2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Diversified Style Generation for Face Anti-Spoofing

Daoyang Lin[1] and Danwei Chen[1]

[1] Nanjing University of Posts and Telecommunications, Yadong New District, Nanjing 210023, China
1392170875@qq.com

Abstract. Face Anti-Spoofing (FAS) is crucial for protecting face recognition systems against various spoofing attacks. However, existing methods still suffer from significant performance degradation when handling unseen domains. To address this challenge, this paper designs a Diversified Style Transformation Network (DSTN) that enhances the domain generalization capability of FAS models through instance-level style augmentation. At the core is a method called Diversified Style Generation (DSG). DSG introduces a set of learnable style bases and uses the Dirichlet distribution to generate dynamic weights for each sample, constructing diversified style-enhanced features. During training, the model is exposed to a broader range of style variations, thereby learning style-invariant features. In addition, this paper designs a content consistency loss and a style diversity loss to preserve semantic information in the augmented features and to encourage diversity among style bases, further improving model robustness. Experiments on multiple standard cross-domain FAS benchmark datasets show that the proposed method outperforms state-of-the-art approaches across various domains, especially in unseen domain tasks, demonstrating stronger generalization capabilities. These results verify the effectiveness and potential of DSG in solving the domain generalization problem.

Keywords: Face Anti-spoofing, Domain Generalization, Style Augmentation.

1 Introduction

Face Anti-Spoofing plays a pivotal role in protecting face recognition systems from various spoofing attacks such as printed images or video replays [1-4]. To tackle these attacks, researchers have developed numerous FAS techniques, including methods based on handcrafted features [5-7] and deep learning approaches for feature extraction [8-11]. Although these methods have achieved excellent performance on specific datasets, their performance typically drops significantly when applied to unseen domains due to changes in data distributions.

To enhance generalization in unseen domains, recent studies have incorporated Domain Generalization (DG) techniques into FAS tasks, employing methods like adversarial learning [12-14] and meta-learning [15,16] to learn domain-invariant representations. Despite the progress of DG-based methods, they often depend on domain labels to align distributions and learn domain-invariant features. However, these domain labels are usually coarse-grained and fail to fully capture the real complexity of domain distributions. Within source domains, variations in lighting conditions, attack types, and background scenes are often overlooked, resulting in fine-grained subdomains. While D2AM [15] attempts to distinguish mixed source domains by assigning pseudo-domain labels, it still requires manually choosing the number of pseudo-domains, thus only partially mitigating the problem. Moreover, aligning domains at the distribution level imposes constraints on the features and cannot guarantee that all feature channels are invariant to instance-specific variations. Consequently, the learned features may still contain information sensitive to these instance-specific styles, limiting generalization performance on new samples from unseen domains.

Inspired by [36], which pioneered the idea of leveraging learnable style bases to project unseen testing samples into a known source domain space, this paper takes a complementary approach by focusing on training-time style augmentation. Instead of relying on test-time updates, this paper proposes a Diversified Style Generation (DSG) method that enriches the model's exposure to diverse style variations during the training phase, thereby improving its generalization capability on unseen domains. Specifically, drawing on the concept of learnable style bases introduced by [36], this paper introduces a set of such bases and dynamically samples weights from a Dirichlet distribution for each training instance. These style bases, when combined through the sampled weights, generate a wide range of style-enhanced features at the instance level. As a result, the model experiences more diverse style patterns during training and learns representations that are less sensitive to style changes. In addition, the proposed approach integrates supervised contrastive and classification losses, thereby forming a fully coherent and comprehensive training framework.

The main contributions can be summarized as follows:

- A new framework is proposed to significantly enhance the cross-domain generalization capability of FAS models by generating instance-level, diversified style-augmented features.
- Leveraging learnable style bases and dynamically generated weights from a Dirichlet distribution, DSG achieves effective style augmentation of original features. To ensure the learning of style-robust features, a style diversity loss is introduced to encourage diversity among style bases, along with a content consistency loss to preserve semantic information in the augmented features.
- Comprehensive experiments on multiple benchmark datasets demonstrate that the proposed method outperforms state-of-the-art approaches in certain aspects of cross-domain FAS tasks, highlighting its effectiveness and potential.

2 Related Work

2.1 Face Anti-Spoofing

Face Anti-Spoofing aims to determine whether an image represents a genuine face or a spoofed attack (such as a printed photo or a replayed video). Early FAS methods mainly relied on handcrafted features [7,17,18] to detect spoofing patterns. With the development of deep learning, various techniques have been explored, including classification-based methods [9,19,20], regression-based methods [21,22], and generative models [23-25], aiming to improve FAS performance. Recently, vision Transformers [26,27] have shown promising results in FAS tasks. However, although these methods perform well in intra-dataset evaluations, they often experience significant performance degradation when applied to different target domains.To address this challenge, domain adaptation techniques [28-30] have been introduced into FAS models. However, such methods rely on accessing target domain data, which may not always be available in real scenarios, limiting their effectiveness. Nevertheless, these methods mostly focus on learning domain-invariant features during training, and still struggle when the target domain differs significantly from the source data.

2.2 Style Augmentation

Style augmentation enriches training data by altering style attributes, enhancing robustness and generalization under diverse variations. It is widely applied in computer vision tasks such as image classification, object detection, and image generation.

Neural network-based style transfer methods separate and recombine content and style by adjusting image statistics. For example, Gatys et al. [32] employed convolutional neural network feature statistics for style transfer, while Huang and Belongie introduced Adaptive Instance Normalization (AdaIN) [33], enabling real-time style manipulation.

In face anti-spoofing, variations in lighting, camera equipment, and attack media (e.g., paper, screen) degrade model performance in unseen domains. To mitigate this, researchers have explored style augmentation for FAS [14,31], generating diverse style transformations to improve adaptability.However, existing FAS style augmentation methods often depend on manually designed or globally applied transformations, lacking the ability to fully capture style complexity and to personalize augmentation at the instance level.

To overcome these limitations, this paper proposes a new approach based on DSG. By introducing learnable style bases and dynamically generating style-enhanced features per sample, this approach achieves instance-level style diversity and significantly improves generalization to unseen domains. Incorporating DSG during training enables the model to learn style-invariant features, thereby strengthening its cross-domain FAS performance.

3 Methodology

Figure 1 presents an overview of the proposed Diversified Style Transformation Network (DSTN), which aims to improve the domain generalization capability of FAS models through instance-level style augmentation. The key component is the DSG module. DSG first learns diverse style bases from training data to provide a rich representation of the style space. Then, a dynamic weighting module uses a Dirichlet distribution to generate style combination weights for each input instance, achieving personalized style augmentation. Additionally, this paper designs two losses to supervise the DSG training.

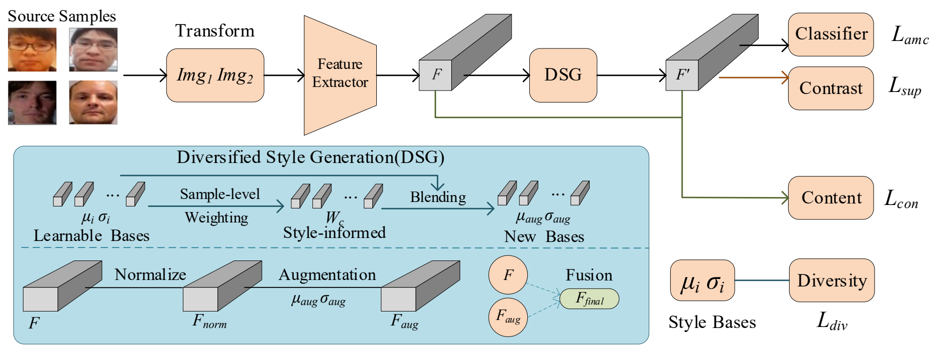


Fig. 1. An overview of the proposed DSTN framework.

3.1 Diversified Style Generation

Introduction of Style Bases. In real-world anti-spoofing tasks, facial data styles are influenced by various factors such as capture devices, lighting, and resolution. Such style diversity imposes great challenges for generalization. If a model cannot adapt to these style changes, its performance on unseen domains typically degrades.

To address this issue, this paper introduces the concept of style bases, which are learnable parameters intended to represent the underlying style patterns in the data. Specifically, we define a set of style bases $\{\mu_b^i, \sigma_b^i\}_{i=1}^N$, where N is the number of style bases, and μ_b^i and σ_b^i denote the mean and standard deviation of the i-*th* style base, respectively. By incorporating these style bases, the model can effectively capture and represent the stylistic characteristics of the data. In particular, μ_b^i encodes the central position of the style, while σ_b^i describes its distributional shape.

By learning these style bases, the model can generate a variety of style features that potentially cover unseen style variations, enhancing robustness and guiding the model toward more semantic features.

Dynamic Weight Generation. Relying solely on fixed style bases is not sufficient for handling diverse inputs. This paper uses a Dirichlet distribution to dynamically generate a weight vector $\mathbf{W}_c \in \mathbb{R}^N$ for each sample, combining different style bases at the instance level.

To encourage style diversity, this paper adjusts the Dirichlet distribution parameters α based on the cosine similarities among the style bases. First, this paper calculates the average similarity of each style base to measure its distinctiveness. Bases with lower similarity are given higher sampling weights, thus increasing their participation in style generation. This design ensures a dynamic balance in style base usage and avoids over-dependence on specific bases while maximizing style coverage.

Formally, this paper first computes the cosine similarity matrix among style bases:

$$S = [cos(\mu_i^b, \mu_j^b)]_{N\times N} \tag{1}$$

Then, it calculates the average similarity for each base:

$$\overline{s}_i = \frac{1}{N-1}\sum_{j\neq i} S_{i,j} \tag{2}$$

Next, it adjusts the Dirichlet parameters α based on the average similarity:

$$\alpha_i = \frac{1}{N} + \lambda(1-\overline{s}_i) \tag{3}$$

where λ is a hyperparameter controlling the adjustment. This hyperparameter governs the extent to which style base similarity influences the sampling weights. Its value was carefully tuned on a held-out validation set. A larger λ more aggressively promotes less similar bases, while a smaller one results in a more uniform sampling distribution. The value $\lambda=0.5$ was selected as it yielded the best cross-domain generalization performance during validation, effectively encouraging style diversity without destabilizing the learning process. Finally, dynamic weights are sampled from the Dirichlet distribution:

$$Wc \sim Dirichlet(\alpha) \tag{4}$$

Generating Style-Enhanced Features. With dynamic weights and style bases, this paper enhances the original features with style. First, it uses the weights to combine style bases and produce augmented means and standard deviations. Then, these statistics are applied to the original features for style transformation.

This approach simulates style perturbations in feature space, improving the model's robustness to style changes and exposing it to a broader style distribution. To preserve the semantic content of the original features, this paper fuses the enhanced features and the original features, striking a balance between augmentation intensity and information retention.

Specifically, let the model-extracted original feature representation be $F_{org} \in \mathbb{R}^{B\times D}$, where B is the batch size and D is the feature dimension. The first step is to normalize these features per sample:

$$F_{norm} = \frac{F_{org} - \mu_{org}}{\sigma_{org} + \epsilon} \tag{5}$$

where $F_{org} \in \mathbb{R}^{B\times 1}$ and $\sigma_{org} \in \mathbb{R}^{B\times 1}$ are the mean and standard deviation of the original features (computed per sample), and ϵ is a small constant to prevent division by zero. Next, using the style bases and the dynamically sampled weights, the augmented mean and standard deviation are generated:

$$\mu_{aug} = W_c \cdot M, \sigma_{aug} = W_c \cdot S \tag{6}$$

where $M = [\mu_b^1, \mu_b^2, \ldots, \mu_b^N]^{\top} \in \mathbb{R}^{N\times D}$ .Afterward, style enhancement is applied to the normalized features:

$$F_{aug} = \sigma_{aug} \odot F_{norm} + \mu_{aug} \tag{7}$$

where $\odot$ denotes element-wise multiplication.

Finally, the enhanced features are fused with the original features to obtain the final representation:

$$F_{final} = F_{org} + \lambda_{aug} \cdot F_{aug} \tag{8}$$

where λ_{aug} controls the augmentation strength.

Style Diversity Loss. When introducing style bases, the absence of additional constraints can cause them to become overly similar, potentially leading to a collapse of the style space. Such a scenario not only diminishes the effectiveness of style augmentation but also restricts the model's generalization capabilities. To address this, a style diversity loss is introduced to promote orthogonality among style bases in the feature space. More specifically, by minimizing the squared difference of the cosine similarities between style bases, their distinctiveness is increased. This design drives the style bases apart in the feature space, resulting in richer style representations.

The core purpose of this loss is to enhance the expressiveness of the style bases so that they can capture more diverse style patterns, ultimately improving both the variety of generated style features and the model's ability to generalize. In practice, cosine similarities are computed between the means and standard deviations of the style bases, and their squared differences are then minimized:

$$L_{diversity} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} [\sin^2(\mu_b^i, \mu_b^j) + \sin^2(\sigma_b^i, \sigma_b^j)] \tag{9}$$

Minimizing $L_{diversity}$ encourages style bases to be orthogonal and enhances style diversity.

Content Consistency Loss. Although style augmentation improves the model's robustness to style variations, overemphasizing stylistic information may cause the model to neglect essential content features of input samples, ultimately affecting the task's primary goal. To avoid this, a content consistency loss is introduced to ensure that the enhanced features incorporate style variations while retaining the original content information. Specifically, the cosine similarity between the enhanced features and the original features is used to measure content consistency, and the loss is minimized accordingly. This approach preserves the semantic information during style augmentation, preventing the model from overfitting to style-related attributes.

First, both the enhanced features and the original features are normalized:

$$F_{aug}^{norm} = Normalize(F_{aug}) \tag{10}$$

Then, compute the cosine similarity matrix between the two sets of features.

$$Z = F_{aug}^{norm} (F_{org}^{norm})^{\bullet} \tag{11}$$

Extract the diagonal elements (these represent the similarity corresponding to each sample).

$$z_{tt} = diag(Z) \tag{12}$$

Finally, compute the content consistency loss:

$$L_{content} = -\frac{1}{B}\sum_{k=1}^{B}\left(z_{tt}^{k} - log\sum_{j=1}^{B} exp(z_{k,j})\right) \tag{13}$$

Minimizing $L_{content}$ ensures that augmented features remain content-consistent.

Theoretical Foundation. DSG effectively treats style variations in FAS tasks as noise, guiding the model through style enhancement and constrained training to learn feature representations that are insensitive to style changes yet sensitive to spoofing detection. This style invariance is the key factor in improving cross-domain generalization capability. Unlike traditional domain generalization methods, DSG does not rely on explicit domain labels but instead handles both intra-domain and inter-domain style variations implicitly through instance-level style enhancement. This fine-grained style manipulation allows the model to better adapt to style variations in unseen domains, significantly improving generalization performance across diverse testing conditions.

3.2 Supervised Contrastive Learning

Beyond style augmentation, a supervised contrastive learning strategy is employed to enhance the model's discriminative ability. Unlike traditional contrastive learning, it leverages labels to define positive and negative sample pairs, resulting in more compact intra-class and more dispersed inter-class distributions—an essential property given the complexity and diversity of attack samples in anti-spoofing tasks. Utilizing multi-view generation from style-enhanced features, multiple views per sample further improve the effectiveness of contrastive learning.

Concretely, for each mini-batch, multiple views (original and style-enhanced) are generated. Positive pairs include views from the same class and the sample's own augmented views, while negative pairs are views from different classes.

The loss function is defined as:

$$L_{SupCon} = \sum_{i=1}^{2B} \frac{-1}{|P(i)|} \sum_{p\in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a\in A(i)} \exp(z_i \cdot z_a / \tau)} \tag{14}$$

where 2B represents the total number of samples (including original features and augmented features), $\mathbf{z}_i$ denotes the normalized feature representation of the i-th sample, $P(i)$ represents the set of positive samples of the same class as sample i, $A(i)$ represents the set of all samples excluding sample i itself, and τ is the temperature coefficient. By minimizing L_{SupCon}, the model can learn more discriminative feature representations.

3.3 Classification Loss

To ensure robust anti-spoofing performance, an Attribute-Aware Multi-Head Classification Loss is introduced. Inspired by domain generalization, this approach uses classification supervision to directly address diverse attack types and distributions, enabling the model to learn subtle, discriminative features. Multiple fully connected layers are employed at the top of the model, each dedicated to a specific data subset. This multi-head structure supports fine-grained learning, capturing the distinct characteristics of different data distributions.

During forward propagation, samples in the batch are partitioned into subsets according to their dataset IDs. For each subset, the corresponding feature representation is extracted and passed through the respective classifier to generate prediction probabilities $\hat{y}_i$.

$$\hat{y}_i = \sigma(FC_i(feat_i) \times scale_i) \tag{15}$$

where σ represents the Sigmoid activation function, and $scale_i$ is the scaling factor used to adjust the output. This process ensures that each classifier focuses only on a specific type of sample, enabling the learning of more targeted discriminative features. For each subset, binary cross-entropy loss is calculated as follows:

$$L_{amc_i} = -\frac{1}{B}\sum_{k=1}^{B}[y_k log\hat{y}_k + (1-y_k)log(1-\hat{y}_k)] \tag{16}$$

where y_k represents the ground truth label of the k-th sample, and $\hat{y}_k$ denotes the probability predicted by the model. To balance the influence of each subset during overall training, the classification losses of all subsets are averaged to obtain the final attribute-aware multi-head classification loss:

$$L_{amc} = \frac{1}{K}\sum_{i=1}^{K} L_{amc_i} \tag{17}$$

where K represents the number of subsets, and in this model K=3. This averaging strategy ensures that each subset's classifier is adequately optimized during training, preventing insufficient training of any subset's classifier due to data imbalance.

The Attribute-Aware Multi-Head Classification Loss considers not only class labels but also the specific attributes of each sample. Combined with the style augmentation module and the contrastive learning module, this improved classification loss contributes to a complete training framework that fosters richer and more detailed feature representations.

3.4 Overall Loss Function

Combining all the losses, this paper's overall loss function is:

$$L_{total} = L_{amc} + \lambda_1 L_{SupCon} + \lambda_2 L_{diversity} + \lambda_3 L_{content} \tag{18}$$

where λ_1, λ_2, and λ_3 are hyperparameters balancing the influence of each loss component.

4 Experiments

4.1 Experimental Settings

Datasets and Protocol: This paper conducts evaluations on four widely used FAS datasets: Oulu-NPU (O) [52], CASIA-MFSD (C) [53], Idiap Replay-Attack (I) [54], and MSU-MFSD (M) [55]. Following standard practice, each dataset is considered a distinct domain, and a leave-one-out protocol is employed for cross-domain evaluation. For instance, O&C&I→M indicates that O, C, and I are used for training, while M serves as the testing domain. Similarly, O&M&I→C, O&C&M→I, and I&C&M→O represent the other leave-one-out configurations.

Implementation Details: Face detection and cropping are performed with MTCNN [56], and all images are resized to 256×256. For a fair comparison with state-of-the-art approaches, this paper adopts ResNet-18 as the backbone. Training is conducted using SGD with an initial learning rate of 5e-3 and a batch size of 96 per domain.

For the DSG model, the number of style bases is set to 64. The weight for the style diversity loss is 0.01, the weight for the content consistency loss is 0.1, and the weight for the supervised contrastive loss is also 0.1. All hyperparameters are tuned on a validation set.

Evaluation Metrics: This paper employs Half Total Error Rate (HTER) and Area Under the Curve (AUC) to evaluate performance. HTER, calculated as the average of the false acceptance rate and false rejection rate, assesses overall classification accuracy. AUC measures the model's discriminative ability across varying decision thresholds. By jointly considering HTER and AUC, this paper provides a comprehensive evaluation of both accuracy and robustness.

Table 1. Evaluated on four widely-used benchmark datasets: CASIA (C), Idiap Replay (I), MSU-MFSD (M), and Oulu-NPU (O).

Methods	O&C&M to I		O&C&I to M		O&M&I to C		I&C&M to O	
	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC
MMD-AAE[34]	31.58	75.18	27.08	83.19	44.59	58.29	40.98	63.08
MADDG[13]	22.19	84.99	17.69	88.06	24.50	84.51	27.98	80.02
NAS-FAS[37]	14.51	93.84	19.53	88.63	16.54	90.18	13.80	93.43
ANRL[40]	16.03	91.04	10.83	96.75	17.83	89.26	15.67	91.90
SSDG-R[12]	11.71	96.59	7.38	97.17	10.44	95.94	15.61	91.54
SSAN-R[14]	8.88	96.79	6.67	**98.75**	10.00	**96.67**	13.72	93.63
PatchNet[41]	13.40	95.67	7.10	98.46	11.33	94.58	11.82	95.07
SA-FAS[42]	6.58	97.54	**5.95**	96.55	8.78	95.37	10.00	96.23
TTDG[36]	6.50	97.98	7.91	96.83	**8.14**	96.49	10.00	95.70
DSTN(ours)	**5.00**	**98.91**	8.57	96.78	9.02	94.95	**9.76**	**96.56**

4.2 Cross-Domain Performance

Cross-domain generalization performance was evaluated under both leave-one-out (LOO) and limited-source-domain conditions to assess the proposed method's effectiveness on unseen domains.

Leave-One-Out (LOO): Under the LOO setting, training was conducted on three source domains, with the remaining dataset serving as the target domain. Methods for comparison were divided into two categories: traditional face anti-spoofing approaches and domain generalization-based face anti-spoofing approaches (DG-FAS). Table 1 presents the performance comparison across four LOO settings. Several observations emerge from Table 1. Traditional face anti-spoofing methods [7,17,18] perform poorly on all four cross-dataset benchmarks because they do not consider learning generalizable features, resulting in performance degradation on unseen domains. Although DG-FAS methods [12-16] improve generalization to some degree, they typically rely on manually defined domain labels for coarse-grained domain alignment. This approach does not guarantee that extracted features remain unaffected by domain-specific styles, thus limiting performance gains. In contrast, the proposed DSTN achieves the best results in two of the test settings. This indicates that the instance-level style augmentation strategy effectively learns style-robust features, significantly enhancing generalization on unseen domains.

Table 2. Comparison results on the constrained source domain.

Methods	M&I to C		M&I to O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)
MS_LBP[43]	51.16	52.09	43.63	58.07
SSDG-M[12]	31.89	71.29	36.01	66.88
ANRL[40]	31.06	72.12	30.73	74.10
SSAN-R[14]	30.00	76.20	29.44	76.62
EBDG[16]	27.97	75.84	25.94	78.28
SA-FAS[42]	25.51	81.41	20.74	86.32
AMEL[44]	24.52	82.12	19.68	87.01
IADG[45]	24.07	85.13	18.47	90.49
DSTN(ours)	**20.79**	**86.78**	**17.85**	**90.56**

Limited-Source Domains: To verify the approach's effectiveness with a small number of source domains, an additional experiment was conducted using only two source domains. Following previous work [45], the MSU-MFSD (M) and Idiap Replay-Attack (I) datasets were used as source domains, while the remaining two datasets, CASIA-MFSD (C) and Oulu-NPU (O), served as target domains for testing. Table 2 provides the results. Even in this highly restrictive scenario, the proposed method still outperforms current state-of-the-art methods. This further confirms the effectiveness of the instance-based domain generalization approach on unseen target domains. Unlike methods that enforce domain alignment across all source domains, instance-level style augmentation enables stronger generalization capability without mandating full domain adaptation.

Performance Variation Analysis: Performance differences (e.g., O&C&M→I: HTER=5.00% vs. O&M&I→C: HTER=9.02%) stem from dataset disparities in acquisition conditions, attack types, and style asymmetry. CASIA-MFSD (C) contains diverse low-quality attacks, while Idiap (I) focuses on high-quality replays. Larger style gaps between source-target domains (e.g., O&M&I→C) challenge DSG's coverage, whereas smaller gaps (O&C&M→I) align better with generated styles. Complex attack types in unseen domains (C) further degrade performance. Despite variations, DSG consistently outperforms existing methods, demonstrating effective multi-domain generalization. Future work will address extreme domain gaps via adaptive style selection.

In summary, the proposed method demonstrates remarkable performance in cross-domain face anti-spoofing tasks, validating the effectiveness of diversified style generation and instance-level alignment strategies.

4.3 Ablation Studies

This section first presents ablation studies to examine the contribution of each component to the model's performance, followed by an investigation of the impact of different contrastive learning losses. All ablation experiments are conducted under the O&C&M→I setting.

Table 3. Ablation study of each component on O&C&M→I.

Baseline	SupCon	Amc	DSG	HTER(%)	AUC(%)
✓	✓			13.95	92.35
✓	✓	✓		8.40	95.78
✓	✓	✓	✓	5.00	98.91

Table 3 shows the ablation results for each component. The baseline uses only the same backbone network (ResNet-18) as in [42] and the SupCon loss , achieving an HTER of 13.95% and an AUC of 92.35%. Incorporating the Amc loss improves performance to an HTER of 8.40% and an AUC of 96.78%. Adding the proposed DSG module further reduces the HTER to 5.00% and increases the AUC to 98.91%. These improvements indicate that each component complements the others, and using them together significantly enhances the model's performance.

Table 4. Ablation study of each style augmentation on O&C&M→I.

Style Augmentation	HTER(%)	AUC(%)
MixStyle[46]	9.00	95.72
SSA[14]	8.29	96.98
SHM[47]	7.18	96.34
CSA[45]	6.86	98.14
DSG(ours)	**5.00**	**98.91**

To validate the effectiveness of DSG, a comparison was made with different style augmentation strategies such as MixStyle and SSA. As shown in Table 4, DSG significantly outperforms other style augmentation methods. This can be attributed to DSG's ability to dynamically generate diverse styles, supported by the style diversity and content consistency losses, which guide the model toward more robust style-invariant representations.

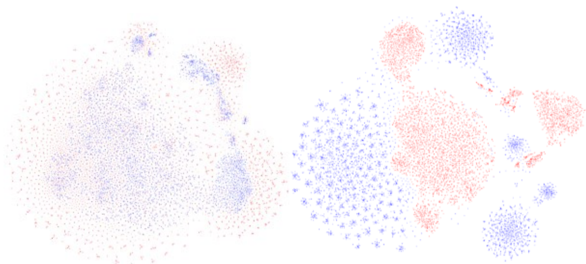
In addition, to confirm the effectiveness of the SupCon loss, a comparison was carried out with different contrastive learning losses, including SimCLR, SimSiam, and the triplet loss. Table 5 shows that while traditional contrastive losses or the triplet loss offer some improvements, they fall short of the performance achieved with SupCon. By leveraging class label information, SupCon constructs richer sets of positive and negative pairs within each batch, facilitating the learning of more discriminative features.

Table 5. Ablation study of each contrastive loss on O&C&M→I.

Style Augmentation	HTER(%)	AUC(%)
SimCLR[48]	12.53	92.42
SimSiam[49]	8.89	95.93
Triplet[50]	7.75	96.11
SupCon[51]	**5.00**	**98.91**

4.4 Visualization and Analysis

To gain deeper insight into the performance of the DSG method during style augmentation, t-SNE (t-distributed Stochastic Neighbor Embedding) was employed to visualize the style-enhanced features. By embedding high-dimensional features into a two-dimensional space, it becomes possible to observe the diversity in style base selection among different methods.



(a) Ours(DSG) (b) SSA

Fig. 2. T-SNE feature visualization of different styles of augmentation. The blue dots represent the original features, while the red ones represent the enhanced features. Compared to SSA [14], DSG generates more diverse features.

As shown in Figure 2, distributions of features generated by DSG appear more scattered and diverse in the two-dimensional space. This indicates that DSG can dynamically learn and select a wide range of style bases, including relatively rare ones. In contrast, other style augmentation methods produce more concentrated feature distributions, suggesting that their style bases are not sufficiently diverse and may be limited to certain dominant styles.

This enhanced diversity can be attributed to the learnable style bases introduced by the DSG approach. Through its dynamic weight generation mechanism, DSG tends to select style bases that differ from the current features, ensuring that the model encounters a broader range of style variations during training. Consequently, the model's generalization capability to unseen domains is improved.

In contrast, traditional style augmentation methods often rely on fixed or predefined style bases, lacking thorough exploration of style diversity and resulting in decreased performance when encountering rare or unseen styles.

5 Conclusion

This study proposes a new DSTN to enhance the generalization capability of face anti-spoofing models. DSG, a plug-and-play module within this framework, can be easily integrated into any deep learning architecture without substantial modifications. By introducing learnable style bases and a dynamic weight generation mechanism, DSG produces instance-level, diversified style-augmented features, thereby enriching feature diversity. This augmentation strategy exposes the model to a wider range of style variations during training, enabling the learning of style-invariant representations and ultimately improving performance on unseen domains.

Experimental results across multiple cross-domain face anti-spoofing benchmarks demonstrate that the proposed approach achieves superior results, confirming both the effectiveness and generalizability of DSG. Ablation studies further highlight the critical role of the DSG module in increasing feature diversity and enhancing model robustness.In essence, DSG provides an effective and flexible solution for improving feature diversity and generalization capability. Due to its modularity, DSG can be broadly applied to other computer vision tasks that require enhanced feature diversity, such as image classification, object detection, and semantic segmentation.

Future work will involve extending DSG to various tasks and models, examining its performance on larger-scale datasets and more complex architectures. Additionally, efforts will be made to integrate DSG with other data augmentation and regularization methods to further enhance model performance and robustness.

Acknowledgments. The author is deeply grateful to their supervisor for the valuable guidance, constructive feedback, and unwavering support, which were instrumental in the completion of this research.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: Arcface: Additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4690–4699. IEEE (2019)

2. Kemelmacher-Shlizerman, I., Seitz, S.M., Miller, D., Brossard, E.: The megaface benchmark: 1 million faces for recognition at scale. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4873–4882. IEEE (2016)
3. Li, S., Xu, J., Xu, X., Shen, P., Li, S., Hooi, B.: Spherical confidence learning for face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 15629–15637. IEEE (2021)
4. Wang, J., Liu, Y., Hu, Y., Shi, H., Mei, T.: Facex-zoo: A pytorch toolbox for face recognition. In: Proceedings of the 29th ACM International Conference on Multimedia (ACM MM), pp. 3779–3782. ACM (2021)
5. Boulkenafet, Z., Komulainen, J., Hadid, A., et al.: Face spoofing detection using colour texture analysis. IEEE Transactions on Information Forensics and Security (TIFS) 11(8), 1818–1830 (2017)
6. Pereira, T.F., Komulainen, J., Anjos, A., Martino, J.D., Hadid, A., Pietikäinen, M., Marcel, S.: Face liveness detection using dynamic texture. Eurasip Journal on Image and Video Processing (JIVP) 2014(1), 1–15 (2014)
7. Pereira, T.d.F., Anjos, A., Martino, J.M.D., Marcel, S.: LBP-top based countermeasure against face spoofing attacks. In: Asian Conference on Computer Vision (ACCV), pp. 121–132. Springer (2012)
8. Cai, R., Li, Z., Wan, R., Li, H., Hu, Y., Kot, A.C.: Learning meta pattern for face antispoofing. IEEE Transactions on Information Forensics and Security (TIFS) 17, 1201–1213 (2022)
9. Feng, L., Po, L.-M., Li, Y., Xu, X., Yuan, F., Cheung, T.C.-H., Cheung, K.-W.: Integration of image quality and motion cues for face antispoofing: A neural network approach. Journal of Visual Communication and Image Representation (JVCIR) 38, 451–460 (2016)
10. Liu, A., Tan, Z., Wan, J., Liang, Y., Lei, Z., Guo, G., Li, S.Z.: Face anti-spoofing via adversarial cross-modality translation. IEEE Transactions on Information Forensics and Security (TIFS) 16, 2759–2772 (2021)
11. Gu, Q., Zhou, Q., Xu, M., Feng, Z., Cheng, G., Lu, X., Shi, J., Ma, L.: Pit: Position-invariant transform for crossfov domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 8761–8770. IEEE (2021)
12. Jia, Y., Zhang, J., Shan, S., Chen, X.: Single-side domain generalization for face anti-spoofing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8484–8493. IEEE (2020)
13. Shao, R., Lan, X., Li, J., Yuen, P.C.: Multi-adversarial discriminative deep domain generalization for face presentation attack detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10023–10031. IEEE (2019)
14. Wang, Z., Wang, Z., Yu, Z., Deng, W., Li, J., Gao, T., Wang, Z.: Domain generalization via shuffled style assembly for face anti-spoofing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4123–4133. IEEE (2022)
15. Chen, Z., Yao, T., Sheng, K., Ding, S., Tai, Y., Li, J., Huang, F., Jin, X.: Generalizable representation learning for mixture domain face antispoofing. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), vol. 35, pp. 1132–1139. AAAI Press (2021)
16. Du, Z., Li, J., Zuo, L., Zhu, L., Lu, K.: Energy-based domain generalization for face anti-spoofing. In: Proceedings of the 30th ACM International Conference on Multimedia (ACM MM), pp. 1749–1757. ACM (2022)
17. Boulkenafet, Z., Komulainen, J., Hadid, A., et al.: Face anti-spoofing based on color texture analysis. In: IEEE International Conference on Image Processing (ICIP), pp. 2636–2640. IEEE (2015)

18. Komulainen, J., Hadid, A., Pietikäinen, M., et al.: Context based face anti-spoofing. In: IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–8. IEEE (2013)
19. Hu, C., Cao, J., Zhang, K.-Y., Yao, T., Ding, S., Ma, L.: Structure destruction and content combination for generalizable anti-spoofing. IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM) 4(4), 508–521 (2022)
20. Li, L., Feng, X., Boulkenafet, Z., Xia, Z., Li, M., Hadid, A.: An original face anti-spoofing approach using partial convolutional neural network. In: International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 1–6. IEEE (2016)
21. Atoum, Y., Liu, Y., Jourabloo, A., Liu, X.: Face anti-spoofing using patch and depth-based cnns. In: IEEE International Joint Conference on Biometrics (IJCB), pp. 319–328. IEEE (2017)
22. Kim, T., Kim, Y., Kim, I., Kim, D.: BASN: Enriching feature representation using bipartite auxiliary supervisions for face anti-spoofing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pp. 494–503. IEEE (2019)
23. Jourabloo, A., Liu, Y., Liu, X.: Face despoofing: Anti-spoofing via noise modeling. In: European Conference on Computer Vision (ECCV), pp. 290–306. Springer (2018)
24. Liu, A., Tan, Z., Wan, J., Liang, Y., Lei, Z., Guo, G., Li, S.Z.: Face anti-spoofing via adversarial cross-modality translation. IEEE Transactions on Information Forensics and Security (TIFS) 16, 2759–2772 (2021)
25. Liu, Y., Liu, X.: Spoof trace disentanglement for generic face anti-spoofing. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 45(3), 3813–3830 (2022)
26. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR) (2020)
27. Mehta, S., Rastegari, M.: Mobilevit: Lightweight, general-purpose, and mobile-friendly vision transformer. arXiv preprint arXiv:2110.02178 (2021)
28. Feng, Z., Zhou, Q., Gu, Q., Tan, X., Cheng, G., Lu, X., Shi, J., Ma, L.: DMT: Dynamic mutual training for semi-supervised learning. Pattern Recognition (PR) 108777 (2022)
29. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International Conference on Machine Learning (ICML), pp. 1180–1189. PMLR (2015)
30. Gu, Q., Zhou, Q., Xu, M., Feng, Z., Cheng, G., Lu, X., Shi, J., Ma, L.: Pit: Position-invariant transform for crossfov domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 8761–8770. IEEE (2021)
31. Zhou, Q., Zhang, K.-Y., Yao, T., Lu, X., Yi, R., Ding, S., Ma, L.: Instance-aware domain generalization for face anti-spoofing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20453–20463. IEEE (2023)
32. Gatys, L.A., Ecker, A.S., Bethge, M., et al.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2414–2423. IEEE (2016)
33. Huang, X., Belongie, S., et al.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1501–1510. IEEE (2017)
34. Li, H., Pan, S.J., Wang, S., Kot, A.C.: Domain generalization with adversarial feature learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5400–5409. IEEE (2018)

35. Wang, G., Han, H., Shan, S., Chen, X.: 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, pp. 6678–6687. IEEE (2020)
36. Shao, R., Lan, X., Yuen, P.C.: Regularized fine-grained meta face anti-spoofing. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 11974–11981. AAAI Press (2020)
37. Yu, Z., Wan, J., Qin, Y., Li, X., Li, S.Z., Zhao, G.: 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)
38. Wang, J., Zhang, J., Bian, Y., Cai, Y., Wang, C., Pu, S.: Self-domain adaptation for face anti-spoofing. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 2746–2754. AAAI Press (2021)
39. Liu, S., Zhang, K.-Y., Yao, T., Sheng, K., Ding, S., Tai, Y., Li, J., Xie, Y., Ma, L.: Dual reweighting domain generalization for face presentation attack detection. arXiv preprint arXiv:2106.16128 (2021)
40. Liu, S., Zhang, K.-Y., Yao, T., Bi, M., Ding, S., Li, J., Huang, F., Ma, L.: Adaptive normalized representation learning for generalizable face anti-spoofing. In: Proceedings of the 29th ACM International Conference on Multimedia, pp. 1469–1477. ACM (2021)
41. Wang, C.-Y., Lu, Y.-D., Yang, S.-T., Lai, S.H.: PatchNet: A simple face anti-spoofing framework via fine-grained patch recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20281–20290. IEEE (2022)
42. Sun, Y., Liu, Y., Liu, X., Li, Y., Chu, W.S.: Rethinking domain generalization for face anti-spoofing: Separability and alignment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 24563–24574. IEEE (2023)
43. Määttä, J., Hadid, A., Pietikäinen, M., et al.: Face spoofing detection from single images using micro-texture analysis. In: Proceedings of the IEEE International Joint Conference on Biometrics (IJCB), pp. 1–7. IEEE (2011)
44. Zhou, Q., Zhang, K.-Y., Yao, T., Yi, R., Ding, S., Ma, L.: Adaptive mixture of experts learning for generalizable face anti-spoofing. In: Proceedings of the 30th ACM International Conference on Multimedia (ACM MM), pp. 6009–6018. ACM (2022)
45. Zhou, Q., Zhang, K.-Y., Yao, T., Lu, X., Yi, R., Ding, S., Ma, L.: Instance-aware domain generalization for face anti-spoofing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20453–20463. IEEE (2023)
46. Zhou, K., Yang, Y., Qiao, Y., Xiang, T.: Domain generalization with mixstyle. In: International Conference on Learning Representations (ICLR) (2021)
47. Zhao, Y., Zhong, Z., Zhao, N., Sebe, N., Lee, G.H.: Style-hallucinated dual consistency learning for domain generalized semantic segmentation. In: European Conference on Computer Vision (ECCV), Part XXVIII, pp. 535–552. Springer (2022)
48. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: Proceedings of the International Conference on Machine Learning, pp. 1597–1607. PMLR (2020)
49. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15750–15758. IEEE (2021)
50. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823. IEEE (2015)

51. Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. In: Advances in Neural Information Processing Systems, vol. 33, pp. 18661–18673. Curran Associates, Inc. (2020)
52. Boulkenafet, Z., Komulainen, J., Li, L., Feng, X., Hadid, A.: Oulu-npu: A mobile face presentation attack database with real-world variations. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 612–618. IEEE (2017)
53. Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., Li, S.Z.: A face antispoofing database with diverse attacks. In: 2012 5th IAPR International Conference on Biometrics (ICB), pp. 26–31. IEEE (2012)
54. Chingovska, I., Anjos, A., Marcel, S.: On the effectiveness of local binary patterns in face antispoofing. In: 2012 BIOSIG-Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG), pp. 1–7. IEEE (2012)
55. Wen, D., Han, H., Jain, A.K.: Face spoof detection with image distortion analysis. IEEE Transactions on Information Forensics and Security 10(4), 746–761 (2015)
56. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters 23(10), 1499–1503 (2016)