A Survey on Unsupervised Domain Adaptation

Pei Wang[1,*]

[1]Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China
Corresponding author: Pei Wang (e-mail: peiwang@kust.edu.cn).

DOI:https://doi.org/10.63619/ijai4s.v1i2.002

Manuscript received March 4, 2025; revised March 30, 2025; published April 13, 2025.

Abstract: The breakthrough progress of deep learning in fields such as computer vision often relies on the support of massive labeled data. However, data annotation is not only time-consuming, but also costly and challenging task. Unsupervised Domain Adaptation (UDA), as a key technical in the field of transfer learning, provides a new research paradigm for solving cross domain generalization problems by constructing a knowledge transfer bridge between source and target domain. Although significant progress has been made in this technology, existing review studies still have shortcomings in terms of systematical and timeliness. To fill this gap, this paper conducts systematic research from three aspects: methodology, dataset, and application practice. Firstly, we conduct a comprehensive and systematic investigation of the existing UDA methods and provide a unified taxonomy framework. Secondly, we systematically reviewed three benchmark datasets and introduced the innovative applications of this technology in cutting-edge fields such as computer vision. Finally, based on the analysis of existing work, we provide new perspectives and technical paths for future research directions in UDA.

Keywords: Transfer Learning, Unsupervised Domain Adaptation, Application, Computer Vision.

1. Introduction

In recent years, machine learning, especially deep learning, particularly deep learning, has demonstrated remarkable success across various domains including computer vision[1], [2], smart healthcare [3], and remote sensing[4], [5], [6], [7]. However, these methods mainly rely on large-scale labeled datasets, where manual annotation processes require a lot of time and cost. A simple method is to train on large-scale data and then test on small-scale data. This strategy often encounters significant performance degradation due to the distribution discrepancy between training data (source domain) and testing data (target domain).

To address this challenge, transfer learning has emerged as an effective paradigm [8], [9]. This methodology enables knowledge transfer from a resource-rich source domain to a distinct but related target domain, allowing models to leverage previously acquired information for new tasks. As illustrated in Fig. 1, transfer learning mimics human analogical reasoning capabilities by adapting existing knowledge to novel situations. For instance, skills developed in bicycle riding can facilitate learning to operate motorcycles through shared balance and coordination mechanisms, while remaining largely inapplicable to automobile driving due to fundamental operational differences.

Unsupervised Domain Adaptation (UDA) [8] is a special type of transfer learning problem, where models leverage labeled data from a source domain and unlabeled data from a target domain. By addressing the domain shift, UDA enables model adaptation to target distributions without supervised information from the target domain, thereby effectively addressing label scarcity issues in real-world applications [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23]. Despite significant advances in UDA

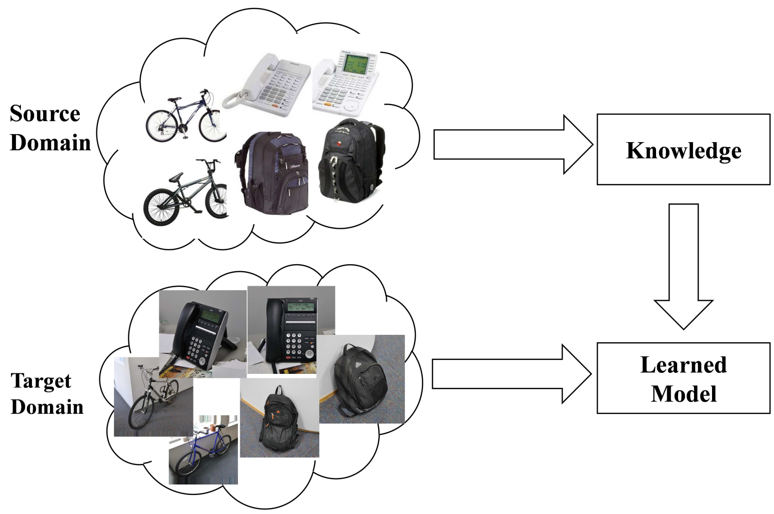

Fig. 1. Schematics of transfer learning.

methodologies, there is currently a lack of a comprehensive and timely review in this field.

To fill this gap, our goal is to timely and comprehensively explore the latest developments in unsupervised domain adaptation. This paper aims to elaborate on the basic concepts of UDA and summarize the latest research results in this field, with a particular emphasis on the innovative elements in different UDA methods, in order to provide a comprehensive understanding of this field and inspire the design and practical application of more UDA methods. The main contributions of this paper can be summarized as follows:

1) We have timely and comprehensively summarized the latest UDA methods and provided a taxonomy for UDA methods, filling the gap in existing literature.

2) We introduced three commonly used benchmark datasets (i.e., Office-31, Office-Home, and VisDA-2017) and provided future research directions.

2. Overview

2.1. Problem Description

The two basic concepts of UDA are domain and task [9]. The domain is represented by data $\mathcal{X}$ and the marginal distribution $P(\mathbf{x})$ that generates the data. Given the domain, task $\mathcal{T}$ consists of label space $\mathcal{Y}$ and the ground-truth function $f(\cdot)$.

Definition 1 (Unsupervised Domain Adaptation, UDA). *Given a source domain dataset $\mathcal{D}_s = \{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^{n_s}$ with n_s labeled samples and the target domain dataset $\mathcal{D}_t = \{(\mathbf{x}_i^t)_{i=1}^{n_t}$ with n_t unlabeled samples, where $y_i^s \in \mathbb{R}^K$ represents the label of sample, K is the number of categories. The key assumption is that the data feature space $\mathcal{X}_s$ and $\mathcal{X}_t$ of the source and target domains be same, while their label space $\mathcal{Y}_s$ and $\mathcal{Y}_t$ also remains consistent. The goal of UDA is to learn the mapping function $f(x)$ by eliminating the discrepancy of joint distributions, so that the learned function $f(x)$ can be well generalized to the target domain, thereby achieving effective knowledge transfer and reuse.*

Formally, the function $f(\mathbf{x}) = C\left(G\left(\mathbf{x}\right)\right)$ contains a classifier C and a feature extractor G, where the feature extractor learns the features of the data $\mathbf{z} = G(\mathbf{x}) \in \mathbb{R}^d$, where d represents the feature dimension, and the classifier learns the predicted output $\mathbf{p} = C(\mathbf{z}) \in \mathbb{R}^K$.

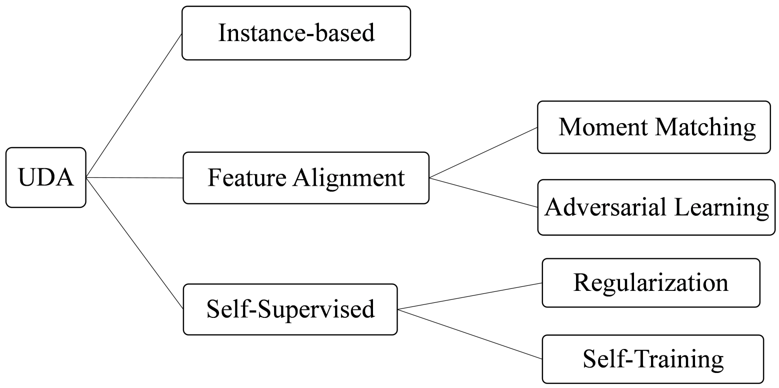

Fig. 2. Taxonomy of UDA methods.

2.2. Taxonomy

UDA is a fundamental method for addressing the scarcity of labeled data, assuming that there is no labeled data in the target domain. This paper provides a comprehensive and systematic review of the relevant work on UDA. By reviewing and evaluating existing work, we hope to provide valuable references and insights for the development of UDA. According to the domain adaptation strategy, existing UDA methods can be roughly divided into the following three categories: instance-based methods [24], [25], [26], [27], feature alignment-based methods [12], [13], [14], [28], [15], [16], [17], and self-supervised methods [29], [30], [31], [32], [33], as shown in Fig. 2.

3. Instance-based Methods

The key idea of the instance-based methods is to assign weights to source domain samples by calculating their similarity to the target domain, and use this to weight the loss function, thereby reducing distribution discrepancy across domains. Most existing research has focused on how to accurately estimate the probability density values between domains [25], [26], [34]. Kernel Mean Matching (KMM) [25] is a well-known method that calculates sample weights by minimizing the Maximum Mean Discrepancy (MMD) [35] between weighted source domain data and target domain data. Another representative work is the KL importance estimation process [27], which uses relative entropy to measure the similarity between source domain samples and target domain, thereby determining the importance of each sample without the need for complex density estimation.

In addition to density estimation, some studies use the predicted values of domain classifiers to evaluate the alignment degree of samples. Tang et al.[36] proposed measuring the similarity between source domain samples and corresponding target clusters (class centers) based on the distance between the two, and assigning weights to different source domain samples accordingly. In the target offset scenario, Zhang et al. [37] used kernel mean matching to estimate the label density ratio and provided the error bound of this method. In addition, inspired by the AdaBoost algorithm, Dai et al. [24] proposed the sample transfer learning method TraAdaBoost. The core idea behind it is to reduce the weight of misclassified source domain samples, as these samples often have significant differences from the target domain.

Instance-based methods require a certain degree of similarity in the distribution between domains. When the distribution discrepancy across domains is too large, the performance of such methods will significantly decrease, which limits their application in practical tasks.

4. Feature Alignment Methods

4.1. Moment Matching Methods

Moment matching method reduces domain shift by matching the high-order statistical moments of source domain and target domain features [13], [14], [17]. This paradigm can learn domain-invariant features to achieve knowledge transfer across domains.

Moment matching methods usually use the maximum mean discrepancy (MMD) [13] that is a non parametric measure, which is defined as:

$$M(P_s, P_t) = \left\| \mathbb{E}_{P_s}[\phi(x^s)] - \mathbb{E}_{P_t}[\phi(x^t)] \right\|_{\mathcal{H}}^2 . \tag{1}$$

The ϕ in the above equation is a nonlinear feature mapping function, $\mathcal{H}$ is the reproducing kernel Hilbert space. Given the kernel function $K(\mathbf{x}^s, \mathbf{x}^t) = \langle \phi(\mathbf{x}^s), \phi(\mathbf{x}^t) \rangle$, where $\langle \cdot, \cdot \rangle$ represents the inner product of two vectors, and $\mathbf{z}$ represents the feature, then the empirical estimate of MMD is redefined as:

$$\begin{aligned} \hat{M}(P_s, P_t) &= \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\mathbf{x}_i^s) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(\mathbf{x}_j^t) \right\|_{\mathcal{H}}^2 \\ &= \left[\frac{1}{n_s^2} \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} K(\mathbf{z}_i^s, \mathbf{z}_j^s) + \frac{1}{n_t^2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} K(\mathbf{z}_i^t, \mathbf{z}_j^t) - \frac{2}{n_s n_t} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} K(\mathbf{z}_i^s, \mathbf{z}_j^t) \right] . \end{aligned} \tag{2}$$

Transfer Component Analysis (TCA) proposed by Pan et al. [14] is a representative moment matching method that learns a mapping function by minimizing the MMD of source and target domain features. However, a major drawback of MMD is that it requires expensive kernel matrix calculations. For this purpose, center moment discrepancy (CMD) [17] utilizes multiple low order moments to define a new distance function by equivalently representing the probability distribution. But, these methods mainly focus on aligning global feature distributions without considering inter-class relationship.

Some works use local maximum mean discrepancy (LMMD) [38], [39], [40] to align feature distributions of the same category (sub-domain) across domains, effectively improving the generalization performance of the learned model. For example, CMMD [39] and LMMD [40] align class-level feature distributions by capturing fine-grained information for each category. In addition, some improved methods based on MMD, such as conditional MMD [41] and joint MMD [42], have been used to measure the distribution discrepancy between domains in the Hilbert space. These methods further improve the performance of the model by estimating the label weight ratio and reweighting the samples.

By integrating deep learning and domain adaptation [43], [44], [45], some works have achieved remarkable performance improvements. For example, DAN [13] and DSAN [40] quantify the distribution discrepancy across domains through MMD and LMMD, respectively. In addition, some methods directly utilize pseudo-labeled features for class-level alignment to improve feature discriminability [45], [39], [40]. For example, TPN [46] uses prototypes (i.e., feature centers for each category) to guide feature alignment. Dynamic Weighted Learning (DWL) [47] can dynamically adjust the proportion of transferability and discriminability of data in the target domain. Xin et al. [48] proposed an end-to-end collaborative alignment framework (CAF) to capture global structural information and local semantic consistency. Furthermore, the researchers proposed weighted MMD [39] and generalized label shift (GLS) [37] to reduce the inconsistency of the label distribution.

To achieve semantic alignment between classes, moving semantic transfer network (MSTN) [44] learns semantic features by aligning the class centers of the source and target domains. In addition, CAN [49] proposes contrastive domain discrepancy, which explicitly models intra- and inter-class discrepancy across domains. However, CAN relies on alternative optimization and class-aware sampling, which greatly increase computational costs. Wang et al. [50] rethink the principle of MMD and proposed a discriminative MMD method that applies trade-off parameters to the intra-class distance hidden in MMD or recalculates the inter-class distance using weights similar to those hidden in MMD. Recently, Wang et al. [51] improved

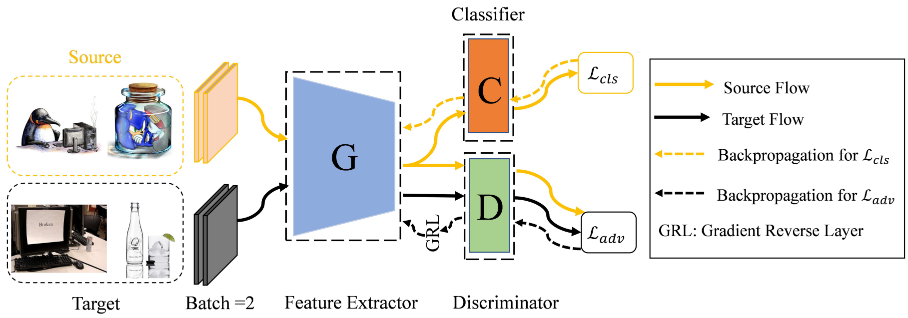

Fig. 3. The architecture of DANN

discriminative ability of the learned features by maximizing the mutual information between features and outputs.

The above methods rely on a specific kernel space to match the high-order statistical moments of the learned features, and cannot characterize any complex probability distribution. In addition, the adaptation of conditional probability distributions based on pseudo labels is susceptible to the influence of noisy pseudo labels, thereby aligning samples to incorrect classes.

4.2. Adversarial Learning Methods

Adversarial learning methods aim to learn domain-invariant features through a min-max two player game between the domain discriminator and feature extractors [10], [11], [12], [15], [16], where the domain discriminator is trained to distinguish whether the input comes from source domain or target domain, while the generator attempts to fool the discriminator. As a pioneering work, DANN [12] proposed a domain adversarial neural network consisting of a domain discriminator D and a feature extractor, as shown in Fig. 3. In addition, the classification loss in source domain should also be minimized [12]. The final objective function is defined as:

$$\begin{aligned} \min_{G,C}\mathcal{L}_{cls} &= \mathbb{E}_{(\mathbf{x}_i^s,\mathbf{y}_i^s)\sim\mathcal{D}_s}\mathcal{L}_{ce}\left(C\left(G\left(\mathbf{x}_i^s\right),\mathbf{y}_i^s\right)\right) \\ \min_G\max_D\mathcal{L}_{adv} &= \mathbb{E}_{\mathbf{x}_i^s\sim\mathcal{D}_s}\log\left[D\left(G\left(\mathbf{x}_i^s\right)\right)\right] + \mathbb{E}_{\mathbf{x}_i^t\sim\mathcal{D}_t}\log\left[1-D\left(G\left(\mathbf{x}_i^t\right)\right)\right], \end{aligned} \tag{3}$$

where, $\mathcal{L}_{ce}\left(\cdot,\cdot\right)$ is the cross-entropy loss function.

CDAN [52] introduces a conditional discriminator to align domain features, which not only considers domain information, but also utilizes discriminative information from prediction output to model the relationship between features and predicted information. Wei et al. [53] proposed MetaAlign, which encourages gradient consistency between feature extractors and discriminators. In addition, GATE [54] uses the similarity between samples to align global and local subgraphs. However, these methods do not explicitly align class-level feature distributions, which may limit their ability to learn discriminative features. To address this issue, ADDA [16] and MADA [28] extend this structure to multiple feature extractors or discriminators to capture multimodal structures and achieve finer grained feature alignment. Similarly, the GVB-GD [55] uses multiple discriminators to learn class-level domain-invariant features. Although such methods can improve the accuracy of feature alignment, multiple feature extractors and discriminators also bring additional computational complexity, making the optimization process more difficult. Recently, Wang et al. [56] proposed class-aware prototypical adversarial network, which uses a single multi-class discriminator (i.e. multi-class classifier) to replace the traditional domain discriminator.

Another adversarial learning methods use the difference between two classifiers as a domain discriminator [57]. Specifically, maximum classifier discrepancy (MCD) [58] quantifies intra-class differences by minimizing the distance between two different classifiers, and learns domain-invariant features through the min-max this discrepancy. To capture intra-class variations, SWD [59] introduced slice Wasserstein distance. However, these methods often overlook the certainty of predictions, which may have a negative impact on distribution consistency. To address this issue, BCDM [60] utilizes classifier determinacy disparity to generate more discriminative features. Although these methods can effectively reduce domain shift, most methods only focus on intra-class differences between predictions, resulting in ambiguous prediction.

In addition, MDD [61] proposed margin disparity discrepancy with a good theoretical support. Based on MDD, i-MDD [62] further introduces task-driven contrastive domain discrepancy. Zhang et al. [57] proposed multi-class scoring disagreement (MCSD) divergence, which can characterize the relations between any pair of multi-class scoring hypotheses. Other works parameterize and integrate a classifier and discriminator into an integrated classifier, achieving joint distributions alignment [63], [15], [64]. For example, Tang et al. [15] proposed discriminative adversarial domain adaptation (DADA), which aligns the joint distribution of source and target domains by jointly parameterizing domain discriminator and classifier. On the one hand, this type of method does not fully utilize the predicted discriminative information, and on the other hand, it requires a complex optimization process, which hinders the learning of discriminative features.To eliminate the influence of semantic irrelevant features, SCDA [65] learns semantic features by min-max the prediction differences of same class samples.

However, the above methods are difficult to handle scenarios where the support sets of two distributions do not overlap completely. Discriminator-free adversarial learning networks (DALN) [11] combines classifiers with nuclear-norm discrepancy directly as domain discriminator to align class-level features by using predicted discriminative information. However, when the batch-size is small or the number of categories is large, this increases the difficulty of calculating the nuclear-norm and hinders domain adaptation. Recently, multi-batch nuclear-norm discrepancy [66] has utilized cache features to eliminate the dependency between nuclear-norm computation and batch-size.

5. Self-Supervised Methods

5.1. Regularization Methods

Some methods aim to further explore the potential of unlabeled data to improve the generalization ability of adaptation models [29], [67], [30], [68], [36], [32]. For example, Long et al. [69] proposed nearest neighbor structure regularization to construct semantic features across domains, which alleviates negative transfer to some extent. EntMin [70] is used to obtain deterministic predictions of target domain samples. Chen et al. [71] proposed maximum squared loss to reduce the impact of easily transferable samples in EntMin on model performance. In addition, MCC [30] solves various domain adaptation scenarios by minimizing the confusion loss of target classification prediction. Further, CC-Loss [72] introduces consistency constraints with different data augmentation, improving the robustness of the confusion matrix to distribution perturbations. Self-ensemble [73] rely on ensemble learning and data augmentation to enhance the generalization ability of the learned model.

Recently, some works have explored the transferability, discriminability, and diversity of the learnedfeatures from the perspective of matrix analysis [74], [29], [67]. For example, BNM [67] utilizes the batch nuclear-norm of the output matrix to improve the discriminability and diversity of prediction outputs. AFN [75] enhances features transferability by increasing feature norm, while BSP [29] balances transferability and discriminability by penalizing the maximum eigenvalue of the feature matrix. For safe transfer learning, Chen et al. [74] proposed batch spectral shrinkage (BSS), which suppresses non-transferable spectral components by penalizing smaller singular values in the feature matrix. In contrast, SENTRY [76] selectively optimizes the entropy of the target sample based on the consistency of multiple random image transformations, improving the generalization performance. In addition, some works use mutual information maximization [77], [78], [79] as the target domain loss to learn more discriminative features. For example, EMDM [32] approximates the ideal objective function by balancing entropy minimization and diversity maximization. Inspired by energy learning, Herath et al. [80] learned domain-invariant features by minimizing the free energy deviation.

Although regularization methods can improve task performance by utilizing unlabeled target domain data, such methods typically require similar spectral properties or inter-class relationships across domains. When there is significant distribution discrepancy, the above requirements are often difficult to establish.

5.2. Self-Training Methods

Some self-training methods train models by generating high-quality pseudo labels to improve model performance [81], [82], [31], [83]. However, due to domain shift in UDA, the generated pseudo labels often contain noisy, which can have a negative impact on the training performance of the model. To address this issue, Saito et al. [31] proposed an asymmetric triple-training method inspired by collaborative training. By selecting two classifiers to predict consistent samples for self training, they guide the other classifier to learn discriminative features of the target domain, which to some extent eliminates the influence of noisy pseudo labels. In addition, SHOT [78] fully utilizes the inherent structure of the target domain, obtains clean pseudo labels through clustering, and uses these pseudo labels for learning the objective function. Gu et al. [82] designed a robust pseudo label loss function in a spherical feature space, which is based on a gaussian-uniform mixture model to estimate the posterior probability of pseudo label correctness, thereby more accurately evaluating the quality of pseudo labels. Cycle self-training [84] is a generalized pseudo label generation method that first trains a target classifier based on pseudo labels, and then allows the classifier to correctly classify on the source domain to learn shared features. BiMem [85] utilizes a bi-directional memory mechanism to learn and remember useful representative information for correcting noisy pseudo labels.

Self-training methods achieve good classification performance by selecting high confidence pseudo labels for supervised learning on target domain. However, due to domain shift, self-training methods inevitably fall into the problem of error accumulation.

6. Datasets

Office-31 [86] is a domain adaptation standard benchmark dataset that includes three different object recognition domains: Amazon (A) for online e-commerce images, Webcam (W) for low resolution images captured by webcams, and DSLR (D) for high-resolution images captured by DSLRs, with a total of 4,110 images and 31 categories. Six domain adaptation tasks were constructed through random pairing to comprehensively evaluate the adaptation performance in different scenarios.

Office-Home [87] is a more challenging dataset in UDA, with a total of 15,588 images across 65 categories. This dataset contains four different domains: art images in various forms such as sketching and painting (A), clip art (C), product images without background (P), and real-world images captured by regular cameras (R). Similarly, based on these four domains, 12 domain adaptation tasks were designed to comprehensively test the performance of the model in diverse scenarios.

VisDA-2017 [88] is a simulation and real-world dataset consisting of two very different domains in UDA: 2D rendering (Synthetic) of 3D model datasets generated under different angles and lighting conditions, and real natural images collected from MSCOCO. Synthetic and Real serve as the source and target domain for domain adaptation tasks, respectively.

7. Application

UDA has demonstrated significant utility across diverse fields such as computer vision, medical image analysis, and time-series modeling, showcasing its versatility in addressing domain shift challenges.

7.1. Computer Vision

UDA plays an important role in computer vision for tasks including but not limited to cross-domain image classification, object detection, and semantic segmentation. Owing to substantial domain discrepancies in illumination conditions, capture angles, background complexity, and spatial resolutions, visual data distributions frequently exhibit divergent feature distributions and statistical discrepancies. UDA research in vision focuses on establishing domain-invariant feature through distribution alignment, enabling effective transfer of discriminative visual knowledge while mitigating domain shift.

7.2. Medical Image Analysis

Compared with computer vision, medical image faces unique challenges in data acquisition and annotation. Medical data usually involves sensitive information and professional knowledge, requiring strict privacy protection and expert-level labeling, which leads to the scarcity and high cost of labeled data. Therefore, how to effectively utilize existing annotated data for knowledge transfer has become an important research direction in UDA [89], [90], such as pneumonia classification [91], [92] and viral hosts prediction [93].

7.3. Time-series Modeling

Time-series data, such as traffic flow, have continuity and dynamism, often involving complex patterns of change and temporal dependencies. UDA demonstrates unique advantages in time-series analysis by addressing non-stationary distribution shifts inherent in dynamic systems, demonstrating great potential and value [94]. This proves particularly valuable for real-time applications including intelligent transportation system optimization and industrial equipment predictive maintenance, where models must dynamically adjust to temporal distribution shift.

8. Future Works

While UDA has demonstrated remarkable success in computer vision and medical image analysis, several fundamental challenges require further exploration and research.

8.1. Generalization Error Bound

Previous work has analyzed the generalization error bound for UDA, which provides new ideas and inspirations for algorithm design. However, the upper bound of generalization error in source-free and open-set domain adaptation still needs further exploration. In addition, the lower bound of generalization error for UDA has not received the attention it deserves. Such analysis would not only quantify the theoretical limits of domain transferability but also reveal the inherent complexity of cross-domain learning through measurable task divergence metrics.

8.2. Diffusion-based Domain Adaptation

The underlying mechanism of diffusion models establishes theoretically-grounded transformations between noise distributions and complex data manifolds through iterative refinement processes. This paradigm is similar to the goal of UDA, both aimed at reducing distribution discrepancy across domains. However, how to apply diffusion models for UDA still needs to be explored.

8.3. Complex Domain Adaptation Scenarios

In practical applications, due to limitations in data privacy protection, source domain data may not be directly accessible, which increases the difficulty of domain adaptation. In addition, the complexity of the real world is also manifested in multiple source and target domains, changes in data categories, limited computing resources, and the demand for online learning.

9. Conclusions

Unsupervised domain adaptation (UDA), as a major research direction in transfer learning, has received increasing attention in recent years. UDA transfers knowledge from labeled source domain to unlabeled target domain, effectively alleviating the dependence that deep learning to labeled data. This paper provides a comprehensive analysis of current UDA methods and proposes a unified taxonomy framework. Then, we provided a detailed introduction to three commonly used benchmark datasets and future research directions in UDA. We believe that this study has the potential to provide valuable inspiration and reference for the development of UDA fields.

Acknowledgements

The authors wish to thank the anonymous reviewers for their valuable suggestions.

References

[1] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[2] K. He, X. Zhang *et al.*, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 770–778.

[3] Y. Yang, Y. Hu, X. Zhang, and S. Wang, "Two-stage selective ensemble of cnn via deep tree training for medical image classification," *IEEE Transactions on Cybernetics*, vol. 52, no. 9, pp. 9194–9207, 2022.

[4] Z. Yu and J. Wang, "Estimating forest carbon stocks from high-resolution remote sensing imagery by reducing domain shift with style transfer," *arXiv preprint arXiv:2502.00784*, 2025.

[5] Z. Yu, J. Wang, H. Chen, and M. Y. I. Idris, "Qrs-trs: Style transfer-based image-to-image translation for carbon stock estimation in quantitative remote sensing," *IEEE Access*, vol. 13, pp. 52 726–52 737, 2025.

[6] Z. Yu, H. Wang, and H. Chen, "A guideline of u-net-based framework for precipitation estimates," *International Journal of Artificial Intelligence for Science (IJAI4S)*, vol. 1, no. 1, 2025.

[7] Y. Luo, J. Wang, X. Yang, Z. Yu, and Z. Tan, "Pixel representation augmented through cross-attention for high-resolution remote sensing imagery segmentation," *Remote Sensing*, vol. 14, no. 21, p. 5415, 2022.

[8] S. Zhao, X. Yue *et al.*, "A review of single-source deep unsupervised visual domain adaptation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 473–493, 2020.

[9] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[10] Z. Cao, K. You, Z. Zhang, J. Wang, and M. Long, "From big to small: Adaptive learning to partial-set domains," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1766–1780, 2023. [Online]. Available: https://doi.org/10.1109/TPAMI.2022.3159831

[11] L. Chen, H. Chen *et al.*, "Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 7181–7190.

[12] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. S. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[13] M. Long, Y. Cao, Z. Cao, J. Wang, and M. I. Jordan, "Transferable representation learning with deep adaptation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 12, pp. 3071–3085, 2018.

[14] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 22, no. 2, pp. 199–210, 2011.

[15] H. Tang and K. Jia, "Discriminative adversarial domain adaptation," in *Thirty-Second AAAI Conference on Artificial Intelligence*, vol. 34, Conference Proceedings, pp. 5940–5947.

[16] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 2962–2971.

[17] W. Zellinger, T. Grubinger, E. Lughofer, T. Natschlger, and S. SamingerPlatz, "Central moment discrepancy (cmd) for domain-invariant representation learning," in *5th International Conference on Learning Representations*, Conference Proceedings.

[18] Z. Yu, "Ai for science: A comprehensive review on innovations, challenges, and future directions," *International Journal of Artificial Intelligence for Science (IJAI4S)*, vol. 1, no. 1, 2025.

[19] Z. Yu and C. S. Chan, "Yuan: Yielding unblemished aesthetics through a unified network for visual imperfections removal in generated images," *arXiv preprint arXiv:2501.08505*, 2025.

[20] P. Wang, "Advances in recommendation systems: From traditional approaches to future trends," *International Journal of Artificial Intelligence for Science (IJAI4S)*, vol. 1, no. 1, 2025.

[21] Z. Yu, M. Y. I. Idris, and P. Wang, "Introduction to the international journal of artificial intelligence for science (ijai4s)," *International Journal of Artificial Intelligence for Science (IJAI4S)*, vol. 1, no. 1, 2025.

[22] Z. Yu and P. Wang, "Capan: Class-aware prototypical adversarial networks for unsupervised domain adaptation," in *2024 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2024, pp. 1–6.

[23] P. Wang, Y. Yang, and Z. Yu, "Multi-batch nuclear-norm adversarial network for unsupervised domain adaptation," in *2024 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2024, pp. 1–6.

[24] W. Dai, Q. Yang, G. Xue, and Y. Yu, "Boosting for transfer learning," in *Twenty-Fourth International Conference on Machine Learning*, ser. ACM International Conference Proceeding Series, Z. Ghahramani, Ed., vol. 227, Conference Proceedings, pp. 193–200.

[25] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Schlkopf, "Correcting sample selection bias by unlabeled data," in *Twentieth Advances in Neural Information Processing Systems*, B. Schlkopf, J. C. Platt, and T. Hofmann, Eds. MIT Press, Conference Proceedings, pp. 601–608.

[26] J. Jiang and C. Zhai, "Instance weighting for domain adaptation in nlp," in *45th Annual Meeting of the Association for Computational Linguistics*, J. Carroll, A. van den Bosch, and A. Zaenen, Eds., Conference Proceedings.

[27] M. Sugiyama, S. Nakajima, H. Kashima, P. von Bnau, and M. Kawanabe, "Direct importance estimation with model selection and its application to covariate shift adaptation," in *Twenty-First Advances in Neural Information Processing Systems*, J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, Eds. Curran Associates, Inc., Conference Proceedings, pp. 1433–1440.

[28] Z. Pei, Z. Cao, M. Long, and J. Wang, "Multi-adversarial domain adaptation," in *Thirty-Second AAAI Conference on Artificial Intelligence*, Conference Proceedings, pp. 3934–3941.

[29] X. Chen, S. Wang, M. Long, and J. Wang, "Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation," in *36th International Conference on Machine Learning*, Conference Proceedings, pp. 1081–1090.

[30] Y. Jin, X. Wang, M. Long, and J. Wang, "Minimum class confusion for versatile domain adaptation," in *16th European Conference on Computer Vision*, Conference Proceedings, pp. 464–480.

[31] K. Saito, Y. Ushiku, and T. Harada, "Asymmetric tri-training for unsupervised domain adaptation," in *34th International Conference on Machine Learning*, D. Precup and Y. W. Teh, Eds., Conference Proceedings, pp. 2988–2997.

[32] X. Wu, S. Zhang, Q. Zhou, Z. Yang, C. Zhao, and L. J. Latecki, "Entropy minimization versus diversity maximization for domain adaptation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 6, pp. 2896–2907, 2023.

[33] W. Zellinger, B. A. Moser, T. Grubinger, E. Lughofer, T. Natschlger, and S. SamingerPlatz, "Robust unsupervised domain adaptation for neural networks via moment alignment," *Information Sciences*, vol. 483, pp. 174–191, 2019.

[34] B. Zadrozny, "Learning and evaluating classifiers under sample selection bias," in *Twenty-first International Conference on Machine Learning*, ser. ACM International Conference Proceeding Series, C. E. Brodley, Ed., vol. 69. ACM, Conference Proceedings. [Online]. Available: https://doi.org/10.1145/1015330.1015425

[35] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schlkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *Twentieth Annual Conference on Neural Information Processing Systems*, B. Schlkopf, J. C. Platt, and T. Hofmann, Eds. MIT Press, Conference Proceedings, pp. 513–520.

[36] H. Tang, K. Chen, and K. Jia, "Unsupervised domain adaptation via structurally regularized deep clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 8722–8732.

[37] R. T. des Combes, H. Zhao, Y. Wang, and G. J. Gordon, "Domain adaptation with conditional distribution matching and generalized label shift," in *34th Advances in Neural Information Processing Systems 33*, Conference Proceedings, pp. 19 276–19 289.

[38] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfer learning," in *IEEE International Conference on Data Mining*, Conference Proceedings, pp. 1129–1134.

[39] H. Yan, Z. Li, Q. Wang, P. Li, Y. Xu, and W. Zuo, "Weighted and class-specific maximum mean discrepancy for unsupervised domain adaptation," *IEEE Transactions on Multimedia*, vol. 22, no. 9, pp. 2420–2433, 2020.

[40] Y. Zhu, F. Zhuang, J. Wang *et al.*, "Deep subdomain adaptation network for image classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1713–1722, 2020.

[41] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *IEEE International Conference on Computer Vision,*, Conference Proceedings, pp. 2200–2207.

[42] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *34th International Conference on Machine Learning*, Conference Proceedings, pp. 2208–2217.

[43] J. Wang, Y. Chen, W. Feng, H. Yu, M. Huang, and Q. Yang, "Transfer learning with dynamic distribution adaptation," *ACM Transactions on Intelligent Systems and Technology*, vol. 11, no. 1, pp. 6:1–6:25, 2020.

[44] S. Xie, Z. Zheng, L. Chen, and C. Chen, "Learning semantic representations for unsupervised domain adaptation," in *35th International Conference on Machine Learning*, Conference Proceedings, pp. 5419–5428.

[45] H. Yan, Y. Ding, P. Li, Q. Wang, Y. Xu, and W. Zuo, "Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 945–954.

[46] Y. Pan, T. Yao, Y. Li, Y. Wang, C. Ngo, and T. Mei, "Transferrable prototypical networks for unsupervised domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 2239–2247.

[47] N. Xiao and L. Zhang, "Dynamic weighted learning for unsupervised domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 15 242–15 251.

[48] B. Xie, S. Li, F. Lv, C. H. Liu, G. Wang, and D. Wu, "A collaborative alignment framework of transferable knowledge extraction for unsupervised domain adaptation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 7, pp. 6518–6533, 2023.

[49] G. Kang, L. Jiang, Y. Yang, and A. G. Hauptmann, "Contrastive adaptation network for unsupervised domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 4893–4902.

[50] W. Wang, H. Li, Z. Ding, F. Nie, J. Chen, X. Dong, and Z. Wang, "Rethinking maximum mean discrepancy for visual domain adaptation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 1, pp. 264–277, 2023.

[51] P. Wang, Y. Yang, Y. L. Xia, K. Wang, X. Y. Zhang, and S. Wang, "Information maximizing adaptation network with label distribution priors for unsupervised domain adaptation," *IEEE Transactions on Multimedia*, vol. 25, pp. 6026–6039, 2023. [Online]. Available: ⟨GotoISI⟩://WOS:001098831500028

[52] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Thirty-Second Advances in Neural Information Processing Systems*, Conference Proceedings, pp. 1647–1657.

[53] G. Wei, C. Lan, W. Zeng, and Z. Chen, "Metaalign: Coordinating domain alignment and classification for unsupervised domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 16 643–16 653.

[54] L. Chen, Y. Lou *et al.*, "Geometric anchor correspondence mining with uncertainty modeling for universal domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 16 134–16 143.

[55] S. Cui, S. Wang, J. Zhuo, C. Su, Q. Huang, and Q. Tian, "Gradually vanishing bridge for adversarial domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 12 452–12 461.

[56] Z. Y. Yu and P. Wang, "Capan: Class-aware prototypical adversarial networks for unsupervised domain adaptation," in *IEEE International Conference on Multimedia and Expo*, Conference Proceedings.

[57] Y. Zhang, B. Deng, H. Tang, L. Zhang, and K. Jia, "Unsupervised multi-class domain adaptation: Theory, algorithms, and practice," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 5, pp. 2775–2792, 2020.

[58] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 3723–3732.

[59] C.-Y. Lee, T. Batra, M. H. Baig, and D. Ulbricht, "Sliced wasserstein discrepancy for unsupervised domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 10 285–10 295.

[60] S. Li, F. Lv, B. Xie, C. H. Liu, J. Liang, and C. Qin, "Bi-classifier determinacy maximization for unsupervised domain adaptation," in *Thirty-Fifth AAAI Conference on Artificial Intelligence*, vol. 35, Conference Proceedings, pp. 8455–8464.

[61] Y. Zhang, T. Liu, M. Long, and M. Jordan, "Bridging theory and algorithm for domain adaptation," in *International Conference on Machine Learning02*, Conference Proceedings, pp. 7404–7413.

[62] M. Li, K. Jiang, and X. Zhang, "Implicit task-driven probability discrepancy measure for unsupervised domain adaptation," in *35th Advances in Neural Information Processing Systems*, Conference Proceedings, pp. 25 824–25 838.

[63] V. K. Kurmi and V. P. Namboodiri, “Looking back at labels: A class based domain adaptation technique,” in *IEEE International Joint Conference on Neural Network*, Conference Proceedings, pp. 1–8.

[64] L. Tran, K. Sohn *et al.*, “Gotta adapt ’em all: Joint pixel and feature-level domain adaptation for recognition in the wild,” in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 2672–2681.

[65] S. Li, M. Xie, F. Lv, C. H. Liu, J. Liang, C. Qin, and W. Li, “Semantic concentration for domain adaptation,” in *IEEE International Conference on Computer Vision*, Conference Proceedings, pp. 9082–9091.

[66] P. Wang, Y. Yang, and Z. Y. Yu, “Multi-batch nuclear-norm adversarial network for unsupervised domain adaptation,” in *IEEE International Conference on Multimedia and Expo*, Conference Proceedings.

[67] S. Cui, S. Wang, J. Zhuo, L. Li, Q. Huang, and Q. Tian, “Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations,” in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 3940–3949.

[68] K. Saito, D. Kim, S. Sclaroff, T. Darrell, and K. Saenko, “Semi-supervised domain adaptation via minimax entropy,” in *IEEE International Conference on Computer Vision*, Conference Proceedings, pp. 8049–8057.

[69] M. Long, J. Wang, G. Ding, D. Shen, and Q. Yang, “Transfer learning with graph co-regularization,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 7, pp. 1805–1818, 2014. [Online]. Available: https://doi.org/10.1109/TKDE.2013.97http://doi.ieeecomputersociety.org/10.1109/TKDE.2013.97https://www.wikidata.org/entity/Q59678224

[70] Y. Grandvalet and Y. Bengio, “Semi-supervised learning by entropy minimization,” in *18th Advances in Neural Information Processing Systems*, Conference Proceedings, pp. 529–536.

[71] M. Chen, H. Xue, and D. Cai, “Domain adaptation for semantic segmentation with maximum squares loss,” in *IEEE International Conference on Computer Vision*, Conference Proceedings, pp. 2090–2099.

[72] Y. Jin, Z. Cao, X. Wang, J. Wang, and M. Long, “One fits many: Class confusion loss for versatile domain adaptation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. Early Access, pp. 1–16, 2024.

[73] G. French, M. Mackiewicz, and M. H. Fisher, “Self-ensembling for visual domain adaptation,” in *6th International Conference on Learning Representations*, Conference Proceedings, pp. 1–15.

[74] X. Chen, S. Wang, B. Fu, M. Long, and J. Wang, “Catastrophic forgetting meets negative transfer: Batch spectral shrinkage for safe transfer learning,” in *33th Advances in Neural Information Processing Systems*, Conference Proceedings, pp. 1906–1916. [Online]. Available: https://proceedings.neurips.cc/paper/2019/hash/c6bff625bdb0393992c9d4db0c6bbe45-Abstract.htmlhttp://papers.nips.cc/paper/8466-catastrophic-forgettin

[75] R. Xu, G. Li, J. Yang, and L. Lin, “Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation,” in *International Conference on Computer Vision*, Conference Proceedings, pp. 1426–1435.

[76] V. Prabhu, S. Khare, D. Kartik, and J. Hoffman, “Sentry: Selective entropy optimization via committee consistency for unsupervised domain adaptation,” in *IEEE International Conference on Computer Vision*, Conference Proceedings, pp. 8538–8547.

[77] T. Li, X. Chen, S. Zhang, Z. Dong, and K. Keutzer, “Cross-domain sentiment classification with contrastive learning and mutual information maximization,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Conference Proceedings, pp. 8203–8207.

[78] J. Liang, D. Hu, and J. Feng, “Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation,” in *37th International Conference on Machine Learning*, Conference Proceedings, pp. 6028–6039.

[79] C. Park, J. Lee, J. Yoo, M. Hur, and S. Yoon, “Joint contrastive learning for unsupervised domain adaptation,” *arXiv preprint arXiv:2006.10297*, vol. 2006.10297, 2020.

[80] S. Herath, B. Fernando, E. Abbasnejad, M. Hayat, S. Khadivi, M. Harandi, H. Rezatofighi, and G. Haffari, “Energy-based self-training and normalization for unsupervised domain adaptation,” in *19th IEEE International Conference on Computer Vision*, Conference Proceedings, pp. 11 619–11 628.

[81] B. Chen, J. Jiang, X. Wang, P. Wan, J. Wang, and M. Long, “Debiased self-training for semi-supervised learning,” in *36th Advances in Neural Information Processing Systems*, Conference Proceedings. [Online]. Available: http://papers.nips.cc/paper_files/paper/2022/hash/d10d6b28d74c4f0fcab588feeb6fe7d6-Abstract-Conference.html

[82] X. Gu, J. Sun, and Z. Xu, “Spherical space domain adaptation with robust pseudo-label loss,” in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 9098–9107.

[83] Z. Zheng and Y. Yang, “Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation,” *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1106–1120, 2021.

[84] H. Liu, J. Wang, and M. Long, “Cycle self-training for domain adaptation,” in *35th Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. N. Dauphin, P. Liang, and J. W. Vaughan, Eds., Conference Proceedings, pp. 22 968–22 981.

[85] J. Zhang, J. Huang, X. Jiang, and S. Lu, “Black-box unsupervised domain adaptation with bi-directional atkinson-shiffrin memory,” in *19th International Conference on Computer Vision*, Conference Proceedings, pp. 11 737–11 748.

[86] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, “Adapting visual category models to new domains,” in *11th European Conference on Computer Vision*, Conference Proceedings, pp. 213–226.

[87] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, “Deep hashing network for unsupervised domain adaptation,” in *IEEE Conference on Computer Vision and Pattern Recognition*, Conference Proceedings, pp. 5385–5394.

[88] X. Peng, B. Usman, N. Kaushik, J. Hoffman, D. Wang, and K. Saenko, “Visda: The visual domain adaptation challenge,” *arXiv preprint arXiv:1710.06924*, 2017.

[89] X. Guo, Y. Zhang, J. Wang, and M. Long, “Estimating heterogeneous treatment effects: Mutual information bounds and learning algorithms,” in *International Conference on Machine Learning*, Conference Proceedings, pp. 12 108–12 121. [Online]. Available: https://proceedings.mlr.press/v202/guo23k.html

[90] Y. Yang, X. Li, P. Wang, Y. Xia, and Q. Ye, “Multi-source transfer learning via ensemble approach for initial diagnosis of alzheimers disease,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, pp. 1–10, 2020.

[91] D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M. K. Prasadha, J. Pei, M. Y. L. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V. A. N. Huu, C. Wen, E. D. Zhang, C. L. Zhang, O. Li, X. Wang,

M. A. Singer, X. Sun, J. Xu, A. Tafreshi, M. A. Lewis, H. Xia, and K. Zhang, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0092867418301545

[92] W. Su, F. Wang, Z. Han, and Y. Yin, "Transferable discriminative learning for medical open-set domain adaptation: Application to pneumonia classification," pp. 1185–1192, December 6-8 2022. [Online]. Available: https://doi.org/10.1109/BIBM55620.2022.9995571

[93] Y. Yang, J. Guo, P. Wang, Y. Wang, M. Yu, X. Wang, P. Yang, and L. Sun, "Reservoir hosts prediction for covid-19 by hybrid transfer learning model," *Journal of Biomedical Informatics*, vol. 117, p. 103736, 2021.

[94] J. Dong, H. Wu, H. Zhang, L. Zhang, J. Wang, and M. Long, "Simmtm: A simple pre-training framework for masked time-series modeling," *arXiv preprint*, vol. arXiv:2302.00861, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2302.00861

Biographies

Pei Wang received the B.S. degree in 2014 from the Jianghuai college, Anhui University, Hefei, China, the M.S. degree in 2018 from Yunnan Normal University, Kunming, China, the Ph.D. degree in 2024 from Yunnan University, Kunming, China. He is currently with the Faculty of Information Engineering and Automation at Kunming University of Science and Technology. His current research interests include transfer learning and large-scale data mining.