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

Supplementary Material

1. Applications of Our Settings

It is challenging to build a generalized training set that adequately covers all possible environmental conditions, user behaviors, demographic diversity, and specific clients' requirements that may sometimes conflict with one another. For example, consider a scenario where Client A authenticates users through mobile phone-based checks, while Client B uses entrance kiosks at fixed locations to capture data from specific angles. These diverse setups introduce variations in the data capture process, including acquisition devices, pre-processing techniques, and challenge-response mechanisms. As a result, there is a domain mismatch between the client data and the training data used to develop the host model. This mismatch can be catastrophic for face anti-spoofing models, as they are particularly sensitive to low-level image features such as task irrelevant noise and artifacts. In our work, we aim to address the above-mentioned problem by building a privileged system that allows for lightweight customization at testing stage by either the host or the client, using only a few labeled samples provided by the client.

Relation to Classical Domain Adaptation.: Classical domain adaptation typically requires explicit access to the source domain data to align the source and target domains. However, in many practical scenarios, access to the source data may be restricted, especially for data involving facial images. Additionally, since the application is related to security, host model parameters cannot be explicitly shared with clients or end-users due to the risks of model theft or white-box adversarial attacks. Therefore, in our setting, both the host model parameters and source training data are not accessible. Instead, only a few data features and prototypes are made available to clients, allowing them to improve the performance for their specific use cases.

2. Interpretation of Geodesic Mix-up

We use optimal transport (OT) to compute intermediate distributions along the geodesic path between the source and target data distributions, which are empirically represented by source prototypes and few-shot client data, respectively. This geodesic path minimizes the cost of transporting probability mass while maintaining a smooth transition between the two distributions. Each intermediate distribution along this path represents a weighted blend of source and target characteristics. By sampling from these intermediate distributions, we generate data that can be interpreted as geodesic mix-up. Geodesic mix-up extends the traditional mix-up concept by shifting from the feature space to the space of probability distributions. More specifically, instead of interpolating between individual data points, it interpolates between entire distributions, offering the following two advantages:

- The **relationships** among real data points can be effectively preserved to the synthetic data points, as geodesic mixup respects the geometric structure of data distributions.
- Interpolating at the distribution level reduces the impact of noise and outliers in individual data points.

Training on these samples allows the classifier to learn how features transition between domains, so that it can adapt to target-specific characteristics while maintaining its knowledge of the source domain.

Solving Equation 5 will get an intermediate distribution along the geodesic path between the source and target distributions. This intermediate distribution represents a weighted blend of source and target characteristics. The parameter $w \in [0, 1]$ in the equation determines the weight of the blend: w close to 0 results in a distribution closer to the source, and w close to 1 results in a distribution closer to the target. Intuitively, Equation 5 aims to find a distribution that transitions smoothly between the source and target by minimizing the transportation cost. If we draw an analogy to Euclidean space, the source, target, and intermediate distributions can be thought of as points, and the intermediate distribution lies on the "straight line" connecting the source and target. However, in Wasserstein space, this "straight line" corresponds to the geodesic path, which captures the optimal transport relationship between the distributions, respecting their geometric structure and preserving relationships among data points.

3. Additional Visualizations

Here, we provide more visualizations. As demonstrated in Fig. 1 (left two), OTA (training-free) effectively calibrate prototypes to adapt to target domain while resisting noisy samples. Geodesic Mixup (right two) generates diverse pseudo distributions which respect geometric information of source and target domains.

4. Scaling Property OTA

Both training-free and lightweight adaptation methods primarily concern few-shot scenarios where the available target domain data are extremely scarce. In application, it is possible that the few-shot number is moderate. To this end, we scale the few-shot number K from 5 to 50 and test OTA accordingly to evaluate OTA's scaling property. As shown

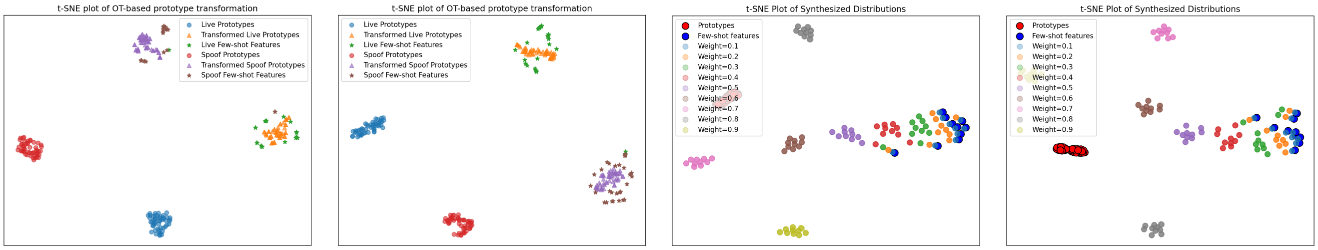

Figure 1. Visualization of OTA in the latent space. Left two plots indicating training-free adaptation. Right two plots resemble the generated synthetic empirical distributions of Geodesic Mixup.

	K=5		**K=10**		**K=20**		**K=50**	
	Avg. HTER↓	Avg. AUC↑	Avg. HTER↓	Avg. AUC↑	Avg. HTER↓	Avg. AUC↑	Avg. HTER↓	Avg. AUC↑
OTA † (training-free)	3.57	99.05	3.21	99.23	3.13	99.27	3.01	99.30
OTA ‡ (lightweight)	**3.28**	**99.29**	**2.91**	**99.39**	**2.82**	**99.48**	**1.97**	**99.65**

Table 1. Results of OTA under different few-shot number K.

Require: Source domains $\{\mathcal{D}_i\}_{i=1}^N$,
Feature extractor $f : \mathcal{X} \to \mathbb{R}^D$,
Target support set $\mathcal{D}_t = \{(\mathbf{x}_{t,j}, y_{t,j})\}_{j=1}^{M_t}$,
Number of centroids K

Train Stage: Learning Prototype-based Framework
Initialize $\mathbf{P} = \{\mathbf{p}^{\text{bona fide}}, \mathbf{p}^{\text{spoof}}\} \in \mathbb{R}^{D \times K \times 2}$
for batch $\mathcal{B}$ in source domains **do**
  $\mathbf{z}_i = f(\mathbf{x}_i)$ for $\mathbf{x}_i \in \mathcal{B}$
  Compute losses: $\mathcal{L}_{\text{proto}}$, $\mathcal{L}_{\text{con}}^{\text{coarse}}$, $\mathcal{L}_{\text{con}}^{\text{fine}}$, $\mathcal{L}_{\text{orth}}$
  Update $\mathbf{P}$ and f using combined loss (Eq. 3)
end for
Ensure: Learned class prototypes $\mathbf{P} = \{\mathbf{p}^{\text{bona fide}}, \mathbf{p}^{\text{spoof}}\}$

Test Stage: Source-free Few-shot Adaptation
Extract features: $\mathbf{Z}_t = \{f(\mathbf{x}_{t,j})\}_{j=1}^{M_t}$

Proposed Method 1: Training-free OT Adaptation
for $c \in \{\text{bona fide}, \text{spoof}\}$ **do**
  Compute $\mathbf{M}_{ij}^c = \|\mathbf{p}_i^c - \mathbf{z}_j^t\|_2^2$; init $\mathbf{a}^c$, $\mathbf{b}^c$
  $\gamma^{c*} = \arg\min_{\gamma \in \Pi(\mathbf{a}^c, \mathbf{b}^c)} \langle \gamma, \mathbf{M}^c \rangle + \lambda \Omega_\alpha(\gamma)$
  Optimal Transform: $\mathbf{p}^{c*} = T_{\gamma^{c*}}(\mathbf{p}^c)$
end for
Ensure: Transformed prototypes $\mathbf{P}^*$ (as classifier)

Proposed Method 2: Lightweight Geodesic mix-up
Initialize a random linear classifier $l : \mathcal{Z} \to \mathbb{R}^2$
while not reach target iterations **do**
  $w \sim \text{Beta}(0.4, 0.4)$
  $\mu_w = \arg\min_\mu [w\mathcal{W}(\mu, \mu_s) + (1-w)\mathcal{W}(\mu, \mu_t)]$
  Optimize classifier l on $\mathbf{Z}_t \cup \mathbf{P} \cup \{\mathbf{u}_i\}_{i=1}^K \overset{\text{i.i.d.}}{\sim} \mu_w$
end while
Ensure: Classifier l

Figure 2. Detailed algorithms for prototype-based backbone training, training-free OT adaptation and light-weight training with Geodesic Mixup.

in Table 1, scaling up few-shot number consistently boost the performance of both training-free and lightweight adaptation OTA methods.

5. Detailed Algorithms

Detailed algorithms can be found in Algorithm 2.