International Journal of Computer Vision
https://doi.org/10.1007/s11263-024-02178-5

Weighted Joint Distribution Optimal Transport Based Domain Adaptation for Cross-Scenario Face Anti-Spoofing

Shiyun Mao[1] · Ruolin Chen[1] · Huibin Li[1]

Received: 25 September 2023 / Accepted: 3 July 2024

Abstract
Unsupervised domain adaptation-based face anti-spoofing methods have attracted more and more attention due to their promising generalization abilities. To mitigate domain bias, existing methods generally attempt to align the marginal distributions of samples from source and target domains. However, the label and pseudo-label information of the samples from source and target domains are ignored. To solve this problem, this paper proposes a Weighted Joint Distribution Optimal Transport unsupervised multi-source domain adaptation method for cross-scenario face anti-spoofing (WJDOT-FAS). WJDOT-FAS consists of three modules: joint distribution estimation, joint distribution optimal transport, and domain weight optimization. Specifically, the joint distributions of the features and pseudo labels of multi-source and target domains are firstly estimated based on a pre-trained feature extractor and a randomly initialized classifier. Then, we compute the cost matrices and the optimal transportation mappings from the joint distributions related to each source domain and the target domain by solving Lp-L1 optimal transport problems. Finally, based on the loss functions of different source domains, the target domain, and the optimal transportation losses from each source domain to the target domain, we can estimate the weights of each source domain, and meanwhile, the parameters of the feature extractor and classifier are also updated. All the learnable parameters and the computations of the three modules are updated alternatively. Extensive experimental results on four widely used 2D attack datasets and three recently published 3D attack datasets under both single- and multi-source domain adaptation settings (including both close-set and open-set) show the advantages of our proposed method for cross-scenario face anti-spoofing.

Keywords Cross-scenario face anti-spoofing · Multi-source domain adaptation · Joint distribution optimal transport · Domain weight learning

1 Introduction

In recent years, face recognition (FR) techniques have been used in various identity authentication scenarios. However, existing FR systems are vulnerable to spoofing attacks such as printed photos, video replay, 3D facial masks, adversarial attacks, etc (Yu et al., 2022). To secure FR systems from various physical attacks, both the communities of industry and academia have paid increasing attention to face anti-spoofing (FAS). In the past two decades, various FAS methods have been proposed including both traditional methods and deep learning-based methods (Yu et al., 2022). Traditional methods based on handcraft descriptors (Komulainen et al., 2013; Patel et al., 2016) can be further classified into texture-based, motion-based, and image analysis-based methods. Subsequently, hybrid (handcrafted + deep learning) (Rehman et al., 2020; Khammari, 2019) and end-to-end deep learning-based methods (Liu et al.,2018 (Yu et al., 2020; Zhang et al., 2020) have also been proposed.

However, the performance of most FAS methods drops significantly in cross-scenario settings due to variations in lighting, facial appearance, or camera quality. In view of this, most existing solutions (Liu et al., 2022; Wang et al., 2020; Jia et al., 2020; Wang et al., 2023, 2022, 2021; Chen et al., 2021; Jiang et al., 2023) focus on improving the cross-scenario capability of deep FAS models by using

Communicated by Segio Escalera.

✉ Huibin Li
huibinli@xjtu.edu.cn

Shiyun Mao
maoshiyun@stu.xjtu.edu.cn

Ruolin Chen
crl1999@stu.xjtu.edu.cn

[1] School of Mathematics and Statistics, Xi’an Jiaotong University, Xi’an 710049, China

Published online: 11 August 2024

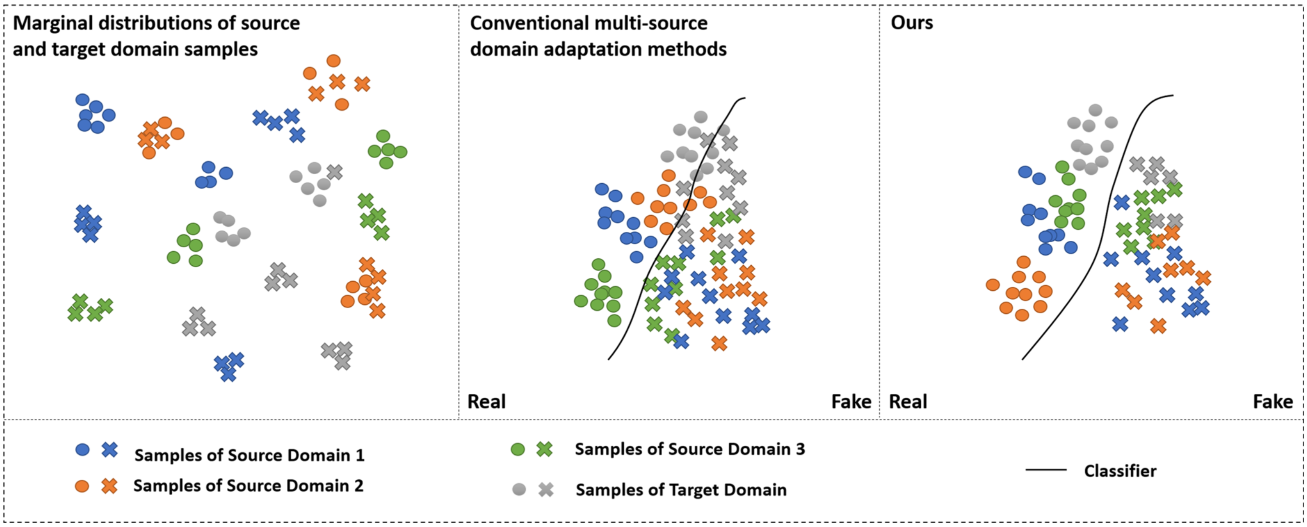

Marginal distributions of source and target domain samples
Conventional multi-source domain adaptation methods
Ours
Real
Fake
Real
Fake
Samples of Source Domain 1
Samples of Source Domain 2
Samples of Source Domain 3
Samples of Target Domain
Classifier

Fig. 1 Left: The original marginal distributions of the samples from three source domains and one target domain. Middle: Conventional multi-source DA methods aim to align the marginal distributions of source and target domain samples to learn a common feature space, which may fail to get a discriminative class boundary. Right: Our proposed multi-source DA method aims to align the joint distributions of sample features and their corresponding labels between source and target domains, which has the potential to learn a more discriminative class boundary

multi-source domain generalization (DG) approach, which assumes that there exists a potential generalized feature space between the given source domains and unseen target domain. By adapting multiple source data to learn a common feature space, the model trained in source domains can be well generalized to the unseen target domain. However, in practice, a large amount of unlabeled facial images are available from existing FR systems, and domain adaptation (DA) forms a natural learning framework for FAS. DA approach attempts to aid cross-scenario FAS by extracting discriminative feature representations from labeled source data and unlabeled target data. Thus, they can exploit rich information in the unlabeled target domain and obtain a more robust decision boundary.

In most DA methods, the distributions of source and target features are matched in a learned feature space, by using Maximum Mean Discrepancy (MMD) (Pei et al., 2018; Rahman et al., 2020), Correlation Alignment (CORAL) (Baochen et al., 2016) or Kullback-Leiber divergence (KL) (Zhuang et al., 2015). Besides, another direction is based on adversarial training (Tzeng et al., 2017), where a discriminator (domain classifier) is trained to distinguish between the source and target representations. However, considering that there are not only large inter-class differences in the samples of each domain, intra-class differences are still obvious, and there are cases where samples with different labels in different domains are closer to each other than samples with the same label in different domains, as illustrated in the left of Fig. 1. So only considering fitting the feature distributions of the source and target domains will have a situation similar to fitting the features of real samples from source domains to fake samples from the target domain, as illustrated in the middle of Fig. 1, which is not conducive to classification. Therefore, different from existing DA-based FAS methods, which attempt to align the marginal distributions in the feature space between the source and target domains. In this paper, we consider the discrepancy in the joint distributions of features and labels of source and target data. In this way, the samples in the source and target domains are aligned based on both features and labels, so that the samples with different labels from the same domain will be separated, while samples with the same label from different domains will be aggregated, as illustrated in the right of Fig. 1.

The main idea of this paper is to find optimal transportation mappings between the product spaces (including features and labels) of each source domain and the product space (including features and pseudo-labels) of the target domain. In this case, we first compute the cost matrices based on the joint distributions of each source domain and the target domain and then compute the optimal transportation mappings while reducing the discrepancy between the joint distributions. The distribution inconsistencies are measured by the Wasserstein distances (Cuturi et al., 2014). After obtaining the optimal transportation mappings, we learn a convex combination of the joint distributions of source domains, which allows us to distribute the masses based on the similarities of the sources with the target, both in the feature and pseudo-label spaces. Domain weights are updated together with the parameters of the feature extractor and classifier by training the weighted transportation loss between each source domain and the target domains, with the weighted source domain classification loss and the target domain entropy loss. Here, the target domain entropy loss is used to adjust the parameters of the feature extractor and classifier adaptively to further fit the distribution of the target domain. Our idea of aligning the joint distributions is reflected in the definition of cost matrix and reacted on the transportation mapping. The reduction of domain discrepancy in our method is reflected in seeking domain-invariant product space of features and labels, rather than feature space. In fact, the single-source domain joint distribution optimal transport is a degeneration of the multi-source domain joint distribution optimal transport, except that the weight of the unique source domain is always equal to 1 and we do not need to train it when training the feature extractor and classifier. Our main contributions to this work can be summarized as follows:

- Facing the cross-scenario FAS problem, we propose to reduce the discrepancy of domain distributions based on the joint distribution, which is dedicated to aligning the joint distributions of both sample features and labels (or pseudo-labels) of source and target domains in the common product space, which is largely different from existing methods.
- To solve the multi-source DA-based FAS, we propose to utilize the Wasserstein distance to measure the distances between the joint distributions, and assign adaptively updated weights to each source domain based on the Wasserstein distances so as to take into account the contributions of different source domains to the target domain.
- Extensive experimental results on four widely used 2D attack datasets and three recently published 3D attack datasets under both single- and multi-source domain adaptation settings (including both close-set and open-set) show the advantages of our proposed method for cross-scenario FAS. Our method achieves state-of-the-art results in all three protocols under the single-source setting, and under the multi-source setting, except for the 2D→2D protocol, which achieves the second-best performance, the remaining two protocols also achieve state-of-the-art results.

2 Related Works

In this section, we will first introduce the DA-based methods for FAS. After that, the focus will be on reviewing the optimal transport-based DA methods and multi-source DA methods that are most relevant to our work.

2.1 Domain Adaptation for Face Anti-spoofing

The basic idea of the DA technique is to mitigate the distribution discrepancy between the source and target domains so that the model trained with the labeled source data can be well adapted to the unlabeled target data. Initially, a maximum mean discrepancy (MMD) based metric learning method is proposed for FAS to align the distributions of source features and target features (Li et al., 2018). Other major developments have focused on the inclusion of adversarial loss functions that drive the inability of CNNs to distinguish whether a sample is from the source or target domain (Wang et al., 2019, 2021; Jia et al., 2021). Specifically, Wang et al. (2021) proposed ML-Net using the combination of center loss and triplet loss jointly to learn a feature representation for source data, then they adapted this representation to the target domain via UDA-Net and DR-Net. Jia et al. (2021) designed a marginal distribution alignment module (MDA) for domain-invariant feature learning and a conditional distribution alignment module (CDA) for centroid alignment of labeled features. In addition, Zhou et al. (2022) reformulated the unsupervised DA-based FAS as a domain stylization problem. The target data is stylized with the source domain style through image translation to directly fit the target data to the source model. Yue et al. (2022) presented a cyclically disentangled feature translation network, and proposed to generate pseudo-labeled images to train a generalizable classifier. Li et al. (2022) proposed a teacher-student framework to improve the cross-domain performance of FAS through single-class DA. Overall, most of these methods require multiple stages of the training process and all of them only consider aligning the marginal distributions, ignoring the role of source labels.

As we know, CASIA-FASD (Zhang et al., 2012), Idiap Replay-Attack (Chingovska et al., 2012), MSU-MFSD (Wen et al., 2015), and OULU-NPU (Boulkenafet et al., 2017) datasets have been widely used to study the DA-based FAS. However, these datasets are limited in data scale and attack types (print and replay) and recorded in controlled indoor scenarios. Recently, many new FAS datasets have been released and there are three major trends in the development of datasets: (1) large-scale data amount, (2) increasing number of novel attack types and complex recording conditions, and (3) multiple modalities. For example, CASIA-SURF 3DMask (Yu et al., 2020) is the first FAS dataset considering outdoor scenes with challenging lighting and it includes three mask decorations (i.e., masks with/without hair and glasses) recorded under six environmental conditions. CASIA-SURF HiFiMask (Liu et al., 2022) dataset contains more than

50,000 videos and it includes 3D mask attacks with three kinds of materials (transparent, plaster, and resin) recorded under six lighting conditions and six indoor/outdoor scenes. And Surveillance High-Fidelity Mask (Fang et al., 2024) dataset is captured under 40 surveillance scenes, and it has 232 3D attacks (high-fidelity masks), 200 2D attacks (posters, portraits, and screens), and 2 adversarial attacks. Besides, CASIA-SURF (Zhang et al., 2019) and CASIA-SURF Cross-ethnicity Face Anti-spoofing (CeFA) (Liu et al., 2021) datasets contain 3 modalities, i.e., RGB, Depth and IR.

Yu et al. (2020) proposed a Neural Architecture Search (NAS)-based approach for FAS. They presented Domain/Type-aware Meta-NAS for leveraging cross-domain/type knowledge for robust searching to improve the transferability of NAS across datasets and unknown attack types. Liu et al. (2022) proposed a training method for supervised FAS tasks, i.e., contrasting context-aware learning framework, which accurately utilizes the rich context information (e.g., subjects, mask material, and illumination) between live face and high-fidelity mask attack pairs. Fang et al. (2024) proposed a Contrastive Quality-Invariance Learning network to mitigate the performance degradation of FAS methods caused by low-quality images in surveillance scenarios. These works have better FAS performance in single dataset scenarios, but have weak generalization ability and cannot effectively solve the DA-based 3D attack FAS. In this paper, we will study the DA-based FAS method dealing with both 2D and 3D attacks, and generalize it to open-set DA in which there are new types of attacks in the target domain that are different from the source domains.

2.2 Optimal Transport Based Domain Adaptation

The optimal transport problem is first introduced by the French mathematician Gaspard Monge in the middle of the 19th century as a way to find a minimal-effort solution to the transport of a given mass of dirt into a given hole. Kantorovich (2006) extended the Monge problem from the viewpoint of transport mapping to transportation plan. Later, new computational strategies have been proposed and make it possible to be used for the problem of DA (Courty et al., 2016, 2017; Damodaran et al., 2018). The core of optimal transport theory applied to the DA problem lies in learning the transformation between domains. In particular, Courty et al. (2016) proposed a regularized unsupervised optimal transport model to align the feature representations of source and target domains. They proposed two regularization schemes to encode the class structure in the source domain while estimating the transportation plan, thus reinforcing the intuition that the samples of the same class must undergo similar transformations. Subsequently, Courty et al. (2017) proposed to minimize the optimal transportation loss between the joint distribution of the source domain and the estimated joint distribution of the target domain. Later, this method was extended to deep learning frameworks (Damodaran et al., 2018) where the feature embedding is simultaneously estimated with the classifier by using an efficient stochastic optimization procedure. An important aspect of joint distribution optimal transport is that the optimization problem involves the joint distribution of both feature embeddings and sample labels, and the simultaneous use of feature and label information is the basis of most generalization bounds (Courty et al., 2017).

2.3 Multi-source Domain Adaptation

For the multi-source DA problem, Yishay et al. (2009) pointed out that learning a weighted combination of multiple source distributions can be better generalized to the target domain under a certain theoretical guarantee. Judy et al. (2018) proposed an algorithm for distribution-weighted combinatorial solutions based on square loss and cross-entropy loss to solve the multi-source DA problem. Recently, many deep learning networks designed specifically for multi-source domains have been proposed to solve the multi-source DA problem. Peng et al. (2019) proposed a multiple source domain adaptive moment matching network (M3SDA), which aims to transfer knowledge learned from multiple labeled source domains to an unlabeled target domain by dynamically aligning the moments of feature distributions. Zhao et al. (2018) proposed the Multi-Source Domain Adversarial Network (MDAN), which approaches the DA problem by optimizing the task-adaptive generalization bounds. Wen et al. (2020) pointed out that in order to achieve the optimal generalization upper bound for the target domain, a trade-off is needed between including all source domains to increase the number of valid samples and excluding less relevant domains to avoid negative transfer. Based on this theory, they proposed a domain aggregation network (DARN), which dynamically adjusts the weights of each source domain during the end-to-end training process. Xu et al. (2018) proposed a deep cocktail network (DCTN) to solve the problem of domain and category transfer between multiple sources. Kang et al. (2020) proposed the Contrast Adaptive Network (CAN), which optimizes a new metric, i.e. the contrast domain variance, explicitly modeling intra-class domain variance and inter-class domain variance. Besides, they utilized the weighting of the inversed classification loss of intra-domain samples as the domain weights for network updates. Li et al. (2021) proposed a multiple-source contribution learning network (MSCLDA) by considering source contributions when predicting a target task. This method can simultaneously learn the similarity and diversity of domains by extracting multi-view features and utilizes a metric based on MMD as the domain weights. Zhao et al. (2020) proposed a multi-source distillation network (MDDA), which not only considers different distances between multiple sources and targets but also investigates the different similarities between source samples and target samples. A metric based on the optimal transport distance is used as the domain weights. Most of these multi-source DA methods are based on feature distributions when measuring the discrepancy between source and target distributions, and these methods are not capable of adaptively adjusting source domain weights. In contrast to the above methods, Turrisi et al. (2022) exploited the diversity of source distributions by adjusting the weights of different source joint distributions to fit the target task, which aims to simultaneously find the optimal transport-based alignment between the source and target joint distributions, as well as the reweighting of the source distributions based on the transportation loss. Inspired by Turrisi et al. (2022), this paper adopts the idea of joint distribution optimal transport to solve the problem of single-source and multi-source DA-based cross-scenario FAS. To the best of our knowledge, this is the first work that uses the idea of weighted joint distribution optimal transport to solve the cross-scenario FAS.

3 Proposed Method

In this paper, we propose a weighted joint distribution optimal transport method for multi-source DA-based FAS (WJDOT-FAS). As shown in Fig. 2, the training phase of the proposed method consists of three modules, namely joint distribution estimation, joint distribution optimal transport, and domain weight optimization. In particular, given the labeled facial samples from K source domains and the unlabeled samples from the target domain, we first estimate the joint distributions of both samples' features and labels (or pseudo labels) for each domain by using a pre-trained feature extractor and a randomly initialized classifier. Then, the cost matrices between the joint distributions of each source domain and the target domain are computed by using a weighted distance metric of both feature space and label space. Once the cost matrices are estimated, we can compute the optimal transportation mappings from the joint distributions of each source domain and the target domain by solving Lp-L1 optimal transport problems. These optimal transportation mappings can map the joint distributions of each source domain and the target domain to a new common space, in which their domain discrepancies can be well-aligned. Considering that different source domains have different contributions to the target domain, domain weights are defined for each source domain, and these weights can be solved by solving a convex optimization problem related to the loss functions of different source domains, target domain, as well as the optimal transportation losses from each source domain to the target domain. Meanwhile, the parameters of the feature extractor and the classifier are also updated, and the learnable parameters and the computations of the three modules are updated alternatively. Once the feature extractor and classifier have been well-trained, they are used to predict the sample labels of the target domain in the testing phase. More details of the

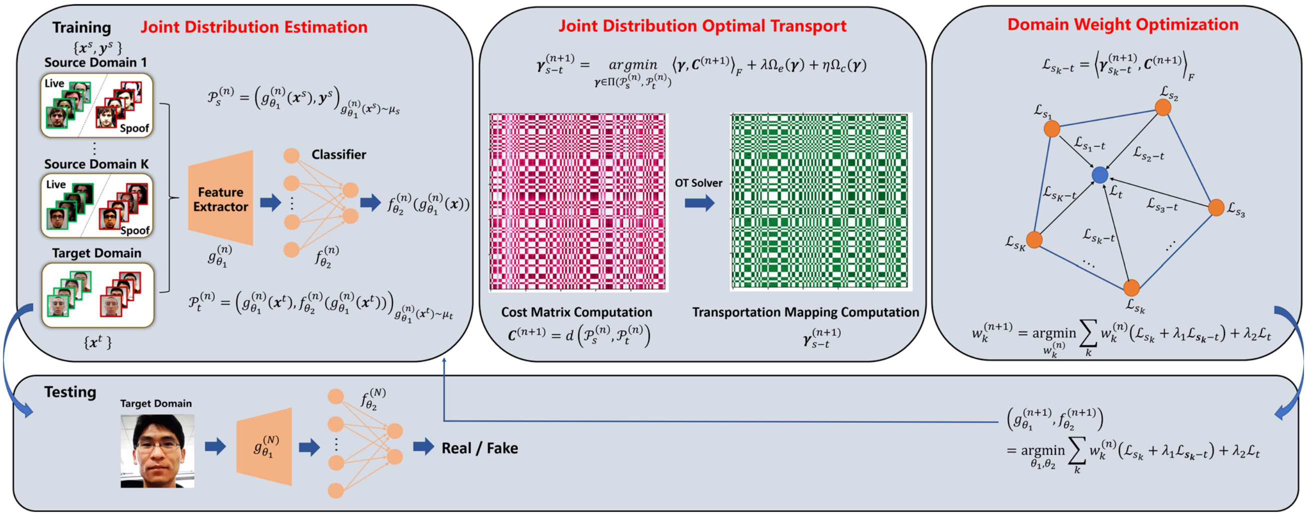

Training
Joint Distribution Estimation
$\{x^s, y^s\}$
Source Domain 1
Live
Spoof
Source Domain K
Live
Spoof
Target Domain
$\{x^t\}$
$\mathcal{P}_s^{(n)} = \left(g_{\theta_1}^{(n)}(x^s), y^s\right)_{g_{\theta_1}^{(n)}(x^s) \sim \mu_s}$
Feature Extractor
$g_{\theta_1}^{(n)}$
Classifier
$f_{\theta_2}^{(n)}$
$f_{\theta_2}^{(n)}(g_{\theta_1}^{(n)}(x))$
$\mathcal{P}_t^{(n)} = \left(g_{\theta_1}^{(n)}(x^t), f_{\theta_2}^{(n)}(g_{\theta_1}^{(n)}(x^t))\right)_{g_{\theta_1}^{(n)}(x^t) \sim \mu_t}$
Joint Distribution Optimal Transport
$\gamma_{s-t}^{(n+1)} = \underset{\gamma \in \Pi(\mathcal{P}_s^{(n)}, \mathcal{P}_t^{(n)})}{argmin} \langle \gamma, C^{(n+1)} \rangle_F + \lambda \Omega_e(\gamma) + \eta \Omega_c(\gamma)$
OT Solver
Cost Matrix Computation
$C^{(n+1)} = d\left(\mathcal{P}_s^{(n)}, \mathcal{P}_t^{(n)}\right)$
Transportation Mapping Computation
$\gamma_{s-t}^{(n+1)}$
Domain Weight Optimization
$\mathcal{L}_{s_k-t} = \left\langle \gamma_{s_k-t}^{(n+1)}, C^{(n+1)} \right\rangle_F$
$\mathcal{L}_{s_1}$
$\mathcal{L}_{s_2}$
$\mathcal{L}_{s_3}$
$\mathcal{L}_{s_K}$
$\mathcal{L}_{s_k}$
$\mathcal{L}_{s_1-t}$
$\mathcal{L}_{s_2-t}$
$\mathcal{L}_{s_3-t}$
$\mathcal{L}_{s_K-t}$
$\mathcal{L}_{s_k-t}$
$\mathcal{L}_t$
$w_k^{(n+1)} = \underset{w_k^{(n)}}{\text{argmin}} \sum_k w_k^{(n)}(\mathcal{L}_{s_k} + \lambda_1 \mathcal{L}_{s_k-t}) + \lambda_2 \mathcal{L}_t$
Testing
Target Domain
$g_{\theta_1}^{(N)}$
$f_{\theta_2}^{(N)}$
Real / Fake
$\left(g_{\theta_1}^{(n+1)}, f_{\theta_2}^{(n+1)}\right) = \underset{\theta_1, \theta_2}{\text{argmin}} \sum_k w_k^{(n)}(\mathcal{L}_{s_k} + \lambda_1 \mathcal{L}_{s_k-t}) + \lambda_2 \mathcal{L}_t$

Fig. 2 An overview of the proposed weighted joint distribution optimal transport method for multi-source DA-based cross-scenario FAS (WJDOT-FAS). The training phase of this method consists of three modules: joint distribution estimation, joint distribution optimal transport, and domain weight optimization. The joint distributions are determined by the feature extractor g_{θ_1} and the classifier f_{θ_2}. Transportation mappings $\boldsymbol{\gamma}_{s_k\text{-}t}$, and domain weights w_k are alternately updated by aligning the joint distributions of each source domain and the target domain. Once the parameters of g_{θ_1} and f_{θ_2} have been well trained, they are used to predict the sample labels of the target domain in the testing phase

proposed method will be introduced in the following paragraphs.

3.1 Joint Distribution Estimation

The aim of the joint distribution estimation module is to estimate the joint distributions of each source domain and the target domain. The joint distribution is defined in the product space of the sample feature space and sample label space. Given the labeled source data $\mathcal{D}_{s_k} = \left\{\boldsymbol{x}_{i_k}^{s_k}, \boldsymbol{y}_i^{s_k}\right\}_{i_k=1}^{n_{s_k}}$ ($k = 1, \dots, K$, K is the number of source domains) and unlabeled target data $\mathcal{D}_t = \left\{\boldsymbol{x}_j^t\right\}_{j=1}^{n_t}$, where n_{s_k} and n_t denote the sample numbers of the k-th source domain and the target domain. Our joint distribution estimation module is composed of two parts: a feature extraction function ($g : \boldsymbol{\mathcal{X}} \rightarrow \boldsymbol{\mathcal{Z}} \subseteq \mathbb{R}^d$) which maps the given facial samples from both source domains and the target domain into their feature space, and a classifier ($f : \boldsymbol{\mathcal{Z}} \rightarrow \boldsymbol{\mathcal{Y}} \subseteq \mathbb{R}^2$) which maps the sample features into their label space. The sample features of the k-th source domain and the target domain can be denoted as $\left\{z_{i_k}^{s_k}\right\}_{i_k=1}^{n_{s_k}}$, i.e. $\left\{g\left(\boldsymbol{x}_{i_k}^{s_k}\right)\right\}_{i_k=1}^{n_{s_k}}$ and $\left\{z_j^t\right\}_{j=1}^{n_t}$, i.e. $\left\{g\left(\boldsymbol{x}_j^t\right)\right\}_{j=1}^{n_t}$ respectively. Suppose we define μ_{s_k} and μ_t as the marginal feature distributions of the k-th source domain and the target domain, since the facial samples are in discrete form, we consider the empirical versions of μ_{s_k} and μ_t, which can be defined in the following forms:

$$\hat{\mu}_{s_k} = \frac{1}{n_{s_k}} \sum_{i_k} \delta_{z_{i_k}^{s_k}}, \tag{1}$$

$$\hat{\mu}_t = \frac{1}{n_t} \sum_j \delta_{z_j^t}, \tag{2}$$

where $\delta_{z_{i_k}^{s_k}}$ and $\delta_{z_j^t}$ are the Dirac functions at points $z_{i_k}^{s_k} \in \mathbb{R}^d$ and $z_j^t \in \mathbb{R}^d$ respectively.

Following above notations, we assume there exit two distinct joint probability distributions $\mathcal{P}_{s_k} = (z^{s_k}, \boldsymbol{y}^{s_k})_{z^{s_k} \sim \mu_{s_k}}$ and $\mathcal{P}_t = \left(z^t, f(z^t)\right)_{z^t \sim \mu_t}$, whose empirical versions can be defined in the following forms:

$$\hat{\mathcal{P}}_{s_k} = \frac{1}{n_{s_k}} \sum_{i_k} \delta_{z_{i_k}^{s_k}, \boldsymbol{y}_{i_k}^{s_k}}, \tag{3}$$

$$\hat{\mathcal{P}}_t = \frac{1}{n_t} \sum_j \delta_{z_j^t, f\left(z_j^t\right)}, \tag{4}$$

where $\delta_{z_{i_k}^{s_k}, \boldsymbol{y}_{i_k}^{s_k}}$ and $\delta_{z_j^t, f(z_j^t)}$ are the Dirac functions at points $\left(z_{i_k}^{s_k}, \boldsymbol{y}_{i_k}^{s_k}\right) \in \mathbb{R}^{d+2}$ and $\left(z_j^t, f\left(z_j^t\right)\right) \in \mathbb{R}^{d+2}$ respectively.

In particular, we use a pre-trained ResNet-18 CNN (or Transformer) backbone to extract the deep features of the given facial samples and use a randomly initialized classifier to compute pseudo labels. The joint distributions of the source samples are estimated by using the sample features extracted by the feature extractor and the true labels; while the joint distribution of the target samples is estimated by using the sample features extracted by the feature extractor and the pseudo-labels computed by the classifier.

3.2 Joint Distribution Optimal Transport

3.2.1 Cost Matrix Computation

Optimal transport (OT) (Cédric et al., 2008) is an efficient way of seeking to transform one distribution into another for a given cost function. It can be used for computing Wasserstein distance between probability distributions. Formally, OT searches a transportation mapping $\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\hat{\mathcal{P}}_s, \hat{\mathcal{P}}_t)$ between two distributions $\hat{\mathcal{P}}_s$ and $\hat{\mathcal{P}}_t$ which yields a minimal displacement cost. In a discrete setting (both distributions are empirical), the Wasserstein distance between $\hat{\mathcal{P}}_s$ and $\hat{\mathcal{P}}_t$ calculated by the OT method can be expressed in the following form:

$$W(\hat{\mathcal{P}}_s, \hat{\mathcal{P}}_t) = \min_{\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\hat{\mathcal{P}}_s, \hat{\mathcal{P}}_t)} \langle \boldsymbol{\gamma}, \boldsymbol{C} \rangle_F. \tag{5}$$

Here, $\langle \cdot, \cdot \rangle_F$ is the Frobenius matrix norm, $\boldsymbol{C} \in \mathbb{R}^{n_s \times n_t}$ is the cost matrix representing the pairwise costs of the joint distributions of source domain samples and the target domain samples. $\boldsymbol{\Pi}(\hat{\mathcal{P}}_s, \hat{\mathcal{P}}_t)$ describes the space of joint probability distributions of source and target domains and $\boldsymbol{\gamma}$ is the transportation mapping which is a matrix of size $n_s \times n_t$.

Joint distribution optimal transport is applied to our method, which is reflected in the definition of the cost matrix $\boldsymbol{C}$. The underlying idea is to align the joint distributions of features and labels from source and target domains instead of only considering the marginal distributions of features. Next, we will illustrate how to calculate $\boldsymbol{C}$ under joint distributions in the case where only one source domain is available. The cost matrix $\boldsymbol{C}$ associated with the product space of features and labels can be expressed as the gap between the joint distributions of the source and target domains, that is:

$$\boldsymbol{C} \triangleq d(\hat{\mathcal{P}}_s, \hat{\mathcal{P}}_t), \tag{6}$$

Specifically, the element of the i-th row and j-th column in $\boldsymbol{C}$ can be expressed as a joint cost measure of costs in the feature and label spaces of the i-th source sample and j-th target sample, combining both the gap between sample features and the discrepancy between sample labels (pseudo labels for the target domain). According to Damodaran et al. (2018), the specific form of $\boldsymbol{C}_{ij}$ is defined as follows:

$$\begin{aligned} \boldsymbol{C}_{ij} &\triangleq c\big(g\left(\boldsymbol{x}_i^s\right), \boldsymbol{y}_i^s; g\left(\boldsymbol{x}_j^t\right), f\left(g\left(\boldsymbol{x}_j^t\right)\right) \\ &= \| g\left(\boldsymbol{x}_i^s\right) - g\left(\boldsymbol{x}_j^t\right) \|^2 + \beta \mathcal{L}_{CE}\left(\boldsymbol{y}_i^s, f\left(g\left(\boldsymbol{x}_j^t\right)\right)\right), \end{aligned} \tag{7}$$

where $\| g(\boldsymbol{x}_i^s) - g(\boldsymbol{x}_j^t) \|^2$ compares the compatibility of the features for source and target samples and it is a l_2^2 distance; while $\mathcal{L}_{CE}(\boldsymbol{y}_i^s, f(g(\boldsymbol{x}_j^t)))$ is a cross-entropy loss, which considers the gap between the true label of the i-th source sample and the pseudo label of the j-th target sample. Parameter β is a scalar value weighing the strength of label cost relative to feature cost. The definition of $\boldsymbol{C}_{ij}$ in Eq. (7) guarantees that our optimal transport is defined under the joint distribution setting. If we only consider aligning the marginal distributions of source and target domain features, then $\boldsymbol{C}_{ij} = \| g(\boldsymbol{x}_i^s) - g(\boldsymbol{x}_j^t) \|^2$, i.e. the basic form of the cost matrix in OT.

3.2.2 Transportation Mapping Computation

In this section, we will introduce how to compute the transportation mapping $\boldsymbol{\gamma}$, considering the case of single-source DA. As shown in Eq. (5), OT searches a transportation mapping $\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\hat{\mathcal{P}}_s, \hat{\mathcal{P}}_t)$ between two distributions $\hat{\mathcal{P}}_s$ and $\hat{\mathcal{P}}_t$, where $\boldsymbol{\Pi}(\hat{\mathcal{P}}_s, \hat{\mathcal{P}}_t)$ can be expressed mathematically in the following form:

$$\begin{aligned}\boldsymbol{\Pi}(\hat{\mathcal{P}}_s, \hat{\mathcal{P}}_t) = \{&\boldsymbol{\gamma} \in (\mathbb{R}^+)^{n_s \times n_t} \mid \\ &\boldsymbol{\gamma}\mathbf{1}_{n_t} = \hat{\mathcal{P}}_s, \boldsymbol{\gamma}^\top \mathbf{1}_{n_s} = \hat{\mathcal{P}}_t\},\end{aligned} \tag{8}$$

where $\mathbf{1}_{n_s}$ and $\mathbf{1}_{n_t}$ are the n_s and n_t-dimension vectors of ones. With the definition of $\boldsymbol{C}_{ij}$ in Eq. (7), we can compute the transportation mapping based on the following equation:

$$\hat{\boldsymbol{\gamma}}_0 = \underset{\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\hat{\mathcal{P}}_s, \hat{\mathcal{P}}_t)}{\arg\min} \ \langle \boldsymbol{\gamma}, \boldsymbol{C} \rangle_F. \tag{9}$$

Equation (9) is a linear programming problem and can be solved by the network simplex algorithm, but solving it becomes difficult when the sample size is large. To solve this problem more efficiently, the entropy regularized version of the above optimal transport problem is proposed (Chingovska et al., 2012) and can be formulated as follows:

$$\hat{\boldsymbol{\gamma}}_0^\lambda = \underset{\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\hat{\mathcal{P}}_s, \hat{\mathcal{P}}_t)}{\arg\min} \ \langle \boldsymbol{\gamma}, \boldsymbol{C} \rangle_F + \lambda \Omega_e(\boldsymbol{\gamma}), \tag{10}$$

where $\Omega_e(\gamma) = \sum_{i,j} \boldsymbol{\gamma}(i,j) \log \boldsymbol{\gamma}(i,j)$ computes the negative entropy of $\boldsymbol{\gamma}$. This regularization is introduced because $\hat{\boldsymbol{\gamma}}_0$, as a solution of the linear program, most of the elements are zero, and thus a smoother version of the transport can be found by increasing the entropy, thus reducing its sparsity. In particular, $\hat{\gamma}_0^\lambda$ can be solved by using Sinkhorn algorithm (Cuturi et al., 2013).

Further, we resort to a class regularization term to estimate a better transport using the source sample label information. Our goal is to penalize the coupling of matching source samples with different labels to the same target sample. Thereby, the new optimization problem can be written in the following form:

$$\hat{\boldsymbol{\gamma}}_0^\eta = \underset{\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\hat{\mathcal{P}}_s, \hat{\mathcal{P}}_t)}{\arg\min} \ \langle \boldsymbol{\gamma}, \boldsymbol{C} \rangle_F + \lambda \Omega_e(\boldsymbol{\gamma}) + \eta \Omega_c(\boldsymbol{\gamma}), \tag{11}$$

where $\eta \geq 0$ and $\Omega_c(\cdot)$ is the class regularization term. In this work, we use group sparse regularization with the aim of making a given target sample receive masses from source samples with the same label. This regularization term is defined as:

$$\Omega_c(\boldsymbol{\gamma}) = \sum_j \sum_{cl} \|\boldsymbol{\gamma}(I_{cl}, j)\|_1^{1/2} \tag{12}$$

where $\|\cdot\|_1$ denotes the l_1 norm and I_{cl} contains the indices of rows in $\boldsymbol{\gamma}$ related to source domain samples of class cl. So, $\boldsymbol{\gamma}(I_{cl}, j)$ is a vector containing coefficients of the j-th column of $\boldsymbol{\gamma}$ associated to class cl. In our case, cl stands for real or fake. This regularization term is called the Lp-L1 regularization term (here, $p = 1/2$) (Courty et al., 2014), and the problem can be transformed into Eq. (10) when the maximization minimization technique is applied on the Lp-L1 parametrization and can be solved by using an efficient Sinkhorn-Knopp algorithm (Courty et al., 2016).

Equations (9), (10) and (11) are called EMD solver, Sinkhorn solver and Lp-L1 solver respectively. After calculating the optimal transportation mapping, the Wasserstein distance between the source and target domain distributions is obtained according to Eq. (5). By computing the transportation mapping under joint distribution optimal transport, samples with similar features and common labels can be matched in the common product space, resulting in better discrimination.

3.3 Domain Weight Optimization

To solve the multi-source DA-based FAS, the weighing of each source domain is an important factor for the generalization ability of the final classifier on the target domain. We propose to assign adaptively updated weights to each source domain based on the Wasserstein distances between the joint distributions of each source domain and the target domain. Besides, for the FAS classification problem, these weights can be computed by solving a convex optimization problem related to the Wasserstein distances (optimal transportation losses) between the joint distributions of each source domain and the target domain and the classification losses of different source domains.

The Wasserstein distances (optimal transportation losses) between the joint distributions of each source domain and the target domain can be computed by solving the optimal transport problems in Eq. (5). It can measure the degree of the joint distribution alignment between the source and target domains. It's not difficult to see that the better the distributions are aligned, the better the generalization effect on the target domain. Specifically, we first compute the cost matrices of the joint distributions between each source domain and the target domain samples by Eq. (7). Then, the optimal transportation mappings of the joint distributions from each source domain to the target domain are computed by Eq. (11). Finally, the Wasserstein distances (optimal transportation losses) between the joint distributions can be computed. The Wasserstein distance (optimal transportation loss) from the k-th source domain to the target domain is defined as:

$$\mathcal{L}_{s_k\text{-}t} = \sum_{i_k} \sum_{j} \hat{\boldsymbol{\gamma}}_{i_k j}^{s_k} d\left(g\left(\boldsymbol{x}_{i_k}^{s_k}\right), \boldsymbol{y}_{i_k}^{s_k}; g\left(\boldsymbol{x}_j^{t}\right), f\left(g\left(\boldsymbol{x}_j^{t}\right)\right)\right). \quad (13)$$

Moreover, to better utilize the source domain information to train the final classifier, we employ the adaptive cross-entropy (AdaCE) loss (Jia et al., 2021) to measure the classification error of the classifier for each source domain. AdaCE loss is defined by adjusting the weight of the cross-entropy loss adaptively based on the classification accuracy. For the k-th source domain, it can be defined as follows:

$$\begin{aligned} \mathcal{L}_{s_k} &= \frac{1}{n_{s_k}} \sum_{i_k} \mathcal{L}_{s_k}\left(\boldsymbol{y}_{i_k}^{s_k}, f\left(g\left(\boldsymbol{x}_{i_k}^{s_k}\right)\right)\right) \\ &= \frac{1}{n_{s_k}} \sum_{i_k} \left(1 - e^{-\mathcal{L}_{CE}\left(\boldsymbol{y}_{i_k}^{s_k}, f\left(g\left(x_{i_k}^{s_k}\right)\right)\right)}\right)^{\alpha}. \quad (14) \\ &\qquad \mathcal{L}_{CE}\left(\boldsymbol{y}_{i_k}^{s_k}, f\left(g\left(\boldsymbol{x}_{i_k}^{s_k}\right)\right)\right), \end{aligned}$$

where $\mathcal{L}_{CE}(\cdot,\cdot)$ is the cross-entropy loss, and α is a hyper-parameter. For the i-th sample of the k-th source domain, $\mathcal{L}_{CE}(\cdot,\cdot)$ is defined as:

$$\mathcal{L}_{CE}\left(\boldsymbol{y}_{i_k}^{s_k}, f\left(g\left(\boldsymbol{x}_{i_k}^{s_k}\right)\right)\right) = -\boldsymbol{y}_{i_k}^{s_k} \log\left(f\left(g\left(\boldsymbol{x}_{i_k}^{s_k}\right)\right)\right). \quad (15)$$

To further refine the parameters of the FAS classifier, we feed the unlabeled target domain data to the classifier and refer to the entropy loss proposed in Jia et al. (2021), which is expressed as follows:

$$\mathcal{L}_t = -\frac{1}{n_t} \sum_{j} f\left(g\left(\boldsymbol{x}_j^{t}\right)\right) \log f\left(g\left(\boldsymbol{x}_j^{t}\right)\right). \quad (16)$$

Once the optimal transportation loss functions from different source domains to the target domain, the classification loss functions related to different source domains as well as the entropy loss function related to the target domain have been defined, we can compute the domain weights and update the network parameters of both feature extractor and classifier by solving the following convex optimization problem:

$$\left(g_{\theta_1}^{(n+1)}, f_{\theta_2}^{(n+1)}, w_k^{(n+1)}\right) = \underset{g_{\theta_1}^{(n)}, f_{\theta_2}^{(n)}, w_k^{(n)}}{\arg\min} \; \mathcal{L}_{total}, \quad (17)$$

$$\mathcal{L}_{total} = \sum_{k} w_k (\mathcal{L}_{s_k} + \lambda_1 \mathcal{L}_{s_k\text{-}t}) + \lambda_2 \mathcal{L}_t, \quad (18)$$

where λ_1 and λ_2 are the trade-off parameters, and w_k denotes the domain weight related to k-th source domain.

The domain weights are continuously updated together with the network parameters. Then, the updated networks are used for the joint distribution estimation of source and target domains again, further for the joint distribution optimal transport, and finally the domain weight optimization. That's to say, the three modules are alternated learned, and updated. Once the network parameters of the feature extractor and classifier have been well-trained, they are used to predict the sample labels of the target domain in the testing phase. The whole training process of the proposed weighted joint distribution optimal transport method for multi-source DA-based FAS is shown in Algorithm 1. It is worth noting that if only one source domain is available, w_k always equals 1.

Algorithm 1 Weighted joint distribution optimal transport method for multi-source DA-based FAS

```
Require: Source data $\mathcal{D}_{s_k}, k = 1, \cdots, K$; target data $\mathcal{D}_t$, hyper-
    parameters $\lambda, \eta, \beta, \lambda_1, \lambda_2$.
Initial $w_k^{(0)} = \frac{1}{K}$, parameters of $g_{\theta_1}^{(0)}$ and $f_{\theta_2}^{(0)}$;
repeat
    Sample mini-batches $\boldsymbol{x}_{i_k}^{s_k}$ from $\mathcal{D}_{s_k}$ and $\boldsymbol{x}_j^{t}$ from $\mathcal{D}_t$;
    Compute features $g_{\theta_1}^{(n)}(\boldsymbol{x}_{i_k}^{s_k})$ and $g_{\theta_1}^{(n)}(\boldsymbol{x}_j^{t})$;
    Compute labels $f_{\theta_2}^{(n)}(g_{\theta_1}^{(n)}(\boldsymbol{x}_{i_k}^{s_k}))$ and $f_{\theta_2}^{(n)}(g_{\theta_1}^{(n)}(\boldsymbol{x}_j^{t}))$;
    Fix $g_{\theta_1}^{(n)}$ and $f_{\theta_2}^{(n)}$, compute $\boldsymbol{C}^{(n+1)}$ by equation (6);
    Compute $\hat{\boldsymbol{\gamma}}_{i_k j}^{s_k (n+1)}$ by equation (11);
    Fix $\hat{\boldsymbol{\gamma}}_{i_k j}^{s_k (n+1)}$, update $g_{\theta_1}^{(n)}$, $f_{\theta_2}^{(n)}$, and $w_k^{(n)}$ by equation (17);
until convergence
```

4 Experimental Results

4.1 Datasets

To evaluate the effectiveness of the proposed WJDOT-FAS method for multi-source DA based FAS, we conducted experiments on four public datasets with only 2D attack types, namely CASIA-FASD (Zhang et al., 2012), Idiap Replay-Attack (Chingovska et al., 2012), MSU-MFSD (Wen et al., 2015) and OULU-NPU (Boulkenafet et al., 2017). For simplicity, they are denoted as C, I, M, and O in the following. Besides, we also conducted experiments on three large-scale public datasets with 3D attack types, namely CASIA-SURF 3DMask (Yu et al., 2020), CASIA-SURF HiFiMask (Liu et

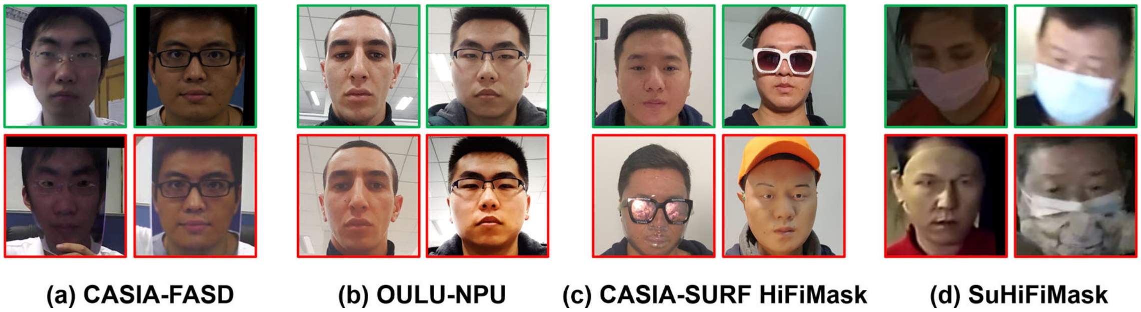


Fig. 3 Examples of the real (the first row) and fake (the second row) face from CASIA-FASD (Zhang et al., 2012), OULU-NPU (Boulkenafet et al., 2017), CASIA-SURF HiFiMask (Liu et al., 2022) and Surveillance High-Fidelity Mask (Fang et al., 2024) databases. It is easy to find that there exists a large inter-domain gap, such as the lighting, background, and types of attack, which results in significant distribution discrepancies among different datasets

al., 2022) and Surveillance High-Fidelity Mask (Fang et al., 2024). For simplicity, they are denoted as 3DMask, HiFiMask, and SuHiFiMask in the following. In addition, we also demonstrated the effectiveness of our WJDOT-FAS method for open-set DA by using datasets containing only 2D attack types as the source domains and a dataset containing 3D attack types as the target domain.

The *CASIA-FASD* (Zhang et al., 2012) dataset consists of 600 videos of real and attack attempts of 50 different subjects. There are three different image qualities in the dataset: low, normal, and high, which are captured with three different cameras (a Sony NEX-5 camera with 1280×720 resolution and two different USB cameras with 640×480 resolution). The face attacks include: distorted photo attacks, cut photo attacks, and video attacks. The *Idiap Replay-Attack* (Chingovska et al., 2012) dataset consists of 1200 videos of real and attack attempts on 50 different subjects. The camera on the MacBook is used to collect the dataset with a resolution of 320 × 240 under two conditions: (i) a control condition with a uniform background and fluorescent lights; and (ii) an unfavorable condition with a non-uniform background and daylight. Three types of deceptive attacks are designed: print attack, mobile attack, and high definition attack. The *MSU-MFSD* (Wen et al., 2015) dataset consists of 440 videos from 55 subjects. The face videos are taken by two types of cameras (MacBook Air camera and Google Nexus 5 Android phone camera). The resolutions are 640 × 480 and 720 × 480. There are mainly two different spoofing attacks, the print photo attack and the replay video attack. The *Oulu-NPU* (Boulkenafet et al., 2017) dataset consists of 4950 real and attack videos from 55 subjects. These videos are recorded with the front cameras of 6 mobile devices (Samsung Galaxy S6 edge, HTC Desire EYE, MEIZU X5, ASUS Zenfone Selfie, Sony XPERIA C5 Ultra Dual and OPPO N3). There are three different lighting conditions and background scenes. The types of presentation attacks are printing and video replay. These attacks are created using two printers and two display devices.

The *CASIA-SURF 3DMask* (Yu et al., 2020) dataset contains 288 real face videos and 864 mask videos from 48 subjects. Six conditions are used for data collection, including normal, back-light, front-light, side-light, outdoor in shadow, and outdoor in sunlight. 3D masks of 48 subjects are collected by 3D printing technology. In addition to the use of naive masks, two more realistic decorative situations (i.e., masks with/without hair and glasses) are considered. The *CASIA-SURF HiFiMask* (Liu et al., 2022) dataset consists of 75 subjects, and each subject provides high-fidelity plaster, resin, and transparent masks. Six different environments, six directional illuminations, and seven recording sensors are applied to the dataset. A total of 54,600 videos (13,650 live videos, 40,950 mask videos) are available in the dataset. The *Surveillance High-Fidelity Mask* (Fang et al., 2024) dataset is captured in 40 real-life surveillance scenarios, such as movie theaters, security gates, and parking lots, representing a wide range of face recognition scenarios. It includes 101 participants of different ages and genders who perform various natural activities in their daily lives. In addition, the dataset contains multiple types of spoofing attacks such as high-fidelity masks, 2D attacks, and adversarial attacks.

In general, there are differences in acquisition devices, acquisition conditions, and attack types for different datasets, which leads to discrepancies among domains; in addition, each dataset is collected by multiple acquisition devices and the attack types are diverse, which leads to a situation where samples of the same label within the dataset are distant from each other, so inter-domain joint distribution metric becomes inevitable. Figure 3 shows some examples of real and fake facial samples in these datasets. It is easy to see that there exists a large inter-domain gap, such as the lighting, background, and types of attack, which results in significant distribution discrepancies among different datasets.

4.2 Experimental Settings and Implementation Details

In this paper, we perform experiments on 2D and 3D attack datasets under single- and multi-source domain settings for (open-set) DA. We set up three protocols under single- and multi-source domain settings, respectively.

Under the **single-source domain** setting:

- **Cross-dataset testing on 2D attack datasets (2D→2D).** We follow the protocols of (Wang et al., 2021; Jia et al., 2021), in which one of the I, C, and M datasets is used as the source domain and the other dataset as the target domain, so there are six sets of experiments. We use the Half Total Error Rate (HTER) (half of the summation of false acceptance rate and false rejection rate) as the evaluation metric. We first compute the Equal Error Rate (EER) and the corresponding threshold on the development set and then utilize the threshold to calculate the HTER on the testing set.
- **Cross-dataset testing on 2D attack datasets (2D→2D).** One of the 3DMask and HiFiMask datasets is used as the source domain and the other dataset as the target domain. We use the HTER and the Area Under the Curve (AUC) as the evaluation metrics.
- **Cross-dataset testing for open-set DA (2D→3D).** One of the C and I datasets is used as the source domain and the SuHiFiMask dataset as the target domain. We also use the HTER and AUC as the evaluation metrics.

Under the **multi-source domain** setting:

- **Cross-dataset testing on 2D attack datasets (2D→2D).** We follow the protocols of (Zhou et al., 2022; Liu et al., 2022), in which three of the four datasets are used as source domains and the remaining one as target domain, so there are four sets of experiments. The HTER and AUC are used as the evaluation metrics.
- **Cross-dataset testing on 3D attack datasets (3D→3D).** Two of the 3DMask, HiFiMask, and SuHiFiMask datasets are used as the source domains, and the other dataset as the target domain. We use the HTER and AUC as the evaluation metrics.
- **Cross-dataset testing for open-set DA (2D→3D).** The C and I datasets are used as the source domains and one of the 3DMask, HiFiMask, and SuHiFiMask datasets is used as the target domain. We also use the HTER and AUC as the evaluation metrics.

In our experiments, we use the MTCNN algorithm (Zhang et al., 2016) for face detection and alignment. We implemented our WJDOT-FAS method on the PyTorch platform and utilized the ResNet-18 (He et al., 2016) and Vision Transformer (ViT) (Touvron et al., 2021)) pre-trained on ImageNet as our backbones. All detected face images are normalized to 256 × 256 × 3 and for the ResNet-18 backbone, we further resize them to 224 × 224 × 3. The network structures of our feature extractor and classifier are the same as (Jia et al., 2021) under ResNet-18 backbone and the same as (Liu et al., 2022) under ViT backbone. Specifically, the feature extractor outputs 512-dimensional features used for optimal transport. Both our JDOT-FAS-ResNet18 and JDOT-FAS-ViT models (including feature extractor and classifier) were trained by using the Adam optimizer with momentum of 1e-4 under the single-source DA setting. Besides, under the multi-source DA setting, our WJDOT-FAS-ResNet18 model was trained by using the Adam optimizer with momentum of 0.06 and weight decay of 2e-4, and our WJDOT-FAS-ViT model was trained by using the Adam optimizer with momentum of 1e-4. For both source and target domains, mini-batch sizes of $n_{s_k} = n_t = 120$ for ResNet-18 and $n_{s_k} = n_t = 60$ for ViT were used, and trained on a single NVIDIA RTX 3090 GPU. The weights of each source domain were also trained by using the Adam optimizer with a momentum of 0.006. The hyperparameters λ, η, β, λ_1, λ_2, and α are set to 0.1, 0.1, 0.1, 5, 0.1, and 2 respectively.

4.3 Comparisons with the State-of-the-Art Methods

4.3.1 Single-source DA Setting

To verify the effectiveness of our JDOT-FAS method under the single-source DA setting, we first compare it with the state-of-the-art FAS methods on the C, I, and M datasets with only 2D attack types. The methods we compare can be divided into three categories: traditional DA methods, including ADDA (Tzeng et al., 2017), DRCN (Ghifary et al., 2016) and DupGAN (Hu et al., 2018), DA based generalized FAS methods, including Li et al. (2018), ADA (Wang et al., 2019), UDA (Wang et al., 2021) and USDAN-Un (Jia et al., 2021), and some novel self-designed DA based deep learning FAS methods, including OCKD (Li et al., 2022), GDA (Zhou et al., 2022), SFDA-FAS (Liu et al., 2022), and CDFTN-R (Yue et al., 2022). The traditional DA methods generally judge a fake or real face by using a simple FC layer-based classifier optimized with cross-entropy loss. The DA-based generalized FAS methods are mainly based on Maximum Mean Discrepancy (MMD) (Li et al., 2018) and adversarial learning methods (Jia et al., 2021; Wang et al., 2019, 2021). The self-designed DA-based deep learning FAS methods are mainly based on some novel deep learning frameworks such as teacher-student learning (Li et al., 2022), generative DA (Zhou et al., 2022), contrastive learning (Liu et al., 2022), and disentangled representation learning (Yue et al., 2022). As shown in Table 1, in general, the DA-based generalized FAS

Table 1 Comparison results (HTER (%)) between the proposed method and the state-of-the-art methods for cross-dataset testing under the single-source DA setting on the C, I, and M datasets

Methods	C→I	C→M	I→C	I→M	M→C	M→I	Average
ADDA (Tzeng et al., 2017)	41.8	36.6	49.8	35.1	39.0	35.2	39.58
DRCN (Ghifary et al., 2016)	44.4	27.6	48.9	42.0	28.9	36.8	38.10
DupGAN (Hu et al., 2018)	42.4	33.4	46.5	36.2	27.1	35.4	36.83
Li et al. (2018)	39.2	14.3	26.3	33.2	10.1	33.3	26.10
ADA (Wang et al., 2019)	17.5	9.3	41.5	30.5	17.7	5.1	20.27
UDA (Wang et al., 2021)	15.6	9.0	34.2	29.0	16.8	3.0	17.93
USDAN-Un (Jia et al., 2021)	16.0	9.2	30.2	25.8	13.3	3.4	16.30
OCKD (Li et al., 2022)	**3.5**	15.0	31.9	20.8	26.7	29.0	21.15
GDA (Zhou et al., 2022)	15.1	**5.8**	29.7	20.8	12.2	2.5	14.40
SFDA-FAS (Liu et al., 2022)	11.5	10.4	**19.6**	24.1	10.0	3.7	13.20
CDFTN-R (Yue et al., 2022)	5.4	14.4	**8.7**	**12.9**	13.5	5.6	**10.08**
JDOT-FAS-ResNet18	9.9	8.3	27.0	**12.9**	**6.1**	**0.0**	10.70
JDOT-FAS-ViT	**3.3**	**5.4**	24.4	**5.8**	**4.4**	**0.5**	**7.30**

The best and the second best values are given in bold

methods perform better than traditional DA methods and the novel self-designed DA-based deep learning FAS methods obtain optimal performance. Our JDOT-FAS method belongs to the second category of the above methods and achieves the best average performance among all the DA-based FAS methods under ViT backbone, reducing the HTER by 2.78% compared to state-of-the-art DA method CDFTN-R (Yue et al., 2022). Our JDOT-FAS-ResNet18 model also has significant advantages over methods that also belong to the second category, reducing the HTER by more than 5.6%. The possible reason is that our method introduces label (pseudo-label) distance into the measure of distribution discrepancy by flexibly defining the cost matrix so that the labels can be taken into account when aligning the source and target distributions. For other DA-based FAS methods, the distributions are defined in the feature space, ignoring the role of source sample labels as well as target sample pseudo-labels in distribution alignment. In particular, the feature extractors of the methods (Wang et al., 2019, 2021) are non-parameter shared, while we use a parameter-shared feature extractor and classifier that map both source data and target data to a shared common product space, thus facilitating the search for a domain-invariant product space of features and labels that is more generalizable. Overall, our JDOT-FAS method achieves competitive results under the single-source DA setting.

Besides, we verify the effectiveness of our JDOT-FAS method under the single-source DA setting on the 3DMask and HiFiMask datasets with 3D attacks. As shown in Table 2, under both ResNet-18 and ViT frameworks, the effect of our JDOT-FAS method is improved compared to the baseline methods, and the effectiveness of our method is superior to the methods in Liu et al. (2022) because they lack the use of information about the target domain samples in the training process. In addition, for the 3D attack datasets, the features extracted using ResNet-18 are more capable of capturing the real and fake information in the face images, and the features are more conducive to the correct classification of the samples in the target domain after the joint distribution optimal transportation mapping. Therefore, our JDOT-FAS method not only has better generalization ability for the DA-based FAS with 2D attacks but also can achieve cross-scenario generalization for the DA-based FAS with 3D attacks.

In addition, to verify the effectiveness of our JDOT-FAS method under the open-set single-source DA setting, we choose the C or I dataset as the source domain which has only 2D attack types, and use the SuHiFiMask dataset with 2D, 3D, and adversarial attacks as the target domain. As shown in Table 3, under both ResNet-18 and ViT frameworks, the effect of our JDOT-FAS method is improved compared to the baseline methods and outperforms the experimental results under all the backbones in Fang et al. (2024). The improvement is more obvious under the ViT framework, with the average HTER reduced by 11.55% and the average AUC improved by 13.4%. This indicates that our JDOT-FAS method is also effective for the open-set DA problem with novel attacks in the target domain, i.e., for the case where the distribution discrepancy between the source and target domains is large, our proposed optimal transportation loss of joint distribution can reduce the domain discrepancy to a certain extent, and improve the classification accuracy on the target domain with novel attacks.

4.3.2 Multi-source DA Setting

To verify the effectiveness of our WJDOT-FAS method under the multi-source DA setting, we first compare it with the state-of-the-art FAS methods on the C, I, M, and O datasets

Table 2 Comparison results (HTER (%) and AUC (%)) between the proposed method and the state-of-the-art methods for cross-dataset testing under the single-source DA setting on the 3DMask and HiFiMask datasets

	HiFiMask→3DMask		3DMask→HiFiMask		Average	
Methods	HTER	AUC	HTER	AUC	HTER	AUC
ResNet50 w/ CCL (Liu et al., 2022)	25.4	82.0	–	–	–	–
CDCN w/ CCL (Liu et al., 2022)	14.0	93.3	–	–	–	–
Aux.(Depth) w/ CCL (Liu et al., 2022)	16.3	90.7	–	–	–	–
Baseline-ResNet18	14.6	91.1	34.8	70.8	24.70	80.95
Baseline-ViT	25.7	83.8	37.4	67.1	31.55	75.20
JDOT-FAS-ResNet18	**6.7**	**98.5**	**25.1**	**84.0**	**15.90**	**91.25**
JDOT-FAS-ViT	**11.1**	**97.0**	**28.3**	**78.0**	**19.70**	**87.50**

The best and the second best values are given in bold

Table 3 Results (HTER (%) and AUC (%)) of testings on the SuHiFiMask dataset with the C or I dataset as the source domain under the single-source DA setting

	C→SuHiFiMask		I→SuHiFiMask		Average	
Methods	HTER	AUC	HTER	AUC	HTER	AUC
ResNet18 (Fang et al., 2024)	44.5	–	42.1	–	43.30	–
ViT (Fang et al., 2024)	42.8	–	45.9	–	44.35	–
CDCN (Fang et al., 2024)	45.9	–	41.4	–	43.65	–
AUX.(Depth) (Fang et al., 2024)	43.8	–	39.6	–	41.70	–
Baseline-ResNet18	47.5	54.0	47.7	52.1	47.60	53.05
Baseline-ViT	45.0	57.3	46.6	55.3	45.80	56.30
JDOT-FAS-ResNet18	**35.0**	**63.3**	**37.0**	**65.1**	**36.00**	**64.20**
JDOT-FAS-ViT	**33.1**	**72.3**	**35.4**	**67.1**	**34.25**	**69.70**

The best and the second best values are given in bold

with only 2D attack types. The methods we compare can be divided into three categories: DG-based FAS methods (Jia et al., 2020; Wang et al., 2022; Liu et al., 2022; Wang et al., 2022; Zhou et al., 2023; Liu et al., 2023; Long et al., 2023), source-free DA-based FAS methods (Liu et al., 2022; Li et al., 2018; Wang et al., 2020; He et al., 2020; Liang et al., 2020; Yang et al., 2021a, b; LV et al., 2021; Wang et al., 2021) and unsupervised DA-based FAS methods (Wang et al., 2021; Zhou et al., 2022; Wang et al., 2019; Quan et al., 2021). As shown in Tables 4 and 5. The DG-based FAS methods are trained without the involvement of target data in the training process. The source-free DA-based FAS methods use the source data for model pre-training and then fine-tune the source domain pre-trained model using the target data, which makes the source-free DA methods take into account the discrepancy between the source and target domains and use meta-learning, contrastive learning, etc. methods to reduce the domain discrepancy. However, most of the source-free DA methods lack the full utilization of source domain information in the alignment of source and target domains. The unsupervised DA-based FAS methods are the most effective methods to reduce domain discrepancy, which align the distributions of source and target domains in the training process. They are based on adversarial learning (Wang et al., 2019, 2021), cross-domain image generation (Zhou et al., 2022) and progressive migration learning (Quan et al., 2021) methods. Our WJDOT-FAS method belongs to the category of unsupervised DA-based FAS methods and it outperforms all the DG-based FAS methods and most of the source-free DA-based FAS methods due to its full utilization of target data information. Besides, our WJDOT-FAS method achieves the best average performance among unsupervised DA methods of the same category. In particular, our WJDOT-FAS-ResNet18 model reduces the average HTER by more than 1.74% and improves the average AUC by more than 0.13% and our WJDOT-FAS-ViT model reduces the average HTER by more than 2.51% and improves the average AUC by more than 0.36%. The possible reasons are as follows, first, other methods only consider the discrepancy in feature distributions, and ignore the effect of image labels. In contrast, we consider the alignment of the joint distributions of features and labels between domains, therefore the target domain classification effect is improved. Second, they treat different source domains uniformly, i.e., each source domain has the same domain weight, which means the target domain can not adaptively choose the source domain that is easier to align in the training process. In contrast, we make more effective use of the source domain information by setting domain

Table 4 Comparison results (HTER (%)) between the proposed method and the state-of-the-art methods for multi-source domain cross-dataset testing on the O, C, I, and M datasets

Methods	O&C& I→ M	O&M& I→ C	O&C&M→ I	I&C&M→ O	Average
SSDG (Jia et al., 2020)	7.4	10.4	11.7	15.6	11.29
PatchNet (Wang et al., 2022)	7.1	11.3	13.4	11.8	10.90
CIFAS (Liu et al., 2022)	6.0	10.7	8.5	13.2	9.57
SSAN (Wang et al., 2022)	6.7	10.0	8.9	13.7	9.82
IADG (Zhou et al., 2023)	5.4	8.7	10.6	8.9	8.40
SA-FAS (Liu et al., 2023)	6.0	8.8	6.6	10.0	7.83
DSCI (Long et al., 2023)	5.5	8.0	5.7	12.6	7.95
AdapBN (Li et al., 2018)	20.5	34.5	27.7	28.2	27.73
FTTA (Wang et al., 2020)	20.1	35.0	27.2	28.3	27.65
SDAN (He et al., 2020)	17.7	25.9	28.2	32.9	26.18
SHOT (Liang et al., 2020)	15.0	20.1	40.2	25.3	25.15
G-SFDA (Yang et al., 2021a)	37.5	38.9	32.6	40.4	37.35
NRC (Yang et al., 2021b)	15.0	47.8	22.1	26.6	27.88
DIPE-FAS (LV et al., 2021)	18.2	25.5	20	17.5	20.30
SDA-FAS (Wang et al., 2021)	15.4	24.5	15.6	23.1	19.65
SFDA-FAS (Liu et al., 2022)	**5.0**	**2.4**	**2.6**	**5.1**	**3.77**
ADA (Wang et al., 2019)	16.9	24.2	23.1	25.6	22.45
UDA (Wang et al., 2021)	16.1	22.2	22.7	24.7	21.43
GDA (Zhou et al., 2022)	9.2	12.2	10.0	14.4	11.45
PTL-FAS (Quan et al., 2021)	7.8±1.21	4.0±0.81	10.4±1.86	14.2±0.98	9.11±1.22
WJDOT-FAS-ResNet18	**5.0**	**1.7**	10.1	7.8	6.15
WJDOT-FAS-ViT	**5.0**	2.8	**9.1**	**4.6**	**5.38**

The best and the second best values are given in bold

weights that can be updated, which is an important reason that improves the effectiveness of cross-dataset testing for the multi-source DA-based FAS with 2D attacks.

Besides, we verify the effectiveness of our WJDOT-FAS method under the multi-source DA setting on the 3DMask, HiFiMask, and SuHiFiMask datasets with 3D attacks. As shown in Tables 6 and 7, under both ResNet-18 and ViT frameworks, the effect of our WJDOT-FAS method is improved compared to the baseline methods. Specifically, under the ResNet-18 framework, the average HTER (AUC) of our WJDOT-FAS method decreases (improves) by 6.07% (8.53%) compared to the baseline method, and under the ViT framework, the average HTER (AUC) of our WJDOT-FAS method decreases (improves) by 6.6% (4.23%) compared to the baseline method. As in the case of single-source DA, using the ResNet-18 network is more capable of capturing the real and fake information of 3D attack datasets. In addition, the generalization effect is best on the 3DMask dataset due to the fact that the 3D masks in the 3DMask dataset are better discriminated compared to the other two 3D attack datasets, and the 3DMask dataset is captured in a simpler acquisition environment. Overall, our WJDOT-FAS method can achieve cross-scenario generalization for DA-based FAS with 3D attacks.

In addition, to verify the effectiveness of our WJDOT-FAS method under the open-set multi-source DA setting, we choose the C and I datasets as source domains which have only 2D attack types and use the 3DMask, HiFiMask, or SuHiFiMask dataset with 3D attacks as the target domain. As shown in Table 8, under both ResNet-18 and ViT frameworks, the effect of our WJDOT-FAS method is improved compared to the baseline methods under the open-set multi-source DA setting. The improvement is more obvious under the C&I→3DMask setting, as the 3DMask dataset has a minor distribution discrepancy with the convex combination of the C and I relative to the other two datasets. The experimental results show that our WJDOT-FAS method is also effective for the open-set multi-source DA problem with novel attacks in the target domain which can reduce the large distributional discrepancy between the convex combination of the source distributions and the target distribution by computing the optimal transportation mappings. The weights of each source domain can be adaptively selected according to the classification loss and the optimal transportation loss,

Table 5 Comparison results (AUC (%)) between the proposed method and the state-of-the-art methods for multi-source domain cross-dataset testing on the O, C, I, and M datasets

Methods	O&C&I→ M	O&M&I→C	O&C&M→I	I&C&M→O	Average
SSDG (Jia et al., 2020)	97.2	95.9	96.6	91.5	95.31
PatchNet (Wang et al., 2022)	97.9	94.9	89.6	93.7	94.01
CIFAS (Liu et al., 2022)	96.3	95.3	97.2	93.4	95.58
SSAN (Wang et al., 2022)	98.8	96.7	96.8	93.6	96.46
IADG (Zhou et al., 2023)	98.2	96.4	94.5	97.1	96.57
SA-FAS (Liu et al., 2023)	96.6	95.4	97.5	96.2	96.42
DSCI (Long et al., 2023)	97.4	97.5	98.4	94.6	96.98
AdapBN (Li et al., 2018)	88.0	72.0	80.3	80.8	80.28
FTTA (Wang et al., 2020)	88.0	71.2	79.6	80.7	79.88
SDAN (He et al., 2020)	90.0	81.3	84.2	75.0	82.63
SHOT (Liang et al., 2020)	87.6	84.3	57.8	78.2	76.98
G-SFDA (Yang et al., 2021a)	67.8	67.2	73.6	63.7	68.08
NRC (Yang et al., 2021b)	87.4	52.4	82.3	78.8	75.23
SDA-FAS (Wang et al., 2021)	91.8	84.4	90.1	84.3	87.65
SFDA-FAS (Liu et al., 2022)	98.0	99.7	**99.5**	**99.0**	**99.04**
GDA (Zhou et al., 2022)	98.0	93.0	96.0	92.6	94.90
PTL-FAS (Quan et al., 2021)	**97.7±1.09**	**99.0±0.77**	**97.2±1.04**	93.7±0.75	96.86±0.91
WJDOT-FAS-ResNet18	**99.1**	**99.8**	95.7	97.0	97.90
WJDOT-FAS-ViT	97.0	99.5	96.9	**99.1**	**98.13**

The best and the second best values are given in bold

Table 6 Results (HTER (%)) for multi-source domain cross-dataset testing on the 3DMask, HiFiMask, and SuHiFiMask datasets

Methods	HiFiMask&SuHiFiMask →3DMask	3DMask&SuHiFiMask →HiFiMask	3DMask&HiFiMask →SuHiFiMask	Average
Baseline-ResNet18	10.2	38.2	38.6	29.00
Baseline-ViT	18.4	42.4	41.7	34.17
WJDOT-FAS-ResNet18	**5.0**	**30.1**	**33.7**	**22.93**
WJDOT-FAS-ViT	**12.6**	**35.9**	**34.2**	**27.57**

The best and the second best values are given in bold

which ultimately improves the classification of the models under the open-set multi-source DA setting.

4.4 Ablation Study

4.4.1 Effectiveness of the Joint Distribution Estimation

Tables 9 and 10 illustrate the advantages of the joint distribution estimation of features and labels (notated as JDOT-FAS or WJDOT-FAS) over the marginal distribution estimation of features (notated as MDOT-FAS or WMDOT-FAS) for the target domain discrimination. It can be seen that JDOT-FAS (or WJDOT-FAS) performs better than MDOT-FAS (or WMDOT-FAS) in the target domain. Specifically, under the single-source domain setting, JDOT-FAS has a 6.48% lower average HTER than MDOT-FAS; under the multi-source domain setting, WJDOT-FAS has a 1.8% lower average HTER and a 0.72% higher average AUC than WMDOT-FAS. This is because optimal transport methods based on joint distributions can achieve alignment of feature distributions of the samples with the same label (pseudo label) in the product space of features and labels, instead of considering only alignment of marginal distributions within the feature space. Therefore, the process of joint distribution alignment between the source and target domains not only allows the target domain to better learn the source domain label information but also allows the target domain to better utilize the information provided by the pseudo-labels, thus achieving more accurate classification results.

To further illustrate the contribution of joint distribution optimal transport-based method to pseudo-labels accuracy improvement, we plot the best HTER variation curves over

Table 7 Results (AUC (%)) for multi-source domain cross-dataset testing on the 3DMask, HiFiMask, and SuHiFiMask datasets

Methods	HiFiMask&SuHiFiMask →3DMask	3DMask&SuHiFiMask →HiFiMask	3DMask&HiFiMask →SuHiFiMask	Average
Baseline-ResNet18	90.4	67.3	66.0	74.57
Baseline-ViT	90.7	60.6	61.6	70.97
WJDOT-FAS-ResNet18	**98.7**	**77.1**	**73.5**	**83.10**
WJDOT-FAS-ViT	**94.5**	**66.3**	**64.8**	**75.20**

The best and the second best values are given in bold

Table 8 Results (HTER (%) and AUC (%)) of testings on the 3DMask, HiFiMask, and SuHiFiMask datasets with the C and I datasets as the source domains under the multi-source DA setting

	C&I→3DMask		C&I→HiFiMask		C&I→SuHiFiMask		Average	
Methods	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC
Baseline-ResNet18	31.0	78.5	41.6	61.5	47.5	53.7	40.03	64.50
Baseline-ViT	39.5	64.9	47.7	54.1	46.5	55.3	44.57	58.10
WJDOT-FAS-ResNet18	**17.3**	**92.5**	**33.6**	**71.4**	**37.4**	**74.1**	**29.43**	**79.33**
WJDOT-FAS-ViT	**17.4**	**86.8**	**39.4**	**60.2**	**36.8**	**65.7**	**31.20**	**70.90**

The best and the second best values are given in bold

Table 9 Evaluations (HTER (%)) of optimal transport methods based on marginal and joint distributions under the single-source DA setting

Methods	C→I	C→M	I→C	I→M	M→C	M→I	Average
MDOT-FAS	16.2	12.5	38.0	22.5	12.6	1.3	17.18
JDOT-FAS	**9.9**	**8.3**	**27.0**	**12.9**	**6.1**	**0.0**	**10.70**

The best and the second best values are given in bold

Table 10 Evaluations (HTER (%) and AUC (%)) of optimal transport methods based on marginal and joint distributions under the multi-source DA setting

	O&C&I→M		O&M&I→C		O&C&M→I		I&C&M→O		Average	
Methods	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC
WMDOT-FAS	7.9	98.8	3.9	99.4	10.6	94.4	9.4	96.1	7.95	97.18
WJDOT-FAS	**5.0**	**99.1**	**1.7**	**99.8**	**10.1**	**95.7**	**7.8**	**97.0**	**6.15**	**97.90**

The best and the second best values are given in bold

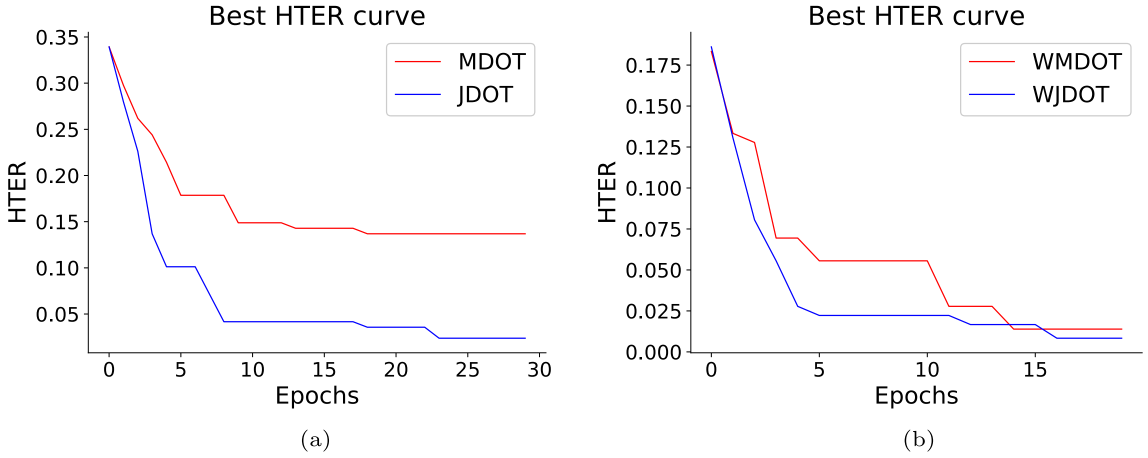
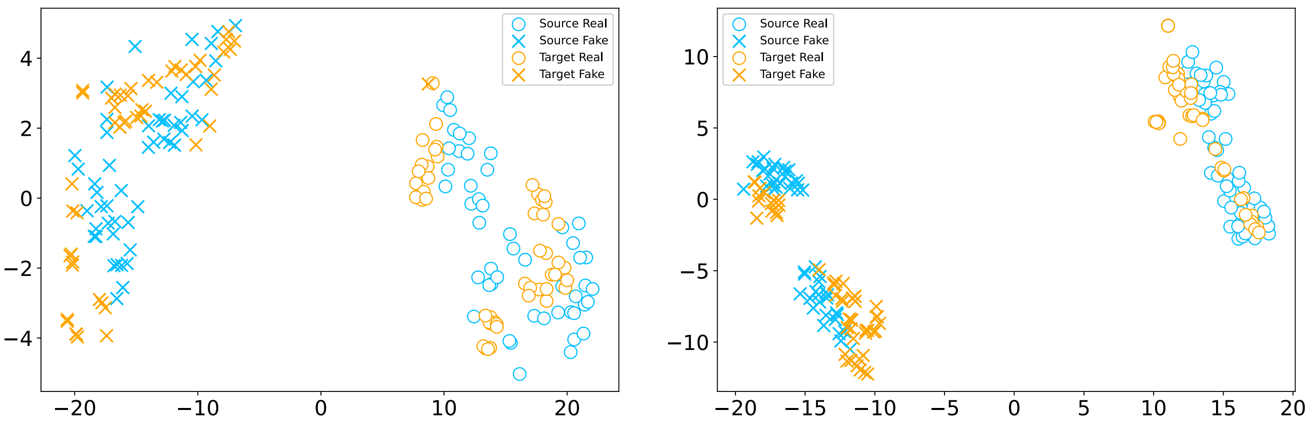
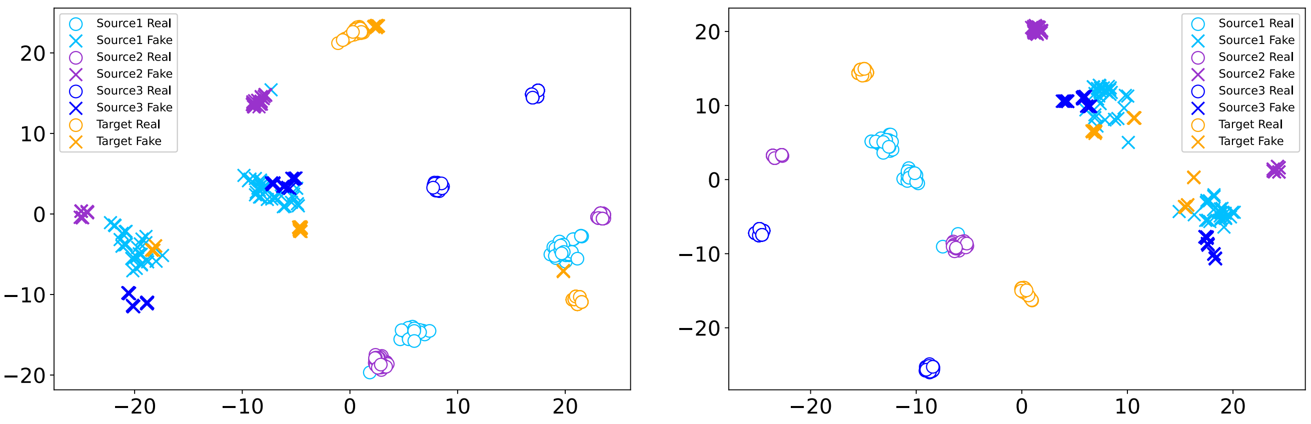


Fig. 4 Best HTER curves under the C→M (a) and O&M&I→C (b) settings. The red lines indicate the optimal transport methods based on marginal distributions, and the blue lines indicate the optimal transport methods based on joint distributions



(a) MDOT-FAS (b) JDOT-FAS

Fig. 5 The t-SNE embeddings of samples for optimal transport methods based on marginal distribution and joint distribution under the C→M setting. Different colors represent different domains; different tokens represent different classes



(a) WMDOT-FAS (b) WJDOT-FAS

Fig. 6 The t-SNE embeddings of samples for optimal transport methods based on weighted marginal distribution and weighted joint distribution under the O&M&I→C setting

the validation sets during the training process under the single- and multi-source DA settings in Fig. 4. As can be seen from the decreasing best HTER curves in both figures, both the marginal and joint distribution optimal transport-based methods can improve the accuracy of target domain pseudo-labels. Both methods continuously update the transportation mappings that minimize the transportation losses so as to align the source and target domain distributions and improve the accuracy of sample estimations in the target domain during the training processes. In addition, we observe that the models trained with joint distribution-based optimal transport methods make the best HTER curves decrease faster than the marginal distribution-based optimal transport methods and eventually drop to lower points, i.e. the final best HTER value drops from 13.69% to 2.38% under the single-source DA setting and from 1.39% to 0.83% under the multi-source DA setting. This is because the discrepancies between the source and target domain distributions can be more accurately characterized based on the joint distributions, and thus the transportation mapping solved using the cost matrix based on the joint distribution can align the source and target domain distributions faster and more accurately in the process of minimizing the optimal transportation loss.

Figures 5 and 6 show the t-SNE (Maaten et al., 2008) visualizations of the source and target domains, which are learned by MDOT-FAS and JDOT-FAS methods under the single-source DA setting, and WMDOT-FAS and WJDOT-FAS methods under the multi-source DA setting, respectively. It can be seen that, for MDOT-FAS and WMDOT-FAS, most of the samples in the target domain are far from the classification boundary, making it easier to distinguish between real and fake faces but there are still a small number of fake faces mixed into the real faces from the source domains in the target domain at the classification boundary. Thus,

Table 11 Evaluations (HTER (%)) of different loss functions for training under the single-source DA setting

Losses	C→I	C→M	I→C	I→M	M→C	M→I	Average
$\mathcal{L}_s$	33.4	32.9	53.1	31.2	15.4	19.1	30.85
$\mathcal{L}_s + \lambda_2 \mathcal{L}_t$	32.6	20.8	51.9	30.0	14.1	16.7	27.68
$\mathcal{L}_s + \lambda_1 \mathcal{L}_{s\text{-}t} + \lambda_2 \mathcal{L}_t$	**9.9**	**8.3**	**27.0**	**12.9**	**6.1**	**0.0**	**10.70**

The best and the second best values are given in bold

Table 12 Evaluations (HTER (%) and AUC (%)) of different loss functions for training under the multi-source DA setting

	O&C&I→M		O&M&I→C		O&C&M→I		I&C&M→O		Average	
Losses	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC
$\sum_k w_k \mathcal{L}_{s_k}$	12.1	93.0	19.8	82.2	21.9	87.8	27.7	77.5	20.38	85.13
$\sum_k w_k \mathcal{L}_{s_k} + \lambda_2 \mathcal{L}_t$	9.6	97.4	5.7	98.8	11.0	92.3	12.6	95.8	9.73	96.08
$\sum_k w_k (\mathcal{L}_{s_k} + \lambda_1 \mathcal{L}_{s_k\text{-}t}) + \lambda_2 \mathcal{L}_t$	**5.0**	**99.1**	**1.7**	**99.8**	**10.1**	**95.7**	**7.8**	**97.0**	**6.15**	**97.90**

The best and the second best values are given in bold

Table 13 Evaluations (HTER (%)) of the hyperparameter β under the single-source DA setting

β	C→I	C→M	I→C	I→M	M→C	M→I	Average
2	25.0	12.1	33.5	13.3	8.1	**0.0**	15.33
1	24.4	9.2	28.9	13.3	7.5	**0.0**	13.88
0.5	15.9	8.7	30.6	**7.5**	6.5	**0.0**	11.53
0.1	**9.9**	**8.3**	**27.0**	12.9	**6.1**	**0.0**	**10.70**
0.05	16.8	10.8	33.5	10.4	8.1	**0.0**	13.27
0.01	12.1	10.8	35.0	13.3	8.3	**0.0**	13.25

The best and the second best values are given in bold

MDOT-FAS and WMDOT-FAS methods have some effect on reducing the discrepancy of domain distributions, but the generalization performance still needs to be improved. For JDOT-FAS and WJDOT-FAS, all fake samples in the source and target domains are concentrated in the lower-left region or in the upper-right region, while all real samples are concentrated in the upper-right region or in the lower-left region, which means that JDOT-FAS and WJDOT-FAS play a greater role in reducing the domain distribution discrepancies compared to MDOT-FAS and WMDOT-FAS. The joint distribution-based optimal transport methods further achieve the minimum intra-class discrepancy and maximum inter-class discrepancy by introducing labels into the domain discrepancy metric. Therefore, joint distribution-based optimal transport methods have stronger discriminative power and better generalization ability by seeking domain-invariant product space.

4.4.2 Effectiveness of the Optimal Transportation Loss

Tables 11 and 12 show the experimental results for each added loss function under the single- and multi-source DA settings, respectively. Three types of models are trained for comparisons, i.e., the models trained with source domain classification loss (source domains weighted classification loss), the models trained with source domain classification loss (source domains weighted classification loss) and target domain entropy loss, and the models trained with source domain classification loss (source domains weighted classification loss), target domain entropy loss and joint distribution optimal transportation loss (weighted joint distribution optimal transportation loss). It can be seen that the test performances under the single- and multi-source DA settings improve with the addition of each loss function and the joint distribution optimal transportation loss plays a larger role under the single-source domain setting, which reduces the average HTER by 16.98%.The entropy loss allows the classifier to directly access unlabeled target data and adaptively adjust its parameters to further adapt to the distribution of the target domain. The joint distribution optimal transportation loss (weighted joint distribution optimal transportation loss) feeds back the distribution discrepancy during transportation into the update of network parameters. This enables the network to train the feature extractor and classifier that make the joint distributions of source and target domains closer to each other. We measure the discrepancy between the joint distributions of source domains and the target domain with the help of the Wasserstein distance, which enables a continuous transformation of the distributions, i.e., we are able to find the optimal transportation mappings that transform the joint distributions of source domains and the target domain into a common product space while preserving the geometrical characteristics of the distributions. The comparison results verify that each loss function of our proposed method con-

Table 14 Evaluations (HTER (%)) of three different transportation mapping solvers under the single-source DA setting

Solvers	C→I	C→M	I→C	I→M	M→C	M→I	Average
EMD	20.5	12.9	34.5	**7.5**	9.3	5.4	15.02
Sinkhorn	18.2	13.8	30.6	16.5	6.5	3.6	14.87
Lp-L1	**9.9**	**8.3**	**27.0**	12.9	**6.1**	**0.0**	**10.70**

The best and the second best values are given in bold

tributes to performance improvement and all loss functions interact with each other to finally achieve the best results.

4.5 Discussion

4.5.1 Influences of the Cost Matrix Computation

In Eq. (7), we generalize the definition of cost matrix in the marginal distribution optimal transport method as the distance between features in the source and target domains to the weighting of features and labels distances in the joint distribution optimal transport method and achieve the alignment of the joint distributions of source and target domain samples by flexibly defining $\boldsymbol{C}$. Table 13 shows the experimental results under different values of β. We conduct experiments under the single-source DA setting, and the experimental results show that the average effect is best when $\beta = 0.1$, so it can be seen that the label cost plays a key role in the total cost matrix to measure the distance of the joint distribution, and the appropriate ratio of feature cost to label cost can make the total cost matrix play a more effective role in aligning the joint distribution of the source and target domains, so that the samples in the target domain can achieve better classification effect.

4.5.2 Influences of the Transportation Mapping Computation

There are three types of solvers that can be used to solve for the transportation mapping, i.e. EMD solver (Eq. 9), sinkhorn solver (Eq. 10) and Lp-L1 solver (Eq. 11). Table 14 shows the experimental results of the target domain test samples with different transportation mapping solvers during training under the single-source DA setting. It can be seen that five of the six experiments achieve the best test results using the Lp-L1 solver. This is because the transportation mapping solution without regularization in Eq. (9) is a linear programming problem, which makes most of the elements of the transportation mapping matrix solved based on the EMD solver to be 0, resulting in a higher sparsity of the transportation mapping matrix and an unsmooth solution. In contrast, the sinkhorn solver (Eq. 10) with entropy regularization can find a smoother version of the transport, thus reducing the sparsity of the transportation mapping matrix. Lp-L1 solver (Eq. 11) makes the target samples receive masses only from source samples with the same label over the smooth transport version. This can further constrain the transportation mapping to transport between samples with the same label in the source and target domains, on the basis of using the joint distribution to define the cost matrix.

4.5.3 Influences of the Domain Weight Optimization

In Table 15, we summarize three categories of domain-weighting strategies for comparison, including using the same weights for each domain, non-learnable weighting strategies based on the distances and optimization-based weighting strategies by using different optimizers. Non-learnable weighting strategies based on distances include classification loss-based (Kang et al., 2020), optimal transport distance-based (Zhao et al., 2020), and MMD based (Li et al., 2021) weighting strategies. Optimization-based weighting strategies include SGD (Saad D, 1998), AdaGrad (Duchi et al., 2011), and Adam (Kingma et al., 2014) algorithms. As can be seen from the comparison results, using the same domain weight for all source domains is equivalent to stitching all source samples into a source distribution and using joint distribution optimal transport on the resulting distribution. This equal treatment of each source domain is not conducive to the model learning the convex combination of source domains that are best adapted to the target domain, therefore using the same weights for each domain is poorly tested on the target domain. In addition, the optimal transport distance (Wasserstein distance) based weighting strategy outperforms the MMD-based weighting strategy, and the classification loss-based weighting strategy is the worst of these three distances-based weighting strategies. This further indicates that the optimal transport distance is a more accurate measure of similarity between domains and that the combination of source domains adapted to the target domain can be better learned with the optimal transport distance-based weighting strategy. Finally, optimization-based weighting strategies by using different optimizers obtain the best average results of the three categories of methods. This is due to the optimization-based weighting strategies allowing for more flexible optimization of domain weights based on the classification performance and transportation performance of each source domain. Models trained with the Adam algorithm achieve the best average test results on the target domain among these three optimiza-

Table 15 Evaluations (HTER (%) and AUC (%)) of domain weight optimization under the multi-source DA setting

	O&C&I→M		O&M&I→C		O&C&M→I		I&C&M→O		Average	
Methods	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC
$w_k = 1/K$	11.3	98.1	2.8	99.7	11.3	90.7	9.9	95.2	8.83	95.93
$w_k \propto 1/\mathcal{L}_{s_k}$	5.8	97.8	5.7	97.3	11.6	91.1	12.4	94.1	8.88	95.08
$w_k \propto e^{-\mathcal{L}^2_{s_k-t}/2}$	5.4	97.4	2.6	99.8	11.5	92.1	10.4	96.4	7.48	96.43
$w_k \propto 1/\mathcal{MMD}(\boldsymbol{x}^{s_k}, \boldsymbol{x}^t)$	7.9	97.1	3.3	**100.0**	11.4	93.3	10.6	94.8	8.30	96.30
optimize with SGD	7.9	98.1	3.3	99.2	11.3	94.8	8.5	96.4	7.75	97.13
optimize with AdaGrad	5.8	**99.1**	3.1	99.8	11.5	93.1	9.9	95.2	7.58	96.80
optimize with Adam	**5.0**	**99.1**	**1.7**	99.8	**10.1**	**95.7**	**7.8**	**97.0**	**6.15**	**97.90**

The best and the second best values are given in bold

tion algorithms due to it determines the parameters update using the first-order moment estimation of the stochastic gradient and determines the adaptive learning rate using the second-order moment estimation of the stochastic gradient. Besides, the Adam algorithm corrects the bias of the first-order moment and second-order moment estimation, which avoids the oscillation during the optimization of the SGD algorithm as well as the low optimization efficiency of the AdaGrad algorithm in the later stage of the optimization process. Overall, the Adam algorithm can better learn the convex combination of source domain distributions that is closest to the target domain distribution. Therefore, we finally choose the Adam algorithm as the domain-weighting strategy.

5 Conclusions and Future Work

In this paper, we introduce a weighted joint distribution optimal transport framework to solve the cross-scenario problem in FAS, which is applicable in both single- and multi-source DA scenarios. This framework can be divided into three parts: joint distribution estimation, joint distribution optimal transport, and domain weight optimization. We compute the optimal transportation mappings from each source domain and the target domain based on the joint distribution cost matrices, and optimize the feature representation, label estimation, and domain weights simultaneously using the weighted optimal transportation loss, weighted source domain classification loss, and the target domain entropy loss. To validate the effectiveness of our method under both single- and multi-source DA settings, we have done extensive experiments on four public FAS datasets with only 2D attacks and three large-scale FAS datasets with 3D attacks. The experimental results show that our method achieves state-of-the-art results in all three protocols under the single-source setting, and under the multi-source setting, our method outperforms most multi-source DA methods. However, there is still room for performance improvement under the O&C&M→I setting. In future work, we will design the optimal transport framework applicable to domain generalization-based and Multi-modal FAS.

Acknowledgements We thank the following professors and their group members: Wan et al., Lei et al., Jain et al., Komulainen et al., and Marcel et al. for the CASIA-FASD, Idiap Replay-Attack, MSU-MFSD, OULU-NPU, 3DMask, HiFiMask and Surveillance datasets. The work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant (No. 61976173).

Data Availability Experimental data from seven datasets: CASIA-FASD (Zhang et al., 2012), Idiap Replay-Attack (Chingovska et al., 2012), MSU-MFSD (Wen et al., 2015), OULU-NPU (Boulkenafet et al., 2017), CASIA-SURF 3DMask (Yu et al., 2020), CASIA-SURF HiFi-Mask (Liu et al., 2022) and Surveillance High-Fidelity Mask (Fang et al., 2024). The data that support the findings of this study are available from third-party institutions (including Idiap Research Institute, Michigan State University, Chinese Academy of Sciences, and University of Oulu) but restrictions apply to the availability of these data, which were used under license for the current study. Data are however available from the authors upon reasonable request and with permission of the above-mentioned third-party institutions.

References

Sun, B., & Saenko, K. (2016). Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision-(ECCV), Amsterdam* (pp. 443–450). Springer.

Boulkenafet, Z., Komulainen, J., Li, L., Feng, X., & Hadid, A. (2017). OULU-NPU: A mobile face presentation attack database with real-world variations. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 612–618). IEEE.

Cédric V. (2008). Optimal transport: Old and new (Grundlehren der mathematischen Wissenschaften). Springer.

Chen, Z., Yao, T., Sheng, K., Ding, S., Tai, Y., Li, J., Huang, F., & Jin, X. (2021). Generalizable representation learning for mixture domain face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Chingovska, I., Anjos, A., & Marcel, S. (2012). On the effectiveness of local binary patterns in face anti-spoofing.*2012 BIOSIG-Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG)* (pp. 1–7).

Courty, N., Flamary, R., & Tuia, D. (2014). Domain adaptation with regularized optimal transport. In *Proc. Eur. Conf. Mach. Learn. Principles Practice Knowl. Discovery Databases* (pp. 1–16).

Nicolas, C., Remi, F., Devis, T., & Alain, R. (2016). Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(9), 1853–1865.

Courty, N., Flamary, R., Habrard, A., & Rakotomamonjy, A. (2017). Joint distribution optimal transportation for domain adaptation. In *Advances in neural information processing systems* (Vol. 30).

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems* (Vol. 26, pp. 2292–2300). Curran Associates, Inc.

Cuturi, M., & Doucet, A. (2014). Fast computation of wasserstein barycenters. In *International Conference on Machine Learning, PMLR* (pp. 685–693).

Damodaran, B. B., Kellenberger, B., Flamary, R., Tuia, D., & Courty, N. (2018). Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 447–463).

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research, 12*, 2121–2159.

Fang, H., Liu, A. J., Wan, J., et al. (2024). Surveillance face anti-spoofing. *IEEE Transactions on Information Forensics and Security, 19*, 1535–1546.

Ghifary, M., Kleijn, W. B., Zhang, M., Balduzzi, D., & Li, W. (2016). Deep reconstruction-classification networks for unsupervised domain adaptation. *Computer Vision-(ECCV), Amsterdam, The Netherlands* (pp. 597–613). Springer

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (pp. 770–778).

He, Y., Carass, A., Zuo, L., Dewey, B. E. & Prince, J. L. (2020). Self domain adapted network. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*.

Hu, L., Kan, M., Shan, S., & Chen, X. (2018). Duplex generative adversarial network for unsupervised domain adaptation. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (pp. 1498–1507).

Jia, Y., Zhang, J., Shan, S. & Chen, X. (2020). Single-side domain generalization for face anti-spoofing. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (pp. 8484–8493).

Jia, Y., Zhang, J., Shan, S., & Chen, X. (2021). Unified unsupervised and semi-supervised domain adaptation network for cross-scenario face anti-spoofing. *Pattern Recognition, 115*, 107888.

Jiang, F., Li, Q., Liu, P., Zhou, X. D., & Sun, Z. (2023). Adversarial learning domain-invariant conditional features for robust face anti-spoofing. *International Journal of Computer Vision* 1–24.

Hoffman, J., Mohri, M., & Zhang, N. (2018). Algorithms and theory for multiplesource adaptation. In *Advances in neural information processing systems* (pp. 8246–8256).

Kang, G., Jiang, L., Wei, Y., et al. (2020). Contrastive adaptation network for single-and multi-source domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(4), 1793–1804.

Kantorovich, L. V. (2006). On the translocation of masses. *Journal of Mathematical Sciences, 133*(4), 1381–1382.

Khammari, M. (2019). Robust face anti-spoofing using CNN with LBP and WLD. *IET Image Processing, 13*, 1880–1884.

Kingma, D. P. & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980

Komulainen, J., Hadid, A., & Pietikäinen, M. (2013). Context based face anti-spoofing. In *IEEE Sixth International Conference on Biometrics* (pp. 1–8).

Li, H., Li, W., Cao, H., Wang, S., Huang, F., & Kot, A. C. (2018). Unsupervised domain adaptation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security, 13*(7), 1794–1809.

Li, Y., Wang, N., Shi, J., Hou, X., & Liu, J. (2018). Adaptive batch normalization for oractical domain adaptation. *Pattern Recognition (PR), 80*, 109–117.

Li, K., Lu, J., Zuo, H., & Zhang, G. (2021). Multi-source contribution learning for domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems, 33*(10), 5293–5307.

Li, Z., Cai, R., Li, H., Lam, K. Y., Hu, Y., & Kot, A. C. (2022). One-class knowledge distillation for face presentation attack detection. *IEEE Transactions on Information Forensics and Security, 17*, 2137–2150.

Liang, J., Hu, D. & Feng, J. (2020). Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning (ICML), PMLR* (pp. 6028–6039).

Liu, Y., Jourabloo, A., & Liu, X. (2018). Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 389–398).

Liu, A., Tan, Z., Wan, J., Escalera, S., Guo, G., & Li, S. Z. (2021). Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 1179–1187).

Liu, Y., Chen, Y., Dai, W., Gou, M., Huang, C. T., & Xiong, H. (2022). Source-free domain adaptation with contrastive domain alignment and self-supervised exploration for face anti-spoofing. In *Computer Vision-ECCV 2022: 17th European Conference, Tel Aviv, Israel* (pp. 511–528). Springer.

Liu, S., Lu, S., Xu, H., Yang, J., Ding, S., & Ma, L. (2022). Feature generation and hypothesis verification for reliable face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, pp. 1782–1791).

Liu, Y., Chen, Y., Dai, W., Li, C., Zou, J., & Xiong, H. (2022). Causal intervention for generalizable face anti-spoofing. In *ICME* (pp. 1–6). IEEE.

Liu, A., Zhao, C., Yu, Z., et al. (2022). Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection. *IEEE Transactions on Information Forensics and Security, 17*, 2497–2507.

Sun, Y., Liu, Y., Liu, X., Li, Y., & Chu, W. S. (2023). Rethinking domain generalization for face anti-spoofing: Separability and alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 24563–24574).

Long, X., Zhang, J., Wu, S., Jin, X., & Shan, S. (2023). Dual sampling based causal intervention for face anti-spoofing with identity debiasing. *IEEE Transactions on Information Forensics and Security (TIFS)*. https://doi.org/10.1109/TIFS.2023.3326370

Lv, L., Xiang, Y., Li, X., Huang, H., Ruan, R., Xu, X. & Fu, Y. (2021). Combining dynamic image and prediction ensemble for cross-domain face anti-spoofing. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2550–2554).

Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research (JLMR), 9*, 2579–2605.

Patel, K., Han, H., & Jain, A. K. (2016). Secure face unlock: Spoof detection on smartphones. *IEEE Transactions on Information Forensics and Security, 11*(10), 2268–2283.

Pei, Z., Cao, Z., Long, M., & Wang, J. (2018). Multi-adversarial domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32).

Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., & Wang, B. (2019). Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1406–1415).

Quan, R., Wu, Y., Yu, X., & Yang, Y. (2021). Progressive transfer learning for face antispoofing. *IEEE Transactions on Image Processing, 30*(3), 3946–3955.

Rahman, M. M., Fookes, C., Baktashmotlagh, M., & Sridharan, S. (2020). On minimum discrepancy estimation for deep domain adaptation. In *Domain adaptation for visual understanding*(pp. 81–94). Springer.

Rehman, Y. A. U., Po, L. M., & Komulainen, J. (2020). Enhancing deep discriminative feature maps via perturbation for face presentation attack detection. *Image and Vision Computing, 94*, 103858.

Saad, D. (1998). Online algorithms and stochastic approximations. *Online Learning, 5*(3), 6.

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & J'egou, H. (2021). Training data-efficient image transformers & distillation through attention. In *ICML, PMLR* (pp. 10347–10357).

Turrisi, R., Flamary, R., Rakotomamonjy, A., & Pontil, M. (2022). Multi-source domain adaptation via weighted joint distributions optimal transport. In *Uncertainty in Artificial Intelligence. PMLR* (pp. 1970–1980).

Tzeng, E., Hoffman, J., Saenko, K., & Darrell, T. (2017). Adversarial discriminative domain adaptation. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (pp. 7167–7176).

Wang, G., Han, H., Shan, S. & Chen, X. (2020). Cross-domain face presentation attack detection via multi-domain disentangled representation learning. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (pp. 6678–6687).

Wang, G., Han, H., Shan, S., & Chen, X. (2019). Improving cross-database face presentation attack detection via adversarial domain adaptation. In *ICB*. IEEE.

Wang, D., Shelhamer, E., Liu, S., Olshausen, B., & Darrell, T. (2020). Tent: Fully test-time adaptation by entropy minimization. arXiv:2006.10726

Wang, G., Han, H., Shan, S., & Chen, X. (2021). Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection. *IEEE Transactions on Information Forensics and Security, 16*, 56–69.

Wang, J., Zhang, J., Bian, Y., Cai, Y., Wang, C., & Pu, S. (2021). Self-domain adaptation for face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)* (Vol. 35, pp. 2746–2754).

Wang, Z., Wang, Z., Yu, Z., Deng, W., Li, J., Gao, T., & Wang, Z. (2022). Domain generalization via shuffled style assembly for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4123–4133).

Wang, C. Y., Lu, Y. D., Yang, S. T., & Lai, S. H. (2022). Patchnet: A simple face anti-spoofing framework via fine-grained patch recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20281–20290).

Wang, Z., Yu, Z., Wang, X., Qin, Y., Li, J., Zhao, C., Liu, X., & Lei, Z. (2023). Consistency regularization for deep face anti-spoofing. *IEEE Transactions on Information Forensics and Security, 8*, 1127–1140.

Wen, D., Han, H., & Jain, A. K. (2015). Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security (TIFS), 10*(4), 746–761.

Wen, J., Greiner, R., & Schuurmans, D. (2020). Domain aggregation networks for multi-source domain adaptation. In *International Conference on Machine Learning. PMLR* (pp. 10214–10224).

Xu, R., Chen, Z., Zuo, W., Yan, J., & Lin, L. (2018). Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3964–3973).

Yang, J., Lei, Z., Yi, D., & Li, S. Z. (2015). Person-specific face anti-spoofing with subject domain adaptation. *IEEE Transactions on Information Forensics and Security (TIFS), 10*, 797–809.

Yang, S., Wang, Y., van de Weijer, J., Herranz, L. & Jui, S. (2021a). Generalized source-free domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 8978–8987).

Yang, S., Van de Weijer, J., Herranz, L., & Jui, S. (2021b). Exploiting the intrinsic neighborhood structure for source-free domain adaptation. *In Advances in neural information processing systems (NeurIPS)* (Vol. 34, pp. 29393–29405).

Mansour, Y., Mohri, M., & Rostamizadeh, A. (2008). Domain adaptation with multiple sources. In *Advances in neural information processing systems* (pp. 1041–1048).

Yu, Z., Zhao, C., Wang, Z., Qin, Y., Su, Z., Li, X., Zhou, F., & Zhao, G. (2020). Searching central difference convolutional networks for face anti-spoofing. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (pp. 5295–5305).

Yu, Z., Wan, J., Qin, Y., Li, X., Li, S. Z., & Zhao, G. (2020). NAS-FAS: Static-dynamic central difference network search for face anti-spoofing. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(9), 3005–3023.

Yu, Z., Qin, Y., Li, X., Zhao, C., Lei, Z., & Zhao, G. (2022). Deep learning for face anti-spoofing: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(5), 5609–5631.

Yue, H., Wang, K., Zhang, G., Feng, H., Han, J., Ding, E., & Wang, J. (2022). Cyclically disentangled feature translation for face anti-spoofing. arXiv preprint arXiv:2212.03651

Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., & Li, S. Z. (2012). A face anti-spoofing database with diverse attacks. In *2012 5th IAPR International Conference on Biometrics (ICB)* (pp. 26–31). IEEE.

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters, 23*, 1499–1503.

Zhang, S., Wang, X., Liu, A., Zhao, C., Wan, J., Escalera, S., Shi, H., Wang, Z., & Li, S. Z. (2019). A dataset and benchmark for large-scale multi-modal face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 919–928).

Zhang, S., Liu, A., Wan, J., Liang, Y., Guo, G., Escalera, S., Escalante, H. J., & Li, S. Z. (2020). CASIA-SURF: A large-scale multi-modal benchmark for face anti-spoofing. *IEEE Transactions on Biometrics, Behavior, and Identity Science, 2*, 182–193.

Zhao, H., Zhang, S., Wu, G., Moura, J. M., Costeira, J. P., & Gordon, G. J. (2018). Adversarial multiple source domain adaptation. In *Advances in neural information processing systems* (Vol. 31).

Zhao, S., Wang, G., Zhang, S., Gu, Y., Li, Y., Song, Z., Xu, P., Hu, R., Chai, H., & Keutzer, K. (2020). Multi-source distilling domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, pp. 12975–12983).

Zhou, Q., Zhang, K. Y., Yao, T., Yi, R., Sheng, K., Ding, S., & Ma, L. (2022). Generative domain adaptation for face anti-spoofing. In *Computer Vision-ECCV, Tel Aviv, Israel* (pp. 335–356). Springer.

Zhou, Q., Zhang, K. Y., Yao, T., Lu, X., Yi, R., Ding, S., & Ma, L. (2023). Instance-aware domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 20453–20463).

Zhuang, F., Cheng, X., Luo, P., Pan, S. J., & He, Q. (2015). Supervised representation learning: Transfer learning with deep autoencoders. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.