A Domain Generalized Face Anti-Spoofing System Using Domain Adversarial Learning

Ching-Yi Chen[1,*], Sin-Ye Jhong[2], Chih-Hsien Hsia[3]

[1]Department of Information and Telecommunications Engineering, Ming Chuan University, Taoyuan, Taiwan, ROC

[2]Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan, ROC

[3]Department of Computer Science and Information Engineering, National Ilan University, Ilan, Taiwan, ROC

Received 28 January 2024; received in revised form 04 June 2024; accepted 05 June 2024

DOI: https://doi.org/10.46604/ijeti.2024.13314

Abstract

This research addresses the enhancement of face anti-spoofing (FAS) in facial recognition systems (FRS) against sophisticated fraudulent activities. Prior methods primarily focus on extracting facial features like color, texture, and dynamic variations, yet these methods struggle to accurately identify common characteristics of forged faces, thereby limiting generalization in practical scenarios. This study aims to propose a novel representation learning framework incorporating adversarial learning algorithms, to segregate features into liveness-specific and domain-specific categories, emphasizing liveness-specific features for FAS advancement. Feature disentanglement is central to this approach, enabling the deep learning models to effectively discern separable latent generating factors, such as identity, liveness, appearance, and texture. This methodology enhances model interpretability, explainability, and generalization. Additionally, Grad-CAM is employed to elucidate the basis for classifications made by the architecture, increasing explainability and trustworthiness. Empirical evaluation across panoply available FAS datasets confirms the superiority, significantly improving performance and robustness over existing technologies.

Keywords: face anti-spoofing, representation learning, adversarial learning algorithms, Grad-CAM

1. Introduction

With the development of facial recognition technology, the authentication of user identity using facial features has been pervasively applied across multifarious domains such as security, financial transactions, smart home technologies, and medical education. Despite the consequent convenience, it is highly susceptible to attacks due to the ease of collecting facial biometrics from quotidian conditions. These vulnerabilities include but are not limited to photo attacks [1], video attacks [2], and 3D mask attacks [3]. In addition, the photo attack presents cost-effectiveness and intuitive operability; attackers can expose stolen or downloaded facial images of the specific person in front of the sensors of the facial recognition system (FRS) or devices, such as smartphones and tablets, to successfully commence the attack.

Apropos such security perils, face anti-spoofing (FAS) in protecting FRS has been validated. As early as 2016, the International Organization for Standardization (ISO/IEC) issued the international standard ISO/IEC 30107. As mentioned in this standard, the main objective is to define and establish a framework for modeling and detecting spoofing attacks on biometric systems, effectively defending against spoofing activities targeting biometric recognition systems. Undeniably, every method of biometric authentication is prone to a certain degree of risk, and face spoofing attacks pose significant threats

[*] Corresponding author. E-mail address: chingyi@mail.mcu.edu.tw

English language proofreader: Chih-Wei Chang

to both personal assets and societal security. FAS, also referred to as liveness detection, is a technique designed to defend FRS against manifold attack vectors including photos, videos, and masks. To address the issue of spoofing attacks, researchers, under the premise that there exist inherent differences between veracious and fake faces, have proposed anti-spoofing methodologies focusing on analyzing texture in color spaces [1, 4], image quality assessment indicators [5], and temporal variations [6].

Yang et al. [7] propose the use of the convolutional neural network (CNN) for FAS. However, due to a limited amount of training data and significant intra-class and inter-class differences in fake facial images, the recognition rate still requires further betterment. In Li et al. [8], on the other hand, CNN and facial depth maps captured by Kinect to extract features are employed, demonstrating high discriminative power between live human faces and 2D fake facial images. Nevertheless, despite the performance improvement facilitated by the introduction of CNN in FAS, FAS is confronted with several challenges. Typically, for instance, when directly applying a network model trained on a source domain dataset to test on an invisible dataset, domain shift issues may arise.

The facial feature representation embodies an entity where commonality and uniqueness coexist. Each face possesses inherent similarities, yet pronounced distinctions emerge in identity between different visages. Such characteristics impede the generalization capabilities of extant FAS methods when confronted with unseen domains. Some researchers propose leveraging multiple existing source datasets to train models to derive a domain-invariant feature space. This, in turn, facilitates the resultant utilization of the learned generalized feature space to augment the generalization performance of the model during testing in unknown domains.

Domain generalization (DG) aims to enhance the robustness of deep learning models and focus on exploring relationships among multiple source domains without direct engagement with target data. The ultimate goal of DG is to foster the development of generalization capabilities that apply to domains not encountered during training, as evidenced in studies [9-11]. Meanwhile, adversarial learning has been implemented to train diverse feature extractors to further learn a universal feature space for the liveness feature, as presented in Shao et al. [9] and Chen et al. [12]. However, the challenge arises from a myriad of attack types in determining a universally applicable feature space for spoofed face detection through DG. Such a hindrance originates from the tendency of the learned representations to assimilate features irrelevant to facial liveness detection, incurring potential overfitting on the training dataset. Consequently, this method is often considered less than ideal in the context of FAS technology.

Machine learning represents a data-driven methodology for modeling problems; however, in the pixel space of images, all variations are intricately entangled, thereby impacting the output. Although deep learning models excel at interpreting statistical patterns between pixel-level data and labels, the performance is rather ineffective in capturing operable causal factors or interpretable image representations from data. The objective of feature disentanglement is to transform these entwined data variations in the original data space into a well-defined representational space. In this space, variations of different elements manifest separability [12-14].

In this study, the latent space of facial images is postulated to be decomposable into two subspaces: the liveness space and the domain space. The liveness features are associated with liveness-related information, while the domain features are indicative of domain-specific information. Given this dichotomy, feature disentanglement techniques are employed to effectively separate the liveness features from the domain features. Subsequently, the extracted liveness information is harnessed for facial liveness detection, with the overarching goal of establishing a more generalized framework for FAS.

The composition of the paper is organized as follows. Section 2 introduces background knowledge on DG, feature disentanglement, and advanced loss functions. Section 3 is concerned with the system architecture. Subsequent experimental results and analysis are presented in Section 4. Finally, Section 5 concludes the paper.

2. Related Approaches

Advancements in FAS necessitate addressing key challenges, such as DG and feature disentanglement, which significantly impact the generalization capabilities and interpretability of models. This section explores notable methodologies and ensuing contributions to improving FAS effectiveness.

2.1. DG for FAS

Concerning deep learning training, a discrepancy arises between the distributions of training and testing data, with training data often sourced from multiple origins, each possessing its distributional shift. Deep learning models inherently tend to overfit the training set, and the resultant domain shift can significantly impair the model's generalization capabilities. DG addresses this challenge by enabling deep learning models to explore relationships among multiple source domains without exposure to any target data. This approach aims to equip the models with generalizability to unseen domains, thus mitigating the impact of distributional discrepancies between training and testing environments.

Augmenting data diversity via methods such as data augmentation or data generation is instrumental in aiding deep learning models to develop more universally applicable representations. Data augmentation strategies often utilize various transformations and adversarial techniques to increase the spectrum of data. On the other hand, data generation methods may include the generation of supplementary samples using approaches like pairwise linear interpolation. Nevertheless, when viewed from the DG perspective in addressing FAS challenges, the aforementioned methodologies do not conform to end-to-end learning paradigms. Moreover, the heterogeneity in types of attacks and data acquisition methods poses a significant challenge in identifying a feature space generalizable across different synthetic facial representations. Consequently, apropos FAS, these techniques are frequently regarded as suboptimal solutions.

2.2. Feature disentanglement

Typically, supervised deep learning models adhere to an end-to-end shortcut learning strategy [15], characterized by their "black-box" nature, rendering the knowledge representation incomprehensible to humans. These models prioritize enhancing prediction accuracy for training samples, enabling the models to autonomously select the most straightforward path for fitting the input/output dataset and adjusting parameters accordingly. The goal of feature disentanglement is to guide deep learning models to extract separable latent generative factors from real-world data in an anthropocentrically understandable manner. These factors include identity, liveness, appearance, texture, etc., thereby establishing a causal model that is not only interpretable and supportive of explanations but also transcends mere pattern recognition. This approach assists in uncovering the internal generative mechanisms of data, thus emerging as a vital tool for enhancing the generalizability and controllability of deep learning models. Representation learning, in this context, involves learning domain-invariant representations or disentangling domain-shared and domain-specific features, consequently bolstering the generalization performance.

2.2.1. Autoencoder and representation learning

The autoencoder utilizes a combination of an encoder and a decoder to reconstruct its features [16]. The encoder compresses input features from a high-dimensional space into a lower-dimensional latent space, while the decoder is responsible for restoring the low-dimensional vector to its original features, as depicted in Fig. 1. The loss function is defined as:

$$L_{MSE} = E_{x-p(x)}\left\{\left[x - Dec\left(Enc(x)\right)\right]^2\right\} \quad (1)$$

where *Enc* represents the encoder, and *Dec* represents the decoder. Autoencoders are commonly employed for learning latent representations of data.

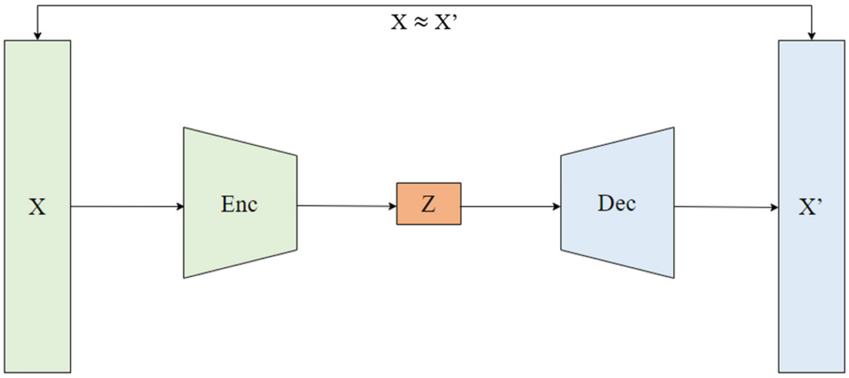

Fig. 1 Architecture of the autoencoder

2.2.2. Generative adversarial network (GAN)

The GAN is a commonly used deep generative model [17]. Technically, GAN consists of a generator G and a discriminator D. The task of the generator is to transform samples from a prior probability distribution into data points, while the discriminator is in charge of distinguishing whether the input is real or generated by the generator. The training process involves maximizing the minimum value of a game between these two components, which is referred to as the min-max game, and ultimately aims to train a generator that can generate data corresponding to real data. The loss function for GAN is expressed as:

$$\min_{G}\max_{D} V\left(D,G\right)=E_{x-p_{data}(x)}\left(\log D(x)\right)+E_{z-p_{G}(z)}\left\{\log\left[1-D\left(G(z)\right)\right]\right\} \tag{2}$$

where $p_G(z) = N(0, 1)$ represents the prior probability distribution, and $p_{data}(x)$ represents the sampling from the training data.

When the encoder and discriminator engage in adversarial training, the discriminator can remove undesired properties from the representations produced by the encoder. This process leads to the emergence of disentangled representations, whose learning seeks to decompose representations to disentangle the latent explanatory factors hidden within observed data. This approach is widely useful across multifarious tasks, enhancing both robustness and controllability.

2.3. Advanced loss functions

Facial recognition algorithms utilize CNNs to extract extensive facial feature vectors. The process of facial identification entails comparing a facial feature vector with those stored in a database. Regarding accuracy, the algorithm relies on consistently capturing similar features of the same person across panoply images while distinguishing features of different individuals. However, traditional CNNs, primarily based on multilayer perceptron networks and employing Softmax loss for classification, demonstrate excellent performance in general image tasks but may lack the necessary discriminative power for tasks with a massive number of label classes. On the other hand, Softmax loss, a combination of the Softmax function and Cross-Entropy loss, is favored in image classification for its ease of optimization and rapid convergence. However, since Softmax loss does not enforce intra-class compactness and inter-class dispersion, the learned features are separable for closed-set classification problems, but not sufficiently discriminative for open-set face recognition problems. Moreover, the size of the linear transformation matrix $W \in \mathrm{R}^{dxn}$ increases linearly with the number of identities n.

Concerning tasks such as retrieval or verification that rely on setting decision thresholds, recent advancements have led to the proposal of more sophisticated loss functions, e.g., large margin cosine loss [18], SphereFace loss [19], and additive angular margin (ArcFace) loss [20]. The underlying principle of these enhanced algorithms is to augment inter-class variance while minimizing intra-class variance. This approach is achieved by reducing the distances between data points within the same category and increasing the margin between different categories. Such a methodology ensures the precise classification of a vast array of distinct categories, thus eliciting the overall effectiveness of the classification process. Concerning the decision boundaries problem of a classification case, the decision margin of cosine loss has a nonlinear angular margin in angular space. The decision margin of SphereFace loss also changes gradually with the angle, while only ArcFace loss possesses a constant linear angular margin throughout the entire range, exhibiting superior geometric characteristics in the issue of decision boundaries.

2.3.1. Cosine loss

In Wang et al. [18], L2 normalization is employed on weight vectors to normalize features and reduce radial variations, Softmax loss is further modified into Cosine loss. The definition of Cosine loss is presented in the formula below:

$$L_{CosFace} = \frac{1}{N}\sum_{i} -\log \frac{e^{s\left(\cos(\theta_{yi})-m\right)}}{e^{s\left(\cos(\theta_{yi})-m\right)} + \sum_{j\neq y_i} e^{s\cos(\theta_{yi})}} \tag{3}$$

$$W = \frac{W^*}{\left\|W^*\right\|} \tag{4}$$

$$x = \frac{x^*}{\left\|x^*\right\|} \tag{5}$$

$$\cos(\theta_j, i) = W_j^T x_i \tag{6}$$

where N is the number of training samples, x_i is the ith feature vector corresponding to the ground-truth class of y_i, the W_j is the weight vector of the jth class, and θ_j is the angle between W_j and x_i.

2.3.2. SphereFace loss

SphereFace loss, as introduced by Liu et al. [19], redefines the Softmax loss by transitioning from an Euclidean distance metric to an angular margin framework. This transformation incorporates an augmented decision margin, which can be denoted as “*m*”, and is further characterized by the constraints $\|W\| = 1$ and $b = 0$. The mathematical formulation of SphereFace loss is presented in the formula below:

$$L_{SphereFace} = \frac{1}{N}\sum_{i} -\log \frac{e^{\|x_i\|\psi(\theta_{yi},i)}}{e^{\|x_i\|\psi(\theta_{yi},i)} + \sum_{j\neq y_i} e^{\|x_i\|\cos(\theta_{yi},i)}} \tag{7}$$

$$\psi(\theta_{y_i,i}) = (-1)^k \cos(m\theta_{y_i,i}) - 2k \tag{8}$$

$$\theta_{y_i,i} \in \left[\frac{k\pi}{m}, \frac{(k+1)\pi}{m}\right] \tag{9}$$

$$k \in [0, m-1] \tag{10}$$

Here, $\psi(\theta,i)$ is a monotonically decreasing angular function obtained by defining the range of cos $(\theta_{y_{i,j}})$, where $m \geq 1$ is an integer controlling the size of the angular margin.

2.3.3. ArcFace loss

In Deng et al. [20], the ArcFace loss is proposed to improve facial recognition performance and stabilize the training procedure. This method entails the use of the arc-cosine function to determine the angle between the current feature vector and the target class center. An additive angular margin is subsequently added to the target angle. Moreover, the cosine function is utilized to compute the target logit, followed by a rescaling of all logits using a fixed feature norm. The computational steps that ensue align with those employed in the traditional Softmax loss framework. The mathematical formulation of ArcFace loss is presented in:

$$L_{ArcFace} = \frac{1}{N}\sum_{i} -\log \frac{e^{s\left(\cos(\theta_{yi}+m)\right)}}{e^{s\left(\cos(\theta_{yi}+m)\right)} + \sum_{j=1,\, j\neq y_i} e^{s\cos\theta_{yi}}} \tag{11}$$

where N is the number of training samples, x_i is the ith feature vector corresponding to the ground-truth class of y_i, and m represents the angular margin. The loss function aims to maximize the angular margin between the correct class and the other classes.

3. Domain-Generalized FAS System

This research framework in this study encompasses two key components, viz., feature disentanglement and DG. First, during the feature disentanglement phase, representations are separated into two distinct elements: liveness-related features and domain-specific features. Second, in the DG phase, the encoder dedicated to liveness is designed to harvest generalized liveness features from a diverse range of domains. Fig. 2 illustrates the proposed approach and provides an overview of the entire learning process. E_L is the liveness encoder, while E_D is the domain encoder. On the other hand, C_L is the liveness classifier, and C_D serves as the domain classifier.

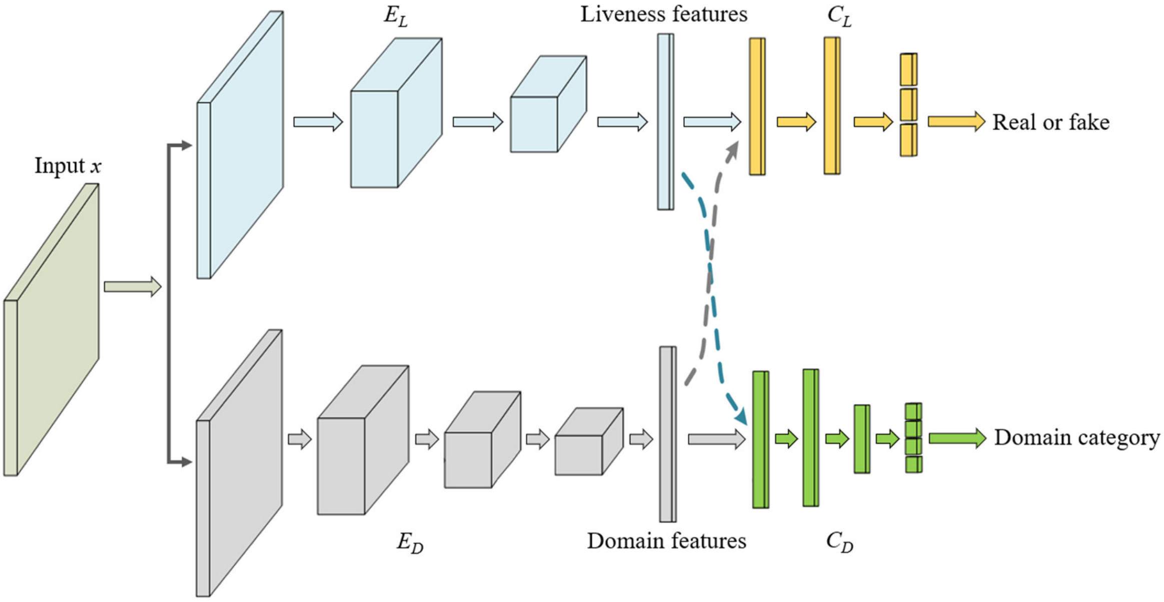

Fig. 2 The overall architecture of the proposed framework

3.1. Overview

To proficiently extract a feature space suitable for real-world FAS scenarios, the primary objective is to segregate domain-specific features, which do not pertain to spoofing, from the liveness features. To accomplish this goal, a proficient framework for disentangled representation learning is introduced, specifically tailored for cross-domain FAS, as illustrated in Fig. 2. This framework is comprised of two integral components: the feature disentanglement process and the symmetric adversarial learning process.

The input, denoted as x, is fed into both encoders, namely E_L and E_D, which represent the facial liveness encoder and the domain encoder, respectively. Here, i symbolizes the number of the different training domain datasets. C_L and C_D signify the liveness classifier and the domain classifier, respectively. The two feature encoders, E_L and E_D, are trained to extract distinct liveness features and domain features. The liveness encoder focuses on identifying features that indicate the presence of a live subject in an image, commonly referred to as liveness features. On the other hand, the domain encoder is responsible for recognizing domain-specific features, which are pertinent to the particular characteristics of the image's environment or context.

During adversarial training, these two encoders operate in a sort of competitive yet complementary manner. The liveness encoder functions to differentiate liveness features from other irrelevant features. Simultaneously, the domain encoder distinguishes domain-specific features, averting the interference with the recognition of liveness features. The key goal of this training process is to refine both encoders' ability to accurately and effectively identify their respective feature sets. Over time, the development of a more robust and discriminative feature space is facilitated. In this space, liveness features are clearly distinguished from domain-specific features, enabling more precise and reliable image analysis, particularly in applications like facial recognition or authentication systems.

3.2. The loss function for learning generalized facial liveness features

In defining the loss function for FAS applications, it is crucial to consider the diversity of potential attacks, which can vary significantly across different domains. For instance, attacks in Domain A might involve photo-based methods, while Domain B could use 3D masks. Relying on traditional Cross-Entropy loss functions in such scenarios would substantially hinder the ability to generalize across different domains. Recent studies have highlighted the limitations of conventional Softmax loss in fully maximizing discriminative power for classification tasks. As a result, a shift has emerged in research efforts towards developing loss functions that focus on maximizing inter-class variance and minimizing intra-class variance, aiming to significantly improve performance in FAS systems.

$L_{Liveness}$ is a measure used to assess whether facial features exhibit characteristics of a live face, as indicated in:

$$L_{Liveness} = L_{ArcFace} \tag{12}$$

L_{Domain}, on the other hand, serves to identify the specific domain to which the facial image x_i belongs, as represented in:

$$L_{Domain} = \sum_{i=1}^{N} \left[C_D \left(E_D(x_i) \right) - d_i \right]^2 \tag{13}$$

Here, N denotes the total count of the image data, where x_i is expressed as the ith feature vector. On the other hand, d_i is designated as the domain label of x_i.

L_{Domain} is designed to ensure that the domain features encapsulate information from various domains. Regarding effective DG, the liveness classifier must be unable to accurately differentiate between these domain features. The error functions for $L_{Conf,L}$ is elaborately detailed in Eq. (8):

$$L_{conf,L} = -\sum_{i=1}^{N} \left[C_D \left(E_L(x_i) \right) - \frac{1}{D} \right]^2 \tag{14}$$

where D represents the total number of different domain datasets.

In line with the principles of adversarial training, the domain classifier must be unable to effectively differentiate liveness features. This inability is a strategic aspect of the training, aimed at the betterment of the model apropos overall robustness and generalizability. The error functions for $L_{Conf,D}$, which encapsulate this aspect of the learning process of the model, are defined as follows below:

$$L_{conf,D} = -\sum_{i=1}^{N} \left[C_L \left(E_D(x_i) \right) - \frac{1}{2} \right]^2 \tag{15}$$

4. Experimental Results and Discussion

This section introduces the experimental setup and results of the proposed framework. To enhance explainability, ablation studies on different loss functions and the proposed framework are included. Furthermore, explainable visualizations using Grad-class activation mapping (CAM) are provided.

4.1. Experimental setup

In this research, the effectiveness of the proposed FAS method was evaluated using four publicly available databases: OULU_NPU [21] (referred to as O), CASIA-FASD [22] (C), Idiap Replay-Attack [23] (I), and MSU-MFSD [24] (M). The mentioned FAS method employs ArcFace [20] as the loss function. During the experiments, each database was alternately designated as the target domain for testing, while the rest served as source domains for training, thereby establishing four distinct testing scenarios: O&M&I to C, O&C&I to M, O&C&M to I, and I&C&M to O. These scenarios were characterized by considerable variations within and across databases, such as in background settings and resolution. Consistent with the methodologies outlined in Shao et al. [9], the area under the curve (AUC) metric was employed to evaluate and compare the performance of different methods under the aforementioned conditions.

4.2. Experimental results

To further substantiate the DG capabilities of the approach, Table 1 presents the significant achievements yielded by the method in a comparative evaluation with other contemporary FAS methods. These results align with the theoretical underpinnings of the model, which prioritizes the effective accentuation of liveness-related facial features through the disentanglement of domain-specific features. Simultaneously, this approach facilitates the extraction of more generalized active features, thereby contributing to an overall enhancement in performance.

Table 1 Comparison of different FAS methods for DG performance (AUC, %)

Method	O&M&I to C	O&C&I to M	O&C&M to I	I&C&M to O
CT [1]	76.89	78.74	62.78	32.71
LBP-TOP [25]	61.05	70.80	49.54	44.09
MS-LBP [26]	44.98	78.50	51.64	49.31
Auxiliary [27]	73.15	85.88	71.69	77.61
The proposed method	87.63	**89.75**	85.13	86.15

Table 2 proffers a comparative analysis of the feature disentanglement method, which utilizes multifarious loss functions. The empirical data demonstrate that the proposed method significantly enhances performance by effectively segregating facial domain features from liveness features. This improvement is consistent across different loss functions, including ArcFace loss and SphereFace loss, with the approach outperforming other existing FAS methods.

Table 2 Comparison of Different facial loss functions (AUC, %)

Method	O&M&I to C	O&C&I to M	O&C&M to I	I&C&M to O
Use ArcFace [20]	87.63	89.75	85.13	86.15
Use SphereFace [19]	92.53	86.72	84.32	83.80

4.3. Explainable visualization with Grad-CAM

Grad-CAM [28], an extension of CAM, is a sophisticated visualization technique that elucidates which segments of a deep neural network are most influential in its predictive output. In addition, Grad-CAM facilitates the identification of specific image regions, thereby rendering the decision-making process of the neural network more transparent and visually interpretable.

Grad-CAM achieves this by generating a coarse localization map using the gradients directed towards any targeted concept within the network, thereby accentuating critical areas in the image that are pivotal for concept prediction. As illustrated in Fig. 3, Grad-CAM is employed to furnish visual interpretative insights into the methodology employed in this study.

In contrast to the conventional binary CNN classifier architecture [7], the model predominantly seeks discriminative cues within the internal facial regions, minimizing its reliance on domain-specific backgrounds. This approach enhances the potential of the model for effective generalization to unfamiliar domains. Moreover, the method exhibits the capability to dynamically adapt its focus to different regions within an image in response to manifold types of face spoofing attacks, tailoring its response to the specific nature of each attack. The visualization results of applying Grad-CAM to the binary CNN method manifest that the regions used to distinguish between real and fake faces are not exclusively concentrated on the facial area but are dispersed across parts of the face and the background. Such a discovery is not very reasonable.

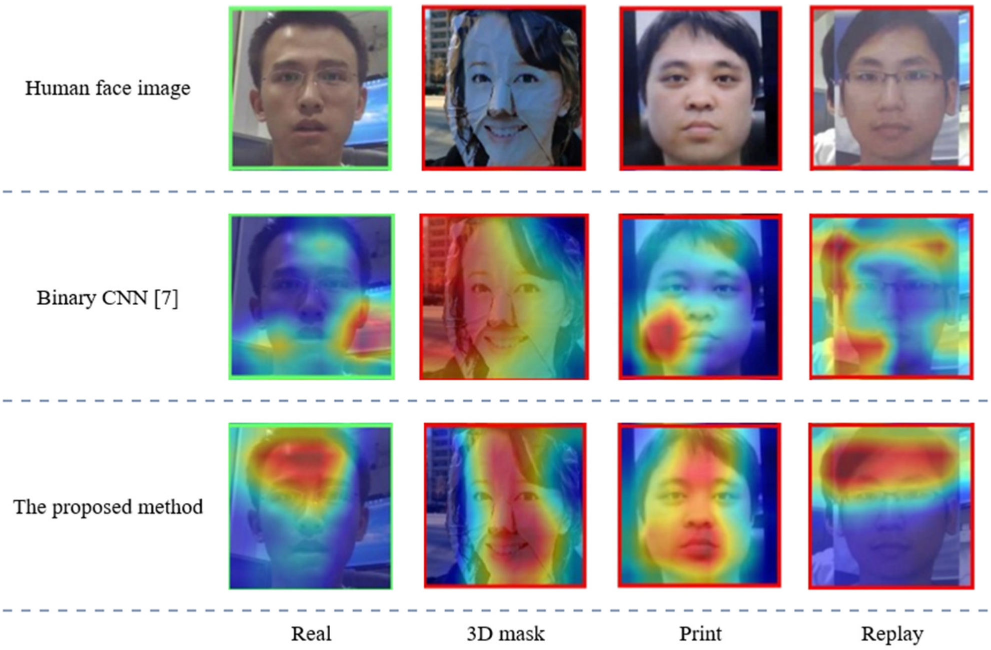

Fig. 3 Grad-CAM results on the four examples of the Binary CNN and the proposed method

5. Conclusion

This research focuses on leveraging feature disentanglement to augment the generalization capacity of FAS systems. The methodology involves decomposing the latent space of facial images into two distinct subspaces: a “liveness” space and a “domain” space. The liveness features extracted from the former are utilized for anti-spoofing classification. This strategy avails enhanced generalization of the model across a broad spectrum of domains. Empirical evidence, which is gathered from testing the proposed framework on four publicly available datasets, manifests superior performance and robust generalization capabilities to the unseen domains. The experimental results of the Grad-CAM visualization technique also demonstrate that the most influential areas for predictive output of the proposed method mainly focus on the internal facial region rather than the background, thereby eliciting a potential for effective generalization to unfamiliar domains.

Looking ahead, future work could potentially evolve the architecture in this research into a multi-frame FAS framework. Such a framework would independently disentangle static liveness information, domain information, and face depth from continuous images. By employing 3D-CNN technology, the development of a depth encoder is envisaged, which is designed to capture depth variations arising from micro-movements within continuous multi-frame images. This approach promises to further enhance the robustness and generalization capability of FAS systems by furnishing a more comprehensive analysis of facial dynamics and depth changes.

Acknowledgment

This work was supported by the National Science and Technology Council, Republic of China under contract MOST 111-2221-E-130-009.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face Spoofing Detection Using Colour Texture Analysis," IEEE Transactions on Information Forensics and Security, vol. 11, no. 8, pp. 1818-1830, August 2016.

[2] U. Muhammad, M. Z. Hoque, M. Oussalah, and J. Laaksonen, "Deep Ensemble Learning With Frame Skipping for Face Anti-Spoofing," Twelfth International Conference on Image Processing Theory, Tools and Applications, pp. 1-6, October 2023.

[3] Q. Guo, S. Liao, Y. Chen, S. Xiao, and J. Ma, "A Deep Learning Based Approach for 3D Mask Face Anti-Spoofing Enhancement," 8th International Conference on Computer and Communication Systems, pp. 1-5, April 2023.

[4] N. Nanthini, N. Puviarasan, and P. Aruna, "An Enhanced Face Anti-Spoofing Model Using Color Texture and Corner Feature Based Liveness Detection," 2nd International Conference on Innovative Practices in Technology and Management, pp. 63-68, February 2022.

[5] K. Karthik and B. R. Katika, "Image Quality Assessment Based Outlier Detection for Face Anti-Spoofing," 2nd International Conference on Communication Systems, Computing and IT Applications, pp. 72-77, April 2017.

[6] R. Shao, X. Lan, and P. C. Yuen, "Deep Convolutional Dynamic Texture Learning With Adaptive Channel-Discriminability for 3D Mask Face Anti-Spoofing," IEEE International Joint Conference on Biometrics, pp. 748-755, October 2017.

[7] J. Yang, Z. Lei, and S. Z. Li, "Learn Convolutional Neural Network for Face Anti-Spoofing," https://doi.org/10.48550/arXiv.1408.5601, August 24, 2014.

[8] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li, and A. Hadid, "An Original Face Anti-Spoofing Approach Using Partial Convolutional Neural Network," Sixth International Conference on Image Processing Theory, Tools and Applications, pp. 1-6, December 2016.

[9] R. Shao, X. Lan, J. Li, and P. C. Yuen, "Multi-Adversarial Discriminative Deep Domain Generalization for Face Presentation Attack Detection," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10015-10023, June 2019.

[10] Y. Jia, J. Zhang, S. Shan, and X. Chen, "Single-Side Domain Generalization for Face Anti-Spoofing," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8481-8490, June 2020.

[11] H. Li, S. J. Pan, S. Wang, and A. C. Kot, "Domain Generalization With Adversarial Feature Learning," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5400-5409, June 2018.

[12] Z. C. Chen, L. H. Tsao, C. L. Fu, and Y. C. Frank Wang, "Disentangled Representation for Domain Generalized Face Anti-Spoofing," Graduate Institute of Communication Engineering, National Taiwan University, Bachelor Thesis Award, 2021.

[13] Y. C. Wang, C. Y. Wang, and S. H. Lai, "Disentangled Representation With Dual-Stage Feature Learning for Face Anti-Spoofing," IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1234-1243, January 2022.

[14] Z. Liu, Z. Feng, Z. Zou, R. Zhang, M. Song, and J. Shen, "Disentangled Representation Based Face Anti-Spoofing," 25th International Conference on Pattern Recognition, pp. 2017-2024, January 2021.

[15] R. Geirhos, J. H. Jacobsen, C. Michaelis, R. Zemel, W. Brendel, M. Bethge, et al., "Shortcut Learning in Deep Neural Networks," Nature Machine Intelligence, vol. 2, pp. 665-673, 2020.

[16] Z. Ren, "The Advance of Generative Model and Variational Autoencoder," IEEE Conference on Telecommunications, Optics and Computer Science, pp. 268-271, December 2022.

[17] K. Sun, Q. Wen, and H. Zhou, "Ganster R-CNN: Occluded Object Detection Network Based on Generative Adversarial Nets and Faster R-CNN," IEEE Access, vol. 10, pp. 105022-105030, 2022.

[18] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, et al., “CosFace: Large Margin Cosine Loss for Deep Face Recognition,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5265-5274, June 2018.

[19] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “SphereFace: Deep Hypersphere Embedding for Face Recognition,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 6738-6746, July 2017.

[20] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 10_Part_1, pp. 5962-5979, October 2022.

[21] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, “Oulu-NPU: A Mobile Face Presentation Attack Database With Real-World Variations,” 12th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 612-618, May-June 2017.

[22] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, “A Face Antispoofing Database With Diverse Attacks,” 5th IAPR International Conference on Biometrics, pp. 26-31, March-April 2012.

[23] I. Chingovska, A. Anjos, and S. Marcel, “On the Effectiveness of Local Binary Patterns in Face Anti-Spoofing,” BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group, pp. 1-7, September 2012.

[24] D. Wen, H. Han, and A. K. Jain, “Face Spoof Detection With Image Distortion Analysis,” IEEE Transactions on Information Forensics and Security, vol. 10, no. 4, pp. 746-761, April 2015.

[25] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, “LBP − TOP Based Countermeasure Against Face Spoofing Attacks,” Computer Vision - ACCV 2012 Workshops: ACCV 2012 International Workshops, pp. 121-132, November 2012.

[26] J. Määttä, A. Hadid, and M. Pietikäinen, “Face Spoofing Detection From Single Images Using Micro-Texture Analysis,” International Joint Conference on Biometrics, pp. 1-7, October 2011.

[27] Y. Liu, A. Jourabloo, and X. Liu, “Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 389-398, June 2018.

[28] K. Duvvuri, S. Chethana, S. S. Charan, V. Srihitha, T. K. Ramesh and K. S. Srikanth, “Grad-CAM for Visualizing Diabetic Retinopathy,” 3rd International Conference for Emerging Technology, pp. 1-4, May 2022.