The Visual Computer (2021) 37:1015–1028
https://doi.org/10.1007/s00371-020-01849-x

ORIGINAL ARTICLE

Understanding deep face anti-spoofing: from the perspective of data

Yujing Sun[1] · Hao Xiong[1] · Siu Ming Yiu[1]

Published online: 15 May 2020

Abstract
Face biometrics systems are increasingly used by many business applications, which can be vulnerable to malicious attacks, leading to serious consequences. How to effectively detect spoofing faces is a critical problem. Traditional methods rely on handcraft features to distinguish real faces from fraud ones, but it is difficult for feature descriptors to handle all attack variations. More recently, in order to overcome the limitation of traditional methods, newly emerging CNN-based approaches were proposed, most of which, if not all, carefully design different network architectures. To make CNN-related approaches effective, data and learning strategies are both indispensable. In this paper, instead of focusing on network design, we explore more from the perspective of data. We present that appropriate nonlinear adjustment and hair geometry can amplify the contrast between real faces and attacks. Given our exploration, a simple convolutional neural network can solve the face anti-spoofing problem under different attack scenarios and achieve state-of-the-art performance on well-known face anti-spoofing benchmarks.

Keywords Face anti-spoofing · Biometrics · Image adjustment · Image processing

1 Introduction

Nowadays, the usage of facial biometrics in various scenarios of business and industry is dramatically increasing and becoming popular in authenticating user identities. One could protect his privacy in electronic devices using face unlocking techniques, to conveniently open bank account remotely via identity verification with webcam, and even to authenticate payment with facial biometrics.

However, it is insecure to use face as a biometric measure for authentication. A recent study [30] on face recognition using commercial matchers shows that face biometric systems can be vulnerable to spoofing attacks, such as fraud photographs, videos, or masks that launch against face authentications or recognition systems, and can lead to inestimable privacy leak and property loss, for instance, private photographs, and sensitive bank information.

Moreover, given the rapid development and prevalence of social media, people are sharing their facial photographs on the internet intentionally. Malicious people can easily obtain such photographs to attack facial recognition systems. Comparing with other biometrics, such as fingerprint and iris, facial images are much more convenient to acquire. As a consequence, the demand to effectively prevent face spoofing attacks is significantly on the rise.

Researchers are taking much effort in recent studies to tackle the problem by investigating different clues. Hardware-based methods take advantages of devices to catch the differences between real and fake faces, such as thermal camera [14] and 3D camera [26]. Despite the satisfactory results achieved, the high cost restricts the practicality of hardware-based approaches. Software-based methods make use of dynamic clues or static image clues. Comparing with dynamic clues, for instance, eye blink [35] and lip movement [11], image-based metrics are more appealing and widely used for that face anti-spoofing process shares the

Yujing Sun and Hao Xiong have contributed equally to this work.

This project is partially supported by a Collaborative Research Fund (CRF, C1008-16G) and Innovation and Technology Support Programme (ITS/173/18FP) of the Hong Kong Government.

✉ Yujing Sun
yjsun@cs.hku.hk

Hao Xiong
hxiong@hku.hk

Siu Ming Yiu
smyiu@cs.hku.hk

[1] Department of Computer Science, The University of Hong Kong, Pok Fu Lam, Hong Kong, China

same information as face recognition procedure and can be easily coupled into it.

Manufacturing procedures, spoofing mediums, or surface geometries can result in various artifacts, such as low quality, image blur, specular reflection, and distorted shading appearance. To locate such artifacts in attacks, image-based methods define various feature descriptors, including texture analysis [51], quality measure [4], and Fisher Vector [5]. However, it is difficult for handcraft descriptors to cover all the variations in different spoofing attacks.

With the success of neural networks in image vision tasks, Yang et al. [50] and Li et al. [29] attempted to use CNNs to perform anti-spoofing automatically. For learning-based approaches to achieve satisfactory performance, data is of the same importance as learning strategy. Nevertheless, most previous CNN-based approaches explore various network architectures while ignore the importance of data understanding. Li et al. [29] made use of the famous VGG-Face architecture and Yang et al. [50] adopted a champion architecture in ImageNet contest. They aim to use the learned deep features to overcome the limitation of traditional features. Unfortunately, without fully understanding the spoofing data, different designs of CNN architecture fail to defeat many handcraft descriptors. Recently, CNNs with more prior knowledge have been proposed. Based on the observation that depth information is critical in detecting 2D attacks, Atoum et al. [1] and Liu et al. [32] presented depth-dependent CNNs. Taking advantage of time coherence, Li et al. [27] introduced a 3D CNN. Additionally, Jourabloo et al. [21] used a noise model-based CNN and Liu et al. [33] presented an unsupervised deep tree network (DTN) to solve the problem. Comparing with previous learning methods, their performance is greatly improved. But they also focus on revising the architectures of CNN and use complex models.

In this paper, we also present a CNN-based method but from a different perspective. Most existing learning-based methods, if not all, focus on exploring various network architectures, however, we devote to understanding the spoofing data. Instead of designing complex and deep architectures as the previous CNN methods do, we use the simplest network design to achieve a promising accuracy, thanks to our better data comprehension by exploring different properties of live faces and attacks. We overview the problem pipeline in Fig. 1. Traditional methods use retrieved features from designed metrics for classification, but the representation power of such feature vectors is limited and will constrain CNN performance [41]; previous CNNs replace the entire procedure in traditional methods with an automatically learned convolution neural network. Our pipeline contains a feature enhancement unit, in which images are enhanced with distinguishable features amplified. Comparing with traditional methods, we adjust desired features without losing any other image information, while comparing with prior

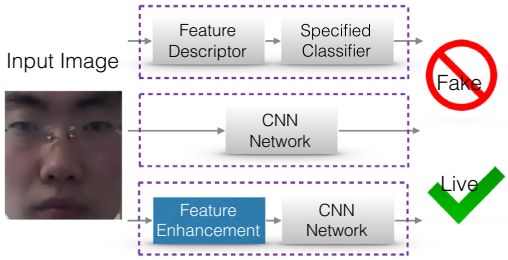


Fig. 1 Pipeline overview. From top to bottom: traditional feature-based approaches, previous CNN methods and our methodology

Fig. 2 Demonstration of using different face types (facial information). Most existing methods uses **a** or **b** while we find that additional hair information **c** can improve detection accuracy

CNN approaches, we focus more on data understanding rather than learning strategy. Based on our explorations, we achieve state-of-the-art accuracy on three well-known data sets with a simple CNN model.

We summarise our contributions as follows:

- We present that appropriate image adjustments can amplify the contrast between live faces and attacks to improve the performance of CNNs. Particularly, we find that nonlinear improvement of both brightness and contrast can greatly improve the detection accuracy.
- We propose that the complex shading of hair geometry can provide important clues for CNN-based spoofing detection. Therefore, type Head (Fig. 2c) is superior to type Face (Fig. 2a) in CNN-based face anti-spoofing.
- We observe that though RGB color space performs poorly in traditional feature-based methods, after our adjustments, it is superior to HSV and YCBCr in CNN-based approaches.
- We demonstrate that given sufficient data understanding, a simple network design with much fewer learning parameters is sufficient to produce decent results.

Most importantly, we demonstrate a different direction to improve the performance of CNNs. A common sense in CNN community that manually selected/ handcrafted features are not preferred due to information loss results in the misconception that preprocessing data can only reduce accuracy. To avoiding information loss, we manipulate each image as a

whole. As a result, preprocessing data based on our finds will amplify desired features without losing information. Besides, the data properties we proposed can be easily generalised to all kinds of CNNs.

2 Literature review

Researches on 2D face anti-spoofing date back over decades. Since then, various methods have been proposed under different attack scenarios. We briefly review the related works in this section. Based on the clues used by different approaches the state-of-arts can be divided into different categories, including CNN-based methods, property analysis, artifacts detection, and external assistance.

2.1 Learning-based methods

Neural networks are proved to be effective in solving many computer vision tasks. Researchers also attempt to tackle the face anti-spoofing problem with CNN solutions. Li et al. [29] and Patel et al. [36] proposed to extract feature based on pre-trained CNN models VGG-face and CaffeNet, respectively, while Xu et al. [49] presented an LSTM-CNN model to make a joint prediction. A CNN architecture is directly used in [50] as a classifier to detect face attacks. However, the performances of approaches mentioned above are unsatisfactory. Most recently, thanks to the success of CNN in estimating depth information from images [22], depth-based CNNs [1,32] are presented, in which depth information is estimated by CNNs and judgements are made accordingly. Taking advantage of temporal information, Li et al. [27] and Gan et al. [19] presented 3D convolutional neural networks to solve this problem. They achieve satisfactory accuracy on video-based databases but cannot deal with image-based databases which contains no temporal information. Due to the network complexity, Li et al. [27] needs a much longer training time as well. Additionally, Jourabloo et al. [21] used a noise model-based CNN and Liu et al. [33] presented an unsupervised deep tree network (DTN) to solve the problem, both of which tackled the face anti-spoofing problem by designing effective network structures. Thanks to our methodologies, our approach can achieve a similar or better performance than existing learning-based methods with a much simpler network architecture on video-based as well as image-based face anti-spoofing databases.

2.2 Property analysis

Real photographs and fake attacks have different properties due to different shooting conditions. Many works make use of such properties to perform face anti-spoofing. Real photographs and fake attacks share different texture property since human faces and spoof mediums reflect light in different manners. The representative texture analysis-based methods include LBP [15,16,34], SIFT [37], HOG [25,51], DoG [38,45], and SURF [5]. Most recently, Zhao et al. [53] proposed to consider volume local binary count patterns to solve the problem. Texture-based methods are fast but may have poor generalizability [37], especially sensitive to illumination variations and different identities. Certain approaches also take advantages of motion clues, such as eye blink [35], and head/lip movement [2] to prevent print attacks. But motion-based method can not deal with video attacks which have facial motions.

2.3 Artifacts detection

Spoofing attacks are recaptured photographs of live faces and thus inevitably contain artifacts caused by color distortion, specular reflection and/or blurriness. Galbally et al. [18], Wen et al. [48] and Garcia and de Queiroz [20] are able to detect fake faces based on image quality analysis. Moreover, observing that attacks lose some low-frequency information while introducing additional higher frequency components (noise signals), Li et al. [28], Pinto et al. [40], and Pinto et al. [39] aim to differentiate live faces from attacks in frequency domain. It is true that artifacts exist in all kinds of attacks, but artifacts changes when attacks vary. Therefore, such approaches possibly perform differently on different data sets.

2.4 External assistance

Hardware-based methods Methods in this category perform face anti-spoofing with the assistance of different hardwares, such as 3D camera [17,26], thermal camera [44], light field camera [23] and flash light [8]. Though hardware-based approaches can achieve higher accuracy, the usage of additional sensors restrict their applications.

User interactive methods User interactive methods require user cooperation to complete verification process. For instance, users have to move head for 3D shape reconstruction [47] and to speak for audiovisual matching purpose [9,10,12]. These methods achieve acceptable accuracy at the cost of user inconvenience and longer recognition time.

3 Methodology

In this section, we explain the mechanisms in details on why image adjustments, hair geometry and RGB color space can improve the face anti-spoofing accuracy. By taking advantage of the observations, a simple network design is able to solve the problem effectively.

3.1 Why real faces and fake ones are different?

Denote the intensity value or luminance channel of a facial image on pixel (x, y) as $I(x, y)$, in which $I(x, y) \in [0, 1]$. Assuming that human face is a Lambertian surface, accordingly, $I(x, y)$ can be expresses as

$$I(x, y) = \Re L_{am} \tag{1}$$

by the Lambertian Reflectance Law, where $\Re$ is the reflectivity of a surface and L_{am} represents the ambient illumination. Real faces and attacks have different $\Re$ due to different surface geometry and texture. But L_{am} is the same under the same lighting condition. Thus $\Re$ is the underlying key factor to differentiate real faces from fraud ones. However, $\Re$ is not always reliable. For different image qualities and lighting conditions, the performance of face liveness detection depending on $\Re$ alone is unstable [31].

3.2 Image adjustment

It is meaningful to introduce extra measures besides $\Re$ to solve the problem more robustly. After introducing flash lights, Chan et al. [8] modifies Eq. 1 with

$$I_f(x, y) = \Re L_{am} + (\Re \cdot \aleph) \cdot L_f, \tag{2}$$

where $\aleph$ is the parameter associated with the flash intensity, direction, and the normals of the face surface. The extra measure provided by the second term of Eq. 2 helps to reduce the error rate. Nevertheless, Chan et al. [8] is unable to deal with images taken under dark environment or without flashes. As a result, it cannot test on the well-known benchmarks we evaluate in Sect. 5.

We propose that image adjustments can work as additional measures in a more general fashion, which could be post-added to all existing images easily.

Define the parameters that control image contrast and brightness as α and β, respectively. Given $I(x, y)$, output $I_e(x, y)$ after adjustment can be written as

$$I_e(x, y) = \alpha \cdot I(x, y) + \beta, \tag{3}$$

where $\alpha \in (0, 2)$ and $\beta \in (-1, 1)$. Contrast and brightness will be changed accordingly when adjusting α and β respectively. The above adjustments can be linear or nonlinear depending on the linearity or nonlinearity of α and β, namely,

$$\alpha, \beta = \begin{cases} C_\alpha, C_\beta & \text{linear} \\ z_\alpha(I(x, y), C_\alpha), z_\beta(I(x, y), C_\beta) & \text{nonlinear,} \end{cases} \tag{4}$$

where C_x is a constant controlling the amount of adjustment and z_x is a nonlinear function of I and C_x. ($x \in \alpha, \beta$.)

Gamma correction γ is another nonlinear adjustment to correct image brightness, where brightness decreases with $\gamma > 1$ while increases when $0 < \gamma < 1$. Introducing γ, the output $I_e(x, y)$ becomes

$$I_e(x, y) = (\alpha \cdot I(x, y) + \beta)^\gamma, \tag{5}$$

which can be rewritten as

$$I_e(x, y) = (\alpha \cdot \Re L_{am} + \beta)^\gamma. \tag{6}$$

Finally, a nonlinear activation function $f(I_e(x, y))$ restricts pixel values within the range $[0, 1]$,

$$f(I_e(x, y)) = \begin{cases} 0 & I_e(x, y) < 0 \\ I_e(x, y) & 0 < I_e(x, y) < 1 \\ 1 & I_e(x, y) > 1. \end{cases} \tag{7}$$

Theoretically, Eq. 6 indicates that

i. Via transforming the original signal $I(x, y)$, α, β, and γ are introducing additional information to facilitate the detection procedure.
ii. with the help of nonlinear activation f in Eq. 7, constant β should be able to affect the method accuracy, because the distribution of extremely dark and white pixels in real and fake faces are explicitly different, for example, real hair is relatively dark but hair in attacks can be over-whitened by reflection.
iii. nonlinear α and β should outperform constant α and β, since real faces and attacks will be obviously more distinguishable with functions $z_x(I(x, y), C_x)$ than with constants C_x.

Visually, the image adjustments can amplify the differences between live and spoofing images:

i. With the brightness enhancement (Fig. 3c), the medium reflection in spoofing images, are obviously amplified and become easier to detect.
ii. Brightness enhancement can also serve to magnify moire artefacts. Note that moire patterns are important clues in video attacks. We observe that moire patterns are more noticeable in brighter conditions. An example of spoofing image with moire artefacts is shown in Fig. 4.
iii. Comparing with increasing brightness, the advantage of improving contrast is less obvious (Fig. 3b). We will demonstrate later that actually reducing contrast tends to produce higher accuracy.

Fig. 3 A visualization on nonlinear brightness and contrast adjustments (+ 35%). From top to bottom real, print attack and video attack from the CASIA-FASD data set, respectively

iv. Though improving contrast alone is not sufficient, contrast improvement can facilitate brightness enhancement to improve visual differentiation (Fig. 3d). Note the difference between the real images in Fig. 3c, d, especially on the hair regions.
v. The introduced adjust helps to amplified the artefacts in spoofing images without damaging the features in live images. As shown in Fig. 3, the live photographs remain realistic after the different adjustments.

3.3 Hair geometry

Real faces and attacks are different in that human faces are in three dimensions with depth information, while either video attacks or print-photograph attacks lose the 3D geometric property, leading to reflect lights in a different manner. To separate real from fake, existing approaches make use of the different shading appearances on faces. Nevertheless, most previous researches discard hair information, as explained in Fig. 2. Comparing with the limited variations on facial features, shading on 3D hair geometry is way more complicated due to large variations on hair textures, colors, and styles. Moreover, human hair has complex BRDF and local lighting effects [7], making it extraordinary hard for fraud attacks to maintain the original appearance. As a result, hair appearance can be an effective measure to verify a genuine face.

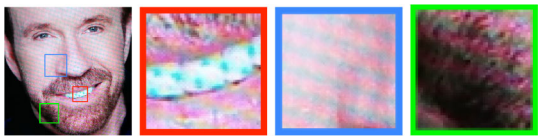


Fig. 4 Moire patterns are more noticeable in brighter part of the image

In Fig. 5, we visualize the feature maps of hair (without adjustment described in Sect. 3.2) after the first convolution layer of a three-layer CNN (Details of CNN architecture will be discussed later in Sect. 3.5). It can be seen clearly that hairs in spoofing images contain less fine geometry/features comparing with that in live photographs, which provides critical information during the detection process. In Fig. 6, we show the feature maps with nonlinear adjustment (+ 35% contrast and + 35% brightness). An important observation is that after adjustment, feature maps of hair in attacks are further smoothed while that in real photographs still preserve the noticeable fine hair strips. That details of hair remain in live photographs but not in attacks can facilitate to differentiate real photographs from fake ones, which we infer is the reason why the performance of using type Head outperforms that of type Face. Please zoom in for best visualization and refer to "Appendix" for feature map visualization of other attack types.

3.4 RGB color space

For that shading information is especially important for spoofing face detection, color spaces with independent luminance channels are expected to perform better. Unsurprisingly, most feature-based methods find that HSV and YCbCr color spaces are preferable that RGB, in which the three color components are highly correlated and the luminance and chrominance are inseparable. Most, if not all, existing CNN methods either keep the convention of traditional methods by using HSV and YCbCr or do not explicitly explore the performance of different color spaces.

Nevertheless, we observe that RGB color space can outperform HSV and YCbCr after the proposed adjustments operation in Sect. 3.2. We suppose the reason might be that real faces and attacks reflect different amount of red, green and blue component. The dependency between luminance and chrominance in RGB color space complexes the problem but simultaneously, introduces more combination possibilities of luminance and chrominance. Via amplifying the distinguishable features with the proposed adjustment, such complicated differences can be better comprehended by CNN but are troublesome for handcraft features to understand.

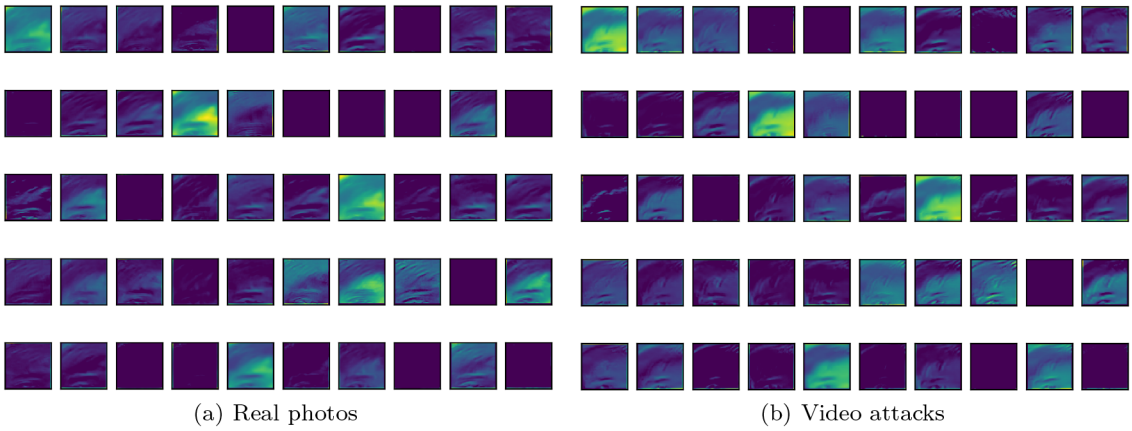


Fig. 5 Visualization of feature maps of hair without adjustment

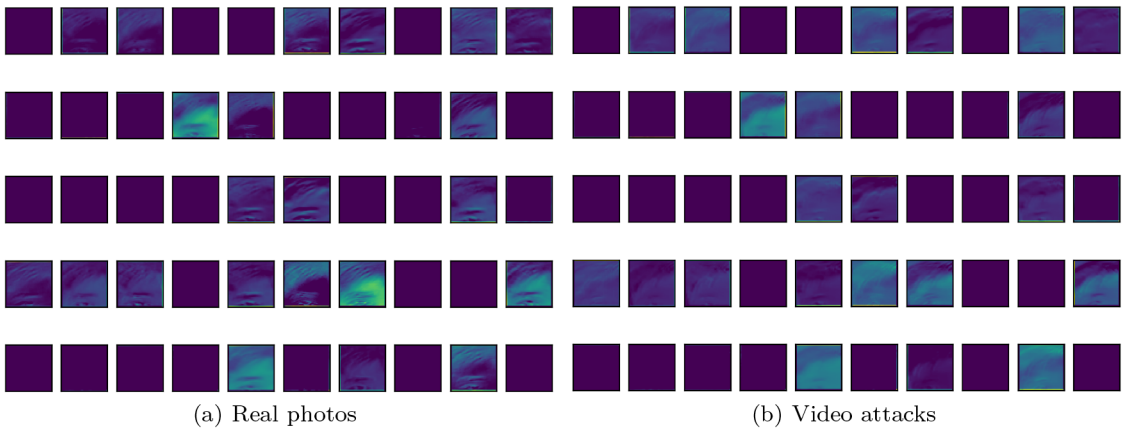


Fig. 6 Visualization of feature maps of hair with nonlinear + 35% contrast adjustment and + 35% brightness adjustment

3.5 Face anti-spoofing using convolutional neural network

Thanks to the data characteristics we described above, a simple CNN can solve the problem effectively. Our network design is demonstrated in Fig. 7. The feature descriptor consists of N blocks, each of which is the combination of a convolutional layer with RELU activation and a feature pooling operation. In experiments, we find that $N = 3$ is sufficient. Following the feature descriptor, a FC layer retrieves the deep histogram of the feature maps, which is then use by another FC layer for fake and real detection. The categorical cross-entropy loss is used to train the network,

$$L = -\sum_i \sum_j t_{i,j} \log(p_{i,j}), \tag{8}$$

where p are the predictions, t are the ground-truth labels, i denotes the data point and j denotes either real faces or attacks.

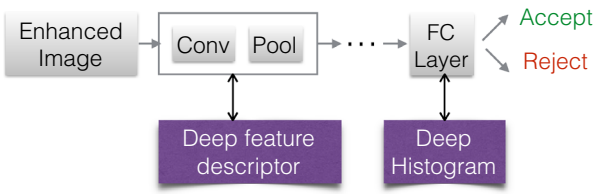


Fig. 7 Network architecture

Setup Since the three benchmarks we evaluate are different in size, we slice the data differently. For the CASIA data set [52], we resize images to 128 × 128 and extract patches from each image with stride 64 and patch size 64 × 64. It takes about 5 min to train the data set with 30 epochs. For the replay attack benchmark [13], images are resized to 96 × 96 and patches are extracted with stride 24 and size 24 × 24. It takes about 200 min to train with 50 epochs. For the MSU USSA database [37], we resize images to 256 × 256 and extract patches with stride 32 and size 64 × 64. It takes about 50 min to train with 50 epochs. For video data sets, we randomly retrieve 200 frames from each video. Patches from

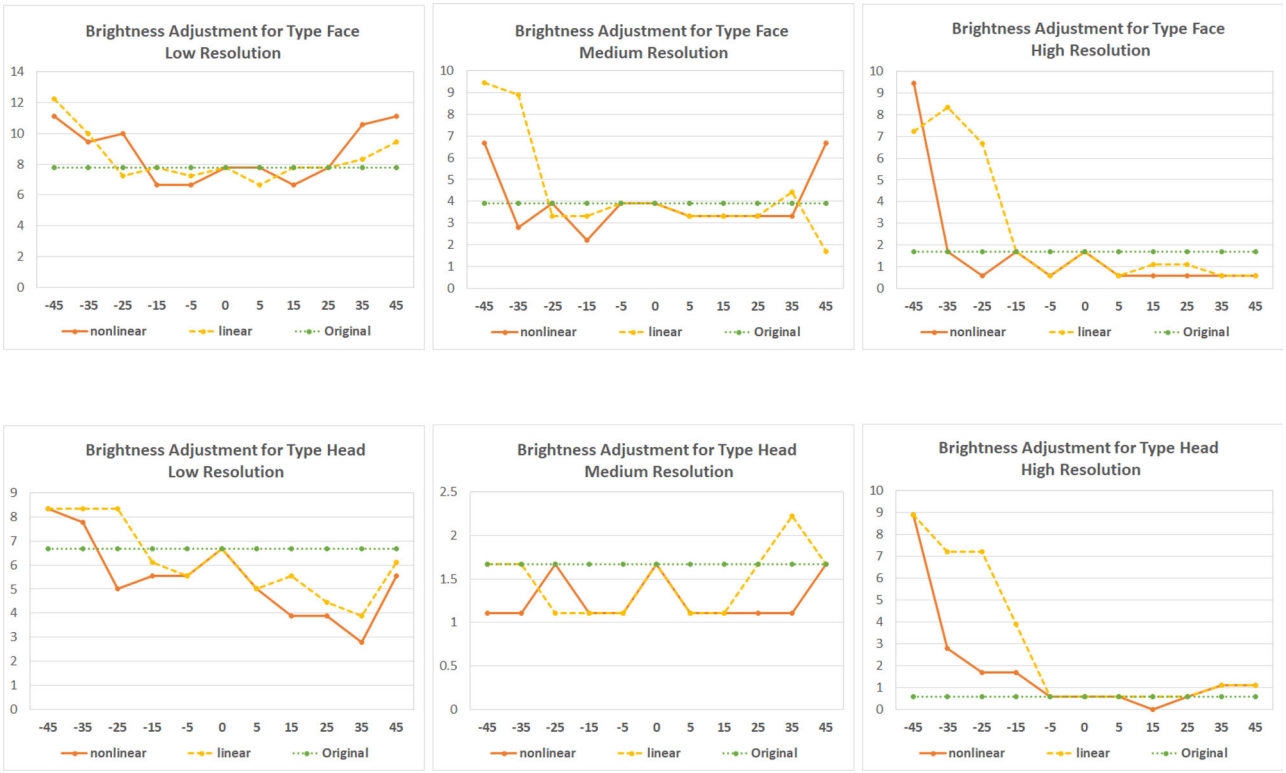


Fig. 8 Brightness adjustment. The x-axis and y-axis show the adjusting amount of β (%) and HTER error (%), respectively

live images/videos are labeled 1 and 0 otherwise. During testing, the predicted label for each video/image will be the average score of all its extracted patches.

Network parameters are learned via Adam [24], with learning rate 0.0001 and batchsize 64. We use 2×2 max pooling with stride 2, $50 \times i$ convolutional filters for block i ($i = 1 \ldots N$), 5×5 convolutional kernels in the first block and 3×3 convolutional kernels in the other blocks. The first FC layer contains 1000 units while the last FC layer consist of 2 units. Dropout with rate 0.5 is used. Our method is implemented using Lasagne on a NVIDIA GeForce GTX 1080 GPU.

4 Analysis

In this section, we investigate the influence of different factors on the CASIA-FASD benchmark [52]. When $\alpha = 1$, $\beta = 0$, and $\gamma = 1$, the image will be kept unchanged. Specifically, we refer $x\%$ adjustment on contrast to $C_\alpha = 1 + \frac{x}{100}$ while $y\%$ adjustment on brightness to $C_\beta = \frac{y}{100}$ and use the fast gain and bias function [42] to interpret nonlinear function z_α and z_β. From now on, we use $+35\%$ nonlinear enhancement on both contrast α and brightness β, no gamma correction ($\gamma = 1$), Head face type, RGB color space, and $N = 3$ blocks to produce the results, except otherwise stated.

4.1 Performance of adjustment strategies

Theoretically, we have explored the functionality of different adjustment in Sect. 3.2. In this section, we explore the influence of proposed strategies experimentally. Figures 8, 9 and 10 quantitatively verify the performance of brightness β, contrast α and Gamma γ adjustments, respectively. In each figure, the x axis indicates the amount of adjustment and the y-axis shows the corresponding HTER error (%).

Overall, the statistics indicates:

- For different adjusting amounts in each type of adjustment (brightness, contrast, or Gamma correction), using type Head (with hair information) produces higher accuracy than using type Face (without hair information).
- Within a reasonable range, both linear and nonlinear adjustments on brightness/contrast reduce the error. Nevertheless, nonlinearity outperforms linearity.
- Within a reasonable range, either improve or reduce brightness/contrast can improve the accuracy (or maintain the similar accuracy as the original). However, for brightness, a positive value is more preferred (improve brightness) while the opposite is true for contrast (reduce contrast).
- Adjusting brightness/contrast to the extreme would increase the error.

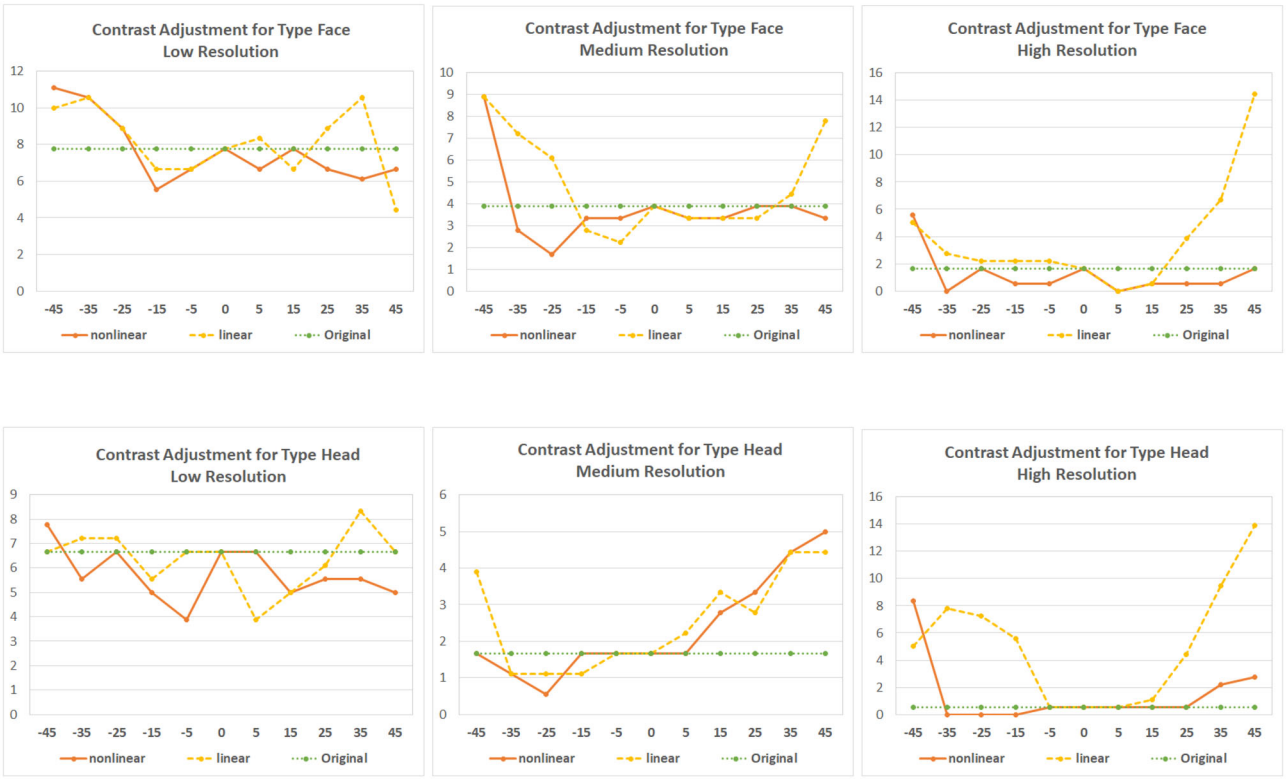


Fig. 9 Contrast adjustment. The x-axis and y-axis show the adjusting amount of α (%) and HTER error (%), respectively

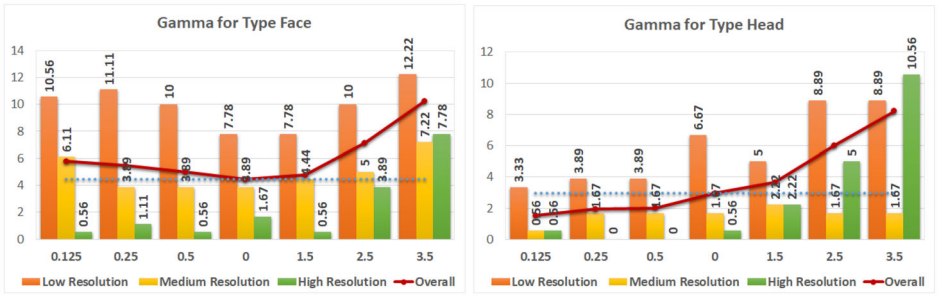


Fig. 10 Gamma correction. The x-axis and y-axis show the value of γ and HTER error (%), respectively. The blue line indicates the error produced by the original data for all image resolutions while the red curve indicates the errors for all image resolutions at every adjustment

– For type Head, Gamma correction that increases image brightness ($\gamma \in (0, 1]$) can reduce the error, which is accordance with the performance of brightness adjustment. But Gamma correction cannot facilitate to reduce error for type Face (Fig. 10).

In experiments, we find that nonlinearly adjusting both brightness and contrast for type Head, can further improve the performance, as demonstrated in Fig. 11.

– A large contrast improvement alone (+ 35) is not able to reduce the error greatly (Fig. 9), but with the help of a large brightness enhancement (+ 35), they together (+ 35, + 35) can achieve a better accuracy.
– Though greatly reducing brightness (− 35) would increase the error to some extent (Fig. 8), while with the assistance of contrast reduction (− 35), they together (− 35, − 35) can have a reasonable performance.
– Reduce contrast and improve brightness to the extreme, such as (− 35, 35), will damage the accuracy.

Generally speaking, on one hand, nonlinearly adjusting both contrast and brightness to the same direction for type Head is preferable; On the other hand, a positive direction tends to produce a lower error rate. Please refer to "Appendix" for the performance of joint nonlinear adjustment on type Face in Fig. 14 and joint linear adjustment in Fig. 15.

We visually demonstrates the influences of nonlinear adjustments on α and β in Fig. 3. Improving brightness amplifies the distortions on attacks, such as texture and color, while, contrast improvement does not have such impact. The best differentiation is achieved by combining them both. Note

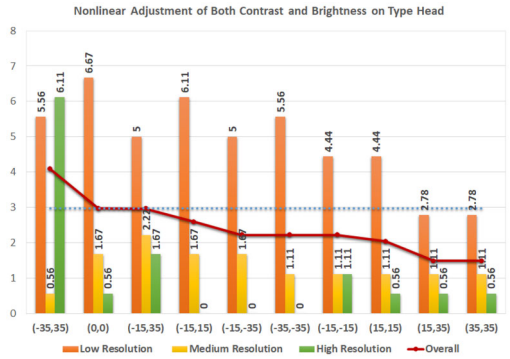


Fig. 11 Nonlinear contrast and brightness adjustment for type Head. The x-axis (α, β) indicates the amount of adjustment (%) and the y axis shows the HTER error (%). The blue line indicates the error produced by the original data for all image resolutions while the red curve indicates the errors for all image resolutions at every adjustment

Table 1 Influence of using different face types (HTER %)

Resolution	Face	Silhouette	Head
Low quality	9.44	5.0	2.78
Medium quality	5.0	1.67	1.11
High quality	0.56	1.67	0.56

that real hair color is over-whitened when increasing brightness alone [the top image on column (c)], which is resolved after introducing additional contrast enhancement [the top image of column (d)] (Fig. 3).

4.2 Influence of hair geometry

Table 1 demonstrates that effect of different face types. Similar to the findings in [37], we observe that making use of the entire face regions (Silhouette) can improve the detection accuracy except for faces in high resolution. However, as expected, the accuracy is further improve for Head face type, thanks to the hair clues. The reason why the most noticeable improvement is achieved on images in low resolution is that low quality faces are taken at a long distance with more loss on depth information, making live faces more "2D-like" and indistinguishable from the ones on attacks. Accordingly, the additional clues introduced by hair geometry contribute to reduce the detection error to a great extent.

Yang et al. [50] observes that CNN performance can be improved by including more background information (Fig. 12d), especially the photograph boundary (Fig. 12b). Nevertheless, a more recent study in [37] indicates that the boundary artifacts can only reduce a bit error and the background hardly improves the performance. Despite of the effectiveness of the artifacts, it is improper to tackle this problem with such non-biometrics, which can be intentionally prevented by malicious users. Different

Fig. 12 Yang et al. [50] uses larger portion of the images to include more background. The former two are from CASIA-FASD and the latter two are from replay attack data set

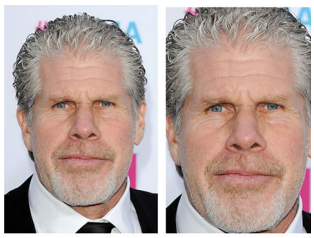


Fig. 13 The detection accuracy decreases when using images with more background information (**a**) rather than without background (**b**)

from [50], our type Head (Fig. 2c) leaves out the obvious artifacts and contains as little background as possible. Our approach outperforms [50] and verifies that hair information is superior to the background/ artifacts in face anti-spoofing.

We also test on the MSU USSA database. When using type Head, our HTER and ERR are 0.52% and 0.40% respectively, but if feeding images with more background information (as demonstrated in Fig. 13) to our CNN, the average HTER and ERR increase to 1.56%, and 1.79%, respectively.

4.3 Effects of color spaces

We then explore how different color spaces influence the network performance. In terms of producing better accuracy, YCbCr and HSV color spaces are preferred in traditional feature-based approaches. Nevertheless, Table 2 demonstrates that RGB color space can achieve the best performance after our data enhancement. Regardless of the different face types, either YCbCr or HSV defeats RGB without adjustment but RGB outperforms them both after enhancing contrast α and brightness β by $+$ 35%.

4.4 Deeper CNNs

Prior CNN-based methods rely on deep models to perform face anti-spoofing, but results in Table 3 indicate that many CNNs with deep architectures have poor performance, such as DPCNN [29] and 3DCNN [19]. We show in Table 4 that the accuracy improvement by deeper archi-

Table 2 Influence of different color spaces (HTER %)

Color space	Face		Head	
	Original	With adj	Original	With adj
YCbCr	3.89	6.29	3.7	4.26
HSV	6.30	5.74	2.59	3.15
RGB	4.45	5.0	2.97	**1.48**

For the results with adjustment, contrast and brightness are nonlinearly enhanced by + 35%
Bold value indicates the best results

tectures is limited. With the number of blocks increasing, the reduction of HTER error declines for all resolutions. Fusing feature maps from block 2, 3, and 4 cannot further improve the performance. We also attempt to replace the last max pooling operation before the FC layers with a spatial pyramid pooling operation, but the accuracy is not improved.

5 Comparison and discussion

We compare our approach with state of the arts on three well-known face anti-spoofing benchmarks. To evaluate all the three data sets, for our methods, we use the same set-ups as that in the experiments of Sect. 4: joint + 35% nonlinear adjustment on both brightness and contrast, type Head, RGB color space, and $N = 3$ blocks. For comparing methods, we use the statistics in the original papers.

5.1 Databases

CASIA-FASD Zhang et al. [52] is the most widely used database in evaluating face anti-spoofing performance. Each subject in the benchmark is taken 3 live videos with different face resolutions under uncontrolled lighting conditions and each live video is prepared with 3 attacks: warp print attack, cut print attack and replay attack. Due to the attacks

Table 3 A comparison between the proposed method and state-of-arts

Methods	CASIA-FASD		Replay-attack		MSU USSA	
	EER (%)	HTER (%)	EER (%)	HTER (%)	EER (%)	HTER (%)
Microtexture analysis [34]	18.2	10.9	13.9	13.8	–	–
Quality assessment [18]	32.4	–	–	15.2	–	–
Visual dynamics [46]	21.8	–	5.3 0	3.8 0	–	–
Spectral cubes [39]	14.0	–	–	2.8 0	–	–
Color LBP [3]	6.20	–	0.40	2.90	–	–
Distortion clue (2016) [37]	5.88	–	–	3.30	3.84	–
Color texture [4]	3.20	–	**0.00**	3.5	–	–
Videolet aggregation [43]	3.14	–	–	–	–	–
Fisher vector [5]	2.80	–	0.1	2.20	–	–
Dynamic texture recognition [53]	6.50	–	1.7	0.8	–	–
CNN [50]	4.92	–	2.14	–	–	–
LSTM-CNN [49]	5.17	5.93	–	–	–	–
DPCNN [29]	4.50	–	2.90	6.10	–	–
3DCNN [19]	–	11.37	–	0.042	–	–
Patch-based CNN [1]	4.44	3.78	2.5	1.25	0.55	0.41
Patch and depth CNN [1]	2.67	2.27	0.79	0.72	**0.35**	**0.21**
Our method	**1.85**	**1.48**	**0.00**	**0.00**	0.52	0.40

Bold values indicate the best results for each column

Table 4 Performance of deeper CNNs (HTER%)

Resolution	Without spatial pyramid pooling				With spatial pyramid pooling			
	2 Blocks	3 Blocks	4 Blocks	Fusion all	2 Blocks	3 Blocks	4 Blocks	Fusion all
Low quality	3.33	2.78	2.78	3.33	5.0	6.11	5.0	5.0
Medium quality	2.22	1.11	1.67	0.56	2.22	2.22	2.22	2.22
High quality	1.1	0.56	0.56	0.56	0.56	0.56	0.56	0.56
Overall	2.22	**1.48**	1.67	**1.48**	2.59	2.96	2.59	2.59

Bold values indicate the best results for resolution 'Overall'

taking facial motions into consideration, previous motion-based approaches will fail on this benchmark, such as eye blinking [35]. It contains 50 subjects in total, with 20 for training and 30 for testing.

Replay attack Chingovska et al. [13] consists of 1300 videos for 50 subjects, including all live and spoofing videos. Both print and replay attacks are covered in this set. Comparing with CASIA-FASD, this data set is much larger in size but is collected under controlled illuminating conditions and backgrounds. The 50 subjects are divided into train, develop, and test set with 15, 15, 20 identities, respectively.

MSU USSA Patel et al. [37] is a recent data set, collecting live images of 1040 in-the-wide celebrities, that are highly diverse in resolution, illumination, chrominance, and so on. It contains 8 different attacks from computer, smartphones, tablet and printed papers.

5.2 Quantitative comparison

We quantitatively compare our method with various state of the arts in Table 3, including CNN [50], LSTM-CNN [49], DPCNN [29], Color LBP [3], Visual Dynamics [46], Spectral Cubes [39], Color Texture [4], Videolet Aggregation [43], Distortion Clue [37], Fisher Vector [5] and Patch and Depth CNN [1]. For fairness, we use the same protocols as specified by the data sets to evaluate our performance. For the first two benchmarks, we use the training sets to learn the network parameters and the testing set to compute the error rates, while a subject-exclusive five fold cross validation for MSU USSA. Replay-attack provides an additional develop set, which we use as a validation set. We calculate ERR on the test set and use the corresponding threshold to compute HTER error. Note that since the MSU USSA is up to date, only a few methods evaluate on it.

We show the comparison in Table 3. Due to the limitations of the representation power, the performance of traditional methods differs on different data sets, such as Color LBP [3]. it is also clear that many feature-based methods, especially the recent Fisher Vector [5], are superior to many CNN methods, such as CNN [50], LSTM-CNN [49], DPCNN [29] and Patch CNN [1], indicating that learning strategy alone is insufficient to well solve the problem. With the assistant of a depth-based CNN, Patch and Depth CNN [1] achieves to reduce the error of using patch CNN alone but still produces higher errors than ours in CASIA and replay attack database. Note that all the CNN-based methods in the comparison use complex and deep model architectures. With a better understanding on the spoofing data, our method can producing competing results on all the three benchmarks with a very simple network design.

5.3 Limitations

Our method acts on image brightness and contrast. Thus, if a training set does not contains enough luminance variation or is taken under different lighting condition from that of the test set, our detection accuracy might be reduced. The Oulu database [6] is the case where images in the training set are taken under one lighting condition while images in the test set are taken under another lighting condition. Our method is not designed to handle such situation. However, if the training set contains enough luminance variations, such as the in-the-wild MSU USSA database, our method can still produce reasonable results.

6 Conclusion and future work

To conclude, we present a systematic study on face anti-spoofing by investigating different image properties. To improve the detection accuracy, data and learning models are both indispensable. Prior CNN methods focus more on exploring different networks but neglecting the importance of data. With a thorough analysis on the spoofing data, we make a simple network architecture achieve state-of-the-art performance. As a future work, We would like to find a way to simulate the data enhancement procedure as a layer in the network, so that a network can be trained from end to end.

Compliance with ethical standards

Funding Prof. Xiu Ming Yiu has received research Grants from the Hong Kong Government.

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

See Figures 14, 15 and 16.

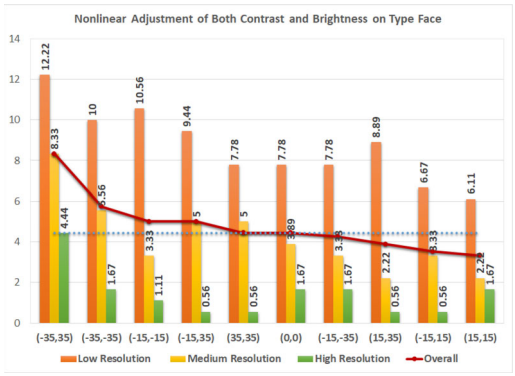


Fig. 14 Nonlinear contrast and brightness adjustment for type Face. The x axis (α, β) indicates the amount of adjustment (%) and the y axis shows the HTER error (%). The blue line indicates the error produced by the original data for all image resolutions while the red curve indicates the errors for all image resolutions at every adjustment

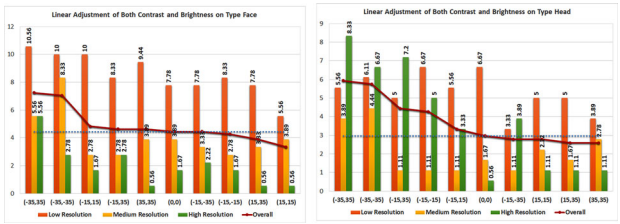
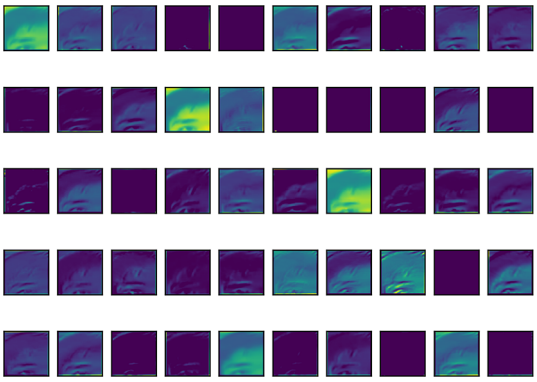
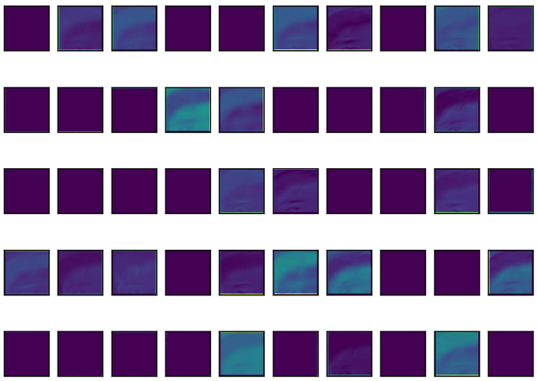


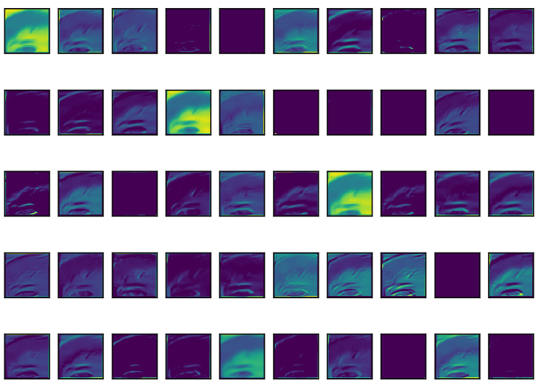
Fig. 15 Linear contrast and brightness adjustment. The blue line indicates the error produced by the original data for all image resolutions while the red curve indicates the errors for all image resolutions at every adjustment



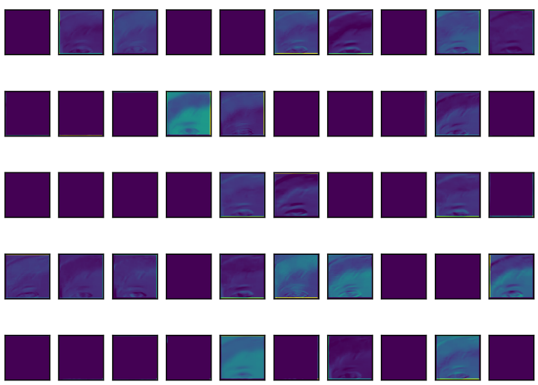
(a) Paper attacks without adjustment



(b) Paper attacks with adjustment



(c) Curved paper attacks without adjustment



(d) Curved paper attacks with adjustment

Fig. 16 Visualization of hair feature maps in paper attacks and curved paper attacks. For the ones with adjustment, the adjustment is + 35% nonlinear improvement on both contrast and brightness

References

1. Atoum, Y., Liu, Y., Jourabloo, A., Liu, X.: Face anti-spoofing using patch and depth-based CNNs. In: 2017 IEEE International Joint Conference on Biometrics (IJCB), pp. 319–328. IEEE (2017)
2. Bharadwaj, S., Dhamecha, T.I., Vatsa, M., Singh, R.: Computationally efficient face spoofing detection with motion magnification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 105–110. IEEE (2013)
3. Boulkenafet, Z., Komulainen, J., Hadid, A.: Face anti-spoofing based on color texture analysis. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 2636–2640. IEEE (2015)
4. Boulkenafet, Z., Komulainen, J., Hadid, A.: Face spoofing detection using colour texture analysis. IEEE Trans. Inf. Forensics Secur. **11**(8), 1818–1830 (2016)
5. Boulkenafet, Z., Komulainen, J., Hadid, A.: Face antispoofing using speeded-up robust features and fisher vector encoding. IEEE Signal Process. Lett. **24**(2), 141–145 (2017)
6. Boulkenafet, Z., Komulainen, J., Li, L., Feng, X., Hadid, A.: OULU-NPU: a mobile face presentation attack database with real-world variations (2017)
7. Chai, M., Luo, L., Sunkavalli, K., Carr, N., Hadap, S., Zhou, K.: High-quality hair modeling from a single portrait photo. ACM Trans. Graphics (TOG) **34**(6), 204 (2015)
8. Chan, P.P., Liu, W., Chen, D., Yeung, D.S., Zhang, F., Wang, X., Hsu, C.C.: Face liveness detection using a flash against 2D spoofing attack. IEEE Trans. Inf. Forensics Secur. **13**(2), 521–534 (2018)
9. Chetty, G.: Biometric liveness detection based on cross modal fusion. In: 12th International Conference on Information Fusion, 2009. FUSION'09, pp. 2255–2262. IEEE (2009)
10. Chetty, G.: Biometric liveness checking using multimodal fuzzy fusion. In: 2010 IEEE International Conference on Fuzzy Systems (FUZZ), pp. 1–8. IEEE (2010)
11. Chetty, G., Wagner, M.: Multi-level liveness verification for face-voice biometric authentication. In: 2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference, pp. 1–6. IEEE (2006)
12. Chetty, G., Wagner, M.: Biometric person authentication with liveness detection based on audio-visual fusion. Int. J. Biom. **1**(4), 463–478 (2009)
13. Chingovska, I., Anjos, A., Marcel, S.: On the effectiveness of local binary patterns in face anti-spoofing. In: 2012 BIOSIG-Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–7. IEEE (2012)
14. Dhamecha, T.I., Nigam, A., Singh, R., Vatsa, M.: Disguise detection and face recognition in visible and thermal spectrums. In: 2013 International Conference on Biometrics (ICB), pp. 1–8. IEEE (2013)
15. de Freitas Pereira, T., Anjos, A., De Martino, J.M., Marcel, S.: Lbp-top based countermeasure against face spoofing attacks. In: Asian Conference on Computer Vision, pp. 121–132. Springer (2012)
16. de Freitas Pereira, T., Anjos, A., De Martino, J.M., Marcel, S.: Can face anti-spoofing countermeasures work in a real world scenario? In: 2013 International Conference on Biometrics (ICB), pp. 1–8. IEEE (2013)
17. Erdogmus, N., Marcel, S.: Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect. In: Biometrics: 2013 IEEE Sixth International Conference on Theory, Applications and Systems (BTAS), pp. 1–6. IEEE (2013)
18. Galbally, J., Marcel, S., Fierrez, J.: Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition. IEEE Trans. Image Process. **23**(2), 710–724 (2014)
19. Gan, J., Li, S., Zhai, Y., Liu, C.: 3D convolutional neural network based on face anti-spoofing. In: 2017 2nd International Conference on Multimedia and Image Processing (ICMIP), pp. 1–5. IEEE (2017)
20. Garcia, D.C., de Queiroz, R.L.: Face-spoofing 2D-detection based on moiré-pattern analysis. IEEE Trans. Inf. Forensics Secur. **10**(4), 778–786 (2015)
21. Jourabloo, A., Liu, Y., Liu, X.: Face de-spoofing: Anti-spoofing via noise modeling. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 290–306 (2018)
22. Karsch, K., Liu, C., Kang, S.: Depth transfer: depth extraction from video using non-parametric sampling. IEEE Trans. Pattern Anal. Mach. Intell. **36**(11), 2144–2158 (2014)
23. Kim, S., Ban, Y., Lee, S.: Face liveness detection using a light field camera. Sensors **14**(12), 22471–22499 (2014)
24. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2014). arXiv preprint arXiv:1412.6980
25. Komulainen, J., Hadid, A., Pietikainen, M.: Context based face anti-spoofing. In: 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–8. IEEE (2013)
26. Lagorio, A., Tistarelli, M., Cadoni, M., Fookes, C., Sridharan, S.: Liveness detection based on 3D face shape analysis. In: 2013 International Workshop on Biometrics and Forensics (IWBF), pp. 1–4. IEEE (2013)
27. Li, H., He, P., Wang, S., Rocha, A., Jiang, X., Kot, A.C.: Learning generalized deep feature representation for face anti-spoofing. IEEE Trans. Inf. Forensics Secur. **13**(10), 2639–2652 (2018)
28. Li, J., Wang, Y., Tan, T., Jain, A.K.: Live face detection based on the analysis of Fourier spectra. In: Biometric Technology for Human Identification, vol. 5404, pp. 296–304. International Society for Optics and Photonics (2004)
29. Li, L., Feng, X., Boulkenafet, Z., Xia, Z., Li, M., Hadid, A.: An original face anti-spoofing approach using partial convolutional neural network. In: 2016 6th International Conference on Image Processing Theory Tools and Applications (IPTA), pp. 1–6. IEEE (2016)
30. Li, Y., Xu, K., Yan, Q., Li, Y., Deng, R.H.: Understanding OSN-based facial disclosure against face authentication systems. In: Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security, pp. 413–424. ACM (2014)
31. Liu, P., Zafar, F., Badano, A.: The effect of ambient illumination on handheld display image quality. J. Digit. Imaging **27**(1), 12–18 (2014)
32. Liu, Y., Jourabloo, A., Liu, X.: Learning deep models for face anti-spoofing: binary or auxiliary supervision (2018). arXiv preprint arXiv:1803.11097
33. Liu, Y., Stehouwer, J., Jourabloo, A., Liu, X.: Deep tree learning for zero-shot face anti-spoofing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4680–4689 (2019)
34. Määttä, J., Hadid, A., Pietikäinen, M.: Face spoofing detection from single images using micro-texture analysis. In: 2011 international Joint Conference on Biometrics (IJCB), pp. 1–7. IEEE (2011)
35. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In: IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007, pp. 1–8. IEEE (2007)
36. Patel, K., Han, H., Jain, A.K.: Cross-database face antispoofing with robust feature representation. In: Chinese Conference on Biometric Recognition, pp. 611–619. Springer (2016)
37. Patel, K., Han, H., Jain, A.K.: Secure face unlock: spoof detection on smartphones. IEEE Trans. Inf. Forensics Secur. **11**(10), 2268–2283 (2016)
38. Peixoto, B., Michelassi, C., Rocha, A.: Face liveness detection under bad illumination conditions. In: 2011 18th IEEE International Conference on Image Processing (ICIP), pp. 3557–3560. IEEE (2011)

39. Pinto, A., Pedrini, H., Schwartz, W.R., Rocha, A.: Face spoofing detection through visual codebooks of spectral temporal cubes. IEEE Trans. Image Process. **24**(12), 4726–4740 (2015)
40. Pinto, A., Schwartz, W.R., Pedrini, H., de Rezende Rocha, A.: Using visual rhythms for detecting video-based facial spoof attacks. IEEE Trans. Inf. Forensics Secur. **10**(5), 1025–1038 (2015)
41. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3D classification and segmentation. In: Proceedings of Computer Vision and Pattern Recognition (CVPR). vol. 1, number 2, p. 4. IEEE (2017)
42. Schlick, C.: Fast alternatives to Perlin's bias and gain functions. Graph. Gems IV **4**, 401–403 (1994)
43. Siddiqui, T.A., Bharadwaj, S., Dhamecha, T.I., Agarwal, A., Vatsa, M., Singh, R., Ratha, N.: Face anti-spoofing with multifeature videolet aggregation. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 1035–1040. IEEE (2016)
44. Socolinsky, D.A., Selinger, A., Neuheisel, J.D.: Face recognition with visible and thermal infrared imagery. Comput. Vis. Image Underst. **91**(1–2), 72–114 (2003)
45. Tan, X., Li, Y., Liu, J., Jiang, L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: European Conference on Computer Vision, pp. 504–517. Springer (2010)
46. Tirunagari, S., Poh, N., Windridge, D., Iorliam, A., Suki, N., Ho, A.T.: Detection of face spoofing using visual dynamics. IEEE Trans. Inf. Forensics Secur. **10**(4), 762–777 (2015)
47. Wang, T., Yang, J., Lei, Z., Liao, S., Li, S.Z.: Face liveness detection using 3D structure recovered from a single camera. In: 2013 International Conference on Biometrics (ICB), pp. 1–6. IEEE (2013)
48. Wen, D., Han, H., Jain, A.K.: Face spoof detection with image distortion analysis. IEEE Trans. Inf. Forensics Secur. **10**(4), 746–761 (2015)
49. Xu, Z., Li, S., Deng, W.: Learning temporal features using LSTM-CNN architecture for face anti-spoofing. In: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), pp. 141–145. IEEE (2015)
50. Yang, J., Lei, Z., Li, S.Z.: Learn convolutional neural network for face anti-spoofing (2014). arXiv preprint arXiv:1408.5601
51. Yang, J., Lei, Z., Liao, S., Li, S.Z.: Face liveness detection with component dependent descriptor. In: 2013 International Conference on Biometrics (ICB), pp. 1–6. IEEE (2013)
52. 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)
53. Zhao, X., Lin, Y., Heikkilä, J.: Dynamic texture recognition using volume local binary count patterns with an application to 2D face spoofing detection. IEEE Trans. Multimed. **20**(3), 552–566 (2018)

Yujing Sun is currently a Postdoctoral Fellow at the University of Hong Kong. She received a bachelor's degree from University of Minnesota, Twin Cities in 2013 and a PhD. in Computer Science from the University of Hong Kong in 2018. Her research interests include image processing, biometrics and Financial visualization.

Hao Xiong is an adjunct assistant professor at the University of Hong Kong. He received a bachelor's degree from Sun Yat-Sen University in 2010 and a PhD. in computer science from the University of Hong Kong in 2013. His research interests include cryptography and blockchain.

Siu Ming Yiu is a full Professor at the University of Hong Kong and the director of Fintech lab. He has Published 100+ papers in referred journals and is being listed as one of the Highly Cited Researchers (globally) in 2019 by the Web of Science Group. He is Conference/program chairs in prestigious conferences in both areas of cryptography and bioinformatics.