# One-Class Face Anti-spoofing via Spoof Cue Map-Guided Feature Learning

Pei-Kai Huang[1], Cheng-Hsuan Chiang[1], Tzu-Hsien Chen[1], Jun-Xiong Chong[1],
Tyng-Luh Liu[2], Chiou-Ting Hsu[1]
[1]National Tsing Hua University [2]Academia Sinica, Taiwan
{alwayswithme,jimmy890718,gapp111062570,jxchong}@gapp.nthu.edu.tw,
liutyng@iis.sinica.edu.tw, cthsu@cs.nthu.edu.tw

2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) | 979-8-3503-5300-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/CVPR52733.2024.00034

## Abstract

*Many face anti-spoofing (FAS) methods have focused on learning discriminative features from both live and spoof training data to strengthen the security of face recognition systems. However, since not every possible attack type is available in the training stage, these FAS methods usually fail to detect unseen attacks in the inference stage. In comparison, one-class FAS, where training data comprise only live faces, aims to detect whether a test face image belongs to the live class or not. In this paper, we propose a novel One-Class Spoof Cue Map estimation Network (OC-SCMNet) to address the one-class FAS detection problem. Our first goal is to learn to extract latent spoof features from live images so that their estimated Spoof Cue Maps (SCMs) should have zero responses. To avoid trapping to a trivial solution, we devise a novel SCM-guided feature learning by combining many SCMs as pseudo ground-truths to guide a conditional generator to create latent spoof features for spoof data. Our second goal is to simulate the potential out-of-distribution spoof attacks approximately. To this end, we propose using a memory bank to dynamically preserve a set of sufficiently "independent" latent spoof features to encourage the generator to probe the latent spoof feature space. Extensive experiments conducted on eight FAS benchmark datasets demonstrate that the proposed OC-SCMNet not only outperforms previous one-class FAS approaches but also achieves performance comparable to the state-of-the-art two-class FAS methods. The code is available at* https://github.com/Pei-KaiHuang/CVPR24_OC_SCMNet.

## 1. Introduction

Face recognition has been widely adopted in everyday situations to facilitate biometric authentication for unlocking devices, making payments, and accessing sensitive data. To prevent facial spoofing attacks by using photos (*i.e.*, print attacks), videos (*i.e.*, replay attacks), or masks (*i.e.*, 3D mask attacks) of other authorized persons, many face anti-spoofing (FAS) methods [2, 9–15, 17, 19, 24, 25, 34, 37] have been developed to distinguish spoof attacks from live images. Most FAS methods adopt the two-class classification method to learn discriminative feature characteristics from both live and spoof images. For example, the authors in [18] decompose any facial image into a live-like image and spoof noise, which are then used to differentiate between live and spoof images. In [8], the authors propose using the spoof cue map for FAS via estimating an all-zero spoof cue map for live images and nonzero ones for spoof images.

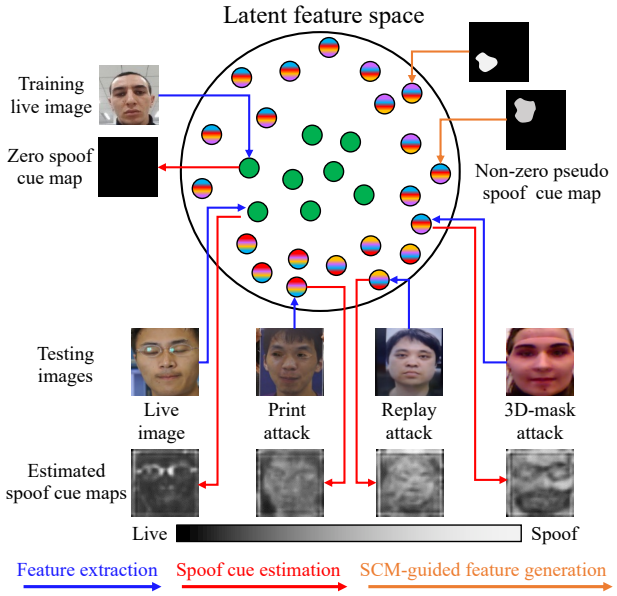



Figure 1. Illustration of the proposed SCM-guided feature learning for one-class face anti-spoofing (FAS). First, under the assumption that live images should yield no spoof cues, we focus on learning to extract latent spoof features so that the estimated Spoof Cue Maps (SCMs) from live images become **0**. Next, to simulate the absent spoof class, we incorporate nonzero pseudo SCMs in the SCM-guided generative network to guide the feature learning.

Compared to two-class methods, one-class FAS methods [2, 15, 20, 24] aim to train the model only from the live class.

2575-7075/24/$31.00 ©2024 IEEE
DOI 10.1109/CVPR52733.2024.00034

Since live faces are collected from real people, different face images in the live class usually exhibit small distribution discrepancies between the training and test domains [17]. Therefore, learning liveness information from the live class alone would be more feasible to identify unseen spoof attacks from the out-of-distribution (OOD) testing domain in the inference stage. However, in the absence of spoof class, the first challenge is how to learn discriminative features from only the live class. Next, unlike other one-class classification problems (such as anomaly detection or novelty detection), one-class FAS deals with highly similar visual characteristics between live and spoof faces and relies heavily on whether the extracted features can characterize the intrinsic difference between the two classes. Finally, due to the constant evolution and variation of new spoof attacks, basic spoof models like those using additive noise from a prior distribution are ineffective in identifying spoof attacks that have not been encountered before.

This study focuses on addressing the above-mentioned challenges in one-class FAS and propose a novel One-Class Spoof Cue Map estimation Network (OC-SCMNet) to effectively detect the out-of-distribution (OOD) occurrences of spoof attacks in inference. Figure 1 illustrates our main idea. First, under the widely acknowledged assumption [8, 18] that live images should contain zero spoof noise or null spoof cues, we propose a novel spoof cue map (SCM) estimation focusing on learning the latent spoof feature representation, which is able to re-produce zero SCMs from live images. Next, to overcome the absence of spoof class, we propose an SCM-guided generative feature learning by combining many nonzero SCMs as pseudo ground-truths of spoof class to guide a conditional generator to generate nontrivial latent spoof features. Finally, to detect OOD occurrences of unseen spoof attacks, we further force the latent feature generation process to continue evolving towards unexplored latent space. In particular, we propose using a fixed-size of sufficiently 'independent' latent features to closely approximate the global latent feature space for probing unseen latent spoof features. We conduct extensive experiments on eight public face anti-spoofing databases to evaluate the effectiveness of the proposed OC-SCMNet. Our experimental results on intra-domain and cross-domain testing demonstrate that the proposed OC-SCMNet not only surpasses existing one-class FAS approaches but also delivers comparable results to state-of-the-art two-class FAS methods.

Our contributions are summarized as follows:

- We introduce a novel one-class face anti-spoof model called OC-SCMNet, focusing on learning discriminative latent spoof features so that their corresponding spoof cue maps (SCMs) can effectively reflect zero and nonzero responses for the live and spoof classes, respectively.
- Under the one-class constraint, we combine nonzero SCMs as pseudo ground-truths for spoof class and propose an SCM-guided generative feature learning to guide the model on generating nontrivial latent spoof features.
- To explore potential spoof attacks, we propose using a fixed-size memory bank to keep a set of representative spoof latent features to approximate the global spoof latent space for probing unseen latent spoof features.
- Our extensive experimental results have shown that OC-SCMNet surpasses previous one-class FAS techniques and attains performances comparable to those of the state-of-the-art two-class FAS methods.

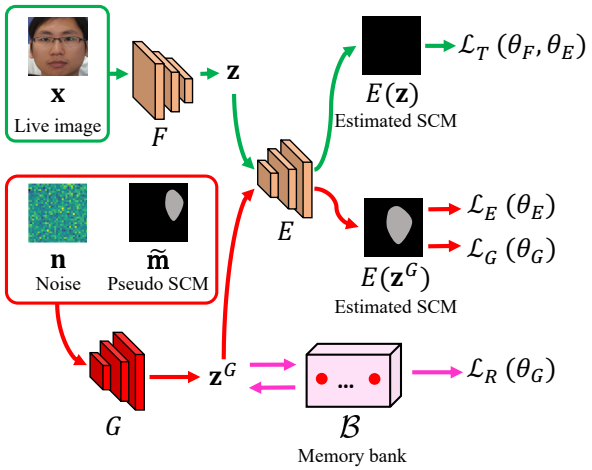



Figure 2. The proposed OC-SCMNet consists of one latent feature extractor $F$, one SCM estimator $E$, and one latent spoof feature generator $G$. We train $F$ and $E$ to produce zero SCMs for the training live images. To avoid trivial solutions for $E$, we use sampled Gaussian noise $\mathbf{n}$ and a pseudo SCM $\widetilde{\mathbf{m}}$ to guide the generator $G$ to learn nontrivial latent features $\mathbf{z}^G$ with corresponding nonzero SCMs. Furthermore, to explore the potential spoof attacks, we use the fixed-size memory bank $\mathcal{B}$ to encourage $G$ to keep generating new latent spoof features towards the unexplored direction.

## 2. Related work

**Two-class face anti-spoofing** Many two-class FAS techniques have been developed to learn discriminative and generalized characteristics in various scenarios, including domain generalization (DG) [9–13, 17, 19, 25–27, 29, 32, 35, 37, 42, 43, 45], domain adaptation (DA) [34], source-free domain adaptation (SFDA) [23], test-time adaptation (TTA) [14], and domain continual learning (DCL) [4]. To yield discriminative features, the authors in [11, 12, 35, 42, 43] propose integrating predefined or learnable descriptors into vanilla convolution to capture gradient-level information to improve the representation capability. Furthermore, the authors in [10, 13, 33, 37] employ disentangled feature learning to distinguish between liveness and domain information to acquire generalized features and domain generalization.

**One-class defect detection/face anti-spoofing** Defect detection with one-class techniques entails training the model exclusively on normal data. These approaches [16, 30] usually analyze distinct characteristics between different regions of a sample to determine whether the sample is normal or abnormal, assuming that abnormal data regions tend to differ from those of normal data. However, because of highly similar visual characteristics between live and spoof faces, the subtle facial differences are no longer discriminative.

In one-class FAS methods [15, 20], the authors propose learning the liveness information through the facial image reconstruction constraint. However, in the absence of spoof faces, the model may simply learn to reconstruct live faces through some general facial features rather than the genuine liveness features. In addition, Gaussian Mixture Models (GMMs) are adopted in [2, 24] to learn the distribution of live images. Specifically, the authors in [24] consider the features of Image Quality Measures introduced in [38] to train the GMMs distribution of live images, and the method in [2] mixes the noise sampled from a Gaussian distribution with live features to create pseudo spoof features for distinguishing between live and spoof images. Alas, take for example that the two approaches [24] and [2] applied to protocol 1 of the dataset **OULU-NPU** [3] achieve ACER [28] values (the lower the better) of 46.95% and 30.242%, respectively. These results are deemed unsatisfactory and less competitive compared to the two-class methods.

# 3. Our method

We tackle the problem of face anti-spoofing (FAS) under the one-class constraint that all training samples in the given dataset are limited to images of live faces. Thus, it is reasonable to expect that the goodness of a learned model for achieving the anti-spoofing task critically depends on how well it can detect the out-of-distribution (OOD) occurrences in inference. To this end, we propose a novel representation learning, guided by the use of spoof cue maps (SCMs), to facilitate the OOD detection for face anti-spoofing.

Our method results in a One-Class Spoof Cue Map estimation Network (OC-SCMNet) to predict whether a face image is `Live` or `Spoof`. Figure 2 shows that OC-SCMNet comprises two key convolutional network modules, denoted as $F$ and $E$, where the former is the feature extractor for the latent feature representation, and the latter is the estimation module for the corresponding spoof cue map. To overcome the difficulty of training a classification model with respect to one-class data, we also include a generative module $G$ to generate latent features with respect to pseudo spoof cue maps. There are totally three sets of parameters, $\boldsymbol{\theta}_F$ of $F$, $\boldsymbol{\theta}_E$ of $E$ and $\boldsymbol{\theta}_G$ of $G$, to be optimized in the course of model training. We next describe the details of our approach.

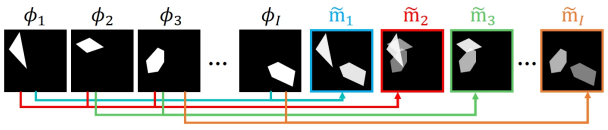



Figure 3. Examples of randomly sampled binary masks $\phi_i$ and the randomly combined pseudo spoof cue maps $\widetilde{\mathbf{m}}_i$.

## 3.1. SCM-guided feature learning

Given a one-class training dataset $D = \{\mathbf{x}\}$, where each $\mathbf{x} \in D$ is an image of live face (or simply referred to as a live image). We denote the proposed FAS model by $T$. As the model $T$ is constructed with a feature extractor $F$ and an SCM estimator $E$, we express the resulting SCM of $\mathbf{x}$ by

$$\mathbf{m} = T(\mathbf{x}) = E(F(\mathbf{x})) = E(\mathbf{z}) \,, \tag{1}$$

where $\mathbf{m}$ is of the same spatial size as $\mathbf{x}$, and $\mathbf{z} = F(\mathbf{x})$ is the latent (unit) feature vector. To link the SCM output, $\mathbf{m}$, with the face antispoofing task, we set the ultimate goal of model training to attain the following two useful properties.

- $T(\mathbf{x}) = \mathbf{0}$, if $\mathbf{x}$ is a live image; $\mathbf{0}$ denotes the null SCM.
- $\|T(\mathbf{x})\|_1 = \|\mathbf{m}\|_1 > \alpha > 0$, if $\mathbf{x}$ is not a live image; $\alpha$ is a scalar/margin to be specified in training. Note that we adopt the entry-wise matrix 1-norm, $\|\mathbf{m}\|_1 = \sum |m_{i,j}|$.

While the above two perspectives of consideration are reasonable, it is indeed not directly applicable under the one-class setting. In particular, learning $T$ with only live images, the optimization could easily lead to a null estimator $E$. On the other hand, without assuming any prior knowledge about spoofing attacks, it is hard to establish a general formulation to generate instances of spoof samples accounting for various scenarios. We instead develop an SCM-guided formulation of generative feature learning to simultaneously resolve the two aforementioned challenging issues.

**Binary masks and SCMs** To avoid trapping in a trivial optimum of a null estimator $E$, we include an SCM-guided generator network $G$ in the OC-SCMNet training stage to produce latent spoof features, denoted as $\mathbf{z}^G$. More importantly, the strategy empowers the resulting SCM estimator $E$ to detect the OOD occurrences for face anti-spoofing.

Recall that we aim to learn the model $T$ that could output an SCM response with $\|\mathbf{m}\|_1 > \alpha$, reflecting a margin-based decision boundary from the null SCM, *i.e.*, $\mathbf{0}$. To generate such spoof cue maps, we randomly sample a collection of binary masks, say, $\{\boldsymbol{\phi}_i\}$ with $\|\boldsymbol{\phi}_i\|_1 > \alpha$ and $\cup\, \boldsymbol{\phi}_i$ covering the whole spatial region. For each training batch, we obtain a set of $N_{\mathcal{M}}$ pseudo spoof cue maps, denoted as $\mathcal{M} = \{\widetilde{\mathbf{m}}\}$ where each $\widetilde{\mathbf{m}}$ is constructed as follows. We first decide on the nonnegative fusion coefficient vector $\mathbf{c} = (c_i)$ by randomly assigning the number of nonzero elements and their respective weights (sum to 1) and then apply the convex combination to form a pseudo spoof cue map by

$$\widetilde{\mathbf{m}} = \sum\nolimits_i c_i \cdot (\mathbf{1} \odot \boldsymbol{\phi}_i) \tag{2}$$

where $\mathbf{1}$ is the all-ones matrix and $\odot$ symbolizes the element-wise multiplication. Figure 3 shows examples of using convex polygons to generate binary masks $\{\phi_i\}$ and the resulting pseudo SCMs $\{\widetilde{\mathbf{m}}_i\}$.

**Generative feature learning** We are now in a position to describe the proposed SCM-guided generative feature learning with the conditional generator $G$. As shown in Figure 2, the process of feature generation by $G$ is conditioned on a given pseudo SCM $\widetilde{\mathbf{m}} \in \mathcal{M}$ and triggered by the noise $\mathbf{n} \sim \mathcal{N}(0, I)$, sampled from a unit Gaussian prior. We express the generative process of latent spoof features by

$$\mathbf{z}^G \leftarrow G(\mathbf{n}|\widetilde{\mathbf{m}}; \boldsymbol{\theta}_G) \;\text{ and }\; \|\mathbf{z}^G\| = 1. \tag{3}$$

With (3), the SCM-guided generative process adds the other aspect of input, *i.e.*, $\mathbf{z}^G$, to enhance the training of the FAS model $T = F \circ E$. Observe from Figure 2 that the SCM-guided feature learning pipeline goes through the generator $G$ and the estimator $E$ and thus involves the parameters $\boldsymbol{\theta}_G$ and $\boldsymbol{\theta}_E$. We adopt the strategy of alternating optimization to learn the parameters $\boldsymbol{\theta}_G$ of $G$ by fixing the parameters $\boldsymbol{\theta}_E$ of $E$. Specifically, we consider minimizing the following loss:

$$\begin{aligned}\mathcal{L}_G(\boldsymbol{\theta}_G) &= \sum\nolimits_{\widetilde{\mathbf{m}}\in\mathcal{M}} \|E(G(\mathbf{n}|\widetilde{\mathbf{m}}; \boldsymbol{\theta}_G)) - \widetilde{\mathbf{m}}\|_2^2 \\ &= \sum\nolimits_{\widetilde{\mathbf{m}}\in\mathcal{M}} \|E(\mathbf{z}^G) - \widetilde{\mathbf{m}}\|_2^2.\end{aligned} \tag{4}$$

So far, we have established an SCM-guided feature learning formulation to generate latent spoof features $\{\mathbf{z}^G\}$. However, such feature generation is driven solely by $\mathcal{M}$ and may not fully explore the whole latent feature space. To better take account of this issue, our formulation keeps a fixed-size memory bank $\mathcal{B}$ that evolves a corresponding latent subspace to encourage globally probing of the feature generation process. At each batch-wise training, a newly generated latent feature $\mathbf{z}^G$ by (3) will be added to the memory bank $\mathcal{B}$, if it satisfies the following criterion:

$$\frac{1}{N_{\mathcal{B}}} \sum\nolimits_{j=1}^{N_{\mathcal{B}}} |\cos(\mathbf{z}^G, \mathbf{z}_j^G)| < \delta, \tag{5}$$

where $\{\mathbf{z}_j^G\}_{j=1}^{N_{\mathcal{B}}}$ are the latent spoof feature vectors currently stored in $\mathcal{B}$, and $\delta$ is a small-value threshold. The inclusion criterion in (5) will memorize a newly generated $\mathbf{z}^G$ if it is sufficiently "independent" to all current entries in $\mathcal{B}$. When $\mathcal{B}$ is fully stored, we adopt the First In First Out (FIFO) scheme to update the memory bank. Since the members in $\mathcal{B}$ define a latent subspace as $\mathrm{Span}(\mathbf{z}_1^G, \ldots, \mathbf{z}_{N_{\mathcal{B}}}^G)$, the event of updating $\mathcal{B}$ implicitly evolves its probing latent subspace.

Thus, to effectively explore the latent space of spoof features in learning the generator $G$, we introduce an additional regularization loss, namely,

$$\mathcal{L}_R(\boldsymbol{\theta}_G) = \sum\nolimits_{\widetilde{\mathbf{m}}\in\mathcal{M}} \sum\nolimits_{j=1}^{N_{\mathcal{B}}} |\cos(\mathbf{z}^G, \mathbf{z}_j^G)|, \tag{6}$$

where $\mathbf{z}^G = G(\mathbf{n}|\widetilde{\mathbf{m}}; \boldsymbol{\theta}_G)$. With the effect of $\mathcal{L}_R$, the model training would drive $G$ to generate latent features not close to the latent subspace spanned by $\mathcal{B}$ and consequently to more effectively probe the latent feature space. Finally, we write out the complete optimization problem for learning the parameters of $G$ as

$$\boldsymbol{\theta}_G^* = \arg\min\nolimits_{\boldsymbol{\theta}_G} \mathcal{L}_G(\boldsymbol{\theta}_G) + \lambda\, \mathcal{L}_R(\boldsymbol{\theta}_G). \tag{7}$$

where $\lambda$ is a parameter to adjust the regularization effect.

### 3.2. Feature-enhanced SCM estimation

Having described the SCM-guided formulation for the generator $G$ of the proposed OC-SCMNet, it remains to show how the remaining parameters are to be optimized to yield the resulting FAS model $T = E \circ F$.

As stated previously, we have set our goal of model training to respect the property $T(\mathbf{x}) = \mathbf{0}$ for any $\mathbf{x} \in D$. Thus, it is reasonable to treat the null SCM, $\mathbf{0}$, as the ground truth of $T(\mathbf{x})$ and consider the following loss function:

$$\begin{aligned}\mathcal{L}_T(\boldsymbol{\theta}_F, \boldsymbol{\theta}_E) &= \sum\nolimits_{\mathbf{x}\in D} \|E(F(\mathbf{x}; \boldsymbol{\theta}_F); \boldsymbol{\theta}_E) - \mathbf{0}\|_2^2 \\ &= \sum\nolimits_{\mathbf{x}\in D} \|E(\mathbf{z}; \boldsymbol{\theta}_E)\|_2^2,\end{aligned} \tag{8}$$

Now, with the latent spoof features $\{\mathbf{z}^G\}$ generated from $G$, the concern of minimizing the above loss easily leading to a null estimator $E$ is no longer an issue. More specifically, the estimator $E$ is also required to predict, for each $\mathbf{z}^G$, the corresponding SCM, $\widetilde{\mathbf{m}}$, as in (3). Thus, we have

$$\mathcal{L}_E(\boldsymbol{\theta}_E) = \sum\nolimits_{\widetilde{\mathbf{m}}\in\mathcal{M}} \|E(\mathbf{z}^G; \boldsymbol{\theta}_E) - \widetilde{\mathbf{m}}\|_2^2, \tag{9}$$

where $\mathcal{M}$ is the set of generated SCMs in each batch-wise training. That is, the number of $\mathbf{z}^G$ considered in (9) equals $N_{\mathcal{M}}$, the size of $\mathcal{M}$. Analogous to (7), by freezing the parameters $\boldsymbol{\theta}_G$ of the generator $G$, the model training of $T = F \circ E$ can be achieved with

$$\boldsymbol{\theta}_F^*, \boldsymbol{\theta}_E^* = \arg\min_{\boldsymbol{\theta}_F, \boldsymbol{\theta}_E} \mathcal{L}_T(\boldsymbol{\theta}_F, \boldsymbol{\theta}_E) + \mathcal{L}_E(\boldsymbol{\theta}_E). \tag{10}$$

### 3.3. Training and testing

**Training** We iteratively optimize the two coupled optimization problems of (7) and (10) in an alternate manner. In each iteration, we first update $F$ and $E$ by minimizing $\mathcal{L}_T$ in (8). Next, we fix $F$ and $E$ and train $G$ by minimizing $\mathcal{L}_G$ and $\mathcal{L}_R$, with a regularization parameter $\lambda$, in (7). Finally, we update $E$ by minimizing $\mathcal{L}_E$ in (9) in terms of $\mathbf{z}^G$.

**Testing** For inference, we apply the FAS model $T = E \circ F$ to a test image $\mathbf{x}$ and calculate its response score by

$$s(\mathbf{x}) = \frac{\sum_{d=1}^{D} \sum_{h=1}^{H} \sum_{w=1}^{W} |T(\mathbf{x})|}{D \cdot H \cdot W}, \tag{11}$$

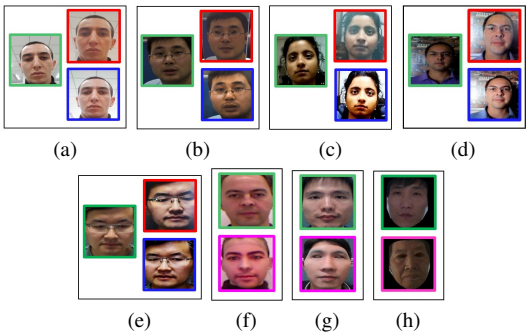



Figure 4. Examples of live faces (boxes in green), `print` attacks (red), `replay` attacks (blue), and `3D mask` attacks (magenta).

where $D, H$, and $W$ refer to the channel number, height, and width of the estimated spoof cue map, respectively. We follow [37, 42] to adopt the Youden Index Calculation [40] for obtaining the threshold of binary classification.

# 4. Experiments

## 4.1. Experiment settings

**Datasets** We conduct extensive experiments on the following eight face anti-spoofing databases: (a) **OULU-NPU** [3] (denoted by **O**), (b) **CASIA-MFSD** [44] (denoted by **C**), (c) **MSU-MFSD** [38] (denoted by **M**), (d) **Idiap Replay-Attack** [6] (denoted by **I**), (e) **SiW** [22] (denoted by **S**), (f) **3DMAD** (denoted by **D**)[7], (g) **HKBU-MARs** (denoted by **H**) [21], and (h) **CASIA-SURF** [41] (denoted by **U**). Examples from each dataset, (a)–(h), are shown in Figure 4.

**Evaluation metrics** For a fair comparison with previous FAS methods, we report the results using the same evaluation metrics, including APCER (%) $\downarrow$ [28], BPCER (%) $\downarrow$ [28], ACER (%) $\downarrow$ [28], HTER (%) $\downarrow$ [1], and AUC (%) $\uparrow$.

**Implementation details** To train OC-SCMNet, we set a constant learning rate of 5e$-$4 with Adam optimizer up to 20 epochs. We set the feature selection threshold $\delta = 0.2$ and the memory bank size $N_{\mathcal{B}} = 16$ for all experiments.

## 4.2. Ablation study

**On different loss terms** In Table 1, we compare using different loss terms to train the proposed OC-SCMNet on the cross-domain protocols **C** $\rightarrow$ **I** and **I** $\rightarrow$ **C**. First, we use the sampled Gaussian noise to replace $\mathbf{z}^G$ in Figure 2 and use only $\mathcal{L}_T$ and $\mathcal{L}_E$ to train the model $T = F \circ E$ as the baseline. Although the baseline model learns to map live images to a zero SCM, its efficacy is constrained since using Gaussian noise alone is not enough to mimic the spoof latent features. Next, we include $\mathcal{L}_G$ to train $G$ to generate latent spoof features $\{\mathbf{z}^G\}$ and use $\{\mathbf{z}^G\}$ to enhance the training of $T$. Because $T$ already learns to estimate zero SCMs from live images, the generator $G$ guided by $\widetilde{\mathbf{m}}$ is able to generate

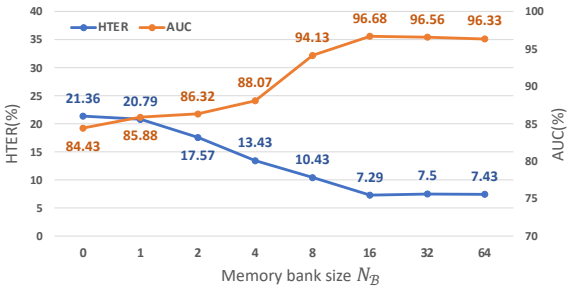



Figure 5. Ablation study on sizes $N_{\mathcal{B}}$ of the memory bank $\mathcal{B}$ under the protocol **C** $\rightarrow$ **I**.

Table 1. Ablation study on the cross-domain protocols **C** $\rightarrow$ **I** and **I** $\rightarrow$ **C**, under different loss combinations.

| Loss Terms | | | C $\rightarrow$ I | | I $\rightarrow$ C | |
|---|---|---|---|---|---|---|
| $\mathcal{L}_T + \mathcal{L}_E$ | $\mathcal{L}_G$ | $\mathcal{L}_R$ | HTER | AUC | HTER | AUC |
| ✓ | | | 30.79 | 63.35 | 39.11 | 61.21 |
| ✓ | ✓ | | 21.36 | 84.43 | 28.78 | 74.25 |
| ✓ | ✓ | ✓ | **7.29** | **96.68** | **17.44** | **82.88** |

latent spoof features capable of producing nonzero SCMs to improve the training of $T$. The performance improvement of $\mathcal{L}_T + \mathcal{L}_E + \mathcal{L}_G$ validates the effectiveness of the proposed generative feature learning. Finally, when further including $\mathcal{L}_R$, we achieve the best performance by continually probing the latent feature space to counter unseen spoof attacks.

**On different sizes of memory bank** In Figure 5, we compare using different memory bank sizes $N_{\mathcal{B}}$ ($N_{\mathcal{B}}$ = 1, 2, 4, 8, 16, 32 and 64) in $\mathcal{B}$ on the protocol **C** $\rightarrow$ **I** and show the results in terms of HTER and AUC. In this experiment, we set $\delta = 0.2$. First, we find that the performance improves steadily as $N_{\mathcal{B}}$ increases from 1 to 16 and achieves the best HTER at $N_{\mathcal{B}} = 16$ . These results demonstrate that larger memory banks indeed better enable the generator to generate additional unseen latent features for extending the dimension of the latent feature subspace. Next, we observe that the performance does not improve as $N_{\mathcal{B}}$ increases from 16 to 64. This performance plateau might indicate that the memory bank has reached a state of overcompleteness. Hence, from this ablation study, we empirically set $N_{\mathcal{B}} = 16$ in OC-SCMNet for subsequent experiments.

**On different selection thresholds** In Figure 6, we set the memory bank size $N_{\mathcal{B}} = 16$ and compare using different selection thresholds $\delta \in [0.1, 1]$ on the protocol **C** $\rightarrow$ **I**. The results show that using $\delta$ ranging from 0.1 to 0.9 all yields better performance than using $\delta = 1$, which means no feature selection. This ablation study verifies that the proposed feature selection mechanism indeed facilitates the generator to generate more unseen latent features to explore the latent space and further enhances the training of $E$. Figure 6 also shows that the performances significantly decline as $\delta$ increases from 0.5 to 1. This outcome indicates that using

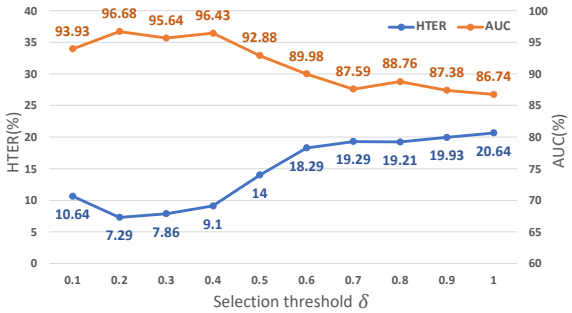



Figure 6. Ablation study on the protocol **C** → **I**, using different thresholds $\delta$ in Equation (5) to select $\mathbf{z}^G$.

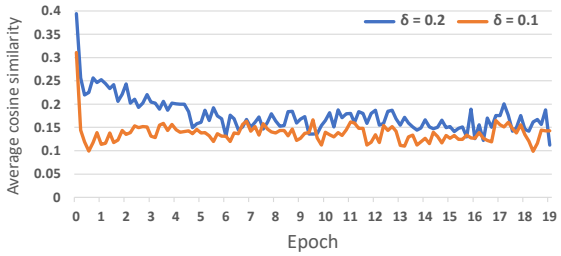



Figure 7. Trend of cosine similarity from the generated latent spoof features under the protocol **C** → **I**.

smaller $\delta$ (*i.e.*, selecting highly dissimilar latent features) can better facilitate $G$ to expand the feature space for enhancing the SCM estimation. Moreover, among all the other settings of $\delta < 0.5$, using $\delta = 0.2$ achieves the best performance with the lowest ACER and the highest AUC. We suspect that $G$ can no longer generate a sufficient number of latent spoof features with similarity lower than 0.1 and thus hinders the improvement of SCM estimation. As shown in Figure 7, although the cosine similarities of the generated features decrease gradually as the training epochs progress, the cosine similarities of the generated features are scarcely lower than 0.1. Therefore, from this ablation study, we empirically set $\delta = 0.2$ in OC-SCMNet for all the following experiments.

### 4.3. Intra-domain and cross-domain testing

**Intra-domain testing** Table 2 shows the intra-domain testing results on **OULU-NPU** [3]. In this experiment, the proposed OC-SCMNet significantly outperforms all the one-class FAS methods [2, 15, 20, 24] with averagely improved 31.02% in ACER. Note that, under intra-domain testing, because spoof attacks exhibit highly similar characteristics between training and test data, two-class FAS methods can better learn live/spoof distinguishing features from two-class data and substantially outperform all the one-class methods.

**Cross-domain testing** We perform cross-domain experiments to evaluate the generalizability of the OC-SCMNet model and report the results in Tables 3, 4, and 5. To begin

Table 2. Intra-domain testing on **OULU-NPU**.

| Training Data | Method | P. | APCER | BPCER | ACER |
|---|---|---|---|---|---|
| Live + Spoof | CDCN [42] (*CVPR 20*) | 1 | 0.4 | 1.7 | 1.0 |
| | CIFL [5] (*TIFS 21*) | | 3.8 | 2.9 | 3.4 |
| | PatchNet [32] (*CVPR 22*) | | 0.0 | 0.0 | **0.0** |
| | LDCN [11] (*BMVC 22*) | | 0.0 | 0.0 | **0.0** |
| | TTN-S [36] (*TIFS 22*) | | 0.4 | 0.0 | 0.2 |
| | LDCformer [12] (*ICIP 23*) | | 0.0 | 0.0 | **0.0** |
| Live | IQM-GMM [24] (*ICB 18*) | | 75.35 | 18.56 | 46.95 |
| | Baweja *et al.* [2] (*IJCB 20*) | | 38.63 | 21.85 | 30.24 |
| | Lim *et al.* [20] (*Access 20*) | | 43.54 | 36.5 | 40.02 |
| | AAE [15] (*CCBR 21*) | | 47.13 | 26.67 | 36.9 |
| | OC-SCMNet (Ours) | | 20.83 | 26.15 | **23.49** |
| Live + 1-Shot Spoof | OC-SCMNet (Ours) | | 20.20 | 10.83 | 15.52 |
| Live + Spoof | CDCN [42] (*CVPR 20*) | 2 | 1.5 | 1.4 | 1.5 |
| | CIFL [5] (*TIFS 21*) | | 3.6 | 1.2 | 2.4 |
| | PatchNet [32] (*CVPR 22*) | | 0.8 | 1.0 | 0.9 |
| | LDCN [11] (*BMVC 22*) | | 0.8 | 1.0 | 0.9 |
| | TTN-S [36] (*TIFS 22*) | | 0.4 | 0.8 | **0.6** |
| | LDCformer [12] (*ICIP 23*) | | 0.0 | 0.0 | **0.0** |
| Live | IQM-GMM [24] (*ICB 18*) | | 41.56 | 27.78 | 34.67 |
| | Baweja *et al.* [2] (*IJCB 20*) | | 51.81 | 19.83 | 35.82 |
| | Lim *et al.* [20] (*Access 20*) | | 72.19 | 18.5 | 45.35 |
| | AAE [15] (*CCBR 21*) | | 37.28 | 39.0 | 38.14 |
| | OC-SCMNet (Ours) | | 22.05 | 28.81 | **25.43** |
| Live + 1-Shot Spoof | OC-SCMNet (Ours) | | 24.13 | 21.44 | 22.79 |
| Live + Spoof | CDCN [42] (*CVPR 20*) | 3 | 2.4±1.3 | 2.2±2.0 | 2.3±1.4 |
| | CIFL [5] (*TIFS 21*) | | 3.8±1.3 | 1.1±1.1 | 2.5±0.8 |
| | PatchNet [32] (*CVPR 22*) | | 1.8±1.47 | 0.56±1.24 | 1.18±1.26 |
| | LDCN [11] (*BMVC 22*) | | 4.55±4.55 | 0.58±0.91 | 2.57±2.67 |
| | TTN-S [36] (*TIFS 22*) | | 1.0±1.1 | 0.8±1.3 | **0.9±0.7** |
| | LDCformer [12] (*ICIP 23*) | | 2.35±2.05 | 0.28±0.68 | 1.31±1.03 |
| Live | IQM-GMM [24] (*ICB 18*) | | 57.17±16.79 | 16.5±6.95 | 36.83±5.35 |
| | Baweja *et al.* [2] (*IJCB 20*) | | 45.39±12.82 | 18.28±16.21 | 31.83±6.99 |
| | Lim *et al.* [20] (*Access 20*) | | 38.51±13.08 | 39.52±11.13 | 39.02±2.16 |
| | AAE [15] (*CCBR 21*) | | 26.62±13.67 | 52.93±16.09 | 39.77±3.74 |
| | OC-SCMNet (Ours) | | 27.10±12.57 | 20.55±11.12 | **23.83±3.14** |
| Live + 1-Shot Spoof | OC-SCMNet (Ours) | | 23.02±12.16 | 11.88±10.8 | 17.45±3.07 |
| Live + Spoof | CDCN [42] (*CVPR 20*) | 4 | 4.6±4.6 | 9.2±8.0 | 6.9±2.9 |
| | CIFL [5] (*TIFS 21*) | | 5.9±3.3 | 6.3±4.7 | 6.1±4.1 |
| | PatchNet [32] (*CVPR 22*) | | 2.5±3.81 | 3.33±3.73 | 2.90±3.00 |
| | LDCN [11] (*BMVC 22*) | | 4.50±1.48 | 3.17±3.49 | 3.83±2.12 |
| | TTN-S [36] (*TIFS 22*) | | 3.3±2.8 | 2.5±2.0 | **2.9±1.4** |
| | LDCformer [12] (*ICIP 23*) | | 1.08±1.28 | 1.17±1.94 | 1.13±1.02 |
| Live | IQM-GMM [24] (*ICB 18*) | | 53.42±14.08 | 16.67±8.38 | 35.04±3.95 |
| | Baweja *et al.* [2] (*IJCB 20*) | | 60.25±16.49 | 10.67±10.37 | 35.46±5.43 |
| | Lim *et al.* [20] (*Access 20*) | | 36.91±10.24 | 20.5±8.01 | 28.07±5.32 |
| | AAE [15] (*CCBR 21*) | | 26.33±18.5 | 40.17±29.04 | 33.12±8.9 |
| | OC-SCMNet (Ours) | | 16.41±14.00 | 11.66±9.42 | **14.04±4.90** |
| Live + 1-Shot Spoof | OC-SCMNet (Ours) | | 4.91±7.45 | 10.0±5.47 | 7.45±3.58 |

with, we follow [25] to conduct cross-domain evaluations on the protocols **[M, I]** → **C** and **[M, I]** → **O** for countering `print` and `replay` attacks. As mentioned in [25], because the datasets **M** and **I** exhibit considerable domain variations, we thus train on these two collections and test on the remaining ones, *i.e.*, **C** and **O**. The results in Table 3 show that OC-SCMNet significantly outperforms all one-class FAS methods, achieving a 22.19% reduction in HTER and an 8.63% increase in AUC. Moreover, we observe that OC-SCMNet attains comparable outcomes with the two-class FAS methods. Our results indicate that due to the cross-domain shift in spoofing characteristics between training and testing data, two-class FAS methods tend to overfit the training data and exhibit significantly reduced performance when applied to unseen domains. In comparison, the proposed OC-SCMNet, without assuming any prior knowledge about the spoof class, demonstrates improved domain generalization ability in detecting unseen spoof attacks for the cross-domain scenario.

Table 3. Cross-domain testing on **[M, I] → C** and **[M, I] → O**.

| Training Data | Method | [M,I] → C | | [M,I] → O | |
|---|---|---|---|---|---|
| | | HTER | AUC | HTER | AUC |
| Live + Spoof | MADDG [25] (*CVPR 19*) | 41.02 | 64.33 | 39.35 | 65.10 |
| | SSDG-M [17] (*CVPR 20*) | 31.89 | 71.29 | 36.01 | 66.88 |
| | SDA [34] (*AAAI 21*) | 32.17 | 72.79 | 28.90 | 73.33 |
| | SSAN-M [37] (*CVPR 22*) | 30.00 | 76.20 | 29.44 | 76.62 |
| | LDCN [11] (*BMVC 22*) | 22.22 | 82.87 | 21.54 | 86.06 |
| | DiVT-M [19] (*WACV 23*) | **20.11** | **86.71** | 23.61 | 85.73 |
| | DFANet [13] (*ICME 23*) | 20.67 | 84.87 | 18.61 | 89.52 |
| | IADG [45] (*CVPR 23*) | 24.07 | 85.13 | **18.47** | **90.49** |
| Live | IQM-GMM [24] (*ICB 18*) | 45.81 | 39.74 | 35.0 | 37.01 |
| | Baweja *et al.* [2] (*IJCB 20*) | 27.33 | 78.50 | 32.01 | 72.19 |
| | Lim *et al.* [20] (*Access 20*) | 43.56 | 53.6 | 39.19 | 64.11 |
| | AAE [15] (*CCBR 21*) | 46.67 | 47.28 | 48.52 | 47.99 |
| | OC-SCMNet (Ours) | **21.67** | **85.30** | **22.03** | **84.28** |
| Live + 1-Shot Spoof | OC-SCMNet (Ours) | 7.56 | 97.29 | 9.86 | 94.09 |

Table 4. Cross-domain testing on **C → I** and **I → C**.

| Training Data | Method | C → I | I → C |
|---|---|---|---|
| | | HTER | HTER |
| Live + Spoof | Auxiliary[22] (CVPR 18) | 27.6 | 28.4 |
| | STASN [39] (CVPR 19) | 31.5 | 30.9 |
| | CDCN [42] (CVPR 20) | 15.5 | 32.6 |
| | CIFL [5] (TIFS 21) | 17.6 | - |
| | AENet [9] (ACPR 21) | 24.7 | 30.9 |
| | PatchNet [32] (*CVPR 22*) | **9.9** | **26.2** |
| Live | IQM-GMM [24] (*ICB 18*) | 37.77 | 48.44 |
| | Baweja *et al.* [2] (*IJCB 20*) | 46.29 | 29.44 |
| | Lim *et al.* [20] (*Access 20*) | 37.36 | 39.78 |
| | AAE [15] (*CCBR 21*) | 20.0 | 26.9 |
| | OC-SCMNet (Ours) | **7.29** | **17.44** |
| Live + 1-Shot Spoof | OC-SCMNet (Ours) | 2.14 | 1.11 |

Table 5. Cross-domain testing on **[O, S] → [D, H, U]**.

| Training Data | Method | [O,S] → D | | [O,S] → H | | [O,S] → U | |
|---|---|---|---|---|---|---|---|
| | | HTER | AUC | HTER | AUC | HTER | AUC |
| Live + Spoof | Auxiliary [22] (*CVPR 18*) | 0.29 | 99.04 | 14.64 | 88.32 | 37.28 | 53.14 |
| | NAS [41] (*TPAMI 20*) | **0.22** | 99.31 | 15.13 | 88.91 | 37.68 | **72.83** |
| | LDCN [11] (*BMVC 22*) | 1.49 | **99.91** | **8.75** | **95.60** | **33.54** | 60.44 |
| Live | IQM-GMM [24] (*ICB 18*) | 43.83 | 43.43 | 19.14 | 80.53 | 38.18 | 66.18 |
| | Baweja *et al.* [2] (*IJCB 20*) | 37.86 | 45.8 | 35.65 | 68.90 | 41.74 | 49.85 |
| | Lim *et al.* [20] (*Access 20*) | 27.69 | 75.47 | 35.19 | 62.98 | 37.34 | 64.24 |
| | AAE [15] (*CCBR 21*) | 22.48 | 78.62 | 31.22 | 73.77 | 45.24 | 53.48 |
| | OC-SCMNet (Ours) | **1.47** | **99.87** | **7.08** | **86.84** | **10.61** | **90.75** |
| Live + 1-Shot Spoof | OC-SCMNet (Ours) | 1.19 | 99.77 | 0.0 | 100.0 | 8.43 | 92.66 |

In Table 4, we follow [22] to perform the single cross-domain testing and assess the outcomes using the datasets **C** and **I**. The dataset **C** comprises both high- and low-resolution images, while **I** consists of low-resolution ones. Hence, the protocol **I** → **C** poses a greater challenge than **C** → **I** because a model trained solely on low-resolution data often struggles with high-resolution samples. Even though the protocol **I** → **C** remains a significant challenge in two-class FAS methods, the results in Table 4 show that OC-SCMNet considerably improves detection performance and outperforms the other one-class/two-class FAS methods under comparison.

Furthermore, in Table 5, we adopt the settings of [41] to conduct cross-domain testing on `3D-mask` attacks by using the live images from **O** and **S** for training and then testing on **D**, **H**, and **U**. Given that the training data for two-class FAS methods only include live images and `print` and `replay` attacks, these methods exhibit limited ability to adapt to new types of attacks, such as `3D-mask` attacks, on the protocol **[O, S]** → **U**. In contrast, OC-SCMNet surpasses other one-class/two-class FAS methods in terms of performance on the **[O, S]** → **U** and **[O, S]** → **H** protocols, and delivers competitive outcomes on the **[O, S]** → **D** protocol.

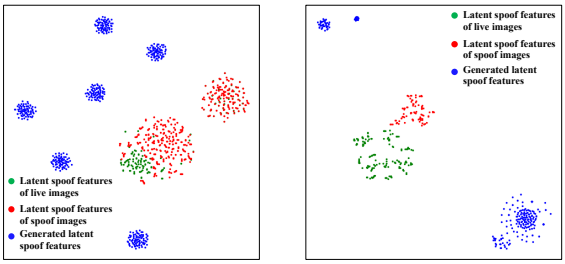
Table 6. New protocols by leave-one-attack-out strategy.

| P. | Unseen attack type | Training subset | Testing subset |
|---|---|---|---|
| 1 | 3D mask | OM(replay + print + live) | DHU (3D mask + live) |
| 2 | | OMCI (replay + print + live) | |
| 3 | print | OM (replay) D (3D mask + live) | OMCI (print + live) |
| 4 | | OMCI (replay) DHU (3D mask + live) | |
| 5 | replay | OM (print) D (3D mask + live) | OMCI (replay+ live) |
| 6 | | OMCI (print) DHU (3D mask + live) | |

Table 7. Experimental comparisons on new protocols.

| Training Data | Method | P. 1 | | P. 2 | | P. 3 | |
|---|---|---|---|---|---|---|---|
| | | HTER | AUC | HTER | AUC | HTER | AUC |
| Live+ Spoof | IADG [45] (*CVPR 23*) | 32.89 | 72.15 | 36.50 | 69.49 | 43.98 | 56.47 |
| | SAFAS [29] (*CVPR 23*) | 38.22 | 63.75 | 34.48 | 65.33 | **30.85** | **75.00** |
| Live | IQM-GMM [24] (*ICB 18*) | 43.58 | 46.99 | 43.82 | 47.18 | 40.25 | 62.02 |
| | Baweja *et al.* [2] (*IJCB 20*) | 39.35 | 61.86 | 42.19 | 57.47 | 41.59 | 61.56 |
| | Lim *et al.* [20] (*Access 20*) | 41.74 | 56.43 | 41.64 | 55.11 | 46.17 | 53.45 |
| | AAE [15] (*CCBR 21*) | 42.85 | 55.97 | 41.07 | 55.35 | 48.50 | 40.94 |
| | OC-SCMNet (Ours) | **24.14** | **74.81** | **20.85** | **85.40** | 37.44 | 63.23 |
| **Training Data** | **Method** | P. 4 | | P. 5 | | P. 6 | |
| | | HTER | AUC | HTER | AUC | HTER | AUC |
| Live+ Spoof | IADG [45] (*CVPR 23*) | 38.56 | 62.14 | 43.85 | 55.75 | 40.04 | 64.13 |
| | SAFAS [29] (*CVPR 23*) | 40.09 | 63.16 | 39.12 | **64.99** | 38.45 | 66.69 |
| Live | IQM-GMM [24] (*ICB 18*) | 47.56 | 41.68 | 37.61 | 64.66 | 48.78 | 41.85 |
| | Baweja *et al.* [2] (*IJCB 20*) | 40.41 | 63.83 | 48.06 | 42.45 | 46.87 | 41.26 |
| | Lim *et al.* [20] (*Access 20*) | 48.29 | 50.30 | 41.32 | 59.08 | 46.45 | 53.71 |
| | AAE [15] (*CCBR 21*) | 42.69 | 57.21 | 46.70 | 53.94 | 37.60 | 64.68 |
| | OC-SCMNet (Ours) | **28.99** | **72.21** | **36.41** | 63.56 | **29.61** | **74.99** |

Finally, we include additional results to highlight the effectiveness of our method against one specific unseen attack type and investigate the impact of varying the number of training datasets. We consider the seven face anti-spoofing datasets **O**, **M**, **C**, **I**, **D**, **H**, and **U**, where the former four (**OMCI**) include `print` and `replay` attacks and the latter three (**DHU**) comprise only `3D mask` attack. As listed in Table 6, we adopt the leave-one-attack-out strategy to form six protocols: P1-P2, P3-P4, P5-P6, each pair of which respectively adopts `3D mask`, `print`, and `replay` as the unseen attack type for two different combinations of training sets. Table 7 shows that OC-SCMNet significantly outperforms the one-class FAS methods [2, 15, 20, 24] and achieves comparable performances to the two-class FAS methods [29, 45] in detecting new attack types. While these outcomes are consistent with the cross-domain testing results in Table 3, they further indicate that two-class FAS methods may overfit the characteristics of seen attack types and exhibit degraded performance when facing unseen attacks. Without assuming a specific form of spoof attack, our method learns the one-class representation of live images by discriminating their null spoof cue maps from those by the generated latent spoof features, which turns out to be general




(a) (b)

Figure 8. $t$-SNE visualizations on the protocols (a) **C** $\rightarrow$ **I**, where **C** and **I** include both `print` and `replay` attacks, and (b) **[O, S]** $\rightarrow$ **D**, where **O**, **S** and **D** include `3D mask` attacks only.

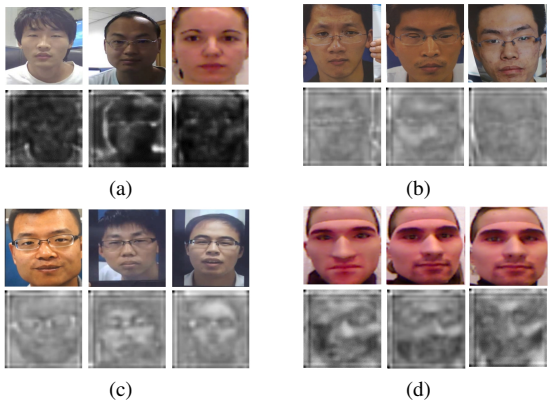
in handling unseen attack types. By comparing the results of experimenting with P1-P2, P3-P4, and P5-P6, it can be seen that the proposed OC-SCMNet model indeed benefits from using more training datasets.

**One-shot testing** The proposed OC-SCMNet is also effective in dealing with the one-shot scenario. Once a single spoof image is available in the training stage, we propose using this one-shot image to guide the generator $G$ to generate latent spoof features as close to the features extracted from this spoof image as possible. To conduct this experiment, we randomly select one spoof image from the training dataset and additionally include a loss term to maximize the similarity between the generated latent spoof features and the extracted latent spoof feature of this spoof image. The experimental results of one-shot testing are shown in Tables 2, 3, 4, and 5. These results show that OC-SCMNet indeed benefits from the generated spoof-like features and achieves improved performance over the one-class setting.

### 4.4. Visualization

**$t$-SNE visualization** In Figure 8, we use $t$-SNE [31] to visualize the latent spoof features extracted from (a) `print` and `replay` attacks, and (b) `3D mask` attacks, on the protocols **C** $\rightarrow$ **I** and **[O, S]** $\rightarrow$ **D**. The visualization results in Figure 8 show that the generated latent spoof features (marked by blue dots) do not simply cluster around those of the live images (marked by green dots) but also largely scatter across the whole latent feature space. Because we use nonzero pseudo SCMs to guide G in generating latent spoof features that do not belong to live images, the generated latent spoof features successfully extend outward from the live training domain and strongly support the model $T$ to generalize to unseen attacks.

**Spoof Cue Map visualization** In Figure 9, we use different examples to demonstrate the effectiveness of the proposed SCM estimation under different types of images: (a) live images, (b) `print` attacks, (c) `replay` attacks, and (d) `3D-mask` attacks. Note that, a lower intensity in the estimated SCM implies a higher probability that the image is a live image. The visualization results in Figure 9 (a) clearly show that the estimated SCMs of live images exhibit almost no spoof cues in contrast to the estimated SCMs from various attacks, as shown in (b) - (d). Next, we see that the estimated SCMs of different attacks exhibit distinct characteristics. In particular, all the SCMs estimated from (b) `print` attacks, (c) `replay` attacks, and (d) `3D-mask` attacks have higher but diversely different responses. In addition, in the case of (c) `replay` attacks, the stronger responses in the estimated SCM highly correspond to the high levels of reflection. As to the case of (d) `3D-mask` attacks, because 3D-masks are designed to closely mimic the topographical and textural intricacies of genuine faces, the estimated SCMs in (d) usually exhibit much subdued responses.



(a) (b) (c) (d)

Figure 9. The estimated spoof cue maps on (a) live images, (b) `print` attacks, (c) `replay` attacks, and (d) `3D-mask` attacks.

## 5. Conclusion

We have introduced a novel One-Class Spoof Cue Map estimation Network (OC-SCMNet) to address the one-class FAS problem. In OC-SCMNet, we first propose learning the zero spoof cue maps (SCM) estimation from live images by enforcing the latent spoof features to have zero SCM. Next, to avoid trapping in a trivial solution, we propose an SCM-guided feature learning to generate latent spoof features that are distinct from the latent spoof features of zero SCMs. Furthermore, we propose using a memory bank to encourage globally probing of the feature generation process to facilitate SCM estimation. Extensive experiments demonstrate that OC-SCMNet outperforms previous one-class FAS methods and achieves competitive performance with state-of-the-art two-class FAS method.

## Acknowledgement

This work was supported in part by the NSTC grants 112-2221-E-007-082-MY3 and 112-2634-F-007-002 of Taiwan.

## References


[1] André Anjos and Sébastien Marcel. Counter-measures to photo attacks in face recognition: a public database and a baseline. In *2011 international joint conference on Biometrics (IJCB)*, pages 1–7. IEEE, 2011. 5

[2] Yashasvi Baweja, Poojan Oza, Pramuditha Perera, and Vishal M Patel. Anomaly detection-based unknown face presentation attack detection. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–9. IEEE, 2020. 1, 3, 6, 7

[3] 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. 3, 5, 6

[4] Rizhao Cai, Yawen Cui, Zhi Li, Zitong Yu, Haoliang Li, Yongjian Hu, and Alex Kot. Rehearsal-free domain continual face anti-spoofing: Generalize more and forget less. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8037–8048, 2023. 2

[5] Baoliang Chen, Wenhan Yang, Haoliang Li, Shiqi Wang, and Sam Kwong. Camera invariant feature learning for generalized face anti-spoofing. *IEEE Transactions on Information Forensics and Security*, 16:2477–2492, 2021. 6, 7

[6] 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

[7] Nesli Erdogmus and Sebastien Marcel. Spoofing face recognition with 3d masks. *IEEE transactions on information forensics and security*, 9(7):1084–1097, 2014. 5

[8] Haocheng Feng, Zhibin Hong, Haixiao Yue, Yang Chen, Keyao Wang, Junyu Han, Jingtuo Liu, and Errui Ding. Learning generalized spoof cues for face anti-spoofing. *arXiv preprint arXiv:2005.03922*, 2020. 1, 2

[9] Pei-Kai Huang, Ming-Chieh Chin, and Chiou-Ting Hsu. Face anti-spoofing via robust auxiliary estimation and discriminative feature learning. In *Asian Conference on Pattern Recognition*, pages 443–458. Springer, 2021. 1, 2, 7

[10] Pei-Kai Huang, Chu-Ling Chang, Hui-Yu Ni, and Chiou-Ting Hsu. Learning to augment face presentation attack dataset via disentangled feature learning from limited spoof data. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2022. 2

[11] Pei-Kai Huang, Hui-Yu Ni, Yan-Qin Ni, and Chiou-Ting Hsu. Learnable descriptive convolutional network for face anti-spoofing. In *BMVC*, 2022. 2, 6, 7

[12] Pei-Kai Huang, Cheng-Hsuan Chiang, Jun-Xiong Chong, Tzu-Hsien Chen, Hui-Yu Ni, and Chiou-Ting Hsu. Ldc-former: Incorporating learnable descriptive convolution to vision transformer for face anti-spoofing. In *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023. 2, 6

[13] Pei-Kai Huang, Jun-Xiong Chong, Hui-Yu Ni, Tzu-Hsien Chen, and Chiou-Ting Hsu. Towards diverse liveness feature representation and domain expansion for cross-domain face anti-spoofing. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2023. 2, 7

[14] 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*, 2023. 2

[15] Xiaobin Huang, Jingtian Xia, and Linlin Shen. One-class face anti-spoofing based on attention auto-encoder. In *Biometric Recognition: 15th Chinese Conference, CCBR 2021, Shanghai, China, September 10–12, 2021, Proceedings 15*, pages 365–373. Springer, 2021. 1, 3, 6, 7

[16] Jeeho Hyun, Sangyun Kim, Giyoung Jeon, Seung Hwan Kim, Kyunghoon Bae, and Byung Jun Kang. Reconpatch: Contrastive patch representation learning for industrial anomaly detection. *arXiv preprint arXiv:2305.16713*, 2023. 3

[17] 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. 1, 2, 7

[18] Amin Jourabloo, Yaojie Liu, and Xiaoming Liu. Face de-spoofing: Anti-spoofing via noise modeling. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 290–306, 2018. 1, 2

[19] Chen-Hao Liao, Wen-Cheng Chen, Hsuan-Tung Liu, Yi-Ren Yeh, Min-Chun Hu, and Chu-Song Chen. Domain invariant vision transformer learning for face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6098–6107, 2023. 1, 2, 7

[20] Seokjae Lim, Yongjae Gwak, Wonjun Kim, Jong-Hyuk Roh, and Sangrae Cho. One-class learning method based on live correlation loss for face anti-spoofing. *IEEE Access*, 8:201635–201648, 2020. 1, 3, 6, 7

[21] Siqi Liu, Pong C Yuen, Shengping Zhang, and Guoying Zhao. 3d mask face anti-spoofing with remote photoplethysmography. In *European Conference on Computer Vision*, pages 85–100. Springer, 2016. 5

[22] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 389–398, 2018. 5, 7

[23] 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 *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XII*, pages 511–528. Springer, 2022. 2

[24] Olegs Nikisins, Amir Mohammadi, André Anjos, and Sébastien Marcel. On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing. In *2018 International Conference on Biometrics (ICB)*, pages 75–81. IEEE, 2018. 1, 3, 6, 7

[25] Rui Shao, Xiangyuan Lan, Jiawei Li, and Pong C Yuen. Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10023–10031, 2019. 1, 2, 6, 7

[26] Rui Shao, Pramuditha Perera, Pong C Yuen, and Vishal M Patel. Federated generalized face presentation attack detection. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[27] Koushik Srivatsan, Muzammal Naseer, and Karthik Nandakumar. Flip: Cross-domain face anti-spoofing with language guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19685–19696, 2023. 2

[28] I Standard. Information technology—biometric presentation attack detection—part 1: Framework. *ISO: Geneva, Switzerland*, 2016. 3, 5

[29] 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, 7

[30] Chin-Chia Tsai, Tsung-Hsuan Wu, and Shang-Hong Lai. Multi-scale patch-based representation learning for image anomaly detection and segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3992–4000, 2022. 3

[31] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 8

[32] 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. 2, 6, 7

[33] Guoqing Wang, Hu Han, Shiguang Shan, and Xilin Chen. Cross-domain face presentation attack detection via multi-domain disentangled representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6678–6687, 2020. 2

[34] 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. 1, 2, 7

[35] Zezheng Wang, Zitong Yu, Chenxu Zhao, Xiangyu Zhu, Yunxiao Qin, Qiusheng Zhou, Feng Zhou, and Zhen Lei. Deep spatial gradient and temporal depth learning for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5042–5051, 2020. 2

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

[37] 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. 1, 2, 5, 7

[38] 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. 3, 5

[39] Xiao Yang, Wenhan Luo, Linchao Bao, Yuan Gao, Dihong Gong, Shibao Zheng, Zhifeng Li, and Wei Liu. Face anti-spoofing: Model matters, so does data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3507–3516, 2019. 7

[40] William J Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950. 5

[41] 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. *IEEE transactions on pattern analysis and machine intelligence*, 43(9):3005–3023, 2020. 5, 7

[42] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5295–5305, 2020. 2, 5, 6, 7

[43] Zitong Yu, Yunxiao Qin, Hengshuang Zhao, Xiaobai Li, and Guoying Zhao. Dual-cross central difference network for face anti-spoofing. In *IJCAI*, 2021. 2

[44] 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

[45] 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. 2, 7