Snapshot Spectral Imaging for Face Anti-Spoofing: Addressing Data Challenges with Advanced Processing and Training

Hui Li, Yaowen Xu*, Zhaofan Zou, Zhixiang He
China Telecom Artificial Intelligence Technology Co. Ltd, Beijing, China
{lih, xuyw1, zouzhf41, hezx3}@chinatelecom.cn

Abstract

Although considerable research progress has been made in the field of face anti-spoofing(FAS), it still faces continuous threats from ultra-realistic face mask attacks. Facing the challenge of highly realistic flexible masks, spectral sensors show great potential in enhancing the safety of FAS systems. However, the application of snapshot spectral imaging (SSI) in FAS is still in its infancy and faces two major challenges: data scarcity and data content differences. To this end, we introduce a data processing and model training method for SSI images. In terms of data processing, we use RandomBorderMask technology and RandomDropChannels strategy to avoid misjudgment of material information, reduce interference from redundant information, and learn from RGB image preprocessing methods to enhance data diversity. Regarding model training, we propose a model integration strategy and semi-supervised learning technology, which combines the prediction results of multiple models and uses pseudo labels to expand the training data to solve the over-fitting problem caused by data scarcity. These innovative methods achieved first-class results in the 5th Face Anti-Spoofing Challenge @CVPR2024, verifying their effectiveness in improving the accuracy and robustness of the FAS system.

1. Introduction

The Face Anti-Spoofing (FAS) system is extensively employed in face recognition systems to safeguard them from the vulnerabilities posed by presentation attacks, including video attacks, print attacks, and 3D masks. Acknowledging the pivotal role of FAS in enhancing security, both academic institutions and industry leaders have conducted extensive research, resulting in remarkable progress in this field[32, 34].

Confronted with the challenge of highly realistic flexible masks crafted from materials like silicone or latex, the integration of advanced spectroscopy sensors into facial recognition systems can markedly improve their ability to detect and distinguish these types of deceptive appearances[20]. The advantages of hyperspectral imaging are particularly noteworthy in this context. Snapshot Spectral Imaging (SSI)[17], an advanced imaging technique, excels at capturing high-dimensional spectral information of a scene in a single exposure. This technique provides a wealth of detailed information, enabling the system to differentiate between authentic faces and deceptive disguises more accurately. However, despite its potential, the application of FAS within SSI remains largely unexplored. Further investigation is necessary to fully harness the capabilities of this cutting-edge technology and ensure robust security in facial recognition systems[30].

The application of hyperspectral images in FAS confronts numerous obstacles[37]. Traditional hyperspectral cameras heavily rely on optics grating and mechanical scanning systems, rendering them prohibitively expensive, bulky, and cumbersome in capturing a single hyperspectral image. Consequently, their practicality in real-world scenarios remains limited. Although researchers have developed the fast computational speed and high reconstruction fidelity make it practical in real-time on-chip hyperspectral imaging systems, such as SSI[21]. However, the SSI data suitable for FAS remains scant. This poses the initial challenge of achieving high performance when working with a limited amount of SSI training data.

The second challenge in the application of SSI images in FAS is the difference in data content. SSI images exhibit significant differences from the widely used RGB image data content. SSI images typically have a higher number of channels but lower resolution, which is the opposite of RGB images. Additionally, SSI images primarily emphasize distinctions in material properties, whereas RGB images reflect the color appearance characteristics of the real world. Researchers must devise innovative data processing techniques that go beyond traditional methods to adapt to these differences. In order to obtain higher accuracy performance from SSI images compared to RGB images, researchers must also improve their training methods and design specialized FAS algorithm models. This involves exploring novel techniques that effectively leverage the spectral information contained

*Corresponding author

Figure 1. SSI examples from HySpeFAS dataset. The first row is the fake face, and the second row is the real face.

in SSI images for enhanced FAS applications.

To delve deeper into spectral face anti-spoofing algorithms that are compatible with SSI images and to fuel research efforts in novel spectroscopic sensor-based face anti-spoofing techniques, a surveillance FAS dataset has been meticulously assembled by Shijie Rao. Dubbed HySpeFAS, this dataset was employed in the 5th Face Anti-spoofing Challenge@CVPR2024. It comprises a comprehensive collection of 6,760 hyperspectral images, which were expertly reconstructed from SSI images utilizing the TwIST algorithm[3]. Each hyperspectral image boasts 30 spectral channels, offering a rich and diverse spectral representation. Figure 1 presents a selection of illustrative samples, providing a glimpse into the dataset's breadth and depth.

Inspired by the above discussion, we introduce a data processing and model training method for SSI images to solve these two problems: data scarcity and data discrepancy. In terms of data processing, we use the unique RandomBorderMask technology to avoid the model's misjudgment of materials such as medical masks. At the same time, we use RandomDropChannels to discard channels and reduce redundant information interference to enhance data diversity. In addition, we also draw on the preprocessing method of RGB images to improve data diversity through various enhancement techniques. In terms of model training, we propose a model ensemble strategy to combine the prediction results of multiple models trained with different augmentations to improve accuracy and robustness. At the same time, we explore semi-supervised learning technology and use model predictions as pseudo-labels to expand the amount of training data, effectively solving the over-fitting problem caused by the scarcity of SSI data. In summary, the main contributions of this paper are summarized as follows:

- Based on the characteristics of SSI images, we proposed a unique data processing method, which effectively avoids misjudgment of material information, reduces redundant information interference, and enhances data diversity.
- We innovatively propose two training methods for SSI images, model ensemble learning, and semi-supervised learning, both of which effectively avoid model overfitting.
- The proposed method won the first place in the "Snapshot Spectral Imaging Face Anti-spoofing Challenge" of the 5th Face Anti-spoofing Challenge@CVPR2024.

2. Related work

FAS methods for RGB images: After several years of intense research and development, the domain of FAS has witnessed remarkable advancements with RGB images. In the initial stages of FAS exploration on RGB images, a plethora of traditional approaches, relying on manual feature extraction were introduced. These traditional methods, such as LBP[5], SIFT[18], SURF[4], and HOG[11], primarily aimed at extracting pertinent spoofing patterns from diverse color spaces. This endeavor often necessitated specific task-related prior knowledge. However, with the swift evolution of deep learning, techniques rooted in Deep Convolutional Neural Networks (CNN) have gradually emerged as the preferred choice for tackling FAS challenges on RGB images. CNNs excel at learning discriminative high-level features from vast datasets, encompassing both color and texture attributes, as well as intricate features like facial contours and edges. Their effectiveness on FAS tasks on RGB images has been unequivocally demonstrated. Prominent CNN-based methodologies include residual learning frameworks[7], and centered differential convolution[33], among others, which have significantly contributed to the field's progress.

Despite the breakthroughs achieved by FAS due to deep learning, it still faces the challenge of domain generalization, such as domain adaptation[12] and domain generalization[10, 24, 31] issues across different scenarios and attack types. Recently, some methods[16, 25, 29] have carefully designed frameworks and loss functions to learn discriminative features and generalized feature spaces. However, these domain generalization-based methods rely excessively on domain labels, which may not accurately reflect the true domain distribution. To address this, Zhou et al.[38] proposed aligning features at the instance level without domain labels, aiming to reduce the sensitivity of features to specific instance styles. Despite progress in domain generalization, a persistent challenge in the RGB FAS field is the increasing realism and decreasing cost of attack props. Many realistic masks and genuine human images cannot be distinguished by models using RGB images alone.

FAS methods for spectral images: With the rapid development of 3D printing technology and bionic silicone rubber technology, the lifelikeness of 3D masks is even sufficient to deceive the human visual perception system[14]. Such highly realistic masks undoubtedly pose significant challenges to the current FAS systems[2, 9]. However, the widely used FAS methods based on additional sensors such as RGBD and NIR often perform poorly in cross-dataset evaluations[22]. This domain adaptation issue leads to difficulties in ensuring the performance of FAS systems in

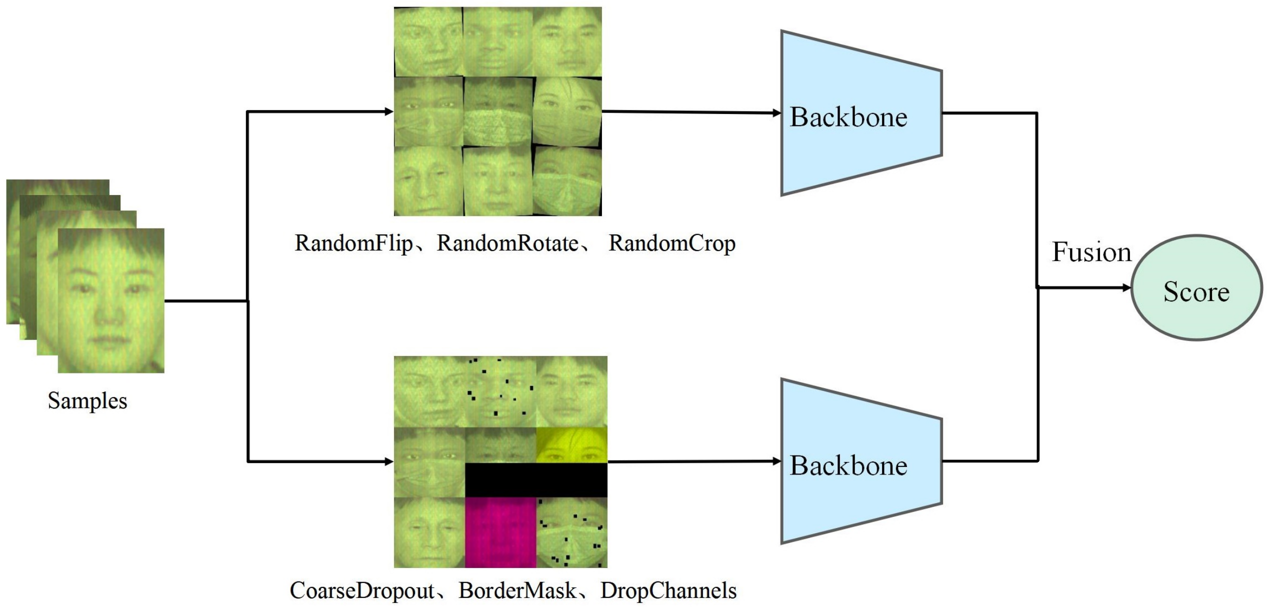


Figure 2. The framework of the proposed method. Each SSI image undergoes two different sets of augmentations and is trained by two branch models, and the outputs are finally integrated for classification.

practical deployment, casting serious doubts on their stability and reliability. To address this issue, researchers have attempted to adopt various advanced sensor technologies, including SWIR[8], thermal cameras[23], light field cameras[13], and polarization cameras[27], to capture more subtle and unique features of genuine and fake faces. While these techniques have improved the recognition capabilities of FAS to a certain extent, their disadvantages are also apparent: high price, bulky size, and inconvenience in operation. These factors hinder their integration into real-world face recognition systems[34].

Spectral analysis, as an effective means of identifying different materials, provides us with a new avenue. Studies have shown that due to the specific absorption effect of hemoglobin in human blood, the reflection spectrum of human skin exhibits two distinct minima at 545nm and 575nm [20]. HSI-based FAS methods leverage this characteristic to distinguish between genuine and fake faces by capturing these subtle spectral differences, often resulting in more reliable and robust performance compared to methods based on RGB cameras[1, 15]. However, traditional hyperspectral cameras, which rely on optical gratings and mechanical scanning systems, are often expensive, bulky, and require a considerable amount of time to capture an HSI. This limits their practical applications in the real world, leading to a scarcity of public FAS datasets based on HSIs[36].

In recent years, with the rapid development of on-chip spectral imaging sensor technology, the utilization of silicon metasurfaces and computational imaging techniques[28, 30] has enabled the realization of on-chip snapshot HSI sensors. These sensors can capture HSI data at video rates, providing the potential for introducing hyperspectral perception into daily life. However, there is still a relative dearth of research on the application of HSI data in FAS, leaving ample room for further exploration in this field.

3. Methodology

In this section, we first introduce an overview of our method in Section 3.1. Then, we elaborate on the proposed data preprocessing method and model training suitable for SSI images in Section 3.2 and Section 3.3 respectively.

3.1. Overview method

In our approach, we recognize the importance of data preprocessing and data augmentation in dealing with the unique characteristics of the HySpeFAS dataset. Furthermore, we also realize the importance of model training methods for scarce data tasks. We utilize two training strategies, model ensemble, and semi-supervised learning, to achieve the best performance on the test set. Figure 2 shows the framework of our approach.

3.2. Data processing method for SSI images

There exist disparities in channels, resolutions, and image content between SSI and RGB images, necessitating a unique approach for FAS. Directly applying the preprocessing methods tailored for RGB images is inadequate. The

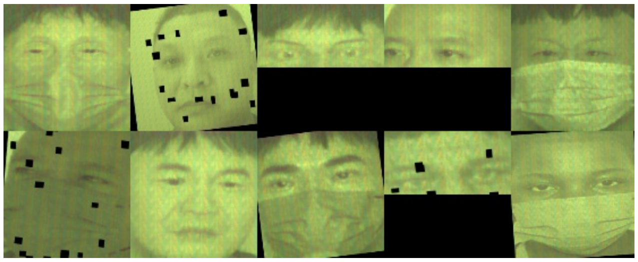


Figure 3. SSI examples after data preprocessing. All images are randomly applied with transformations: flipping, cropping, rotation, CoarseDropout, RandomBorderMask, RandomDropChannels.

key advantage of SSI images lies in their heightened sensitivity to diverse materials. While there are conspicuous disparities in the materials of authentic faces and 3D face masks, the introduction of medical masks worn by real individuals can obfuscate the distinction between real faces and face masks. Consequently, to extract reliable material information from SSI images, we employ the **RandomBorderMask** technique to occlude the lower portion of the face, thereby preventing the model from erroneously interpreting mask material.

We attempted to transform the 30-channel SSI image into a 3-channel image utilizing the averaging technique and discovered that both methods exhibited comparable classification accuracy in the FAS task. This observation suggests that SSI channels contain redundant information. To prevent the model from unduly focusing on a specific layer of channels, we employed the **RandomDropChannels** method to randomly discard some channels. This strategy enhances data diversity and mitigates the risk of data loss or corruption.

In addition to the aforementioned operations, we also draw inspiration from the preprocessing methods utilized for RGB images to meticulously preprocess our data, thereby ensuring optimal model performance. Precisely, we normalize the input data to guarantee that each feature possesses a comparable scale, preventing any specific feature from dominating the learning process due to its larger magnitude. Furthermore, we adopt diverse augmentation techniques tailored to the unique characteristics of our data, including random flipping, random cropping, and random rotation. Additionally, we introduce a technique called **CoarseDropout**, which randomly masks larger areas of the input image to simulate occlusion or dropout scenarios, further enriching the data diversity. Figure 3 is examples after data preprocessing

3.3. Model training method for SSI images

Due to the limited availability of SSI data, with existing datasets containing merely a few thousand samples, model training is highly susceptible to overfitting. To address this issue arising from the scarcity of SSI data, we employed two innovative training strategies, both achieving a perfect score of 0 ACER of Challenge@CVPR2024.

1) Model Integration. Instead of relying solely on a single model, we combined the predictions of multiple models to enhance accuracy and robustness. Each model was trained using a distinct data augmentation technique. During the prediction phase, we averaged the outputs of these models, capitalizing on the diversity of their predictions.

2) Semi-supervised Learning Technique. We initially trained the model using labeled data and augmented the dataset with various techniques to broaden its diversity. Once the model was trained, we employed it to predict labels for the test dataset. These predicted labels were subsequently treated as ground truth labels, and the model was further trained using this expanded labeled dataset. By leveraging the model's predictions as pseudo-labels, we effectively increased the amount of labeled data available for training, ultimately enhancing performance.

4. Experiments

4.1. Experimental Settings

HySpeFAS datase. The HySpeFAS dataset consists of a total of 6,760 images, which are divided into three parts: training, validation, and testing sets. The training set comprises 3,900 images, and 936 images for validation, and 1,924 images for the testing set, which serves as the ultimate test to evaluate the performance of the algorithms. Each hyperspectral image in the dataset includes 30 spectral channels, and each image provides detailed information across a broad range of wavelengths.

Evaluation metrics. We adopt established standards within the field of FAS for evaluating the performance in the competition, including Attack Presentation Classification Error Rate (APCER), Bona Fide Presentation Classification Error Rate (BPCER), and Average Classification Error Rate (ACER) as the evaluation metric, in which APCER and BPCER are used to measure the error rate of fake or live samples, respectively. They can be formulated as:

$$\begin{aligned} APCER &= \frac{FP}{FP+TN}, \\ BPCER &= \frac{FN}{FN+TP}, \\ ACER &= \frac{APCER+BPCER}{2}, \end{aligned} \tag{1}$$

where FP, FN, TN, and TP denote the counts of false positive, false negative, true negative, and true positive instances, respectively. ACER is used to determine the final ranking in Snapshot Spectral Imaging Face Anti-spoofing Challenge@CVPR2024.

Data preprocess. Since the dataset has already been cropped to focus on the face area, we directly load the dataset without the need for additional cropping. Additionally, we normalize the data to ensure all features or channels have a similar scale as each channel represents a specific spectral band, and we rescale the values of each channel to a range between 0 and 1 independently.

Architecture details. We use resnest14[35] which has a shallower version of the popular resnest architecture since the training set is relatively small for this task to mitigate overfiting and reduce the risk of poor generalization.

Training details. Our proposed method is implemented with Pytorch. In the training phase, we use the stochastic gradient descent (SGD) optimizer with a momentum of 0.9, and the initial learning rate is $1e^{-2}$, and employ a cosine learning rate schedule to adjust the learning rate during training. We resized the images to a size of 112×112 based on the observation that the mean dimensions of the images are close to 112 pixels in both width and height. Various data augmentation techniques are applied to expand training data, including RandomCrop, RandomFlip, RandomRotate, RandomCoarseDropout, RandomBorderMask, and RandomDropChannels. For RandomCoarseDropout, we set the probability is 0.1, the size of the erasure area ranges from 8×8 to 32×32 pixels, and the number of areas varies between 8 to 16. For RandomBorderMask, we apply masking with a probability of 0.3, where the mask's range varies between 0.5 and 1. For RandomDropChannels, we drop out 10% of channels with a probability of 0.1.

We choose the Cross Entropy loss function which is commonly used for classification tasks. We employ balancing strategies to handle sample imbalance in the training set. The model is trained on two V100 GPUs for 50 epochs with a batch size of 80.

Semi-supervised and model ensemble learning details. For the semi-supervised learning method, after training a model with various data augmentation techniques, then 100% of predictions are incorporated into the training set to refine the model, and we select the final model from the last epoch of the iterative process. For the model ensemble method, we train two models with different data augmentation techniques, we merge the two models with a weight ratio of 1:1 as the final output.

4.2. Comparison with State-of-the-art Methods

In this section, we compare the performance of the proposed method with other teams. Table 1 summarizes the comparison results of three metrics: APCER, BPCER, and ACER. Our method achieves the highest performance on all metrics when employing semi-supervised learning and model ensemble methods, and the result demonstrates the effectiveness of our method for face anti-spoofing on the HySpeFAS dataset.

Team	APCER(%) ↓	BPCER(%) ↓	ACER(%) ↓
ctyun-ai	0.310	0.641	0.475
Ricardozzf	0.124	0.320	0.031
hexianhua	0.372	**0**	0.186
SeaRecluse	0.062	**0**	0.031
Ours(semi-supervise)	**0**	**0**	**0**
Ours(model ensemble)	**0**	**0**	**0**

Table 1. Comparing results with other teams on the test set of the HySpeFAS Dataset. Our method achieves 0 on APCER(%), BPCER(%), and ACER (%) metrics when using semi-supervised learning and model ensemble methods.

4.3. Ablation Study

In this section, ablation studies are conducted to demonstrate the importance of the choice of data augmentations, backbone, and input size. We implement a series of testing sets under different settings and detailed performance is provided under the metrics of APCER, BPCER, and ACER.

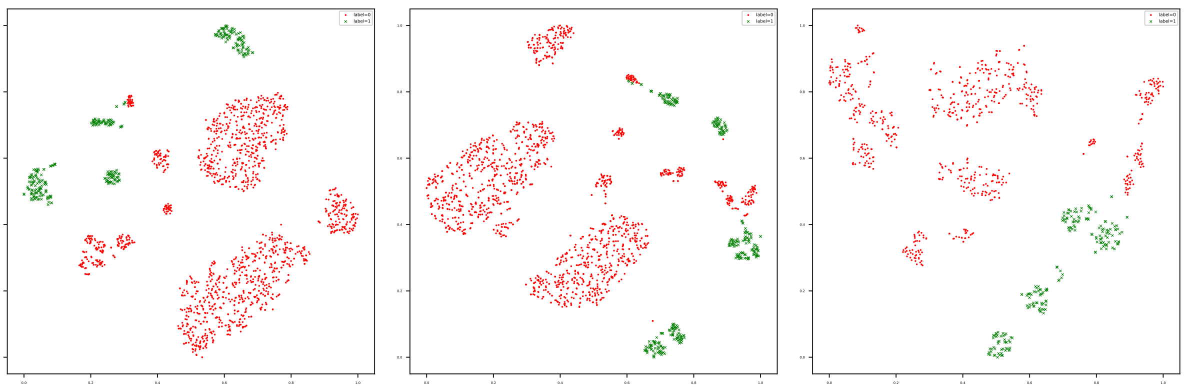
Data Augmentation We explore the performance of using different data augmentation strategies. Table 2 reports detailed results. We observe that incorporating data augmentation techniques such as RandomFlip, RandomCrop, and RandomRotate leads to improvements in ACER, APCER, and BPCER metrics to 0.962%, 1.786%, and 1.374%, respectively. Furthermore, from our experiments, we found that the combination of CoarseDropout, RandomBorderMask, and RandomDropChannels has resulted in the best performance, although these individual enhancements may not have shown improvements when used separately. These performance improvements indicate the importance of utilizing suitable data augmentation strategies on SSI images.

Backbone Selection We compare the performance of resnest which are variants of the resnest network architecture, differentiated based on the depth and the number of layers they possess. As presented in Table 3, using resnest14 achieved best ACER(1.065%), best APCER(0.687%) and best BPCER (1.442%). These results indicate that selecting an appropriate network improves the performance of our method.

We analyze the performance using different model structures and sizes including inceptionv4[26], regnety32[19], resnet18[7], and resnest14. According to the performance presented in table 4, resnest14 achieves the best performance in terms of ACER and APCER metrics with values of 0.687% and 1.065%, respectively. The experimental results suggest

Exp.	Flip	Crop	Rotate	CoarseDropout	BorderMask	DropChannels	APCER(%) ↓	BPCER(%) ↓	ACER(%) ↓
1	✗	✗	✗	✗	✗	✗	6.731	4.808	5.769
2	✓	✗	✗	✗	✗	✗	4.533	4.808	4.670
3	✓	✓	✗	✗	✗	✗	0.824	1.923	1.374
4	✓	✓	✓	✗	✗	✗	1.786	0.962	1.374
5	✓	✓	✓	✓	✗	✗	1.923	0.962	1.442
6	✓	✓	✓	✓	✓	✗	2.610	0.481	1.545
7	✓	✓	✓	✓	✓	✓	**0.687**	1.442	**1.065**
8	✓	✓	✓	✗	✗	✓	5.907	2.404	4.155
9	✓	✓	✓	✓	✗	✓	7.005	**0**	3.503
10	✓	✓	✓	✗	✓	✓	9.066	**0**	4.533

Table 2. The result of different data augmentation strategies on model performance.



(a) visualization of Exp.6 in Table 2 (b) visualization of Exp.9 in Table 2 (c) visualization of Exp.7 in Table 2

Figure 4. Visualization of different data augmentation.

backbone	APCER(%) ↓	BPCER(%) ↓	ACER(%) ↓
resnest50	3.709	2.404	3.056
resnest26	3.846	1.923	2.885
resnest14	**0.687**	**1.442**	**1.065**

Table 3. The performance on different depths of model.

backbone	Macs	APCER(%) ↓	BPCER(%) ↓	ACER(%) ↓
inceptionv4	1.66G	12.225	**0.481**	6.353
regnety32	0.84G	14.835	4.327	9.581
resnet18	0.75G	7.967	0.962	4.464
resnest14	0.73G	**0.687**	1.442	**1.065**

Table 4. The result of different model structures.

backbone	input size	APCER(%) ↓	BPCER(%) ↓	ACER(%) ↓
resnest14	224×224	3.709	3.846	3.777
resnest14	64×64	7.967	2.885	5.426
resnest14	112×112	**0.687**	**1.442**	**1.065**

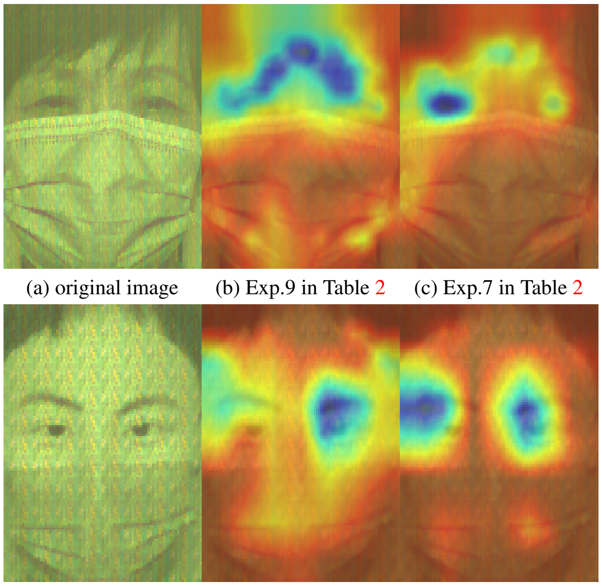
Table 5. The result of different input sizes.

that the structure of the model we choose is more suitable for the HySpeFAS dataset.

Impact of image size We explore the impact of different input image sizes on the model's performance by testing the model with dimensions of 224×224, 112×112, and 64×64. From the results presented in Table 5, it is observed that the use of 112×112 dimensions achieve the best performance and achieve metrics of 0.687% on APCER, 1.442% on APCER, and 1.065% on ACER. The experimental findings indeed underscore the advantage of utilizing input image sizes comparable to the original dimensions employed during the model's training phase, significantly enhancing the overall performance of our proposed method.

4.4. Visualizations

Impact of data augmentation We employ distributed Stochastic Neighbor Embedding (t-SNE) to visualize the model's performance after applying different data augmentation techniques. As shown in Figure 4, combining different data augmentation separates positive and negative samples more effectively. We use GradCam[6], a gradient localization-based visual interpretation technique to observe the effectiveness of BorderMask. As shown in Figure 5, our model focuses on the relevant characteristics and ignores irrelevant or redundant information to achieve good performance.



(a) original image (b) Exp.9 in Table 2 (c) Exp.7 in Table 2

(d) original image (e) Exp.9 in Table 2 (f) Exp.7 in Table 2

Figure 5. GradCam of the effectiveness on BorderMask augmentation.

5. Conclusion

In this paper, we deeply explore the application of SSI images in the field of FAS and propose a series of innovative solutions to address the two major challenges of data scarcity and data content differences faced in this field. In terms of data processing, we adopt the RandomBorderMask technology and RandomDropChannels strategy, which effectively avoids misjudgment of material information, reduces the interference of redundant information, and significantly enhances the diversity of data. In terms of model training, we proposed two methods, model ensemble learning, and semi-supervised learning, which not only improved the accuracy and robustness of the model but also successfully solved the over-fitting problem caused by data scarcity. These innovative methods have achieved remarkable results in practical applications. Our method won first place in the ”Snapshot Spectral Imaging Face Anti-spoofing Challenge” of the 5th Face Anti-spoofing Challenge@CVPR2024. In summary, our method provides important theoretical support and practical guidance for the application of SSI images in the field of FAS.

References

[1] Mohsen Ardabilian, Abdel-Malek Zine, and Shiwei Li. Multi-, hyper-spectral biometrics modalities. *Hidden Biometrics: When Biometric Security Meets Biomedical Engineering*, pages 127–153, 2020. 3

[2] Sushil Bhattacharjee, Amir Mohammadi, and Sébastien Marcel. Spoofing deep face recognition with custom silicone masks. In *2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS)*, pages 1–7. IEEE, 2018. 2

[3] José M Bioucas-Dias and Mário AT Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image processing*, 16 (12):2992–3004, 2007. 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. 2

[5] 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. 2

[6] Jacob Gildenblat and contributors. Pytorch library for cam methods. `https://github.com/jacobgil/pytorch-grad-cam`, 2021. 6

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2, 5

[8] Guillaume Heusch, Anjith George, David Geissbühler, Zohreh Mostaani, and Sébastien Marcel. Deep models and shortwave infrared information to detect face presentation attacks. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(4):399–409, 2020. 3

[9] Shan Jia, Guodong Guo, and Zhengquan Xu. A survey on 3d mask presentation attack detection and countermeasures. *Pattern recognition*, 98:107032, 2020. 2

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

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

[12] Yunsheng Li, Lu Yuan, Yinpeng Chen, Pei Wang, and Nuno Vasconcelos. Dynamic transfer for multi-source domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10998–11007, 2021. 2

[13] Mengyang Liu, Hong Fu, Ying Wei, Yasar Abbas Ur Rehman, Lai-man Po, and Wai Lun Lo. Light field-based face liveness detection with convolutional neural networks. *Journal of Electronic Imaging*, 28(1):013003–013003, 2019. 3

[14] Siqi Liu, Baoyao Yang, Pong C Yuen, and Guoying Zhao. A 3d mask face anti-spoofing database with real world variations. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 100–106, 2016. 2

[15] Yifei Liu, Min Zheng, and Qingyong Li. Face liveness verification based on hyperspectrum analysis. In *2017 31st international conference on advanced information networking*

and applications workshops (WAINA), pages 612–615. IEEE, 2017. 3

[16] Yongluo Liu, Yaowen Xu, Zhaofan Zou, Zhuming Wang, Bowen Zhang, Lifang Wu, Zhizhi Guo, and Zhixiang He. Adversarial domain generalization for surveillance face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 6352–6360, 2023. 2

[17] Andrew McClung, Sarath Samudrala, Mahsa Torfeh, Mahdad Mansouree, and Amir Arbabi. Snapshot spectral imaging with parallel metasystems. *Science advances*, 6(38):eabc7646, 2020. 1

[18] 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. 2

[19] Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, and Piotr Dollár. Designing network design spaces. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10428–10436, 2020. 5

[20] Shijie Rao, Yidong Huang, Kaiyu Cui, and Yali Li. Anti-spoofing face recognition using a metasurface-based snapshot hyperspectral image sensor. *Optica*, 9(11):1253–1259, 2022. 1, 3

[21] Shijie Rao, Kaiyu Cui, Yidong Huang, Jiawei Yang, Sheng Xu, and Yali Li. Fast deep learning reconstruction algorithm for on-chip snapshot hyperspectral imaging. In *2023 Opto-Electronics and Communications Conference (OECC)*, pages 1–3, 2023. 1

[22] Arian Sabaghi, Marzieh Oghbaie, Kooshan Hashemifard, and Mohammad Akbari. Deep learning meets liveness detection: recent advancements and challenges. *arXiv preprint arXiv:2112.14796*, 2021. 2

[23] Jongwoo Seo and In-Jeong Chung. Face liveness detection using thermal face-cnn with external knowledge. *Symmetry*, 11(3):360, 2019. 3

[24] 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. 2

[25] 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

[26] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI conference on artificial intelligence*, 2017. 5

[27] Yu Tian, Kunbo Zhang, Leyuan Wang, and Zhenan Sun. Face anti-spoofing by learning polarization cues in a real-world scenario. In *Proceedings of the 4th International Conference on Advances in Image Processing*, pages 129–137, 2020. 3

[28] Zhu Wang, Soongyu Yi, Ang Chen, Ming Zhou, Ting Shan Luk, Anthony James, John Nogan, Willard Ross, Graham Joe, Alireza Shahsafi, et al. Single-shot on-chip spectral sensors based on photonic crystal slabs. *Nature communications*, 10 (1):1020, 2019. 3

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

[30] Jian Xiong, Xusheng Cai, Kaiyu Cui, Yidong Huang, Jiawei Yang, Hongbo Zhu, Wenzheng Li, Bo Hong, Shijie Rao, Zekun Zheng, et al. Dynamic brain spectrum acquired by a real-time ultraspectral imaging chip with reconfigurable metasurfaces. *Optica*, 9(5):461–468, 2022. 1, 3

[31] Shijie Yu, Feng Zhu, Dapeng Chen, Rui Zhao, Haobin Chen, Shixiang Tang, Jinguo Zhu, and Yu Qiao. Multiple domain experts collaborative learning: Multi-source domain generalization for person re-identification. *arXiv preprint arXiv:2105.12355*, 2021. 2

[32] 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. 1

[33] 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

[34] 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*, 45(5):5609–5631, 2022. 1, 3

[35] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Haibin Lin, Zhi Zhang, Yue Sun, Tong He, Jonas Mueller, R Manmatha, et al. Resnest: Split-attention networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2736–2746, 2022. 5

[36] 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

[37] Xianyi Zhang and Haitao Zhao. Hyperspectral-cube-based mobile face recognition: A comprehensive review. *Information Fusion*, 74:132–150, 2021. 1

[38] 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