Face Anti-spoofing based on Deep Stack Generalization Networks

Xin Ning[1,2], Weijun Li[1,2], Meili Wei[2], Linjun Sun[1,2] and Xiaoli Dong[1,2]
[1]*Institute of Semiconductors, Chinese Academy of Sciences, 100083, Beijing, China*
[2]*Cognitive Computing Technology Wave Joint Lab, 100083, Beijing, China*

Keywords: Face Anti-spoofing, Convolutional Neural Networks, Stacked Generalized Approach, Intra-class Variations.

Abstract: Thanks for the recent development of Convolutional Neural Networks (CNNs), the performance of face anti-spoofing methods has been improved by extracting more distinguishing features between genuine and fake faces than the hand-crafted texture features. As known, the way of fraud is diverse, thus the fake class has large intra-class variations, so training as a binary classification problem is hard to learn the distinguishing features. In this work, our contribution is a novel model fusion approach for face anti-spoofing, which can reduce the intra-class variations. According to the type of fraud, we firstly train different models for face anti-spoofing problem by CNN, thus the intra-class variations of fake class has reduced during training each model. Distinguishing features can be learned more easily. Then the stacked generalized method is used for combining the lower models to achieve better predictive accuracy. For perfecting the generalized accuracy, the stacked generalized approach changes the weight of each model's prediction, so that the model after fusion can predict precisely whether the face image is fake or genuine. Meanwhile, the experimental results indicate our method can obtain excellent results compared to the state-of-the-art methods.

1 INTRODUCTION

Face recognition has achieved great success during the past decades, and has been widely applied in access control system, login system and so on. However, a high security requirement for face authentication is urgent, because only a photo, video replay, or 3D-mask of valid user can easily spoof a face recognition system to access secure information illegally. With the popularity of electronic devices, people can easily get photos and videos belonging to others through the network, which caused a lot of face recognition system for security risks. Therefore, anti-spoofing problem for face biometric system has gained great attention to the scholars and companies.

The fragility of face recognition systems to face spoof attacks has motivated a number of studies on face anti-spoofing, such as LBP (Maatta et al., 2011), HOG (Komulainen et al., 2013), LBP-TOP (Pereira, 2012), Image quality assessment (Wen et al., 2015), CNN (Yang et al., 2014), etc.

However, published studies are limited in their scope, because these methods are more to train a common model to prevent attack from photo, video or 3D-mask. Because of the diversity of fraud, a general model cannot be learned in such complexities, may not work when facing specific fraud. Besides, these methods regard face anti-spoofing as a binary classification problem, that is to say the all kinds of fake face is one class, and the real face is another class. As we all know, the fake face can be a photo, a video replay, or a 3D-mask. Due to different ways of fraud, the fake class has large intra-class variations. These existing classifiers in identifying sample work individually, as is well known, when making critical decisions, wise people often take into account the opinions from several experts rather than only one. In this paper, we do not address 3D-mask attacks, which are more costly to launch, we focus on photo and replayed video attacks. We learn a CNN model for each type of fraud, then the stacked generalization method will be use to integrate the two models, which will decide whether the input face image is real or not. Stacked generalization is a general method of using a high-level model to combine lower-level models to achieve greater predictive accuracy.

The remainder of the paper is organized as follows: Section 2 briefly reviews the relevant work on face anti-spoofing. Section 3 presents our

Ning, X., Li, W., Wei, M., Sun, L. and Dong, X.
Face Anti-spoofing based on Deep Stack Generalization Networks.
DOI: 10.5220/0006568103170323
In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018)*, pages 317-323
ISBN: 978-989-758-276-9

approach. The experimental setup and results are discussed in Section 4. Finally, in Section 5 we summarize this work highlighting its main contributions.

2 RELATED WORK

Because of the diversity of spoofing attacks, existing traditional face anti-spoofing approaches can be mainly categorized into four categories: (i) motion based methods, (ii) texture based methods, (iii) method based on image quality analysis, and (iv) methods based on other cues. The motion based methods was designed primarily to counter printed photo attacks. Eye blinking (Pan et al., 2007) or lip movement (Kollreider et al., 2007) are used for face anti-spoofing. Given that motion is a relative feature across video frames, these methods are expected to have better generalization ability than the texture based methods. However, motion based methods need a relatively long time to accumulate stable vitality features for face spoof detection. The texture based methods include LBP (Maatta et al., 2011), HOG (Komulainen et al., 2013), etc. Pereira et al. (2012) used LBP-TOP to extract spatial and time domain features from three orthogonal planes. Unlike motion based methods, texture based methods need only a single image to detect a spoof. However, the generalization ability of many texture based methods has been found to be poor. A recent work (Galbally et al., 2014) proposed a biometric liveness detection method for iris, fingerprint and face images using 25 image quality measures, including 21 full-reference measures and 4 non-reference measures.

Different from traditional methods, CNNs can extract distinguishing end-to-end features directly from raw data, and has been proved efficient in many other vision fields. Yang et al., (2014) extract features by CNN, then feeding them to a SVM classifier. Xu et al., (2016) proposed LSTM-CNN architecture to learn the temporal structure from videos.

Those works consider the face anti-spoofing as a binary classification problem, all real face is one class, and the other is all kinds of fake face. Because of the variety of fake face, photo attacks and video attacks will be different on the texture, reflect illumination, resolution, etc., thus the large intra-variance will increase the difficulty of classification. Each model is heterogeneous and has strong classification ability, therefore, the integration model with stacked generalization method will make full use of the advantages of different models, complement each other, thus will achieve better predictive accuracy.

3 PROPOSED METHOD

For reducing the intra-class variations, we train different models according to the type of fraud face. Each model can learn the deep and distinguishing feature for classifying fake or real. The stacked generalization method takes full advantage of each model's prediction and change the weights of each prediction. Then a wiser decision would be made for maximizing generalize accuracy. With the stacked generalized method, the training difficulty of anti-spoofing problem is decreased than training a general model. Besides, the model can converge more easily.

3.1 Stacked Generalization

Stacked generalization (Wolpert, 1992; Ting and Witten, 1997) is a general method of using a high-level model to combine lower-level models to achieve greater predictive accuracy. It's a scheme for minimizing the generalization error rate of one or more classifiers, and works by reducing the biases of the classifiers with respect to a provided learning set. When fusing the multiple classifiers, stacked generalization exploited a strategy more sophisticated for combining the individual classifiers. Stacked generalization tries to learn which classifiers are reliable ones, and use a higher-level learning algorithm, the so-called "meta-classifier", to discover the best way of how to combine the outputs of the base classifiers. As shown in Figure 1, there are two kinds of classifiers: several base classifiers (leavel-0 classifiers) and one meta-classifier (level-1 classifier). The output class probabilities generated by level-0 models are used to form level-1 data. Then a multivariable linear regression model (MLR) is adapted for classification tasks for level-1 classifier.

3.1.1 Level-0 Generalizers

As shown before (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; He et al., 2016), deeper and better pre-training networks lead to better performance. ResNet (He et al., 2016) won the 1st place on the ILSVRC 2015 classification task. The depth of representations is of central importance for many visual recognition tasks, especially in face anti-spoofing. In the task of face anti-spoofing, the intra-class variations are large, and are mainly caused by the appearance of different people, the different ways of fraud, and the different resolutions of faces images captured by different camera, such as the photo attacks can be printed on different types of paper, the video attacks can be played on different electronic equipment and so on. Therefore, the deeper networks can extract more useful and distinguishing features. In this paper, we use a ResNet50 (He et al., 2016) model as the level-0 generalizers.

Given two training data sets $Q=\{(y_n,x_n),n=1,...,N_1\}$ and $P=\{(y_n,x_n),n=1,...,N_2\}$, where y_n is the class value and x_n represents the attribute values of the SS instance, Q defines the set of real face and video attack face, and P defines the set of real face and photo attack face. The specific type of fraud is included in each dataset, causing the small intra-class variations and large inter-class variations. So the training difficulty can be decreased. Then randomly split the data sets into two parts Q_1,P_1 and Q_2,P_2 . Define $Q_1=\{(y_n,x_n),n=1,...,N_1-M\}$, $P_1=\{(y_n,x_n),n=1,...,N_2-M'\}$ and $Q_2=\{(y_n,x_n),n=1,...,M\}$, $P_2=\{(y_n,x_n),n=1,...,N_2-M'\}$ to be the training and validation sets. The Q_1 and P_1 datasets are used for training level-0 data, and the Q_2 and P_2 datasets are used for forming the level-1 data. Given ResNet50 called level-0 generalizers, training on the data in the training set Q_1 and P_1 to induce a model, for $k=1,...,K$, which are called level-0 models. The level-1 data assembled from the outputs of the K models is $Q_2^{'}=\{(y_n,z_{11n},...,z_{1In},...,z_{k1n},...,z_{kIn},...,z_{K1n},...,z_{KIn}),n=1,...,M\}$, $P_2^{'}=\{(y_n,z_{11n},...,z_{1In},...,z_{k1n},...,z_{kIn},...,z_{K1n},...,z_{KIn}),n=1,...,M'\}$, where $z_{kin}=P_{ki}(x_n)$ is the prediction from the level-0 models, and $P_{ki}(x_n)$ denote the probability of the ith output class, and in the ResNet50, $P_{ki}(x_n)$ is the class probability of the Softmax layer's output, and $\sum_{i=1}^{I}P_{ki}(x_n)=1$.

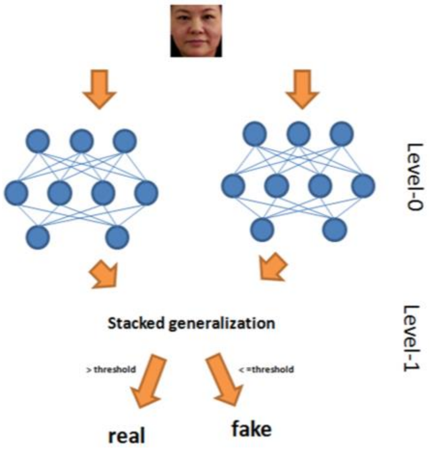

Figure 1: The illustration of stacked generalization.

3.1.2 Level-1 Generalizers

After obtaining the level-1 data, we use MLR (Breiman, 1996) to derive from this data a model $\tilde{M}$ in this work. MLR which Breiman (Breiman, 1996) used in regression settings, is an adaptation of a least-square linear regression algorithm. If the classification problem is with real-valued attributes, it can be transformed into a multi-response regression problem. In the face anti-spoofing task, they are two classes, they can be converted into two separate regression problems, where the problem for class ℓ has instances with responses equal to 1 when they have class 1 and 0 otherwise.

The linear regression for class 1 is simply:

$$LR_\ell(z)=\sum_{k}^{K}\sum_{i}^{I}\alpha_{ki\ell}P_{ki}(z) \qquad (1)$$

We solve this problem to choose the linear regression coefficients $\{\alpha_{ki\ell}\}$ to minimize:

$$\sum_{j}\sum_{(y_n,z_n)\in Q_2^{'}}\left(y_n-\sum_{k}\sum_{i}\alpha_{ki\ell}P_{ki}(z_n)\right)^2 \qquad (2)$$

where y is equal to 1 when the instance's label is corresponding with the class 1,s and 0 otherwise. These regression coefficients can be the weight of each level-0 model's prediction probability. In the test phase, after the fusion, the probability of each sample predicting fake or real is readjusted, so that according to the probability, the sample is classifying correctly. In the experimental stage, we compare the accuracy with the AND rule. As shown in Table 2, our method is better than the AND rule with a large margin. According the AND rule, if both the two model's prediction is real, then the final prediction is real, otherwise is fake. We believe the reason is that our method learns which model is more reliable and which not. Because each level-0 model is heterogeneous, for each specific sample, the classification accuracy rates of these models are quite different. Therefore, the weights can adjust the predictions so that the more wisdom decision can be given.

When classifying a new instance x, compute $LR_\ell(x)$ for the two classes, and assign the instance to class real which has the greatest value. That is to say, if $LR_{real}(x) > LR_{fake}(x)$, then we believe the sample is a real face, otherwise fake face.

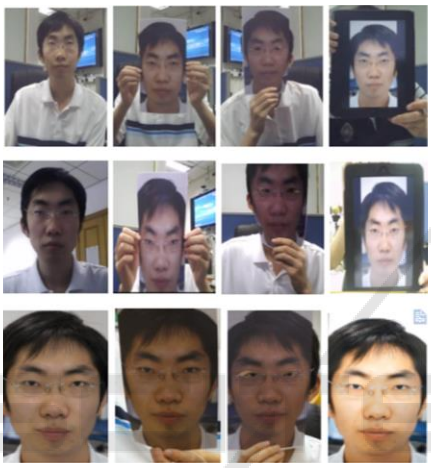

Figure 2: Examples of real access and spoofing attempts in the CASIA-FASD database. The first row is low-quality, the second row is middle-quality and the last row is high-quality. The first column is genuine, and the others form left to right are warp-photo, cut-photo and video fraud.

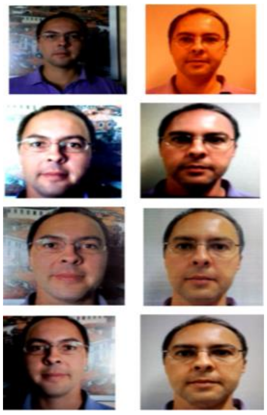

Figure 3: Examples of real access and spoofing attempts in the REPLAY-ATTACK database. The four rows from top to down are real video, print fraud video, mobile fraud video and high-definition fraud video. And the first column is adverse environment and the second column is controlled.

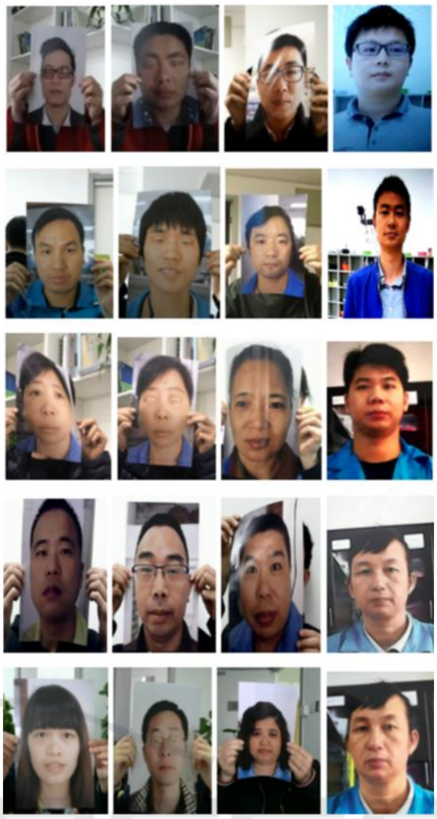

Figure 4: Examples of real access and spoofing attempts in our database. The five rows are captured by different cameras, from top to down are Phone Low Resolution front camera, Pad front camera, Phone Normal Resolution front camera, USB Low Resolution and USB High Resolution. The four columns are different type of fraud, from left to right are warp A4 photo fraud, cut eye A4 photo, warp copper photo and video replay fraud.

4 EXPERIMENTS

Our proposed method is successful on the face anti-spoofing data sets. We evaluated it in this section on two different datasets, including CASIA (Zhang et al., 2012), and Replay-attack (Chingovska et al., 2012).

4.1 Implementation Details

In this work, during the training, we first separate the video attacks and photo attacks into different sets. For each video, we selected one frame every three frames, thus forming the training, validation and test sample sets. For each frame, the face can be detected by using Viola-Jones algorithm. To provide precise face locations, we implement the face alignment algorithm proposed in Sun et al., (2014) after a common Viola-Jones face detection. After obtaining the landmarks, their bounding box is regarded as the final face location. According to Yang et al. (2014), beyond the conventional face region, the backgrounds are helpful for the classification as well. Therefore, we enlarge the original ones with re-scaling ratio 1.8. Finally, all input images are resized to 224*224. For the CNN, we use Caffe toolbox and adopt a commonly used structure ResNet50, which was ever used in He et al. (2016). In the training of the ResNet50 for video attack face and photo attack face, the learning rate is 0.0001; decay rate is 0.001; and the momentum is 0.9. Before fed into the ResNet50, the data are first centralized by the mean of training data.

4.2 Datasets

4.2.1 CASIA-FASD Database

The database collects 600 short videos from 50 clients (Zhang et al., 2012). For each subject, there different quality videos are captured by different resolution camera. For each camera, one real video and corresponding three different kinds of attacks were recorded. The three kinds of attacks are warped photo attack, cut photo attack and electronic screen attack. Therefore, each subject has 12 sequences (3 genuine and 9 fake ones). Three different imaging quality conditions were recorded using an imaging device of (1) High-quality, (2) Middle-quality, and (3) Low-quality. Example frames from genuine and fake videos are shown in Figure 2. For evaluation, the total set of videos is divided into two non-overlapping subsets for training and testing.

4.2.2 REPLAY-ATTACK Database

This database contains 1200 short videos from 50 subjects (Chingovska et al., 2012), including both real accesses and face spoofing attacks. Each person in the database was recorded the videos in two illumination conditions: controlled and adverse. A high resolution pictures and videos were taken for each subject under the same condition. There are three attacks: print attacks, mobile attacks and high-definition attacks. And the videos were divided into hand based attacks and fixed based attacks. Example frames from genuine and fake videos are shown in Figure 3.For evaluation, the total set of videos is divided into three non-overlapping subsets for training, development and testing.

4.2.3 Our Database

The database contains 1500 short videos from 100 subjects, including both real accesses and face spoofing attacks. For each subject, there different quality videos are captured by different resolution camera. For each camera, one real video and corresponding three different kinds of attacks were recorded. The three kinds of attacks are warped photo attack, cut photo attack and electronic screen attack. Example frames from genuine and fake videos are shown in Figure 4. Therefore, each subject has 15 sequences (1 genuine and 14 fake ones). For evaluation, the total set of videos is divided into two non-overlapping subsets for training.

4.3 Experimental Results

The performance of the proposed stack generalized based face anti-spoofing approach was evaluated on the three databases. All these results are given in Table 1. For a performance comparison, the results of the state-of-the-art countermeasures and the baseline algorithms in databases to face spoofing attacks are listed in Table 1. On the CASIA-FASD database, best performance in previous work was achieved by the LBPs form three orthogonal planes (LBP-TOP) method, exploring the spatial and temporal LBP distributions simultaneously. Our method achieved an EER of 3.42%, which is better than the LBP-TOP method. On the REPLAY-ATTACK, our method achieved an EER of 0.13% and HTER of 0.63% respectively, both of which are superior to the others.

Besides, the performance of our method was compared with the AND rule method on the REPLAY-ATTACK, CASIA-FASD and our database. The results are listed in Table 2.From the results, the proposed stacked generalized method achieved a huge performance improvement in liveness detection compared with the AND rule method.

On the REPLAY-ATTACK, CASIA-FASD and our database, our method achieved a huge performance improvement in face anti-spoofing problem. These results illustrate the effectiveness of the proposed stacked generalized face liveness detection approach.

Table 1: Comparison between the proposed countermeasure and state-of-the-art methods based on the REPLAY-ATTACK, CASIA-FASD database.

Approach	Replay-attack-dev(EER)%	Replay-attack-test(HTER)%	CASIA-FASD-test-EER (%)
IQA based (Galbally and Marcel, 2014)	---	---	32.40
Motion (Pereira et al., 2013)	11.60	11.70	26.60
LBP + SVM (Yang et al., 2013)	8.55	11.75	18.50
LBP-TOP + SVM (Pereira et al., 2012)	7.88	7.60	10.00
SBIQF+NN (Feng et al., 2016)	3.83	6.13	15.5
YCbCr + HSV + LBP (Boulkenafet et al., 2015)	0.4	2.9	6.2
LSTM (Xu et al., 2016)	---	---	5.17
Our Method	0.13	0.63	3.42

Table 2: Comparison between the proposed countermeasure and the AND rule method based on the REPLAY-ATTACK, CASIA-FASD database.

Approach	Replay-attack-dev(EER)%	Replay-attack-test(HTER)%	CASIA-FASD-test(EER)%
AND Rule	0.38	2.63	5.83
Our Method	0.13	0.63	3.42

5 CONCLUSIONS

With the rapid development of face anti-spoofing techniques, the threats of spoofing attacks will also increase in the diversity. It's hard to learn a model to detect all types of fraud. Hence, the integration of several countermeasures is a promising approach. The proposed method is a way of combination. And our method can be easily combined with other algorithms, as long as these algorithms are helpful for liveness detection. In future work, other advance neural networks will be investigated to improve face anti-spoofing performance, such as the long short-term memory (LSTM) network, which may be more effective in learning face liveness features.

ACKNOWLEDGEMENTS

This work was supported by the National Nature Science Foundation of China (No. 61572458), and the China Scholarship Council (Grant No.201404910237).

REFERENCES

Maatta J, Hadid A, Pietikainen M., 2011. Face spoofing detection from single images using micro-texture analysis[C]// *International Joint Conference on Biometrics. IEEE*.

Komulainen J, Hadid A, Pietikainen M., 2013. Context based face anti-spoofing[C]// *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems. IEEE*.

Pereira T D F, Anjos A, Martino J M D, et al., 2012. LBP－TOP Based Countermeasure against Face Spoofing Attacks [J].

Wen D, Han H, Jain A K., 2015. Face Spoof Detection with Image Distortion Analysis [J]. *IEEE Transactions on Information Forensics & Security*.

Yang J, Lei Z, Li S Z., 2014. Learn Convolutional Neural Network for Face Anti-Spoofing [J]. *Computer Science*.

Pan G, Sun L, Wu Z, et al., 2007. Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Webcamera[C]// *IEEE, International Conference on Computer Vision*.

Kollreider K, Fronthaler H, Faraj M I, et al., 2007. Real-Time Face Detection and Motion Analysis with Application in “Liveness” Assessment [J]. *Information Forsssensics & Security IEEE Transactions on*.

Galbally J, Marcel S, Fierrez J., 2014. Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint, and Face Recognition [J]. *IEEE Transactions on Image Processing a Publication of the IEEE Signal Processing Society*.

Xu Z, Li S, Deng W., 2016 Learning temporal features using LSTM-CNN architecture for face anti-spoofing[C]// *Pattern Recognition. IEEE*.

Wolpert D H., 1992. Original Contribution: Stacked generalization [J]. *Neural Netw*.

Ting K M, Witten I H., 1997. Stacked generalization: when does it work? [C]// *Fifteenth International Joint Conference on Artificial Intelligence*. Morgan Kaufmann Publishers Inc.

Krizhevsky A, Sutskever I, Hinton G E., 2012. ImageNet classification with deep convolutional neural networks[C]// *International Conference on Neural Information Processing Systems*. Curran Associates Inc.

Simonyan K, Zisserman A., 2014 Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. *Computer Science*.

He K, Zhang X, Ren S, et al., 2016. Deep Residual Learning for Image Recognition[C]// *Computer Vision and Pattern Recognition. IEEE.*

Breiman L., 1996. Stacked regressions [J]. *Machine Learning*.

Chingovska I, Anjos A, Marcel S., 2012. On the effectiveness of local binary patterns in face anti-spoofing[C]// *Biometrics Special Interest Group. IEEE*.

Zhang Z, Yan J, Liu S, et al., 2012. A face ant spoofing database with diverse attacks[C]// *Iapr International Conference on Biometrics*.

Galbally J, Marcel S., 2014. Face Anti-spoofing Based on General Image Quality Assessment[C]// *International Conference on Pattern Recognition. IEEE Computer Society*.

Pereira T D F, Anjos A, Martino J M D, et al., 2013. Can face anti-spoofing countermeasures work in a real world scenario? [C]// *International Conference on Biometrics. IEEE*.

Yang J, Lei Z, Liao S, et al., 2013. Face liveness detection with component dependent descriptor[C]// *International Conference on Biometrics. IEEE.*

Pereira T D F, Anjos A, Martino J M D, et al., 2012. LBP − TOP, Based Countermeasure against Face Spoofing Attacks[C]// *International Conference on Computer Vision*.

Feng L, Po L M, Li Y, et al., 2016. Integration of image quality and motion cues for face anti-spoofing [J]. *Journal of Visual Communication & Image Representation*.

Boulkenafet, Zinelabidine, Jukka Komulainen, and Abdenour Hadid., 2015 "Face anti-spoofing based on color texture analysis.*" Image Processing (ICIP), 2015 IEEE International Conference on. IEEE*.

Xu Z, Li S, Deng W., 2016. Learning temporal features using LSTM-CNN architecture for face anti-spoofing[C]// *Pattern Recognition. IEEE*.

Sun J, Wen F, Wei Y, et al., 2014. Face alignment by Explicit Shape Regression [J]. *International Journal of Computer Vision.*