A Patch-Based CNN Built on the VGG-16 Architecture

Subjects: Computer Science, Artificial Intelligence

Contributor: Dewan Ahmed Muhtasim , Monirul Islam Pavel , Siok Yee Tan

Facial recognition is a prevalent method for biometric authentication that is utilized in a variety of software applications. This technique is susceptible to spoofing attacks, in which an imposter gains access to a system by presenting the image of a legitimate user to the sensor, hence increasing the risks to social security. Consequently, facial liveness detection has become an essential step in the authentication process prior to granting access to users. A patch-based convolutional neural network (CNN) with a deep component for facial liveness detection for security enhancement was developed, which was based on the VGG-16 architecture.

biometric liveness detection social security artificial intelligence CNN VGG-16 LSTM

1. Introduction

Biometric systems have been utilized in various security applications in recent years due to ongoing research into their implementation [1][2]. Facial recognition-based liveness detection is one of the major branches of biometric technology that have been effectively applied in e-commerce, device security and organizational attendance, as well as for ensuring top-notch security, especially in the era of the IR 4.0. The core role of liveness detection is to verify whether the source of a biometric sample is a live human being or a fake representation. This process provides more safety and improvements to traditional facial recognition-based security systems [3], which use a person's unique biometric information, such as their face, to allow that individual to access specific systems or data. However, one of the primary impediments to biometric identification systems is the risk of spoofing attacks [4]. A facial spoofing attack is an attempt by an unauthorized person to circumvent the facial authentication protocol and the facial verification process by employing deception techniques, such as identity forgery [5]. A printed image of an authorized face or a recorded video from a display may provide sufficient unique data to deceive the system [6][7]. As a result, the resiliency of these security systems can be diminished.

A multitude of applications use facial biometric authentication, such as automated teller machines (ATMs), smart security systems and other similar systems [8][9]. Advanced artificial intelligence models increase the data processing capability of biometric technology, which results in more effective biometric security systems [10]. Nevertheless, spoofing attacks are a common form of attack that reduce the effectiveness of biometric authentication systems [11]. A facial spoofing attack is an attempt by an illegal user to circumvent a facial authentication system and facial verification system using deception methods, such as counterfeiting the identity of

an authorized user [5]. Therefore, the implementation of liveness detection systems that are powered by artificial intelligence has the potential to make existing biometric-based security systems more sustainable and secure while also providing better social security.

Recently, numerous effective facial anti-spoofing methods have been developed [4][5][6][7][8]. They can be generally divided into fixed feature-based facial anti-spoofing algorithms [12] and automated learnable feature-based facial anti-spoofing algorithms [13]. Facial anti-spoofing techniques utilize hand-crafted features from actual and counterfeit faces to detect spoofing. The features of actual and false faces are determined before training facial anti-spoofing algorithms. Motion, texture, image quality, 3D shape and multi-spectral reflectance are examples of fixed feature-based algorithms. Automated learnable feature-based facial anti-spoofing methods distinguish between legitimate and fake faces using deep learning approaches, such as convolutional neural networks (CNNs). CNNs learn the properties of real and fake faces during training. By transmitting raw pixels through hidden layers, CNNs translate the raw pixels of images into probabilities. The number of hidden layers determines the depth of a CNN [14][15]. Although deep CNNs are the optimal solution for most applications, their usage in facial anti-spoofing applications has been restricted due to a lack of training data.

2. Facial Anti-Spoofing Approaches

2.1. Feature-Based Facial Anti-Spoofing Approaches

Liu et al. proposed an improved local binary pattern for face maps that could be used as a classification feature [16]. When these characteristics were put into a support vector machine (SVM) classifier, the face maps could be classified as genuine or false. Tan et al. proposed a technique for formulating the anti-spoofing objective as a binary classification problem [17]. They implemented Difference of Gaussian (DoG) filtering to eliminate noise from 2D Fourier spectra and then extended a sparse logistic regression classifier both nonlinearly and spatially for the classification. Matta et al. utilized multiscale local binary patterns to examine the texture of facial images [18]. Afterward, the macrotexture patterns were encoded into an improved feature histogram, which was then input into an SVM classifier.

Parveen et al. proposed a texture descriptor that they described as the dynamic local ternary pattern, in which the textural features of skin were examined using an adaptive threshold configuration and the classification was performed using a support vector machine with a linear kernel [19]. Das et al. proposed a method that was based on a frequency and texture analysis to distinguish between real and artificial faces [20]. The frequency evaluation was accomplished by Fourier transforming the images into the frequency domain and then calculating the frequency descriptor to detect dynamic variations in the faces. LBP was employed to analyze the texture and an SVM classifier with a radial basis function kernel was used to classify the generated feature vectors.

The method that was proposed by Luan et al. involved three characteristics: blurriness, the specular reflection ratio and color channel distribution characteristics [21]. The images were classified using an SVM, according to these three characteristics. Chan et al. developed a technique that utilized flash to defend against 2D spoofing attacks

[22]. This technique captured two images per individual: one with flash and the other without. Three additional descriptors were employed to extract the textural information of the faces: a descriptor that was based on uniform LBP and the standard deviation and mean of the grayscale difference between the two recorded images of each person. The difference between the pictures with and without flash, as assessed by the four descriptors, was used to classify the images. Kim et at. proposed a method that defined the differences between the surface characteristics of live and fake faces by calculating the diffusion speeds and extracting anti-spoofing characteristics that were based on the local patterns of the diffusion speeds [23]. These characteristics were input into a linear SVM classifier to assess the liveness of the images. Yeh et al. developed a method that utilized digital focus properties with different depths of field to accomplish liveness detection [24]. The nose and the lower right portion of the cheek were examined for preprocessing. Due to the impact of the depth of field, the degree of blurriness differed between real and false images. The k-nearest neighbor method was used to classify the results. Although hand-crafted feature extraction was also utilized in [12][13], researchers adopted a patch-based CNN that was built on VGG-16 architecture to conduct feature extraction automatically, which removed the need for hand-crafted feature extraction.

2.2. Deep Learning-Based Facial Anti-Spoofing Approaches

Deep CNN models have been employed in recent facial liveness identification studies because they provide more accurate liveness detection than the previously presented strategies [7][8][14][15]. For facial anti-spoofing detection, Atoum et al. suggested a two-stream CNN-based method that included a patch-based CNN and a depth-based CNN [25]. While the first CNN extracted local features from face image patches, the second extracted depth features by computing the depth from the entire image and then using an SVM for feature extraction. This two-stream network could be trained end-to-end to discriminate between real and fake faces based on their rich appearance attributes by including these parameters. Rehman et al. proposed a technique that concentrated on data randomization in mini-batches to train deep CNNs for liveness detection [26].

The method that was proposed by Alotaibi et al. used a mix of input facial diffusion accompanied by a three-layer CNN architecture [27]. For facial liveness detection, Alotaibi et al. suggested an approach that employed nonlinear diffusion, accompanied by a tailored deep convolutional network [28]. By rapidly diffusing the input image, nonlinear diffusion aided in differentiating between fake and genuine images. As a result, the edges of flat images faded away while the edges of genuine faces stayed visible. Furthermore, to extract the most significant properties for classification, a customized deep convolutional neural network was suggested. Koshy et al. proposed a method that combined nonlinear diffusion with three architectures: CNN-5, ResNet50 and Inception v4. They found that the Inception v4 architecture was the best [29]. Jourabloo et al. addressed the facial anti-spoofing problem as an image denoising problem, which resulted in the development of an anti-spoofing method that was based on CNNs to achieve the facial anti-spoofing objective [30]. In the first layer of a Lenet-5-based neural network model, De Souza et al. applied the LBP descriptor, which improved the accuracy of facial spoofing detection [31]. An improved version of LBPnet, which is called n-LBPnet, was suggested to achieve higher accuracy for real-time facial spoofing detection by integrating the local response normalization (LRN) step into the second layer of the network.

Xu et al. demonstrated that facial anti-spoofing in videos could be achieved using a deep architecture that integrated LSTM and a CNN [32]. The LSTM obtained the temporal correlations in the input sequences while the CNN retrieved the local and dense features. Tu et al. proposed that facial spoofing detection in video sequences could be achieved using a CNN–LSTM network, which concentrated on motion cues throughout video frames [33]. To enhance the facial emotions of the humans in the videos, they used Eulerian motion magnification. Highly discriminative features were extracted from the video frames using a CNN and LSTM, which were also used to capture the temporal dynamics from the videos. Recently, Khade et al. proposed an iris liveness detection technique that used several deep convolutional networks [34]. Densenet121, Inceptionv3, VGG-16, EfficientNetB7 and Resnet50 were employed in that study to identify iris liveness using transfer learning approaches. The limited dataset necessitated the use of a transfer learning approach to prevent overfitting. As stated above, numerous well-referenced CNN models have demonstrated the ability to distinguish between real and fake faces. Therefore, this entry presents a patch-based CNN architecture for training complicated and differentiating features to improve the security of present facial recognition-based biometric authentication systems against printed images and replay attacks.

3. Architecture of the Proposed System

3.1. Liveness Detection

A patch-based CNN that was built on the VGG-16 architecture with a deep aspect was proposed for liveness detection to improve security. After the patches are constructed, the input images are transmitted sequentially to the CNN, which serves as the front-end of the architecture. The CNN output is then passed to an LSTM, which identifies temporal features in the sequence and determines whether the dense layer in the neural network output is real or fake. The workflow of the proposed method is shown **Figure 1**.

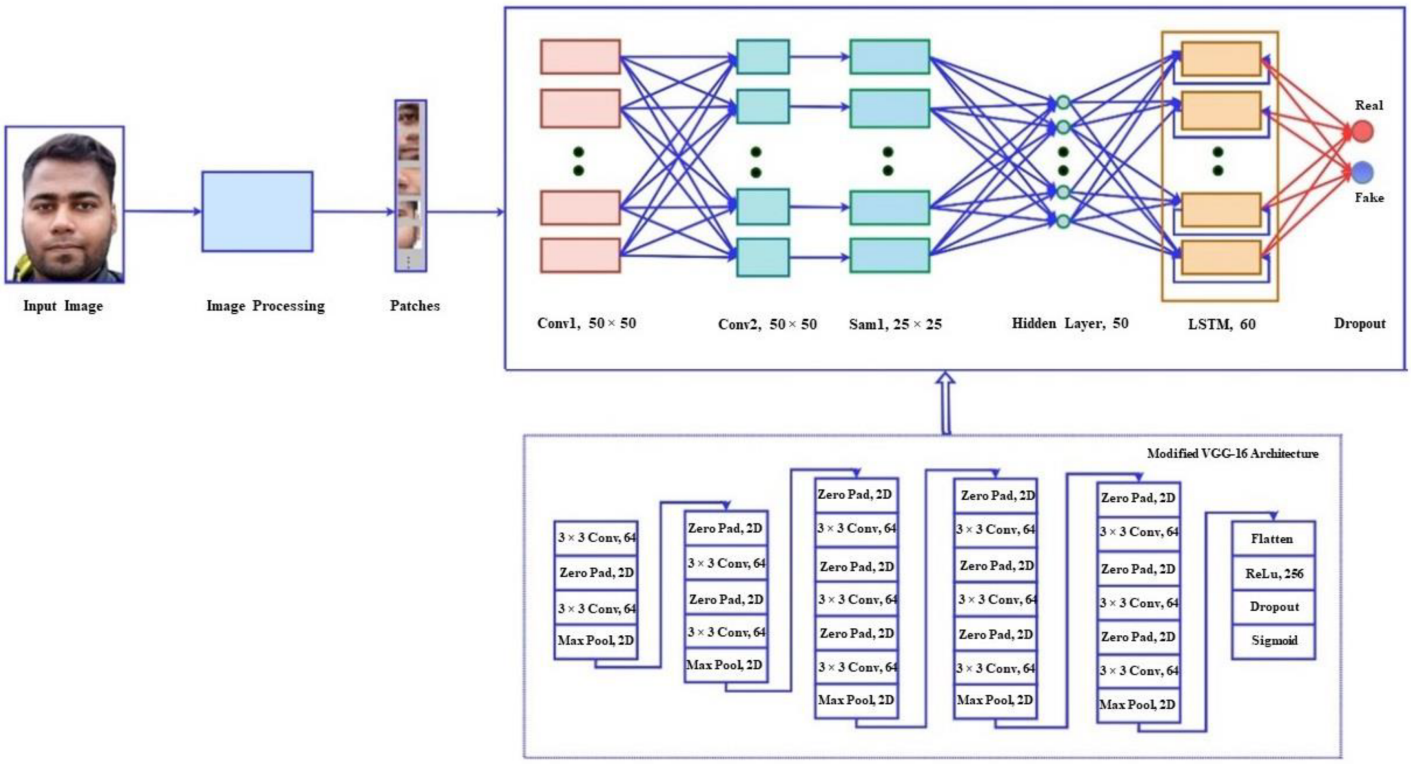

Figure 1. The workflow of the patch-based CNN–LSTM architecture with the modified VGG-16.

3.2. Patch-Based CNN

Before applying the patch-based approach [23][24][35] for real-time testing purposes, an image processing algorithm (LBPH) was adopted using OpenCV [36], which allowed facial boundaries to be detected within images. Further, there were several motivations for the proposed CNN to use patches instead of the whole face. Firstly, the number of CNN learning samples needed to be increased due to the limited number of samples that were available for training in all of the accessible anti-spoofing datasets. Although hundreds of faces could be taken from individual videos by cropping faces frame by frame, overfitting posed a huge problem while training the CNN because of the significant similarities between the images.

Secondly, classic CNNs need to redimension faces when employing full facial photos as inputs because of the various resolutions of the photos. These shifts in size could lead to a drop in discrimination between images. In contrast, by adopting local patches, the native resolution of the original images can be maintained and the discriminatory capability of the system can be preserved. To solve this, a patch-based CNN was deployed in which each frame was converted into patches that were then classified separately. As the spoof-specific discriminatory information was present in the whole facial region, patch-level inputs were used to enable the CNN to detect this information irrespective of the patch position, despite this being a more complicated process than using the entire facial image.

Furthermore, the input features were selected by converting HSV color space to obtain the discriminative descriptors as the scope of anti-spoofing methods that use RGB images is limited [25] and color space can also be

utilized for chrominance and luminance information. Then, pixel-wise LBP maps were randomly extracted for spatial texture information rather than using them as traditional histogram descriptors [37][38][39]. Using a pre-trained Haar cascade model for front facial detection, the maps were transformed into feature representations, as well as fixed sized patches, for processing using the VGG-16-based CNN–LSTM model.

References

1. Dronky, M.R.; Khalifa, W.; Roushdy, M. A review on iris liveness detection techniques. In Proceedings of the 2019 IEEE 9th International Conference on Intelligent Computing and Information Systems (ICICIS 2019), Cairo, Egypt, 8–10 December 2019; pp. 48–59.
2. Nsaif, A.K.; Ali, S.H.M.; Jassim, K.N.; Nseaf, A.K.; Sulaiman, R.; Al-Qaraghuli, A.; Wahdan, O.; Nayan, N.A. FRCNN-GNB: Cascade faster R-CNN with Gabor filters and naïve bayes for enhanced eye detection. IEEE Access 2021, 9, 15708–15719.
3. Chan, M.; Delmas, P.; Gimel'farb, G.; Leclercq, P. Comparative study of 3D face acquisition techniques. In International Conference on Computer Analysis of Images and Patterns; Springer: Berlin/Heidelberg, Germany, 2005; Volume 3691, pp. 740–747.
4. Raheem, E.A.; Ahmad, S.M.S.; Adnan, W.A.W. Insight on face liveness detection: A systematic literature review. Int. J. Electr. Comput. Eng. 2019, 9, 5165–5175.
5. Chen, F.M.; Wen, C.; Xie, K.; Wen, F.Q.; Sheng, G.Q.; Tang, X.G. Face liveness detection: Fusing colour texture feature and deep feature. IET Biom. 2019, 8, 369–377.
6. Khade, S.; Gite, S.; Thepade, S.D.; Pradhan, B.; Alamri, A. Detection of iris presentation attacks using hybridization of discrete cosine transform and haar transform with machine learning classifiers and ensembles. IEEE Access 2021, 9, 169231–169249.
7. Khade, S.; Gite, S.; Thepade, S.D.; Pradhan, B.; Alamri, A. Detection of iris presentation attacks using feature fusion of thepade's sorted block truncation coding with gray-level co-occurrence matrix features. Sensors 2021, 21, 7408.
8. Muley, A.; Bendre, A.; Maheshwari, P.; Kumbhar, S.; Dhakulkar, B. Survey on biometric based ATMs. Int. J. Sci. Res. Sci. Technol. 2021, 8, 292–297.
9. Kowshika, A.; Sumathi, P.; Sandra, K.S. Facepin: Face biometric authentication system for ATM using deep learning. NVEO-Nat. Volatiles Essent. OILS J. 2022, 9, 1859–1872.
10. Waymond, R. Artificial Intelligence in a Throughput Model: Some Major Algorithms; CRC Press: Boca Raton, FL, USA, 2020.
11. Hadid, A.; Evans, N.; Marcel, S.; Fierrez, J. Biometrics systems under spoofing attack. IEEE Signal Process. Mag. 2015, 32, 20–30.

12. 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 Trans. Inf. Forensics Secur. 2015, 10, 864–879.
13. 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. J. Vis. Commun. Image Represent. 2016, 38, 451–460.
14. Abdullah, A.; En Ting, W. Orientation and scale based weights initialization scheme for deep convolutional neural networks. Asia-Pac. J. Inf. Technol. Multimed. 2020, 9, 103–112.
15. Zavvar, M.; Garavand, S.; Nehi, M.R.; Yanpi, A.; Rezaei, M.; Zavvar, M.H. Measuring reliability of aspect-oriented software using a combination of artificial neural network and imperialist competitive algorithm. Asia-Pac. J. Inf. Technol. Multimed. 2016, 5, 75–85.
16. Liu, X.; Lu, R.; Liu, W. Face liveness detection based on enhanced local binary patterns. In Proceedings of the 2017 Chinese Automation Congress (CAC 2017), Jinan, China, 20–22 October 2017; pp. 6301–6305.
17. Tan, X.; Li, Y.; Liu, J.; Jiang, L. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In European Conference on Computer Vision; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6316, pp. 504–517.
18. Määttä, J.; Hadid, A.; Pietikäinen, M. Face spoofing detection from single images using micro-texture analysis. In Proceedings of the International Joint Conference on Biometrics (IJCB), Washington, DC, USA, 11–13 October 2011; pp. 1–7.
19. Parveen, S.; Ahmad, S.M.S.; Abbas, N.H.; Adnan, W.A.W.; Hanafi, M.; Naeem, N. Face liveness detection using dynamic local ternary pattern (DLTP). Computers 2016, 5, 10.
20. Das, D.; Chakraborty, S. Face liveness detection based on frequency and micro-texture analysis. In Proceedings of the 2014 International Conference on Advances in Engineering & Technology Research (ICAETR 2014), Unnao, India, 1–2 August 2014; pp. 3–6.
21. Luan, X.; Wang, H.; Ou, W.; Liu, L. Face liveness detection with recaptured feature extraction. In Proceedings of the 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Shenzhen, China, 15–17 December 2017; pp. 429–432.
22. Chan, P.P.K.; Liu, W.; Chen, D.; Yeung, D.S.; Zhang, F.; Wang, X.; Hsu, C.C. Face liveness detection using a flash against 2D spoofing attack. IEEE Trans. Inf. Forensics Secur. 2018, 13, 521–534.
23. Kim, W.; Suh, S.; Han, J.J. Face liveness detection from a single image via diffusion speed model. IEEE Trans. Image Process. 2015, 24, 2456–2465.

24. Yeh, C.H.; Chang, H.H. Face liveness detection with feature discrimination between sharpness and blurriness. In Proceedings of the 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), Nagoya, Japan, 8–12 May 2017; pp. 398–401.
25. Atoum, Y.; Liu, Y.; Jourabloo, A.; Liu, X. Face anti-spoofing using patch and depth-based CNNs. In Proceedings of the International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; pp. 319–328.
26. Rehman, Y.A.U.; Po, L.M.; Liu, M. LiveNet: Improving features generalization for face liveness detection using convolution neural networks. Expert Syst. Appl. 2018, 108, 159–169.
27. Alotaibi, A.; Mahmood, A. Enhancing computer vision to detect face spoofing attack utilizing a single frame from a replay video attack using deep learning. In Proceedings of the 2016 International Conference on Optoelectronics and Image Processing (ICOIP), Warsaw, Poland, 10–12 June 2016; pp. 1–5.
28. Alotaibi, A.; Mahmood, A. Deep face liveness detection based on nonlinear diffusion using convolution neural network. Signal Image Video Process. 2017, 11, 713–720.
29. Koshy, R.; Mahmood, A. Optimizing deep CNN architectures for face liveness detection. Entropy 2019, 21, 423.
30. Jourabloo, A.; Liu, Y.; Liu, X. Face de-spoofing: Anti-spoofing via noise modeling. In Lecture Notes in Computer; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11217, pp. 297–315.
31. Souza, G.B.D.; Member, S.; Felipe, D.; Pires, R.G.; Marana, A.N.; Papa, J.P. Deep texture for roboutface spoofing detection. IEEE Trans. Circuits Syst. II Express Briefs 2017, 64, 1397–1401.
32. Xu, Z.; Li, S.; Deng, W. Learning temporal features using LSTM-CNN architecture for face. In Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 3–6 November 2015; pp. 141–145.
33. Tu, X.; Zhang, H.; Xie, M.; Luo, Y.; Zhang, Y.; Ma, Z. Enhance the Motion Cues for Face Ani-Spoofing using CNN-LSTM Architecture. arXiv 2019, arXiv:1901.05635.
34. Khade, S.; Gite, S.; Pradhan, B. Iris Liveness Detection Using Multiple Deep Convolution Networks. Big Data Cogn. Comput. 2022, 6, 67.
35. Pereira, T.D.F. Lbp-top based countermeasure against face spoofing attacks. In Asian Conference on Computer Vision; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; pp. 121–132.
36. Pavel, M.I.; Chowdhury, I.; Akther, A. Raspberry Pi and image processing based person recognition system for visually impaired people. Int. Res. J. Eng. Technol. 2018, 5, 809–813.
37. Yu, Z.; Li, X.; Shi, J.; Xia, Z.; Zhao, G. Revisiting pixel-wise supervision for face anti-spoofing. IEEE Trans. Biom. Behav. Identity Sci. 2021, 3, 285–295.

38. Cai, R.; Li, Z.; Wan, R.; Li, H.; Hu, Y.; Kot, A.C. Learning meta pattern for face anti-spoofing. IEEE Trans. Inf. Forensics Secur. 2022, 17, 1201–1213.
39. Khan, H.A.; Jue, W.; Mushtaq, M.; Mushtaq, M.U. Brain tumor classification in MRI image using convolutional neural network. Math. Biosci. Eng. 2020, 17, 6203–6216.