Evaluation of Local Descriptors and Deep CNN Features for Face Anti Spoofing

Sandan Priya, Sanjay Pawar, Akanksha Joshi

ABSTRACT--- Recently facial recognition Technology are being habitual for various access control requirements and spoof detection in such a system has drawn growing attention. In this paper, we represent by comparison analysis of different local descriptors and off the shelf deep networks for feature extraction-Local Binary Pattern (LBP), SIFT, Histogram of Oriented Gradients (HOG), Shallow CNN, VGG16 and Inception-Resnet-V2 for face spoofing detection. Furthermore, we evaluated three Classifiers-Decision Tree, Artificial Neural Network (ANN) and Support Vector Machine (SVM) over the feature extracted through local descriptors and deep networks. The evaluation has been conducted using publicly available YALE face database containing real and fake facial images. Real dataset consists of 5121 entries and fake dataset has 7508 images. The analysis results demonstrate that the best prediction accuracy of real and spoof is obtained with Inception_ResnetV2 features when classified with ANN and about 96.23% accuracy is achieved.

Keywords—LBP, SIFT, HOG, ANN, SVM, CNN, face quality

I. INTRODUCTION

In recent times, the security plays a very important role in all scenarios. There has been a lot of growth and development for automatic authentication system. In the traditional authentication system, it uses password which can be easily broken. Three different types of authentication system used as of now are face, fingerprint and iris; which are completely dependent on human features.

We face many challenges against spoofing when developing any automatic authentication system. In spoofing, the attackers use fake images similar to real images to access the victim data or services. Spoofing is done in numerous ways such as: i. Photo attack: attacker uses fake image of victim to access data/services. ii. Replay attack: attacker uses looped video of victim image. Replay attack is more natural than print attack. iii. Mask Attack: attacker wears a mask to fool the system to access victim data or services. In all biometric, face is a promising means of authentication due one to its convenience, its low cost of acquisition and acceptability by users, and also because it is very suitably used in a wide variety of environments, including mobile phones [4]. In many research works researchers have proposed various methods for face spoof detection [5, 6, 11-12, 14, 16].

In this paper we compared different local image descriptors, and deep ConvNet (CNN) to extract all face region features from the given image. In modern time, the ConvNet (CNN) has shown promising results in various computer visiontasks because of its automatic learning ability and very good performance. They are beingapplied to solve the issues of image classification, object detection, in medical stream and biometric field and many more places. In our research, we have used a shallow CNN, VGG16 and Inception ResNet V2, and we have provided a comparative analysis of their performance for spoof detection. We utilized pretrained networks on ImageNet Dataset and by doing this we try to give an idea which off-the-shelf architecture can give improved accuracy. Apart from three deep networks, we have also compared features extracted using well known local image descriptors- LBP, SIFT, and HOG.

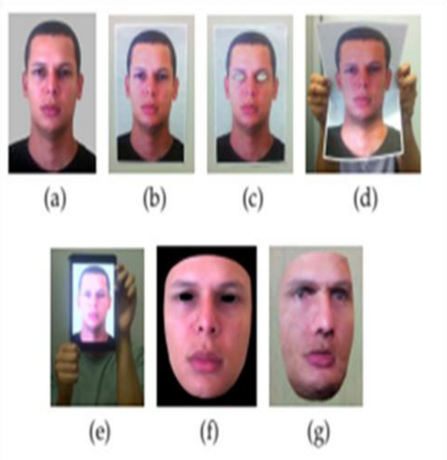


Figure 1: Different types of face spoof attack: (a) Real Image; (b)printed image; (c) eye-cut image; (d) warped image; (e) video playback; (f) life-size wearable mask and (g) paper-cut mask.

Local Binary Pattern (LBP) [9,17,24,33] is a grayscale local texture descriptor and it converts selected pixels into binary code. SIFT is used to detect and describes all possible local features, Histogram of Oriented Gradient (HOG) has been previously used as image descriptor for image classification, object detection. After extracting all these features we get a feature vector, it is given to the classifier to detect the class of the input image. We compared total three classifiers- Decision Tree, Artificial Neural Network (ANN) and Support Vector Machine (SVM).

The earlier Face Anti-Spoofing primarily focuses on quality parameters, texture, motion and frequency to identify real and non-real or spoof face.

Revised Manuscript Received on August 19, 2019.

SandanPriya,(E-mail: sandhanapriyanadar@gmail.com)
Sanjay Pawar,(E-mail: drsanjayspawar@gmail.com)
Akanksha Joshi, (E-mail: akanksha@cdac.in)

In NUAA benchmark [23] of 15 subjects, Määttä et al. [24] compared Local Phase Quantization (LPQ), Gabor wavelet based descriptor and Local Binary Patterns (LBP) for face spoofing and reported that LBP performed the best for the task. Next, they demonstrated that combinations of three LBP descriptors with different settings are more accurate than single configuration LBP. Di, Wen et.al [7]used Image Distortion Analysis (IDA) for spoofing detection. They selected color diversity, blurriness, reflection and chromaticfeatures. The combined features are classified using Support Vector Machine (SVM).J. Galbally et.al [19] presented a face Anti-Spoofing based on Image Quality Assessment (IQA) Algorithm. They first transfer image into grayscale followed by filtering with a low-pass Gaussian filter to generate distorted version of the input image. Further the quality between the grayscale and distorted image is compared by using Image Quality Assessment metric such as Pixel Difference measured by computing distortion between two images. Correlation measured by computing angles between pixels, Edge measures. They used LDA for classification of real images and spoof images.

J. Komulainen et.al [20] used Histogram of Oriented Gradient (HOG) descriptor to detect close-up spoofed faces.In the approach they detected upper body part and if upper body is not present in the image then authors consider non presence of the real face. Then the image is given to a spoofing medium detector to find whether the input image is spoof or real image. Anjos et.al [21] presented an approach to detect real and fake by countermeasures based on texture and motion. Linear Discriminant Analysis (LDA) for classification. J. Li et.al [22] presented an approach to detect the real and spoof face images by analysis of the Fourier Spectra of the input image. They basically demonstrated that the high frequency components of the non-real images are lesser than compared to real images. They classified the input face as spoofed if the median is smaller than a threshold.

David, Menotti [2] used two deep learning approaches to detect spoofing through automatic feature learning by convolutional neural network and learning the weight of network via back propagation.In [2] they have used two approaches to detect spoofing via Convolutional neural architecture and back propagation. Image distortion analysis extracted features like colour diversity specular , reflection, blurriness and chromatic moment are given to SVM classifier and it detected spoofing was seen in [7].CNN based approaches for face anti-spoofing was proposed in [6].

In this work, our aim is to perform comparative analysis by using different types of descriptors and CNN network architecture along with different classifiers. We used publically available database NUAA. It contains around 12000 images of real and spoofed image.

The rest of the paper flow is as follows. Section II, provides details for CNN architectures and local descriptors system.Section III, gives experimental setup and results, here we demonstrate the comparison between the two architectures and their accuracy and error rate. At last, we conclude our research in section V.

II. METHODOLOGY

In this section we first describe the pipeline adopted for spoof detection in this paper. Then we provide details of local image descriptors and deep architectures used for comparative analysis in this work. The methodology for face spoof detection is shown in Figure 3.

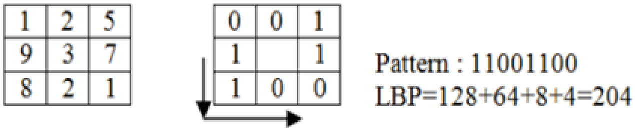
1. Face Detection and Extraction

Face Detection is a part of object detection and the first and essential step for face spoof detection, and it is used to identify faces in the images. In this Paper, we used deep Learning based face detector present on Dlib [25]. Face detection and extractionmethod uses a Maximum-Margin Object Detector (MMOD) with ConvNet based features.

2. Local Feature Extractors

A. Local Binary Pattern (LBP) [28]

LBP calculated the local features from an image by comparing the central pixel value from the local pixel value. Each neighbor pixels' value compared with central pixel value, If the central pixel value is less than neighbor pixel value then replaces the neighbor pixel by one (1) or else zero (0). Later we obtain the binary sequence. At the last stage transformation takes place from binary to decimal transformation. The basic LBP computation is shown inbelow Figure 2.



1	2	5
9	3	7
8	2	1

0	0	1
1		1
1	0	0

Pattern : 11001100
LBP=128+64+8+4=204

Figure 2: Computation of LBP operator

B. Histogram of Oriented Gradients (HOG) [8]

It is a feature extraction technique, to detect object from an image. It represents image as a single feature vector as conflicting to a set of feature vector. It is computed by sliding window approach over an image. This descriptor is computed for each pixel in an image. Initially it divides images into cells for each cells histogram of gradient direction is determined. All these histogram is represented in single vectors. These vector the concatenates the components of the normalized histograms from all the block regions.

C. SIFT (Scale-Invariant Feature Transform) [10]

David G. Lowe proposed SIFT in 1999. It is a feature Extraction Techniques in Image processing. SIFT detects and Define the local Features of an Images. This algorithm can be explained in four steps namely, Feature descriptor generation , Feature point localization, Orientation assignment,keypoint detection . Keypoints are essential and salient and well identifiable features like blobs. These Keypoints are recognized with Gaussian filter by smoothing the image and thus this unwanted noise is eliminated. However the key point remains the same. Once the keypoint has been detected the next part is to define the region around them for feature matching. In SIFT , first we measure the gradient of image them histogram of orientation is calculated. This histogram of orientation measures how powerful the gradients in each direction. By determining multiple histograms and combining themwe get a final histogram.

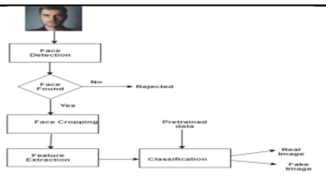


Figure 3: System architecture

3. Deep CNN Features

There are a number of DCNN architectures well known within the deep learning community. We utilized a shallow CNN (modified Alex Net [27]), VGG16 [18], and InceptionResNet-V2 [1] for the intent of comparison. Training a deep neural networks such as VGG16 which has hundreds of millions parameters is a non-trivial task that not only takes several days but also require huge amount of labeled training data. A much cheaper approach in terms of both required amount of training data and time, is to extract features using existing CNN models and then training a classifier such as SVM on these features.

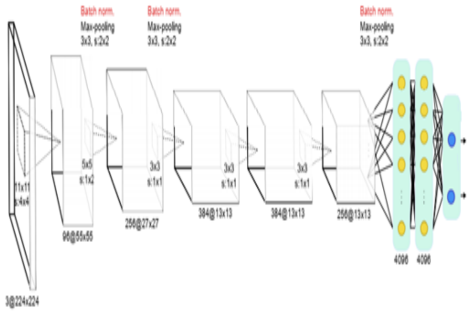


Figure 4: Shallow Convolutional Neural

Shallow CNN

It contains 5 convolutional layers, 3 fully connected andmax-pooling layers. Before the first and second layer of fully connected dropout is applied. After everyfully connected and convolutional layersRelu is applied. The size of image in the following architecture represented in 224 x 224. The architecture follows Alexnet [27]. However, We also applied batch normalization [15] after each convolutional and fully connected layer. We performed a pre-training of the network on ImageNet dataset. The shallow CNN architecture is shown in figure 4.

VGG16

VGG is developed by visual geometry group by Oxford in 2014.the layer range is from 11 to 19. In this paper, we used VGG 16 pre-trained on ImageNet dataset. The ImageNet dataset is a public dataset consists of more than 14 million images. It has 16 layers with trainable parameters and about 138 million parameters.

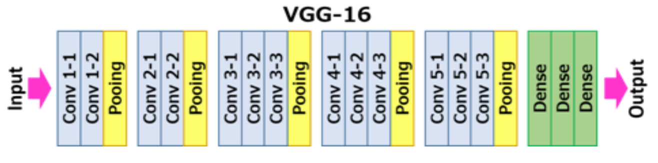


Figure 5:Architecture of VGG 16

It is a deep convolutional neural network consists of 16 layers. It contains convolutional layer of 3*3, soft max layer, max pooling layer 2*2and the fully connected layer. The Activation function i.e.ReLu is used. Here by removing the fully connected layer, the VGG16 is act as a Descriptor, which is going to produce the vector from the extracted feature value. Later these vectors are given to different type of classifiers such as Decision Tree, ANN and SVM to differentiate the dataset as a spoofed image and real image. Figure 4 shows the architecture of VGG 16.

Inception_ResNet_V2

The key thought behind Inception-ResNet-V2 [1] network is to gather two recent ideas in ILSVRC challenge i.e.Residual connections andcurrentInception architecture proposedin [12], [1] presented the residual connection in which they give convincing hypothetical and functional evidence for the benefit of utilizing additive merging of signals both for image recognition, and especially for object detection. The authors debate that residual connections are inherently necessary for training very deep convolutional models.

In [1], the authors have demonstrated the use of residual connections seems to great training speed, which is alone a great argument for their use. The Inception deep convolutional architecture was presented in [13] and was called by GoogLeNet or Inception-v1, After that it refined in many ways, first by Inception-v2 [15] by kumar et al in which they used batch normalization. Later the architecture was refinedby additionalideasreferred asInception-v3in the iteration[1] .

In Inception-ResNet-V2, instead of usingoriginal Inceptionauthors used cheaper Inception blocksand each block consist of filer expansion layer i.e.1*1 Convolutional without activation. Inception blocks is used for scaling up the dimensionality of the filter bank before the addition to match the depth of the input. The batch-normalization is used on top of the traditional layers.

4. Classifiers

Support Vector Machine (SVM) [26]

It is used to decide spoofed image and real image. SVM is trained by extracting feature values and types of input sample. Once the training is done, the SVM generates a classification matrix consisting of spoofed or real class images.

For the SVM classifier, we performed a grid search for both hyperparameters C and gamma. A smaller C leads to a larger decision margin at the cost of more incorrect training classifications; a larger gamma leads to a faster influence drop around support vectors. We let C vary logarithmically between 10 and 10000; and gamma between 0.00001 and 10. Figure 5 shows the average validation scores for hyperparameter values. The best values obtained are- C = 100:0 and gamma= 0:01.

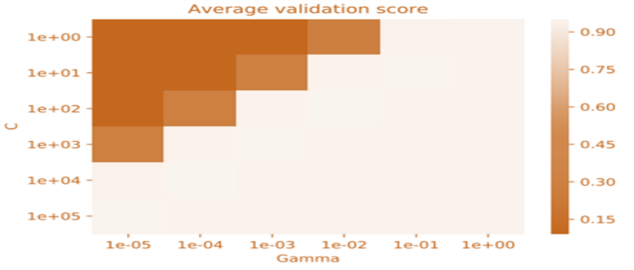


Figure 6: SVM Hyperparameter Search

Decision Tree (DT)

Decision Tree Classifier is simple and widely used for classifying images. Classifying problems using the method of Decision Tree (DT) is pretty straightforward. It habitual a tree-like graph or model of decisions and their possible consequences, regard chance event outcomes, utility and resource costs. Decision Tree Classifier poses a series of carefully planned questions about the attributes of the recorded testing set. It is one way to display an algorithm that only contains conditional control statements.

Artificial Neural Network (ANN)

It is a computational model. ANN comprises of structures and functions of biological neural networks. The structure of ANN gets affected by the important features that flow all over the network. Since ANN is a neural network it can learn, in a sense - based on that input and output. ANNs have three layers that are interconnected. The first layer consists of input neurons. These neurons send data on to the second layer, which in turn sends the output neurons to the third layer. ANNs are considered nonlinear statistical data modeling tools where the complex relationships between inputs and outputs are modeled or patterns are found.

III. EXPERIMENTAL ANALYSIS& RESULTS

A. Experimental Setup

For CNN feature extraction, we used Tesla K40m GPU with 12GB memory installedin an Intel Corei7 server with 16 GB RAM. The OS is Ubuntu 16.04 with Cuda 8.0 and cuDNN 5.0. We used Python-3.5 for all our experiments. and for evaluation of deep neural networks the Tensorflow-1.12 deep learning framework is used.

Initially, we divided the dataset into training, validation and test sets by randomly selected 100% images for training, 10% for validation and 10% for testing purposes.

Accuracy of our approaches is computed as follows:

$$Accuracy\ is\ equal,$$

$$\frac{100\text{ x }(\text{Number of Correctly Predicted Real Instances} + \text{Number of Correctly Predicted Fake Instances})}{\text{TotalNumber ofInstances}}\ldots.(1)$$

B. Data Set

For the purpose of this paper, we have used NUAA [23] Dataset. This dataset is publicly available and is proposed by Tan et al. [23]. The image in the dataset is extracted from videos of 15 subjects captured in three sections and contains attempts of attack based on hand-held printed photos. We divided the dataset, into training and test sets. The training set has 1743 live images and 1748 printed images, and the test set contains 3362 live and 5761 non-live face images. For testing purpose, 5007 real entries are used and 7394 fake entries are used. For training purpose, 113 real entries are used and 113 fake entries are used.

Initially the images are resized into 160*160 resolutions as per deep network.

C. Results

To conduct various experiments, we used Local Binary Pattern (LBP), SIFT, Shallow CNN, Histogram of Oriented Gradients (HOG),VGG16 and Inception-Resnet-V2 as our feature extraction approaches and SVM, Decision Tree and ANN as our classifiers. We trained the classifiers on the training set and conducted a testing on the test data, and measured the MAP (Mean Average Precision) values. The results of the preliminary evaluation are given in Table 1.

According to the results, Inception_ResnetV2 + Decision Tree method has higher accuracy results (i.e. 99.52%). All results indicate that deep learning based descriptors outperforms compared to the traditional local image descriptors. It is also seen that among the local image descriptors, LBP + SVM with 82.74% has given the best accuracy. It can be deduced from the results that Inception_ResnetV2 features are more robust compared to all other approaches. Another thing to be noted is that, the Computational cost for deep learning based networks for image feature extraction is quite high. The results in Table 1 also demonstrates that on an average, ANN based classifier is more robust and show less variation compared to other classifier methods.

Table 1: Accuracies and Computation Costs

Feature Extraction Approach	Classifier	Accuracy(%)	Time (ms)
LBP	SVM	82.74	57
HOG	SVM	50.44	306
SIFT	SVM	52.65	903
Shallow CNN	SVM	83.08	112
VGG16	SVM	93.36	213
Inception_ResnetV2	SVM	99.11	223
LBP	DT	52.21	55
HOG	DT	50.88	299

SIFT	DT	76.99	874
Shallow CNN	DT	80.03	108
VGG16	DT	82.30	213
Inception_ResnetV2	DT	**99.52**	204
LBP	ANN	63.71	**49**
HOG	ANN	87.61	298
SIFT	ANN	89.38	896
Shallow CNN	ANN	81.03	99
VGG16	ANN	92.47	199
Inception_ResnetV2	ANN	92.92	209

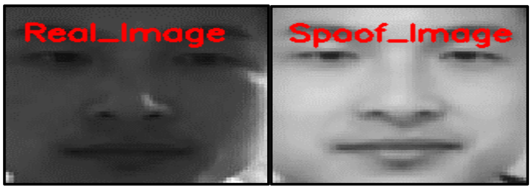


Figure 7 : Results

IV. CONCLUSION AND FUTURE WORK

The face authentication modality is the most reliable system as its unique characteristics. For authenticating user, face spoof detection plays a vital role in today's world. As the growth of upcoming technology, authentication of particular user increased in demand. In the last few years, the researcher has given more attention to detecting fake and real faces. However, we have performed an empirical study on a publically available dataset using six different descriptors: LBP, HOG, SIFT, Simple CNN, VGG16, and inception-ResNet V2 with SVM, Decision Tree and ANN classifier.From the results, it can see that Inception_ResnetV2 performs significantly better than other methods described in this paper for face spoof detection. LBP descriptors are easy to compute, do not need training, and are low dimensional.

In future, we would extend this study on more datasets and will also try fusion of local descriptors with CNN descriptors.

REFERENCES

1. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna,"Rethinking the inception architecture for computer vision,"arXiv preprint arXiv:1512.00567 , 2015.
2. Menotti, David &Chiachia, Giovani& Pinto, Allan & Schwartz, William &Pedrini, Helio&Falcão, Alexandre& Rocha, Anderson, "Deep Representations for Iris, Face, and Fingerprint Spoofing Attack Detection," 10.1109/TIFS.2015.2398817, 2014.
3. Mhou, Kudzaishe, "Face Spoof Detection Using Light Reflection in Moderate to Low Lighting,"10.1109/ACIRS.2017.7986063,2017.
4. Patel, Keyurkumar& Han, Hu & K. Jain, Anil, "Secure Face Unlock: Spoof Detection on Smartphones," IEEE Transactions on Information Forensics and Security. 11. 10.1109/TIFS.2016.2578288, 2016.
5. Boulkenafet, Zinelabidine&Komulainen, Jukka&Hadid, Abdenour, "Face Spoofing Detection Using Colour Texture Analysis," IEEE Transactions on Information Forensics and Security. 11. 1-1. 10.1109/TIFS.2016.2555286, 2016.
6. Atoum, Yousef& Liu, Yaojie&Jourabloo, Amin & Liu, Xiaoming, "Face Anti-Spoofing Using Patch and Depth-Based CNNs," 10.1109/BTAS.2017.8272713, 2017.
7. Wen, Di & Han, Hu & K. Jain, Anil, "Face Spoof Detection With Image Distortion Analysis," IEEE Transactions on Information Forensics and Security, 10. 10.1109/TIFS.2015.2400395, 2015.
8. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego,CA,USA,2005,pp.886-893vol.1.doi:10.1109/CVPR.2005.177 ,2005.
9. Akanksha Joshi, AbhishekGangwar, Zia Saquib, "Person recognition based on fusion of iris and periocular biometrics," 12th International Conference on Hybrid Intelligent Systems (HIS), vol., no., pp.57,62, 4-7 Dec. 2012, DOI: 10.1109/HIS.2012.6421309,2012.
10. Lowe, David G,"Object recognition from local scale-invariant features," Proceedings of the International Conference on Computer Vision, pp. 1150–1157. doi:10.1109/ICCV.1999.790410, 1999.
11. Arashloo, Shervin& Kittler, Josef & Christmas, William,"An Anomaly Detection Approach to Face Spoofing Detection," A New Formulation and Evaluation Protocol,IEEE Access. PP. 1-1, 2017.
12. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv preprint arXiv:1512.03385, 2015.
13. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–9, 2015.
14. Li, Xiaobai&Komulainen, Jukka& Zhao, Guoying& Yuen, Pong-Chi &Pietikainen, Matti, "Generalized face anti-spoofing by detecting pulse from face videos," 4244-4249. 10.1109/ICPR.2016.7900300, 2016.
15. S. Ioffe and C. Szegedy. Batch normalization, "Accelerating deep network training by reducing internal covariate shift," In Proceedings of The 32nd International Conference on Machine Learning , pages 448–456, 2015.
16. Sun, Zhonglin& Sun, Li & Li, Qingli, "Investigation in Spatial-Temporal Domain for Face Spoof Detection," 1538-1542. 10.1109/ICASSP.2018.8461942, 2018.
17. Akanksha Joshi, AbhishekGangwar, Renu Sharma and Zia Saquib, "Periocular Feature Extraction Based on LBP and DLDA," International Conference on Computer Science, Engineering & Applications, Springer, Delhi, India, May 25-27, 2012. DOI: 10.1007/978-3-642-30157-5_101
18. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," coRR, vol. Abs/1409.1556, Sep. 2014.
19. J. Galbally, S. Marcel, "*Face anti-spoofing based on general image quality assessment*" in Proc. IAPR/IEEE Int. Conf. on Pattern Recognition, ICPR, 2014.
20. J.Komulainen, A.Hadid, M.Pietik¨ainen, "*Context based face Antispoofing,*" in Proc. the International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2013.
21. A. Anjos, J. Komulainen, A. Hadid, S. Marcel, M.Pietik¨ainen,"*Complementary countermeasures for detecting scenic face spoofing attacks,*" in IAPR International Conference on Biometrics, 2013.
22. J. Li, Y. Wang, T. Tan, A. K. Jain, "*Live face detection based on the analysis of Fourier spectra,*" in Biometric Technology for Human Identification, 2004.
23. X. Tan, Y. Li, J. Liu, L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in Proc. ECCV, pp. 504–517, 2010.
24. J. Määttä, A. Hadid, and M. Pietikainen,"Face spoofing detection from single images using micro-texture analysis," In International Joint Conference on Biometrics, pages 1–7, 2011.
25. https://github.com/davisking/dlib-models
26. Cortes, C., Vapnik, V, "Support-vector networks," Machine learning **20**(3), 273{297, 1995.
27. A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," Proc. Advances in Neural Information Processing Systems, pp. 1097– 1105, 2012.
28. T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, no. 7, pp. 971–987, 2002.