Real-time Multiclass Face Spoofing Recognition Through Spatiotemporal Convolutional 3D Features

Salvatore Giurato[0000−0002−7230−2425], Alessandro Ortis[0000−0002−7230−2425], and Sebastiano Battiato[0000−0001−6127−2470]

Image Processing Laboratory, Dipartimento di Matematica e Informatica, Universita' degli studi di Catania, Viale A. Doria 6, Catania - 95125 Italy
salvatore.giurato@phd.unict.it {ortis, battiato}@dmi.unict.it

Abstract. Face recognition is used in numerous authentication applications, unfortunately they are susceptible to spoofing attacks such as paper and screen attacks. In this paper, we propose a method that is able to recognise if a face detected in a video is not real and the type of attack performed on the fake video. We propose to learn the temporal features exploiting a 3D Convolution Network that is more suitable for temporal information. The 3D ConvNet, other than summarizing temporal information, allows us to build a real-time method since it is so much more efficient to analyse clips instead of analyzing single frames. The learned features are classified using a binary classifier to distinguish if the person in the clip video is real (i.e. live) or not, multi class classifier recognises if the person is real or the type of attack (screen, paper, ect.). We performed our test on 5 public datasets: Replay Attack, Replay Mobile, MSU-MSFD, Rose-Youtu, RECOD-MPAD.

Keywords: Antispoofing Attack · 3D Features · Multi-Class detection · liveness.

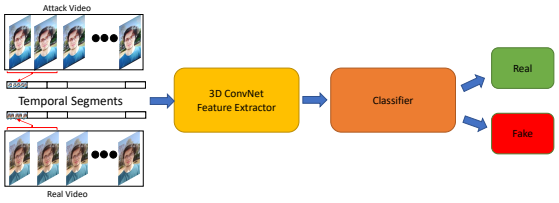
1 Introduction

Face Recognition is a biometric system that has been deployed in real life applications such as recognizing people's identity. Face Recognition made significant progress during the past years with DeepFace [13], DeepIDs [14], VGG Face [15], FaceNet [16], SphereFace [17] and ArcFace [18]. However the more Face Recognition is popular the newer spoofing attacks appears. Common attacks can be categorized as video replay attacks, photo attacks, and 3D mask attacks. Common Face Recognition methods are unable to detect the difference between real faces and attack faces. In literature there are different approaches on how to distinguish real face or attack face. One approach is to detect the motion in a video in order to prevent photo attack, Li et al. [19] used the Fourier spectra to estimate the temporal changes due to the motion. In contrast to Li et al. [19], Kollreider et al [20] and Kollreider et al. [21] worked on RGB space instead of frequency domain to estimate respectively the 2D and 3D motion. The work in

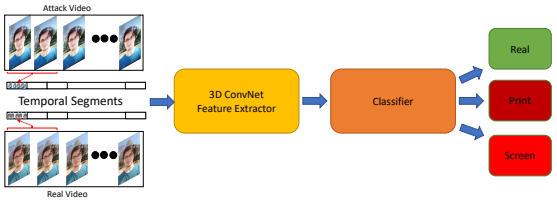
Pan et al. [22] and Sun et al. [23] exploited the human physiological behaviour to detect anomalies in eye blinking, in case of anomalies the video is considered a spoofing attack. Bao et al. [24] used optical flow to distinguish between 3D photo and planar photo attack. A different approach was proposed by Li et al. [25] and Nowara et al. [26] that exploit remote PhotoPlethysmoGraphy (rPPG) that detect blow flood using RGB images, this method is able to detect photo based and 3D mask attack. The first work that used a machine learning approach through Convolutional Neural Networks (CNN) was Yang et al. [27] where an AlexNet [9] architecture with the last layer replaced by an SVM binary classifier was employed. Patel et al. [28] also used an AlexNet architecture replacing the 1000-way SoftMax with a binary classifier, in this case the network is pretrained on ImageNet [11] and WebFace [12] and fine tuned on the facial spoofing dataset. Li et al. [29] proposed a CNN based on VGG-Face [15] pretrained on massive dataset and fine-tuned. George et al. [30] proposed Deep Pixelwise Binary Supervision (DeepPixBiS) based on DenseNet [10] that uses two losses during the training. While the motions method works on the video the convolutional methods works on the single frame of the video, it means that they are so much more time consuming. To reduce the complexity of the video analysis Tran et al. [7] proposed a 3D Convolutional Network to solve the action recognition problem. Sultani et al. [8] exploited the model of Tran et al. [7] to build a new model for anomaly detection. Sultani et al. [8] using the 3D Convolutional neural network trained by Tran et al. [7] obtained optimal results on the anomaly detection problem. Our proposed approach is to build a novel method to detect anti-spoofing attacks exploiting 3D Convolutional Network, and then we will present our studies of a novel method that is able to distinguish the type of attack. To the best of our knowledge this is the first work that recognise whether the video is real or a print attack or replay attack or mask attack. To exploit well the 3D Convolutional Network we divide the videos in clips, from now on we will refer as segment when we talk about clips. The problem of spoofing detection still does not have a benchmark dataset, some dataset that has been used in previous works are not available, some others are new so they have not been tested in newer works. There is not an anti-spoofing dataset in the wild which means that all the videos in all datasets are collected in a protected environment, this means that each dataset has different characteristics. For this reason, we decided to work on 5 different datasets: Idiap Replay Attack [1], Idiap Replay Mobile [2], Recod-MPAD [3], MSU-MFSD [4], Rose-Youtu Dataset [5][6].

2 Proposed Method

This section presents the keystone of our approach. First, the videos are segmented in temporal video clips of a fixed duration, to properly analyse the temporal changes in the scenes. As previously mentioned, face detection and face recognition are deployed in different real-life applications. Once the face is detected, a recognition algorithm is applied to the portion of image depicting the face, discarding the background information. In order to get the same



(a) Binary Classification



(b) Multi-class classification

Fig. 1: Considered binary and multiclass pipelines.

information of the face recognition algorithm, the Face Detection procedure is performed on each frame of each segment. The proposed method exploits two neural network or machine learning models: the first network extracts the feature for each segment (i.e. short video clip), the second model is a binary classifier or a multiclass classifier. The first network is a pretrained 3D Convolutional Network [7], by which the features of the segments are extracted, in this way it is generated a 3D array of features that contains a single feature representing a segment of video. The 3D Convolutional Network is fed with a segment in order to get the temporal features, which help reducing the computational complexity of the method, since a single segment is able to summarise the information of a certain number of frames. Such features of segments represent instances in a bag of features, which will be fed to a classifier. Once all the features of the segments are extracted the second network is trained for spoofing attack detection in a binary or multiclass classification setting. In the binary classifier we consider the features bag as negative for real videos or as positive for attack videos, regardless the type of attack. In the multiclass classifier we consider the features bag as negative (0) for real videos, and as positive screen (1) for screen attack videos and as positive print (2) for print attack videos, and as positive mask (3) for mask attack videos. It is also possible to compute an interpolation to the feature bag to double the features dimensions and try to get more information from them. The sequence of features extracted from a video segment can be eventually interpolated. The result is a sequence of features doubled in number. Such a technique is successfully applied in [8] to detect anomalies in videos. The underlying idea is that the interpolation (i.e., average) of two temporally subsequent features results in a more smooth feature, in which anomaly signals may emerge. In the context of our work, we expect that feature anomalies in attack

attempts videos may be detected in a similar way. After the classification of the interpolated network to have only one result as in the other classifiers, or a different number of results, for each feature bag, it is computed the extrapolation process. In Figure 1a and 1b it is possible to observe the pipeline of how our method works.

Table 1: Information of the dataset (Attack: Print (P), Screen (S), Mask (M))

Dataset	Train	Segment Train	Test	Segment Test	Attack
	Real/Fake	Real/Fake	Real/Fake	Real/Fake	
Replay Attack	60/300	1380/4200	80/400	1840/5600	P/S
Replay Mobile	120/192	2108/3413	110/192	1931/3399	P/
Recod-MPAD	250/1000	824/1185	200/800	670/952	P/S
MSU-MFSD	30/90	516/1492	40/120	671/1986	P/S
Rose-Youtu	147/398	1081/4989	171/468	1521/5229	P/S/M

3 Evaluation

3.1 Dataset and Evaluation Metrics

We evaluated the performance of our model with 5 known datasets: Idiap Replay Attack [1], Idiap Replay Mobile [2], Recod-MPAD [3], MSU-MFSD [4], Rose-Youtu Dataset [5][6]. In Table 1 a summary of the details about the datasets. The videos are divided in train/test sets as recommended by the authors of each dataset. It is worth highlighting that there is no person that appears in both train and test sets. Note that the chosen extraction process forces the frames of the video to be resized in 112×112 format before the feature extraction. In both binary and multiclass classifiers we build a confusion matrix from the classification. From the confusion matrix we extract the value of True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) Therefore, we consider:

- True Positive (TP) the elements correctly classified as "Attack" in the binary classifier and as "Print" and "Screen" in the multiclass classifier.
- True Negative (TN) the elements correctly classified as "Real" in both binary classifier and multiclass classifier.
- False Positive (FP) the elements incorrectly classified as "Attack" in the binary classifier and as "Print" and Screen" in the multiclass classifier.
- False Negative (FN) the elements incorrectly classified as "Real" in both binary classifier and multiclass classifier.

The metrics we used to evaluate our classification are:

– Accuracy represents the percentage of correct predictions in the classification:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}. \quad (1)$$

– False Acceptance Rate is the ratio of False Positive predictions:

$$FAR = \frac{FP}{FP + TP}. \quad (2)$$

– False Rejection Rate is the ratio of False Negative predictions:

$$FRR = \frac{FN}{FN + TN}. \quad (3)$$

– HTER (Half Total Error Rate) [38] is a measure of the error, it is one of the most common metrics in biometric system to evaluate the system performance:

$$HTER = \frac{FAR + FRR}{2}. \quad (4)$$

Another metric similar to HTER is EER (Equal Error Rate) that is the threshold where FAR = FRR, however this metric can be used only to evaluate a training, but cannot be used to measure the performance of a model [39].

Table 2: Results of binary classifiers on All Datasets

Method	Binary		Multiclass	
	Acc	HTER	Acc	HTER
K-NN	75.7	29.1	65.9	36.7
LDA	80.3	24.7	**76.7**	**26.4**
SVM linear	78.8	27.0	73.6	30.0
Random Forest	**82.8**	**19.8**	74.2	27.7
Decision Tree	69.7	37.8	59.8	43.7
GBoost	79.5	26.1	75.6	27.0
3L-NN	79.8	25.7	/	/

3.2 Implementation Details

To start the pre-processing of the videos, we detect the face in each frame using FaceNet [16]. After the detection, the face is cropped from the image, and we store in the memory only the face for each frame. To continue the faces are resized in 112×112 and stored in segments containing 16 frames each. We extract features from the fully connected (FC) layer FC6 of the C3D network [7], which have a feature of dimensions of 4096. As a classifier we used different methods:

- K-NN with K = 3
- LDA
- SVM using the linear kernel
- Random Forest
- Decision Tree
- GBoost
- 3L-NN: A 3 Fully Connected Neural Network, each fully connected layer is followed by a dropout layer. The output of the 3 Fully Connected layers is respectively 512, 32 and 1. The last output is passed through a sigmoid function to convert to 0 and 1. The optimizer used is Adam with a learning rate = 0.001.

Table 3: Binary Classifier intra-dataset analysis:Replay Attack (I-RA), Replay-Mobile (I-RM), MSU-MFSD (MFSD), Recod-MPAD (MPAD), Rose-Youtu (Y)

Method	I-RA		I-RM		MFSD		MPAD		Y	
	Acc	HTER	Acc	HTER	Acc	HTER	Acc	HTER	Acc	HTER
K-NN	87.8	16.5	86.6	14.6	72.9	35.0	68.3	30.7	84.3	22.3
LDA	**91.4**	11.0	**93.7**	**6.3**	67.5	36.0	68.9	31.8	83.0	24.2
SVM linear	90.6	16.5	91.8	7.1	80.8	25.3	77.3	23.1	**84.5**	**21.3**
Random Forest	87.0	**9.0**	89.2	11.1	75.5	32.3	76.1	24.2	84.3	21.5
Decision Tree	83.9	19.4	78.0	23.6	67.1	40.0	64.5	36.2	78.9	30.4
GBoost	87.5	14.9	88.6	12.3	72.5	35.2	**77.8**	**22.6**	82.5	24.4
3L-NN	88.29	11.1	88.5	12.2	78.2	29.0	66.2	35.0	84.0	21.8
3L-NN$_{interp}$	91.2	11.7	88.3	8.7	**81.2**	**22.4**	72.7	26.2	83.1	**21.3**

3.3 Binary Classifier

In this section we summarise the experiments made with the binary classifier. For the sake of comparison, we performed the classification on the features of the video with different classifiers (K-NN, Support Vector Machine, Random Forest, Decision Tree, GBoost, 3L-NN). In Table 2 there is the comparison between these methods and a simple binary classifier. To perform a fair comparison the features are extracted from C3D Network, all these methods are trained with the train set of all the dataset, to get the results shown in the table the test set is used on the trained models. However, considering all the dataset together is not a good idea because they are from different protected environments, and they contain different information so this may lead a model to be more error prone. In addition, the presence of video attack in Dataset is higher than the presence of real video, this may cause overfit in models and networks. In Table

3 there are the results of the intra-dataset analysis using different datasets and different classifiers. As we can see for each dataset the classifiers have a different behaviour, this is because of the difference between the datasets. From the results we notice that some dataset are "easier" to generalise with the classifier such as Replay-Attack and Replay Mobile and more difficult to generalise such as MSU-MFSD and Recod-MPAD. In bold are highlighted the best results for each dataset.

Table 4: Multi Class Classifier: Replay Attack (I-RA), Replay-Mobile (I-RM), MSU-MFSD (MFSD), Recod-MPAD (MPAD), Rose-Youtu (Y)

Method	I-RA		I-RM		MFSD		MPAD		Y	
	Acc	HTER	Acc	HTER	Acc	HTER	Acc	HTER	Acc	HTER
K-NN	70.5	28.7	80.9	19.2	65.5	40.1	62.6	36.3	85.3	15.3
LDA	**87.0**	**13.1**	**92.1**	**6.8**	66.5	38.2	61.6	36.8	87.6	14.5
SVM linear	83.6	15.7	91.5	7.7	**74.1**	**31.4**	71.7	28.4	**87.9**	**13.2**
Random Forest	77.3	16.4	87.2	13.8	69.3	36.4	68.7	31.4	87.0	14.2
Decision Tree	69.5	28.9	67.9	32.5	61.3	45.1	58.3	40.2	87.0	14.2
GBoost	81.0	17.2	88.7	11.8	70.4	35.0	**71.9**	**28.2**	79.3	23.8

3.4 Multiclass Classifier

In this section we summarise the experiments made with the multiclass classifier. We performed the classification on the features of the video with different classifiers (i.e. K-NN, Support Vector Machine, Random Forest, Decision Tree, GBoost). This is the first time that a method is able to distinguish the type of attacks. In Table 2 there is the comparison between different models, to perform a fair comparison the features are extracted from C3D Network, all these methods are trained on with the train set of all Dataset, to get the results we use the test set. In Table 4 there are the results on intra-dataset analysis on different models using different multiclass classifiers, the aim is to distinguish different types of attack. As we can see the classifiers work in a different way, for each dataset. However, we observe from Table 3 and Table 4 that the behaviour of binary and multiclass classifiers is coherent.

3.5 Comparison with the State-of-the-art

To evaluate our method, we decided to use more than one classifier. In order to obtain a fair comparison, we performed intra dataset and cross dataset tests on binary classifiers. As mentioned above, the datasets contain videos that have different characteristics and different conditions because they have been recorded

Table 5: Comparison with the state of the art: Replay Attack (I-RA), Replay-Mobile (I-RM), MSU-MFSD (M), Rose-Youtu (Y)

Method	I-RA	I-RM	M	Y
	HTER	HTER	HTER	HTER
LBP + SVM [1]	13.87	N/A	N/A	N/A
SVM RBF [2]	5.28	7.8	N/A	N/A
SURF (Gray) [31]	21.2	N/A	N/A	N/A
CSURF (RGB) [31]	13.5	N/A	N/A	N/A
CSURF (HSV) [31]	11.5	N/A	N/A	N/A
CSURF (YCbCr) [31]	8.9	N/A	N/A	N/A
CSURF (HSV + YCbCr) [31]	8.2	N/A	N/A	N/A
MRCNN [32]	1.6	N/A	N/A	N/A
Patch-Based CNN [33]	1.25	N/A	N/A	N/A
Depth-Based CNN [33]	0.75	N/A	N/A	N/A
Patch-Depth CNN [33]	**0.72**	N/A	N/A	N/A
DeepPixBiS [30]	N/A	**0.0**	N/A	N/A
CoALBP (HSV) [4, 31]	3.7	N/A	9.8	26.6
CoALBP (YCbCr) [4, 31]	1.4	N/A	8.1	17.1
LPQ (HSV) [4, 31]	7.9	N/A	12.2	30.4
LPQ (YCbCr) [4, 31]	6.3	N/A	7.4	27.6
Deep Learning Features [4, 27]	2.1	N/A	**5.8**	**8.0**
Our_{LDA}	11.0	6.3	36.0	24.2
Our_{SVM}	16.5	7.1	25.3	21.3
$\text{Our}_{RandomForest}$	9.0	11.1	32.3	21.5
$\text{Our}_{3L-NN_{interp}}$	11.7	8.7	22.4	21.3

in "under a controlled environment" so the cross-dataset, so the expectation is to have higher HTER than in intra-dataset. In Table 5 there is the comparison of our method with other state of art methods, in intra-dataset analysis, existing methods outperform our method. Although these methods outperform our method they analyse the videos frame by frame, on the other hand we analyse our video using segments of 16 frames, this means that our method has a less computation complexity. In Table 6 there is the comparison of our method with other state of art methods in cross-dataset analysis. As mentioned for the intra-dataset analysis all methods work in a different way for each dataset. In both Table 5 and Table 6 we choose methods that to the best of our knowledge are the best state of art methods, we reported the results that they presented in their papers, so in case of missing information we reported "N/A". We observe that Our K-NN method is competitive with most of other state of art models when trained on Replay-Attack and tested on MSU-MFSD, our SVM method is competitive when trained on Rose-Youtu and tested on Replay Attack and our Random Forest Method is competitive when trained on Rose-Youtu and tested on MSU-MFSD.

Table 6: Cross Analysis: Replay Attack (I), MSU-MFSD (M), Rose-Youtu (Y)

Method	I -> M	I -> Y	M -> I	M -> Y	Y -> I	Y -> M
	HTER	HTER	HTER	HTER	HTER	HTER
SURF (Gray) [31]	43.8	N/A	48.2	N/A	N/A	N/A
CSURF (RGB) [31]	44.1	N/A	47.8	N/A	N/A	N/A
CSURF (HSV) [31]	44.3	N/A	54.6	N/A	N/A	N/A
CSURF (YCbCr) [31]	31.8	N/A	53.8	N/A	N/A	N/A
CSURF (HSV + YCbCr) [31]	33.0	N/A	50.6	N/A	N/A	N/A
SA [4, 27]	33.2	42.8	33.3	30.0	36.2	24.9
KSA [4, 27]	33.3	40.1	34.9	30.4	38.8	26.1
ADA [34]	30.5	N/A	5.1	N/A	N/A	N/A
PAD-GAN [35]	23.2	N/A	8.7	N/A	N/A	N/A
ML-Net [36]	35.3	42.8	11.5	34.6	30.7	32.6
UDA [36]	29.0	**39.8**	3.0	**29.7**	**23.7**	**24.4**
Li,Zi et al [37]	**20.8**	N/A	**2.9**	N/A	N/A	N/A
Our$_{K-NN}$	32.1	47.3	46.4	40.2	52.0	40.7
Our$_{SVM}$	35.0	45.5	36.9	42.5	25.3	39.1
Our$_{RandomForest}$	37.9	55.3	37.1	42.0	51.1	29.2

3.6 Ablation Study

In our experiment we investigated the effectiveness of using temporal features instead of using single features for each frame. This study is performed on the dataset Idiap Replay-Attack. In Table 7 are reported the results of this study considering the classes "real" and "attack" for the binary classifier and the class "real", "print attack" and "screen attack" for the multiclass classifier. We observe from the Table 7 that the single features have slightly better results; however, the time of training and classification are so much higher than the model based on temporal features.

4 Conclusion

In this study we proposed different ways to recognise face spoofing attacks, taking into account either the classic binary task (i.e., real vs attack) and the multi-class task concerning different types of spoofing attack. This method is able to execute the real time analysis in both binary and multiclass classifiers thanks to the 3D features. We performed a big series of test on different classifiers (K-NN, Support Vector Machine, Random Forest, Decision Tree, GBoost, 3L-NN) and on different datasets (Replay-Attack, Replay-Mobile, MSU-MFSD, RECOD-MPAD, Rose-Youtu) As far as we know this is the first work able to distinguish between the type of attack and obtaining closer results to the binary classifier. In addition this is the first time that an anti-spoofing method has used the 3D

Table 7: Ablation Study

Method	Binary				Muliclass			
	Temporal		Single Frame		Temporal		Single Frame	
	Accuracy	HTER	Accuracy	HTER	Accuracy	HTER	Accuracy	HTER
K-NN	87.8	14.9	88.7	14.7	70.5	25.7	75.4	23.5
LDA	**91.4**	**11.6**	**92.7**	**9.8**	**87.0**	**13.3**	**90.3**	**10.9**
SVM linear	90.6	0.132	91.6	11.7	83.6	17.8	87.4	13.9
Random Forest	87.4	17.5	87.9	49.0	/	/	/	/
Decision Tree	83.0	22.6	83.9	22.0	/	/	/	/
GBoost	87.5	17.0	88.0	16.6	/	/	/	/

Convolutional Neural Network to extract the features. Even though our method already has a good computational complexity, the aim in the future is to optimize the classification and to improve the results of our model using a more complex classifier and a new 3D features extractor. In future work, a comprehensive study will be presented on the complexity of the anti-spoofing methods.

Acknowledgements This work is partially funded by TIM S.p.A. through its UniversiTIM granting program.

Portions of the research in this paper used the Replay-Attack Dataset made available by the Idiap Research Institute, Martigny, Switzerland.

Portions of the research in this paper used the Replay-Mobile Dataset made available by the Idiap Research Institute, Martigny, Switzerland. Such Corpus was captured in collaboration with the Galician R and D Center on Advanced Telecommunications (GRADIANT), Vigo, Spain.

References

1. Chingovska, Ivana, André Anjos, and Sébastien Marcel. "On the effectiveness of local binary patterns in face anti-spoofing." 2012 BIOSIG-proceedings of the international conference of biometrics special interest group (BIOSIG). IEEE, 2012.
2. Costa-Pazo, Artur, et al. "The replay-mobile face presentation-attack database." 2016 international conference of the Biometrics Special Interest Group (BIOSIG). IEEE, 2016.
3. Almeida, Waldir R., et al. "Detecting face presentation attacks in mobile devices with a patch-based CNN and a sensor-aware loss function." PloS one 15.9 (2020): e0238058.
4. Wen, Di, Hu Han, and Anil K. Jain. "Face spoof detection with image distortion analysis." IEEE Transactions on Information Forensics and Security 10.4 (2015): 746-761.
5. Li, Haoliang, et al. "Unsupervised domain adaptation for face anti-spoofing." IEEE Transactions on Information Forensics and Security 13.7 (2018): 1794-1809.

6. Li, Zhi, et al. "One-class knowledge distillation for face presentation attack detection." IEEE Transactions on Information Forensics and Security 17 (2022): 2137-2150.
7. Tran, Du, et al. "Learning spatiotemporal features with 3d convolutional networks." Proceedings of the IEEE international conference on computer vision. 2015.
8. Sultani, Waqas, Chen Chen, and Mubarak Shah. "Real-world anomaly detection in surveillance videos." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
9. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Communications of the ACM 60.6 (2017): 84-90.
10. Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
11. Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.
12. Yi, Dong, et al. "Learning face representation from scratch." arXiv preprint arXiv:1411.7923 (2014).
13. Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
14. Sun, Yi, Xiaogang Wang, and Xiaoou Tang. "Deeply learned face representations are sparse, selective, and robust." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
15. Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." (2015).
16. Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
17. Liu, Weiyang, et al. "Sphereface: Deep hypersphere embedding for face recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
18. Deng, Jiankang, et al. "Arcface: Additive angular margin loss for deep face recognition." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
19. Li, Jiangwei, et al. "Live face detection based on the analysis of fourier spectra." Biometric technology for human identification. Vol. 5404. SPIE, 2004.
20. Kollreider, Klaus, Hartwig Fronthaler, and Josef Bigun. "Evaluating liveness by face images and the structure tensor." Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05). IEEE, 2005.
21. Kollreider, Klaus, Hartwig Fronthaler, and Josef Bigun. "Non-intrusive liveness detection by face images." Image and Vision Computing 27.3 (2009): 233-244.
22. Pan, Gang, et al. "Eyeblink-based anti-spoofing in face recognition from a generic webcamera." 2007 IEEE 11th international conference on computer vision. IEEE, 2007.
23. Sun, Lin, et al. "Blinking-based live face detection using conditional random fields." Advances in Biometrics: International Conference, ICB 2007, Seoul, Korea, August 27-29, 2007. Proceedings. Springer Berlin Heidelberg, 2007.
24. Bao, Wei, et al. "A liveness detection method for face recognition based on optical flow field." 2009 International Conference on Image Analysis and Signal Processing. IEEE, 2009.

25. Li, Xiaobai, et al. "Generalized face anti-spoofing by detecting pulse from face videos." 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016.
26. Nowara, Ewa Magdalena, Ashutosh Sabharwal, and Ashok Veeraraghavan. "Ppgsecure: Biometric presentation attack detection using photopletysmograms." 2017 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017). IEEE, 2017.
27. Yang, Jianwei, Zhen Lei, and Stan Z. Li. "Learn convolutional neural network for face anti-spoofing." arXiv preprint arXiv:1408.5601 (2014).
28. Patel, Keyurkumar, Hu Han, and Anil K. Jain. "Cross-database face antispoofing with robust feature representation." Biometric Recognition: 11th Chinese Conference, CCBR 2016, Chengdu, China, October 14-16, 2016, Proceedings 11. Springer International Publishing, 2016.
29. Li, Lei, et al. "An original face anti-spoofing approach using partial convolutional neural network." 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, 2016.
30. George, Anjith, and Sébastien Marcel. "Deep pixel-wise binary supervision for face presentation attack detection." 2019 International Conference on Biometrics (ICB). IEEE, 2019.
31. Boulkenafet, Zinelabidine, Jukka Komulainen, and Abdenour Hadid. "Face antispoofing using speeded-up robust features and fisher vector encoding." IEEE Signal Processing Letters 24.2 (2016): 141-145.
32. Ma, Yukun, Lifang Wu, and Zeyu Li. "A novel face presentation attack detection scheme based on multi-regional convolutional neural networks." Pattern Recognition Letters 131 (2020): 261-267.
33. Atoum, Yousef, et al. "Face anti-spoofing using patch and depth-based CNNs." 2017 IEEE International Joint Conference on Biometrics (IJCB). IEEE, 2017.
34. Wang, Guoqing, et al. "Improving cross-database face presentation attack detection via adversarial domain adaptation." 2019 International Conference on Biometrics (ICB). IEEE, 2019.
35. Wang, Guoqing, et al. "Cross-domain face presentation attack detection via multi-domain disentangled representation learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
36. Wang, Guoqing, et al. "Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection." IEEE Transactions on Information Forensics and Security 16 (2020): 56-69.
37. Li, Zhi, et al. "One-class knowledge distillation for face presentation attack detection." IEEE Transactions on Information Forensics and Security 17 (2022): 2137-2150.
38. Hong, Yuxuan. Performance evaluation metrics for biometrics-based authentication systems. Diss. 2021.
39. Bengio, Samy, et al. "Confidence measures for multimodal identity verification." Information Fusion 3.4 (2002): 267-276.