to capture the joint distribution and bridge the two domains. Then a large amount of additional source samples are transferred to the target domain through pseudo labelling for further target domain feature learning.

4.3 Unlabelled Target Dataset

This problem assumes no labelled target domain data is available. We name this problem as *unsupervised heterogeneous domain adaptation*. In this problem, the feature spaces could be completely different between datasets. It can also be assumed that the source data consist of multiple modalities while the target data only contain one of the modalities, or vice versa.

Statistical Approach. Chen et al. [25] and Li et al. [134] assume the source datasets contain multiple modalities and target dataset only contains one modality and the distribution shift between datasets also exists. Specifically, the statistical approach (e.g. MMD) is used such that the source and target common modalities are projected to a shared subspace to reduce the distribution mismatch. In the meantime, the multiple source modalities are also transformed to the same representation in the shared space. They iteratively refine the shared space and the robust classifier.

Correspondence Approach. The co-occurrence data between different feature spaces or modalities have been employed for heterogeneous domain adaptation [180, 265].

Hybrid Approach. The correspondence approach or statistical approach are generally incorporated into higher-level representation approach for transferring between data modalities or feature spaces.

Canonical Correlation Analysis (CCA)[5] is a standard approach to learning two linear projections of two sets of data that are maximally correlated. Neither supervised data nor the paired data are required. Many cross-modal recognition or retrieval methods incorporate the idea of CCA[6, 61, 262] into deep models. Cross-media multiple deep networks (CMDN)[176] jointly preserve the intra-media and inter-media information and then hierarchically combine them for learning the rich cross-media correlation. CastrejÃşn et al. [23] introduce a cross-modal representation method across RGB modality, sketch modality, clipart, and textual descriptions of indoor scenes. The cross-modal convolutional neural networks are regularized using statistical regularization so that they have a shared representation that is invariant to different modalities.

The paired correspondence data are used in [89], where a cross-modal supervision transfer method is proposed. The deep CNNs are pre-trained on the source data (e.g. a large-scale labelled RGB dataset). Then the paired target data (unlabelled RGB and depth image pairs) are used for transferring the source parameters to the target networks by constraining the paired samples from different modalities to have the similar representations.

A line of research focuses on the task of translation between different domains. For example, in machine translation between languages, the sentence pairs are presented in the form of a parallel training corpus for learning the translation system. Traditional translation system [115] is generally phrase-based, whose sub-components are usually learned separately. Differently, a newly emerging approach, named Neural machine translation [8, 31, 108, 215], constructs and trains a neural network that inputs a sentence and outputs the translated sentence.

Similarly, in the computer vision domain, image-to-image translation [98] has also been extensively exploited, which aims at converting an image from one representation of a given scene to another (e.g. texture synthesis [130], sketch to photograph [98], RGB to depth [89], time hallucination [98, 122, 201], image to semantic labels [57, 146, 252], stimulated to real image [203], style transfer [72, 107, 130, 239, 278], and general image-to-image translation [13, 98, 112, 129, 144, 145, 268, 289]). The key idea for tackling these tasks is to learn a translation model between paired (correspondence approach) or unpaired samples (statistical approach) from different domains. The recent deep learning based techniques have greatly advanced the image-to-image translation task. For example, the deep convolutional neural networks based methods [57, 72, 107, 146, 252, 278], and the Generative Adversarial Networks (GANs [84]) based methods [13, 98, 112, 129, 130, 144, 145, 203, 239, 268, 289] have been exploited for learning the translation model. Though the original purposes of some of these work on translation between domains may not be cross-dataset recognition, the ideas can be borrowed for cross-modality or cross feature spaces recognition. If a proper translation between domains can be obtained, the target task can be boosted by the translated source domain data.

5 HETEROGENEOUS LABEL SPACES

This section discusses the problems that $\mathcal{X}_S = \mathcal{X}_T$ and $\mathcal{Y}_S \neq \mathcal{Y}_T$. For example, in the classification tasks, when the label spaces between datasets are different, there still exists shared knowledge between previous categories (e.g. horse) and new categories (e.g. zebra) that can be used for learning new categories. The source domain is assumed to be labelled except for the last sub-problem (Section 5.5).

5.1 Labelled Target Dataset

This setting is commonly used in deep learning context. In practice, the deep networks are rarely trained from scratch (with random initialization), since the target datasets rarely have sufficient labelled data. Thus, transfer learning is

generally used. The pre-trained deep models from a very large source dataset are used either as an initialization (then fine-tune the model according to the target data) or as a fixed feature extractor for the target, which is generally different from the original task (i.e. different label spaces).

The fine-tuning procedure is similar to *one-shot learning* or *few-shot learning*. The key difference is that the available target data are sufficient for the target task in fine-tuning but in few-shot learning, the target data are generally rare (e.g. only one sample per class in the extreme case). The few-shot learning also has close connection with multi-task learning. The difference is that one-shot learning emphasizes on the recognition of the target data with limited labelled data while the objective of multi-task learning is to improve all the tasks with good training data in each task.

Higher-level Representation Approach. Since the training of deep learning models requires a large scale dataset to avoid overfitting, the transfer learning techniques [269] can be used for small scale target datasets. The most commonly used transfer learning technique is to initialize the weights from a pre-trained model and then the target training data are used to fine-tune the parameters for the target task. When the pre-trained source model is used as the initialization, two strategies can be employed. First is to fine-tune all the layers of the deep neural network, while the second strategy is to freeze several earlier layers and only fine-tune the later layers to reduce the effects of overfitting. This is inspired by the observation that the features extracted from the early layers show more general features (e.g. edge or color) that are transferable to different tasks. However, the later layers are gradually more specific to the details of the original source tasks. Other transfer methods [50, 189] directly use the pre-trained deep convolutional nets (normally after removing the last one or two fully connected layers) on a large dataset (e.g. ImageNet [45]) as a fixed feature extractor for the target data.

Note that when the pre-trained deep models are used as an initialization or a fixed feature extractor in the deep learning frameworks, only the pre-trained weights need to be stored without the need of storing the original large scale source data, which is appealing.

Class-based Approach. Patricia and Caputo [174] treat the pre-trained models from multi-source domains as experts to augment the target features. The output confidence values of prior models are treated as features and the features from the target samples are augmented with these confidence values to build a target classifier. Several classifier-based methods are proposed to transfer the parameters of classifiers using generative models [60, 123], or discriminative models [7, 106, 156, 220]. The key idea is using source models as prior knowledge to regularize the models of the target task. These methods are also called the Hypothesis Transfer Learning (HTL) since it assumes no explicit access to the source domain data and only uses source models learned from a source domain. The HTL has been theoretically analysed [51, 121, 241]

Hybrid Approach. Recently, the deep learning based approaches have been proposed for few-shot learning, most of which are metric learning based methods. One early neural network approach to one-shot learning was provided by Siamese networks [113], which employs a structure to rank similarity between inputs. Vinyals et al. [230] propose the matching networks, where a differentiable neural attention mechanism is used over a learned embedding of the limited labelled target data. This method can be considered as a weighted nearest-neighbour classifier in an embedded space. Snell et al. [205] transform the input into an embedding space by proposing a prototypical network and the prototype from each class is taken as the mean of the embedded support set. Differently, Ravi and Larochelle [188] propose a meta-learning-based few-shot learning method, where a meta-learner LSTM [91] model is used to produce updates for training the few-shot neural network classifier. Given a few target labelled examples, this approach can generalize well on the target set.

5.2 Unlabelled Target Dataset

Some researches also try to tackle the heterogeneous label space problem by assuming that only unlabelled target data are presented. This problem can be named as *unsupervised transfer learning*.

Higher-level Representation. The higher-level representation approach is generally used for this problem. Two different scenarios are considered in literature.

The first scenario assumes that only the label spaces between datasets are disjoint while the distribution shift is not considered. Since no labelled target data are available, the unseen class information is generally gained from a higher level semantic space shared between datasets. For example, some research assumes that the human-specified high-level semantic space (e.g. attributes [168], or text descriptions [190]) shared between datasets are available. Given a defined attribute or text description ontology, a vector in the semantic space can be used for representing each class. However, it is expensive to acquire the attribute annotations or text descriptions. Hence, to avoid human involved annotations, another strategy learns the semantic space by borrowing the large and unrestricted, but freely available, text corpora (e.g. Wikipedia) to derive a word vector space [64, 158, 206]. The related work on semantic space (e.g. attributes, text descriptions, or word vector) will be further discussed in Section 5.4, since the target data are generally not required when the semantic space is involved.

The second scenario assumes that apart from the different label spaces, the domain shift (i.e. the distribution shift of features) also exists between datasets [66, 114, 139, 234, 259, 267, 282]. This is named the *projection domain shift* problem by [66]. For example, as illustrated in Figure 8, both zebra and pig have the same attribute 'hasTail', but the visual appearances and the distributions of the tails of zebra and pig are very different. To reduce the domain shift explicitly, the training data (unlabelled) in the target domain are generally required to be available. For example, Fu et al. [66] introduce a multi-view embedding space in a transductive setting, such that different semantic views are aligned. Kodirov et al. [114] propose a regularised sparse representation framework that utilizes the target class prototypes estimated from target images to regularise the projections of the target data and thus overcomes the projection domain shift problem.

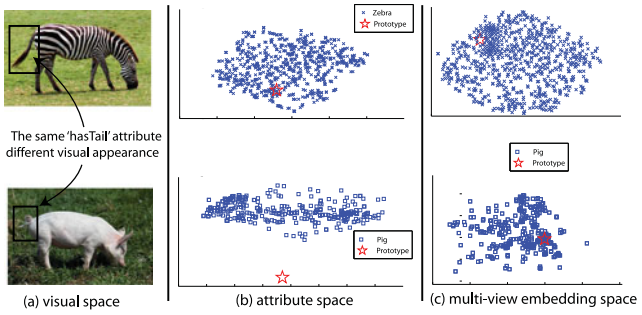

Fig. 8. Examples of projection domain shift.(Figure used courtesy of [66])

5.3 Sequential/Online Labelled Target Data

This problem assumes the target data are sequential and can be from different classes, which is also called *sequential/online transfer learning*, and closely related to *lifelong learning* [141, 192, 218]. Both concepts focus on the continuous learning processes for evolving tasks. However, sequential/online transfer learning emphasizes on how to improve the target domain performance (without sufficient target training data), but lifelong learning tries to improve the future target

task (with sufficient target training data) as well as all the past tasks [28]. Also, the lifelong learning can be seen as incremental/online multi-task learning.

Self Labelling. Nater et al. [164] address an action recognition scenario where the unseen activities to be recognized only have one labelled sample per new activity. They build a multi-class model which uses the prior knowledge of seen classes and progressively learns the new classes. Then the newly labelled activities are integrated into the previous model to update the activity model. Zhao and Hoi [284] propose an ensemble learning based online transfer learning method (OTL) that learns a classifier in an online fashion using the target data, and combines it with the pre-learned source classifier. The combination weights are tuned dynamically based on the loss between the ground-truth label of the incoming sample and the current prediction. Tommasi et al. [221] then extended OTL [284] and addressed the case of online transfer learning from multiple sources.

5.4 Unavailable Target Data

This problem is also named *zero-shot learning* in literature, where unseen target categories are to be recognized without having access to the target data. Different from *domain generalization* (see Section 3.7), the categories of unseen target data are different from the source categories in *zero-shot learning*. As mentioned in Section 5.2, the unseen categories can be generally connected via some auxiliary information, such as a common semantic space.

Higher-level Representation. Most of the methods for this problem rely on the existence of a labelled source dataset of seen categories and the prior knowledge about the semantic relationship between the unseen and seen categories. In general, the seen and unseen categories are correlated in a high-level semantic space. Such a semantic space can be an attribute space [168], text description space [190], or a word vector space [64, 158, 206]. Since multiple semantic spaces are often complementary to each other, some methods are proposed to fuse multiple semantic spaces [3, 277].

The attribute space is the most commonly used intermediate semantic space. The attributes are defined as properties observable in images, which are described with human-designated names such as “white”, “hairy”, “four-legged”. Hence, in addition to label annotation, the attribute annotations are required for each class. However, the attributes are annotated per-class rather than per-image. Thus, the effort to annotate a new category is small. Two main strategies are proposed for recognizing unseen object categories using attributes. The first is recognition using independent attributes, consists of learning an independent classifier per attribute [120, 124, 143, 168, 172]. At test time, the attribute values for test data are predicted using the independent classifiers and the labels are then inferred. Since attribute detectors are expected to generalize well on both seen and unseen categories, some research is devoted to discovering discriminant attributes [24, 182, 187], or modelling the uncertainty of attributes [101, 240], or robustly detecting attributes from images [19, 68]. However, Akata et al. [2] argue that the attribute classifiers in previous works are learned independently of the end-task, and thus they may be able to predict the attributes from new images but may not be able to effectively infer the classes. Hence, the second strategy is recognition by assuming a fixed transformation (W) between the attributes and the class labels [1, 3, 140, 181, 191, 248, 280, 281] to learn all attributes simultaneously: $F(x, y; W) = \theta(x)^T W \phi(y)$, where $\theta(x)$ and $\phi(y)$ represent image and class embeddings, both are given. To sum up, the attribute-based zero-shot learning methods are promising for recognizing unseen classes, while with a key drawback that the attribute annotations are still required for each class. Instead of using attributes, the second semantic space is image text descriptions [190], which provides a natural language interface. However, similar to attribute space, the expensive manual annotation is required for obtaining the good performance. The third semantic space is the word vector space [64, 128, 158, 206], which is derived from a huge text corpus and generally learned by a deep neural network. The word vector space is attractive since extensive annotations are not required for obtaining the semantic space.

5.5 Unlabelled Source Dataset

This problem assumes that the source data are unlabelled but the contained information (e.g. basic visual patterns) can be used for target tasks, which is known as *self-taught learning*.

Higher-level Representation. Raina et al. [185] firstly presented the idea of "self-taught learning". They learn the sparse coding from the source data to extract higher-level features. Some variations of Raina et al. [185]'s method are proposed either by generalizing the Gaussian sparse coding to exponential family sparse coding [126] , or by taking the supervision information contained in labelled images into consideration [235]. Moreover, Kumagai [119] provide a theoretical analysis for self-taught learning with the focus on discussing the learning bound of sparsity-based methods.

The idea of self-taught learning has also been used in deep learning framework, where the unlabelled data are used for pre-training the network to obtain good starting point of parameters [69, 117, 125]. For instance, Gan et al. [69] use the unlabelled samples to pre-train the first layer of Convolutional deep belief network (CDBN) for initializing the network parameters. Kuen et al. [117] extract the domain-invariant features from unlabelled source image patches for the tracking tasks using stacked convolutional autoencoders.

6 HETEROGENEOUS FEATURE SPACES AND LABEL SPACES

In this section, a more challenging scenario is discussed, where $\mathcal{X}_S \neq \mathcal{X}_T$ and $\mathcal{Y}_S \neq \mathcal{Y}_T$. There is little work regarding this scenario due to the challenges and the common assumption that sufficient source domain labelled data is available.

6.1 Labelled Target Dataset

This problem assumes the labelled target data are available. We name this problem as *heterogeneous supervised transfer learning*.

Higher-level Representation. Rather than assuming completely different feature spaces, most methods in this setting assume that the source domain contains data with multi-modality but the target domain only has one of the source domain modalities. Ding et al. [48] propose to uncover the missing target modality by finding similar data from the source domain, where a latent factor is incorporated to uncover the missing modality based on the low-rank criterion (as illustrated in Figure 9). Similarly, Jia et al. [103] propose to transfer the knowledge of RGB-D (RGB and depth) data to the dataset that only has RGB data. They applied the latent low-rank tensor method to discover the common subspace of the two datasets.

Hybrid Approach. Hu and Yang [95] assume the feature spaces, the label spaces, as well as the underlying distributions are all different between source and target datasets and propose to transfer the knowledge between different activity recognition tasks by learning a mapping between different sensors. They adopt the similar idea of translated learning [41] to find a translator between different feature spaces using statistical approach (e.g. JS divergence). Then the Web knowledge is used to link the different label spaces using self-labelling.

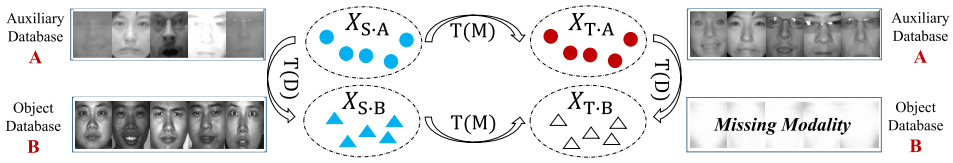

Fig. 9. Example of multiple source modalities and one target modality.(Figure used courtesy of [48])

6.2 Sequential/Online Labelled Target Data

This problem assumes the sequential/online target data have different feature space with source data, which is named as *heterogeneous sequential/online transfer learning*.

Self Labelling. As mentioned in Section 5.3, Zhao and Hoi [284] propose the OTL method for online transfer learning. They also consider the case of heterogeneous feature spaces by assuming the source domain feature space to be a subspace of the target domain feature space. Then a multi-view approach is proposed by adopting a co-regularization principle of online learning of two target classifiers simultaneously from the two views (the source domain feature space and the new space). The unseen target example is classified by the combination of the two target classifiers.

7 DATASETS

Table 1 lists the commonly used visual datasets for transfer learning. They are categorised into object recognition, Hand-Written digit recognition, face recognition, person re-identification, scene categorization, action recognition and video event detection. In the table, the ✓indicates the dataset has been evaluated on the corresponding problem while the # indicates the datasets that have the potential to be used in the evaluation of the algorithms for the problem though reported results are not publicly available to our knowledge. Due to the page limit, readers are referred to the supplementary material and the references for more detailed information of the datasets.

8 CHALLENGES AND FUTURE DIRECTIONS

Transfer learning is a promising and important approach to cross-dataset visual recognition and has been extensively studied in the past decades with much success. Figure 1 shows the problem-oriented taxonomy and the statistics on the number of papers for each problem has showed that most previous works concentrate on a subset of problems presented in Figure 1. Specifically, only nine out of the seventeen problems are relatively well studied where the source and target domains share at least either their feature spaces or label spaces, the source domain data are labelled and balanced, target domain data are balanced and non-sequential. The rest eight problems especially those where the target data is imbalanced and sequential are much less explored. Such a landscape together with the recent fast-advancing deep learning approach has revealed many challenges and opened many future opportunities as elaborated below for cross-dataset visual recognition.

8.1 Deep Transfer Learning

As deep learning advances, transfer learning is also shifted from traditional shallow-learning based approaches to deep neural network based approaches. In practice, the deep networks for the target task are rarely trained from scratch (i.e. with random initialization), since the target datasets rarely have sufficient samples. Thus, transfer learning is generally used. The pre-trained deep models from a very large source dataset are used either as an initialization [269] (then fine-tune the model according to the target data) or a fixed feature extractor for the target task of interest [50, 189].

Similarly, in deep domain adaptation, the deep models are either used as feature extractors (then shallow-based domain adaptation methods are used for further adaptation) [36, 73, 116, 224, 266, 273], or used in an end-to-end fashion (i.e. the domain adaptation module is integrated into the deep model) [17, 71, 148, 151, 152, 225, 226]. It is still unclear which approach would perform better. The advantage of using deep models as feature extractors is that the computational cost is much lower since shallow-based DA methods are generally much faster than deep learning-based methods. Another advantage is that many shallow-based methods have a global optimum value. The drawback is that the degree of adaptation may be insufficient in the shallow-based methods to fully leverage the deeply extracted features. On the other hand, the advantage of integrating an adaptation module into deep models is two-fold. First, it is

Table 1. Suitability of the widely used datasets where the ✓indicates the dataset has been used the corresponding problems while the # indicates the datasets can be potentially used for the problem. Problem notations: P3.1, supervised domain adaptation (DA); P3.2, Semi-supervised DA; P3.3, Unsupervised DA; P3.4, Supervised online DA; P3.5, Supervised online DA; P3.6, Unsupervised online DA; P3.7, Domain generalization; P4.1, Supervised Heterogeneous DA; P4.2, Semisupervised Heterogeneous DA; P4.3, Unsupervised Heterogeneous DA; P5.1, Few-shot Learning; P5.2, Unsupervised transfer learning (TL); P5.3, Online TL; P5.4, Zero-shot Learning; P5.5, Self-taught Learning; P6.1, Heterogeneous TL; P6.2, Heterogeneous online TL.

	Datasets	P3.1	P3.2	P3.3	P3.4	P3.5	P3.6	P3.7	P4.1	P4.2	P4.3	P5.1	P5.2	P5.3	P5.4	P5.5	P6.1	P6.2
Object	Office[193]	✓	✓	✓	✓			✓	✓	✓	✓							
	Office+Caltech[81]	✓	✓	✓	✓			✓	#	#	✓							
	Cross-dataset testbed[222]	✓	✓	✓	#			#	#	#	#							
	Office-Home[229]	✓	✓	✓				#										
	VLCS[111]	✓	#	✓				✓										
	ImageCLEF-DA[22]	#	#	✓	✓			#										
	PACS[132]	#	#	#				✓										
	CIFAR-10 v.s. STL-10[63]	#	#	✓														
	RGB-D→Caltech256[25]	✓	#	#					#	#	✓							
	Syn Signs v.s. GTSRB[70]	#	#	✓														
	NUS-WIDE[33]								✓	#	#							
	Wikipedia dataset[177]								✓	#	#							
	Pascal Sentence[186]								✓	#	#							
	MSCOCO[142]								✓	#	#							
	aP&Y[59]											✓	✓		✓			
	AwA[124]											✓	✓		✓			
	Caltech-UCSD CUB[231]											✓	✓		✓			
	Caltech-256[221]													✓				
	Car over time[92]						✓											
	STL-10 dataset[34]															✓		
	LabelMe → NUS-WIDE[235]															✓		
	Outdoor scene v.s. Caltech101[185]															✓		
Digit&Character	MNIST v.s. MNIST-M[70]	✓	#	✓														
	MNIST v.s. SVHN[70]	✓	#	✓														
	USPS v.s. SVHN[70]	✓	#	✓														
	SYN DIGITS v.s. SVHN[70]	✓	#	✓														
	Omniglot[123]											✓						
	Digits v.s. English characters[185]															✓		
	English characters v.s. Font characters[185]															✓		
Face	CMU Multi-PIE[88]	✓	✓	✓														
	CMU-PIE v.s. Yale B[48]																✓	
	Oulu-CASIA NIR&VIS v.s. BUAA-VisNir[48]																✓	
	CUHK Face Sketch[242]								✓	#	✓							
	CASIA NIR-VIS 2.0[133]								✓	#	#							
	ePRIP VIS-Sketch[159]								✓	#	#							
Person	VIPeR[87]			✓														
	CUHK02[137]			✓														
	PRID[90]			✓														
	ILIDS[286]			✓														
	CAVIAR[29]			✓														
	3DPeS[11]			✓														
Scene	CMPlaces[23]								✓	#	#							
	SUN Attribute[175]											✓	✓		✓			
	Scene over time[92]						✓											
	NYUD2[89]										✓							
Action	UCF YouTube v.s. HMDB51[288]											✓						
	KTH v.s. UCF YouTube[156]											✓						
	KTH v.s. CareMedia[156]											✓						
	KTH → MSR Action[20]						✓											
	HumanEva v.s. KSA[156]											✓						
	A combination of KTH, Weizmann, UIUC[143]														✓			
	Multiview IXMAS dataset [244]			✓				✓										
	N-UCLA Multiview Action3D[236]			✓														
	ACT4^2 dataset [30, 166]			✓				✓										
	MSR pair action 3D → MSR daily[103]																✓	
	Transferring Activities[164]												✓					
Event	TRECVID 2005[264]	✓																
	TRECVID 2010&2011[155]	✓	✓															
	TRECVID MED 13[258]													✓				
	ImageNet→TRECVID 2011[217]								#	#	✓							
	ImageNet→LabelMe Video[217]								#	#	✓							

end-to-end trainable. Secondly, the adaptation can be performed in multiple levels of features. While the drawbacks are the computational cost and the local optimum. To date, these two approaches have produced similar performance on some datasets [116, 152, 162, 274] though the end-to-end deep systems involve more parameters and require more computational costs. One of the missing study in the literature is a systematic study and comparison of the two approaches under same or similar conditions. For instance, both deep and shallow-based methods can use MMD metric between distributions as a constraint to the objective function. Thus, the comparison between the two approaches using MMD metric may be conducted.

The adversarial nets derived from GANs [84] are appealing in deep learning-based transfer methods. The adversarial loss measures the JS divergence between two sets of data. In practice, the adversarial loss achieves better results and requires smaller batch sizes compared to the MMD loss [71, 131]. Currently, the adversarial nets-based transfer methods have been used on many transfer learning tasks, such as domain adaptation [17, 70, 71, 145, 226], partial domain adaptation [21, 272], cross-modal transfer [13, 98, 112, 129, 130, 144, 145, 203, 239, 268, 289], and zero-shot learning [249, 292]. However, some of the drawbacks of GANs may also remain in adversarial nets-based transfer methods, such as unclear stopping criteria and hard training.

8.2 Partial Domain Adaptation

Partial domain adaptation aims at adapting from a source dataset to an unlabelled target dataset whose label space is known to be a subspace of that of the source [21, 94, 272] or in a more general and challenging setting where only a subset of the label spaces between the source and target is overlapping [171]. The former may be considered to be a special case of transfer learning between heterogeneous label spaces and a typical and practical example is to transfer from a large source dataset with more classes to a small target dataset with less classes. The latter is a problem bearing both domain adaptation and zero shot learning. Generally, the distribution shift is caused not only by label space difference but also by the intrinsic divergence of distributions (i.e. the distribution shifts exist even on shared classes between source and target). Partial domain adaptation has a more realistic setting than conventional unsupervised domain adaptation. Solutions to this problem would expand the applications of domain adaptation and provide a basic mechanism for online transfer learning and adaptation. However, few papers have been found on partial domain adaptation.

8.3 Transfer Learning from Multiple Sources

The multi-source domain adaptation (MSDA) [53, 55, 79, 93, 214, 257] refers to adaptation from multiple source domains that have exactly the same label space as the target domain. Intuitively, the MSDA methods should be able to obtain superior performance compared to the single source setting. However, in practice, the adaptation from multiple sources generally can only give similar or even worse results compared to transferring from one of the source domains (though not every one of them) [102, 198]. This is probably due to the negative transfer issue. In addition, most source data contains multiple unknown latent domains [79, 93] in the real-world applications. Thus, how to discover latent domains and how to measure the domain similarities are still fundamental issues.

A more realistic setting is incomplete multi-source domain adaptation (IMSDA) [49, 257] here each source label space is only a subset in the target domain and the union of the multiple source label spaces covers the target label space. IMSDA is a more challenging problem compared with MSDA, since the distribution shifts among the sources as well as the target domain are harder to be reduced due to the incompleteness of each source domain. In addition, when the number of sources increases, this problem will become challenging.

Multiple sources can be generalised to a target task, referred to as domain generalization [16, 58, 73, 74, 111, 132, 162, 163, 207] without the need of any target data. Domain generalization is of practical significance, but less addressed in the previous research. Since there is no target data available, domain generalization often has to learn semantically meaningful model shared across different domains.

8.4 Sequential/Online Transfer Learning

In sequential/online transfer learning [284], source data may not be fully available when the adaptation or transfer learning is being performed and/or the target data may also arrive sequentially. In addition, the source or even the

target data cannot be fully stored and revisited in the future learning process. The adapted model is often required to perform well not only on the new target data but also to maintain its performance on the source data or previously seen data. Such a setting is sometimes known as incremental learning or transfer learning without forgetting under certain assumptions [127, 141, 202]. Few studies on this problem have been reported as shown in Figure 1.

8.5 Data Imbalance

The issue of data imbalance in the target dataset has been much neglected in the previous research, while imbalanced source data may be converted to balanced ones by discarding or re-weighting the training (source) data during the learning procedure. However, the target data can hardly follow such a process especially when the target data is insufficient. Data imbalance can be another source of distribution divergence between datasets and is ubiquitous in real-world applications. So far, there has been little study on how the existing algorithms for cross-dataset recognition would perform on imbalanced target data or how the imbalance would affect the algorithm performance.

8.6 Few-shot and Zero-shot Learning

Few-shot learning and Zero-shot learning are interesting and practical sub-problems in transfer learning which aim to transfer the source models efficiently to the target task with only a few (few-shot) or even no target data (zero-shot). In few-shot learning, the target data are generally rare (i.e. only one training sample is available for each class in the extreme case). Thus, the standard supervised learning framework could not provide an effective solution for learning new classes from only few samples [60, 123]. This challenge becomes more obvious in the deep learning context, since it generally relies on larger datasets and suffers from overfitting with insufficient data [205, 230].

Compared to few-shot learning, zero-shot learning does not require any target data. A key challenge in zero-shot learning is the issue of *projection domain shift* [65], which is neglected by most previous work. Since the source and target categories are disjoint, the projection obtained from the source categories is biased if they are applied to the target categories directly. For example, both zebra (one of the source class) and pig (one of the target class) have the same attribute 'hasTail', but the visual appearances of the tails of zebra and pig are very different (as shown in Figure 8). However, to deal with the projection domain shift problem, the unlabelled target data are generally required. Thus, further exploration of new solutions to reduce the projection domain shift is useful for effective zero-shot learning. Another future direction is the exploration of more high-level semantic spaces for connecting seen and unseen classes. The most frequently used high-level semantics are manually annotated attributes or text descriptions. Some recent work [64, 128, 158, 206] employs the word vector as semantic space without relying on human annotation, but the performance of zero-shot learning using word vector is generally poorer than that using manually labelled attributes.

A recent work [250] presents a comprehensive analysis of the recent advances in zero-shot learning. They critically compare and analyse the state-of-the-art methods and unifies the data splits of training and test sets as well as the evaluation protocols for zero-shot learning. Their evaluation protocol emphasizes on the *generalized zero-shot learning*, which is considered more realistic and challenging. The traditional zero-shot learning generally assumes that the training categories do not appear at test time. By contrast, the generalized zero-shot setting relaxes this assumption and generalizes to the case where both seen and unseen categories are presented in the test stage, which provides standard evaluation protocols and data splits for fair comparison and realistic evaluation in the future.

8.7 Cross-modal Recognition

The cross-modal transfer, a sub-problem of heterogeneous domain adaptation and heterogeneous transfer learning as shown in Figure 1, refers to transfer between different data modalities (e.g. text v.s. image, image v.s. video, RGB

v.s. Depth, etc.). Compared to cross-modal retrieval [232] and translation [98], fewer works are dedicated to cross-modal recognition through adaptation or transfer learning. The recognition across data modalities is ubiquitous in the real-world applications. For instance, the depth images acquired by the newly released depth cameras are much rarer compared to RGB images. Effectively using rich and massive labelled RGB images to help the recognition of depth images can reduce the extensive efforts of data collection and annotation. Some preliminary works can be found in [89, 103, 129, 134, 238].

8.8 Transfer Learning from Weakly Labelled Web Data

The data on the Internet are generally weakly labelled. Textual information (e.g., caption, user tags, or description) can be easily obtained from the web as additional meta information for visual data. Thus, effectively adapting the visual representations learned from the weakly labelled data (e.g. web data) or co-existent other modality data to new tasks is interesting and practically important. A recent work releases a large scale weakly labelled web image dataset (WebVision [136]).

8.9 Self-taught Learning

A natural assumption among most of the literature is that the source data are extensive and labelled. This may be because the source data are generally treated as the auxiliary data for instructing or teaching the target task and the unlabelled source data could be unrelated and may lead to negative transfer. However, some research works argue that the redundant unlabelled source data can still be a treasure as a good starting point of parameters for target task as mentioned in Section 5.5. How to effectively leverage the massively available unlabelled source data to improve the transfer learning approaches is an interesting problem.

8.10 Large Scale and Versatile Datasets for Transfer Learning

The development of algorithms usually depends very much on the available datasets for evaluation. Most of the current visual datasets for cross-dataset recognition are small scale in terms of either number of classes or number of samples and they are especially not suitable for evaluating deep learning algorithms. An establishment of truly large scale versatile (i.e. suitable for different problems) and realistic dataset would drive the research a significant step forward. As well known, the creation of a large scale dataset may be unaffordably expensive. Combinations and re-targeting of existent datasets can be an effective and economical way as demonstrated in [275]. As shown in Table 1, there are few visual recognition datasets designed for online transfer learning (e.g. P3.5, P3.6, P5.3, and P6.2). Most of the current online transfer learning deals with the detection tasks[255] or text recognition tasks[284]. To advance the transfer learning approaches for more broad and realistic applications, it is essential to create a few large scale datasets for online transfer learning.

9 CONCLUSION

Transfer learning from previous data for current tasks has a wide range of real-world applications. Many transfer learning algorithms for cross-dataset visual recognition have been developed in the last decade as reviewed in this paper. A key question that often puzzles a practitioner or a researcher is that which algorithm should be adopted for a given task. This paper intends to answer the question by providing a problem-oriented taxonomy of transfer learning for cross-dataset recognition and a comprehensive survey of the recently developed algorithms with respect to the taxonomy. Specifically, we believe the choice of an algorithm for a given target task should be guided by the attributes of both source and target datasets and the problem-oriented taxonomy offers an easy way to look up the problem and

the methods that are likely to solve the problem. In addition, the problem-oriented taxonomy has also shown that many challenging problems in transfer learning for visual recognition have not been well studied. It is likely that research will focus on these problems in the future.

Though it is impossible for this survey to cover all the published papers on this topic, the selected works have well represented the recent advances and in-depth analysis of these works have revealed the future research directions in transfer learning for cross-dataset visual recognition.

ACKNOWLEDGMENTS

This work is partially supported by the Australian Research Council Future Fellowship under Grant FT180100116.

REFERENCES

[1] Zeynep Akata, Mateusz Malinowski, Mario Fritz, and Bernt Schiele. 2016. Multi-cue zero-shot learning with strong supervision. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 59–68.

[2] Zeynep Akata, Florent Perronnin, Zaid Harchaoui, and Cordelia Schmid. 2013. Label-embedding for attribute-based classification. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 819–826.

[3] Zeynep Akata, Scott Reed, Daniel Walter, Honglak Lee, and Bernt Schiele. 2015. Evaluation of output embeddings for fine-grained image classification. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 2927–2936.

[4] Rahaf Aljundi, Rémi Emonet, Damien Muselet, and Marc Sebban. 2015. Landmarks-based Kernelized Subspace Alignment for Unsupervised Domain Adaptation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 56–63.

[5] Theodore Wilbur Anderson. 1984. *An introduction to multivariate statistical analysis*. Vol. 2.

[6] Galen Andrew, Raman Arora, Jeff Bilmes, and Karen Livescu. 2013. Deep canonical correlation analysis. In *Proc. International Conference on Machine Learning*. 1247–1255.

[7] Yusuf Aytar and Andrew Zisserman. 2011. Tabula rasa: Model transfer for object category detection. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 2252–2259.

[8] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proc. International Conference on Learning Representations*.

[9] Mahsa Baktashmotlagh, Mehrtash T Harandi, Brian C Lovell, and Mathieu Salzmann. 2013. Unsupervised domain adaptation by domain invariant projection. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 769–776.

[10] Mahsa Baktashmotlagh, Mehrtash T Harandi, Brian C Lovell, and Mathieu Salzmann. 2014. Domain adaptation on the statistical manifold. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2481–2488.

[11] Davide Baltieri, Roberto Vezzani, and Rita Cucchiara. 2011. 3dpes: 3d people dataset for surveillance and forensics. In *Proc. joint ACM workshop on Human gesture and behavior understanding*. ACM, 59–64.

[12] Oscar Beijbom. 2012. *Domain adaptations for computer vision applications*. Technical Report. University of California San Diego.

[13] Sagie Benaim and Lior Wolf. 2017. One-Sided Unsupervised Domain Mapping. In *Advances in Neural Information Processing Systems*.

[14] Yoshua Bengio. 2012. Deep Learning of Representations for Unsupervised and Transfer Learning. *Unsupervised and Transfer Learning Challenges in Machine Learning, Volume 7* (2012), 19.

[15] Adeleh Bitarafan, Mahdieh Soleymani Baghshah, and Marzieh Gheisari. 2016. Incremental Evolving Domain Adaptation. *IEEE Transactions on Knowledge and Data Engineering* 28, 8 (Aug 2016), 2128–2141.

[16] Gilles Blanchard, Gyemin Lee, and Clayton Scott. 2011. Generalizing from several related classification tasks to a new unlabeled sample. In *Proc. Advances in Neural Information Processing Systems*. 2178–2186.

[17] Konstantinos Bousmalis, Nathan Silberman, David Dohan, Dumitru Erhan, and Dilip Krishnan. 2017. Unsupervised Pixel-Level Domain Adaptation with Generative Adversarial Networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

[18] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. 2016. Domain separation networks. In *Advances in Neural Information Processing Systems*. 343–351.

[19] Maxime Bucher, Stéphane Herbin, and Frédéric Jurie. 2016. Improving Semantic Embedding Consistency by Metric Learning for Zero-Shot Classiffication. In *Proc. European Conference on Computer Vision*. Springer, 730–746.

[20] Liangliang Cao, Zicheng Liu, and Thomas S Huang. 2010. Cross-dataset action detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1998–2005.

[21] Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Michael I Jordan. 2017. Partial Transfer Learning with Selective Adversarial Networks. *arXiv preprint arXiv:1707.07901* (2017).

[22] Barbara Caputo and Novi Patricia. 2014. Overview of the imageclef 2014 domain adaptation task. In *ImageCLEF 2014: Overview and analysis of the results*.

[23] Lluís Castrejón, Yusuf Aytar, Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. 2016. Learning Aligned Cross-Modal Representations from Weakly Aligned Data. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2940–2949.

[24] Chao-Yeh Chen and Kristen Grauman. 2014. Inferring analogous attributes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 200–207.

[25] Lin Chen, Wen Li, and Dong Xu. 2014. Recognizing RGB images by learning from RGB-D data. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1418–1425.

[26] Minmin Chen, Kilian Q Weinberger, and John Blitzer. 2011. Co-training for domain adaptation. In *Proc. Advances in Neural Information Processing Systems*. 2456–2464.

[27] Minmin Chen, Zhixiang Xu, Fei Sha, and Kilian Q Weinberger. 2012. Marginalized Denoising Autoencoders for Domain Adaptation. In *Proc. International Conference on Machine Learning*. 767–774.

[28] Zhiyuan Chen, Nianzu Ma, and Bing Liu. 2015. Lifelong learning for sentiment classification. In *Association for Computational Linguistics*.

[29] Dong Seon Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino. 2011. Custom Pictorial Structures for Re-identification. *Proc. British Machine Vision Conference*.

[30] Zhongwei Cheng, Lei Qin, Yituo Ye, Qingming Huang, and Qi Tian. 2012. Human daily action analysis with multi-view and color-depth data. In *Proc. European Conference on Computer Vision*. Springer, 52–61.

[31] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. *Syntax, Semantics and Structure in Statistical Translation* (2014), 103.

[32] Sumit Chopra, Suhrid Balakrishnan, and Raghuraman Gopalan. 2013. DLID: Deep learning for domain adaptation by interpolating between domains. In *Proc. ICML Workshop on Challenges in Representation Learning*. Citeseer.

[33] Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng. 2009. NUS-WIDE: a real-world web image database from National University of Singapore. In *Proc. ACM international conference on image and video retrieval*. ACM, 48.

[34] Adam Coates, Andrew Ng, and Honglak Lee. 2011. An analysis of single-layer networks in unsupervised feature learning. In *Proc. international conference on artificial intelligence and statistics*. 215–223.

[35] Diane Cook, Kyle D Feuz, and Narayanan C Krishnan. 2013. Transfer learning for activity recognition: A survey. *Knowledge and information systems* 36, 3 (2013), 537–556.

[36] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. 2016. Optimal transport for Domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2016).

[37] Gabriela Csurka. 2017. A comprehensive survey on domain adaptation for visual applications. In *Domain Adaptation in Computer Vision Applications*. Springer, 1–35.

[38] Gabriela Csurka, Boris Chidlovskii, and Florent Perronnin. 2014. Domain adaptation with a domain specific class means classifier. In *Proc. European Conference on Computer Vision Workshops*. Springer, 32–46.

[39] Zhen Cui, Hong Chang, Shiguang Shan, and Xilin Chen. 2014. Generalized unsupervised manifold alignment. In *Proc. Advances in Neural Information Processing Systems*. 2429–2437.

[40] Zhen Cui, Wen Li, Dong Xu, Shiguang Shan, Xilin Chen, and Xuelong Li. 2014. Flowing on Riemannian manifold: Domain adaptation by shifting covariance. *IEEE Transactions on Cybernetics* 44, 12 (2014), 2264–2273.

[41] Wenyuan Dai, Yuqiang Chen, Gui-Rong Xue, Qiang Yang, and Yong Yu. 2009. Translated learning: Transfer learning across different feature spaces. In *Proc. Advances in Neural Information Processing Systems*. 353–360.

[42] Wenyuan Dai, Gui-Rong Xue, Qiang Yang, and Yong Yu. 2007. Transferring naive bayes classifiers for text classification. In *Proc. AAAI Conference on Artificial Intelligence*, Vol. 22. 540.

[43] Wenyuan Dai, Qiang Yang, Gui-Rong Xue, and Yong Yu. 2007. Boosting for transfer learning. In *Proc. International Conference on Machine Learning*. ACM, 193–200.

[44] Hal Daumé III. 2007. Frustratingly easy domain adaptation. In *Proc. Annual Meeting of the Association of Computational Linguistics*. 256ï£¡–263.

[45] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 248–255.

[46] Tom Diethe, David R Hardoon, and John Shawe-taylor. 2008. Multiview Fisher discriminant analysis. In *In NIPS Workshop on Learning from Multiple Sources*. Citeseer.

[47] Zhengming Ding, Ming Shao, and Yun Fu. 2015. Deep Low-Rank Coding for Transfer Learning. In *Proc. International Joint Conference on Artificial Intelligence*. 3453–3459.

[48] Zhengming Ding, Ming Shao, and Yun Fu. 2015. Missing Modality Transfer Learning via Latent Low-Rank Constraint. *IEEE Transactions on Image Processing* 24, 11 (2015), 4322–4334.

[49] Zhengming Ding, Ming Shao, and Yun Fu. 2016. Transfer learning for image classification with incomplete multiple sources. In *Proc. International Joint Conference on Neural Networks*. IEEE, 2188–2195.

[50] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2014. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. In *Proc. International Conference on Machine Learning*. 647–655.

[51] Simon S Du, Jayanth Koushik, Aarti Singh, and Barnabas Poczos. 2017. Hypothesis Transfer Learning via Transformation Functions. In *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). 574–584.

[52] Lixin Duan, Ivor W Tsang, and Dong Xu. 2012. Domain transfer multiple kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 3 (2012), 465–479.

[53] Lixin Duan, Ivor W Tsang, Dong Xu, and Tat-Seng Chua. 2009. Domain adaptation from multiple sources via auxiliary classifiers. In *Proc. International Conference on Machine Learning*. ACM, 289–296.

[54] Lixin Duan, Dong Xu, IW-H Tsang, and Jiebo Luo. 2012. Visual event recognition in videos by learning from web data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 9 (2012), 1667–1680.

[55] Lixin Duan, Dong Xu, and Ivor W Tsang. 2012. Domain adaptation from multiple sources: A domain-dependent regularization approach. *IEEE Transactions on Neural Networks and Learning Systems* 23, 3 (2012), 504–518.

[56] Lixin Duan, Dong Xu, and Ivor W Tsang. 2012. Learning with Augmented Features for Heterogeneous Domain Adaptation. In *Proc. International Conference on Machine Learning*. 711–718.

[57] David Eigen and Rob Fergus. 2015. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proc. IEEE International Conference on Computer Vision*. 2650–2658.

[58] Chen Fang, Ye Xu, and Daniel Rockmore. 2013. Unbiased metric learning: On the utilization of multiple datasets and web images for softening bias. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 1657–1664.

[59] Ali Farhadi, Ian Endres, Derek Hoiem, and David Forsyth. 2009. Describing objects by their attributes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1778–1785.

[60] Li Fei-Fei, Rob Fergus, and Pietro Perona. 2006. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 4 (2006), 594–611.

[61] Fangxiang Feng, Xiaojie Wang, and Ruifan Li. 2014. Cross-modal retrieval with correspondence autoencoder. In *Proc. ACM international conference on Multimedia*. ACM, 7–16.

[62] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars. 2013. Unsupervised visual domain adaptation using subspace alignment. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 2960–2967.

[63] Geoffrey French, Michal Mackiewicz, and Mark Fisher. 2017. Self-ensembling for domain adaptation. *arXiv preprint arXiv:1706.05208* (2017).

[64] Andrea Frome, Greg S Corrado, Jon Shlens, Samy Bengio, Jeff Dean, Tomas Mikolov, et al. 2013. Devise: A deep visual-semantic embedding model. In *Advances in neural information processing systems*. 2121–2129.

[65] Yanwei Fu, Timothy M Hospedales, Tao Xiang, Zhenyong Fu, and Shaogang Gong. 2014. Transductive multi-view embedding for zero-shot recognition and annotation. In *European Conference on Computer Vision*. Springer, 584–599.

[66] Yanwei Fu, Timothy M Hospedales, Tao Xiang, and Shaogang Gong. 2015. Transductive multi-view zero-shot learning. *IEEE transactions on pattern analysis and machine intelligence* 37, 11 (2015), 2332–2345.

[67] Jo ao Gama, Indr.e vZliobait.e, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. 2014. A survey on concept drift adaptation. *Comput. Surveys* 46, 4 (2014), 44.

[68] Chuang Gan, Tianbao Yang, and Boqing Gong. 2016. Learning Attributes Equals Multi-Source Domain Generalization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 87–97.

[69] Junying Gan, Lichen Li, Yikui Zhai, and Yinhua Liu. 2014. Deep self-taught learning for facial beauty prediction. *Neurocomputing* 144 (2014), 295–303.

[70] Yaroslav Ganin and Victor Lempitsky. 2015. Unsupervised Domain Adaptation by Backpropagation. In *Proc. International Conference on Machine Learning*. 1180–1189.

[71] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *Journal of Machine Learning Research* 17, 59 (2016), 1–35.

[72] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2016. Image style transfer using convolutional neural networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 2414–2423.

[73] Muhammad Ghifary, David Balduzzi, W Bastiaan Kleijn, and Mengjie Zhang. 2016. Scatter Component Analysis: A Unified Framework for Domain Adaptation and Domain Generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP, 99 (2016), 1–1.

[74] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, and David Balduzzi. 2015. Domain Generalization for Object Recognition with Multi-task Autoencoders. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 2551–2559.

[75] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, David Balduzzi, and Wen Li. 2016. Deep Reconstruction-Classification Networks for Unsupervised Domain Adaptation. In *European Conference on Computer Vision*. Springer, 597–613.

[76] Behnam Gholami, Ognjen (Oggi) Rudovic, and Vladimir Pavlovic. 2017. PUnDA: Probabilistic Unsupervised Domain Adaptation for Knowledge Transfer Across Visual Categories. In *Proc. IEEE International Conference on Computer Vision*.

[77] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proc. International Conference on Machine Learning*. 513–520.

[78] Boqing Gong, Kristen Grauman, and Fei Sha. 2013. Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation. In *Proc. International Conference on Machine Learning*. 222–230.

[79] Boqing Gong, Kristen Grauman, and Fei Sha. 2013. Reshaping visual datasets for domain adaptation. In *Proc. Advances in Neural Information Processing Systems*. 1286–1294.

[80] Boqing Gong, Kristen Grauman, and Fei Sha. 2014. Learning kernels for unsupervised domain adaptation with applications to visual object recognition. *International Journal of Computer Vision* 109, 1-2 (2014), 3–27.

[81] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. 2012. Geodesic flow kernel for unsupervised domain adaptation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2066–2073.

[82] Mingming Gong, Kun Zhang, Tongliang Liu, Dacheng Tao, Clark Glymour, and Bernhard Schölkopf. 2016. Domain adaptation with conditional transferable components. In *Proc. International Conference on Machine Learning*. 2839–2848.

[83] Yunchao Gong, Qifa Ke, Michael Isard, and Svetlana Lazebnik. 2014. A multi-view embedding space for modeling internet images, tags, and their semantics. *International journal of computer vision* 106, 2 (2014), 210–233.

[84] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

[85] Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. 2011. Domain adaptation for object recognition: An unsupervised approach. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 999–1006.

[86] Raghavan Gopalan, Ruonan Li, and Rama Chellappa. 2014. Unsupervised adaptation across domain shifts by generating intermediate data representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 11 (2014), 2288–2302.

[87] Douglas Gray, Shane Brennan, and Hai Tao. 2007. Evaluating appearance models for recognition, reacquisition, and tracking. In *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, Vol. 3. 1–7.

[88] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. 2010. Multi-pie. *Image and Vision Computing* 28, 5 (2010), 807–813.

[89] Saurabh Gupta, Judy Hoffman, and Jitendra Malik. 2016. Cross modal distillation for supervision transfer. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 2827–2836.

[90] Martin Hirzer, Csaba Beleznai, Peter M Roth, and Horst Bischof. 2011. Person re-identification by descriptive and discriminative classification. In *Proc. Scandinavian conference on Image analysis*. Springer, 91–102.

[91] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[92] Judy Hoffman, Trevor Darrell, and Kate Saenko. 2014. Continuous manifold based adaptation for evolving visual domains. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 867–874.

[93] Judy Hoffman, Brian Kulis, Trevor Darrell, and Kate Saenko. 2012. Discovering latent domains for multisource domain adaptation. In *Proc. European Conference on Computer Vision*. Springer, 702–715.

[94] Tzu Ming Harry Hsu, Wei Yu Chen, Cheng-An Hou, Yao-Hung Hubert Tsai, Yi-Ren Yeh, and Yu-Chiang Frank Wang. 2015. Unsupervised Domain Adaptation With Imbalanced Cross-Domain Data. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 4121–4129.

[95] Derek Hao Hu and Qiang Yang. 2011. Transfer learning for activity recognition via sensor mapping. In *Proc. International Joint Conference on Artificial Intelligence*, Vol. 22. 1962–1967.

[96] De-An Huang and Yu-Chiang Frank Wang. 2013. Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 2496–2503.

[97] Jiayuan Huang, Arthur Gretton, Karsten M Borgwardt, Bernhard Schölkopf, and Alex J Smola. 2006. Correcting sample selection bias by unlabeled data. In *Proc. Advances in Neural Information Processing Systems*. 601–608.

[98] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

[99] Vidit Jain and Erik Learned-Miller. 2011. Online domain adaptation of a pre-trained cascade of classifiers. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 577–584.

[100] Nathalie Japkowicz and Shaju Stephen. 2002. The class imbalance problem: A systematic study. *Intelligent Data Analysis* 6, 5 (2002), 429–449.

[101] Dinesh Jayaraman and Kristen Grauman. 2014. Zero-shot recognition with unreliable attributes. In *Proc. Advances in Neural Information Processing Systems*. 3464–3472.

[102] I-Hong Jhuo, Dong Liu, DT Lee, Shih-Fu Chang, et al. 2012. Robust visual domain adaptation with low-rank reconstruction. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2168–2175.

[103] Chengcheng Jia, Yu Kong, Zhengming Ding, and Yun Raymond Fu. 2014. Latent tensor transfer learning for rgb-d action recognition. In *Proc. ACM International Conference on Multimedia*. ACM, 87–96.

[104] Wenhao Jiang, Hongchang Gao, Fu-lai Chung, and Heng Huang. 2016. The l2, 1-Norm Stacked Robust Autoencoders for Domain Adaptation. In *Proc. AAAI Conference on Artificial Intelligence*.

[105] Wei Jiang, Eric Zavesky, Shih-Fu Chang, and Alex Loui. 2008. Cross-domain learning methods for high-level visual concept classification. In *Proc. IEEE International Conference on Image Processing*. IEEE, 161–164.

[106] Luo Jie, Tatiana Tommasi, and Barbara Caputo. 2011. Multiclass transfer learning from unconstrained priors. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 1863–1870.

[107] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *European Conference on Computer Vision*.

[108] Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent Continuous Translation Models. In *Proc. Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

[109] Meina Kan, Shiguang Shan, and Xilin Chen. 2015. Bi-shifting Auto-Encoder for Unsupervised Domain Adaptation. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 3846–3854.

[110] Meina Kan, Shiguang Shan, Haihong Zhang, Shihong Lao, and Xilin Chen. 2012. Multi-view discriminant analysis. In *Proc. European Conference on Computer Vision*. Springer, 808–821.

[111] Aditya Khosla, Tinghui Zhou, Tomasz Malisiewicz, Alexei A Efros, and Antonio Torralba. 2012. Undoing the damage of dataset bias. In *Proc. European Conference on Computer Vision*. Springer, 158–171.

[112] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. 2017. Learning to Discover Cross-Domain Relations with Generative Adversarial Networks. In *Proc. International Conference on Machine Learning*, Doina Precup and Yee Whye Teh (Eds.), Vol. 70. International Convention Centre, Sydney, Australia, 1857–1865.

[113] Gregory Koch, TORONTO EDU, Richard Zemel, and Ruslan Salakhutdinov. 2015. Siamese Neural Networks for One-shot Image Recognition. In *Proc. ICML Deep Learning Workshop*.

[114] Elyor Kodirov, Tao Xiang, Zhenyong Fu, and Shaogang Gong. 2015. Unsupervised domain adaptation for zero-shot learning. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 2452–2460.

[115] Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proc. Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 48–54.

[116] Piotr Koniusz, Yusuf Tas, and Fatih Porikli. 2017. Domain Adaptation by Mixture of Alignments of Second-or Higher-Order Scatter Tensors. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

[117] Jason Kuen, Kian Ming Lim, and Chin Poo Lee. 2015. Self-taught learning of a deep invariant representation for visual tracking via temporal slowness principle. *Pattern Recognition* 48, 10 (2015), 2964–2982.

[118] Brian Kulis, Kate Saenko, and Trevor Darrell. 2011. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1785–1792.

[119] Wataru Kumagai. 2016. Learning Bound for Parameter Transfer Learning. In *Advances in Neural Information Processing Systems*. 2721–2729.

[120] Neeraj Kumar, Alexander C Berg, Peter N Belhumeur, and Shree K Nayar. 2009. Attribute and simile classifiers for face verification. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 365–372.

[121] Ilja Kuzborskij and Francesco Orabona. 2013. Stability and Hypothesis Transfer Learning. In *Proc. International Conference on Machine Learning*. 942–950.

[122] Pierre-Yves Laffont, Zhile Ren, Xiaofeng Tao, Chao Qian, and James Hays. 2014. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Transactions on Graphics* 33, 4 (2014), 149.

[123] Brenden Lake, Ruslan Salakhutdinov, Jason Gross, and Joshua Tenenbaum. 2011. One shot learning of simple visual concepts. In *Proc. Cognitive Science Society*, Vol. 33.

[124] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. 2009. Learning to detect unseen object classes by between-class attribute transfer. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 951–958.

[125] Quoc V Le, Will Y Zou, Serena Y Yeung, and Andrew Y Ng. 2011. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3361–3368.

[126] Honglak Lee, Rajat Raina, Alex Teichman, and Andrew Y Ng. 2009. Exponential family sparse coding with applications to self-taught learning. In *Proc. International Joint Conference on Artifical intelligence*. Morgan Kaufmann Publishers Inc., 1113–1119.

[127] Sang-Woo Lee, Jin-Hwa Kim, Jaehyun Jun, Jung-Woo Ha, and Byoung-Tak Zhang. 2017. Overcoming catastrophic forgetting by incremental moment matching. In *Advances in Neural Information Processing Systems*. 4655–4665.

[128] Jimmy Lei Ba, Kevin Swersky, Sanja Fidler, et al. 2015. Predicting deep zero-shot convolutional neural networks using textual descriptions. In *Proc. IEEE International Conference on Computer Vision*. 4247–4255.

[129] Chunyuan Li, Hao Liu, Changyou Chen, Yuchen Pu, Liqun Chen, Ricardo Henao, and Lawrence Carin. 2017. Alice: Towards understanding adversarial learning for joint distribution matching. In *Advances in Neural Information Processing Systems*. 5501–5509.

[130] Chuan Li and Michael Wand. 2016. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *European Conference on Computer Vision*.

[131] Chun-Liang Li, Wei-Cheng Chang, Yu Cheng, Yiming Yang, and Barnabás Póczos. 2017. MMD GAN: Towards deeper understanding of moment matching network. In *Advances in Neural Information Processing Systems*. 2200–2210.

[132] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. 2017. Deeper, Broader and Artier Domain Generalization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 5542–5550.

[133] Stan Z Li, Dong Yi, Zhen Lei, and Shengcai Liao. 2013. The casia nir-vis 2.0 face database. In *Proc. IEEE Conference on Computer vision and pattern recognition workshops*. IEEE, 348–353.

[134] Wen Li, Lin Chen, Dong Xu, and Luc Van Gool. 2017. Visual Recognition in RGB Images and Videos by Learning from RGB-D Data. *IEEE transactions on pattern analysis and machine intelligence* (2017).

[135] Wen Li, Lixin Duan, Dong Xu, and Ivor W Tsang. 2014. Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 6 (2014), 1134–1148.

[136] Wen Li, Limin Wang, Wei Li, Eirikur Agustsson, and Luc Van Gool. 2017. WebVision Database: Visual Learning and Understanding from Web Data. *arXiv preprint arXiv:1708.02862* (2017).

[137] Wei Li and Xiaogang Wang. 2013. Locally aligned feature transforms across views. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 3594–3601.

[138] Wen Li, Zheng Xu, Dong Xu, Dengxin Dai, and Luc Van Gool. 2017. Domain generalization and adaptation using low rank exemplar svms. *IEEE transactions on pattern analysis and machine intelligence* (2017).

[139] Xin Li, Yuhong Guo, and Dale Schuurmans. 2015. Semi-supervised zero-shot classification with label representation learning. In *Proc. IEEE International Conference on Computer Vision*. 4211–4219.

[140] Yanan Li, Donghui Wang, Huanhang Hu, Yuetan Lin, and Yueting Zhuang. 2017. Zero-Shot Recognition using Dual Visual-Semantic Mapping Paths. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

[141] Zhizhong Li and Derek Hoiem. 2016. Learning without forgetting. In *European Conference on Computer Vision*. Springer, 614–629.

[142] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.

[143] Jingen Liu, Benjamin Kuipers, and Silvio Savarese. 2011. Recognizing human actions by attributes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3337–3344.

[144] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. 2017. Unsupervised Image-to-Image Translation Networks. In *Advances in Neural Information Processing Systems*.

[145] Ming-Yu Liu and Oncel Tuzel. 2016. Coupled generative adversarial networks. In *Advances in Neural Information Processing Systems*. 469–477.

[146] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 3431–3440.

[147] Mingsheng Long, Guiguang Ding, Jianmin Wang, Jiaguang Sun, Yuchen Guo, and Philip S Yu. 2013. Transfer sparse coding for robust image representation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 407–414.

[148] Mingsheng Long and Jianmin Wang. 2015. Learning Transferable Features with Deep Adaptation Networks. In *Proc. International Conference on Machine Learning*. 97–105.

[149] Mingsheng Long, Jianmin Wang, Guiguang Ding, Sinno Jialin Pan, and Philip S Yu. 2014. Adaptation regularization: A general framework for transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 26, 5 (2014), 1076–1089.

[150] Mingsheng Long, Jianmin Wang, Guiguang Ding, Jiaguang Sun, and Philip S Yu. 2013. Transfer feature learning with joint distribution adaptation. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 2200–2207.

[151] Mingsheng Long, Jianmin Wang, and Michael I Jordan. 2017. Deep transfer learning with joint adaptation networks. In *Proc. International Conference on Machine Learning*.

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

[153] Hao Lu, Lei Zhang, Zhiguo Cao, Wei Wei, Ke Xian, Chunhua Shen, and Anton van den Hengel. 2017. When Unsupervised Domain Adaptation Meets Tensor Representations. In *Proc. IEEE International Conference on Computer Vision*.

[154] Jie Lu, Vahid Behbood, Peng Hao, Hua Zuo, Shan Xue, and Guangquan Zhang. 2015. Transfer learning using computational intelligence: a survey. *Knowledge-Based Systems* 80 (2015), 14–23.

[155] Zhigang Ma, Yi Yang, Yang Cai, Nicu Sebe, and Alexander G Hauptmann. 2012. Knowledge adaptation for ad hoc multimedia event detection with few exemplars. In *Proc. ACM International Conference on Multimedia*. ACM, 469–478.

[156] Zhigang Ma, Yi Yang, Feiping Nie, Nicu Sebe, Shuicheng Yan, and Alexander G Hauptmann. 2014. Harnessing lab knowledge for real-world action recognition. *International Journal of Computer Vision* 109, 1-2 (2014), 60–73.

[157] Anna Margolis. 2011. *A literature review of domain adaptation with unlabeled data.* Technical Report.

[158] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[159] Paritosh Mittal, Aishwarya Jain, Gaurav Goswami, Richa Singh, and Mayank Vatsa. 2014. Recognizing composite sketches with digital face images via SSD dictionary. In *Proc. IEEE International Joint Conference on Biometrics*. IEEE, 1–6.

[160] Jose G Moreno-Torres, Troy Raeder, RocíO Alaiz-RodríGuez, Nitesh V Chawla, and Francisco Herrera. 2012. A unifying view on dataset shift in classification. *Pattern Recognition* 45, 1 (2012), 521–530.

[161] Saeid Motiian, Quinn Jones, Seyed Iranmanesh, and Gianfranco Doretto. 2017. Few-Shot Adversarial Domain Adaptation. In *Advances in Neural Information Processing Systems*. 6673–6683.

[162] Saeid Motiian, Marco Piccirilli, Donald A Adjeroh, and Gianfranco Doretto. 2017. Unified Deep Supervised Domain Adaptation and Generalization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 5715–5725.

[163] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. 2013. Domain Generalization via Invariant Feature Representation. In *Proc. International Conference on Machine Learning*. 10–18.

[164] Fabian Nater, Tatiana Tommasi, Helmut Grabner, Luc Van Gool, and Barbara Caputo. 2011. Transferring activities: Updating human behavior analysis. In *Proc. IEEE International Conference on Computer Vision Workshops*. IEEE, 1737–1744.

[165] Jie Ni, Qiang Qiu, and Rama Chellappa. 2013. Subspace interpolation via dictionary learning for unsupervised domain adaptation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 692–699.

[166] Li Niu, Wen Li, and Dong Xu. 2015. Multi-view domain generalization for visual recognition. In *Proc. IEEE International Conference on Computer Vision*. 4193–4201.

[167] Li Niu, Wen Li, and Dong Xu. 2015. Visual Recognition by Learning from Web Data: A Weakly Supervised Domain Generalization Approach. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2774–2783.

[168] Mark Palatucci, Dean Pomerleau, Geoffrey E Hinton, and Tom M Mitchell. 2009. Zero-shot learning with semantic output codes. In *Proc. Advances in Neural Information Processing Systems*. 1410–1418.

[169] Sinnojialin Pan, Ivor W Tsang, James Tin Yau Kwok, and Qiang Yang. 2009. Domain Adaptation via Transfer Component Analysis. In *Proc. International Joint Conference on Artificial Intelligence*. 1187.

[170] Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 10 (2010), 1345–1359.

[171] Pau Panareda Busto and Juergen Gall. 2017. Open Set Domain Adaptation. In *Proc. IEEE International Conference on Computer Vision*.

[172] Devi Parikh and Kristen Grauman. 2011. Relative attributes. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 503–510.

[173] Vishal M Patel, Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. 2015. Visual domain adaptation: A survey of recent advances. *IEEE Signal Processing Magazine* 32, 3 (2015), 53–69.

[174] Novi Patricia and Barbara Caputo. 2014. Learning to learn, from transfer learning to domain adaptation: A unifying perspective. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1442–1449.

[175] Genevieve Patterson, Chen Xu, Hang Su, and James Hays. 2014. The sun attribute database: Beyond categories for deeper scene understanding. *International Journal of Computer Vision* 108, 1-2 (2014), 59–81.

[176] Yuxin Peng, Xin Huang, and Jinwei Qi. 2016. Cross-media shared representation by hierarchical learning with multiple deep networks. In *Proc. International Joint Conference on Artificial Intelligence*. AAAI Press, 3846–3853.

[177] Jose Costa Pereira, Emanuele Coviello, Gabriel Doyle, Nikhil Rasiwasia, Gert RG Lanckriet, Roger Levy, and Nuno Vasconcelos. 2014. On the role of correlation and abstraction in cross-modal multimedia retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 3 (2014), 521–535.

[178] David N Perkins and Gavriel Salomon. 1992. Transfer of learning. *International Encyclopedia of Education* 2 (1992), 6452–6457.

[179] Michaël Perrot and Amaury Habrard. 2015. A Theoretical Analysis of Metric Hypothesis Transfer Learning. In *Proc. International Conference on Machine Learning*. 1708–1717.

[180] Guo-Jun Qi, Charu Aggarwal, and Thomas Huang. 2011. Towards semantic knowledge propagation from text corpus to web images. In *Proc. international conference on World wide web*. ACM, 297–306.

[181] Ruizhi Qiao, Lingqiao Liu, Chunhua Shen, and Anton van den Hengel. 2016. Less is more: zero-shot learning from online textual documents with noise suppression. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 2249–2257.

[182] Jie Qin, Li Liu, Ling Shao, Fumin Shen, Bingbing Ni, Jiaxin Chen, and Yunhong Wang. 2017. Zero-shot action recognition with error-correcting output codes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

[183] Brian Quanz and Jun Huan. 2009. Large margin transductive transfer learning. In *Proc. ACM conference on Information and knowledge management*. ACM, 1327–1336.

[184] Joaquin Quionero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D Lawrence. 2009. *Dataset shift in machine learning*. The MIT Press.

[185] Rajat Raina, Alexis Battle, Honglak Lee, Benjamin Packer, and Andrew Y Ng. 2007. Self-taught learning: transfer learning from unlabeled data. In *Proc. International Conference on Machine learning*. ACM, 759–766.

[186] Cyrus Rashtchian, Peter Young, Micah Hodosh, and Julia Hockenmaier. 2010. Collecting image annotations using Amazon's Mechanical Turk. In *Proc. NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Association for Computational Linguistics, 139–147.

[187] Mohammad Rastegari, Ali Farhadi, and David Forsyth. 2012. Attribute Discovery via Predictable Discriminative Binary Codes. In *European Conference on Computer Vision*. 876–889.

[188] Sachin Ravi and Hugo Larochelle. 2017. Optimization as a Model for Few-Shot Learning. In *Proc. International Conference on Learning Representations*.

[189] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. 2014. CNN features off-the-shelf: an astounding baseline for recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 512–519.

[190] Scott Reed, Zeynep Akata, Honglak Lee, and Bernt Schiele. 2016. Learning deep representations of fine-grained visual descriptions. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 49–58.

[191] Bernardino Romera-Paredes and Philip Torr. 2015. An embarrassingly simple approach to zero-shot learning. In *Proc. International Conference on Machine Learning*. 2152–2161.

[192] Paul Ruvolo and Eric Eaton. 2013. ELLA: An Efficient Lifelong Learning Algorithm. In *Proc. International Conference on Machine Learning*. 507–515.

[193] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. 2010. Adapting visual category models to new domains. In *Proc. European Conference on Computer Vision*. Springer, 213–226.

[194] Kuniaki Saito, Yoshitaka Ushiku, and Tatsuya Harada. 2017. Asymmetric Tri-training for Unsupervised Domain Adaptation. In *Proc. International Conference on Machine Learning*.

[195] Ling Shao, Fan Zhu, and Xuelong Li. 2015. Transfer learning for visual categorization: a survey. *IEEE Transactions on Neural Networks and Larning Systems* 26, 5 (2015), 1019–1034.

[196] Ming Shao, Carlos Castillo, Zhenghong Gu, and Yun Fu. 2012. Low-rank transfer subspace learning. In *Proc. IEEE International Conference on Data Mining*. IEEE, 1104–1109.

[197] Ming Shao, Dmitry Kit, and Yun Fu. 2014. Generalized transfer subspace learning through low-rank constraint. *International Journal of Computer Vision* 109, 1-2 (2014), 74–93.

[198] Sumit Shekhar, Vishal M Patel, Hien V Nguyen, and Rama Chellappa. 2013. Generalized domain-adaptive dictionaries. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 361–368.

[199] Sumit Shekhar, Vishal M Patel, Hien Van Nguyen, and Rama Chellappa. 2015. Coupled projections for adaptation of dictionaries. *IEEE Transactions on Image Processing* 24, 10 (2015), 2941–2954.

[200] Yuan Shi and Fei Sha. 2012. Information-Theoretical Learning of Discriminative Clusters for Unsupervised Domain Adaptation. In *Proc. International Conference on Machine Learning*. 1079–1086.

[201] Yichang Shih, Sylvain Paris, Frédo Durand, and William T Freeman. 2013. Data-driven hallucination of different times of day from a single outdoor photo. *ACM Transactions on Graphics* 32, 6 (2013), 200.

[202] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. 2017. Continual learning with deep generative replay. In *Advances in Neural Information Processing Systems*. 2994–3003.

[203] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Josh Susskind, Wenda Wang, and Russ Webb. 2017. Learning from simulated and unsupervised images through adversarial training. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.

[204] Si Si, Dacheng Tao, and Bo Geng. 2010. Bregman divergence-based regularization for transfer subspace learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 7 (2010), 929–942.

[205] Jake Snell, Kevin Swersky, and Richard S Zemel. 2017. Prototypical Networks for Few-shot Learning. In *Proc. International Conference on Learning Representations*.

[206] Richard Socher, Milind Ganjoo, Christopher D Manning, and Andrew Ng. 2013. Zero-shot learning through cross-modal transfer. In *Proc. Advances in Neural Information Processing Systems*. 935–943.

[207] Dimitris Stamos, Samuele Martelli, Moin Nabi, Andrew McDonald, Vittorio Murino, and Massimiliano Pontil. 2015. Learning with dataset bias in latent subcategory models. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3650–3658.

[208] Masashi Sugiyama, Shinichi Nakajima, Hisashi Kashima, Paul V Buenau, and Motoaki Kawanabe. 2008. Direct importance estimation with model selection and its application to covariate shift adaptation. In *Proc. Advances in Neural Information Processing Systems*. 1433–1440.

[209] Sanatan Sukhija, Narayanan C Krishnan, and Gurkanwal Singh. 2016. Supervised Heterogeneous Domain Adaptation via Random Forests. In *Proc. International Joint Conference on Artificial Intelligence*. AAAI Press.

[210] Baochen Sun, Jiashi Feng, and Kate Saenko. 2016. Return of Frustratingly Easy Domain Adaptation. In *Proc. AAAI Conference on Artificial Intelligence*.

[211] Baocheng Sun and Kate Saenko. 2015. Subspace Distribution Alignment for Unsupervised Domain Adaptation. In *Proc. British Machine Vision Conference*.

[212] Baochen Sun and Kate Saenko. 2016. Deep CORAL: Correlation Alignment for Deep Domain Adaptation. In *Proc. Transferring and Adapting Source Knowledge in Computer Vision (TASK-CV) in conjunction with the ECCV*.

[213] Qian Sun, Rita Chattopadhyay, Sethuraman Panchanathan, and Jieping Ye. 2011. A two-stage weighting framework for multi-source domain adaptation. In *Proc. Advances in Neural Information Processing Systems*. 505–513.

[214] Shiliang Sun, Honglei Shi, and Yuanbin Wu. 2015. A survey of multi-source domain adaptation. *Information Fusion* 24 (2015), 84–92.

[215] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. 3104–3112.

[216] Songbo Tan, Xueqi Cheng, Yuefen Wang, and Hongbo Xu. 2009. Adapting naive bayes to domain adaptation for sentiment analysis. In *Proc. European Conference on Information Retrieval*. Springer, 337–349.

[217] Kevin Tang, Vignesh Ramanathan, Li Fei-Fei, and Daphne Koller. 2012. Shifting weights: Adapting object detectors from image to video. In *Advances in Neural Information Processing Systems*. 638–646.

[218] Sebastian Thrun. 1998. Lifelong learning algorithms. *Learning to learn* 8 (1998), 181–209.

[219] Tatiana Tommasi and Barbara Caputo. 2013. Frustratingly easy nbnn domain adaptation. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 897–904.

[220] Tatiana Tommasi, Francesco Orabona, and Barbara Caputo. 2010. Safety in numbers: Learning categories from few examples with multi model knowledge transfer. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3081–3088.

[221] Tatiana Tommasi, Francesco Orabona, Mohsen Kaboli, and Barbara Caputo. 2012. Leveraging over prior knowledge for online learning of visual categories. In *Proc. British Machine Vision Conference*. 1–11.

[222] Tatiana Tommasi and Tinne Tuytelaars. 2014. A testbed for cross-dataset analysis. In *Proc. European Conference on Computer Vision*. Springer, 18–31.

[223] Antonio Torralba and Alexei Efros. 2011. Unbiased look at dataset bias. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1521–1528.

[224] Yao-Hung Hubert Tsai, Yi-Ren Yeh, and Yu-Chiang Frank Wang. 2016. Learning Cross-Domain Landmarks for Heterogeneous Domain Adaptation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 5081–5090.

[225] Eric Tzeng, Judy Hoffman, Trevor Darrell, and Kate Saenko. 2015. Simultaneous deep transfer across domains and tasks. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 4068–4076.

[226] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. 2017. Adversarial discriminative domain adaptation. *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2017).

[227] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. 2014. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474* (2014).

[228] Hemanth Venkateswara, Shayok Chakraborty, and Sethuraman Panchanathan. 2017. Deep-Learning Systems for Domain Adaptation in Computer Vision: Learning Transferable Feature Representations. *IEEE Signal Processing Magazine* 34, 6 (2017), 117–129.

[229] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. 2017. Deep Hashing Network for Unsupervised Domain Adaptation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 5018–5027.

[230] Oriol Vinyals, Charles Blundell, Tim Lillicrap, Daan Wierstra, et al. 2016. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*. 3630–3638.

[231] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. 2011. *The Caltech-UCSD Birds-200-2011 Dataset*. Technical Report CNS-TR-2011-001. California Institute of Technology.

[232] Bokun Wang, Yang Yang, Xing Xu, Alan Hanjalic, and Heng Tao Shen. 2017. Adversarial Cross-Modal Retrieval. In *Proc. ACM on Multimedia Conference*. ACM, 154–162.

[233] Chang Wang and Sridhar Mahadevan. 2011. Heterogeneous domain adaptation using manifold alignment. In *Proc. International Joint Conference on Artificial Intelligence*. 1541–1546.

[234] Donghui Wang, Yanan Li, Yuetan Lin, and Yueting Zhuang. 2016. Relational knowledge transfer for zero-shot learning. In *Proc. AAAI Conference on Artificial Intelligence*. AAAI Press, 2145–2151.

[235] Hua Wang, Feiping Nie, and Heng Huang. 2013. Robust and discriminative self-taught learning. In *International Conference on Machine Learning*. 298–306.

[236] Han Wang, Xinxiao Wu, and Yunde Jia. 2014. Video Annotation via Image Groups from the Web. *IEEE Transactions on Multimedia* 16, 5 (2014), 1282–1291.

[237] Pichao Wang, Wanqing Li, Zhimin Gao, Chang Tang, Jing Zhang, and Philip Ogunbona. 2015. ConvNets-Based Action Recognition from Depth Maps through Virtual Cameras and Pseudocoloring. In *Proc. ACM Conference on Multimedia Conference*. ACM, 1119–1122.

[238] Pichao Wang, Wanqing Li, Jun Wan, Philip Ogunbona, and Xinwang Liu. 2018. Cooperative Training of Deep Aggregation Networks for RGB-D Action Recognition. In *Proc. AAAI Conference on Artificial Intelligence*. AAAI Press.

[239] Xiaolong Wang and Abhinav Gupta. 2016. Generative image modeling using style and structure adversarial networks. In *European Conference on Computer Vision*.

[240] Xiaoyang Wang and Qiang Ji. 2013. A unified probabilistic approach modeling relationships between attributes and objects. In *Proc. IEEE International Conference on Computer Vision*. 2120–2127.

[241] Xuezhi Wang and Jeff Schneider. 2014. Flexible transfer learning under support and model shift. In *Advances in Neural Information Processing Systems*. 1898–1906.

[242] Xiaogang Wang and Xiaoou Tang. 2009. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 11 (2009), 1955–1967.

[243] Pengfei Wei, Yiping Ke, and Chi Keong Goh. 2016. Deep Nonlinear Feature Coding for Unsupervised Domain Adaptation. In *Proc. International Joint Conferences on Artificial Intelligence*.

[244] Daniel Weinland, Edmond Boyer, and Remi Ronfard. 2007. Action recognition from arbitrary views using 3d exemplars. In *Proc. IEEE International Conference on Computer Vision*. 1–7.

[245] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. 2016. A survey of transfer learning. *Journal of Big Data* 1, 3 (2016), 1–40.

[246] RS Woodworth and EL Thorndike. 1901. The influence of improvement in one mental function upon the efficiency of other functions.(I). *Psychological Review* 8, 3 (1901), 247.

[247] Yue Wu and Qiang Ji. 2016. Constrained Deep Transfer Feature Learning and Its Applications. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.

[248] Yongqin Xian, Zeynep Akata, Gaurav Sharma, Quynh Nguyen, Matthias Hein, and Bernt Schiele. 2016. Latent embeddings for zero-shot classification. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 69–77.

[249] Yongqin Xian, Tobias Lorenz, Bernt Schiele, and Zeynep Akata. 2018. Feature Generating Networks for Zero-Shot Learning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.

[250] Yongqin Xian, Bernt Schiele, and Zeynep Akata. 2017. Zero-shot learning-The Good, the Bad and the Ugly. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 3077–3086.

[251] Min Xiao and Yuhong Guo. 2015. Feature space independent semi-supervised domain adaptation via kernel matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 1 (2015), 54–66.

[252] Saining Xie and Zhuowen Tu. 2015. Holistically-nested edge detection. In *Proc. IEEE international conference on computer vision*. 1395–1403.

[253] Hongyu Xu, Jingjing Zheng, and Rama Chellappa. 2015. Bridging the Domain Shift by Domain Adaptive Dictionary Learning. In *Proc. British Machine Vision Conference*.

[254] Jiaolong Xu, Sebastian Ramos, David Vazquez, and Antonio M Lopez. 2014. Domain adaptation of deformable part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 12 (2014), 2367–2380.

[255] Jiaolong Xu, Sebastian Ramos, David Vázquez, and Antonio M López. 2014. Incremental Domain Adaptation of Deformable Part-based Models.. In *Proc. British Machine Vision Conference*.

[256] J. Xu, D. Vï£¡zquez, K. Mikolajczyk, and A. M. Lï£¡pez. 2016. Hierarchical online domain adaptation of deformable part-based models. In *2016 IEEE International Conference on Robotics and Automation*. 5536–5541.

[257] Ruijia Xu, Ziliang Chen, Wangmeng Zuo, Junjie Yan, and Liang Lin. 2018. Deep Cocktail Network: Multi-source Unsupervised Domain Adaptation with Category Shift. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

[258] Xun Xu, Timothy Hospedales, and Shaogang Gong. 2017. Transductive zero-shot action recognition by word-vector embedding. *International Journal of Computer Vision* 123, 3 (2017), 309–333.

[259] Xing Xu, Fumin Shen, Yang Yang, Dongxiang Zhang, Heng Tao Shen, and Jingkuan Song. 2017. Matrix Tri-Factorization with Manifold Regularizations for Zero-shot Learning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

[260] Zheng Xu, Wen Li, Li Niu, and Dong Xu. 2014. Exploiting low-rank structure from latent domains for domain generalization. In *Proc. European Conference on Computer Vision*. Springer, 628–643.

[261] Makoto Yamada, Leonid Sigal, and Yi Chang. 2014. Domain Adaptation for Structured Regression. *International Journal of Computer Vision* 109, 1-2 (2014), 126–145.

[262] Fei Yan and Krystian Mikolajczyk. 2015. Deep correlation for matching images and text. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3441–3450.

[263] Hongliang Yan, Yukang Ding, Peihua Li, Qilong Wang, Yong Xu, and Wangmeng Zuo. 2017. Mind the Class Weight Bias: Weighted Maximum Mean Discrepancy for Unsupervised Domain Adaptation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

[264] Jun Yang, Rong Yan, and Alexander G Hauptmann. 2007. Cross-domain video concept detection using adaptive svms. In *Proc. ACM International Conference on Multimedia*. ACM, 188–197.

[265] Liu Yang, Liping Jing, Jian Yu, and Michael K Ng. 2016. Learning transferred weights from co-occurrence data for heterogeneous transfer learning. *IEEE transactions on neural networks and learning systems* 27, 11 (2016), 2187–2200.

[266] Ting Yao, Yingwei Pan, Chong-Wah Ngo, Houqiang Li, and Tao Mei. 2015. Semi-supervised Domain Adaptation with Subspace Learning for Visual Recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2142–2150.

[267] Meng Ye and Yuhong Guo. 2017. Zero-Shot Classification with Discriminative Semantic Representation Learning. *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2017).

[268] Zili Yi, Hao Zhang, Ping Tan Gong, et al. 2017. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation. In *Proc. IEEE International Conference on Computer Vision*.

[269] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks?. In *Advances in neural information processing systems*. 3320–3328.

[270] Werner Zellinger, Thomas Grubinger, Edwin Lughofer, Thomas Natschläger, and Susanne Saminger-Platz. 2017. Central moment discrepancy (CMD) for domain-invariant representation learning. In *Proc. International Conference on Learning Representations*.

[271] Deming Zhai, Bo Li, Hong Chang, Shiguang Shan, Xilin Chen, and Wen Gao. 2010. Manifold Alignment via Corresponding Projections. In *Proc. British Machine Vision Conference*. BMVA Press, 3.1–3.11.

[272] Jing Zhang, Zewei Ding, Wanqing Li, and Philip Ogunbona. 2018. Importance Weighted Adversarial Nets for Partial Domain Adaptation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.

[273] Jing Zhang, Wanqing Li, and Philip Ogunbona. 2017. Joint Geometrical and Statistical Alignment for Visual Domain Adaptation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

[274] Jing Zhang, Wanqing Li, and Philip Ogunbona. 2018. Unsupervised Domain Adaptation: A Multi-task Learning-based Method. *arXiv preprint arXiv:1803.09208* (2018).

[275] Jing Zhang, Wanqing Li, Pichao Wang, Philip Ogunbona, Song Liu, and Chang Tang. 2016. A Large Scale RGB-D Dataset for Action Recognition. In *Proc. International Workshop on Understanding Human Activities through 3D Sensors (UHA3DSâĂŹ16) in conjunction with International Conference on Pattern Recognition*.

[276] Kun Zhang, Krikamol Muandet, Zhikun Wang, et al. 2013. Domain adaptation under target and conditional shift. In *Proc. International Conference on Machine Learning*. 819–827.

[277] Li Zhang, Tao Xiang, and Shaogang Gong. 2017. Learning a Deep Embedding Model for Zero-Shot Learning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

[278] Richard Zhang, Phillip Isola, and Alexei A Efros. 2016. Colorful image colorization. In *European Conference on Computer Vision*.

[279] Yu Zhang and Dit-Yan Yeung. 2010. Transfer metric learning by learning task relationships. In *Proc. ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1199–1208.

[280] Ziming Zhang and Venkatesh Saligrama. 2015. Zero-shot learning via semantic similarity embedding. In *Proc. IEEE International Conference on Computer Vision*. IEEE, 4166–4174.

[281] Ziming Zhang and Venkatesh Saligrama. 2016. Zero-shot learning via joint latent similarity embedding. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 6034–6042.

[282] Ziming Zhang and Venkatesh Saligrama. 2016. Zero-shot recognition via structured prediction. In *European Conference on Computer Vision*. Springer, 533–548.

[283] Zhong Zhang, Chunheng Wang, Baihua Xiao, Wen Zhou, Shuang Liu, and Cunzhao Shi. 2013. Cross-view action recognition via a continuous virtual path. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2690–2697.

[284] Peilin Zhao and Steven C Hoi. 2010. OTL: A framework of online transfer learning. In *Proc. International Conference on Machine Learning*. 1231–1238.

[285] Jingjing Zheng, Zhuolin Jiang, P Jonathon Phillips, and Rama Chellappa. 2012. Cross-View Action Recognition via a Transferable Dictionary Pair. In *Proc. British Machine Vision Conference*, Vol. 1. 1–11.

[286] W-S Zheng. 2009. Associating Groups of People. *Proc. British Machine Vision Conference*.

[287] Joey Tianyi Zhou, Ivor W Tsang, Sinno Jialin Pan, and Mingkui Tan. 2014. Heterogeneous Domain Adaptation for Multiple Classes. In *Proc. International Conference on Artificial Intelligence and Statistics*. 1095–1103.

[288] Fan Zhu and Ling Shao. 2014. Weakly-supervised cross-domain dictionary learning for visual recognition. *International Journal of Computer Vision* 109, 1-2 (2014), 42–59.

[289] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Proc. IEEE International Conference on Computer Vision*.

[290] Y. Zhu, W. Chen, and G. Guo. 2014. Evaluating spatiotemporal interest point features for depth-based action recognition. *Image and Vision Computing* 32, 8 (2014), 453–464.

[291] Yin Zhu, Yuqiang Chen, Zhongqi Lu, Sinnojialin Pan, Guirong Xue, Yong Yu, and Qiang Yang. 2011. Heterogeneous Transfer Learning for Image Classification. In *Proc. AAAI Conference on Artificial Intelligence*.

[292] Yizhe Zhu, Mohamed Elhoseiny, Bingchen Liu, and Ahmed Elgammal. 2018. Imagine it for me: Generative Adversarial Approach for Zero-Shot Learning from Noisy Texts. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.