ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/317549190

Transfer Learning Using Convolutional neural networks for Face Anti-Spoofing

Chapter · July 2017

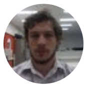
DOI: 10.1007/978-3-319-59876-5

CITATIONS
0

READS
1,494

6 authors, including:

Oeslle Lucena
University of Campinas
7 PUBLICATIONS **3** CITATIONS
SEE PROFILE



Vitor Hugo G Moia
University of Campinas
12 PUBLICATIONS **1** CITATION
SEE PROFILE

Roberto Souza
The University of Calgary
19 PUBLICATIONS **90** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Medical Imaging Processing View project

Project Sampling and similarity hashes on digital forensics field View project

All content following this page was uploaded by Oeslle Lucena on 13 June 2017.

The user has requested enhancement of the downloaded file.

Transfer Learning Using Convolutional Neural Networks for Face Anti-Spoofing

Oeslle Lucena[1], Amadeu Junior[1], Vitor Moia[1], Roberto Souza[1], Eduardo Valle[1] and Roberto Lotufo[1]

[1]University of Campinas, Campinas, Brazil
{oeslle,amadeu,vghmoia,rmsouza,dovalle,lotufo}@dca.fee.unicamp.br

Abstract. Face recognition systems are gaining momentum with current developments in computer vision. At the same time, tactics to mislead these systems are getting more complex, and counter-measure approaches are necessary. Following the current progress with convolutional neural networks (CNN) in classification tasks, we present an approach based on transfer learning using a pre-trained CNN model using only static features to recognize photo, video or mask attacks. We tested our approach on the REPLAY-ATTACK and 3DMAD public databases. On the REPLAY-ATTACK database our accuracy was 99.04% and the half total error rate (HTER) of 1.20%. For the 3DMAD, our accuracy was of 100.00% and HTER 0.00%. Our results are comparable to the state-of-the-art.

Keywords: Face anti-spoofing, Transfer learning, Deep learning, Face recognition

1 Introduction

In the last few years, the usage of face recognition systems for surveillance and authentication increased significantly due to the advances in computer vision technologies. As usage grows, the complexity of spoofing attacks also arises, and more complex counter-measure approaches are built. For instance, some of them consist of presenting to the vision sensor a fake image, video or even a 3D mask.

Deep Learning methods are representation-learning methods with multiple levels of representation along the neurons in a deep neural network (DNN) [1]. In a DNN, each level utilizes a non-linear module that transforms the representation at one level into a higher and more abstract representation at the next level [2]. Also, in contrast to conventional machine learning algorithms, DNNs are fed with raw data, and they discover the representations needed for detection or classification [1].

Inspired by this ability, this work proposes a method to identify spoofing attacks based on transfer learning using a pre-trained convolutional neural network (CNN). The main contributions of this paper are 1) a CNN approach based on transfer learning, using only static features, in other words no time relation between frames was used 2) evaluation of half total error rate (HTER) on two public databases that outperformed or at least matched the state-of-the-art.

This work is organized as follow: Section 2 presents the related work with research that motivated our study of face anti-spoofing. The description of the proposed CNN method is detailed in Section 3. The experiments and results to validate our architecture, including pre-processing methods and evaluation metrics are reported in Section 4. Finally, the conclusions and future works are presented in Section 5.

2 Related Work

Due to many spoofing attack forms, elaborated approaches have been developed to identify and block these attacks. Some of them rely on extra sensors [3,4]. Others can be categorized into two main groups: feature level static and dynamic [5].

Feature level static methods focus on the analysis of images without considering the time relation between them. Those comprise techniques that use Fourier analysis, Lambertian models, Difference-of-Gaussian (DoG) and Local Binary Patterns (LBP). Li *et al.* [6] analyze the Fourier spectrum and introduced a high-frequency descriptor to identify spoofing attacks through images sequences. Using Fourier analysis combined with Lambertian model features, Tan et al. [7] were capable of extracting reflectance to recognize attacks. Peixoto et al. [8] used DoG filters and Sparse Logistic Regression Model to improve previous results in extreme light environments. Erdogmus et al. [9] proposed a solution based on LBP to recognize attacks on the 3D-MAD database, and their results showed an HTER of 0.95%.

Feature level dynamic methods explore the time relation between sequential frames of a video. In many works, the authors have used motion on detected faces as cues to recognize the attack (e. g.: eye blink movement [10,11], tracking face natural movements [12] or lips movement [13]). Anjos *et al.* [14] used correlation between background and foreground optical flows. Pereira et al. [15] used LBP-TOP operator combining space and time into a single texture descriptor. In this work, the authors reported an HTER of 7.6% on the REPLAY-ATTACK database. Komulainen et al. [16] presented a method that combined through a linear logistic regression the LBP operator and the correlation between background and face movements. The authors reported an HTER of 5.11% on the REPLAY-ATTACK database. Feng *et al.* [17] presented a neural network that fuses features such as ~~shearlet-based~~ image quality, face motion, and scene motion clues. The neural network is a pre-trained layer-wise sparse autoencoder. In this approach, the neural network is fine-tuned with a softmax layer classifier and labeled data using backpropagation. The HTER for REPLAY-ATTACK and 3DMAD database were of 0.00%.

Except for Feng *et al.*'s [17] work, aforementioned methods rely on handcrafted features to determine attacks. Other approaches take advantage of CNNs abilities to identify features from images, thus recognizing attacks. Yang *et al.* [18] proposed an approach that uses a CNN architecture of AlexNet [19] for feature extraction, and Support Vector Machines (SVM) for classification. Also, the authors deployed a different method for pre-processing images varying faces' bounding boxes sizes, and number of successive frames used on the CNN. The HTER in their best case scenario on REPLAY-ATTACK database was of 2.81%.

Menotti *et al.* [20] proposed architecture optimization (AO) and filter optimization (FO). The second approach uses fine-tuning on CifarNet CNN [20]. However, their best case scenario was using AO that seeks for an optimal architecture of CNNs with filters weights set randomly. This approach achieve an HTER of 0.76% on REPLAY-ATTACK and 0.00% on the 3DMAD database. Alotaibi *et al.* [21] used a non-linear diffusion operator as a pre-processing step, then the processed image was applied to a custom six layers CNN. Using the REPLAY-ATTACK database the HTER was of 10.00%.

3 Proposed Method

Our approach is based on transfer learning a pre-trained CNN [22–24]. Transfer learning passes learned "knowledge" from a Machine Learning model trained on one task to another one [25]. For CNNs, there are two approaches to do transfer learning. The simpler uses the source model as a "off-the-shelf" feature extractor [26], using the output of a chosen layer as input for the target model, which is the only one trained for the new task. A more sophisticated approach "fine-tunes" the source model, in whole or in part, retrain its weights via backpropagation.

Transfer learning may be used to avoid overfitting a large network if there is not enough data to train it from scratch [25]. Transfer learning also saves computational resources, since training from scratch may take from days to weeks. In our case, we chose the CNN architecture VGG-16 that was pre-trained on ImageNet database [27], and its team secured the first and the second places in the localization and classification tasks of the ImageNet ILSVRC-2014 [28]. As the transfer learning approach, we used the fine-tuning.

3.1 CNN Architecture

The proposed face anti-spoofing network (FASNet) follows the VGG-16 architecture except for the top layers. The FASNet architecture is shown in Figure 1 with its top layers highlighted in yellow. Our model was built based on Keras tutorials [29]. The FASNet code is available at `https://github.com/OeslleLucena/FASNet`.

The VGG-16 architecture is a 2D CNN for a 224×224 image size as input. In total, it has 16 convolutional layers, 64 for filters in the first block, 128 filters in the second, 256 in the third, 512 in the fourth and fifth blocks. Each convolution has a kernel of size 3×3. All max-pooling layers are performed in a 2×2 window, with stride 2. The activation functions are rectifier linear unit (ReLU). There are three fully connected (FC) layers, the sizes of each one are 4096, 1000, and 1000. The FASNet changes compared to the VGG-16 are only at the top layers. One FC layer was removed, and the size of the other two was modified to 256 and 1. Moreover, we adopted the Adam optimizer [30] with the configurations provided by the paper, changing learning rate to 10^{-4} and weight decay to 10^{-6}. Also, the decision function has been modified from a softmax to a sigmoid, which often performs better for binary classification [31].

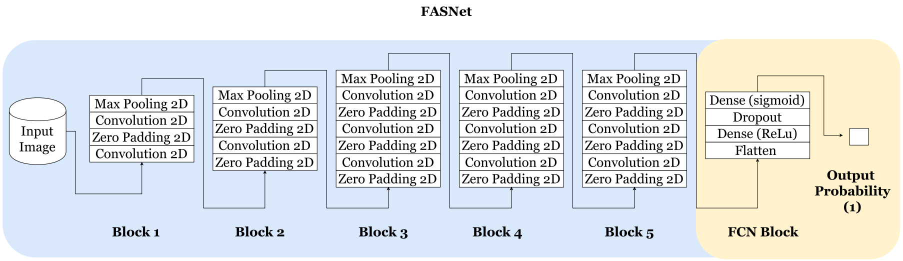


Fig. 1. FASNet architecture. As previously mentioned, the top layers of VGG-16 were changed to perform the binary classification for the anti-spoofing task. The layers highlighted in yellow represent the modified top layers

3.2 Databases

3DMAD This database is comprised of 76,500 frames of 17 persons in a total of 255 videos (300 frames per video) recorded using a Kinect for both genuine subjects and mask attacks [32]. Each frame consists of a depth image, corresponding RGB image, and manually annotated eye positions. For our purposes, only color images were used. All data is split into 3 sessions, which the first two sessions only have genuine subjects, and the last session has the mask attacks. The two first sessions have both 7 videos while the third session has only 5 videos. Some samples of this benchmark are shown in Figure 2.

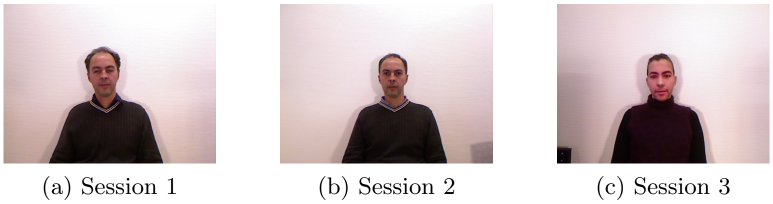
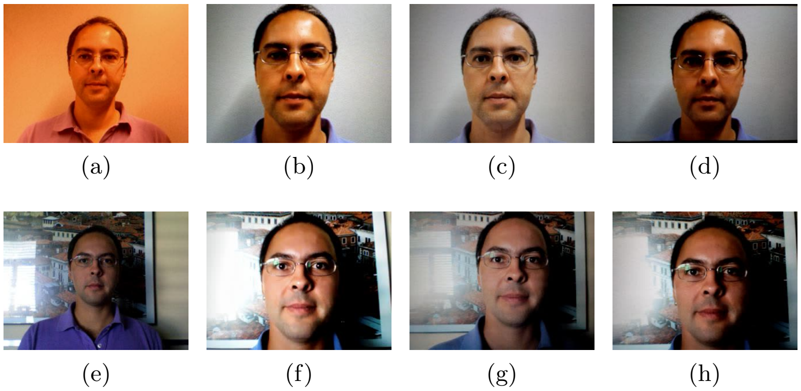

(a) Session 1
(b) Session 2
(c) Session 3

Fig. 2. Example of images from 3DMAD database from each session [32]. (a) and (b) are genuine images and (c) is a mask attack.

REPLAY-ATTACK This database consists of 1,300 video clips of photo and video attack attempts of 50 subjects [33]. The subjects of this database were collected under two different illumination conditions, controlled and adverse. The first condition was with office light turned on, blinds were down and homogeneous background. The second condition had the following characteristics: blinds up, complex backgrounds, and office lights were out. Also, the attack protocols are classified according to the type of device used to generate the attack, for example: print, mobile (phone), and high-definition (tablet). Some samples of this benchmark are shown in Figure 3.


(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 3. Example of images from REPLAY-ATTACK database [33]. The first row represent controlled images and the second row represents adverse images. In the first column is shown examples of genuine images. From the second column to the fourth are shown example of the following protocols of attacks: print, mobile and high definition.

4 Experiments and Results

Our experiments were conducted on the train and test folders provided by each face anti-spoofing database. Since the public databases are composed of video clips and our architecture is 2D, some pre-processing was needed. All the details of the adopted procedures and evaluation metrics used are found in Section 4.1.

To train the FASNet we froze the weights from the bottom layers up to the third block of our face anti-spoofing network (FASNet), and we fine-tuned weights from the fourth block up to the top layers via backpropagation. Then, we evaluate our method on REPLAY-ATTACK and 3DMAD test folders. The comparison with the state-of-the-art results are discussed in Section 4.2.

The implementation of the algorithm was built using Keras [29] with Theano [34] as backend. Furthermore, all analyses were conducted using a computer equipped with Intel(R) Xeon(R) E5506 2.13GHz (6GB RAM), using a NVIDIA Tesla K40c GPU (12GB).

4.1 Pre-processing and Metrics

The pre-processing steps adopted were two: 1) subsampling and 2) face detection. Step one consisted of extracting half of the frames per second in each video. In step two, we first used the OpenFace face detector [35] algorithm to find the region-of-interest (ROI) corresponding to a face. Next, using OpenFace algorithms we cropped to a window sized 96 pixels and aligned the faces to center based on the nose and eyes position.

Our evaluation was based on accuracy (ACC) and half total error rate (HTER), which are often used for assessing biometrics systems. Since the CNN output scores are probabilities [20], the computation of HTER was done assuming $\tau = 0.5$, and its estimation was not needed.

4.2 Results

We compare our method with six other approaches, three of them based on conventional machine learning algorithms [9, 15, 16], one based on shallow neural networks [17], and other two based on CNNs [20, 21]. Our method outperformed HTER for all of the conventional machine learning methods submitted on the 3DMAD and REPLAY-ATTACK databases, which are reported in the first three lines of Table 1. FASNet has beaten almost all the state-of-art methods, losing only for the Multi-cues Integration approach [17] on REPLAY-ATTACK benchmark, and our method reached an HTER of 0.00% on 3DMAD database. All comparative results are shown in Table 1. Remark that our approach only uses static features, while Multi-cues Integration approach, for example, combined both static and dynamic features.

Table 1. Comparative test results for different databases among FASNet and state-of-the-art methods.

Databases	3DMAD		REPLAY-ATTACK	
	ACC (%)	HTER (%)	ACC (%)	HTER (%)
LBP-TOP + SVM [15]	-	-	-	7.60
Texture-based countermeasures [9]	-	0.95	-	-
Context Based [16]	-	-	-	5.11
Spoofnet [20]	100.00	0.00	98.75	0.70
Multi-cues Integration [17]	-	0.00	-	**0.00**
Non-linear Diffusion [21]	-	-	-	10.00
FASNet	100.00	0.00	**99.04**	1.20

5 Conclusions

In this paper, we introduced a new approach that uses transfer learning in convolutional neural networks to address face anti-spoofing methods. Our experimental results showed a HTER on REPLAY-ATTACK and 3DMAD databases equal to 1.20% and 0.00%, respectively, outperforming almost all state-of-the-art methods. For future work, we intend to investigate deeper CNN architectures, for example ResNet [36], incorporate dynamic features, and also explore other complex benchmarks.

References

1. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553) (05 2015) 436–444
2. Schmidhuber, J.: Deep learning in neural networks: An overview. Neural Networks **61** (2015) 85–117
3. Seal, A., Ganguly, S., Bhattacharjee, D., Nasipuri, M., Basu, D.K.: Automated thermal face recognition based on minutiae extraction. CoRR **abs/1309.1000** (2013)
4. Zhang, Z., Yi, D., Lei, Z., Li, S.Z.: Face liveness detection by learning multispectral reflectance distributions. In: Face and Gesture 2011. (March 2011) 436–441
5. Galbally, J., Marcel, S., Fierrez, J.: Biometric antispoofing methods: A survey in face recognition. IEEE Access **2** (2014) 1530–1552
6. Li, J., Wang, Y., Tan, T., Jain, A.K.: Live face detection based on the analysis of fourier spectra. Volume 5404. (2004) 296–303
7. Tan, X., Li, Y., Liu, J., Jiang, L. In: Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model. Springer Berlin Heidelberg, Berlin, Heidelberg (2010) 504–517
8. Peixoto, B., Michelassi, C., Rocha, A.: Face liveness detection under bad illumination conditions. In: 2011 18th IEEE International Conference on Image Processing. (Sept 2011) 3557–3560
9. Erdogmus, N., Marcel, S.: Spoofing 2d face recognition with 3d masks. IEEE Transactions on Information Forensics and Security **9**(7) (July 2014) 1084–1097
10. Pan, G., Sun, L., Wu, Z., Lao, S.: Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In: 2007 IEEE 11th International Conference on Computer Vision. (Oct 2007) 1–8
11. Li, J.W.: Eye blink detection based on multiple gabor response waves. In: 2008 International Conference on Machine Learning and Cybernetics. Volume 5. (July 2008) 2852–2856
12. Liting, W., Xiaoqing, D., Chi, F.: Face live detection method based on physiological motion analysis. Tsinghua Science & Technology **14**(6) (2009) 685 – 690
13. Kollreider, K., Fronthaler, H., Bigun, J.: Verifying liveness by multiple experts in face biometrics. In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. (June 2008) 1–6
14. Anjos, A., Chakka, M.M., Marcel, S.: Motion-based counter-measures to photo attacks in face recognition. IET Biometrics **3**(3) (Sept 2014) 147–158
15. Pereira, T.F., Anjos, A., De Martino, J.M., Marcel, S. In: LBP TOP Based Countermeasure against Face Spoofing Attacks. Springer Berlin Heidelberg, Berlin, Heidelberg (2013) 121–132
16. Komulainen, J., Hadid, A., Pietikäinen, M.: Context based face anti-spoofing. In: 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS). (Sept 2013) 1–8
17. Feng, L., Po, L.M., Li, Y., Xu, X., Yuan, F., Cheung, T.C.H., Cheung, K.W.: Integration of image quality and motion cues for face anti-spoofing: A neural network approach. Journal of Visual Communication and Image Representation **38** (2016) 451 – 460
18. Yang, J., Lei, Z., Li, S.Z.: Learn convolutional neural network for face anti-spoofing. CoRR **abs/1408.5601** (2014)
19. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. (2012) 1097–1105

20. Menotti, D., Chiachia, G., Pinto, A., Schwartz, W.R., Pedrini, H., Falcão, A.X., Rocha, A.: Deep representations for iris, face, and fingerprint spoofing detection. IEEE Transactions on Information Forensics and Security **10**(4) (April 2015) 864–879
21. Alotaibi, A., Mahmood, A.: Deep face liveness detection based on nonlinear diffusion using convolution neural network. Signal, Image and Video Processing (2016) 1–8
22. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Region-based convolutional networks for accurate object detection and segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(1) (Jan 2016) 142–158
23. Noh, H., Hong, S., Han, B.: Learning deconvolution network for semantic segmentation. In: The IEEE International Conference on Computer Vision (ICCV). (December 2015)
24. Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging **35**(5) (May 2016) 1285–1298
25. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., eds.: Advances in Neural Information Processing Systems 27. Curran Associates, Inc. (2014) 3320–3328
26. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: Cnn features off-the-shelf: An astounding baseline for recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. (June 2014)
27. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV) **115**(3) (2015) 211–252
28. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR **abs/1409.1556** (2014)
29. Chollet, F.: keras. https://github.com/fchollet/keras (2015)
30. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR **abs/1412.6980** (2014)
31. Duch, W., Jankowski, N.: Survey of neural transfer functions. Neural Computing Surveys **2** (1999) 163–213
32. 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 Biometrics Special Interest Group (BIOSIG). (Sept 2012) 1–7
33. Erdogmus, N., Marcel, S.: Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect. In: 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS). (Sept 2013) 1–6
34. Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I.J., Bergeron, A., Bouchard, N., Bengio, Y.: Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop (2012)
35. Amos, B., Ludwiczuk, B., Satyanarayanan, M.: Openface: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science (2016)
36. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (June 2016)

View publication stats