IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 August 2023, accepted 29 August 2023, date of publication 11 September 2023, date of current version 22 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3313623

RESEARCH ARTICLE

RDAOT: Robust Unsupervised Deep Sub-Domain Adaptation Through Optimal Transport for Image Classification

OBSA GILO[1], JIMSON MATHEW[1], (Senior Member, IEEE), SAMRAT MONDAL[1], (Senior Member, IEEE), AND RAKESH KUMAR SANODIYA[2], (Member, IEEE)

[1]Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna, Bihar 801106, India
[2]Indian Institute of Information Technology Sri City, Chittoor 517646, India

Corresponding author: Obsa Gilo (obsa_1921cs33@iitp.ac.in)

ABSTRACT In traditional machine learning, the training and testing data are assumed to come from the same independent and identical distributions. This assumption, however, does not hold up in real-world applications, as differences between the training and testing data may have different distributions. Domain adaptation has emerged as a solution that enables the transfer of knowledge between domains with distinct distributions. In this paper, we primarily utilize domain adaptation in the context of visual recognition tasks despite its growing application in diverse domains. Earlier studies have mainly aimed at minimizing the differences in global distributions between the domains and failed to capture the local, pertinent features crucial for domain alignment. Furthermore, models struggle to perform well and generalize to target data when outliers or noise exist in the datasets. This work addresses these problems and provides unique strategies for unsupervised domain adaptation using RDAOT (Robust Deep Adaptation via Optimal Transport). To capture local information by utilizing LMMD (Local Maximum Mean Discrepancy) to minimize the divergence of the feature distributions between the domains. We examine label noise robustness in the source domain and ROT (Robust Optimal Transport) loss to preserve robustness in domain adaptation, which lessens the cost of transporting source distributions to the target distributions. The significance of our presented technique was assessed through extensive experiments on six different visual recognition domain adaptation datasets. The results demonstrate that our method outperforms the current state-of-the-art techniques, indicating superior performance. Our approach was evaluated against several baselines, and the results significantly improved average accuracy across various datasets. Specifically, the average accuracy improved from on the OfficeCaltech10 (91.8% to 96.85%), OfficeHome (67.7% to 68.10%), Office31 (88.17% to 88.92%), IMAGECLEF-DA (87.9% to 90.24%), PACS (69.08% to 85.72%), and VisDA-2017 (80.2 % to 89.43%) datasets, respectively.

INDEX TERMS Domain adaptation, noisy labels, optimal transport, sub-domain adaptation.

I. INTRODUCTION

Recently, deep learning models have been bellowing in performance in diverse applications. However, it needs a massive quantity of annotated data for training. Getting labeled data and labeling is too expensive and time-consuming in real-world applications. Also, traditional machine learning assumes that training and testing come from identical distributions (I.I.D). In real-world cases, these assumptions may fail because training and testing data may be drawn from different distributions due to domain shifts. For instance, domain shifts may occur by different factors like background clutter, viewpoint, camera quality, environment, dataset bias, etc. Training the model directly the model in the presence of

The associate editor coordinating the review of this manuscript and approving it for publication was Rongbo Zhu.

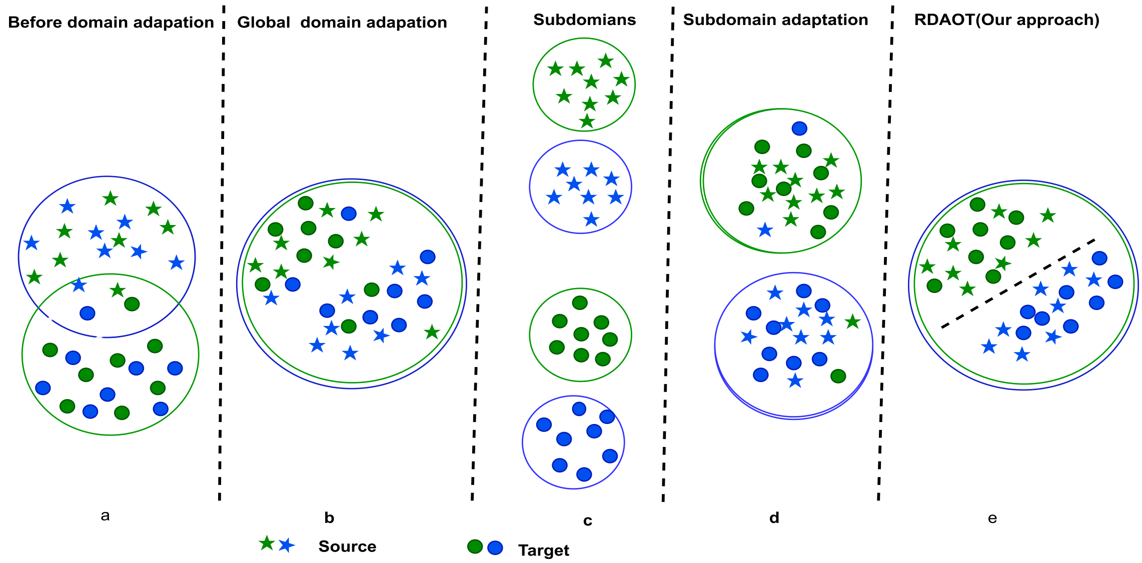


FIGURE 1. Motivational examples of the proposed approach. a) Before domain adaptation (BDA), b) Global domain adaptation (GDA), c) Subdomains, d) Subdomain adaptation (SDA), and e) The proposed model(RDAOT) approach.

domain shift may lead to poor performance [1]. To overcome the above issues lately, domain adaptation approaches [2], [3], [4], [5] drew researchers' awareness because of the capability to handle the absence of annotated data and reduce divergences between the domains, depending on the data on the target label's availability. Domain adaptation divides categories into three groups: supervised, which implies the availability of labeled data; semi-supervised, which occurs when both labeled and unlabeled data are present; and unsupervised, which requires no labeled data.

This paper addresses an unsupervised domain adaptation scenario, where labeled data are abundant in the source domains, and unlabeled data in the target domain [3], [6], [7]. The study delves into the domain of visual recognition tasks, particularly emphasizing this area. The objective is to develop a classifier that can be trained using the labeled data from the source domain and generalize effectively to the target domain. Previous research has introduced numerous techniques for domain adaptation [7], [8], [9], [10], [11], [12], [13]. To reduce the discrepancy between the source and target domains, most domain adaptation strategies concentrate on discovering domain-invariant features [4], [8], [14], [15], [16]. These methods can be divided into three categories: 1) Instance-based [93], [94], [94], gain insights into the significance of labeled data in the source domain through instance selection and weighting. The approach considers that certain source instances might not be relevant to the target domain, even in the shared subspace. To tackle this issue, it aims to minimize the distribution divergence by adjusting the weights of source samples or selecting specific landmark instances. Subsequently, the model learns from those samples that exhibit more remarkable similarity to the target domain samples. 2) Feature-based, which learns domain invariant features to reduce the distribution discrepancy [4], [8], [15]. 3) Adversarial based, which involves training a discriminator (domain classifier) to discriminate between the source and target representations, aims to promote domain confusion [9], [17], [18]. The most existent discrepancy-based unsupervised domain adaptation (UDA) is Maximum Mean Discrepancy (MMD) [4], which metric minimizes variations between two example trials by computing instance mean discrepancies and Correlation Alignment (CORAL) [14], [19], finds a correlation between domains by utilizing second-order (Covariance) statistics of various distributions are intended to match one another. In contrast to JAN [21], which utilizes JMMD loss to minimize discrepancy across the domains, ADDA [20] recognizes domain loss by computing discriminator and feature extractor in GANs model. HOMM [16] to capture relevant information more accurately, extending to higher orders is necessary. This expanded approach allows for a more precise representation of the underlying data. LMMD can catch inherent properties specific to related class [22], [23]. Most of the strategies discussed above are founded on transferable features, and GANs [24] investigated increasing transferable by considering the discriminant features [17]. The discriminability of features to alleviate conditional distributions between domains has also been explored in the context of curriculum Pseudo Labelling and CDANs [26]. Optimal Transport (OT), also known as Wasserstein distance or Earth Mover's Distance (EMD), is a mathematical framework that has gained significant attention in various fields, including machine learning, computer vision, and statistics. It was originally introduced in economics and operations research in the 18th century by the French mathematician Gaspard Monge [27]. In domain adaptation and machine learning, OT measures the similarity between probability distributions representing the source and target domains. It provides a principled way to quantify the discrepancy or distance between these distributions, allowing us to align the domains and transfer knowledge effectively [27], [28], [68]. One of the key advantages of OT is its ability to preserve the underlying structure of the data. Unlike traditional distance metrics, such as Euclidean distance or Kullback-Leibler (KL) divergence, OT considers the spatial arrangement of the data points, making it more suitable for datasets with complex geometric structures. DeepJDOT [29] introduced deep joint optimum transport techniques that evaluate domain consistent attributes and maintain handle label information to minimize the cost of transferring between domains. The Wasserstein distance measure is used by WDGRL [30] motivated by GAN [31] in the optimal transport concept. Training DNNs to achieve high accuracy has become crucial in deep learning, given the prevalence of noisy labels in large-scale datasets. Labeling such datasets is a costly and error-prone process, leading to the inclusion of incorrect labels even in high-quality datasets. Consequently, effectively training DNNs in the presence of noisy labels has gained significant practical importance in the field [32]. In domain adaptation tasks, noisy source labels can significantly hinder the model's performance and ability to generalize well to target domains. While cross-entropy loss is widely used for minimizing risk during the training of deep neural networks (DNNs), it has limitations when handling noisy labels in datasets. One of the challenges is that cross-entropy loss tends to make models over-fit easy samples and under-fit hard samples [32], [33]. In the OT, Wasserstein distance is mainly used for minimizing transport costs between domains. However, OT is sensitive to the outlier for the examples, which with large noise is critical problems in domain adaptation due to unfavorable transfer performance of our model degradable [34]. This will happen because every example is given equal weight due to the marginal constraints in transportation plans. The following is a summary of our significant contributions:

1) We introduce the RDAOT strategy, which aims to maintain robustness and match feature distribution across domains. We use SCE (Symmetrical Cross Entropy) loss to handle noisy labels at source domains and Robust Optimal Transport (ROT) to handle outliers that are robust to noise.
2) In addition, our method takes into account the local distributions of both domains, adjusting subdomains across different subcategories of the same class using the Local Maximum Mean Discrepancy (LMMD). When data from different contexts has variations in the same class, obtaining important features for alignment becomes easier.
3) To show the importance of our technique and how it exceeds other domain adaptation strategies, we took out comprehensive experimental and ablation procedures.

The remainder of the document is organized as follows: Section II discussed the most pertinent related works, and Section III provided a brief explanation of the research's goals. Section IV further describes the suggested framework and its elements. Comparison of our approach novelty with different approaches discussed in Section VI. Section V details experimental analyses, datasets, and results. The conclusion and directions for future research are addressed in VII.

II. RELATED WORKS

A common transfer learning strategy is domain adaptation [3]. When working with data from various distributions, domain shift in machine learning can be a challenging problem. Unsupervised domain adaptation (UDA) is a technique that utilizes labeled training data and unlabeled testing data to address this challenge. UDA is beneficial when domains are selected from several distributions due to the environment, camera, background, viewpoint, illumination, image quality, and more.

A. UNSUPERVISED DOMAIN ADAPTATION

This technique has been suggested in previous works such as [35], [36], [91], and [92]. There are three categories that can be used to group previous domain adaptation strategies. Instance-based methods fall under the first category. These methods assign different weights to source samples based on how well they match the distribution of the target domain. The classifier is then trained on the weighted samples. This approach has been put forth in several assignments, including those [37], [38], [39]. Feature-based methods, popular and effective deep-learning techniques for feature representation, fall under the second category of domain adaptation techniques. By mapping them to a shared latent space with a similar feature distribution, these approaches seek to identify domain-invariant features from both domains [1], [2], [4], [10], [14], [16], [28], [40], [41], [42]. Generative adversarial-based techniques, or GANs, are the third type based on the work of GANs [31]. By using a mini-max game to learn feature representations, these techniques combine two networks generator and discriminator [7], [17], [20], [43] for works that describe major domain adaption advancements made utilizing GAN-based techniques. These techniques employ domain confusion loss, which is simulated as a mini-max game. Our study centers on unsupervised domain adaptation scenarios, primarily employing feature-based techniques. We emphasize the significance of preserving robustness in domain adaptation by extracting local information features from subdomains and effectively handling label noise present in the labeled source domains, as it significantly impacts model performance. Furthermore, we investigate the application of optimal transport in domain adaptation to capture the underlying geometric structures of the data. We extend this approach to incorporate robust optimal transport, effectively addressing outliers and noise in the data during knowledge transfer between domains using optimal transport plans.

B. DOMAIN ADAPTATION DISSIMILARITY MEASURES

Discrepancy measures are a key component of domain adaptation, as they help to reduce the variations between the source and target domain distributions. There are many different discrepancy metrics that have been explored, but we will focus on the most related and appropriate to our work.

A well-liked metric that calculates the mean discrepancies between two instance trials is MMD [4], [44]. It assumes that the mean is zero if two samples come from the same distribution. CORAL [14] suggests straightforward and effective distances that compute covariance while considering domain correlation. DeepCORAL [15] is a CORAL extension that is especially useful to hold non-linear operations with the help of deep learning prototypes. Additionally, it uses the second moment (i.e., variances) to compute feature distribution gaps at network bottleneck layers. HOMM [16] is incompatible with LMMD and CORAL, two methods for handling fundamental moment matching. Through the use of clustering algorithms, JDDA [45] permits the learning of discriminative features for both instance-based and center-based approaches. They concentrate on how compact the distributions are within and between classes. The MMD discrepancy measures have a version called LMMD [22], [23]. When the local information of domains is ignored, using the Maximum Mean Discrepancy (MMD) measure diminishes the overall distribution disparity between the domains. In contrast, Local Maximal Margin Distribution (LMMD) addresses this challenge by incorporating local data from multiple domains that share the same categories and effectively collecting pertinent information. The domain adaption method proposed by JDOT [28] improves the alignment between the feature and label distributions while making it easier to optimal transport coupling. Using deep learning architectures, DeepJDOT [29] reduces the gaps across the domains by combining common features and label distributions.

C. LEARNING WITH NOISY LABELS AND OUTLIERS

Robustness to label noise and outliers is an essential consideration in domain adaptation, as noisy labels and outliers can significantly affect the model's performance. Several approaches have been proposed to address this issue. One approach is to use robust loss functions that are less sensitive to label noise. For example, the Huber loss function [46] is less susceptible to outliers than the mean squared error loss function. Another approach is to use label smoothing [47], which replaces the hard labels with soft labels that assign some probability mass to other classes. Another approach is to use self-training [48], where the model is trained on the labeled source domain data and then used to generate pseudo-labels for the unlabeled target domain data. The model is then retrained on the combined labeled and pseudo-labeled data. This approach effectively reduces the impact of label noise in domain adaptation [49]. SCE [50] is a loss function designed to be more robust to label noise than the standard cross-entropy loss function. It achieves this by using a symmetric version of the cross-entropy loss that considers the similarity between the predicted and ground-truth labels. This makes the loss function less sensitive to label noise and more robust to mislabeled data. ROT [34] is a method that uses optimal transport to align the distributions of the source and target domains while simultaneously minimizing the impact of label noise. It achieves this by using a robust loss function less sensitive to label noise and incorporating a regularization term that encourages the model to produce smooth and consistent predictions. ROOT [51] is an extension of the ROT method that incorporates a robust loss function that is less sensitive to outliers and label noise. It achieves this by using a modified Wasserstein distance robust to outliers and a regularization term that encourages the model to produce smooth and consistent predictions. ROOT effectively improves the robustness of domain adaptation models to label noise and outliers, and it has been applied successfully in several computer vision tasks, such as object recognition and semantic segmentation. By employing ROOT, domain adaptation models can better handle noisy and mislabeled data and outliers and produce more accurate and reliable predictions. The method [52] uses optimal transport to align the distributions of the source and target domains while minimizing the impact of label noise. It also incorporates a noisy label robust loss function less sensitive to label noise and a mixup regularization term that encourages the model to produce smooth and consistent predictions.

Also, some recent works have proposed using adversarial training to improve the robustness of domain adaptation models to label noise. For example, the Adversarial Discriminative Domain Adaptation (ADDA) [20] method uses a domain discriminator to distinguish between the source and target domains and an adversarial loss to encourage the feature extractor to produce domain-invariant features. The technique effectively reduces the impact of label noise in domain adaptation [53]. In our work, we investigate the impact of label noise in the source domain on the performance of domain adaptation models and explore techniques to improve the robustness of these models to noisy labels. It is well-known that outliers and noisy labels in the training data can significantly affect the performance of machine learning models. To address this issue, we propose using robust outlier optimal transport, which computes the transport matrix using the Sinkhorn algorithm with Wasserstein distance to minimize the cost of transportation from the source to the target domains. Previous works in domain adaptation have focused on reducing the global distribution gap between the source and target domains to enhance domain-invariant features. However, these methods often fail to capture relevant, most informative local information for transferring knowledge between domains, especially considering category-wise information. To address this issue, we propose minimizing the domain's feature distributions using local maximum mean discrepancy (LMMD).

D. DOMAIN ADAPTATION FOR OPTIMAL TRANSPORT

Machine learning researchers are paying close attention to optimal transport, which has been expanded to include domain adaptability. It interacts with various disciplines, including probability, geometry, and optimization theory. It is also a full geometrical toolkit that computes the probability of distributions and optimal mapping to minimize cost functions in metric space [54]. The concept of optimal transport was first put forth in 1781 [55] and later expanded by Kantorovich, by looking at optimal coupling and duality, loosens up Monge's formulation and optimal resource location. Recent studies on optimal transport have focused on machine learning, Computer vision, NLP, image retrieval, and other domains. Prior investigations suggested that domain adaptation could benefit from using optimal transport [28], [29]. Using the domain-specific probability distributions, the Wasserstein distance (WD) is a well-liked technique for reducing transportation costs.

WD [27] has its origins in optimal transport (OT) theory and functions as a measure between two probability estimations. These studies were motivated by Generative Adversarial Network (GAN) [30]. Robust optimal transport has been proposed as a method for outlier detection in various applications, including computer vision, image processing, and machine learning. The idea is to use optimal transport to estimate the distance between likelihood distributions and identify outliers that deviate significantly from the expected distribution. However, OT is very sensitive to outliers in the data, as every sample, including outliers, is weighed similarly in its objective function due to the marginal constraints. This can lead to suboptimal results and reduced performance in applications with outliers [34], [51]. The epsilon-contamination model and the Huber loss function have been widely used in various applications, including computer vision, machine learning, and signal processing, to handle data outliers and improve statistical models' robustness [56]. While previous studies have explored the use of robust methods for handling noisy labels, they often focus on global features and do not take into account local features. This can lead to misclassification of data points and reduced performance when dealing with unseen data. Additionally, robust methods for handling noisy labels may not generalize well to new datasets, limiting their effectiveness. To address these limitations, we focused on robust domain adaptation, which considers local information and noisy robustness to enhance the model's capacity to generalize to new datasets.

E. SUBDOMAIN ADAPTATION

Subdomain adaptation is a recent area of research that focuses on matching the distributions of subdomains between the source and target domains. This approach has been explored in several works, such as [17], [43], [57], [58], and [59]. MSTN [58] is a subdomain technique that creates semantic information for unlabeled target data. It does this by first creating multiple unique attribute spaces, one for each source domain. Then, it independently aligns the source and target distributions in each space. Finally, it encourages alignments between these spaces to determine class predictions on unlabeled target cases. Co-DA [59] is a similar technique that creates multiple unique attribute spaces. However, unlike MSTN, Co-DA does not independently align each space's source and target distributions. Instead, it encourages alignments between these spaces by using a contrastive loss function. MSTN and Co-DA have been shown to be effective in improving the performance of transfer learning models on unlabeled target data. However, Co-DA has been shown to be more effective than MSTN in some cases.

DSAN [57] proposed a subdomain adaptation technique that uses local maximum mean discrepancy (LMMD) to align the distributions. LMMD is a distance metric that measures the difference between the means of the distributions in a local region of the feature space. DSAN showed that LMMD is more effective than adversarial loss in capturing the geometry of the data. Our proposed approach, RDAOT (Robustness Feature Domain Adaptation via Optimal Transport), considers the model's robustness to handle noisy labels and outliers. This is achieved by minimizing the domain distribution discrepancy using local maximum mean difference (LMMD). Unlike previous subdomain adaptation techniques that required additional networks, RDAOT uses LMMD to efficiently capture the geometry of the data without the need for additional networks. By considering the robustness of the model, RDAOT can handle noisy labels and outliers, which are common issues in domain adaptation tasks.

III. OBJECTIVE

We present an innovative RDAOT model that can effectively adapt to unsupervised domain situations is presented in this research. Prior methods [16], [35], [36], [45] in this field have been demonstrated to be insufficiently robust for practical applications and unable to grasp fine-grained data. For instance, the MDD method [10] aims to minimize global distribution divergences between domains. However, it tends to overlook local information and robustness issues, which are crucial considerations when dealing with complex data distributions. Another approach, CORAL [14], focuses on aligning distribution divergences using a covariance matrix. While this approach can help with distribution alignment, it lacks consideration for geometrical information, local information, and noise in the data. WD [30] techniques address the geometrical structure of data, which is a step in the right direction. However, these techniques still fail to capture local information and ensure robust knowledge transfer between domains. Furthermore, methods based on KL [87] divergence aim to measure the probability distribution between source and target domains. However, these methods tend to struggle to capture local information and are particularly sensitive to outliers in the data, which can negatively impact their performance. To address these problems, RDAOT uses a subdomain adaptation approach to extract relevant features using LMMD, and it also uses a robust noise-label approach to deal with noisy labels at the source domain. In addition, RDAOT introduces a new metric called ROT that is robust to outliers and noise. This metric helps to preserve geometrical information when transferring knowledge from the source domain to the target domain. This helps to avoid negative transfer, which can significantly impact the model's performance. RDAOT has been evaluated on various benchmark datasets and has been shown to outperform state-of-the-art methods [1], [7], [14], [45], [72] on all of them. This suggests that RDAOT is a promising new approach for unsupervised domain adaptation.

IV. PROPOSED APPROACH

The RDAOT techniques, created specifically for unsupervised domain adaptation, are introduced in this section. To facilitate understanding, we present a notation with descriptions used in the paper in Table 1. Subsection IV-A description precisely specifies the problem statement. We outline the different elements of the RDAOT model and explain the mathematical reasoning behind each one. The end-to-end architecture of the overall components is depicted in Fig. 2, along with a brief description in the text. The following sections will give an in-depth analysis of each component.

A. PROBLEM DEFINITION

The deep neural network $h(\mathbf{x})$ is built using robust domain adaptation and can handle label noise in the source domain, minimize discrepancies across domains by considering subdomains, and identify common features across data sets to reduce the risk of errors in predictions. Given a labelled source domain, represented as $\mathcal{K}_s$, which consists of n_{sr} labelled examples, each depicted as $(\mathbf{x}_i^{sr}, \mathbf{y}_i^{sr})$, where $\mathbf{y}_i^{sr}$ is a one-hot vector representing the label of $\mathbf{x}_i^{sr}$, meaning that $y_{i,j}^{sr} = 1$ means $\mathbf{x}_i^{sr}$ belongs to the jth class. In addition, a n_{tr} unlabeled instance target domain denoted as $\mathcal{K}_t$ controls unlabeled samples. The feature space and label space for $\mathcal{K}_s$ and $\mathcal{K}_t$ are the same despite examples from additional data allotments. However, there may be large discrepancies between the marginal probability distributions of $q_s(X_{sr})$ and $p_t(X_{tr})$.

B. LEARNING ROBUSTNESS FOR NOISY LABELS

Learning robustness for noisy labels is a challenging problem in deep learning. One approach to address this issue is using domain adaptation techniques to learn transferable features robust to label noise. Symmetric Cross Entropy (SCE) is a loss function that effectively reduces the impact of label noise in the source domain. n is defined as:

$$L_{CE} = -\sum_{i=1}^{n} y_i \log(q_{si}) \tag{1}$$

$$L_{RCE} = -\sum_{i=1}^{n} q_{si} \log(y_i) \tag{2}$$

$$L_{sym} = \alpha L_{CE} + \beta L_{RCE} \tag{3}$$

TABLE 1. Dictionary of notations.

Definition	Notation
Examples of Source and target	K_t, K_s
Total samples of target/source samples	n_{tr}, n_{tr}
Source /target domain marginal distribution	$q_s(x_{sr}), p_t(x_{tr})$
Source domain with labels	X_{sr}, Y_{sr}
Target examples with their labels	X_{tr}, Y_{tr}
Balanced parameters (Trade-off)	λ, λ_s
Batch-size	b
Cross entropy loss for classification	$L(\cdot, \cdot)$
Identity matrix	I
Feature map	$\phi(\cdot)$
Target and source column vectors are all weights of one.	$1_t, 1_s$
Provide data that includes information from both the source and destination	X
class, Number of class	c,C
The importance of each domain is assigned a weight.	w^c
Gamma	γ
Epsilon	ϵ
Beta	β
Entropic regularization	$H(\gamma)$

$$SCE(p_s, y) = -\alpha \sum_{i=1}^{N} y_i \log q_{si} - (1-\beta) \sum_{i=1}^{N} q_{si} \log y_i \tag{4}$$

where n is the number of data points in the source domain, y_i is the ground truth label, and p_i is the predicted probability for data point i. The symmetric loss is a weighted sum of the cross-entropy loss and the reverse cross-entropy loss. The α and β are hyperparameters that control the trade-off between the cross-entropy loss and the reverse cross-entropy loss.

The Symmetric Cross Entropy (SCE) loss is a robust loss function that can be utilized to train a model for classifying data from a source domain that has noisy labels. This loss function is capable of handling the noise in the labels by taking into account both the predicted distribution and the ground truth distribution.

C. SUBDOMAIN ALIGNMENT

Recent studies have provided evidence that deep neural networks (DNNs) [36] outperform traditional feature techniques in terms of faster learning of transferable models [60], [61]. Conversely, global domain adaptation [9], [62] encompasses the comprehensive alignment of all aspects of domains simultaneously and may fail to collect appropriate information in each domain. This can lead to misclassification of some features and poor outcomes. Additionally, to reduce distribution divergences between fields, past domain adaptation algorithms neglected the topological information in the data and instead focused on universal domain-oriented matching loss [21], [62]. To solve these problems, we suggest the RDAOT method. The source domain's classification loss and the domain alignment loss, respectively, are expressed in the following mathematical formulae.

$$\hat{\mathcal{L}}_{\mathrm{CE}}(X^{sr}, Y^{sr}) = \min_h \frac{1}{n_{sr}} \sum_{i=1}^{n_{sr}} (h(\mathbf{x}_i^{sr}), \mathbf{y}_i^{sr}) \tag{5}$$

$$\min_h \frac{1}{n_{sr}} \sum_{i=1}^{n_{sr}} L\left(h\left(\mathbf{x}_i^{sr}\right), \mathbf{y}_i^{sr}\right) + \lambda \hat{d}(q_s, p_t) \tag{6}$$

In this approach, we employ global domain adaptation (GDA) to align the source and target domains. GDA operates on a global level and does not take into account subdomains that may exist within the same classes across different fields. To balance the classification and domain adaptation loss, we introduce a parameter denoted as λ>0.

Methods for subdomain adaptation, such as those in [22], [23], and [57], concentrate on the LMMD loss. In order to facilitate this process, we partition the labeled source domain, denoted as $\mathcal{K}_s$, and the unlabeled target domain, denoted as $\mathcal{K}_t$, into C subdomains. Each subdomain is identified as $\mathcal{K}_s^{(c)}$ and $\mathcal{K}_t^{(c)}$, respectively, where c begin from 1 to C and represents a specific class. Additionally, we have access to the category labels $\mathcal{K}_s^{(c)}$ and $\mathcal{K}_t^{(c)}$ for the subdomains. The allocations of these subdomains are represented as $q^{(c)}$ and $p^{(c)}$, respectively, indicating the distribution of data within each subdomain. SDA aims to approximate the distribution of data points within a subdomain, given a set of samples with the same label. This is accomplished via the subdomain loss, which consists of both a classification loss and an adaption loss for the subdomain [22], [57].

$$\min_h \frac{1}{n_{sr}} \sum_{i=1}^{n_{sr}} L\left(h\left(\mathbf{x}_i^{sr}\right), \mathbf{y}_i^{sr}\right) + \lambda \mathbf{J}_c\left[\hat{d}\left(q^{(c)}, p^{(c)}\right)\right] \tag{7}$$

where the mean of the category is represented by $\mathbf{J}_c[\cdot]$. The mismatch in distribution between subdomains can be quantified by computing the disparity in equation (7) between the subdomain allocations obtained through the Maximum Mean Discrepancy (MMD) [4] and the Localized Maximum Mean Discrepancy (LMMD) [57], along with the trade-off criteria controlled by the parameter λ.

1) LOCAL MAXIMUM MEAN DISCREPANCY

The LMMD strategy, as detailed in [22] and [57], focuses on matching the same category's origin and destination (i.e., target) domains and yields valuable information. To address this issue, the LMMD technique is offered, which takes into account the relevance of samples based on their classification.

$$d_{\mathcal{H}}(q,p) \triangleq \mathbf{E}_c \left\| \mathbf{E}_{q^{(c)}}\left[\phi\left(\mathbf{x}^{sr}\right)\right] - \mathbf{E}_{p^{(c)}}\left[\phi\left(\mathbf{x}^{tr}\right)\right] \right\|_{\mathcal{H}}^2 \tag{8}$$

where $\mathbf{x}^{sr}$ and $\mathbf{x}^{tr}$ are the instances in $\mathcal{K}_{sr}$ and $\mathcal{K}_t$, and $q^{(c)}$ and $p^{(c)}$ are the domains of $\mathcal{K}_s^{(c)}$ and $\mathcal{K}_t^{(c)}$, consequently. The individual instance is assigned to a unique classification based on its corresponding importance weight, denoted as w^c. Then, as (8), obtain an impartial estimator.

$$\hat{\mathcal{L}}_{\mathrm{LMMD}}(q,p) = \frac{1}{\mathbb{C}} \sum_{c=1}^{\mathrm{c}} \left\| \sum_{x_i \in K^s} w_i^{Sc} \phi\left(x_i^{sr}\right) - \sum_{x_j \in K^t} w_j^{\mathrm{Tc}} \phi\left(x_j^{tr}\right) \right\|_{\mathcal{H}} \tag{9}$$

where where w_i^{Sc} and w_j^{Tc} are the importance weights of x_i^{sr} and x_j^{tr} that belongs to c, consecutively and provided sample, x_i. For instance, x_i, the importance is calculated as follows:

$$w_i^c = \frac{y_{ic}}{\sum_{(x_j,y_j)\in K} y_{jc}} \tag{10}$$

where y_{ic} is represents the labeled of x_i and the c^{th} vector of the component y_c. In the specified source domain, the importance weight w_i^{sc} for the sample x_i^{sr} is computed using its true label y_i^{sr}. Similarly, for an unlabeled sample x_j^{tr} in the target domain, the importance weight w_j^{Tc} is computed using the corresponding pseudo label $\hat{y}_j^t$.

D. ROBUST OPTIMAL TRANSPORT

Optimal transport refers to a mathematical problem that seeks the most effective manner of transferring mass from one distribution to another [64], [65]. When applied to domain adaptation, we can envision the source and target domains as two distinct distributions, and our objective is to determine the optimal approach to transport the labeled data from the source domain to the target domain in the most efficient manner. The Sinkhorn algorithm [64], [66], [67] is an iterative computational technique commonly employed to solve the optimal transport problem. It provides an effective means of finding solutions for the transportation task by iteratively adjusting the mass transportation plan between the source and target distributions. The algorithm works by iteratively updating a transport plan, which is a matrix that specifies how much mass should be transported from each point in the source domain to each point in the target domain. Robust optimal transport is a variation of optimal transport designed to be more vital to outliers [68]. Outliers are data points very different from the rest of the data. In domain adaptation, outliers can be caused by noise or differences between the source and target domains [34], [51]. To Compute the cost matrix C that captures the pairwise distances or dissimilarities between the source and target samples using Wasserstein distance with entropic regularization as follows:

$$OT_\epsilon(w_s, \gamma) = \min_\gamma \langle C, \gamma\rangle + \epsilon \cdot H(\gamma) \tag{11}$$

To extend the OT objective function by incorporating a penalty term for outliers. This penalty term discourages the transportation of samples that are considered outliers. The extended objective function becomes:

$$OT_\epsilon(w_s, \gamma) = \min_\gamma \langle C, \gamma\rangle + \epsilon \cdot H(\gamma) + \lambda \cdot \mathcal{R}(\gamma) \tag{12}$$

Here, $\lambda > 0$ is the penalty parameter, and $\mathcal{R}(\gamma)$ is a robustness term that quantifies the presence of outliers in the transport plan. The specific form of $\mathcal{R}(\gamma)$ depends on the chosen robustness measure. It is worth noting that the particular choice of the robustness measure and penalty parameter update strategy may vary depending on the application and problem characteristics. Further research and experimentation may also be necessary to determine the most suitable robustness measure and penalty parameter settings for a particular domain adaptation scenario. To compute the transport plan, we utilize Sinkhorn algorithms with outlier using penalty at algorithm 1.

In robust domain adaptation, the goal is to align the feature distributions of the source and target domains while also being robust to label noise and outliers. This is achieved by jointly minimizing eq. (3), eq. (9), and eq. (12).

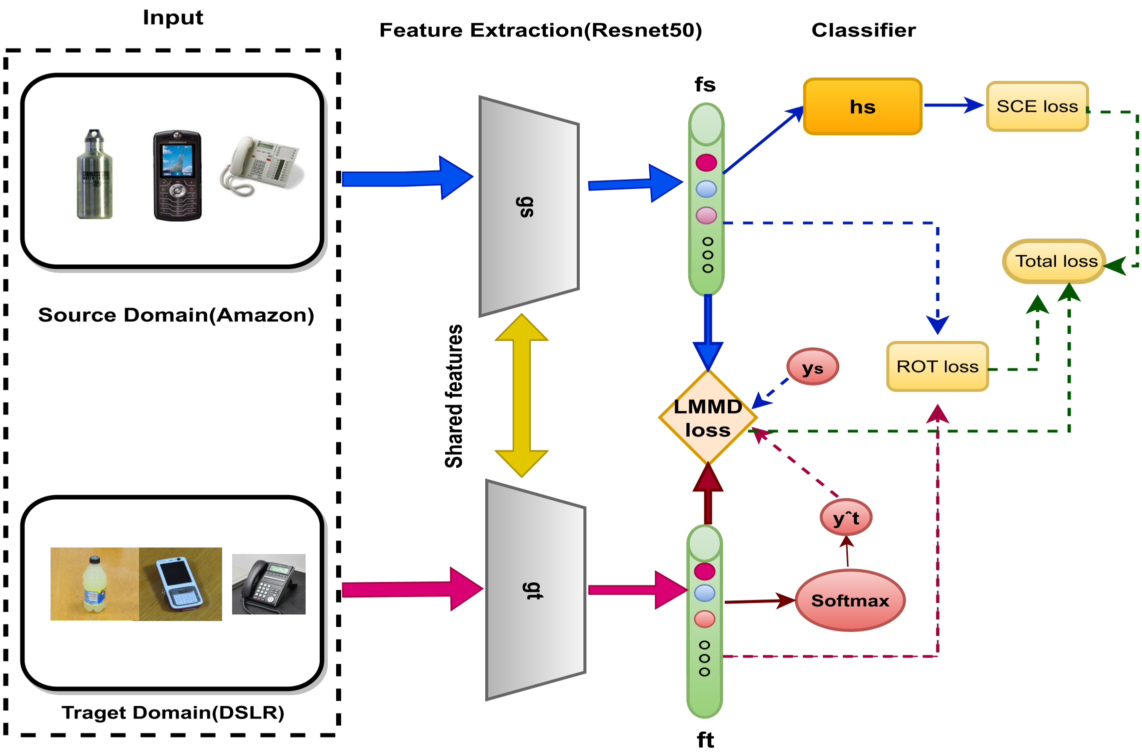


FIGURE 2. The proposed RDAOT framework for UDA. For extracts a feature of the f_s and f_t domains employing the Resnet50 model. Our approach involves sharing features between domains and minimizing divergences using LMMD loss. The Robust Optimal Transport (ROT) method is used in the fully connected layers to diminish the probability distributions of the cross domains. ROT uses the Sinkhorn algorithm with Wasserstein distance to be robust to outliers. To mitigate the impact of noise labels in the source domain, we utilize the Symmetric Cross-Entropy (SCE) loss function at the classifier layer. The ground truth labels for the source domain are Y_s, and the predicted labels for the target domain are Y_t.

Algorithm 1 Sinkhorn Algorithm With Outlier Robust Using Penalty

Require: C: Cost matrix, m: Number of source samples, n: Number of target samples, p: Order of the distance, λ: Hyperparameter

Ensure: T: Transport plan

Initialize $T \leftarrow \frac{1}{mn}\mathbf{1}\mathbf{1}^T$

while Not converged **do**

Compute row and column scaling factors $r \leftarrow \frac{1}{m}\sum_{j=1}^{n} T_{ij}$ and $c \leftarrow \frac{1}{n}\sum_{i=1}^{m} T_{ij}$

Update $T \leftarrow \frac{1}{mn}\text{diag}(r)C^p\text{diag}(c)T + \lambda\left(T - \frac{1}{mn}\mathbf{1}\mathbf{1}^T\right)$

end while

Return T

The total loss functions can be expressed as follows:

$$L_{Total} = L_{sym} + \lambda\hat{\mathcal{L}}_{\text{LMMD}}(q, p) + \lambda OT_{\epsilon}(w_s, \gamma) \quad (13)$$

Here, L_{sym} to learn noisy labels exhibits an SCE loss in the labeled source domain. $\hat{\mathcal{L}}_{\text{LMMD}}$ denotes the LMMD loss, which measures the difference between the feature distributions of the source and target domains. Finally, $OT_{\epsilon}(w_s, \gamma)$ is the robust optimal transport, which controls outliers and noisiness in the dataset using a transport plan matrix and computing cost matrix. The tradeoff parameter λ is used for LMMD and OT loss.

By minimizing the total loss functions jointly, the model can learn domain-invariant features that are robust to label noise and outliers, improving the model's performance on the target domain. The symmetric cross-entropy loss helps the model learn from the labeled source domain, while the LMMD loss and robust optimal transport help the model align the feature distributions of the source and target domains and handle label noise and outliers.

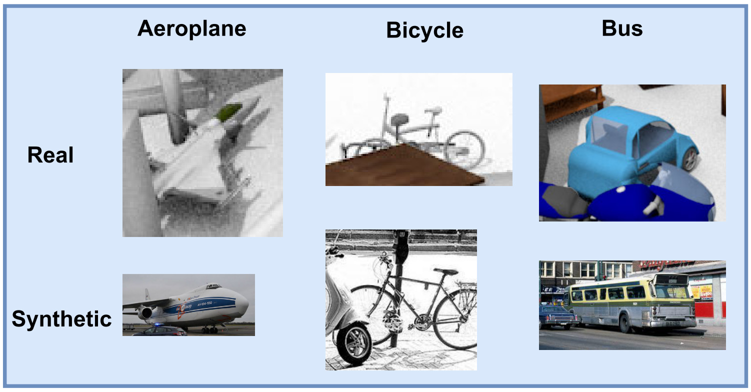
V. EXPERIMENTS

To evaluate the effectiveness of our proposed model (RDAOT), we compare its performance to cutting-edge transfer learning and deep learning techniques.

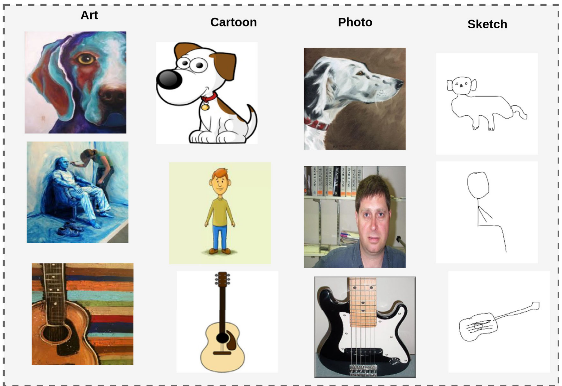
A. DATASETS

To validate our results, we conducted experiments using widely used domain adaptation datasets commonly employed in the field.

Office31 [2] is a common example for optical domain adaptation. It comprises 4,652 images and 31 classes from three domains: Amazon, DSLR, and Webcam. The Amazon dataset comprises images obtained from multiple sources, including amazon.com, DSLR cameras, and



(a) VisDA-2017



(b) PACS

(c) Office31

FIGURE 3. Examples of a) VisDA-2017, b) PACS, and c) Office31.

webcams. It encompasses a diverse collection of photographs captured in various settings using both web and SLR cameras. We evaluated our approaches on six distinct transfer tasks as described in the study [1], [9], [57].

OfficeHome [69] is a common benchmark for domain adaptation. It consists of 15,588 images and 65 classes from 4 domains: Art (A), Clipart (C), Product (P), and Real-World (R). Art contains creative images such as drawings, paintings, and decorations. Clipart is a collection of clipart images. The product contains images of objects without a background. Real-World contains images of objects captured with a regular camera. On all 12 transfer tasks, we examined our techniques.

ImageCLEF-DA is a dataset for domain adaptation consisting of three domains: Pascal VOC 2012 (P), ImageNet, Caltech-256 (C), and ILSVRC 2012 (I). There are 600 images in each domain, and these images are further divided into 12 categories, with 50 images allocated to each category. To create diverse transfer tasks, we formed six combinations of tasks. It is a valuable addition to the Office31 dataset, providing more consistent opportunities for research investigations. Unlike Office31, which comprises multiple varying-sized domains, ImageCLEF-DA consists of three similar domains. This balanced distribution of domains allows for more standardized and reliable studies in the domain adaptation field[1] [21].

VisDA-2017 [70] is a standard dataset that serves as a standardized evaluation framework for assessing domain adaptation algorithms in the context of visual recognition tasks. The dataset comprises a total of 280,000 images distributed across 12 classes. Specifically, it consists of 152,397 synthetic images in the source domain and 55,388 real-world images in the target domain. In the source domain, the dataset includes artificial 2D renderings of 3D models captured under different angles and lighting situations. The target domain, on the other hand, comprises real images shot in real-world circumstances. In this benchmark, the objective is to train models using synthetic photos and subsequently assess their performance on real-world photographs. This setup aims to replicate the common domain shift observed in practical applications.

OfficeCaltech10 is a benchmark dataset for visual recognition tasks in domain adaptation. It is a supplement to the well-known Office dataset, which contains images from three distinct domains: W(Webcam), A(Amazon), and D (Dslr). The OfficeCaltech10 dataset incorporates ten extra object and

[1] http://imageclef.org/2014/adaptation

TABLE 2. The most typical six datasets for UDA (C=class and D=domian.

datasets	Min no. images	max no. images	C	D	Total
Office31	7	100	31	3	4110
OfficeCaltech10	8	151	10	4	2533
OfficeHome	15	99	65	4	15588
ImageCLEF-DA	50	50	12	4	2400
PACS	80	820	7	4	7,900
VisDA-2017	207	7500	12	2	280,000

image classes from the Caltech-256 dataset. This inclusion represents a separate and more demanding domain (C) within the dataset. It has ten classes and 12 combination tasks for four domains.

PACS is a benchmark for domain generalization and adaptation in the case of visual recognition tasks. It consists of four domains: cartoons (Cl), photos (Pr), art images (Ar), and sketches (Sc). Each domain contains seven categories: dog, elephant, giraffe, guitar, horse, house, and person. The dataset was created to address domain generalization and adaptation challenges. Domain generalization is training a model on a set of source domains and then applying it to a new target domain. Domain adaptation is training a model on a set of source domains and then applying it to a new target domain similar to one of the source domains. PACS is a challenging dataset because it contains various image styles and content. This makes it difficult for models to generalize to new domains. However, PACS is also a valuable dataset because it can be used to evaluate the performance of domain generalization and domain adaptation methods. We collected the publicly accessible datasets utilized in our research for this paper from the following URL[2] for **VisDA-2017, Office-Home, Office31, OfficeCaltech10, and PACS**

B. IMPLEMENTATION DETAILS

In our experiments, we utilized the PyTorch framework [72] and employed the ResNet50 model [71] as the network backbone. The ResNet50 model was initially trained on the standard ImageNet dataset [75]. To refine the model, we conducted fine-tuning on the pre-trained model's convolutions and pooling layers. Simultaneously, the classifier layer was trained using back-propagation to optimize its performance. We utilized a mini-batch size 32, L2 weight decay 5×10^{-4}, early stopping maximum 40, and gradient descent (SGD) with a momentum of 0.9 for the training procedure. We follow an approach for annealing depicted in [9] and set the learning rate 10^{-4}.

We conducted an investigation and ablation analysis for the ROT (Robust Optimal Transport) loss employing λ set 0.1 to 0.3. In order to assess the effectiveness of the ROT loss function, we utilized the average classification accurateness of the target domain, specifically evaluating its performance on unlabeled target examples. Based on three random trials for each activity, we presented the findings. Our analysis and ablation study showed that ROT loss is an effective technique for improving the performance of domain adaptation models. Specifically, we found that selecting the value of λ impacted classification accuracy on the target domain.

C. STATE-OF-THE-ART COMPARISON STRATEGIES

We compare our proposed RDAOT techniques to several baseline deep learning and transfer learning methods.

- Deep CORAL [15] determines the covariance between the source and target domains using variances.
- ResNet [72] is a classic CNN that employs residual operations.
- DANNs [7] uses an inverse gradient layer and adversarial loss to reduce domain discrepancy.
- DAN [1] discrepancy between the source and destination are adjusted using multi-kernel MMD.
- RTN [21] reduces the Maximum Mean Difference between the two domains using the fusion elements.
- ADDA [20] utilizes adverse and discriminatory loss to align the distribution gap of training and testing domains
- MADA [43] uses numerous domain discriminators to distinguish diverse components to align different data distributions precisely.
- JANs [21] uses JMMD loss to reduce the discrepancy between the domains.
- CAN and iCAN [81] learn domain-invariant representations by combining the losses from all blocks using collaborative learning.
- CDAN and CDAN+E [17] to improve transferability and discriminability and regularise the uncertainty of the predictions made by the classifier. The classifier predictions are controlled for uncertainty and transferability via entropy minimization.
- JDDA [45] concentrates on joint alignment and distinguishing characteristics.
- DSAN [57] is built on using LMMD loss for subdomain adaptability.
- TMDA [23] relates to the local distribution disparity in each manifold as assessed by the manifold maximum mean discrepancy (MMD).
- OTAdapt [88] uses Gromov–Wasserstein distance loss to reduce the disparity between the domains.

D. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section covers evaluating the RDAOT method, including comparing its performance with other competitive strategies. Additionally, we discuss the feature visualization approaches used in RDAOT and present the results of ablation experiments.

1) COMPARATIVE ANALYSIS OF PERFORMANCE WITH THE BASELINE METHODS

We compared our approach to different baseline domain adaption approaches to assess its effectiveness. Our evaluation results demonstrate the superiority of our approach

[2]https://github.com/jindongwang/transferlearning/tree/master/data

TABLE 3. The OfficeHome dataset's accuracy (%) when using the ResNet50 model in a UDA environment.

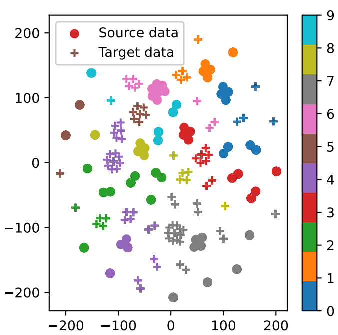
Methods	A → P	A → C	C → A	A → R	C → R	C → P	P → C	P → A	R → A	P → R	R → P	R → C	Average
ResNet [72]	50	34.9	37.4	58	46.2	41.9	31.2	38.5	53.9	60.4	59.9	41.2	46.12
DANN [7]	59.3	45.6	47	70.1	60.9	58.5	43.7	46.1	63.2	68.5	76.8	51.8	57.625
DAN [1]	57	43.6	45.8	67.9	60.4	56.5	43.6	44	63.1	67.7	74.3	51.5	56.28
CDAN [17]	69.3	49	54.4	74.5	68.4	66	48.3	55.6	68.4	75.9	80.5	55.4	63.80
DeepJDOT [29]	50.41	39.73	39.52	62.49	53.15	54.35	39.24	36.72	52.29	63.55	70.4	45.43	50.60
AWDAN [75]	**72.9**	51.1	60.8	**77.4**	**73.8**	**73**	52.2	62.9	68.6	80.4	83.1	55.9	67.67
JAN [21]	61.2	45.9	50.4	68.9	61	59.7	43.4	45.8	63.9	70.3	76.8	52.4	58.30
DSAN [57]	70.8	54.4	60.4	75.4	68	67.8	55.9	62.6	73.8	78.5	83.1	60.6	67.60
SCA [74]	64.6	46.7	53.1	71.3	65.2	65.3	47.2	54.6	68.2	72.7	80.2	56	62.09
RDAOT	71.05	**55.17**	59.33	75.67	68.21	68.37	**56.31**	61.31	**74.29**	77.74	82.95	**61.83**	67.68
RDAOT(SCE)	70.22	55.1	**61.23**	75.51	69.18	69.14	55.85	**64.07**	73.63	**78.82**	**83.69**	60.85	**68.10**

TABLE 4. Accuracy (%) achieved on the OfficeCaltech10 for UDA.

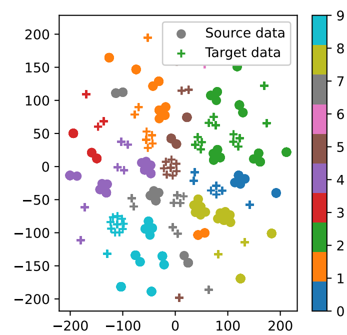
Methods	A → D	A → C	C → A	A → W	D → A	C → W	D → C	D → A	W → A	D → W	W → D	W → C	Average
Source [77]	85.4	82.7	91.5	78.3	83.1	88.5	74.6	80.6	77	99	100	69.6	84.19
CORAL [15]	80.8	85.3	91.1	76.3	81.1	86.5	80.4	88.7	82.1	99.3	100	78.7	85.85
GFK [11]	84.7	78.1	89.1	76.3	80.3	88.5	78.4	89	83.9	99.3	100	76.2	85.31
DMP [79]	90.4	86.6	92.8	91.3	88.5	93	85.3	91.4	91.9	97.7	100	85.6	91.20
KGOT [78]	86.6	85.7	91.4	82.4	87.1	92.4	85.6	91.8	89.7	99.3	100	85	89.75
OT_IT [68]	84.1	83.3	88.7	77.3	88.5	90.5	84	83.3	88.9	98.3	99.4	79.1	87.11
CKB [80]	93.6	87.0	93.4	90.2	90.8	93.6	83.5	92.7	92.4	100	100	84.3	92.22
RDAOT	**96.82**	**95.28**	**95.51**	**98.64**	**97.63**	**94.9**	**96.03**	**95.82**	**95.81**	100	100	**95.81**	**96.85**

TABLE 5. The accuracy (%) results for UDA on the IMAGECLEF-DA dataset using the ResNet50 model.

Methods	I → P	P → I	I → C	C → I	C → P	P → C	Average
ResNet [72]	74.8	83.9	91.5	78.0	65.5	91.3	80.7
DANN [7]	75	86	96.2	87	74.3	91.5	85
DAN [1]	75	86.2	93.3	84.1	69.8	91.3	81.8
JAN [21]	76.8	88	94.7	89.5	74.2	91.7	85.8
D-CORAL [15]	76.9	88.5	93.6	86.4	74	91.6	85.2
CAN [81]	78.2	87.5	94.2	89.5	75.8	89.2	85.7
MADA [43]	75	87.9	96	88.8	75.2	92.2	85.8
iCAN [81]	79.5	89.7	94.7	89.9	78.5	92	87.4
CDAN+E [17]	77.7	90.7	**97.7**	91.3	74.2	94.3	87.7
CDAN [17]	76.7	90.6	97	90.5	74.5	93.5	87.1
SCA [74]	78.1	89.2	96.8	91.3	78.2	94	87.9
SWD [82]	78.1	89.6	95.2	89.3	73.4	92.8	86.4
TDMA [23]	78.74	92.31	94.48	88.71	75.83	91.45	86.92
RDAOT(ours)	**80.37**	**93.33**	97.5	**93.67**	**82.06**	**94.5**	**90.24**



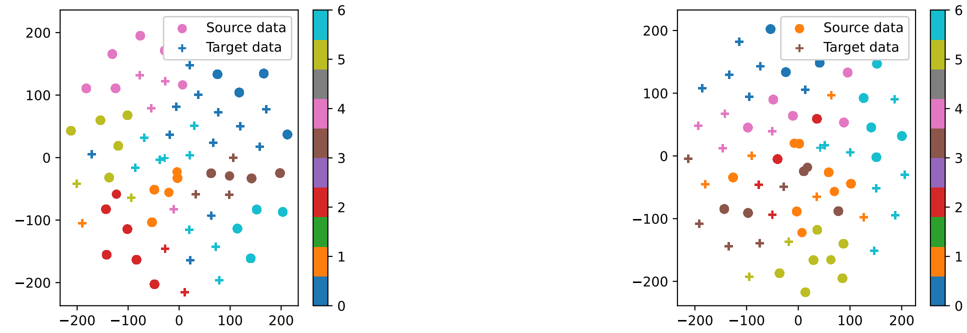
(a) A → W before DA



(b) A → W after DA

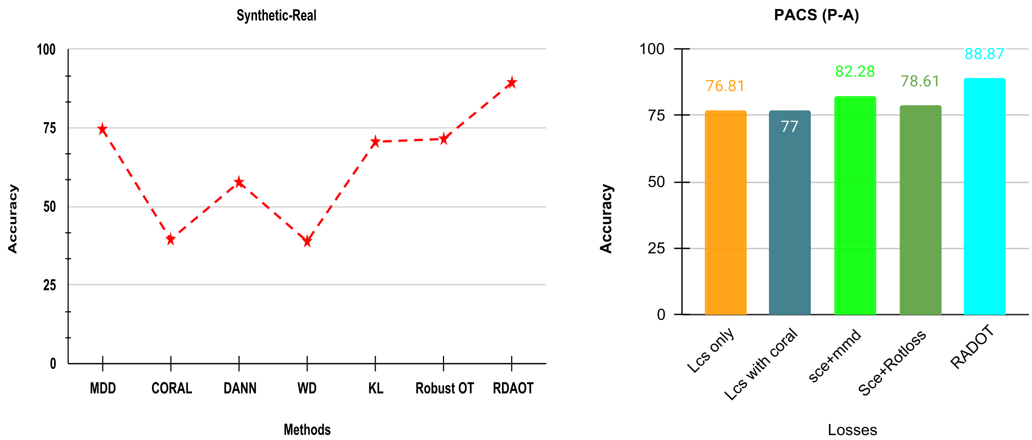
FIGURE 4. Using t-SNE, features from the Office-Caltech10 data sets for A → W tasks are visualized employing the BDA (source only) model and the ADA (RDAOT) approach. target domain: '+', Source domain: 'o'; class represent by colors.

over these baselines, indicating its efficacy in addressing the challenges of the robustness of domain adaptation. These findings are supported by significant research in the field, highlighting the importance of developing practical, robust domain adaptation techniques to improve the performance of machine learning models in real-world scenarios.



(a) P → A before DA (b) P → A after DA

FIGURE 5. For the P → A task, the feature visualization source-only model and RDAOT approach were used on the PACS data sets. Source domain: 'o'; destination domain: '+'. Diverse colors describe classes of features.



(a) (b)

FIGURE 6. a) For Synthetic → Real tasks from the VisDA-2017 dataset, and b) Ablation study on PACS datasets for P → A with different losses.

TABLE 6. Accuracy (%) for UDA (Synthetic=S and Real=R) on VisDA-2017.

Methods	S → R
MDD [10]	74.6
GTA [83]	69.50
CDAN [17]	70
MCD [84]	69.9
DANN [7]	57.7
CORAL [14]	39.5
MEDM [86]	79.6
ERM [85]	39.10
ERM(prob) [85]	37.20
WD [30]	38.9
Robust OT [34]	71.5
MEDM-LS [86]	80.2
OTAdapt [88]	71.88
KL [87]	70.6
RDAOT(Ours)	**89.43**

TABLE 7. The accuracy results (%) achieved on the Office31 dataset for the UDA setting using the ResNet50 model.

Methods	A → W	D → W	W → D	A → D	D → A	W → A	Average
ResNet [72]	68.4	96.7	99.3	68.9	62.5	60.7	76.1
D-CORAL [15]	77.7	97.6	99.7	81.1	64.6	64	80.8
DAN [1]	83.8	96.8	99.5	78.4	66.7	62.7	81.3
ADDA [20]	86.2	96.2	98.4	77.8	69.5	68.9	82.9
DANN [7]	82	96.9	99.1	79.7	68.2	67.4	82.2
JAN [21]	85.4	97.4	99.8	84.7	68.6	70	84.3
GTA [83]	89.5	97.9	99.8	87.7	72.8	71.4	86.6
MADA [43]	90	97.4	99.6	87.8	70	66.4	85.2
CDAN [17]	93.1	98.2	100	89.8	70.1	68	86.6
CAN [81]	81.5	98.2	99.7	88.5	65.9	63.4	82.4
iCAN [81]	92.5	98.8	100	90.1	72.1	69.2	87.2
JDDA-I [45]	82.1	95.2	99.7	76.1	56.9	65.1	79.18
DSAN [57]	92.08	98.74	100	**91.16**	74.55	72.49	88.17
SCA [74]	93.6	98	100	89.5	72.6	72.4	87.68
JDDA-C [45]	82.6	95.2	99.7	79.8	57.4	66.7	80.2
ETD [89]	92.1	**100**	100	88	71	67.8	86.48
SWD [82]	90.4	98.7	100	94.7	70.3	70.5	87.4
TMDA [23]	86.21	97.09	99.8	83.32	65.61	64.84	82.81
OTAdapt [88]	75.35	74.57	73.69	73.83	75.48	72.37	74.22
AWDAN [75]	92	98.5	99.9	90.2	75.3	69.5	87.57
RDAOT(Ours)	**94.09**	98.99	100.00	89.56	**76.46**	**74.41**	**88.92**

a: OFFICEHOME

Table 3 We evaluated the effectiveness of our method on OfficeHome, which consists of 65 categories and four domains. We tested our methodology on 12 transfer assignments and compared it to state-of-the-art methods. On six of the twelve transfer tasks, our method outperformed the other approaches, with an average accuracy that exceeded the different approaches. In addition, we conducted

TABLE 8. Accuracy (%) for UDA setting on PACS datasets.

Methods	Ar → Cl	Ar → Pr	Ar → Sc	Cl → Ar	Cl → Sc	Cl → Pr	Pr → Ar	Pr → Cl	Pr → Sc	Sc → Ar	Sc → Cl	Sc → Pr	Average
DANN [7]	71	94.5	58.6	76.4	78.6	76.1	68	50.7	29.3	39.2	64.3	44.3	62.58
MMD [10]	79.5	94.5	62.1	79.5	80.8	74.1	67.7	47.4	59.7	40	65.7	45.1	66.34
ERM [85]	66.1	94.3	53.6	69.7	82	72.2	65.7	29.1	38	41.3	66.7	49.3	60.67
CORAL [14]	62.7	86.3	46.2	75.9	78.3	56.9	70	47.5	15.8	39.1	59.9	37.4	56.33
ERM(prob) [85]	63.5	93.5	60.9	70.8	81.5	70.4	63.3	27.2	35.9	40.9	67.9	46	60.15
KL [87]	73.1	95.4	67.4	83.3	83.1	68.2	75.5	67.7	64.5	48.2	63.5	39.1	69.08
WD [30]	76.2	92.4	53.9	69	72.9	48.7	62.6	56.1	22.3	36.1	60.5	38.5	57.43
RDAOT	**87.84**	**98.92**	**69.94**	**89.84**	**96.71**	**77.53**	**88.82**	**82.42**	**73.38**	**84.57**	**87.03**	**91.62**	**85.72**

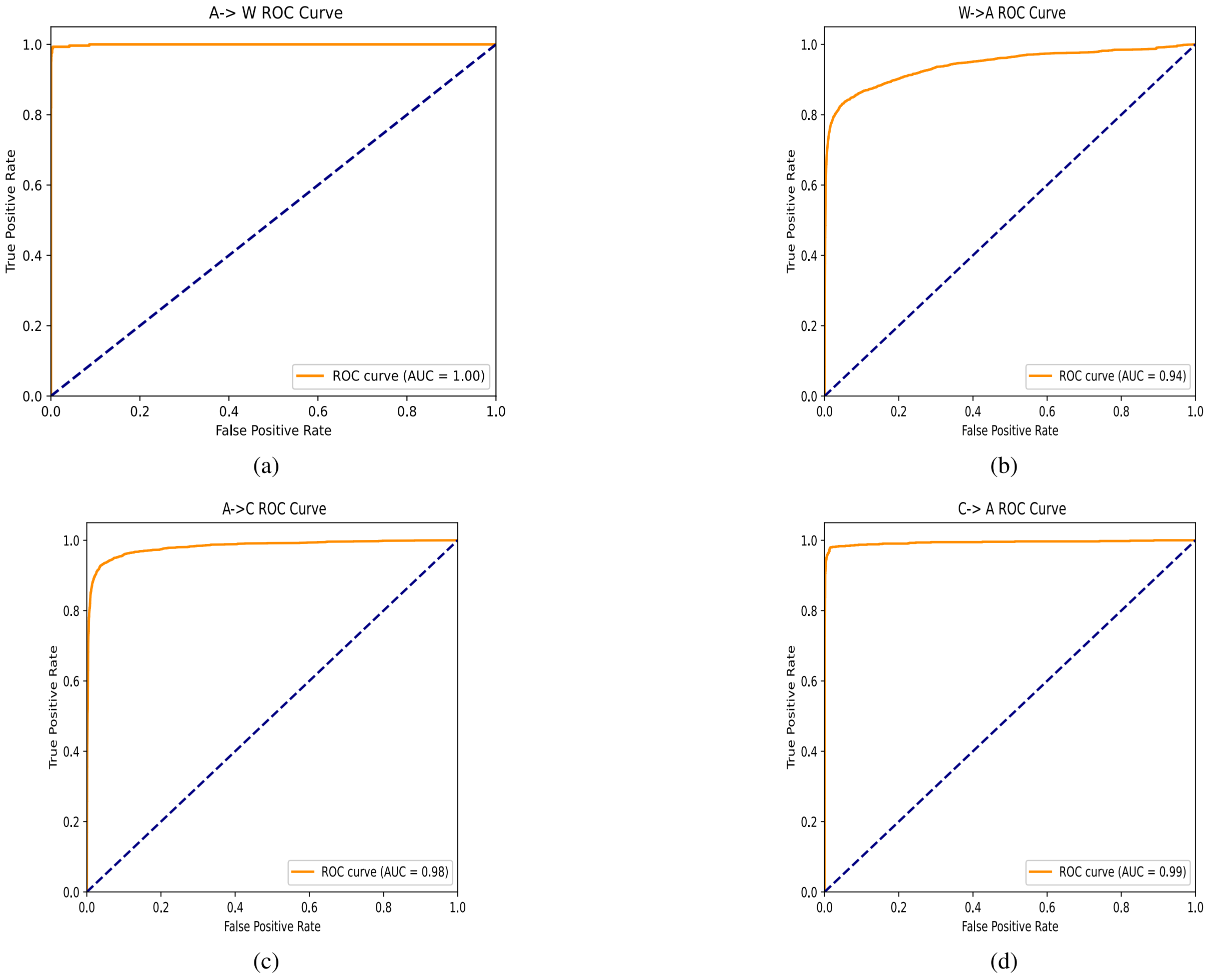


FIGURE 7. ROC AUC plots for different experiments. a) Using Officecaltech10 (A-W) tasks, b) Office31 (W-A) tasks, c) PACS dataset (A->C) tasks, and d) Imageclef (C->A).

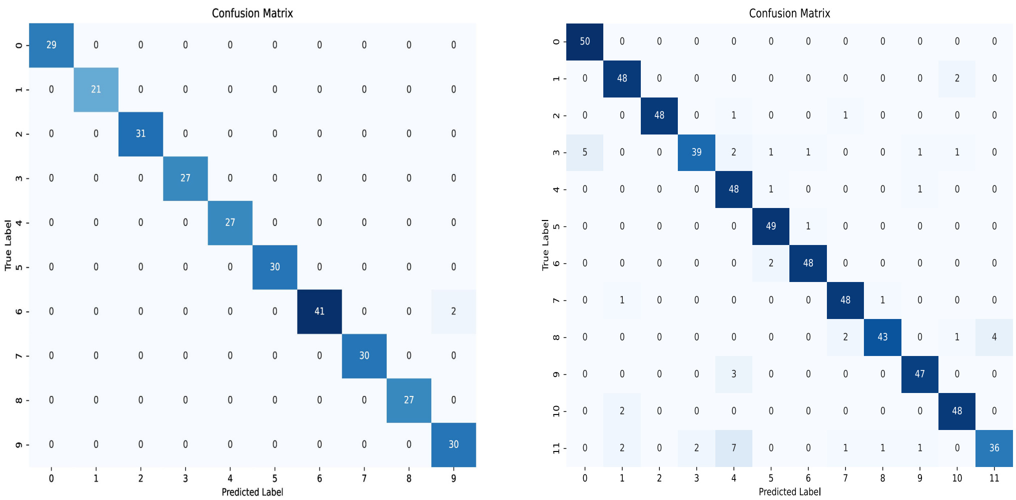
TABLE 9. The noise rate ratio, alpha, beta, and outlier ratio of the PACS dataset were the subject of an ablation investigation.

noise rate ratio	alpha	beta	outlier ratio	Acc(A → C)	Acc (A → S)	Acc (C → A)
0.3	0.1	0.3	0.2	87.84	69.94	89.84
0.2	0.1	0.3	0.2	87.16	74.65	90.28

experiments with two variants of our method: RDAOT and RDAOT(SCE). In our study, for RDAOT(SCE), we utilized the SCE category loss [50] to address noisy labels in the source domain. We observed that for certain tasks in RDAOT, employing robust optimal transport with symmetric cross-entropy was crucial [34]. Our investigation indicated that robust symmetric cross-entropy was effective in handling noisy labels, and as a result, RDAOT(SCE) exhibited superior performance.

b: OFFICECALTECH10

The results are shown in Table 4. We assessed the significance of our method on the OfficeCaltech10 dataset and compared our approach with several baseline domain



(a) (b)

FIGURE 8. a) Confusion matrix for A → W tasks from the OfficeCaltech10 dataset, and b) Confusion matrix on ImageCLEF-DA datasets for P → A tasks.

TABLE 10. Comparison of our approach RDAOT with other baselines.

	Criteria			
Approach	Criteria-1	Criteria-2	Criteria-3	Criteria-4
MDD [10]	✓	✗	✗	✗
CORAL [14]	✗	✓	✗	✗
DANN [7]	✓	✗	✗	✗
WD [30]	✗	✗	✓	✗
KL [87]	✓	✗	✗	✗
DSAN [57]	✗	✓	✗	✗
JDDA-I [45]	✓	✓	✗	✗
ETD [89]	✗	✗	✓	✗
SWD [82]	✗	✗	✓	✗
DeepJDOT [29]	✗	✗	✓	✗
HOT-DA [76]	✗	✓	✓	✗
OT_IT [68]	✗	✗	✓	✗
KGOT [78]	✗	✗	✓	✗
CKB [80]	✗	✓	✗	✗
Robust OT [34]	✗	✗	✓	✓
OTAdapt [88]	✗	✗	✓	✗
RDAOT	✓	✓	✓	✓

adaptation methods. Our experimental results demonstrated that our method achieved superior performance compared to other techniques in ten out of the twelve transfer tasks. The accuracy of our method surpassed that of various state-of-the-art strategies, further validating its effectiveness. Our results and visualizations confirmed that robust domain adaptation, which involves handling noisy labels in the source domain and maintaining outliers in the datasets using optimal transport with subdomain adaptation, is an effective solution for domain adaptation problems. These findings highlight the importance of developing robust and efficient domain adaptation techniques to improve the performance of machine-learning models in real-world scenarios.

c: IMAGECLEF-DA

Table 5 shows the results of our method exceed most baselines. The average accuracy shows a 2% improvement compared with the baseline. Our experiments verified that handling robust noisy labels and outliers in the datasets improved the performance of our model. Moreover, we found that focusing on local attributes in each domain class, also known as subdomain adaptation, was particularly effective. These findings highlight the importance of developing domain adaptation techniques that can handle noisy labels and outliers while leveraging local features to improve the performance of machine learning models in real-world scenarios.

d: VISDA-2017

Table 6 presents the results of our method on the VisDa-2017 dataset, a challenging and unbalanced dataset consisting of two domains, synthetic and real images. We used the synthetic domain as the source and the natural domain as the target. We compared our RDAOT method to numerous baseline domain adaption approaches and discovered that it surpassed the state-of-the-art methods, with an accuracy of 89.43%. These findings highlight the importance of developing robust and efficient domain adaptation techniques that can handle challenging and unbalanced datasets to improve the performance of machine learning models in real-world scenarios.

e: OFFICE31

Table 7 presents the evaluation results on the Office 31 benchmark dataset, which consists of three domains and 31 classes. The evaluation involved six transfer tasks with unbalanced datasets. Our method demonstrated high performance in three out of the six transfer tasks. Our approach performed comparably in the remaining three transfer tasks to the different baselines. Notably, our method excelled in the challenging tasks of W→A, D→A, and A→W, producing outstanding results. Our method surpasses other state-of-the-art techniques with an overall average accuracy of 88.92%. The observed improvements clearly demonstrate that our proposed framework can greatly enhance the performance of adaptation when combined with existing strategies.

f: PACS

We assessed our approach on the PACS dataset, which consists of four domains and 12 transfer tasks, and the results are presented in Table 8. Our method outperformed the state-of-the-art methods in 12 out of the 12 transfer tasks, which is a remarkable achievement. These results demonstrate the effectiveness of our approach in addressing the challenges of domain adaptation and improving the performance of machine learning models in real-world scenarios. Our findings emphasize the importance of developing robust and efficient domain adaptation techniques to handle diverse and challenging datasets to achieve state-of-the-art performance.

2) FEATURE VISUALIZATION

We employ t-SNE [90], creating a visualization that helps to show clear ideas of our method contribution by comparing with before and after domain adaption. Figure 4a depicts cross-entropy loss (source alone), while Figure 4b displays the outcomes of our RDAOT approach. Both figures were created using A $\rightarrow$ W tasks from Office-Caltech10 data sets. Before DA, features were dispersed and did not overlap, but after using our domain alignment approaches, features from multiple domains were brought together. Figure 5a and Figure 5b illustrate before and after domain adaptation using the PACS dataset for domain transfer tasks from domain P to A tasks. We present the ROC-AUC plot in Figure 7 utilizing the Micro-average ROC AUC method to compute the ROC AUC score for multi-class classification tasks. This approach considers all classes together, calculating the ROC curve and area under the curve (AUC) instead of treating each class separately. Figure 8 depicts the confusion matrix for the OfficeCaltech10 (A $\rightarrow$ W) tasks and the ImageCLEF-DA (P $\rightarrow$ A) tasks.

3) ABLATION INVESTIGATIONS

For our strategies, we tested with various losses and sensitivity values. We compared the Source alone loss (Lcs), Maximum Mean Discrepancy (MMD), CORAL, SCE, and Local Maximum Mean Discrepancy (LMMD) losses in domain adaptation to our Robust Domain Adaptation through Optimal Transport (RDAOT) technique. We did a thorough ablation analysis on the PACS dataset for the P$\rightarrow$A task, and the results are shown in Figure 6b. Our method outperformed all other approaches.

This result demonstrates that RDAOT is more effective at aligning the distributions of the source and target domains. The ablation study removed different components from the RDAOT method to assess their contributions. We discovered that each procedure component was essential for achieving high performance, indicating that RDAOT is a thorough method of domain adaptation. Figure 6a contrasts our findings with other techniques, including CORAl, MDD, WD, DANN, Robust OT, and KL. Our methodology outperforms all others with more accuracy. We have also examined other crucial variables that impact the outcome. We include investigated the effect of the noise rate ratio in Table 9. We use PACS datasets including transfer tasks (A $\rightarrow$ C, A $\rightarrow$ S, and C $\rightarrow$ A). This analysis confirmed that selecting a noise ratio is critical for model performance.

VI. COMPARED NOVELTY OF OUR METHOD WITH OTHER DA APPROACHES

We compare the significance of our approaches to others' baselines by employing four criteria to determine whether or not they satisfy:-

- **Criteria-1** Preserve global information by minimizing marginal distribution discrepancy.
- **Criteria-2** Retains local information for capturing relevant features using LMMD.
- **Criteria-3** Holds geometrical information of the domains.
- **Criteria-4** keeps robustness during knowledge transfer from the source domain to target domains to handle noise or outliers.

From Table 10, we analyze the superiority of our techniques with state-of-the-art domain adaption methods and compare whether they satisfy or not the given criteria. As we observe, the majority of methods focus on minimizing global distributions and neglect local information, which is more informative. In addition, only a few methods consider robustness and geometrical information criteria. Our approach considers all the cases and shows surpasses the average performance in all six benchmark datasets.

VII. CONCLUSION AND FUTURE DIRECTIONS

This work presents a novel approach to unsupervised domain adaptation (UDA) called RDAOT (Robust Deep Adaptation via Optimal Transport). Our technique, which includes subdomain adaptation, emphasizes the importance of local features and the robustness of noisy labels in domain adaptation. In addition, we leverage robust optimal transport methodologies, specifically employing Sinkhorn algorithms, to handle outliers or noise within domains. We conducted rigorous experiments and thorough analysis to evaluate the effectiveness of our approach on six benchmark datasets for domain adaptation. Finally, the proposed method (RDAOT) has shown promising results in improving the performance of domain adaptation models on both shallow transfer learning and deep-domain adaptation scenarios. In our future work, we envision extending the applicability of our approach to diverse domains, with a specific focus on textual domains, time series, healthcare, and beyond. Additionally, we aspire to enhance our methodology to facilitate smooth adaptation across multiple domains, even in cases where the source data originates from multiple diverse domains.

DECLARATION OF THE INTEREST

The authors claim no financial or other conflicts of interest that could have influenced the study's findings.

REFERENCES

[1] M. Long, Y. Cao, J. Wang, and M. Jordan, ''Learning transferable features with deep adaptation networks,'' in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 97–105.

[2] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, ''Adapting visual category models to new domains,'' in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 213–226.

[3] S. Pan and Q. Yang, ''A survey on transfer learning,'' *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Nov. 2010.

[4] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, ''A kernel two-sample test,'' *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 723–773, Jan. 2012.

[5] G. Liu and J. Wu, ''Unsupervised person re-identification by intra–inter camera affinity domain adaptation,'' *J. Vis. Commun. Image Represent.*, vol. 80, Oct. 2021, Art. no. 103310.

[6] W. M. Kouw and M. Loog, ''A review of domain adaptation without target labels,'' 2019, *arXiv:1901.05335*.

[7] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, and H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, ''Domain-adversarial training of neural networks,'' *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, 2016.

[8] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, ''A theory of learning from different domains,'' *Mach. Learn.*, vol. 79, nos. 1–2, pp. 151–175, May 2010.

[9] Y. Ganin and V. Lempitsky, ''Unsupervised domain adaptation by back-propagation,'' in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1180–1189.

[10] X. Chen, S. Wang, M. Long, and J. Wang, ''Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation,'' in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 1081–1090.

[11] B. Gong, Y. Shi, F. Sha, and K. Grauman, ''Geodesic flow kernel for unsupervised domain adaptation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2066–2073.

[12] X. Zhang, F. Cai, X. Hu, J. Zheng, and H. Chen, ''A contrastive learning-based task adaptation model for few-shot intent recognition,'' *Inf. Process. Manage.*, vol. 59, no. 3, May 2022, Art. no. 102863.

[13] J. Tao, S. Wen, and W. Hu, ''Robust domain adaptation image classification via sparse and low rank representation,'' *J. Vis. Commun. Image Represent.*, vol. 33, pp. 134–148, Nov. 2015.

[14] B. Sun, J. Feng, and K. Saenko, ''Correlation alignment for unsupervised domain adaptation,'' in *Domain Adaptation in Computer Vision Applications*. Cham, Switzerland: Springer, 2017, pp. 153–171.

[15] B. Sun and K. Saenko, ''Deep coral: Correlation alignment for deep domain adaptation,'' in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 443–450.

[16] C. Chen, Z. Fu, Z. Chen, S. Jin, Z. Cheng, X. Jin, and X. S. Hua, ''HoMM: Higher-order moment matching for unsupervised domain adaptation,'' in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 4, pp. 3422–3429.

[17] M. Long, Z. Cao, J. Wang, and M. I. Jordan, ''Conditional adversarial domain adaptation,'' in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 1647–1657.

[18] L. Meng, H. Su, C. Lou, and J. Li, ''Cross-domain mutual information adversarial maximization,'' *Eng. Appl. Artif. Intell.*, vol. 110, Apr. 2022, Art. no. 104665.

[19] B. Sun, J. Feng, and K. Saenko, ''Return of frustratingly easy domain adaptation,'' in *Proc. AAAI Conf. Artif. Intell.*, vol. 30, no. 1, 2016.

[20] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, ''Adversarial discriminative domain adaptation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2962–2971.

[21] M. Long, H. Zhu, J. Wang, and M. I. Jordan, ''Deep transfer learning with joint adaptation networks,'' in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2208–2217.

[22] Z. Wang, X. He, B. Yang, and N. Li, ''Subdomain adaptation transfer learning network for fault diagnosis of roller bearings,'' *IEEE Trans. Ind. Electron.*, vol. 69, no. 8, pp. 8430–8439, Aug. 2022.

[23] P. Wei, Y. Ke, X. Qu, and T.-Y. Leong, ''Subdomain adaptation with manifolds discrepancy alignment,'' *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 11698–11708, Nov. 2022.

[24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ''Generative adversarial networks,'' *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[25] W. Zhang and D. Wu, ''Discriminative joint probability maximum mean discrepancy (DJP-MMD) for domain adaptation,'' in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.

[26] J. Choi, M. Jeong, T. Kim, and C. Kim, ''Pseudo-labeling curriculum for unsupervised domain adaptation,'' in *Proc. Brit. Mach. Vis. Conf. (BMVC)*. Springer, 2019.

[27] C. Villani, *Optimal Transport: Old and New*, vol. 338. Cham, Switzerland: Springer, 2009.

[28] N. Courty et al., ''Optimal transport for domain adaptation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1853–1865, 2016.

[29] B. B. Damodaran, B. Kellenberger, R. Flamary, D. Tuia, and N. Courty, ''DeepJDOT: Deep joint distribution optimal transport for unsupervised domain adaptation,'' in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 447–463.

[30] J. Shen, Y. Qu, W. Zhang, and Y. Yu, ''Wasserstein distance guided representation learning for domain adaptation,'' in *Proc. 32nd AAAI Conf. Artif. Intell. 30th Innov. Appl. Artif. Intell. Conf. 8th AAAI Symp. Educ. Adv. Artif. Intell.*, 2018, pp. 4058–4065.

[31] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ''Generative adversarial nets,'' *Stat*, vol. 1050, p. 10, 2014.

[32] Y. Wang, X. Ma, Z. Chen, Y. Luo, J. Yi, and J. Bailey, ''Symmetric cross entropy for robust learning with noisy labels,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 322–330.

[33] D. Tanaka, D. Ikami, T. Yamasaki, and K. Aizawa, ''Joint optimization framework for learning with noisy labels,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5552–5560.

[34] Y. Balaji, R. Chellappa, and S. Feizi, ''Robust optimal transport with applications in generative modeling and domain adaptation,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 12934–12944.

[35] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, ''Analysis of representations for domain adaptation,'' in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, vol. 19. Cambridge, MA, USA: MIT Press, 2007, p. 137.

[36] Y. Bengio, A. Courville, and P. Vincent, ''Representation learning: A review and new perspectives,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[37] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, ''Transfer joint matching for unsupervised domain adaptation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1410–1417.

[38] Q. Gu, Z. Li, and J. Han, ''Joint feature selection and subspace learning,'' in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, vol. 2, Jul. 2011, pp. 1294–1299.

[39] J. Huang, A. Gretton, K. Borgwardt, B. Scholkopf, and A. Smola, ''Correcting sample selection bias by unlabeled data,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19, 2006.

[40] J. Hoffman, E. Rodner, J. Donahue, B. Kulis, and K. Saenko, ''Asymmetric and category invariant feature transformations for domain adaptation,'' *Int. J. Comput. Vis.*, vol. 109, nos. 1–2, pp. 28–41, Aug. 2014.

[41] W. M. Kouw, L. J. van der Maaten, J. H. Krijthe, and M. Loog, ''Feature-level domain adaptation,'' *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 5943–5974, 2016.

[42] M. Kandemir, ''Asymmetric transfer learning with deep Gaussian processes,'' in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 730–738.

[43] Z. Pei, Z. Cao, M. Long, and J. Wang, ''Multi-adversarial domain adaptation,'' in *Proc. 32nd AAAI Conf. Artif. Intell. 30th Innov. Appl. Artif. Intell. Conf. 8th AAAI Symp. Educ. Adv. Artif. Intell.*, Feb. 2018, pp. 3934–3941.

[44] W. Wang, H. Li, Z. Ding, F. Nie, J. Chen, X. Dong, and Z. Wang, ''Rethinking maximum mean discrepancy for visual domain adaptation,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 1, pp. 264–277, Jan. 2023, doi: 10.1109/TNNLS.2021.3093468.

[45] C. Chen, Z. Chen, B. Jiang, and X. Jin, ''Joint domain alignment and discriminative feature learning for unsupervised deep domain adaptation,'' in *Proc. Nat. Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 3296–3303.

[46] P. J. Huber, ''Robust estimation of a location parameter,'' in *Breakthroughs in Statistics: Methodology and Distribution*, 1992, pp. 492–518.

[47] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, ''Rethinking the inception architecture for computer vision,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[48] H. Scudder, ''Probability of error of some adaptive pattern-recognition machines,'' *IEEE Trans. Inf. Theory*, vol. IT-11, no. 3, pp. 363–371, Jul. 1965.

[49] Y. Zou, Z. Yu, X. Liu, B. V. K. V. Kumar, and J. Wang, ''Confidence regularized self-training,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5981–5990.

[50] Z. Zhang and M. R. Sabuncu, ''Generalized cross-entropy loss for training deep neural networks with noisy labels,'' in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, Dec. 2018, pp. 8792–8802.

[51] D. Mukherjee, A. Guha, J. M. Solomon, Y. Sun, and M. Yurochkin, ''Outlier robust optimal transport,'' in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 7850–7860.

[52] K. Fatras, H. Naganuma, and I. Mitliagkas, ''Optimal transport meets noisy label robust loss and mixup regularization for domain adaptation,'' in *Proc. Conf. Lifelong Learn. Agents*, 2022, pp. 966–981.

[53] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, ''CSPNet: A new backbone that can enhance learning capability of CNN,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 390–391.

[54] I. Redko, A. Habrard, and M. Sebban, ''Theoretical analysis of domain adaptation with optimal transport,'' in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, 2017, pp. 737–753.

[55] R. Durrett, ''Probability metrics and the stability of stochastic models (Sveltozar T. Racheu),'' *SIAM Rev.*, vol. 34, no. 1, p. 147, Mar. 1992.

[56] P. J. Huber, *Robust Statistics* (Wiley Series in Probability and Mathematics Statistics). 1981, pp. 309–312.

[57] Y. Zhu, F. Zhuang, J. Wang, G. Ke, J. Chen, J. Bian, H. Xiong, and Q. He, ''Deep subdomain adaptation network for image classification,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1713–1722, Apr. 2021.

[58] S. Xie, Z. Zheng, L. Chen, and C. Chen, ''Learning semantic representations for unsupervised domain adaptation,'' in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 5423–5432.

[59] A. Kumar, P. Sattigeri, K. Wadhawan, L. Karlinsky, R. Feris, W. T. Freeman, and G. Wornell, ''Co-regularized alignment for unsupervised domain adaptation,'' in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, Dec. 2018, pp. 9367–9378.

[60] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, ''How transferable are features in deep neural networks?'' in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, vol. 2, Dec. 2014, pp. 3320–3328.

[61] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, ''Learning and transferring mid-level image representations using convolutional neural networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1717–1724.

[62] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, ''Simultaneous deep transfer across domains and tasks,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4068–4076.

[63] Q. Wang and T. Breckon, ''Unsupervised domain adaptation via structured prediction-based selective pseudo-labeling,'' in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 6243–6250.

[64] G. Peyre and M. Cuturi, ''Computational optimal transport: With applications to data science,'' *Found., Trends® Mach. Learn.*, vol. 11, nos. 5–6, pp. 355–607, 2019.

[65] L. Kantorovitch, ''On the translocation of masses,'' *Manage. Sci.*, vol. 5, no. 1, pp. 1–4, Oct. 1958.

[66] M. Cuturi, ''Sinkhorn distances: Lightspeed computation of optimal transport,'' in *Proc. 26th Int. Conf. Neural Inf. Process. Syst.*, vol. 2, Dec. 2013, pp. 2292–2300.

[67] M. Cuturi and A. Doucet, ''Fast computation of Wasserstein barycenters,'' in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 685–693.

[68] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, ''Optimal transport for domain adaptation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1853–1865, Sep. 2017.

[69] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, ''Deep hashing network for unsupervised domain adaptation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5385–5394.

[70] X. Peng, B. Usman, N. Kaushik, D. Wang, J. Hoffman, and K. Saenko, ''VisDA: A synthetic-to-real benchmark for visual domain adaptation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 2021–2026.

[71] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and L. Antiga, ''PyTorch: An imperative style, high-performance deep learning library,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.

[72] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[73] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, ''ImageNet large scale visual recognition challenge,'' *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[74] W. Deng, L. Zheng, Y. Sun, and J. Jiao, ''Rethinking triplet loss for domain adaptation,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 29–37, Jan. 2021.

[75] Y. Zhang, N. Wang, and S. Cai, ''Adversarial sliced Wasserstein domain adaptation networks,'' *Image Vis. Comput.*, vol. 102, Oct. 2020, Art. no. 103974.

[76] M. El Hamri, Y. Bennani, and I. Falih, ''Hierarchical optimal transport for unsupervised domain adaptation,'' *Mach. Learn.*, vol. 111, no. 11, pp. 4159–4182, Nov. 2022.

[77] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''ImageNet classification with deep convolutional neural networks,'' *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[78] Z. Zhang, M. Wang, and A. Nehorai, ''Optimal transport in reproducing kernel Hilbert spaces: Theory and applications,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 7, pp. 1741–1754, Jul. 2020.

[79] Y.-W. Luo, C.-X. Ren, D.-Q. Dai, and H. Yan, ''Unsupervised domain adaptation via discriminative manifold propagation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1653–1669, Mar. 2022.

[80] Y.-W. Luo and C.-X. Ren, ''Conditional Bures metric for domain adaptation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13984–13993.

[81] W. Zhang, W. Ouyang, W. Li, and D. Xu, ''Collaborative and adversarial network for unsupervised domain adaptation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3801–3809.

[82] C.-Y. Lee, T. Batra, M. H. Baig, and D. Ulbricht, ''Sliced Wasserstein discrepancy for unsupervised domain adaptation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10277–10287.

[83] S. Sankaranarayanan, Y. Balaji, C. D. Castillo, and R. Chellappa, ''Generate to adapt: Aligning domains using generative adversarial networks,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8503–8512.

[84] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, ''Maximum classifier discrepancy for unsupervised domain adaptation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3723–3732.

[85] O. Bousquet, S. Boucheron, and G. Lugosi, ''Introduction to statistical learning theory,'' in *Summer School on Machine Learning*. Cham, Switzerland: Springer, 2003, pp. 169–207.

[86] X. Wu, S. Zhang, Q. Zhou, Z. Yang, C. Zhao, and L. J. Latecki, ''Entropy minimization versus diversity maximization for domain adaptation,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 6, pp. 2896–2907, Jun. 2021.

[87] A. T. Nguyen, T. Tran, Y. Gal, P. H. Torr, and A. G. Baydin, ''KL guided domain adaptation,'' 2021, *arXiv:2106.07780*.

[88] T.-D. Truong, R. T. N. Chappa, X.-B. Nguyen, N. Le, A. P. G. Dowling, and K. Luu, ''OTAdapt: Optimal transport-based approach for unsupervised domain adaptation,'' in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 2850–2856.

[89] M. Li, Y.-M. Zhai, Y.-W. Luo, P.-F. Ge, and C.-X. Ren, ''Enhanced transport distance for unsupervised domain adaptation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13933–13941.

[90] L. Van der Maaten and G. Hinton, ''Visualizing data using t-SNE,'' *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–12, 2008.

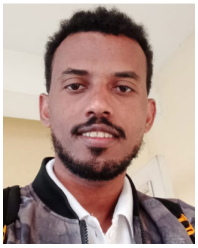
[91] R. K. Sanodiya, J. Mathew, B. Paul, and B. A. Jose, ''A kernelized unified framework for domain adaptation,'' *IEEE Access*, vol. 7, pp. 181381–181395, 2019.

[92] O. Gilo, J. Mathew, S. Mondal, and R. K. Sanodiya, ''Unsupervised sub-domain adaptation using optimal transport,'' *J. Vis. Commun. Image Represent.*, vol. 94, Jun. 2023, Art. no. 103857.

[93] S. Masashi, M. Krauledat, and K.-R. Müller, ''Covariate shift adaptation by importance weighted cross validation,'' *J. Mach. Learn. Res.*, vol. 8, no. 5, pp. 1–20, 2007.

[94] S. Bickel, M. Brückner, and T. Scheffer, ''Discriminative learning under covariate shift,'' *J. Mach. Learn. Res.*, vol. 10, pp. 2137–2155, 2009.

[95] Y. H. Tsai, Y.-R. Yeh, and Y. F. Wang, ''Learning cross-domain landmarks for heterogeneous domain adaptation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5081–5090.



OBSA GILO received the B.Sc. and M.Sc. degrees in computer science from Wallaga University, Ethiopia, in 2014 and 2018, respectively. He is currently pursuing the Ph.D. degree in computer science and engineering with the Indian Institute of Technology Patna, India. He was a Lecturer for more than four years with Wallaga University. He published and contributed to reputable journals and conferences. His current research interests include computer vision, machine learning, pattern recognition, transfer learning, and domain adaptation.

JIMSON MATHEW (Senior Member, IEEE) received the master's degree in computer engineering from Nanyang Technological University (NTU), Singapore, and the Ph.D. degree in computer engineering from the University of Bristol, Bristol, U.K. Throughout his career, he has held positions at various prestigious institutions, including the Centre for Wireless Communications, National University of Singapore; Bell Laboratories Research, Lucent Technologies North Ryde, Australia; the Royal Institute of Technology KTH, Stockholm, Sweden; and the Department of Computer Science, University of Bristol. He was the Head of the Department of Computer Science and Engineering, Indian Institute of Technology Patna (IIT Patna), India. He is currently a Professor with the Computer Science and Engineering Department, IIT Patna. He has made significant contributions to the field of computer engineering and has a strong academic portfolio. He holds multiple patents, has coauthored three books, and has published over 100 papers in renowned international journals and conferences. His research interests include fault-tolerant computing, hardware security, large-scale integration design, and design automation. His expertise and contributions have had a notable impact on computer engineering. He is a member of The Institution of Engineering and Technology (IET).

SAMRAT MONDAL (Senior Member, IEEE) received the Ph.D. degree in computer science and engineering from the School of Information Technology, Indian Institute of Technology Kharagpur, Kharagpur, India, in 2010. He was a Visiting Faculty with the University of Denver, Colorado, CO, USA, for 11 months. He was a Visiting Research Scholar with National Semiconductor Corporation, Santa Clara, CA, USA. Since December 2010, he has been a Faculty Member with the Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna, India. He received Research Grants from the Science and Engineering Research Board, Government of India, on multiple occasions. He has authored or coauthored several research papers in reputed international journals and conferences. His research interests include security and privacy, database and data mining applications, and smart energy management-related applications.

RAKESH KUMAR SANODIYA (Member, IEEE) received the Ph.D. degree from the Indian Institute of Technology (IIT), Patna, in 2019, where he developed state-of-the-art domain adaptation approaches for image classification. He was a Research Assistant Professor with the National Taipei University of Technology, Taiwan. He is currently an Assistant Professor with the Indian Institute of Information Technology Sri City, Chittoor. He is an accomplished academician with expertise in the field of domain adaptation for image classification. His research has focused on developing feature and instance re-weighting based on semi-supervised and unsupervised domain adaptation frameworks for visual image classification. He has published his research results in several reputable journals, including *Knowledge-Based Systems*, *Expert Systems with Applications*, *Soft Computing*, *Applied Intelligence*, *Image and Vision Computing*, *Sensors*, and IEEE ACCESS, among others. He has presented his research findings at various prestigious conferences, such as IEEE CEC, IEEE IJCNN, ICONIP, INDICON, and many others.