Enhancing Face Liveness Detection: Novel Deep CNN Architectures for Anti-Spoofing

Swapnil R. Shinde
Department of Computer Science and Information Technology, Symbiosis Institute of Technology, Symbiosis International (Deemed University) Lavale, Pune, Maharashtra, India | Department of Engineering and Technology, Bharati Vidyapeeth (Deemed University), Kharghar, Navi Mumbai, Maharashtra, India
swapnil.shinde.phd2019@sitpune.edu.in

Anupkumar M. Bongale
Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology, Symbiosis International (Deemed University) Lavale, Pune, Maharashtra, India
anupkumar.bongale@sitpune.edu.in (corresponding author)

Deepak Dharrao
Department of Computer Science and Engineering, Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune, Maharashtra, India
deepak.dharrao@sitpune.edu.in

Dipti Jadhav
Department of Computer Science and Engineering, Ramrao Adik Institute of Tecnology, D Y Patil deemed to be University, Nerul,Navi Mumbai, Maharashtra, India
dipti.jadhav@rait.ac.in

Nilesh Yadav
Department of Computer Engineering, K. J. Somaiya Institute of Technology, Sion, Mumbai, Maharashtra, India
nilesh.yadav@somaiya.edu

Received: 29 May 2025 | Revised: 7 July 2025, 20 July 2025, and 24 July 2025 | Accepted: 27 July 2025

 DOI: https://doi.org/10.48084/etasr.12431

ABSTRACT

In today's technologically advanced environment, security is of paramount importance, and various methods are employed to ensure robust protection and implement strong security measures. Biometric security can be achieved using various means such as strong keys, efficient key-value pair challenges, etc. Biometric authentication systems use physiological or behavioral mechanisms to offer robustness in identity verification. In terms of physical aspects of authentication security, faces, fingerprints, palms, etc. are used as primary modalities. Face is the most widely used modality due to its ease of availability, but it is easily susceptible to spoof attacks. Attacks on face systems are broadly categorized into 2D and 3D attacks. Leveraging deep learning methods with CNN architectures has shown efficacy in detecting spoofing attempts. However, since pre-trained models have complex architecture and involve large training times, there is a need for simple and efficient architectures with lower computational cost for face liveness detection. This paper introduces two novel deep CNN architectures that have fewer parameters, require less computational resources for execution, and achieve improved results. A comparative analysis with existing methods was conducted using the NUAA Imposter Database and the 3D MAD dataset on RGB and YCBCR color spaces. One proposed architecture achieved an impressive accuracy of 99.87% in detecting 3D face spoofing attacks, with a Half Total Error Rate (HTER) of just 0.19%, outperforming existing methods. The proposed CNN architectures exhibit promising outcomes in enhancing the generalization of attack detection systems with lower computational cost.

Keywords-biometric security; Convolutional Neural Network (CNN); face spoofing; HTER; pre-trained architecture

I. INTRODUCTION

Biometrics stands as a paramount approach for authenticating individuals due to its inherent security and robustness, which encompasses two main mechanisms: physiological and behavioral. Physiological traits [1] utilized in biometric authentication consider features such as palm, fingerprint, face, retina, and iris. Among these, face recognition has been a prominent topic in various high-impact application areas. Face Recognition Systems (FRS) are susceptible to presentation attacks, which can lead to data and privacy losses [2, 3]. Presentation attacks in FRS are mainly categorized as 2D and 3D [4], which include methods such as photo printing, image display, video replay, and mask attacks using a 3D face. Detection of 3D attacks involves extracting pertinent features from the input images, often using State-Of-The-Art (SOTA) mechanisms, with an emphasis on the extraction of shape- and texture-based features from the input 3D images [5]. The systems designed for the detection of presentation attacks are broadly categorized into hardware- or software-based, with some being hybrid combinations. Software-based systems focus mainly on extracting meaningful features from data using diverse methods and evaluating performance through machine learning approaches [6]. Deep learning methods that apply Convolutional Neural Networks (CNNs) [7] have presented promising results for feature extraction and classification tasks, surpassing traditional machine learning methods. Previous studies have proposed systems that integrate handcrafted features with deep learning-based features to enhance face spoof detection. These hybrid approaches leverage the strengths of both techniques, handcrafted methods for robust feature extraction and deep models for powerful classification. Some systems generated feature vectors using handcrafted techniques and then applied transfer learning architectures for classification. This combination has shown promising results in improving detection accuracy across various attack scenarios.

In the past decade, several studies have explored the application of transfer-learning models for the detection of face spoofing attacks, showcasing innovative approaches that leverage advanced techniques to enhance detection accuracy and robustness. A notable study [8] introduced a mechanism that used two color spaces, CIELUV and YCbCr, along with the VGG-Face model for the detection of spoof attacks. This method involved denoising face samples with conversion to specified color spaces before the application of the VGG-Face model. Face identification was achieved through a multi-CNN model, while non-local means denoising was employed for noise reduction. The VGG-Face model was an optimized version of VGG-16 that replaces regular pooling with average pooling for feature extraction. In [9], features were extracted using the AlexNet and VGG16 transfer learning models, and the activation features of fully connected layers were combined and passed to an SVM classifier. In [10], a unique system integrated face liveness detection into the authentication process before face recognition. This method trained a Siamese network on image pairs, followed by feature extraction and classification using AlexNet.

In [11], a deformable convolution layer was applied in the MobileNetV2 architecture, helping the model adapt to dynamically changing spatial variation to capture distinct and robust features from the input and demonstrate improved results. Although pre-trained models offer performance improvements, their extensive layers and training times introduce significant overhead. Thus, some CNN architectures have been proposed to address the issue. Recent advances in CNNs have led to the development of multi-channel architectures [12], accommodating diverse input sources such as grayscale, RGB, infrared, and thermal images. In [13], an attention-based method utilized a dual-stream CNN (TSCNN), fusing features extracted from RGB and Multi-Scale Retinex Space (MSR) by applying the ResNet architecture. By incorporating attention mechanisms, the system effectively classified and matched features, yielding significant improvements over traditional handcrafted feature models. In [14], a fusion architecture integrated handcrafted with deep features to enhance face spoofing detection. This approach combined texture features, extracted using LBP variants, with deep features obtained from a deep CNN architecture, yielding improved generalization capabilities across different datasets.

These studies presented hybrid models that combined transfer learning and CNN architectures. However, these hybrid models involve a large number of parameters, introducing training overhead and high computational cost. This study introduces two CNN architectures that enhance performance and address the concerns of computational requirements. The contributions of this study are as follows:

- Face Liveness Network-1 (FLNet-1) has multiple hidden layers for generalization in face spoofing attack detection and is evaluated on two color spaces.
- Face Liveness Network-2 (FLNet-2) is an improvement over FLNet-1, with an additional convolution block for improved performance in face spoofing attack detection.
- Presents a comparative analysis between existing methods and the proposed architectures, accompanied by a detailed discussion.

II. PROPOSED FACE LIVENESS DETECTION ARCHITECTURE

This study presents simplified Deep Neural Network (DNN) models that leverage the strengths of convolution layers while mitigating their drawbacks. Figures 1 and 2 show the proposed network architecture diagrams.

A. FLNet-1

The proposed FLNet-1 consists of two convolution layers, two max-pooling layers, one fully connected layer, and one dropout layer. The FLNet-1 architecture uses two convolution layers with increasing filters (32 and 64) and 3×3 kernels, which helps to extract low to mid-level features, such as edges, shapes, textures, etc., from the images. The max-pooling layers help to reduce dimensionality and control model complexity, making it translation invariant. The flatten layer creates a 1D

vector for the dense layer to extract high-level features. The dropout layer is added before the classification task to reduce overfitting and pass only distinct and effective features to the next step. The layers and their parameters are selected based on multiple iterations after a thorough analysis of their functioning characteristics. FLNet-1 uses L2 regularization and the RMSPROP optimizer with a Learning Rate (LR) of 10^{-4}.

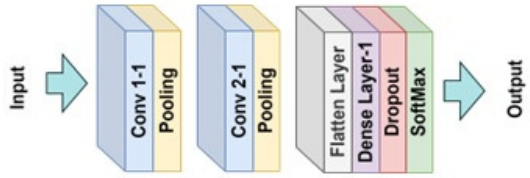


Fig. 1. Proposed Face Liveness Network 1 (FLNet-1) architecture

B. FLNet-2

The proposed FLNet-2 has three max-pooling layers, three convolution layers, two fully connected layers, and two dropout layers. FLNet-2 uses three convolution layers with an increasing number of filters (32, 64, 128) and the same kernel size of 3×3. The additional convolution and max-pooling layers help the model extract deeper and more abstract features from the input image. A balanced increase in depth and regularization boosts the model's generalization with increasing computational cost. Multiple dropout layers reduce overfitting and add to the model's performance.

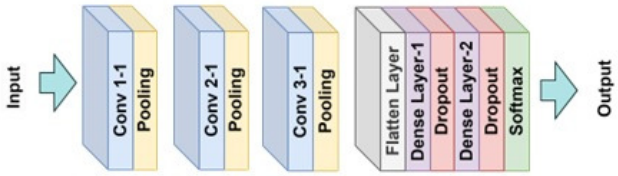


Fig. 2. Proposed Face Liveness Network (FLNet-2) Architecture

The mathematical representation for the layers is given in the following section, along with the FLNet-1 and FLNet-2 layer-wise details. Both architectures share some initial layers, with the remaining being unique for each architecture.

C. Detailed Architecture of the Models

1) Common CNN Layers (Steps 1-3)

Input: $X \in \mathbb{R}^{\{64\times64\times C\}}$, where C is the number of input channels.

- Step 1: Apply 32 convolution filters of size 3×3 followed by ReLU activation:

  $F_i = ReLU(X * W_i + b_i), i = 1,\ldots,32$

  Output: $F \in \mathbb{R}^{\{64\times64\times32\}}$

- Step 2: Apply max pooling with a 2×2 window:

  $P_{\{i,j,k\}} = max(F_{\{2i:2i+1,2j:2j+1,k\}})$

  Output: $P \in \mathbb{R}^{\{32\times32\times32\}}$

- Step 3: Apply 64 convolution filters and max-pooling:

  Output: $P' \in \mathbb{R}^{\{16\times16\times64\}}$

2) FLNet-1: Unique Steps (Steps 4-7)

- Step 4: Flatten the feature map:

  $Flatten(P') \in \mathbb{R}^{\{16384\}}$

- Step 5: Dense layer with ReLU activation:

  $h_1 = ReLU(W_{dense1} \times Flatten(P') + b_{dense1})$

- Step 6: Apply dropout with a rate of 20%:

  $h_1^{drop} = Dropout(h_1, p = 0.2)$

- Step 7: Final dense layer and softmax activation:

  $z = W_{dense2} \times h_1^{drop} + b_{dense2}$

  $\hat{y}_i = exp(z_i) / (exp(z_0) + exp(z_1)),\ i = 0,1$

3) FLNet-2: Unique Steps (Steps 4-9)

- Step 4: Apply 128 convolution filters and max pooling:

  Output: $P' \in \mathbb{R}^{\{8\times8\times128\}}$

- Step 5: Flatten the feature map:

  $Flatten(P') \in \mathbb{R}^{\{8192\}}$

- Step 6: Apply dropout with a rate of 40%:

  $h_1^{drop} = Dropout(h_1, p = 0.4)$

- Step 7: Second dense layer with ReLU activation:

  $h_2 = ReLU(W_{dense2} \times h_1^{drop} + b_{dense2})$

- Step 8: Apply dropout with a rate of 20%:

  $h_2^{drop} = Dropout(h_2, p = 0.2)$

- Step 9: Final dense layer and softmax activation:

  $z = W_{dense3} \times h_2^{drop} + b_{dense3}$

  $\hat{y}_i = \frac{exp(z_i)}{\exp(z_0)} + \exp(z_1),\ i = 0,1$

where W_i denotes the weight matrices used in the respective layers, h_i denotes the output of the hidden layers, and F and P represent the output feature map for the convolution and max-pooling layers.

Table I compares both architectures in terms of standard parameters.

TABLE I. COMPARISON OF PROPOSED ARCHITECTURES

Parameters/ Model	FLNeT-1	FLNeT-2
Number of blocks	5	8
Input size	64×64×3	64×64×3
Output size (No of parameters)	258	514
Trainable parameters	2,116,930	1,175,490
Optimizer	RMS Prop	Adam
Learning rate	10^{-4}	10^{-4}
FLOPs (per image)	~24.4 MFLOPs	~84 MFLOPs

D. Computational Analysis

Table I shows the number of parameters and the FLOPs for the proposed architectures. Table IV shows a comparison with existing methods. These two tables show that both FLNeT architectures require fewer computational resources and execute the training process faster than existing models.

III. EXPERIMENTATION

The proposed approach was tested on the 3D MAD dataset [15, 16] and the NUAA photo imposter dataset [17, 18]. In 3D MAD, three sessions were considered for recording the videos, where two sessions represent legitimate access, and one session is for illegitimate access, i.e., mask. The dataset comprises images of 17 subjects, with each session containing five videos per subject. Each recording comprises 300 frames, recorded at a frame rate of 30 fps. The Kinect 3D sensor was used to record the dataset images. The NUAA dataset contains 15 subjects, with 5,105 genuine cases and 7,509 attack cases. The dataset was captured in various locations and with changing light conditions. The attack images were obtained by capturing HD images with a digital camera.

The proposed architectures were implemented on a system with a Core i7 CPU at 3.60 GHz, 16 GB of RAM, and a 6 GB GPU. Training was performed with a batch size of 64 and for 20 and 25 epochs, respectively, for both color spaces (RGB and YCBCR). The evaluation parameters considered were Accuracy (ACC), Bona fide Presentation Classification Error Rate (BPCER) [19], Half Total Error Rate (HTER) [20], and Attack Presentation Classification Error Rate (APCER). The formulas for all the evaluation parameters are:

$$\text{ACC} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (1)$$

$$\text{APCER} = \frac{\text{FP}}{\text{TN+FP}} \quad (2)$$

$$\text{BPCER} = \frac{\text{FN}}{\text{TP+FN}} \quad (3)$$

$$\text{HTER} = \frac{\text{FPR+FNR}}{2} \quad (4)$$

IV. RESULTS AND DISCUSSION

A. 3D Attack Dataset(3D MAD) Results

The proposed FLNet-1 model achieved best performance at Epoch 25 with an accuracy of 99.35% for the RGB color space, whereas FLNet-2 demonstrated superior performance at Epochs 20 and 25, achieving a peak accuracy of 99.87% for the same color space, as can be seen in Figure 3. In terms of performance metrics for FLNet-1, the lowest values for APCER, BPCER, and HTER were observed at Epoch 25 with 1.18%, 0.39%, and 0.78%, respectively, as illustrated in Figure 4. The best performance metrics for FLNet-2 were obtained for Epochs 20 and 25, with 0.39%, 0%, and 0.19% for APCER, BPCER, and HTER, respectively, as shown in Figure 5.

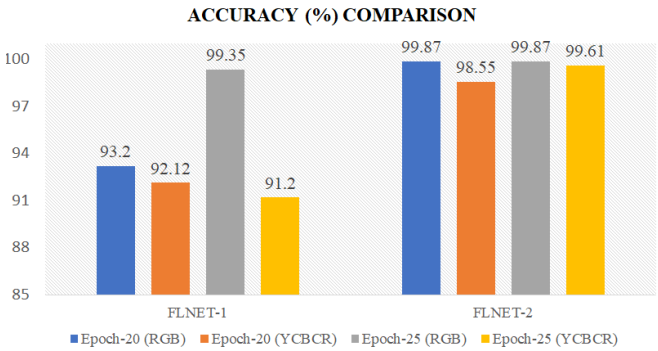


Fig. 3. Accuracy comparison of the proposed system for RGB and YCBCR color spaces on the 3D MAD dataset.

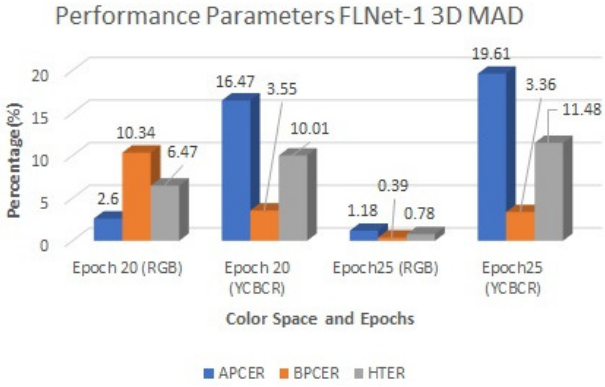


Fig. 4. FLNet-1 results for standard parameters on 3D MAD dataset.

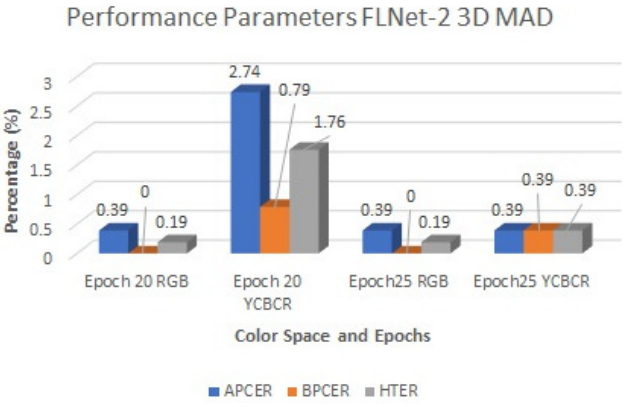


Fig. 5. FLNet-2 results for standard parameters on the 3D MAD dataset.

B. 2D Attack Dataset(NUAA) Results

The proposed FLNet-1 performed best at Epoch 25, where it achieved the best Accuracy of 95.32% for the YCBCR color space, while FLNet-2 achieved an accuracy of 98.61% on the same color space, as can be seen in Figure 6. At Epoch 25, FLNet-1 achieved error rates of 3.86%,5.46% and 4.66% for APCER, BPCER, and HTER, respectively, as shown in Figure 7. At Epoch 25, FLNet-2 obtained error rates of 1.42%,1.36%, and 1.39% for APCER, BPCER, and HTER, respectively, as can be seen in Figure 8.

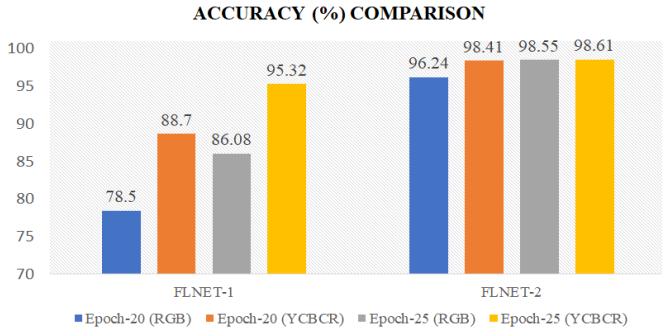


Fig. 6. Accuracy comparison of the proposed models for RGB and YCBCR color spaces on the NUAA dataset.

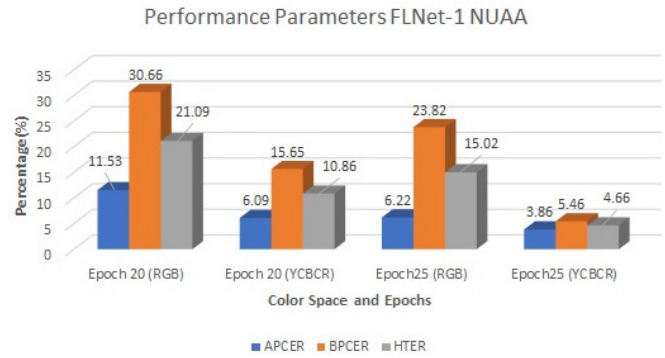


Fig. 7. FLNet-1 results for standard parameters on NUAA dataset.

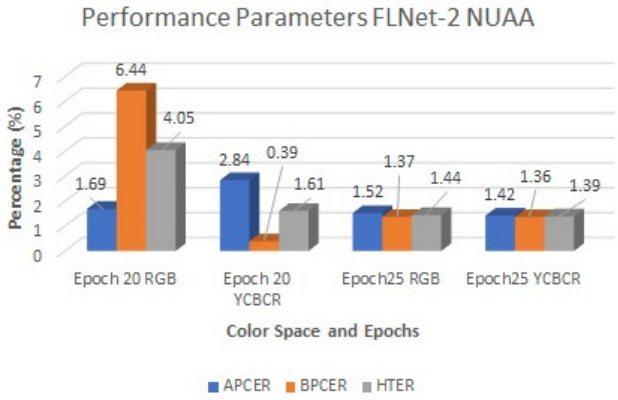


Fig. 8. FLNet-2 results for standard parameters on NUAA dataset.

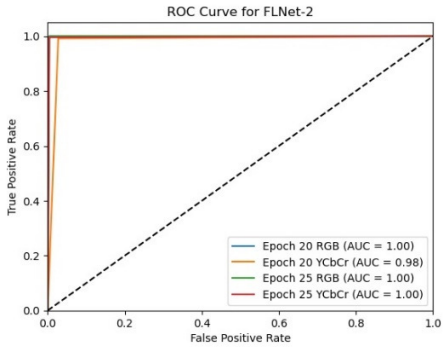
Table II shows cross-dataset results for the FLNet-2 architecture at 25 Epochs, as it was the best-performing model. When trained on the 3D MAD dataset and tested on NUAA, it achieved the best accuracy of 65.68%, while trained on NUAA and tested on 3D MAD, it achieved a satisfactory result of 52.47%. The HTER value is only 23%, indicating a lower error rate and making it more suitable for generalization in attack detection.

Figure 9 shows the ROC curves for FLNet-2 on 3D MAD and NUAA. The ROC curves show an AUC of 1.0, which is the best performance of FLNet-2 in the classification of real and fake face images.

TABLE II. CROSS-DATASET RESULTS FOR FLNET-2

Training dataset	Metrics	Testing dataset
		NUAA
3D MAD	Accuracy	65.68%
	APCER	12.65%
	BPCER	34.68%
	HTER	23.32%
		3D MAD
NUAA	Accuracy	52.47%
	APCER	36.54%
	BPCER	52.48%
	HTER	44.52%

(a)



(b)

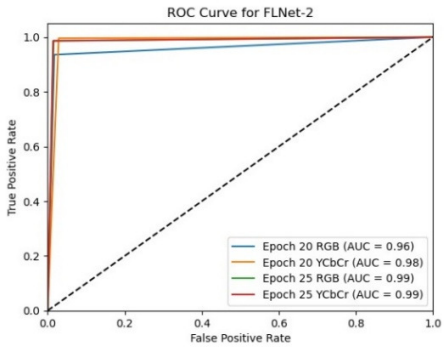


Fig. 9. FLNet-2 ROC curves: (a) 3D MAD, (b)NUAA bottom row

C. *Ablation Study*

An ablation study was performed to evaluate the contribution of each architectural component in the FLNet-1 model. The model achieved an accuracy of 99.35% with the lowest value for HTER of 0.78%. Table III presents the ablation study results. Removing the dropout layer reduced the performance drastically, indicating its contribution to avoiding overfitting. Dense layer and ReLU activation also highlight their contribution to feature extraction and adding non-linearity. These results clearly validate the effectiveness of each component in the architecture and justify their inclusion in the model.

TABLE III. ABLATION STUDY RESULTS FOR FLNET-1

Model Variant	Accuracy (%)	APCER (%)	BPCER (%)	HTER (%)
A: FLNet-1	99.35	1.18	0.39	0.78
B: No Dropout	98.10	2.70	1.60	2.15
C: No Dense layer	97.45	3.20	2.00	2.60
D: No ReLU activation	96.20	4.00	2.90	3.45

To validate the performance improvement, a paired t-test was conducted between the accuracy values of FLNet-1 and FLNet-2. The results indicate a statistically significant difference in accuracy, with FLNet-2 outperforming FLNet-1 ($t = -3.2020, p = 0.0493$), confirming the effectiveness of the additional layers.

D. Error Analysis

As can be seen in Figure 4, FLNet-1 has high error rates on the 3D MAD dataset for the YCbCr color space, with APCER being quite high and indicating that it cannot easily reject spoof attacks. Misclassified spoof samples often featured realistic skin textures and smooth lighting, confusing the model. The NUAA dataset consists of normalized grayscale images, thus limiting spatial cues, and hence, in the RGB color space, the model achieved higher error rates. Higher BPCER indicates that it misclassified the real images due to changes in lightning conditions. Figure 10 shows some sample misclassified images.

Fig. 10. Misclassified samples: NUAA (top row), 3D MAD (bottom row).

E. Comparison of Existing Deep CNN Models with the Proposed

In the literature, both transfer learning and CNN models have been widely employed for face spoofing detection, with their performance typically evaluated using standard metrics, such as Accuracy, HTER, etc. Table IV shows a comparison of the proposed models with existing ones for face spoofing detection. FLNet-2 achieved 99.87% accuracy on the 3D MAD dataset with the lowest HTER of 0.19%, which is the best among the existing methods shown in Table IV. FLNet-2 also obtained the best accuracy of 98.61% and the lowest HTER of 1.39% compared to other models on the NUAA dataset. FLNet-1 achieved the best accuracy of 99.35% on the 3D MAD dataset with the lowest HTER of 0.78%, whereas on the NUAA dataset, it achieved 95.32% accuracy and an HTER of 4.66%. The proposed FLNeT architectures perform well on both 3D and 2D attack datasets, with fewer parameters. Since their parameters are almost 1/10 of pre-trained and 1/5 of other architectures, the training time was drastically reduced, and improved results were achieved for face spoofing detection. The table also presents the number of parameters and FLOPs utilized by the respective architectures for achieving the results for the above metrics. The number of FLOPs for the proposed architectures is in the order of MFLOPs, whereas the architectures in literature require GFLOPs for execution, thus validating the claim of being computationally cost-effective.

TABLE IV. COMPARISON WITH SOTA METHODS

Face anti-spoofing method	# Parameters / FLOPs	Dataset	Accuracy (%)	HTER (%)
VGG16+ LBP [21]	~13.8M / ~16 G	3D MAD	75.25	25.65
VGG16 [22]	~13.8M / ~16 G	NUAA	73.19	28.41
VGG19[23]	~14.4M / ~20 G	NUAA	78.56	18.7
Dual stream rPPG network [24]	~6M / ~5G	3D MAD	97.23	2.32
LeTSrPPG [25]	~7M / ~8 G	3D MAD	95.53	12.35
VGG16[26]	~13 M / ~16 G	3D MAD	99.04	1.20
Proposed FLNET-1	~2.1 M / ~24.4 M	3D MAD, NUAA	**99.35** **95.32**	**0.78** **4.66**
Proposed FLNET-2	~1.1 M / ~84 M	3D MAD. NUAA	**99.87** **98.61**	**0.19** **1.39**

V. LIMITATIONS

This paper proposes simple deep CNN architectures for face spoofing detection; however, the evaluation is limited to only two datasets, which cover a narrow range of attack scenarios and vectors. For developing a more generalized and robust spoofing detection system, future research should consider evaluation on multiple diverse datasets, encompassing a wider variety of spoofing attacks, with cross-dataset evaluation. In addition, vision transformers have gained significant attention in recent years, often outperforming CNNs in various vision tasks. Applying vision transformer-based architectures for face spoofing detection can further reduce errors.

VI. CONCLUSION

Face spoofing is associated with presentation attacks and can be performed using a printed photo, a photo display on a screen, a video replay, or a 3D mask. Pre-trained models have been previously used for the detection of spoofing for 2D and 3D attacks. This study proposed two models for the detection of face presentation attacks in both 2D and 3D scenarios. The evaluation was presented for both RGB and YCbCr based on standard performance parameters. FLNet-2 achieved an excellent accuracy of 99.87% with 0.19% HTER on 3D attack samples, and FLNet-1 achieved an accuracy of 99.35% with HTER of 0.78%. FLNet-1 and FLNet-2 were compared with existing models regarding accuracy and HTER. The proposed architectures surpass existing SOTA methods and demonstrate strong generalizability in detecting presentation attacks. The lower numbers of parameters and FLOPs for execution with improved results demonstrate the superiority of the proposed architectures in the detection of 2D and 3D attacks. Researchers can explore other publicly available PAD datasets to test the proposed architectures and perform modifications. Modifications to existing pre-trained models utilizing their strength of feature extraction in combination with the addition of new representation layers can enhance detection rates and improve generalizability in the detection of spoofing attacks.

REFERENCES

[1] S. Bhattacharjee and S. Marcel, "What You Can't See Can Help You - Extended-Range Imaging for 3D-Mask Presentation Attack Detection," in *2017 International Conference of the Biometrics Special Interest*

Group (BIOSIG), Darmstadt, Germany, Sep. 2017, pp. 1–7, https://doi.org/10.23919/BIOSIG.2017.8053524.

[2] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face Spoofing Detection Using Colour Texture Analysis," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1818–1830, Dec. 2016, https://doi.org/10.1109/TIFS.2016.2555286.

[3] A. B. S. Salamh and H. I. Akyüz, "A Novel Feature Extraction Descriptor for Face Recognition," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 8033–8038, Feb. 2022, https://doi.org/10.48084/etasr.4624.

[4] N. Erdogmus and S. Marcel, "Spoofing Face Recognition With 3D Masks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 7, pp. 1084–1097, Jul. 2014, https://doi.org/10.1109/TIFS.2014.2322255.

[5] T. Edmunds and A. Caplier, "Motion-based countermeasure against photo and video spoofing attacks in face recognition," *Journal of Visual Communication and Image Representation*, vol. 50, pp. 314–332, Jan. 2018, https://doi.org/10.1016/j.jvcir.2017.12.004.

[6] A. Mahore and M. Tripathi, "Detection of 3D Mask in 2D Face Recognition System Using DWT and LBP," in *2018 IEEE 3rd International Conference on Communication and Information Systems (ICCIS)*, Singapore, Dec. 2018, pp. 18–22, https://doi.org/10.1109/ICOMIS.2018.8644807.

[7] S. Hashemifard and M. Akbari, "A Compact Deep Learning Model for Face Spoofing Detection." arXiv, Jan. 12, 2021, https://doi.org/10.48550/arXiv.2101.04756.

[8] K. Balamurali, S. Chandru, M. S. Razvi, and V. Sathiesh Kumar, "Face Spoof Detection Using VGG-Face Architecture," *Journal of Physics: Conference Series*, vol. 1917, no. 1, Mar. 2021, Art. no. 012010, https://doi.org/10.1088/1742-6596/1917/1/012010.

[9] V. P. Vishwakarma, R. Gupta, and A. K. Yadav, "A Novel Non-Iterative Deep Convolutional Neural Network with Kernelized Classification for Robust Face Recognition," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16460–16465, Oct. 2024, https://doi.org/10.48084/etasr.8229.

[10] M. Pei, B. Yan, H. Hao, and M. Zhao, "Person-Specific Face Spoofing Detection Based on a Siamese Network," *Pattern Recognition*, vol. 135, Mar. 2023, Art. no. 109148, https://doi.org/10.1016/j.patcog.2022.109148.

[11] S. M. Ibrahim, M. S. Ibrahim, S. Khan, Y. W. Ko, and J. G. Lee, "Improving Face Presentation Attack Detection Through Deformable Convolution and Transfer Learning," *IEEE Access*, vol. 13, pp. 31228–31238, 2025, https://doi.org/10.1109/ACCESS.2025.3541546.

[12] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel, "Biometric Face Presentation Attack Detection With Multi-Channel Convolutional Neural Network," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 42–55, 2020, https://doi.org/10.1109/TIFS.2019.2916652.

[13] H. Chen, G. Hu, Z. Lei, Y. Chen, N. M. Robertson, and S. Z. Li, "Attention-Based Two-Stream Convolutional Networks for Face Spoofing Detection," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 578–593, 2020, https://doi.org/10.1109/TIFS.2019.2922241.

[14] F. M. Chen, C. Wen, K. Xie, F. Q. Wen, G. Q. Sheng, and X. G. Tang, "Face liveness detection: fusing colour texture feature and deep feature," *IET Biometrics*, vol. 8, no. 6, pp. 369–377, 2019, https://doi.org/10.1049/iet-bmt.2018.5235.

[15] N. Erdogmus and S. Marcel, "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)*, Arlington, VA, USA, Sep. 2013, pp. 1–6, https://doi.org/10.1109/BTAS.2013.6712688.

[16] N. Erdoğmuş and S. Marcel, "3D Mask Attack Dataset (3DMAD)." Zenodo, Apr. 23, 2013, [Online]. Available: https://zenodo.org/records/4068477.

[17] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model," in *Computer Vision – ECCV 2010*, 2010, pp. 504–517, https://doi.org/10.1007/978-3-642-15567-3_37.

[18] "NUAA Imposter Database." [Online]. Available: https://parnec.nuaa.edu.cn/_upload/tpl/02/db/731/template731/pages/xtan/NUAAImposterDB_download.html.

[19] P. Jaswanth, P. Y. Chowdary, and M. V. S. Ramprasad, "Deep learning based intelligent system for robust face spoofing detection using texture feature measurement," *Measurement: Sensors*, vol. 29, Oct. 2023, Art. no. 100868, https://doi.org/10.1016/j.measen.2023.100868.

[20] D. Menotti *et al.*, "Deep Representations for Iris, Face, and Fingerprint Spoofing Detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 864–879, Apr. 2015, https://doi.org/10.1109/TIFS.2015.2398817.

[21] P. K. Das, B. Hu, C. Liu, K. Cui, P. Ranjan, and G. Xiong, "A New Approach for Face Anti-Spoofing Using Handcrafted and Deep Network Features," in *2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, Zhengzhou, China, Nov. 2019, pp. 33–38, https://doi.org/10.1109/SOLI48380.2019.8955089.

[22] F. Abdullakutty, E. Elyan, P. Johnston, and A. Ali-Gombe, "Deep Transfer Learning on the Aggregated Dataset for Face Presentation Attack Detection," *Cognitive Computation*, vol. 14, no. 6, pp. 2223–2233, Nov. 2022, https://doi.org/10.1007/s12559-022-10037-z.

[23] S. D. Thepade, M. Dindorkar, P. Chaudhari, and S. Bang, "Face presentation attack identification optimization with adjusting convolution blocks in VGG networks," *Intelligent Systems with Applications*, vol. 16, Nov. 2022, Art. no. 200107, https://doi.org/10.1016/j.iswa.2022.200107.

[24] R. Sun, X. Yu, H. Feng, F. Wang, and X. Zhang, "Motion-robust mask face presentation attack detection via dual-stream texture-rPPG network," *The Visual Computer*, vol. 41, no. 7, pp. 4517–4532, May 2025, https://doi.org/10.1007/s00371-024-03675-x.

[25] S. Q. Liu, X. Lan, and P. C. Yuen, "Learning Temporal Similarity of Remote Photoplethysmography for Fast 3D Mask Face Presentation Attack Detection," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 3195–3210, 2022, https://doi.org/10.1109/TIFS.2022.3197335.

[26] O. Lucena, A. Junior, V. Moia, R. Souza, E. Valle, and R. Lotufo, "Transfer Learning Using Convolutional Neural Networks for Face Anti-spoofing," in *Image Analysis and Recognition*, vol. 10317, F. Karray, A. Campilho, and F. Cheriet, Eds. Springer International Publishing, 2017, pp. 27–34.