International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

An Efficient ROI-Based Transfer Learning to Discriminate Spoof Attacks

[1]Hirendra R. Hajare, [2]Asha Ambhaikar

Submitted: 29/12/2023 **Revised**: 05/02/2024 **Accepted**: 13/02/2024

Abstract: Modern video processing tools and advanced computer-aided systems have made it difficult to distinguish between real and fake identities. The spoofing tools nowadays are very powerful to fool the best antispoofing framework. The article presents a simple but robust dual-unit framework for preprocessing and blind feature extraction plus classification to discriminate between real and fake subjects. A low-complexity preprocessing unit to enhance the image details and a quality feature extraction module using YOLOV5s are introduced using transfer learning. The dual unit framework is evaluated over IDIAP-Replay attack dataset images obtained 98.39% classification accuracy for distinguishing the authentic and the fooled samples. The work does not use data augmentation, or face alignments, and performs well on imbalance classes having uneven foreground illumination samples.

Keywords*: Advanced computer-aided systems, antispoofing, dual unit framework, YOLOV5s, transfer learning, IDIAP-Replay Attack, and data augmentation.*

1. Introduction

Biometric-based authentication in several computing applications has gained momentum as a reason for its immense popularity in the digital era with growing advanced biometric technologies. Due to its accuracy, simplicity, and security reasons [3][4], it overpassed the traditional approaches and is now considered an active research area [1][2], These biometric systems are concerned with unique human biological, chemical, and physiological features which include different external and internal elements such as thumb, voice, signatures, palm, iris, eyes, ears, DNA pattern, etc. [5][6]. Face-based authentication is the most popular infrastructure among all due to its employment in numerous commercial, forensic, military, government, banks, smartphones, and home security applications.

Every citizen of India today is been recognized using a unique Adhar ID (Unique identification system) that today covers almost 1.3 billion people across. Many state and central government schemes are availed using the unique ID including medical facilities, agriculture subsidies, subsidies, etc. The multimodal approach comprising three different traits includes human face, iris, and fingerprints for authentication [7]. Despite having such a robust authentication system, human faces are breached using forged faces thus challenging the best authentication system. The overall success rate for the breaching identities amounts to 70% using spoofed faces through fake images [8][9].

To sustain or defend the forged attacks in face biometrics, it is essential to concatenate biometric authentications with the antidotes. Presentation attacks which are the most crucial concerns are constructed using either a subject photo to fool the authentication system, pre-recoded video for simulating live subjects, and a 3D mask. Except for the last, the first two forgeries are commonly used for face spoofing due to low cost and have grown immensely [10]. The face attacks are combated using either active or passive techniques. The active methods are robust against photo attacks and pre-recorded video attacks while they show low intensity against 3D mask attacks. Passive methods employ reflectance analysis, motion analysis, and texture. Dedicated hardware for sensing the temperature can easily monitor the disparity between the subject and object but due to their non-availability to end users, they remain unfit and less accessible. Therefore, most commonly they are used in coordination with image relying techniques [11][12].

Earlier work focussed on traditional conventional features [13-17] based on liveliness cues which required task-aware knowledge for design. However, the liveness cues are inconvenient since they are obtained from long-term interactive videos. Also, they are susceptible to video attacks making them unreliable. Handcrafted features extracted over various color spaces are effective spoofing elements that carry texture,

[1,*]*Ph. D. Scholar, Computer Science & Engineering,*
Kalinga University, Chhattisgarh, India
Email ID: hirendrahajare@gmail.com
ORCID ID : 0000-0003-2177-5190
[2]*Professor, Computer Science and Engineering,*
Kalinga University, Chhattisgarh, India
drambhaikar@gmail.com

structure, and image quality details of the face region. However, they are prone to higher computational complexity, and illumination conditions, and are found unsuitable for inter datasets. Work proposed in [13][18][19] used liveness clues, gaze tracking in [22][23], physical movement of head and face in [20][21] and remote physiological indicators in [14][24][25][26] while handcrafted descriptors such as LBP, HOG, SIFT, DoG and SURF were part of [15], [16], [17], [28] and [27] respectively.

Despite the remarkable achievements of earlier face antispoofing schemes on intra-domain datasets, they performed poorly on inter-domain images. This is because inter-domain dataset images offer distinct characteristics that remain unaddressed related to internal relations. Therefore, the earlier models lack generalization ability and deep networks with supervised learning are an undistinguished part of most of the literature for antispoofing techniques.

The article contributes in the following aspects:

1. The face anti-spoofing dual unit framework offers a simple and robust preprocessing unit to enhance the face details while preserving the edges.

2. The ill-illuminated face images are contrast corrected by measuring the current contrast and then correcting it.

3. A small YOLOV5s network is used to extract blind quality features and classify authentic subjects from fake ones.

The next section deals with recent research contributions. Materials and a detailed description of the proposed face anti-spoofing framework are presented later followed by the experimental results combined with discussions. The last section concludes the article in a lucid manner with the merits and limitations of the proposed face anti-spoofing framework.

2. Related Work

The work proposed in [29] used 4-step preprocessing of the images from four different datasets including the OULU-NPU, MSU-MFSD, NUAA, and the Replay Attack datasets. The preprocessing primarily eliminated unwanted regions from the photo using a face cropping Dlib [] face detector to extract the face region. The next stage involved aligning the face using translation, scaling, and rotation operations about the line between the eyes. Further, redundant samples (frames) were discarded by sub-sampling the frames thus lowering the frame rate to examine the consequences of fewer training images. The redundant frames were eliminated based on the structural similarity metric. To reduce the amount of time required for training the images, transfer learning was used with VGG [30] previously trained for similar tasks for face antispoofing. Modified VGG16 trained on the ImageNet dataset was over another network due to its original accuracy for classifying the real and the spoofed faces. The performance was evaluated by considering images under controlled lighting conditions.

Face anti-spoofing based on facial landmarks detection and eye liveliness using a convolutional neural network classifier was proposed in [30]. The authors used a modified MobileNetV2 model to train the samples from the LCC FASD dataset. The reality of the person is determined using facial information such as posture, the opening of the mouth, the condition of the eyes, and direction of the eyeballs, and so on. The Dlib library was used to grab the 68 face liveness details from the face region and provided to the KNN model for training over 1942 real and 16885 spoofed faces. They performed data augmentation and the RGB channels of the images were averaged. A custom network involving a convolutional layer and a fully connected layer was used along the MobileNetV2 network to train and classify real and spoofed images. They obtained an accuracy of 98% over controlled lighting conditions.

The work proposed in [31] used the detected face for feature extraction after denoising the face image and converting the image to two different color spaces. They used Ycbcr and CIELuv color spaces for extracting features using the VGG network to output a 512-dimension vector. Further, they concatenated the features obtained from the color spaces using a pooling layer replacing the classification layer in the VGG network. They obtained an accuracy of 99.6% using SVM on the NUAA dataset. The face region from the photos of the dataset was detected using a Multitask cascaded convolutional neural network (MTCNN) [32] which comprises a P-net for prediction of face position and respective bounding boxes, R-net to eliminate false positive samples and the O-net to refine the bounding boxes thus improving the accuracy.

The authors in [33] used a similar tri-modal architecture to extract features from RGB, Depth, and IR. They used a convolutional module, and 3-RS blocks in each of the branches. Shallow features from the first RS block are spliced with the middle features of the third RS block. The features from all the branches are then concatenated after they are squeezed excited and fed to the fourth RS block. Finally, the features from the fourth RS block are fed to the classifier after they are passed through the global averaging pooling unit. They used the CASIA SURF dataset for evaluating their model and showed that their suggested face anti-spoofing model can preserve the details and enhance the representation of the enhanced features.

The work proposed in [34] introduced a multi-domain feature alignment framework called MADG to improve the generalization ability to unseen domains. An adversarial learning framework extracts features across cross-domain and is used collectively to constitute a multi-domain alignment technique. They used triplet mining to collect the differences between the real and the fake images by aligning the features from different domains. They tested their model on four different datasets which included MSU-MFSD, CASIA-FASD, OULU-NPU, and the IDIAP Replay Attack. The face regions were extracted using the MTCNN technique, rotated, and resized to 256x256x3 dimensions. They used modified RESNet-18 architecture for feature extraction and a fully connected CNN model for classification. Although their work was concentrated on feature alignment, they explored multi-domain problems.

An unobtrusive method to detect spoof attacks was presented in [35]. They used a Bi-lateral filter (Gaussian filter) to remove unwanted noise from the resized input image (256x256) and detected edges in the face region using the LoG filter (Laplacian of Gaussian) after enhancing the image. An Edge-Net Autoencoder was used to extract the dominant features from the enhanced face region. The dimension of the extracted features is reduced and classified using fully connected CNN. The model was evaluated on three datasets including the IDIAP, CASIA FASD, and self-generated Edge-Net dataset. They classified real and fake images with an accuracy of about 99% and 100% on publicly available datasets and self-generated datasets respectively.

3. Materials and Method

The IDIAP dataset with Replay Attack under consideration consists of 4000 real images and 9950 spoofed images belonging to 80 real and 199 fake subjects converted from videos respectively. Each subject either real or fake has a distinct number of images. We have taken the first 200 images approximately from each folder corresponding to each of the subjects. The details are provided in [36]. Figure 1 and 2 shows images of real face images and spoofed face images for the same subjects. The proposed System for Classification is shown in Figure 3 below.

We partitioned our system into two stages: The pre-processing unit, and the feature extraction followed by the classification unit. The input images from the dataset are either real or spoofed corresponding to 80 real and 199 fake subjects. The pre-processing stage comprises two parallel sections converging at the averaging unit as shown in Figure 3. The first section computes the contrast and corrects it to improve the quality of the image to obtain quality features while the next section filters the image for edge preservation and enhancement.

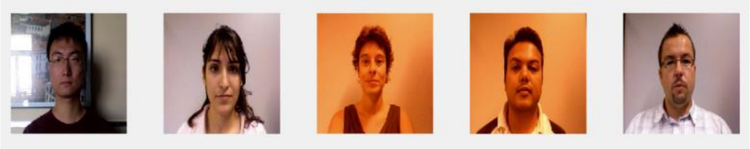

Fig. 1. Real images extracted from the dataset videos

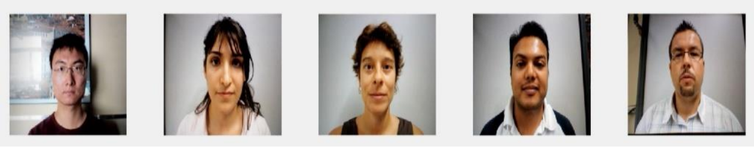

Fig. 2. Spoofed images extracted from the dataset videos

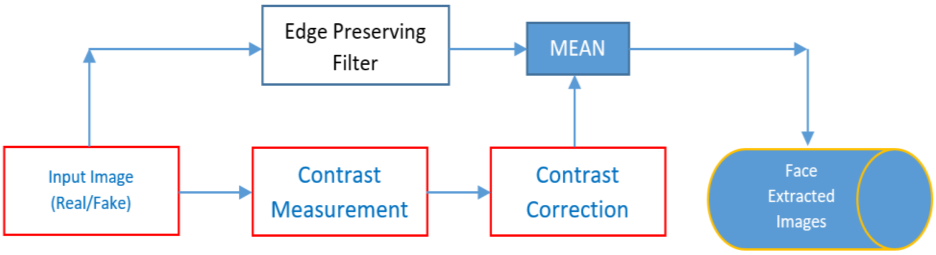

Fig. 3. The Pre-processing & Face extraction system

The perceived contrast of an image is influenced by viewing conditions and the spatial arrangements of the image and measurement of such contrast is not so simple. The parameters that affect the image contrast involve color, contents, illumination, viewing distance, resolution, etc. Thus only measuring the difference between the brightest and the darkest point measures the perceived contrast [37]. Out of many such local and global contrast measuring classic approaches, Tadmor and Tolhurst's [38] global approach has been used for measuring the contrast of images. Figure 4 shows the output of modified DOG filtering. The concept is modified and adapted to the difference of the Gaussian (DOG) model. They proposed equation (1) for measuring the contrast.

$$C^{\varpi}(x,y) = \frac{R_C(x,y)-R_S(x,y)}{R_C(x,y)+R_S(x,y)} \quad (1)$$

Where the output of the central component is,

$$R_c(x,y) = \sum_{i=x-3r_c}^{i=x+3r_c} * \sum_{j=y-3r_c}^{j=y+3r_c} Center\ (i-x, j-y) I(i,j) \quad (2)$$

While the output of the surround component is,

$$R_s(x,y) = \sum_{i=x-3r_c}^{i=x+3r_c} * \sum_{j=y-3r_c}^{j=y+3r_c} Surround\ (i-x, j-y) I(i,j) \quad (3)$$

The center and surrounding components of the receptive field are given by,

$$Center\ (x,y) = exp\left[-\left(\frac{x}{r_c}\right)\left(\frac{x}{r_c}\right) - \left(\frac{y}{r_c}\right)\left(\frac{y}{r_c}\right)\right] \quad (4)$$

(x,y) is the spatial coordinates of the receptive field, and r_c is the radius at which the sensitivity decreases to *1/e* w. r. t. the peak level.

$$Surround\ (x,y) = 0.85\left(\frac{r_c}{r_s}\right) exp\left[-\left(\frac{x}{r_s}\right)\left(\frac{x}{r_s}\right) - \left(\frac{y}{s}\right)\left(\frac{y}{r_s}\right)\right] \quad (5)$$

Such that $r_s > r_c$.

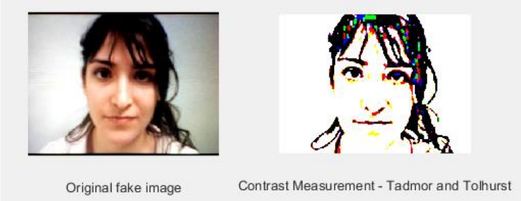

Fig. 4. Contrast measurement using the Tadmor and Tolhurst method

We enhanced the contrast using the following technique. The result showed improved and acceptable contrast over the parent images. The images were then converted to grayscale and concatenated to form a 4D array to reduce the processing time for feature extraction. We applied the following correction technique to the grayscale image and all three frames of the color image independently.

$$M = 255*CM \quad (6)$$

$$Factor = 259 * \frac{(M+255)}{(255*(259-M))} \quad (7)$$

$$G = (Factor * (I - 128)) + 128 \quad (8)$$

The work in [39] introduced an edge-preserving and denoising filter for 2D and 3D images and extended it to patches for feature extraction. The framework considers a color image in a hybrid special-spectral 5D space *{x, y, R, G, B}*. The filter requires two tuning parameters and includes the time step for stability (Usually set to the reciprocal of the squared number of dimensions) and iterations for which the filter operates. Beltrami filter is capable of removing aliasing and weak textures while preserving the edge's fine structure. We used the filter on each of the color channels of the input image *A* separately with 20 iterations and a time step of 0.5 to obtain the filtered image *C.*

The filtered and contrast-corrected images *G* and *C* were considered and the average of the two images *(F)* was used to segment the face region of the original image to discard any unwanted region that may remain due to poor contrast or blurred edges. Thus the expression for image *F* is given by the following expression (9), *'i'* representing the color channel for color image.

$$F = \frac{1}{2}[G_i + C_i] \quad (9)$$

The face region I_{face} was extracted using the bounding box algorithm in MATLAB which covers the region from head to neck so that significant features could be extracted for better accuracy. The extracted face region was resized to the dimension of *[120 120 3]* since the bounding boxes for each individual were of varying sizes and could lead to variable feature sizes. Figure 5 shows the outputs of the preprocessing stages. Perceptually being similar, the PSNR (Peak Signal to Noise Ratio) between the output and the original image (first image) is indicated below each output. The PSNR value (female image) is 42.3671 for the filtered image, 33.1709 for the Contrast corrected image, and 38.5852 for the mean image concerning the original image. Likewise, it is 41.9414, 38.9798, and 42.8129 respectively for the male image. Figure 6 represents the extraction of the region of interest (face) using the Bounding Box algorithm and Figure 7 shows real face images from the training set about different subjects obtained using the proposed approach.

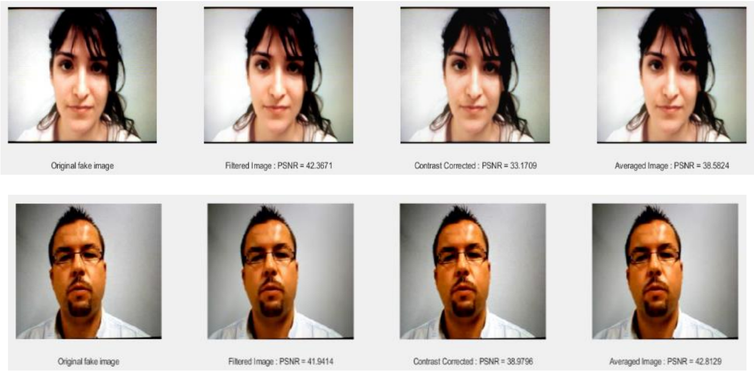
Original fake image
Filtered Image : PSNR = 42.3671
Contrast Corrected : PSNR = 33.1709
Averaged Image : PSNR = 38.5824
Original fake image
Filtered Image : PSNR = 41.9414
Contrast Corrected : PSNR = 38.9796
Averaged Image : PSNR = 42.8129

Fig. 5. The preprocessing stage. Original input image, Filtered image, Contrast corrected image, and the Averaged image. The PSNR values reflect pixel value changes in each stage.

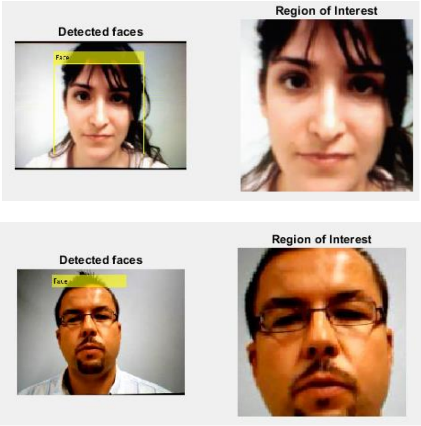

Fig. 6. Region of Interest (FACE) using Bounding Box Algorithm

Fig. 7. Face region extracted using the Bounding Box Algorithm

All the images in the dataset are pre-processed likewise and the automatic face cropped images are stored in a separate dataset. Experimental analysis showed that our pre-processing approach failed for some subject sample images in the real folder. The percentage of failure to extract the actual face from the sample images was negligible as compared to the total available images in the dataset belonging to the real class. Therefore we neglected those samples due to their poor contrast and interference of other objects making the framework unfit to extract the face region. The first challenge was to handle data imbalance (80 real and 199 fake subjects) and later the number of samples to be considered from both categories. To check the robustness of our pre-processing and feature extraction plus classification units, we decided to select 200 samples from the real class and 50 samples from the fake class. After pre-processing, we obtained 16768 samples for the real class and 9950 samples for the fake class (Neglecting the failure samples from the real class). We partitioned the face samples into real and fake categories as depicted in Table 1.

Table 1. Training, testing, and validation samples from real and fake classes after pre-processing.

Class	Total samples	Training samples	Validation samples	Test samples
Real	16768	11742	1678	3348
Fake	9950	6965	995	1990

Blind feature extraction and classification were performed using YOLOV5s deep network through transfer learning. The 120x120x3 face images from the training set and the validation sets were provided to the YOLOV5s network for 25 epochs. The weights obtained through transfer learning were used to classify the test samples using the trained YOLOV5s network. The following Figure 8 shows the feature extraction plus classification unit.

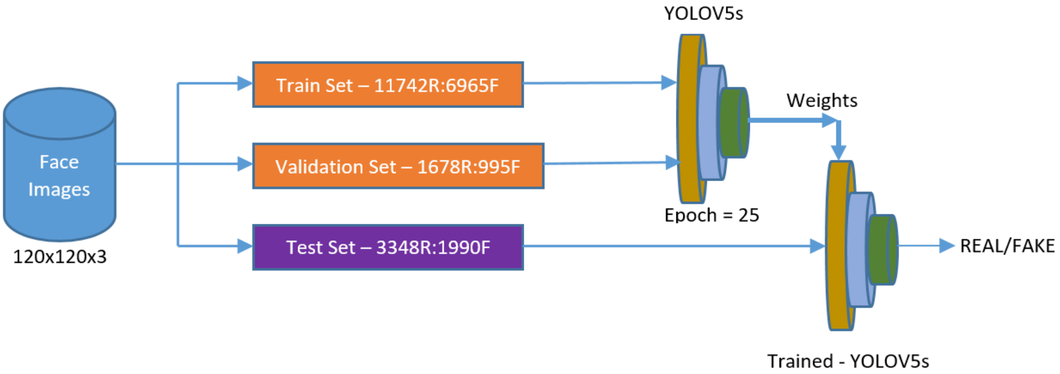

Fig. 8. The feature extraction plus classification unit.

4. Results and Discussion

The two units-based face antispoofing framework was developed on two separate platforms. The pre-processing unit was carried out on MATLAB 2021b while the feature extraction plus classification was subjected to the YOLOV5s network on Python 3.9-based SYPDER. Both the units were installed on a Windows 11 environment with an i5 processor (2.71 GHz), 16 GB RAM, and 512 GB SSD. We obtained 97.48% classification accuracy over the real samples and 98.92% on the fake samples when independently tested using the YOLOV5s network. The overall accuracy was 98.39% when random samples were subjected and tested for 20 iterations as a part of cross-validation from all face images obtained after the pre-processing. The performance of the YOLOV5s network was limited due to data imbalance concerning the number of real and fake subjects as well as the number of samples considered for training the network. We used approximately 200 samples for each real subject while only 50 samples belonged to the fake images. The original dataset folders belonging to each of the classes contained an uneven number of samples related to each subject. As a result, the only solution to balance the data was through augmentation which was not part of our work.

Occlusions due to aspects, scarf, and hairs, partial poor contrast, and incomplete face region significantly contributed to the complexity. Figure 9 shows examples of spectacle over face, ROI surrounded by a scarf, poor lightning over partial ROI, and incomplete face features. The proposed framework for antispoofing achieved higher detection accuracy due to an efficient pre-processing mechanism irrespective of various face occlusions, incomplete face details, and uneven illuminations. Examples of uneven foreground illuminations are shown in Figure 10. We set the bounding box coordinates with an offset to cover the utmost details of the face region for better results by experimenting over a large number of samples and fixed the threshold values for the offset. The offset was a compromise to fit several samples from the dataset against losing details for a few samples. We found that the details such as ears, lower chin, and forehead were eliminated in a few samples. Also, eliminating the non-ROI region improved the detection accuracy. No face alignment strategy was adopted in our method.

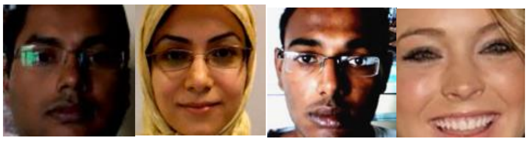

Fig. 9. Face occlusions and incomplete face details.

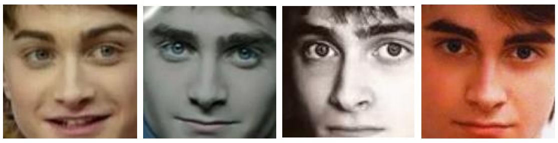

Fig. 10. Different foreground lightning for a single subject.

The following Table 2 shows the comparison of our proposed face antispoofing framework using the YOLOV%s network and other recent techniques found in the literature. The performance of the proposed face antispoofing framework is superior to work in [30] and [40] and nearer to techniques suggested in [35] and [41] while nearer to work suggested in [[42]. However, the computational complexity of our proposed model concerning the extraction of face region is low as well we have used the small YOLOV5 model for feature extraction and classification. The work proposed in [35] uses two different deep networks: the Edge-Net Autoencoder and the CNN in the preprocessing and the classification stage respectively. Work in [41] used a motion amplification algorithm for enhancing the frames 20 times and a two-input CNN for feature extraction and classification.

Table 2. Performance comparison of proposed face antispoofing framework.

Ref.	Year	Network	Dataset	% Acc.
[30]	2023	MobileNetV2	LCC-FASD	98
[40]		-	Self-Generated	97.3
[35]		Edge-Net Autoencoder	Replay Attack	99.5
[41]		CNN	IDIAP - Replay Attack	99.34
[42]		Deep CNN	FPAD - Replay Attack	98.67
Ours		YOLOV5s	IDIAP - Replay Attack	98.39

5. Conclusion

The proposed face antispoofing framework offers a simple but efficient mechanism to distinguish between fake and authentic faces. It is robust to foreground illumination variations, face occlusions, and incomplete face information. One of the preprocessing units not only enhances the details but also preserves the details while the other calculates the current contrast level and corrects the contrast of the input image. The averaging unit holds the quality details present in the image and helps the deep network YOLOV5s model to discriminate between authentic and fake images with higher accuracy. Despite the data imbalance between the subjects and the samples the proposed framework can perform well. The system offers low computational complexity and possesses generalization ability over the Replay attacks. Moreover, it fails to preprocess some samples from the real face datasets as a consequence of the presence of other background objects and poor contrast. We considered imbalanced samples from both classes due to the scarcity of samples in the fake folder. As seen from Figures 9 and 10, a generalized contrast correction algorithm is difficult to design. Also, a region under occlusion may be omitted for performance improvement. Data augmentation can be performed to balance the two classes using various operations.

References

[1] Jain, A.K. and Ross, A., “Handbook of biometrics”, pp. 1–22. Springer, London (2008)

[2] Sharma, D., and Selwal, A., “FinPAD: state-of-the-art of fingerprint presentation attack detection mechanisms, taxonomy and future perspectives”, Pattern Recognit. Lett. Volume 152, Issue 1, pp. 225–252, 2021.

[3] Jain A. K., Ross A., and Prabhakar S., “An introduction to biometric recognition”, IEEE Trans. Circuits Syst. Video Technol., Volume 14, Issue 1, pp. 4–20, 2004.

[4] Selwal A., Gupta S. K., and Kumar S., “A scheme for template security at feature fusion level”, Adv. Sci. Technol. Res. J., Volume 10, Issue 31, pp. 23–30, 2016.

[5] Jain A.K., Ross A., Pankanti S., and Member S., “Biometrics: a tool for information security”, IEEE Trans. Inf. Foren. Secure, Volume 1, Issue 2, pp. 125–143, 2006.

[6] Manzoor S. I. and Selwal A., “An analysis of biometric-based”, In: 2018 Fifth Int. Conf. Parallel, Distrib. Grid Comput., no. 4, pp. 306–311, 2018.

[7] Sharma D. and Selwal A., “An intelligent approach for fingerprint presentation attack detection using ensemble learning with improved local image features”, No 0123456789. Springer, 2021.

[8] Kresimir D. and Mislav G., “A survey of biometric recognition methods”, In: 46th Int. Symposium Electron. Mar. ELMAR-2004. 16–18 June, 2004. Zadar. Croat. A, no. June, pp. 184–193, 2004.

[9] Akhtar Z. and Foresti G. L., “Face spoof attack recognition using discriminative image patches”, J. Electr. Comput. Eng. 2016.

[10] S. Kumar, S. Singh, and J. Kumar, “A comparative study on face spoofing attacks”, in 2017 International Conference on Computing, Communication, and Automation (ICCCA), pp. 1104–1108, IEEE, 2017.

[11] L. Sun, W. Huang, and M. Wu, “Tir/vis correlation for liveness detection in face recognition”, in International Conference on Computer Analysis of Images and Patterns, pp. 114–121, Springer, 2011.

[12] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, “Face liveness detection by learning multispectral reflectance distributions”, in 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), pp. 436– 441, IEEE, 2011.

[13] G. Pan, L. Sun, Z. Wu, and S. Lao, “Eyeblink-based anti-spoofing in face recognition from a generic web camera”, in Proc. IEEE Int. Conf. Comput. Vis., pp. 1–8, 2007.

[14] X. Li, J. Komulainen, G. Zhao, P.-C. Yuen, and M. Pietikainen, “Generalized face anti-spoofing by detecting pulse from face videos”, in Proc. IEEE 23rd Int. Conf. Pattern Recognit., pp. 4244–4249, 2016.

[15] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, “LBP - TOP based countermeasure against face spoofing attacks”, in Proc. Asian Conf. Comput. Vis., pp. 121–132, 2012.

[16] J. Komulainen, A. Hadid, and M. Pietikainen, “Context-based face anti-spoofing”, in Proc. IEEE 6th Int. Conf. Biometrics: Theory Appl. Syst., pp. 1–8, 2013.

[17] K. Patel, H. Han, and A. K. Jain, “Secure face unlock: Spoof detection on smartphones”, IEEE Trans. Inf. Forensics Security, vol. 11, no. 10, pp. 2268–2283, Oct. 2016.

[18] H.-K. Jee, S.-U. Jung, and J.-H. Yoo, “Liveness detection for embedded face recognition system”, Int. J. Biol. Med. Sci., vol. 1, pp. 235–238, 2006.

[19] J.-W. Li, “Eye blink detection based on multiple Gabor response waves”, in Proc. IEEE Int. Conf. Mach. Learn. Cybern., pp. 2852–2856, 2008.

[20] L. Wang, X. Ding, and C. Fang, “Face live detection method based on physiological motion analysis”, Tsinghua Sci. Technol., vol. 14, no. 6, pp. 685–690, 2009.

[21] W. Bao, H. Li, N. Li, and W. Jiang, “A liveness detection method for face recognition based on optical flow field”, in Proc. IEEE Int. Conf. Acoust. Speech Signal Processing, pp. 233–236, 2009.

[22] J. Bigun, H. Fronthaler, and K. Kollreider, “Assuring liveness in biometric identity authentication by real-time face tracking”, in Proc. IEEE Int. Conf. Comput. Intell. Homeland Security. Pers. Saf., pp. 104–111, 2004.

[23] Ali, F. Deravi, and S. Hoque, “Liveness detection using gaze collinearity”, in Proc. IEEE 3rd Int. Conf. Emerg. Secure. Technol., pp. 62–65, 2012.

[24] Y. Liu, A. Jourabloo, and X. Liu, “Learning deep models for face anti-spoofing: Binary or auxiliary supervision”, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 389–398, 2018.

[25] B. Lin, X. Li, Z. Yu, and G. Zhao, “Face liveness detection by rPPG features and contextual patch-based CNN”, in Proc. 3rd Int. Conf. Biometric Eng. Appl., pp. 61–68. 2019.

[26] Z. Yu, W. Peng, X. Li, X. Hong, and G. Zhao, “Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement”, in Proc. IEEE Int. Conf. Comput. Vis., pp. 151–160, 2019.

[27] Z. Boulkenafet, J. Komulainen, and A. Hadid, “Face antispoofing using speeded-up robust features and fisher vector encoding”, IEEE Signal Process. Lett., vol. 24, no. 2, pp. 141–145, Feb. 2017.

[28] X. Tan, Y. Li, J. Liu, and L. Jiang, “Face liveness detection from a single image with the sparse low rank bilinear discriminative model”, in Proc. Eur. Conf. Comput. Vis., pp. 504–517, 2010.

[29] Sandoval Verissimo, Guilherme Gadelha, Leonardo Batista, Joao Janduy and Fabio Falcao, “Transfer Learning for Face Anti-Spoofing Detection,” IEEE Latin America Transactions, Volume 21, No. 4, April 2023.

[30] Ruchi Zawar and Vrishali Chakkarwar, “Real-Time Face Liveliness Detection and Face Anti-spoofing Using Deep Learning”, ACVAIT 2022, AISR 176, pp. 626-636, 2023.

[31] K Balamurali et al., “Journal of Physics: Conference Series”, 1917, 012010, 2021.

[32] Zhang K., Zhang Z., Li Z. and Qiao Y., “Joint face detection and alignment using multitask cascaded convolutional networks”, IEEE Signal Process. Letters, Volume 23, pp. 1499–1503, 2016.

[33] Chunyan Li, Zhiyong, Jianhong Sun and Rui Li, “Middle-shallow feature aggregation in multimodality for face anti-spoofing”, Scientific Reports, Volume 13, 9870, 2023.

[34] Zhang S. and Nie W., “Multi-Domain Feature Alignment for Face Anti-Spoofing”, Sensors, Volume 23, 4077, 2023.

[35] Amal H. Alharbi, S. Karthick, K. Venkatachalam, Mohamed Abouhawwash and Doaa Sami Khafaga, “Spoofing Face Detection Using Novel Edge-Net Autoencoder for Security”, Intelligent Automation & Soft Computing, Volume 35, No. 3, pp. 2773-2787, 2023.

[36] Pereira Tiago de Freitas, Komulainen Jukka, Anjos Andre, Martino Jose Mario De, Hadid Abdenour, Pietikainen Matti and Marcel Sebastien, “Face liveness detection using dynamic texture”, EURASIP Journal on Image and Video Processing, volume 2, 2014.

[37] Simone Gabriele, Pedersen Marius, and Hardeberg John Yngve, “Measuring perceptual contrast in digital images”, Journal of Vis. Communication R., volume 23, pp. 491-506, 2012.

[38] Tadmor Y. and Tolhurst D., “Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes.”, Vision Research, Volume 40, Issue 22, pp. 3145–3157, 2000.

[39] Wetzler A. and Kimmel R., “Efficient Beltrami Flow in Patch-Space. In: Bruckstein, A.M., ter Haar Romeny, B.M., Bronstein, A.M., Bronstein, M.M. (eds) Scale Space and Variational Methods in Computer Vision”, SSVM 2011. Lecture Notes in Computer Science, volume 6667. Springer, Berlin, Heidelberg, pp. 134-143, 2012.

[40] Cuong Nguyen The, Vladimir Ivanovich Syryamkin, Thang Nguyen Chien, Trang Nguyen Hoang Thuy and Lyanshenko Dmtriy, “Evaluating methods of anti-spoofing of living entities and propose solutions”, E3S Web of Conferences, 389, 02023, 2023.

[41] Xin Chen, Jingmei Zhou, Xiangmo Zhao, Hongfei Wang and Yuqi Li., “A presentation attack detection network based on dynamic convolution and multilevel feature fusion with security and reliability”, Future Generation Computer Systems, Volume 146, September 2023, pp. 114-121, 2023.

[42] S. D. Thepade, M. R. Dindorkar, P. R. Chaudhari and S. V. Bang, “Enhanced Face Presentation Attack Prevention Employing Feature Fusion of Pretrained Deep CNN Model and Thepade’s Sorted Block Truncation Coding. IJE Transactions A: Basics, Volume 36, No. 04. pp. 807-816, 2023.