Optimal Transport-Guided Source-Free Adaptation for Face Anti-Spoofing

Zhuowei Li[1*†], Tianchen Zhao[2*], Xiang Xu[2], Zheng Zhang[2], Zhihua Li[2],
Xuanbai Chen[2], Qin Zhang[2], Alessandro Bergamo[2], Anil K. Jain[2], Yifan Xing[2]
[1]Rutgers University [2]AWS AI Labs

Abstract

Developing a face anti-spoofing model that meets the security requirements of clients worldwide is challenging due to the domain gap between training datasets and diverse end-user test data. Moreover, for security and privacy reasons, it is undesirable for clients to share a large amount of their face data with service providers. In this work, we introduce a novel method in which the face anti-spoofing model can be adapted by the client itself to a target domain at test time using only a small sample of data while keeping model parameters and training data inaccessible to the client. Specifically, we develop a prototype-based base model and an optimal transport-guided adaptor that enables adaptation in either a lightweight training or training-free fashion, without updating base model's parameters. Furthermore, we propose geodesic mixup, an optimal transport-based synthesis method that generates augmented training data along the geodesic path between source prototypes and target data distribution. This allows training a lightweight classifier to effectively adapt to target-specific characteristics while retaining essential knowledge learned from the source domain. In cross-domain and cross-attack settings, compared with recent methods, our method achieves average relative improvements of 19.17% in HTER and 8.58% in AUC, respectively.

1. Introduction

Traditional approaches for developing face anti-spoofing models involve collecting extensive training datasets designed to cover a broad spectrum of real user interactions and spoof attempts. These datasets aim to represent diverse environmental conditions, user behaviors, demographic variations, and client-specific properties, such as artifacts introduced by image acquisition devices. However, these approaches have proven ineffective in practice [88], as the variety of user conditions is virtually limitless. Moreover, client-specific properties are dynamic and evolve over time, for example, upon the introduction of new image sensors and acquisition technologies. The conventional strategies struggle to keep up with these rapid advancements, as collecting new training datasets and retraining models on a daily basis is impractical.

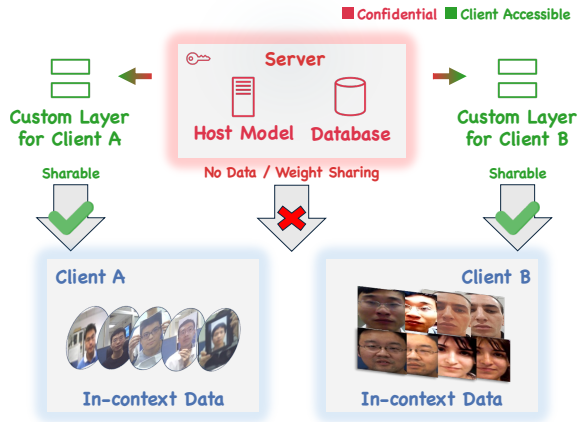

Figure 1. Overview of our setting. Unlike Few-shot Learning (FSL), Domain Adaptation (DA), and Domain Generalization (DG), our approach uses few-shot client data at test stage to adapt a customizable layer for each client's needs, while keeping the host model and source training data confidential. This layer refers to prototypes and a lightweight classifier in our training-free and lightweight training approaches.

Furthermore, conventional approaches involving tuning the model to meet client requirements using a small set of labeled end-user examples face several critical challenges. First, legal and privacy restrictions concerning human faces and security models often limit the quantity of available reference data and prohibit explicit sharing of host model parameters and source training data with clients for improvement. Second, fine-tuning a separate model at the service host side to address specific requirements for each client is impractical, as it is costly to maintain and manage future updates [16, 30, 66]. Third, clients often desire a quick customization service that can avoid the inefficiencies of back-and-forth communication with the host in order to meet

*Equal contribution.
†Work done during internship at AWS AI Labs.

ever-evolving changes in their use cases.

In this work, we consider the problem of building a privileged system that allows for lightweight user-specific customization at testing time by the client, using only a few labeled samples, without the need for the host to share model parameters or training samples. These samples might be from different domains (cross-domain) or represent new spoofing instruments by adversaries (cross-attack). This differs from traditional Domain Adaptation (DA) and Domain Generalization (DG) methods, where DA requires both source and target domain data during training [38, 47, 78, 102] and DG relies solely on static parameters to generalize to downstream scenarios [7, 22, 25, 41, 49, 67, 104, 105].

We address these challenges by introducing a customizable layer on top of the host model that can adapt to the target domain using a limited amount of client data, without accessing the source training data and host model parameters. This customizable layer can be maintained and updated by the client in a source-free fashion, providing maximum flexibility to address each client's specific requirements [70]. The adaptation is realized by first learning, during the model training phase, a set of source prototypes that encode information about distributions of source domain data. At test time, we adapt the model through optimal transportation (OT) of the learned prototypes either in a training-free fashion or using a lightweight training method. OT is particularly suitable for our application where the size of empirical data available to represent the true underlying distribution is limited [13, 18], as it leverages Wasserstein distance to effectively exploit the geometry of the underlying feature space, faithfully aligning source and target distributions even when the data is sparse or unevenly sampled.

Concretely, in our training-free approach, we map source prototype features into the target domain with an optimal transport transformation that requires no learnable parameters. This transformation aligns the source prototypes with the target distribution's structure, allowing them to capture the unique characteristics of the target domain and make inference with the target data features. Furthermore, inspired by our training-free approach, we propose an alternative method that trains a lightweight classifier on top of the frozen feature extractor, taking as input the source prototypes and a few target data features. To improve the learning of the decision boundary under a low-data regime, we introduce geodesic mixup, an OT-based synthesis method that generates augmented training data. Unlike traditional mixup data augmentations [71, 93] that perform point-wise linear interpolation between pairs of data features, the synthetic data generated along the geodesic path between source and target distributions guides the classifier to better capture the underlying feature manifold of both domains. By training on these samples, the classifier learns how features transition between domains, adapting to target-specific characteristics while maintaining knowledge about the source domain.

In summary, our main contribution is a novel framework for face anti-spoofing that uses a source-free few-shot adaptation approach based on prototypes and does not modify the parameters of the backbone host model. Our method leverages optimal transport for adaptation (OTA), where we employ the Wasserstein distance to estimate the true underlying data distribution based on a sparse sample of data, and a novel data augmentation method, geodesic mixup, to improve domain generalization performance. Our method achieves average relative improvements of 19.17% in HTER and 8.56% in AUC under cross-domain and cross-attack scenarios compared to state-of-the-art methods.

2. Related Works

Face anti-spoofing (FAS) is a crucial security component in face recognition systems that has been extensively researched. Early efforts relied on handcrafted features like LBP [5, 14, 39], HOG [32, 90], SIFT [58], and others [8, 31, 56, 87]. Deep learning has significantly advanced the field, with CNN-based methods [22, 44–46, 73, 88, 100] integrating feature extraction and classification into unified frameworks. However, many approaches struggle to generalize across domains in cross-dataset evaluations. Domain Adaptation (DA) [20, 38, 47, 51, 75, 77, 86, 102] adapts models from source to target domains using labeled or unlabeled target data [38, 47, 78, 102], but it requires access to both source and target data at the training time, which is infeasible in our setting due to privacy concerns. Domain Generalization (DG) leverages multiple source domains to train models that generalize to unseen domains [7, 22, 25, 41, 49, 67, 104, 105], but it underperforms for significant domain shifts. Some methods [25, 62, 76] employ adversarial training for domain-invariant features, while others [7, 63, 67, 78, 80] use meta-learning and domain separability to improve generalization.

Recent DG advances include instance-level domain-specific strategies like CIFAS [48], which uses causal intervention to mitigate domain bias, and AMEL [101], which incorporates domain-specific features. Recently, Liu et al. [49] proposed unsupervised DG frameworks that leverage unlabeled data to learn generalizable features. While DA and DG methods have achieved success in FAS, both fall short in addressing practical scenarios where hosts must offer tailored solutions to clients with only a few labeled target samples at test time, without explicitly sharing the source model parameters. Our work investigates this constrained setting to highlight its practical importance and to draw research attention to this under-explored topic.

Source-Free Domain Adaptation (SFDA) [28] decouples

domain adaptation from direct source data usage by leveraging pre-trained source models. Liang et al. [40] learns target-specific feature extraction with pseudo labels guided by class-wise prototypes. Zhang et al. [96] and Lee et al. [37] also consider few-shot SFDA scenarios, but they do not restrict the access to the source model. In FAS, Lv et al. [52] and Mao et al. [54] employ pseudo labels for self-training but suffer from accumulated errors. Liu et al. [47] considers a similar setting but also allows for unrestricted host model access. Huang et al. [23] focuses on online adaptation to unlabeled target data without source data. Our method provides a lightweight solution for source-free few-shot adaptation without modifying the backbone host model parameters.

Optimal Transport (OT) has been widely adopted in domain adaptation [4, 11, 35, 36, 55, 64, 85, 99] and generative modeling [3, 19] due to its ability to compute Wasserstein distances between probability distributions by effectively exploiting the underlying metric space geometry. OT is well-suited to our setting because its theoretical formulation adapts straightforwardly to discrete cases, working directly with empirical distribution estimates without assumptions on source and target distribution supports [10, 11, 36, 99]. Regularized OT addresses overfitting when few samples are available [13, 18]. While Mix-up [71, 93] has been used in Adversarial Domain Adaptation [84, 89] to improve discriminator decision boundaries, our proposed geodesic mix-up serves a different purpose: it generates synthetic distributions by interpolating along a geodesic path in feature space with OT, and we train on data sampled from these generated synthetic distributions to learn feature transitions between domains, adapting to target characteristics while maintaining source domain understanding. Xu et al. [83] also considers applying OT to the FAS task; nevertheless, they employ linear mix-up to augment the training set before applying JDOT [12], a well-established DA method, which stands in contrast to our proposed geodesic mix-up.

3. Methodology

We propose optimal transport-guided source-free few-shot adaptation for face anti-spoofing (OTA), which trains a privileged model that supports convenient customization at the test stage by either host or client. Different from standard classification training, OTA is built upon a prototype-based training framework [69] that learns a set of source prototype features as both last-layer classifiers and surrogates for source domain distributions. A detailed description of this framework is provided in Section 3.2. At the testing stage, OTA offers two approaches for test-stage adaptation through optimal transportation (OT) of the learned prototypes: a training-free method and a lightweight training method, introduced in Section 3.3 and in Section 3.4, respectively. We refer to Fig. 2 for an overview of OTA.

3.1. Problem Formulation

During the training stage, we have access to N labeled source-domain datasets: $\{\mathbb{D}_i\}_{i=1}^{N}$, where each dataset $\mathbb{D}_i$ consists of M_i labeled training samples $\{(x_{ij}, y_{ij})\}_{j=1}^{M_i}$. Here, $x_{ij} \in \mathbb{X}_i$ is an image from the i^{th} dataset and $y_{ij} \in \{0, 1\}$ is its corresponding binary label, indicating whether the image is bona fide or a spoof, respectively. At test time, we assume the existence of a few-shot labeled dataset $\mathbb{D}_t = \{(x_{tj}, y_{tj})\}_{j=1}^{M_t}$ from the target domain, which is not accessible during the training of the feature extractor f. The size of $\mathbb{D}_t$ is significantly smaller than that of each source dataset, *i.e.*, $M_t \ll M_i, i \in \{1, 2, ..., N\}$. By default, we assume $\mathbb{D}_t$ contains samples from both bona fide and spoof classes; we will investigate the one-class scenario in our ablation studies. The objective is to build a customizable layer on top of f that adapts to the target domain using $\mathbb{D}_t$ during the adaptation phase of the test stage, without using $\{\mathbb{D}_i\}_{i=1}^{N}$ or modifying f.

3.2. Prototype-based Framework

The primary challenge in our setting arises from the lack of explicit access to the source-domain data and host model at the test stage due to privacy and proprietary considerations. To this end, we adopt a prototype-based framework. Instead of learning a classifier, we train a multi-centroid prototype for each class, with $\boldsymbol{p}^{\text{bona fide}} \in \mathbb{R}^{D\times K}$ for the bona fide class and $\boldsymbol{p}^{\text{spoof}} \in \mathbb{R}^{D\times K}$ for the spoof class, where K denotes the number of sub-centers (determined prior to training) and D indicates the feature dimension.

Given an image embedding $\boldsymbol{z} = f(x) \in \mathbb{R}^D$, classification is performed by calculating the mean cosine similarity between $\boldsymbol{z}$ and each set of prototypes. The label is then assigned based on the highest mean similarity over the K prototypes:

$$c = \underset{c\in\{\text{bona fide},\text{spoof}\}}{\arg\max} \frac{1}{K}\sum_{k=1}^{K} \frac{\boldsymbol{z}\cdot\boldsymbol{p}_k^c}{\|\boldsymbol{z}\|_2\|\boldsymbol{p}_k^c\|_2}. \tag{1}$$

In this framework, prototypes not only serve as classifiers but also encapsulate the feature distributions of the source domains. We use prototypes with sub-centroids to improve their expressiveness. At the test stage, the feature extractor f is treated as a black box that takes as input the test image and generates its corresponding feature vector. The learned prototypes $\boldsymbol{P} = \{\boldsymbol{p}^{\text{bona fide}}, \boldsymbol{p}^{\text{spoof}}\} \in \mathbb{R}^{D\times K\times 2}$ are accessible since they are lightweight and respect privacy.

Training. Intuitively, an embedding feature $\boldsymbol{z}_i$ of image x_i should be close to its corresponding prototype sub-centroids while being distinctly separated from mismatched prototypes. Inspired by the ArcFace loss [15], we enforce an additive margin m in the angular space. Let $\boldsymbol{S} \in \mathbb{R}^{2\times K} =$

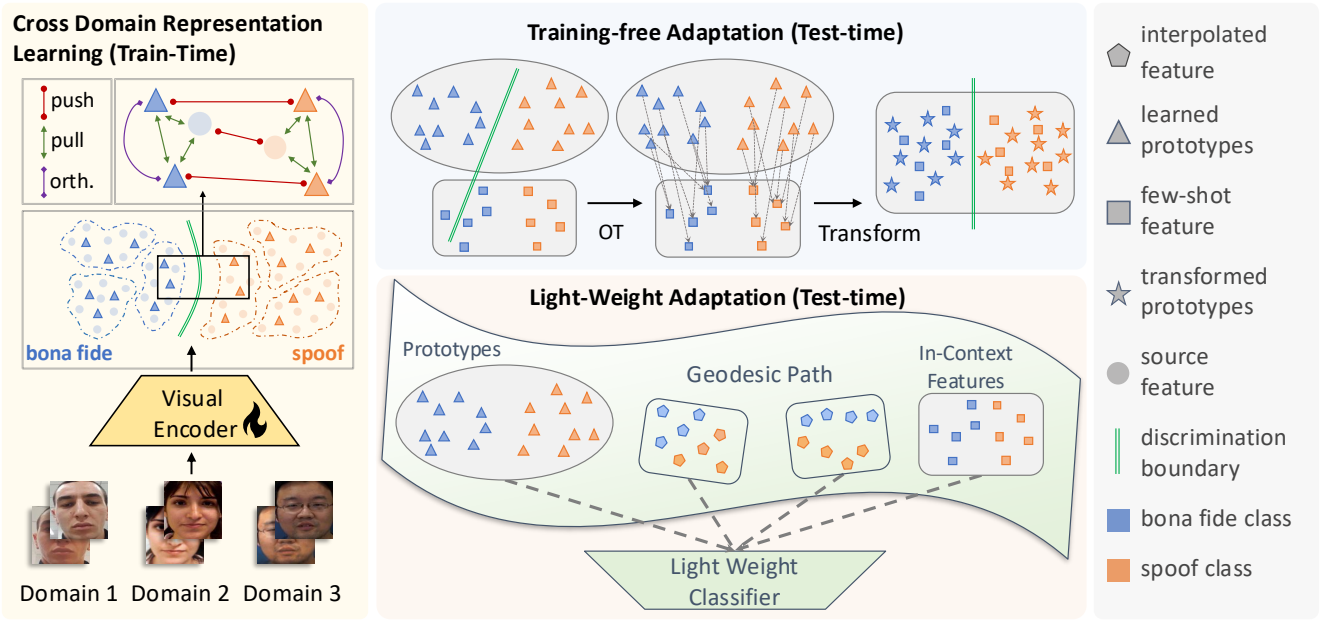

Figure 2. Overview of OTA. OTA learns a feature extractor and prototypes across multiple source domains during training. At test time, it adapts to few-shot client data via two approaches: a training-free method using optimal transport to shift source prototypes without learnable parameters, and a lightweight method training a classifier on synthetic data generated along the geodesic path, preserving source-domain understanding while adapting to target specifics.

$\boldsymbol{P}^T \boldsymbol{z}_i$ represent the centroid-wise similarity between an image embedding $\boldsymbol{z}_i \in \mathbb{R}^D$ and the prototypes $\boldsymbol{P} \in \mathbb{R}^{D \times K \times 2}$. We minimize the following loss function:

$$\mathcal{L}_{\text{proto}} = -\frac{1}{|I|} \sum_{i \in I} \log \frac{e^{s \cos(\theta_{y_i} + m)}}{e^{s \cos(\theta_{y_i} + m)} + e^{s \cos \theta_{1-y_i}}}, \quad (2)$$

where $\theta_{y_i} = \arccos\left(\frac{1}{K} \sum_{k=1}^{K} (\boldsymbol{P}_{y_i k}^T \boldsymbol{z}_i)\right)$, $y_i \in \{0, 1\}$ is the label for the image x_i, and $I = \{1, 2, \ldots, B\}$ is the index set of a batch. Unlike the original ArcFace loss, which minimizes the angular distance between a sample and its nearest sub-center, we reduce the cosine distance between a data embedding and its corresponding group of sub-centroids. The margin m is set as a learnable parameter. To avoid trivial mode collapse among sub-centroids, we regularize sub-centroids via a class-wise orthogonal loss: $\mathcal{L}_{\text{orth}} = \left\| \boldsymbol{p}^T \boldsymbol{p} - \mathbf{I} \right\|_2^2$, where $\mathbf{I}$ is the identity matrix.

The loss function in Eq. 2 establishes the relationship between data instances and their corresponding prototypes. To further improve intra-class alignment and inter-class separability, as well as preserve the geometric information among different source domains, we introduce an instance-wise supervised contrastive loss [27]:

$$\mathcal{L}_{\text{con}} = \sum_{i \in I_{\text{multi}}} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_p / \tau)}{\sum_{a \in A(i)} \exp(\boldsymbol{z}_i \cdot \boldsymbol{z}_a / \tau)},$$

where $I_{\text{multi}} = \{1, 2, \ldots, 2B\}$ is the index set of the multiviewed batch [27], $A(i) = I / \{i\}$ and $P(i) = \{p \in A(i) : \tilde{y}_p = \tilde{y}_i\}$. Our use of supervised contrastive loss is two-fold: a coarse supervision $\mathcal{L}_{\text{con}}^{\text{coarse}}$ involving only two classes—bona fide and spoof—and a fine-grained version $\mathcal{L}_{\text{con}}^{\text{fine}}$ that treats different source domains and attack methods as distinct sub-classes. Our final learning objective is:

$$\mathcal{L} = \mathcal{L}_{\text{proto}} + \alpha \mathcal{L}_{\text{con}}^{\text{coarse}} + \beta \mathcal{L}_{\text{con}}^{\text{fine}} + \eta \mathcal{L}_{\text{orth}}, \quad (3)$$

where the scalars α, β, and η regulate the strength of the coarse contrastive loss, fine-grained contrastive loss, and orthogonal term, respectively.

3.3. In-Context Adaptation Using Optimal Transportation

OTA considers a scenario where only the prototypes $\boldsymbol{p}$ and a few target-domain examples $\mathbb{D}_t$ are available at the adaptation time during the test stage. Our approach is to identify a transformation function that can adjust the prototypes $\boldsymbol{p}$ appropriately based on $\mathbb{D}_t$. A critical property of prototypes is that they encode rich information about the source domains within their geometric structure. Motivated by this, we propose to use Optimal Transportation (OT), which respects the geometries of both source and target distributions [72], to generate the transformation function.

Specifically, we derive this transformation function by solving the following regularized OT optimization problem:

$$\gamma^* = \underset{\gamma \in \mathbb{R}_+^{2K \times M_t}}{\arg\min} \sum_{i,j} \gamma_{i,j} M_{i,j} + \lambda \Omega_\alpha(\gamma) \quad \text{s.t. } \gamma \mathbf{1} = \boldsymbol{a};\ \gamma^T \mathbf{1} = \boldsymbol{b};\ \gamma \geq 0, \quad (4)$$

where γ^* represents the optimal transportation plan, and $\boldsymbol{a}, \boldsymbol{b}$ are the weights (summing to 1) of the source and target distributions, respectively. M is the cost matrix of size $K \times M_t$, where $2K$ is the number of prototypes, and M_t is the number of data points in the target domain. Each entry $M_{i,j}$ denotes transportation cost from source unit i to target unit j. Here, unit i refers to a sub-center from learned prototypes, and j corresponds to a feature from few-shot data extracted using f. The cost metric is defined as the cosine distance between i and j, aligning with our classification criteria in Sec. 3.2. Ω_α is a Laplacian regularization term that aims to preserve data structure during transport [18].

Once the OT plan γ^* is obtained, we use it as the transformation function to shift the learned prototypes toward the region where the few-shot data features reside, while preserving their original geometric information. Specifically, each shifted prototype center $\boldsymbol{p}^*$ is generated by a barycentric projection: $\boldsymbol{p}^* = \sum_{j=1}^{M_t} \pi_{i,j} \boldsymbol{z}_{t,j}$, where $\pi_{i,j} = \frac{M_{i,j}}{\sum_{j=1}^{M_t} M_{i,j}}$ is the normalized transport plan and $\boldsymbol{z}_{t,j}$ is the latent feature of the j-th target-domain data sample. These transformed prototypes $\boldsymbol{P}^* = \{\boldsymbol{p}_1^*, \ldots, \boldsymbol{p}_{2K}^*\}$ are then deployed as final classifiers. Note that although OT formulation is initially designed for unsupervised transformation [11], we perform class-wise transformation based on the label information to retain the discrimination capability of transformed prototypes.

3.4. Geodesic Mixup

In this section, we explore an alternative approach by conducting a lightweight training procedure that uses optimal transport in a different fashion. Specifically, we treat both the prototypes and few-shot target-domain data features as training data in the latent space and learn a separate decision boundary from scratch. It is well-known that classifiers trained on few-shot data can lead to skewed decision boundaries, which in turn result in poor performance [30, 66]. A common remedy for this issue involves generating synthetic data as augmentations [17, 61, 94]. Among these methods, mixup [71, 93] is a simple option that can also be applied in the embedding space. Nevertheless, such instance-wise interpolation does not account for global information (*i.e.*, geometry) of a distribution, which is critical for our scenario, generating inferior results due to unexpected artifacts [33].

To address this challenge, we propose leveraging discrete Wasserstein barycenters with free support [1] to generate synthesized data. In our approach, the problem involves only two distributions, empirically represented by prototypes from the source domains and few-shot data features from the target domain. Searching for barycenters is equivalent to interpolating along the Wasserstein geodesic between these source and target distributions. Concretely, given a mixing coefficient $w \in [0, 1]$, we look for a distribution μ that minimizes its Wasserstein distance to both the source and target distributions:

$$\mu = \min_{\mu} \big[wW(\mu, \mu_s) + (1 - w)W(\mu, \mu_t) \big], \tag{5}$$

where $W(\cdot, \cdot)$ denotes the Wasserstein distance between two distributions. In our case, μ_s is substantiated by learned source prototypes and μ_t contains few-shot target-domain features. μ is modeled by Q equally important supports: $\mu = \sum_{i=1}^{Q} \frac{1}{Q} \delta_{x_i}$, where δ_{x_i} refers to the Dirac function at position x_i. In practice, we set Q to be K, the total number of sub-centers within the learned prototypes for either bona fide or spoof. We then generate μ on the fly at each iteration as an augmented data batch according to a randomly sampled w from a beta distribution: $w \sim \text{Beta}(0.4, 0.4)$. Unlike point-wise interpolation (*i.e.*, standard mixup strategies), using generated μ respects the geometries of both source and target distributions, thus leading to a better discrimination boundary.

In implementation, we solve the entropy-regularized version of Eq. 5 that supports the efficient Sinkhorn algorithm [65] and reduces sparsity. To further encourage vicinity risk minimization, we also perturb few-shot data features at each iteration according to common data augmentation techniques (*e.g.*, color jittering, random Gaussian noise).

4. Experimental Results

In this section, we compare OTA with state-of-the-art FAS methods using the widely adopted cross-domain and cross-attack evaluation protocol. To further validate the effectiveness of OTA, we benchmark it against several test-time few-shot adaptation methods from general domains. Additionally, we conduct a series of ablation studies to evaluate the functionality of each component in OTA.

4.1. Experiment Settings

Datasets. We evaluate our method on four standard benchmarks: `Idiap Replay Attack` [9] (**I**), `OULU-NPU` [6] (**O**), `CASIA-MFSD` [98] (**C**), and `MSU-MFSD` [82] (**M**). Following the approach of previous works, we treat each dataset as a distinct domain and employ a leave-one-out testing protocol to evaluate cross-domain performance. For example, the protocol **OCI** $\rightarrow$ **M** indicates training on `OULU-NPU`, `CASIA-MFSD`, and `Idiap Replay Attack`, and testing on `MSU-MFSD`.

Evaluation Metrics. We use three commonly adopted metrics to quantify performance: Half Total Error Rate (HTER), Area Under the Receiver Operating Characteristic Curve (AUC), and True Positive Rate (TPR) at a False Positive Rate (FPR) of 1% (TPR@FPR=1%).

Implementation Details. We initialize the feature extractor using the pretrained ViT-B/16 model from CLIP [59]. The number of sub-centroids K for each class is set to 50 by

Method	OCI→M			OMI→C			OCM→I			ICM→O			Avg.
	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓
MADDG [62]	17.69	88.06	-	24.50	84.51	-	22.19	84.99	-	27.98	80.02	-	23.09
DR-MD-Net [76]	17.02	90.10	-	19.68	87.43	-	20.87	86.72	-	25.02	81.47	-	20.64
RFMeta [63]	13.89	93.98	-	20.27	88.16	-	17.30	90.48	-	16.45	91.16	-	16.97
NAS-FAS [91]	19.53	88.63	-	16.54	90.18	-	14.51	93.84	-	13.80	93.43	-	16.09
D^2AM [7]	12.70	95.66	-	20.98	85.58	-	15.43	91.22	-	15.27	90.87	-	16.09
SDA [78]	15.40	91.80	-	24.50	84.40	-	15.60	90.10	-	23.10	84.30	-	19.65
DRDG [43]	12.43	95.81	-	19.05	88.79	-	15.56	91.79	-	15.63	91.75	-	15.66
ANRL [42]	10.83	96.75	-	17.83	89.26	-	16.03	91.04	-	15.67	91.90	-	15.09
SSDG-R [25]	7.38	97.17	-	10.44	95.94	-	11.71	96.59	-	15.61	91.54	-	11.28
SSAN-R [80]	6.67	98.75	-	10.00	96.67	-	8.88	96.79	-	13.72	93.63	-	9.81
PatchNet [73]	7.10	98.46	-	11.33	94.58	-	13.40	95.67	-	11.82	95.07	-	10.91
SA-FAS [67]	5.95	96.55	-	8.78	95.37	-	6.58	97.54	-	10.00	96.23	-	7.82
IADG [103]	5.41	98.19	-	8.70	96.44	-	10.62	94.50	-	8.86	97.14	-	8.39
GDA [102]	9.20	98.00	-	12.20	93.00	-	10.00	96.00	-	14.40	92.60	-	11.45
HPDR [21]	4.58	96.02	-	11.30	94.42	-	11.26	92.49	-	9.93	95.26	-	9.27
SDA-FAS [47]	5.00	97.96	-	2.40	99.72	-	2.62	99.48	-	5.07	99.01	-	3.77
TTDG [105]	4.16	98.48	-	7.59	98.18	-	9.62	98.18	-	10.00	96.15	-	7.84
CFPL [41]	3.09	99.45	94.28	2.56	99.10	66.33	5.43	98.41	85.29	3.33	99.05	90.06	3.60
OTA * (zero-shot)	2.62	99.34	92.38	2.22	99.49	90.67	5.32	98.44	89.00	3.56	99.34	88.64	3.43
ViTA [22] (5-shot)	4.75	98.84	76.67	5.00	99.13	82.14	5.37	98.57	76.15	7.16	97.97	73.24	5.57
ViTAF [22] (5-shot)	3.42	99.30	88.33	1.40	99.85	95.71	3.74	99.34	85.38	7.17	98.26	71.97	3.93
OTA † (training-free)	2.38	99.42	93.33	2.67	99.49	91.11	5.19	98.56	88.22	3.03	99.45	90.66	3.21
OTA ‡ (lightweight)	**2.14**	**99.47**	**95.23**	**2.00**	**99.75**	**93.79**	**4.85**	**98.81**	**91.30**	**2.61**	**99.52**	**92.30**	**2.91**

Table 1. The results (%) of cross-domain evaluation on MSU-MFSD (M), CASIA-FASD (C), ReplayAttack (I), and OULU-NPU (O) datasets. Note that symbols $*$, †, and ‡ indicate three versions of OTA: zero-shot (Section 3.2), training-free domain adaptation (Section 3.3), and lightweight training domain adaptation (Section 3.4), respectively. Baseline results are sourced from Liu et al. [41]. Both of our approaches outperform existing DG methods and achieve competitive performance against FSL methods (colored in gray). Among our proposed approaches, the lightweight training method delivers the best performance, as it trains with augmented feature data from geodesic mixup to more effectively capture the geometry of the feature space.

Methods	AUC↑
SVM1+IMQ [2]	$70.23^{12.69}$
CDCN++ [92]	$87.53^{10.90}$
SSAN [81]	$88.01^{9.93}$
TTN-S [79]	$89.71^{9.17}$
UDG-FAS [50]	$92.43^{6.86}$
GAC-FAS [34]	$93.39^{4.27}$
OTA * (zero-shot)	$\mathbf{98.32}^{0.24}$
OTA † (training-free)	$\mathbf{98.38}^{0.10}$
OTA ‡ (lightweight)	$\mathbf{98.54}^{0.29}$

(a) Unseen 2D attack

Method	AUC↑
Saha *et al.* [60]	79.20
Panwar *et al.* [57]	80.00
SSDG-R [26]	82.11
CIFAS [48]	83.20
UDG-FAS [50]	87.26
GAC-FAS [34]	$89.27^{0.58}$
OTA * (zero-shot)	$\mathbf{98.62}^{0.52}$
OTA † (training-free)	$\mathbf{98.75}^{0.79}$
OTA ‡ (lightweight)	$\mathbf{99.63}^{0.11}$

(b) Unseen 3D attack

Table 2. Evaluation on cross-attack protocol following Arashloo et al. [2]. Baseline results are sourced from Le and Woo [34].

default, and the strength scalars α, β, and η are set to 0.01, 0.01, and 1.0, respectively. At the test stage, we use 10 randomly selected images per class from the target domain as few-shot examples, and the strength λ of the Laplacian regularization term is fixed at 100. For learning a separate decision boundary, we optimize a linear classifier for 100 iterations using standard cross-entropy loss and the Adam optimizer [29] with a learning rate of 0.01.

4.2. Comparison to SoTA methods

Cross-domain Evaluation. Table 1 presents the cross-domain performance of OTA under a comprehensive evaluation protocol. As shown, although our primary focus is not on improving domain generalization (DG) performance, the proposed prototype-based framework achieves state-of-the-art (SoTA) DG performance (indicated by $*$), demonstrating the effectiveness of our proposed representation learning described in Section 3.2. Without additional training, OTA improves performance on the target domain by leveraging few-shot examples (denoted by †), particularly excelling in the most challenging metric, TPR@FPR=1%. Our lightweight training with geodesic mixup (indicated by ‡) improves performance further, surpassing existing methods by a clear margin. Unlike previous few-shot approaches in the FAS domain, such as ViTAF [22], which assumes that few-shot data from the target domain is available during the training phase, our approach treats few-shot data as examples that can only be incorporated at the test stage. This makes OTA a more practical yet challenging setting.

Cross-attack Evaluation. In real-world applications, security models must contend with increasingly sophisticated attack methods. Evaluating OTA against unseen attacks is critical to validate its robustness. Following the protocol proposed in [2], we simulate unseen 2D attacks by training on two domains from **I**, **C**, and **M**, and testing on an unseen attack from an unseen domain. For 3D attacks, we train on **O**, **C**, and **M**, and evaluate using a subset from the CelebA-Spoof dataset [97]. As shown in Table 2, OTA exhibits strong generalization performance when confronted with unseen attacks. Additionally, both the training-free and lightweight adaptation variants further improve performance by incorporating few-shot examples of the new attack during test time.

Method	OCI→M			OMI→C			OCM→I			ICM→O			Avg.
	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓
NCM [53]	3.25	98.16	29.68	3.15	98.23	31.48	7.97	96.60	31.57	3.28	99.43	89.03	4.41
Tip-Adaptor [95]	3.73	98.87	85.08	3.37	99.28	89.83	9.28	93.44	29.53	3.05	**99.46**	88.87	4.84
OTA † (training-free)	**2.38**	**99.42**	**93.33**	**2.67**	**99.49**	**91.11**	**5.19**	**98.56**	**88.22**	**3.03**	99.45	**90.66**	**3.21**
Linear Probe [59]	2.14	99.46	78.57	2.22	99.64	92.89	5.38	98.74	88.00	3.12	99.41	90.80	3.22
Manifold Mixup [71]	2.14	99.45	69.05	2.16	99.63	93.10	5.32	98.80	90.1	2.82	99.42	91.19	3.04
OTA ‡ (lightweight)	**2.14**	**99.47**	**95.23**	**2.00**	**99.75**	**93.79**	**4.85**	**98.81**	**91.30**	**2.61**	**99.52**	**92.30**	**2.91**

Table 3. Comparison to few-shot adaptation methods. We compare OTA against several representative few-shot adaptation methods in the FAS context. Our method achieves overall better performance in HTER.

Method	OCI→M			OMI→C			OCM→I			ICM→O			Avg.
	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓
SSDG-R [26]	$22.84^{1.14}$	$78.67^{1.31}$	-	$28.76^{0.89}$	$80.91^{1.10}$	-	$14.65^{1.21}$	$91.93^{1.35}$	-	$15.83^{1.29}$	$92.13^{0.96}$	-	20.52
SSAN-R [81]	$21.79^{3.68}$	$84.06^{3.78}$	-	$26.44^{2.91}$	$78.84^{2.83}$	-	$35.39^{8.04}$	$70.13^{9.03}$	-	$25.72^{3.74}$	$79.37^{4.69}$	-	27.34
PatchNet [74]	$25.92^{1.13}$	$83.43^{0.87}$	-	$36.26^{1.98}$	$71.38^{1.89}$	-	$29.75^{2.76}$	$80.53^{1.35}$	-	$23.49^{1.80}$	$84.62^{1.92}$	-	28.86
SA-FAS [68]	$14.36^{1.10}$	$92.06^{0.53}$	-	$19.40^{0.66}$	$88.69^{0.67}$	-	$11.48^{1.10}$	$95.74^{0.55}$	-	$11.29^{0.32}$	$95.23^{0.24}$	-	14.13
GAC-FAS [34]	$12.29^{1.29}$	$95.35^{0.57}$	-	$15.37^{1.52}$	$91.67^{1.67}$	-	$12.51^{3.03}$	$93.03^{2.24}$	-	$9.89^{0.47}$	$96.44^{0.18}$	-	12.52
OTA *(zero-shot)	$\mathbf{4.52}^{0.70}$	$\mathbf{99.18}^{0.25}$	$\mathbf{86.19}^{5.10}$	$\mathbf{4.33}^{0.89}$	$\mathbf{99.23}^{0.20}$	$\mathbf{88.00}^{1.75}$	$\mathbf{7.98}^{1.33}$	$\mathbf{96.64}^{0.67}$	$\mathbf{84.97}^{1.33}$	$\mathbf{4.36}^{0.15}$	$\mathbf{99.09}^{0.09}$	$\mathbf{86.50}^{1.71}$	**5.29**

Table 4. Evaluation at convergence. Following the evaluation method proposed in [68], we compare the zero-shot version of OTA with baseline methods upon convergence. The baseline results are sourced from [34]. OTA consistently exhibits superior convergence performance measured by HTER, AUC, and TPR@FPR=1%.

4.3. Comparison to Few-shot Adaptation Methods

In this section, we examine the effectiveness of OTA by comparing it with other comparable test-time few-shot adaptation methods. Specifically, we implement the following four baseline methods, covering both training-free and lightweight adaptation schemes.

Nearest Class Mean (NCM) [53]. We implement NCM using only class mean feature from few-shot data as the classifier. As shown in Table 3, NCM lags behind OTA by a significant margin, highlighting the importance of knowledge transfer from source domains under few-shot conditions.

Tip-Adaptor [95]. We compare OTA with the state-of-the-art training-free few-shot adaptation method. Tip-Adaptor relies on an additional validation set for hyperparameter selection. Hence, we build a validation set with 256 randomly selected images from existing source domains. As shown in Table 3, OTA demonstrates a clear advantage over Tip-Adaptor and even matches its performance when a large additional validation set is provided for Tip-Adaptor.

Linear Probe [59]. Training a linear probe over a frozen feature extractor is a simple yet competitive method [24] when lightweight training is allowed. Here, we train a linear probe over a mixture of our learned prototypes and few-shot features. As shown in Table 3, linear probes achieve strong performance, outperforming other training-free adaptation methods. However, OTA demonstrates a noticeable improvement over a simple linear probe.

Manifold Mixup [71]. We compare OTA with manifold mixup, which also performs latent space augmentation to boost performance over a linear probe. However, unlike this instance-wise augmentation, our proposed geodesic mixup generates distribution-wise augmentations, which is more effective under our setup.

4.4. Ablation Studies

Performance Upon Convergence. A recent study [67] points out that a single snapshot of performance on a test set may not accurately reflect a detection model's generalization capability. Following their approach, we report the average performance of our model across the last 10 evaluations in Table 4. OTA consistently exhibits superior convergence performance measured by HTER, AUC, and TPR@FPR=1%, while showing significantly less fluctuation than other methods. For example, as shown in Table 1, the converged HTER score of SA-FAS increases from 7.82 to 14.13. In contrast, OTA achieves HTER scores of 3.43 at the snapshot and 5.29 at convergence.

Effectiveness of Learning Objectives. Our proposed prototype-based framework combines a prototype-based margin loss, coarse-to-fine supervised contrastive loss, and an orthogonal loss for supervision. Using the cross-attack (3D) protocol as an example, we show in Table 6 the effectiveness of each proposed loss component. Each term effectively improves the DG performance of our prototype-based framework.

Sub-centroid Number. To increase the expressiveness of prototypes and better capture the geometric information of source domains, we employ a multi-centroid strategy. Using the cross-attack (3D) protocol as an example, we run experiments with the number of sub-centers $K \in \{1, 5, 10, 30, 50\}$, yielding HTER scores of $\{9.40, 6.33, 5.21, 5.07, 5.20\}$, respectively. Increasing the number of centroids effectively improves domain generalization (DG) performance, and the improvement plateaus around 50 sub-centers; we observed similar trends in other

Method	OCI→M			OMI→C			OCM→I			ICM→O		
	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑	HTER↓	AUC↑	TPR@ FPR=1%↑
OTA (DG)	2.62	99.34	90.49	2.22	99.49	90.67	**5.32**	**98.44**	**88.13**	3.56	99.34	88.64
OTA (spoof only)	2.38	99.40	90.49	2.33	99.49	**91.11**	5.83	98.24	88.00	3.53	99.35	88.41
OTA (bona fide only)	**2.38**	**99.42**	**92.87**	**2.22**	**99.53**	90.89	6.42	98.29	85.50	**3.38**	**99.36**	**89.65**
OTA (both)	2.38	99.42	93.33	2.67	99.49	91.11	5.19	98.56	88.22	3.03	99.45	90.66

Table 5. **Empirical results of OTA under the one-class setting.** Providing few-shot data from the bona fide class proves to be more effective than from the spoofing class in most scenarios. One exception is OCM→I, where I focuses exclusively on replay attacks under controlled environments. Consequently, both the bona fide and spoof data of **I** differ markedly from those of the other three datasets, and OTA needs both bona fide and spoof data to achieve satisfactory adaptation results.

proto.	orth.	coarse con.	fine con.	HTER(%)↓	AUC(%)↑	TPR(%) ↑ @FPR=1%
✓	-	-	-	5.20	98.37	81.34
✓	✓	-	-	5.14	98.38	82.68
✓	✓	✓	-	4.87	98.51	84.81
✓	✓	✓	✓	**4.19**	**98.62**	**87.35**

Table 6. Ablation of learning objectives introduced in Section 3.2. Each loss term effectively improves the DG performance of our prototype-based framework.

experiments. Consequently, we set $K = 50$ for all subsequent experiments.

Visualizations. We visualize the transformed prototypes and synthesized distributions generated by geodesic mixup in the latent space to provide a clear illustration of our method. As shown in Fig. 3 (L), the transformed prototypes are relocated closer to the regions where the few-shot client data features reside, while preserving the geometric information of the original prototypes learned from the source domains. For geodesic mixup, Fig. 3 (R) shows that synthesized distributions are gradually transformed from the source distribution toward the target distribution as we vary the weight w from 0.1 to 0.9.

Efficiency. The training-free variant of OTA introduces zero additional learnable parameters and requires a minimal adaptation time of only 0.17 ± 0.03 minutes, estimated over 10 trials. For lightweight adaptation with geodesic mixup, the additional parameter load is 3.9 KB, and the adaptation process completes in 22.73 ± 3.51 minutes, also estimated over 10 trials. Our customizable layer offers maximum flexibility and convenience for customization and maintenance by both the host and clients.

4.5. One-class Scenario

We extend the training-free version of our approach to the one-class setting, where only a few data points from either the bona fide or spoof class are provided by the client. We transform only the prototypes corresponding to the class that has associated few-shot data, while keeping the other class's prototypes unchanged. As shown in Table 5, providing few-shot data from either the bona fide or spoof category helps improve performance compared to DG in most cases. Interestingly, under the one-class setting, providing few-shot data from the bona fide class appears to be more effective than from the spoofing class. We hypothesize that this is because the bona fide class is inherently more compact, so fewer data samples can better represent the target bona fide distribution.

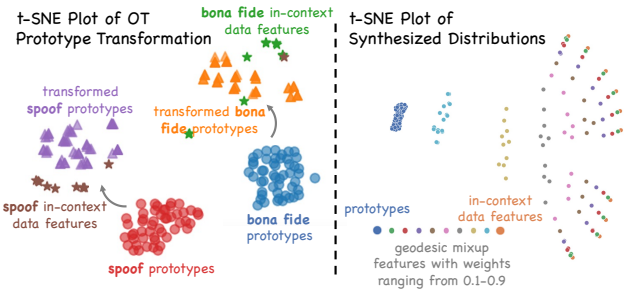

Figure 3. (L) The transformed prototypes are relocated closer to the regions where the few-shot client data features reside, while preserving the geometric information of the original prototypes learned from the source domains. (R) Synthesized distributions are gradually transformed from the source prototypes toward the target few-shot data features, as weight w is varied from 0.1 to 0.9.

5. Conclusion

In this work, we explore a practical yet under-explored scenario in the face anti-spoofing literature, where only a limited set of labeled target-domain data is available at the test stage, and the client portal has no access to the host model. To address this challenge, we propose a prototype-based backbone model, on top of which two efficient adaptation modules are crafted for training-free and lightweight training settings. In particular, we introduce geodesic mixup, an optimal transport-guided synthesis method that generates pseudo empirical distributions as augmentations, thereby improving the learned decision boundary. Extensive empirical results on four benchmarks show that our prototype-based framework achieves performance on par with state-of-the-art domain generalization methods. Furthermore, its performance can be boosted by incorporating few-shot target-domain data at test time, leveraging either the proposed training-free or lightweight training modules. Our approach, which can be integrated with other face recognition algorithms, establishes a simple yet strong baseline for future research.

References

[1] Pedro C Álvarez-Esteban, E Del Barrio, JA Cuesta-Albertos, and C Matrán. A fixed-point approach to barycenters in wasserstein space. *Journal of Mathematical Analysis and Applications*, 441(2):744–762, 2016. 5

[2] Shervin Rahimzadeh Arashloo, Josef Kittler, and William Christmas. An anomaly detection approach to face spoofing detection: A new formulation and evaluation protocol. *IEEE Access*, 2017. 6

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *ICML*. PMLR, 2017. 3

[4] Yogesh Balaji, Rama Chellappa, and Soheil Feizi. Normalized wasserstein for mixture distributions with applications in adversarial learning and domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6500–6508, 2019. 3

[5] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face anti-spoofing based on color texture analysis. In *2015 IEEE international conference on image processing (ICIP)*, pages 2636–2640. IEEE, 2015. 2

[6] Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. 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)*, pages 612–618. IEEE, 2017. 5

[7] Zhihong Chen, Taiping Yao, Kekai Sheng, Shouhong Ding, Ying Tai, Jilin Li, Feiyue Huang, and Xinyu Jin. Generalizable representation learning for mixture domain face anti-spoofing. In *Proceedings of the AAAI conference on artificial intelligence*, pages 1132–1139, 2021. 2, 6

[8] Girija Chetty. Biometric liveness checking using multimodal fuzzy fusion. In *International Conference on Fuzzy Systems*, pages 1–8. IEEE, 2010. 2

[9] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG)*, pages 1–7. IEEE, 2012. 5

[10] Nicolas Courty, Rémi Flamary, and Devis Tuia. Domain adaptation with regularized optimal transport. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part I 14*, pages 274–289. Springer, 2014. 3

[11] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1853–1865, 2016. 3, 5

[12] Nicolas Courty, Rémi Flamary, Amaury Habrard, and Alain Rakotomamonjy. Joint distribution optimal transportation for domain adaptation. *Advances in neural information processing systems*, 30, 2017. 3

[13] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013. 2, 3

[14] Tiago de Freitas Pereira, André Anjos, José Mario De Martino, and Sébastien Marcel. Lbp- top based countermeasure against face spoofing attacks. In *ACCV*, 2012. 2

[15] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019. 3

[16] Aditya Deshpande, Alessandro Achille, Avinash Ravichandran, Hao Li, Luca Zancato, Charless Fowlkes, Rahul Bhotika, Stefano Soatto, and Pietro Perona. A linearized framework and a new benchmark for model selection for fine-tuning. *arXiv:2102.00084*, 2021. 1

[17] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. *arXiv:1708.04552*, 2017. 5

[18] Sira Ferradans, Nicolas Papadakis, Gabriel Peyré, and Jean-François Aujol. Regularized discrete optimal transport. *SIAM Journal on Imaging Sciences*, 7(3):1853–1882, 2014. 2, 3, 5

[19] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in neural information processing systems*, 30, 2017. 3

[20] Xiao Guo, Yaojie Liu, Anil Jain, and Xiaoming Liu. Multi-domain learning for updating face anti-spoofing models. In *European Conference on Computer Vision*, pages 230–249. Springer, 2022. 2

[21] Chengyang Hu, Ke-Yue Zhang, Taiping Yao, Shouhong Ding, and Lizhuang Ma. Rethinking generalizable face anti-spoofing via hierarchical prototype-guided distribution refinement in hyperbolic space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1032–1041, 2024. 6

[22] Hsin-Ping Huang, Deqing Sun, Yaojie Liu, Wen-Sheng Chu, Taihong Xiao, Jinwei Yuan, Hartwig Adam, and Ming-Hsuan Yang. Adaptive transformers for robust few-shot cross-domain face anti-spoofing. In *European conference on computer vision*, pages 37–54. Springer, 2022. 2, 6

[23] Pei-Kai Huang, Chen-Yu Lu, Shu-Jung Chang, Jun-Xiong Chong, and Chiou-Ting Hsu. Test-time adaptation for robust face anti-spoofing. In *BMVC*, pages 379–380, 2023. 3

[24] Yunshi Huang, Fereshteh Shakeri, Jose Dolz, Malik Boudiaf, Houda Bahig, and Ismail Ben Ayed. Lp++: A surprisingly strong linear probe for few-shot clip. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23773–23782, 2024. 7

[25] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Single-side domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8484–8493, 2020. 2, 6

[26] Yunpei Jia, Jie Zhang, Shiguang Shan, and Xilin Chen. Single-side domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vi-*

sion and Pattern Recognition, pages 8484–8493, 2020. 6, 7

[27] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33: 18661–18673, 2020. 4

[28] Youngeun Kim, Donghyeon Cho, Kyeongtak Han, Priyadarshini Panda, and Sungeun Hong. Domain adaptation without source data. *IEEE Transactions on Artificial Intelligence*, 2(6):508–518, 2021. 2

[29] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2014. 6

[30] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, pages 1–30. Lille, 2015. 1, 5

[31] Klaus Kollreider, Hartwig Fronthaler, Maycel Isaac Faraj, and Josef Bigun. Real-time face detection and motion analysis with application in "liveness" assessment. *IEEE Transactions on Information Forensics and Security*, 2(3):548–558, 2007. 2

[32] Jukka Komulainen, Abdenour Hadid, and Matti Pietikäinen. Context based face anti-spoofing. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2013. 2

[33] Alex Lamb, Vikas Verma, Juho Kannala, and Yoshua Bengio. Interpolated adversarial training: Achieving robust neural networks without sacrificing too much accuracy. In *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security*, pages 95–103, 2019. 5

[34] Binh M Le and Simon S Woo. Gradient alignment for cross-domain face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 188–199, 2024. 6, 7

[35] Tien-Nam Le, Amaury Habrard, and Marc Sebban. Deep multi-wasserstein unsupervised domain adaptation. *Pattern Recognition Letters*, 125:249–255, 2019. 3

[36] Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10285–10295, 2019. 3

[37] Suho Lee, Seungwon Seo, Jihyo Kim, Yejin Lee, and Sangheum Hwang. Few-shot fine-tuning is all you need for source-free domain adaptation. *arXiv preprint arXiv:2304.00792*, 2023. 3

[38] Haoliang Li, Wen Li, Hong Cao, Shiqi Wang, Feiyue Huang, and Alex C Kot. Unsupervised domain adaptation for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 13(7):1794–1809, 2018. 2

[39] Jiangwei Li, Yunhong Wang, Tieniu Tan, and Anil K Jain. Live face detection based on the analysis of fourier spectra. In *Biometric technology for human identification*, pages 296–303, 2004. 2

[40] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International conference on machine learning*, pages 6028–6039. PMLR, 2020. 3

[41] Ajian Liu, Shuai Xue, Jianwen Gan, Jun Wan, Yanyan Liang, Jiankang Deng, Sergio Escalera, and Zhen Lei. Cfpl-fas: Class free prompt learning for generalizable face anti-spoofing. pages 222–232, 2024. 2, 6

[42] Shubao Liu, Ke-Yue Zhang, Taiping Yao, Mingwei Bi, Shouhong Ding, Jilin Li, Feiyue Huang, and Lizhuang Ma. Adaptive normalized representation learning for generalizable face anti-spoofing. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 1469–1477, 2021. 6

[43] Shubao Liu, Ke-Yue Zhang, Taiping Yao, Kekai Sheng, Shouhong Ding, Ying Tai, Jilin Li, Yuan Xie, and Lizhuang Ma. Dual reweighting domain generalization for face presentation attack detection. *arXiv:2106.16128*, 2021. 6

[44] Yaojie Liu and Xiaoming Liu. Spoof trace disentanglement for generic face anti-spoofing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3813–3830, 2022. 2

[45] Yaojie Liu, Joel Stehouwer, Amin Jourabloo, and Xiaoming Liu. Deep tree learning for zero-shot face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4680–4689, 2019.

[46] Yaojie Liu, Joel Stehouwer, and Xiaoming Liu. On disentangling spoof trace for generic face anti-spoofing. In *ECCV*, pages 406–422. Springer, 2020. 2

[47] Yuchen Liu, Yabo Chen, Wenrui Dai, Mengran Gou, Chun-Ting Huang, and Hongkai Xiong. Source-free domain adaptation with contrastive domain alignment and self-supervised exploration for face anti-spoofing. In *European Conference on Computer Vision*, pages 511–528. Springer, 2022. 2, 3, 6

[48] Yuchen Liu, Yabo Chen, Wenrui Dai, Chenglin Li, Junni Zou, and Hongkai Xiong. Causal intervention for generalizable face anti-spoofing. In *ICME*, 2022. 2, 6

[49] Yuchen Liu, Yabo Chen, Mengran Gou, Chun-Ting Huang, Yaoming Wang, Wenrui Dai, and Hongkai Xiong. Towards unsupervised domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20654–20664, 2023. 2

[50] Yuchen Liu, Yabo Chen, Mengran Gou, Chun-Ting Huang, Yaoming Wang, Wenrui Dai, and Hongkai Xiong. Towards unsupervised domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20654–20664, 2023. 6

[51] Yuchen Liu, Yabo Chen, Wenrui Dai, Mengran Gou, Chun-Ting Huang, and Hongkai Xiong. Source-free domain adaptation with domain generalized pretraining for face anti-spoofing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2

[52] Lingling Lv, Youjun Xiang, Xianfeng Li, Hanye Huang, Rongju Ruan, Xiaoyan Xu, and Yuli Fu. Combining dynamic image and prediction ensemble for cross-domain face anti-spoofing. In *ICASSP 2021-2021 IEEE Interna-*

tional Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2550–2554. IEEE, 2021. 3

[53] Zheda Mai, Ruiwen Li, Hyunwoo Kim, and Scott Sanner. Supervised contrastive replay: Revisiting the nearest class mean classifier in online class-incremental continual learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3589–3599, 2021. 7

[54] Shiyun Mao, Ruolin Chen, and Huibin Li. Weighted joint distribution optimal transport based domain adaptation for cross-scenario face anti-spoofing. *International Journal of Computer Vision*, pages 1–21, 2024. 3

[55] Eduardo Fernandes Montesuma and Fred Maurice Ngole Mboula. Wasserstein barycenter for multi-source domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16785–16793, 2021. 3

[56] Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *ICCV*, 2007. 2

[57] Ankush Panwar, Pratyush Singh, Suman Saha, Danda Pani Paudel, and Luc Van Gool. Unsupervised compound domain adaptation for face anti-spoofing. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–8. IEEE, 2021. 6

[58] Keyurkumar Patel, Hu Han, and Anil K Jain. Secure face unlock: Spoof detection on smartphones. *IEEE TIFS*, 2016. 2

[59] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 5, 7

[60] Suman Saha, Wenhao Xu, Menelaos Kanakis, Stamatios Georgoulis, Yuhua Chen, Danda Pani Paudel, and Luc Van Gool. Domain agnostic feature learning for image and video based face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 802–803, 2020. 6

[61] Ali Shafahi, Mahyar Najibi, Mohammad Amin Ghiasi, Zheng Xu, John Dickerson, Christoph Studer, Larry S Davis, Gavin Taylor, and Tom Goldstein. Adversarial training for free! *Advances in neural information processing systems*, 32, 2019. 5

[62] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C. Yuen. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *CVPR*, 2019. 2, 6

[63] Rui Shao, Xiangyuan Lan, and Pong C Yuen. Regularized fine-grained meta face anti-spoofing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11974–11981, 2020. 2, 6

[64] Jian Shen, Yanru Qu, Weinan Zhang, and Yong Yu. Wasserstein distance guided representation learning for domain adaptation. In *Proceedings of the AAAI conference on artificial intelligence*, 2018. 3

[65] Richard Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, 74(4):402–405, 1967. 5

[66] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017. 1, 5

[67] Yiyou Sun, Yaojie Liu, Xiaoming Liu, Yixuan Li, and Wen-Sheng Chu. Rethinking domain generalization for face anti-spoofing: Separability and alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 24563–24574, 2023. 2, 6, 7

[68] Yiyou Sun, Yaojie Liu, Xiaoming Liu, Yixuan Li, and Wen-Sheng Chu. Rethinking domain generalization for face anti-spoofing: Separability and alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24563–24574, 2023. 7

[69] Korawat Tanwisuth, Xinjie Fan, Huangjie Zheng, Shujian Zhang, Hao Zhang, Bo Chen, and Mingyuan Zhou. A prototype-oriented framework for unsupervised domain adaptation. *Advances in Neural Information Processing Systems*, 34:17194–17208, 2021. 3

[70] Chandra Thapa, Pathum Chamikara Mahawaga Arachchige, Seyit Camtepe, and Lichao Sun. Splitfed: When federated learning meets split learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8485–8493, 2022. 2

[71] Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Yoshua Bengio. Manifold mixup: Better representations by interpolating hidden states. In *International conference on machine learning*, pages 6438–6447. PMLR, 2019. 2, 3, 5, 7

[72] Cédric Villani et al. *Optimal transport: old and new*. Springer, 2009. 4

[73] Chien-Yi Wang, Yu-Ding Lu, Shang-Ta Yang, and Shang-Hong Lai. Patchnet: A simple face anti-spoofing framework via fine-grained patch recognition. pages 20281–20290, 2022. 2, 6

[74] Chien-Yi Wang, Yu-Ding Lu, Shang-Ta Yang, and Shang-Hong Lai. 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*, pages 20281–20290, 2022. 7

[75] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Improving cross-database face presentation attack detection via adversarial domain adaptation. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019. 2

[76] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection. *TIFS*, 2020. 2, 6

[77] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 16:56–69, 2020. 2

[78] Jingjing Wang, Jingyi Zhang, Ying Bian, Youyi Cai, Chunmao Wang, and Shiliang Pu. Self-domain adaptation for

face anti-spoofing. In *Proceedings of the AAAI conference on artificial intelligence*, pages 2746–2754, 2021. 2, 6

[79] Zhuo Wang, Qiangchang Wang, Weihong Deng, and Guodong Guo. Learning multi-granularity temporal characteristics for face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 17:1254–1269, 2022. 6

[80] Zhuo Wang, Zezheng Wang, Zitong Yu, Weihong Deng, Jiahong Li, Tingting Gao, and Zhongyuan Wang. Domain generalization via shuffled style assembly for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4123–4133, 2022. 2, 6

[81] Zhuo Wang, Zezheng Wang, Zitong Yu, Weihong Deng, Jiahong Li, Tingting Gao, and Zhongyuan Wang. Domain generalization via shuffled style assembly for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4123–4133, 2022. 6, 7

[82] Di Wen, Hu Han, and Anil K Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015. 5

[83] Bingrong Xu, Zhigang Zeng, Cheng Lian, and Zhengming Ding. Few-shot domain adaptation via mixup optimal transport. *IEEE Transactions on Image Processing*, 31:2518–2528, 2022. 3

[84] Minghao Xu, Jian Zhang, Bingbing Ni, Teng Li, Chengjie Wang, Qi Tian, and Wenjun Zhang. Adversarial domain adaptation with domain mixup. In *Proceedings of the AAAI conference on artificial intelligence*, pages 6502–6509, 2020. 3

[85] Pengcheng Xu, Prudhvi Gurram, Gene Whipps, and Rama Chellappa. Wasserstein distance based domain adaptation for object detection. *arXiv preprint arXiv:1909.08675*, 2019. 3

[86] Xiang Xu, Xiong Zhou, Ragav Venkatesan, Gurumurthy Swaminathan, and Orchid Majumder. d-sne: Domain adaptation using stochastic neighborhood embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2497–2506, 2019. 2

[87] Xiang Xu, Yuanjun Xiong, and Wei Xia. On improving temporal consistency for online face liveness detection system. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 824–833, 2021. 2

[88] Xiang Xu, Tianchen Zhao, Zheng Zhang, Zhihua Li, Jon Wu, Alessandro Achille, and Mani Srivastava. Principles of designing robust remote face anti-spoofing systems. *arXiv:2406.03684*, 2024. 1, 2

[89] Shen Yan, Huan Song, Nanxiang Li, Lincan Zou, and Liu Ren. Improve unsupervised domain adaptation with mixup training. *arXiv:2001.00677*, 2020. 3

[90] Jianwei Yang, Zhen Lei, Shengcai Liao, and Stan Z Li. Face liveness detection with component dependent descriptor. In *ICB*, 2013. 2

[91] Zitong Yu, Jun Wan, Yunxiao Qin, Xiaobai Li, Stan Z. Li, and Guoying Zhao. Nas-fas: Static-dynamic central difference network search for face anti-spoofing. In *TPAMI*, 2020. 6

[92] Zitong Yu, Chenxu Zhao, Zezheng Wang, Yunxiao Qin, Zhuo Su, Xiaobai Li, Feng Zhou, and Guoying Zhao. Searching central difference convolutional networks for face anti-spoofing. In *CVPR*, 2020. 6

[93] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv:1710.09412*, 2017. 2, 3, 5

[94] Ruixiang Zhang, Tong Che, Zoubin Ghahramani, Yoshua Bengio, and Yangqiu Song. Metagan: An adversarial approach to few-shot learning. *Advances in neural information processing systems*, 31, 2018. 5

[95] Renrui Zhang, Rongyao Fang, Wei Zhang, Peng Gao, Kunchang Li, Jifeng Dai, Yu Qiao, and Hongsheng Li. Tip-adapter: Training-free clip-adapter for better vision-language modeling. *arXiv:2111.03930*, 2021. 7

[96] Wenyu Zhang, Li Shen, Wanyue Zhang, and Chuan-Sheng Foo. Few-shot adaptation of pre-trained networks for domain shift. *arXiv preprint arXiv:2205.15234*, 2022. 3

[97] Yuanhan Zhang, ZhenFei Yin, Yidong Li, Guojun Yin, Junjie Yan, Jing Shao, and Ziwei Liu. Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*, pages 70–85. Springer, 2020. 6

[98] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z Li. A face antispoofing database with diverse attacks. In *2012 5th IAPR international conference on Biometrics (ICB)*, pages 26–31. IEEE, 2012. 5

[99] Han Zhao, Remi Tachet Des Combes, Kun Zhang, and Geoffrey Gordon. On learning invariant representations for domain adaptation. In *International conference on machine learning*, pages 7523–7532. PMLR, 2019. 3

[100] Tianchen Zhao, Xiang Xu, Mingze Xu, Hui Ding, Yuanjun Xiong, and Wei Xia. Learning self-consistency for deepfake detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15023–15033, 2021. 2

[101] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Ran Yi, Shouhong Ding, and Lizhuang Ma. Adaptive mixture of experts learning for generalizable face anti-spoofing. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6009–6018, 2022. 2

[102] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Ran Yi, Kekai Sheng, Shouhong Ding, and Lizhuang Ma. Generative domain adaptation for face anti-spoofing. In *European Conference on Computer Vision*, pages 335–356. Springer, 2022. 2, 6

[103] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Ran Yi, Shouhong Ding, and Lizhuang Ma. Instance-aware domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20453–20463, 2023. 6

[104] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Shouhong Ding, and Lizhuang Ma. Test-time domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 175–187, 2024. 2

[105] Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Shouhong Ding, and Lizhuang Ma. Test-time domain generalization for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 175–187, 2024. 2, 6