Dynamic Feature Queue for Surveillance Face Anti-spoofing via Progressive Training

Keyao Wang[*,1,†], Mouxiao Huang[*,1,2,3], Guosheng Zhang[*,1], Haixiao Yue[1], Gang Zhang[1], Yu Qiao[2]
[1]Department of Computer Vision Technology (VIS), Baidu Inc.
[2]ShenZhen Key Lab of Computer Vision and Pattern Recognition,
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
[3]University of Chinese Academy of Sciences
{wangkeyao, huangmouxiao, zhangguosheng, yuehaixiao, zhanggang03}@baidu.com,
{mx.huang, yu.qiao}@siat.ac.cn

Abstract

In recent years, face recognition systems have faced increasingly security threats, making it essential to employ Face Anti-spoofing (FAS) to protect against various types of attacks in traditional scenarios like phone unlocking, face payment and self-service security inspection. However, further exploration is required to fully secure FAS in long-distance settings. In this paper, we propose two contributions to enhance the security of face recognition systems: Dynamic Feature Queue (DFQ) and Progressive Training Strategy (PTS). DFQ converts the conventional binary classification task into a multi-classification task. It treats live samples as a closed set and attack samples as an open set by using a dynamic queue that stores the features of spoofing samples and updates them. On the other hand, PTS targets difficult samples and iteratively adds them in batches for training. The proposed PTS divides the entire training set into blocks, trains only a small portion of the data, and gradually increases the training data with each stage while also incorporating low-scoring positive samples and high-scoring spoof samples from the test set. These two contributions complement each other by enhancing the model's ability to generalize and defend against various types of attacks, making the face recognition system more secure and reliable. Our proposed methods have achieved top performance on ACER metric with 4.73% on the SuHiFiMask dataset [11] and won the first prize in Surveillance Face Anti-spoofing track of the Challenge@CVPR 2023.

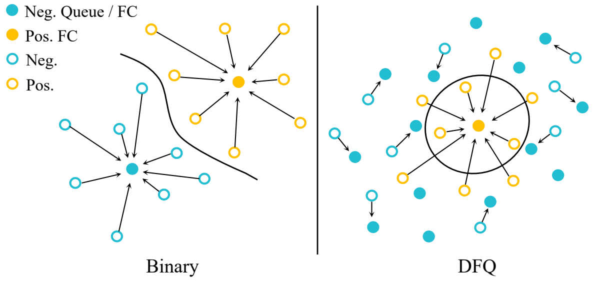

Figure 1. Our proposed method treats each attack case as a separate attack type and models the traditional liveness binary classification task as a multi-class task, based on the assumption that liveness samples are a closed set and attack samples are an open set.

1. Introduction

Face recognition technology [14, 16, 31] has become an integral part of many security and surveillance systems [26, 39]. However, its reliability is constantly challenged by the threat of face spoofing attacks, where an attacker uses a fake face to deceive the system into granting access or authentication, such as replay-attack [7], print-attack [42] and face-mask [10]. Therefore, the development of effective FAS methods has become a critical research direction.

Early FAS methods mainly relied on handcrafted features [1, 4, 8, 17, 28, 29], which required prior knowledge and human liveness cues to distinguish between live and spoof faces. While there have been notable advancements in the performance of face presentation attack detection (PAD) technology in short-distance scenarios [5, 18, 21, 25, 33, 36, 40], such as phone unlocking, face payment, and self-service security inspection, it remains sensitive to face quality and falls short in long-distance applications. This

[*]Equal contribution. This work is done while Mouxiao Huang is an intern at Baidu and being mentored by Keyao Wang and Guosheng Zhang.
†Correspondence to: Keyao Wang (wangkeyao@baidu.com)

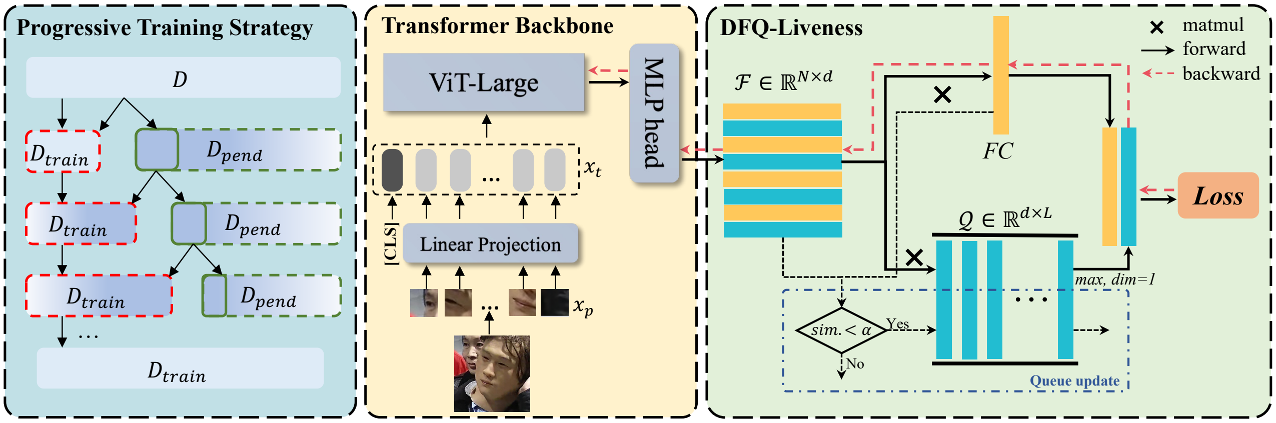

Figure 2. The overview of the proposed Dynamic Feature Queue and Progressive Training Strategy has three parts: (1) Progressive Training Strategy: divides the training set into blocks, trains only a portion of the data at each stage, and increases it with each subsequent stage. (2) Transformer Backbone: divides input images into patches and extracts features. (3) DFQ-Liveness: uses a dynamic queue to store the features of spoofing samples and treats attack samples as an open set during classification.

limitation obstructs the deployment of FAS in surveillance scenarios, where the ability to detect face spoofing attacks is crucial for ensuring the security and integrity of the system. With the emergence of deep learning, FAS methods based on deep neural networks have achieved state-of-the-art performance. However, most existing methods are still vulnerable to unknown spoofing attacks. As the use of surveillance systems in long-distance scenarios becomes more prevalent, detecting face spoofing attacks has become an even more significant challenge [11]. In these scenarios, low-quality faces are common, and they often do not provide sufficient detail for fine-grained feature-based FAS tasks. Therefore, developing effective FAS systems that can handle these challenging scenarios is crucial for ensuring the security and reliability of surveillance applications.

To address these challenges, we propose Dynamic Feature Queue (DFQ) and Progressive Training Strategy (PTS) method for FAS in surveillance scenarios. The DFQ is designed to improve the model's ability to generalize to unknown attack types. By modeling the traditional binary classification task into a multi-classification task, the proposed method can effectively model the unknown characteristics of attack samples and improve the generalization ability of the model. Additionally, the PTS method is designed to refine and extract massive training data. By adding difficult samples in batches for training through the loop iterative training method, the proposed method can improve the model's ability to learn from non-stationary data and adapt to new spoofing attacks. In this paper, we present a comprehensive evaluation of our proposed methods on a challenging surveillance scenarios dataset SuHiFiMask [11] that is first work to extend FAS to real surveillance scenes rather than mimicking low-resolution images and surveillance environments. The experimental results demonstrate the effectiveness and robustness of our proposed methods in detecting various types of face spoofing attacks. And the main contributions of this paper are summarized below:

- We propose the Dynamic Feature Queue (DFQ) and Progressive Training Strategy (PTS) methods for Face Anti-spoofing (FAS) in surveillance scenarios.
- We evaluate the proposed methods on a challenging surveillance dataset, SuHiFiMask [11], which extends FAS to real surveillance scenes.
- The experimental results demonstrate the effectiveness and robustness of our proposed methods in detecting various types of face spoofing attacks.

2. Related Work

Face Anti-spoofing Methods In recent years, FAS has received increasing attention due to its critical role in ensuring the security and reliability of facial recognition systems. Various approaches have been proposed to address this problem, including handcrafted feature-based methods and deep learning-based methods. During the early stages of FAS research, many traditional handcrafted feature-based methods were proposed, which required task-specific prior knowledge. These methods were designed based on human liveness cues, such as gaze tracking [1], facial or head movements [3] and eye-blinking [28], for dynamic discrimination. However, capturing these cues from videos is inconvenient for practical deployment. In addition, classical handcrafted descriptors such as LBP [8], SIFT [29], SURF [4], HOG [17] were developed to extract effective spoofing patterns from various color spaces, which also required task-specific prior knowledge. Recently, deep learning-based methods [2, 12, 19, 22–24, 34, 35, 37, 38, 41] have shown promising results in FAS. These methods use convolutional neural networks (CNNs) to learn highly discriminative features from large-scale datasets. CNN-based methods can capture complex patterns and relationships between features, making them highly effective for FAS. Examples of CNN-based methods include residual-learning frameworks [12], central difference convolution [38], and LSTM [12] or GRU-based [37] methods. Despite the significant progress made in FAS, challenges remain, such as domain adaptation and generalization across different scenarios and attack types. Future research directions include developing more robust and efficient FAS methods that can adapt to different scenarios and attack types.

Face Attacks in Surveillance Scenarios The field of face recognition in surveillance scenarios has garnered significant attention from researchers, who have concentrated on data collection and algorithm design. To facilitate research in this area, several face recognition datasets have been released, including SCface [13], QMUL-Survface [6], and IJB-C [27], which aggregate face images from various sources with the objective of improving the global population's representation. These datasets provide researchers with large-scale, real-world data to facilitate the development and evaluation of face recognition algorithms specifically designed for surveillance scenarios. Face recognition algorithms have been developed for surveillance scenarios utilizing these datasets. This work [20] implemented adversarial generative networks and fully convolutional architectures for the supervised discriminative learning of ground-resolution faces. SFace [43] proposed the sigmoid-constrained hypersphere loss to reduce intra-class distance of high-quality samples while avoiding over-fitting label noise. AdaFace [16] introduced an adaptive marginal function to prioritize the role of clean samples in classification by adjusting the importance of different samples.

3. Methodology

This section presents the proposed methods for enhancing the security of face recognition systems, which include Dynamic Feature Queue (DFQ) and Progressive Training Strategy (PTS). To begin, we introduce the problem formulation in Section 3.1, followed by detailed explanations of DFQ in Section 3.2 and PTS in Section 3.3. An overview of our methods can be found in Figure 2.

3.1. Problem Formulation

Algorithm 1 Pseudocode of DFQ in a Paddle-like style.

```
# encoder: encoder network for feature extraction
# fc: linear layer for classification
# queue: feature queue for negative (spoof) samples
# scale: scaling factor for logit
# alpha: similarity threshold for enqueued features

# load a minibatch data with N samples
for imgs, labels in loader:
    feat = encoder(imgs) # features Nxd
    feat = normalize(feat) # features normalization

    # similarity the with fc: Nx1
    log0 = matmul(feat, normalize(fc.weight))

    # similarity the with queue: NxQ
    sim = matmul(feat, queue.detch())

    # choose the most similar sample: Nx1
    log1 = topk(sim, 1, axis=1, largest=True)[0]

    # logits: Nx2
    logits = concat((log0, log1), axis=1)

    # cross entropy loss
    loss = CrossEntropyLoss(logits * scale, labels)

    # updata encoder and fc layer
    loss.backward()
    update([encoder.params, fc.weight])

    # choose easy negative features
    neg = feat[(label > 0) & (log0[:, 0] < alpha)]

    # update queue
    enqueue(queue, neg)
    dequeue(queue)
```

The problem of face anti-spoofing in surveillance scenarios can be formulated as a classification task, where a given face image needs to be classified as either belonging to an authorized individual or an unauthorized individual. Mathematically, the dataset is formulated as $\{x_i \in X, y_i \in Y\}$, where $\{x_i \in X, i = \{1, 2, ..., N\}\}$ is the input space of face images and $\{y_i \in Y = \{0, 1\}\}$ is the output space of binary labels where 0 represents the class of authorized individuals and 1 represents the class of unauthorized individuals. Here, $x_i \in \mathbb{R}^{h \times w \times 3}$ is a face image with height h, width w and 3 channels, and N is the number of identities in the dataset. The goal of the FAS system is to learn a mapping function $y = f(x) : X \to Y$ that can accurately predict the class label of a given face image. Specifically, we aim to learn a feature extractor $\Phi(x; \phi) \in \mathbb{R}^d$ with learnable parameters ϕ to encode images into feature embeddings and then a fully connected layer $W \in \mathbb{R}^{d \times N}$ maps the features to the corresponding label y_i. The encoder is trained by minimizing the Cross Entropy loss:

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(W^T \cdot \Phi(x_i; \phi), y_i) \tag{1}$$

3.2. Dynamic Feature Queue

To address the issue of limited defense capability of FAS systems against unknown attack types in surveillance scenarios, we introduce a novel dynamic queue training algorithm called Dynamic Feature Queue (DFQ), as shown in Figure 2 and Figure 3. Our approach is based on the hypothesis that live samples form a closed set, while attack samples form an open set, modeling the traditional binary classification task into a multi-classification task. The main

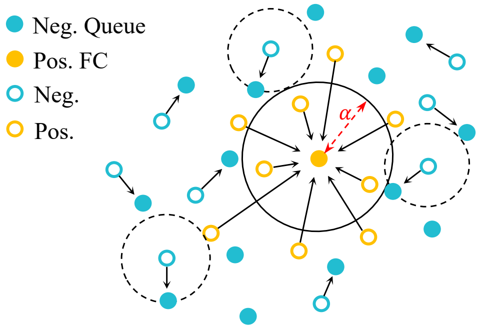

Figure 3. The process of samples learning during DFQ training. Live samples are considered a closed set, while attack samples are treated as an open set.

idea behind DFQ is to utilize a dynamic queue to update the model with negative (spoof) samples.

The pseudocode of DFQ is presented in Algorithm 1. The algorithm takes a mini-batch of images and their corresponding labels as input. It extracts the features for each image and normalizes them. It then calculates the similarity between the features and the linear layer's weight (the center of positive samples), giving a similarity score of each sample with the weight. Next, it calculates the similarity between the features and the feature queue using matrix multiplication, which results in a similarity score for each sample with each negative sample in the queue. It chooses the most similar negative sample for each image, and concatenates the similarity score of the image with the similarity score of its corresponding negative sample to form a 2-dimensional tensor of logits. The Cross Entropy loss can be calculated and then update the encoder and linear layer. After that, it selects easy negative features by filtering those with similarity scores lower than a certain threshold α and updates the feature queue by adding the new negative features and removing the oldest ones.

The DFQ algorithm iteratively trains the encoder network and linear layer and updates the feature queue to generate negative samples for better training. The use of negative samples from the queue provides additional information for the model to learn the feature space better, improving the model's performance. Additionally, DFQ model inference calculates feature similarity with positive FC, eliminating the need for extra model parameters.

3.3. Progressive Training Strategy

To address the problem of hard samples in deep networks under massive non-stationary data, we propose a Progressive Training Strategy (PTS) algorithm. PTS is a training strategy for deep metric learning that aims to address the problem of hard sample mining. The PTS algorithm

Algorithm 2 Pseudocode of PTS in a Paddle-like style.

```
# isr: initial sampling rate
# hsr: hard sample rate
# dr: decay rate
# pss: progressive step size
# infos: [(img_path, label),...]
# ds: dataset to train

# initial training
p_l, n_l = len(ds.p_infos), len(ds.n_infos)
p_infos = random.sample(ds.p_infos, (1 - isr) * p_l)
n_infos = random.sample(ds.p_infos, (1 - isr) * n_l)

pend_ds = copy(ds)
pend_infos = p_infos + n_infos
ds.infos = list(set(ds.infos) - set(pend_infos))
loader = build_loader(ds)

# start training
for ep in range(max_epoch):

    train_epoch(loader, model)

    if (ep + 1) % pss == 0:
        p_l, n_l = len(p_infos), len(n_infos)

        # update val_ds and val_loader
        pend_ds.infos = p_infos + n_infos
        pend_loader = build_loader(pend_ds)

        with paddle.no_grad():
            preds = test_epoch(pend_loader, model)

        # sort by prediction score
        p_idxs = argsort(preds[:p_l])
        n_idxs = argsort(-preds[n_l:])

        # mining hard samples
        ph_infos = p_infos[p_idxs[:hsr * p_l]]
        nh_infos = n_infos[n_idxs[:hsr * n_l]]

        # update p_infos and n_infos
        p_infos = p_infos[p_idxs[hsr * p_l:]]
        n_infos = n_infos[n_idxs[hsr * n_l:]]

        # update ds and loader
        ds.infos += ph_infos + nh_infos
        loader = build_loader(ds)

        hsr *= dr # hard sample rate decay
```

consists of two key components: initial sampling rate and progressive step size. The initial sampling rate determines the percentage of easy samples to be used during the initial training, while the progressive step size is the number of epochs after which the hard samples are progressively added to the training set.

The pseudocode for the PTS algorithm is shown in Algorithm 2. PTS begins by randomly selecting a subset of easy positive and negative samples from the training dataset, based on the given initial sampling rate. The remaining samples are used for validation. During training, the algorithm progressively adds hard samples to the training set, based on the validation loss. Specifically, after every pss epochs, the algorithm evaluates the model on the validation set and sorts the validation samples by their predicted scores. Then, the algorithm mines the top hsr percentage of hard samples and adds them to the training set. Finally, the sampling rate is reduced by a decay rate factor, dr, to ensure that the algorithm progressively focuses more on hard samples as the training progresses.

In general, PTS offers a framework for training deep neural networks on massive non-stationary data, while also preserving the capability to identify previously learned classes through hard sample data mining.

4. Experiments

4.1. Experimental Setup

Datasets To assess the efficacy of our proposed method in surveillance scenarios, we utilized a large-scale dataset called SuHiFiMask [11] as our primary dataset. SuHiFiMask comprises 40 real-life surveillance scenes, such as movie theaters, security gates, and parking lots, representing a diverse range of face recognition scenarios. It includes 101 participants of different ages and genders, engaged in various natural activities of daily life. The dataset also incorporates several types of spoofing attacks, including high-fidelity masks, 2D attacks, and adversarial attacks. The data was collected under realistic outdoor conditions, capturing diverse weather and lighting conditions. The SuHiFiMask [11] dataset is divided into three subsets: {training, dev, and test}, with {159,063, 89,276, and 161,882} images, respectively. The partitioning of images into these subsets is based on their quality scores, which are assigned according to a specific range for each subset. Specifically, SuHiFiMask [11] assigns images with quality scores ranging from [0.4, 1] to the training set, scores from [0.3, 0.4) to the dev set, and scores from [0, 0.3) to the test set. This dataset provides a comprehensive and varied set of data for evaluating and refining FAS algorithms in surveillance settings. As seen in Figure 4, the SuHiFiMask dataset exhibits significant differences in image quality scores across its three parts. Images in the train set are generally of good quality, with clear facial features that are easy to identify. Faces in the dev set are of lower quality, with a range of noise types such as masks and occlusions. In contrast, the test set has the poorest quality faces with the highest amount of noise compared to the train and dev sets, including various types of noise such as motion blur, lens flare, and low lighting conditions. This implies that the proposed FAS method must be robust enough to handle these types of noise to be effective in real-world surveillance settings.

Evaluation Metrics To assess the efficacy of our proposed approach for FAS, we utilize the widely accepted metrics, namely the Attack Classification Error Rate (ACER) and Area Under the Curve (AUC), concurrently. The ACER quantifies the FAS system's ability to accurately identify legitimate faces and fake faces by averaging the attack presentation classification error rate (APCER) and the bona fide presentation classification error rate (BPCER) at a specific decision threshold. The APCER and BPCER are instrumental in determining the accuracy of classifying an image as either live or spoof and in establishing the balance between security and convenience. The APCER evaluates the percentage of presentation attacks that are misidentified as bona fide examples, indicating the level of security vulnerability. Conversely, the BPCER gauges the percentage of bona fide examples that are misidentified as presentation attacks, indicating the degree of user inconvenience. ACER metric is defined as follows:

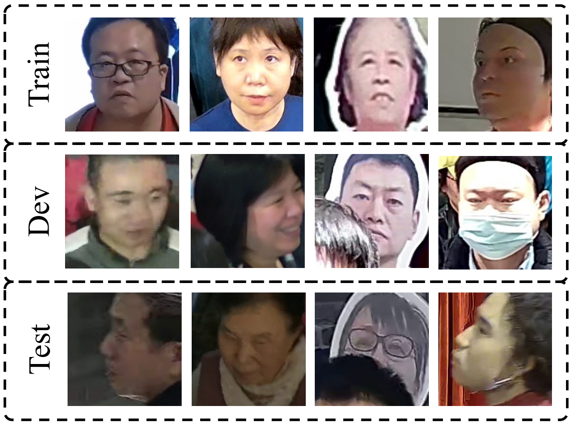

Figure 4. Samples from SuHiFiMask [11] dataset. Images are assigned with quality scores ranging from [0.4, 1] to the training set, scores from [0.3, 0.4) to the dev set, and scores from [0, 0.3) to the test set.

$$ACER = \frac{APCER + BPCER}{2} \tag{2}$$

ACER is a widely used performance metric for FAS systems, and it is often reported alongside other metrics such as AUC, which measures the ability of a model to distinguish between positive and negative classes. AUC metric is defined as follows:

$$AUC = \int_{0}^{1} TPR(t) dFPR(t) \tag{3}$$

where $TPR(t)$ and $FPR(t)$ are the true positive rate and false positive rate, respectively, at a given classification threshold t.

Implementation Details To train our model, we utilized 8 V100 GPUs with 32G memory each. For feature extraction, we employed a ViT-Large backbone [9] with 300 million parameters, which was pretrained on the ImageNet-1K dataset [30]. The input images are resized to $224 \times 224 \times 3$ and normalized using the mean and standard deviation computed from the SuHiFiMask [11] dataset. The feature embedding dimension is set to 768. We use a batch size of 64 and use SGD with momentum 0.9. The total number of epochs is 120, with a warmup strategy applied for the first 2000 iterations. We set the initial learning rate to 0.01 and use Cosine decay to gradually reduce the learning rate. The hyperparameters in PTS are set to $isr = 0.2$, $hsr = 0.15$, $dr = 0.5$, respectively. To address the poor quality of face images in surveillance scenarios, as shown in Figure 4, which often suffer from blurriness, changes in illumination, occlusions, compression artifacts, and other issues, we applied various image augmentation techniques including random flip, random rotation, random crop, photometric distortion, and blurs. To ensure that the distribution of the training set data is as close as possible to that of the test set, we also used low-quality augmentations like motion blur, which was found to be particularly effective in monitoring scenarios where people are in motion. To further improve the robustness of our models, we utilize Test-Time Augmentation (TTA) [32], a technique that enhances the accuracy and generalization performance of deep learning models. We apply three augmentations to each test image, including random flip, rotation, and crop, and average the predictions of the original and augmented images to yield the final prediction. By leveraging TTA, our models are better equipped to handle variations in the test data, leading to improved performance and generalization capabilities.

4.2. Experimental Results

Comparison of Backbones To ensure effective feature extraction and superior performance in a face anti-spoofing model, selecting a suitable backbone network is crucial. We conducted a comprehensive evaluation of popular and effective models [9, 15], including ResNet50, ResNet101, ViT-S, ViT-B, and ViT-L, on the SuHiFiMask validation set without employing any training strategies or tricks. The results in Table 1 demonstrate that ViT-L outperforms other models in terms of ACC (accuracy), AUC, and ACER metrics. Additionally, as the face images in the test set are usually of lower quality than those in the validation set, we selected ViT-L as the backbone of our face anti-spoofing model to achieve superior performance on low-quality images.

Comparison with SOTA Methods We compare the performance of our proposed method with state-of-the-art (SOTA) methods (teams) on the SuHiFiMask dataset [11]. Table 2 summarizes the results of the comparison in terms of four metrics: AUC, APCER, BPCER, and ACER. Among the SOTA methods, CTEL_AI achieves the lowest BPCER with a value of 1.90%. However, our proposed method achieves the highest performance on the AUC and ACER metrics, with values of 98.38% and 4.73%, respectively. In addition, our method achieves the lowest APCER with a value of 5.07%, which is significantly lower than the other methods. The experimental results confirm the effectiveness of our proposed method that integrates the DFQ and PTS methods to enhance the performance of face anti-spoofing. Our approach surpasses SOTA methods on the SuHiFiMask dataset, which is specifically designed for surveillance scenarios with complex challenges such as diverse types of attacks, as well as variations in illumination, occlusion, noise, and other issues. These results highlight the potential of our method to improve the accuracy and reliability of face anti-spoofing in real-world scenarios.

Models	ACC ↑	AUC ↑	ACER ↓
ResNet50	97.91	99.12	2.09
ResNet101	96.73	98.83	3.27
ViT-S	98.06	99.19	1.94
ViT-B	98.22	99.21	1.78
ViT-L	**98.37**	**99.25**	**1.63**

Table 1. Performance comparison of different backbone networks on the SuHiFiMask [11] validation set (dev split), based on ACC, AUC, and ACER (%) metrics.

Team	AUC ↑	APCER ↓	BPCER ↓	ACER ↓
OPDAI	97.38	9.18	5.13	7.16
hexianhua	97.83	11.21	2.94	7.08
horsego	96.97	8.17	4.26	6.22
CTEL_AI	98.21	9.20	**1.90**	5.56
Ours	**98.38**	**5.07**	4.38	**4.73**

Table 2. Comparing results on the test set of the SuHiFiMask [11] dataset. Our method achieves the highest performance on the AUC, APCER, and ACER (%) metrics.

Method	AUC ↑	APCER ↓	BPCER ↓	ACER ↓
baseline	97.72	8.37	6.52	7.44
w/ DFQ	97.88	7.27	**4.07**	5.67
w/ PTS	98.06	**4.84**	6.92	5.88
w/ DFQ & PTS	**98.38**	5.07	4.38	**4.73**

Table 3. The effectiveness of DFQ and PTS was evaluated on the test set of the SuHiFiMask [11], and the results showed significant improvements in face anti-spoofing performanc (%).

Effect of DFQ and PTS To assess the effectiveness of our proposed DFQ and PTS methods, we conducted extensive experiments on the SuHiFiMask dataset [11]. The baseline model was trained without DFQ or PTS, while the other models were trained with either DFQ, PTS, or a combination of both. Table 3 shows the results of our experiments, indicating that incorporating DFQ and PTS significantly improved face anti-spoofing performance, as measured by the AUC, APCER, BPCER, and ACER metrics. Notably, the model trained with both DFQ and PTS achieved the best performance on AUC and ACER metrics. These findings demonstrate the effectiveness of our proposed methods in mitigating the impact of spoofing attacks on face recognition systems in surveillance scenarios.

5. Conclusion

In conclusion, we have proposed a novel method for FAS in surveillance scenarios by utilizing DFQ and PTS algorithms. Our approach achieved SOTA performance on the SuHiFiMask [11] dataset, which is a challenging benchmark due to variations in illumination, pose, and various types of attacks, and won the first prize in Surveillance Face Anti-spoofing track of the Challenge@CVPR 2023. The results demonstrate the effectiveness of our method in enhancing the security of face recognition systems. This work presents new opportunities for future research on FAS in real-world scenarios and provides a solid foundation for further development and exploration in the field of FAS.

References

[1] Asad Ali, Farzin Deravi, and Sanaul Hoque. Liveness detection using gaze collinearity. In *2012 Third International Conference on Emerging Security Technologies*, pages 62–65. IEEE, 2012. 1, 2

[2] Yousef Atoum, Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Face anti-spoofing using patch and depth-based cnns. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 319–328. IEEE, 2017. 3

[3] Wei Bao, Hong Li, Nan Li, and Wei Jiang. A liveness detection method for face recognition based on optical flow field. In *2009 International Conference on Image Analysis and Signal Processing*, pages 233–236. IEEE, 2009. 2

[4] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face antispoofing using speeded-up robust features and fisher vector encoding. *IEEE Signal Processing Letters*, 24(2):141–145, 2016. 1, 2

[5] 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*, volume 35, pages 1132–1139, 2021. 1

[6] Zhiyi Cheng, Xiatian Zhu, and Shaogang Gong. Surveillance face recognition challenge. *arXiv preprint arXiv:1804.09691*, 2018. 3

[7] 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. 1

[8] Tiago de Freitas Pereira, André Anjos, José Mario De Martino, and Sébastien Marcel. Lbp- top based countermeasure against face spoofing attacks. In *Computer Vision-ACCV 2012 Workshops: ACCV 2012 International Workshops, Daejeon, Korea, November 5-6, 2012, Revised Selected Papers, Part I 11*, pages 121–132. Springer, 2013. 1, 2

[9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021. 5, 6

[10] Nesli Erdogmus and Sébastien Marcel. Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect. In *2013 IEEE sixth international conference on biometrics: theory, applications and systems (BTAS)*, pages 1–6. IEEE, 2013. 1

[11] Hao Fang, Ajian Liu, Jun Wan, Sergio Escalera, Chenxu Zhao, Xu Zhang, Stan Z. Li, and Zhen Lei. Surveillance face anti-spoofing, 2023. 1, 2, 5, 6, 7

[12] 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. 3

[13] Mislav Grgic, Kresimir Delac, and Sonja Grgic. Scface–surveillance cameras face database. *Multimedia tools and applications*, 51:863–879, 2011. 3

[14] Jianzhu Guo, Xiangyu Zhu, Chenxu Zhao, Dong Cao, Zhen Lei, and Stan Z Li. Learning meta face recognition in unseen domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6163–6172, 2020. 1

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 6

[16] Minchul Kim, Anil K Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18750–18759, 2022. 1, 3

[17] 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. 1, 2

[18] Bi Li, Teng Xi, Gang Zhang, Haocheng Feng, Junyu Han, Jingtuo Liu, Errui Ding, and Wenyu Liu. Dynamic class queue for large scale face recognition in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3763–3772, 2021. 1

[19] Lei Li, Zhaoqiang Xia, Xiaoyue Jiang, Fabio Roli, and Xiaoyi Feng. Compactnet: learning a compact space for face presentation attack detection. *neurocomputing*, 409:191–207, 2020. 3

[20] Pei Li, Loreto Prieto, Domingo Mery, and Patrick J Flynn. On low-resolution face recognition in the wild: Comparisons and new techniques. *IEEE Transactions on Information Forensics and Security*, 14(8):2000–2012, 2019. 3

[21] Ajian Liu and Yanyan Liang. Ma-vit: Modality-agnostic vision transformers for face anti-spoofing. 1

[22] Ajian Liu and Yanyan Liang. Ma-vit: Modality-agnostic vision transformers for face anti-spoofing. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 1180–1186, 2022. 3

[23] Ajian Liu, Zichang Tan, Jun Wan, Yanyan Liang, Zhen Lei, Guodong Guo, and Stan Z Li. Face anti-spoofing via adversarial cross-modality translation. *IEEE Transactions on Information Forensics and Security*, 16:2759–2772, 2021. 3

[24] Ajian Liu, Chenxu Zhao, Zitong Yu, Jun Wan, Anyang Su, Xing Liu, Zichang Tan, Sergio Escalera, Junliang Xing,

Yanyan Liang, et al. Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 17:2497–2507, 2022. 3

[25] 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. 1

[26] Zuheng Ming, Muriel Visani, Muhammad Muzzamil Luqman, and Jean-Christophe Burie. A survey on anti-spoofing methods for facial recognition with rgb cameras of generic consumer devices. *Journal of Imaging*, 6(12):139, 2020. 1

[27] Hajime Nada, Vishwanath A Sindagi, He Zhang, and Vishal M Patel. Pushing the limits of unconstrained face detection: a challenge dataset and baseline results. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–10. IEEE, 2018. 3

[28] Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *2007 IEEE 11th international conference on computer vision*, pages 1–8. IEEE, 2007. 1, 2

[29] Keyurkumar Patel, Hu Han, and Anil K Jain. Secure face unlock: Spoof detection on smartphones. *IEEE transactions on information forensics and security*, 11(10):2268–2283, 2016. 1, 2

[30] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. 5

[31] Lei Shang, Mouxiao Huang, Wu Shi, Yuchen Liu, Yang Liu, Fei Wang, Baigui Sun, Xuansong Xie, and Yu Qiao. Improving training and inference of face recognition models via random temperature scaling, 2022. 1

[32] Divya Shanmugam, Davis Blalock, Guha Balakrishnan, and John Guttag. Better aggregation in test-time augmentation, 2021. 6

[33] 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

[34] Xiao Song, Xu Zhao, Liangji Fang, and Tianwei Lin. Discriminative representation combinations for accurate face spoofing detection. *Pattern Recognition*, 85:220–231, 2019. 3

[35] Jun Wan, Sergio Escalera, Hugo Jair Escalante, Guodong Guo, and Stan Z Li. Special issue on face presentation attack detection. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(3):282–284, 2021. 3

[36] 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, 2022. 1

[37] 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*, volume 35, pages 2746–2754, 2021. 3

[38] Zitong Yu, Xiaobai Li, Xuesong Niu, Jingang Shi, and Guoying Zhao. Face anti-spoofing with human material perception. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*, pages 557–575. Springer, 2020. 3

[39] Zitong Yu, Yunxiao Qin, Xiaobai Li, Chenxu Zhao, Zhen Lei, and Guoying Zhao. Deep learning for face anti-spoofing: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 1

[40] Haixiao Yue, Keyao Wang, Guosheng Zhang, Haocheng Feng, Junyu Han, Errui Ding, and Jingdong Wang. Cyclically disentangled feature translation for face anti-spoofing. *arXiv preprint arXiv:2212.03651*, 2022. 1

[41] Shifeng Zhang, Ajian Liu, Jun Wan, Yanyan Liang, Guodong Guo, Sergio Escalera, Hugo Jair Escalante, and Stan Z Li. Casia-surf: A large-scale multi-modal benchmark for face anti-spoofing. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(2):182–193, 2020. 3

[42] 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. 1

[43] Yaoyao Zhong, Weihong Deng, Jiani Hu, Dongyue Zhao, Xian Li, and Dongchao Wen. Sface: Sigmoid-constrained hypersphere loss for robust face recognition. *IEEE Transactions on Image Processing*, 30:2587–2598, 2021. 3