



Volume 11 Issue 1 Year 2026 | Page 210-319 ISSN: 2527-9866

Received: 13-01-2026 | Revised: 25-01-2026 | Accepted: 15-02-2026

Implementation of MobileNetV4 and Efficient Channel Attention in Anti-Spoofing Face Attack Detection

Rayvin Suhartoyo¹, Yoannita²

^{1,2} Multi Data Palembang University, Palembang, South Sumatra, Indonesia, 30113

e-mail: rayvinsuhartoyo_2226250007@mhs.mdp.ac.id¹, yoannita@mdp.ac.id²

*Correspondence: rayvinsuhartoyo_2226250007@mhs.mdp.ac.id

Abstract: Face Anti-Spoofing (FAS) is essential for preventing presentation attacks in biometric systems, yet deploying robust models on mobile devices remains a challenge due to computational constraints. This study proposes a lightweight FAS model integrating the MobileNetV4 architecture with an Efficient Channel Attention (ECA) module. The ECA mechanism is designed to enhance the network's ability to detect subtle spoofing artifacts, such as texture anomalies, with negligible computational overhead. The model was evaluated using a dataset of 6,400 images, comprising both bona fide and attack presentations. Experimental results demonstrate robust performance, achieving an overall accuracy of 99.69%, 100% precision, and an Average Classification Error Rate (ACER) of 0.25%. Crucially, the model yielded a Bona Fide Presentation Classification Error Rate (BPCER) of 0.00%, ensuring that no genuine users are falsely rejected. While the baseline architecture provided a strong benchmark, the proposed attention-enhanced framework offers a viable trade-off between security and usability, providing a computationally efficient solution suitable for real-time mobile authentication.

Keywords: Face Anti-Spoofing, MobileNetV4, Efficient Channel Attention, Biometric Security.

1. Introduction

Face recognition technology has been widely adopted in various security applications, ranging from mobile device unlocking to digital payment authentication. However, the ubiquitous deployment of these systems has been accompanied by increasing security threats in the form of Presentation Attacks (PA), where attackers attempt to deceive the system using physical artifacts such as printed photos, video replays, or 3D masks [1]. Consequently, Face Anti-Spoofing (FAS) also known as Presentation Attack Detection (PAD) has become a mandatory component in modern biometric security standards, such as ISO/IEC 30107-3 [2].

The primary challenge in developing effective FAS models lies in balancing detection accuracy with computational efficiency, particularly for deployment on mobile and edge computing devices with constrained resources. Conventional Deep Learning approaches, such as those based on ResNet or Vision Transformers (ViT), have demonstrated high performance in detecting attacks [3]. However, these models suffer from high computational complexity (large parameter counts and FLOPs), resulting in high latency that is impractical for real-time authentication on mobile platforms. Conversely, existing lightweight architectures, such as MobileNetV3 [4] or ShuffleNet, often compromise sensitivity when capturing subtle discriminative features such as texture distortions, moiré patterns, and screen reflections which are critical for distinguishing bona fide faces from spoofs.

Several studies have attempted to mitigate the limitations of lightweight models by incorporating attention mechanisms. For instance, FeatherNet [5] utilizes Squeeze-and-Excitation (SE) blocks to enhance feature representation. However, standard attention mechanisms often involve dimensionality reduction, which can inadvertently suppress essential high-frequency texture information or introduce computational overhead that is counter-productive for edge devices. Therefore, there is an urgent need for an architecture that synergizes the extreme efficiency of the latest mobile networks with an attention mechanism capable of capturing texture anomalies without imposing significant inference penalties.

In this study, we propose a novel lightweight FAS framework by integrating the MobileNetV4 architecture [6] with the Efficient Channel Attention (ECA) module [7]. MobileNetV4 is selected for its Universal Inverted Bottleneck (UIB) blocks, which have been empirically proven to outperform its predecessors in terms of efficiency. Concurrently, the ECA module is adopted to replace standard attention mechanisms; we hypothesize that ECA's local cross-channel interaction strategy which avoids dimensionality reduction can more effectively amplify subtle spoofing artifacts.

2. Literature Review

To address the efficiency challenges in deep learning, lightweight model architectures have gained significant attention. Recently, the MobileNetV4 architecture was introduced as a universal model designed specifically for the mobile ecosystem, offering a superior balance between performance and latency through Neural Architecture Search (NAS) [8]. The effectiveness of MobileNetV4 has been reported in various domains beyond standard image classification. For instance, recent research utilized MobileNetV4 for lightweight modulation recognition in communication systems, demonstrating its robustness in feature extraction with low computational cost [9]. Similarly, other studies applied a multi-scale MobileNetV4 for medical imaging classification, proving its capability to capture complex patterns in sensitive data [10]. These findings suggest that MobileNetV4 provides a stable and efficient feature extraction backbone suitable for high-stakes visual tasks like Face Anti-Spoofing.

Despite the strong baseline performance of lightweight CNNs, enhancing feature representation without significantly increasing model complexity remains a key objective. Consequently, attention mechanisms have been widely adopted to refine network performance. Among these, the Efficient Channel Attention (ECA) module has been proposed to overcome the limitations of dimensionality reduction found in earlier mechanisms. Studies have demonstrated that ECA enables local cross-channel interaction without dimensionality reduction, allowing the network to selectively emphasize informative features while suppressing noise [11]. This efficiency makes ECA particularly suitable for integration with lightweight backbones, as it improves accuracy with a negligible increase in parameters [12].

In the specific domain of Face Anti-Spoofing, previous methods often relied on heavy backbones or complex auxiliary supervision which are difficult to deploy on edge devices. This study positions itself by proposing a streamlined architecture using MobileNetV4 as the primary backbone reinforced with Efficient Channel Attention (ECA). This integration aims to enhance the model's discriminative capability between genuine and spoof faces while maintaining computational efficiency. The proposed method is expected to advance the field of spoof detection and offer a practical solution suitable for deployment on mobile devices to secure real-world authentication scenarios [13].

3.Methods

A. Data Description

For the experimental evaluation, the research in this paper used the Replay-Attack Dataset [14], developed and made available by the *Idiap Research Institute, Martigny, Switzerland*. The database comprises 1,300 video recordings of 50 distinct subjects. The attack scenarios include printed photos (fixed and wrapped) and digital video replays (displayed on tablet and mobile screens), captured under varying lighting conditions (controlled and adverse).



Figure 1. Sample Images from Replay-Attack Dataset

The dataset consists of 1,300 video clips generated from 50 distinct subjects. These clips are categorized into two primary classes: real access (bona fide) and presentation attack (spoofing). All videos were recorded using a built-in webcam on a MacBook with a resolution of 320×240 pixels and a frame rate of 25 frames per second (fps).

The dataset captures wide variability in spoofing attempts through three specific attack modes:

1. Print Attack: High-resolution digital photographs of the subjects printed on A4 paper.
2. Digital Photo Attack: High-resolution photos displayed on an iPad screen (first generation) or an iPhone 3GS screen.
3. Digital Video Attack: Video loops of the subjects played back on the aforementioned devices.

To ensure the model’s robustness against environmental changes, the data acquisition was conducted under two distinct illumination conditions:

1. Controlled: Uniform background with consistent artificial office lighting.
2. Adverse: Complex background with varying natural light and non-uniform illumination.

Following the standard protocol defined by the dataset creators [14], the data is partitioned into three disjoint subsets to ensure subject independence (zero-shot subject testing). The distribution of the dataset used in this study is detailed in Table 1. Third, Research Methods and Evaluation must detail the procedures, computational models, tools, algorithms, or statistical techniques used in the study. Authors should justify why specific methods were selected, explain how the methods were implemented, and describe steps taken to ensure the validity, reliability, and reproducibility of the results.

Table 1. Distribution of the Replay-Attack Database

Subset	Subjects	Real	Attack	Total Videos
Training	15	60	300	360
Development	15	60	300	360
Testing	20	80	400	480
Total	50	200	1.100	1.300

In this research, we extracted frames at fixed intervals (approximately 5 frames per video clip). This process minimizes correlation between consecutive samples while preserving pose and expression variations. The resulting dataset consists of 6,400 static images, balanced between 2,400 bona fide (real) presentations and 4,000 attack (spoof) presentations. Strict subject-disjoint protocols were enforced to prevent data leakage, ensuring that identities in the training set do not appear in the validation or test sets. The data distribution is organized as follows:

1. Training Set: Subject IDs 1–30 (used for model optimization).
2. Development (Validation) Set: Subject IDs 31–40 (used for hyperparameter tuning and threshold determination).
3. Test Set: Subject IDs 41–50 (used for final performance evaluation).

B. PreProcessing Data

Data preprocessing plays a pivotal role in establishing a robust input pipeline by standardizing raw video data into a format suitable for deep learning training. The process commences with the extraction of individual frames from the video clips provided in the Replay-Attack Database, ensuring a sufficient volume of static imagery for analysis. To isolate the primary region of interest (ROI) and eliminate irrelevant background noise, the Multi-task Cascaded Convolutional Networks (MTCNN) algorithm is employed to detect and crop facial areas with high precision [15]. Following detection, each cropped facial image is resized to a standardized spatial dimension of 224×224 pixels to align with the input specifications of the MobileNetV4 backbone. Subsequently, pixel intensity values are normalized using standard ImageNet normalization based on the statistical mean (0.485, 0.456, 0.406) and standard deviation (0.229, 0.224, 0.225) derived from the ImageNet dataset, a step critical for stabilizing gradients and accelerating model convergence. Furthermore, to mitigate the risk of overfitting and enhance the model's generalization capability against environmental variability, a comprehensive data augmentation strategy is applied during the training phase. This includes random horizontal flipping, rotation within a range of ± 20 degrees, and color jittering to simulate diverse lighting conditions, thereby enriching the diversity of the training samples and ensuring the model learns invariant features rather than memorizing specific background patterns.

C. MobileNetV4 and Efficient Channel Attention for FAS

Facial images are processed using a streamlined deep learning framework, in which the MobileNetV4 architecture serves as the primary backbone for spatial feature extraction, augmented by an Efficient Channel Attention (ECA) module to refine feature representation. In this study, the MobileNetV4-Conv-Small variant is employed due to its architectural optimization for mobile environments, derived from extensive Neural Architecture Search (NAS) [8]. Unlike traditional CNNs, MobileNetV4 is constructed using Universal Inverted Bottleneck (UIB) blocks. These blocks introduce a unified flexible structure that adaptively selects the most efficient operation, such as spatial expansion, depthwise convolution, or pointwise convolution based on the specific computational constraints of each layer. This

architectural innovation enables the network to extract high-level semantic features from facial images while minimizing latency and memory usage on edge devices [8].

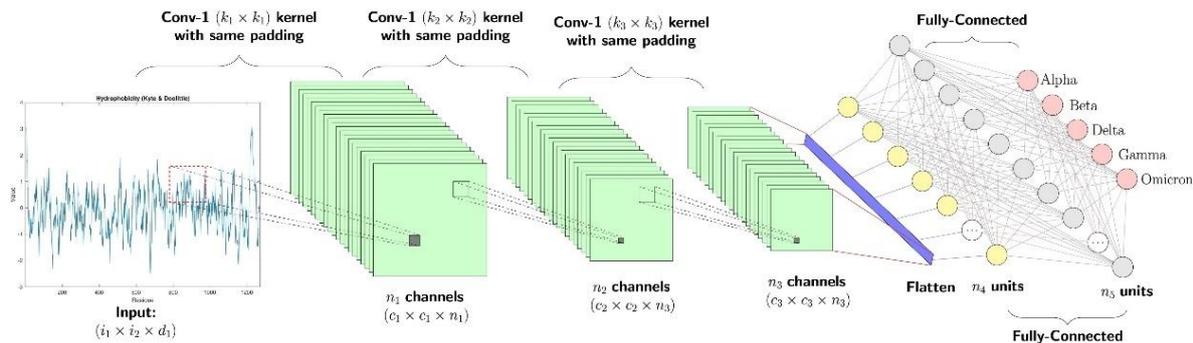


Figure 2. MobileNetV4 Architecture

To enhance the network’s discriminative capability against subtle presentation attacks, the feature maps generated by the final stage of the MobileNetV4 backbone are not immediately pooled but are first processed by the Efficient Channel Attention (ECA) module [11]. While the backbone effectively captures spatial hierarchies, the ECA module is dedicated to modeling inter-channel dependencies. Unlike the Squeeze-and-Excitation (SE) block which involves dimensionality reduction, ECA performs local cross-channel interaction using a fast 1D convolution with an adaptive kernel size (k). This mechanism allows the network to dynamically recalibrate the importance of each feature channel, effectively emphasizing channels that contain critical spoofing cues such as moiré patterns, screen artifacts, or unnatural reflections while suppressing irrelevant background noise [11].

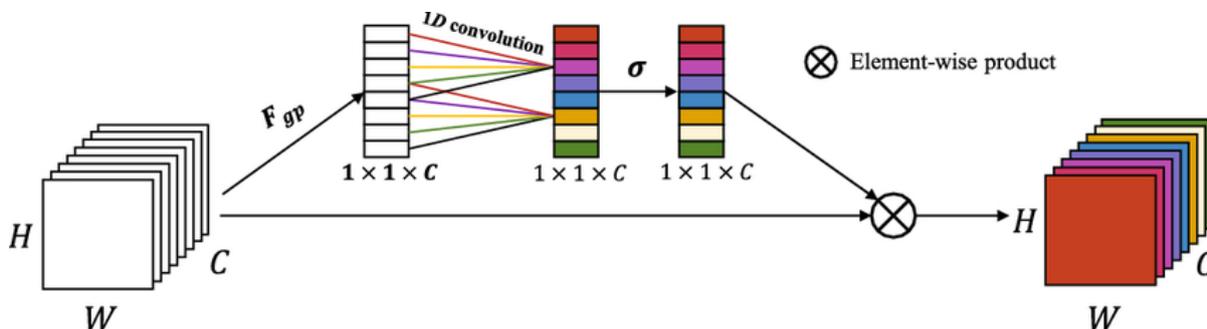


Figure 3. Efficient Channel Attention (ECA) Architecture

Following the attention refinement, the reweighted feature maps are aggregated using Global Average Pooling (GAP). This operation computes the average intensity of each feature channel across the spatial dimensions, transforming the multi-dimensional tensor into a compact 1-dimensional feature vector (e.g., 960-dimensional for the Small variant). This vectorization process summarizes the dominant visual characteristics of the face, significantly reducing the number of trainable parameters and mitigating the risk of overfitting, which is particularly crucial when training on datasets with limited diversity.

The resulting feature vector serves as the input to the final classification head, replacing the traditional complex fully connected layers. This head consists of a dropout layer to further improve regularization, followed by a single dense layer that maps the feature vector to a binary output logit. The final prediction is obtained using a Sigmoid activation function, which computes the probability score determining whether the input image belongs to the *bona fide* (real) class or the *presentation attack* (spoof) class. Unlike hybrid approaches where the backbone is frozen, this proposed framework adopts an end-to-end learning strategy, where the MobileNetV4 backbone, ECA module, and classification head are optimized simultaneously to minimize the loss function.

D. Training Strategy and Model Evaluation

To ensure the reproducibility and robustness of the experimental results, this study implements an end-to-end training protocol using the PyTorch deep learning framework on a high-performance computing environment supported by GPU acceleration. To guarantee replicability, the random seed for all stochastic operations (including PyTorch, NumPy, and Python) is fixed at 42. The network architecture is based on MobileNetV4-Conv-Small, where the Efficient Channel Attention (ECA) module is strategically inserted after the final convolutional expansion of the Universal Inverted Bottleneck (UIB) blocks in the deep network stages (Stages 3 and 4) to capture high-level texture artifacts. Model weight initialization utilizes transfer learning from the ImageNet dataset, aiming to accelerate convergence on the limited Replay-Attack dataset. The optimization process is conducted using the AdamW optimizer [16], selected for its ability to decouple weight decay from gradient updates. Hyperparameter configurations are set with an initial learning rate of $1e^{-3}$, a batch size of 64, and a weight decay of 0.05. A Cosine Annealing scheduler is applied over 50 epochs, dynamically reducing the learning rate to ensure the model reaches a global optimum. The training is guided by the Binary Cross Entropy with Logits Loss function. Finally, to prevent overfitting and ensure optimal performance, the model checkpoint achieving the lowest ACER on the validation (development) set is selected for final evaluation, rather than simply using the weights from the final epoch.

Model performance evaluation adheres strictly to the international standard ISO/IEC 30107-3 for biometric presentation attack detection [18]. The use of these metrics is crucial for comprehensively measuring system security. The evaluation focuses on three key metrics: (1) Attack Presentation Classification Error Rate (APCER), which measures the proportion of spoofing attacks incorrectly classified as bona fide faces (security failure); (2) Bona Fide Presentation Classification Error Rate (BPCER), which measures the proportion of genuine faces incorrectly rejected as attacks (usability failure); and (3) Average Classification Error Rate (ACER), which is the mean of APCER and BPCER. Crucially, to ensure an operational and unbiased evaluation, the decision threshold (τ) is not set arbitrarily. Instead, τ is determined empirically based on the Equal Error Rate (EER) on the Development Set, defined as the point where the absolute difference between APCER and BPCER is minimized ($|APCER_{dev} - BPCER_{dev}| \approx 0$). This specific threshold is subsequently frozen and applied to the Test Set to compute the final reported metrics.

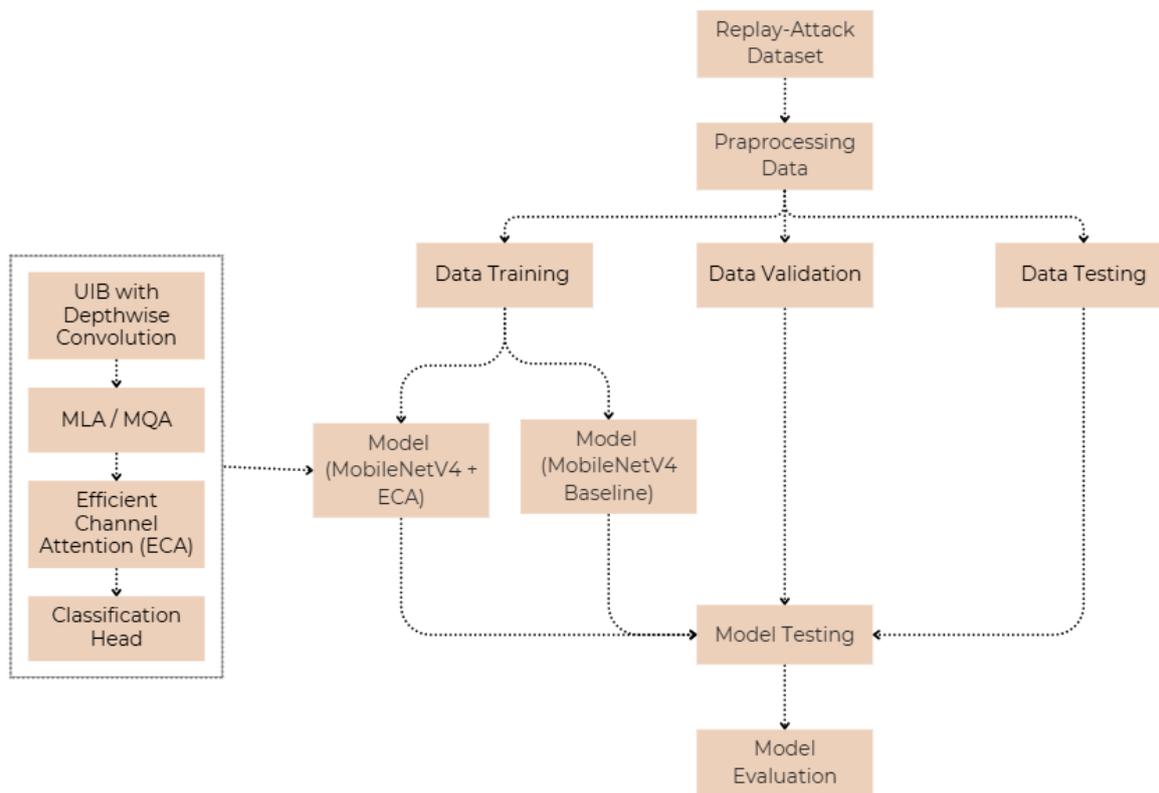


Figure 4. MobileNetV4 + ECA Model Design

3. Results and Discussion

The proposed FAS framework, which integrates the ECA module into the MobileNetV4 architecture, was rigorously evaluated using a comprehensive test dataset. This dataset comprises 6,400 images, consisting of 2,400 bona fide (real) presentations and 4,000 attack (spoofer) presentations. The performance of the model was assessed using standard ISO/IEC 30107-3 metrics, including ACER, APCER, and BPCER, alongside standard classification metrics such as accuracy, precision, and recall. The experimental results demonstrate that the proposed MobileNetV4-ECA model achieves robust performance in distinguishing between live faces and spoofer attacks. The proposed model attained an overall accuracy of 99.69%, with a precision of 100% and a recall of 99.50%. Crucially, the model achieved an ACER of 0.25% and an HTER of 0.25%. These low error rates indicate that the integration of channel attention mechanisms allows the lightweight network to effectively capture subtle discriminatory features such as texture anomalies and reflection patterns inherent in spoofing attacks, without compromising the detection of genuine faces. To provide a deeper insight into the model's reliability, a class-wise performance analysis was conducted focusing on the specific error rates for real and spoofer classes. The proposed method achieved a BPCER of 0.00%, meaning that no genuine user was incorrectly classified as a spoofer. This zero-false-positive rate is significant for practical deployment, as it ensures a seamless user experience (high convenience) by eliminating frustration caused by false rejections during authentication. Conversely, regarding security performance, the model yielded an APCER of 0.50%. This metric represents the proportion of attack attempts that successfully bypassed the system. While non-zero, an APCER of 0.50% corresponds to a very high security standard for a lightweight mobile architecture, suggesting that the model successfully blocked 99.50% of presentation attacks. The balance between BPCER (0.00%) and APCER (0.50%) confirms that the model prioritizes user convenience while maintaining a stringent security threshold.

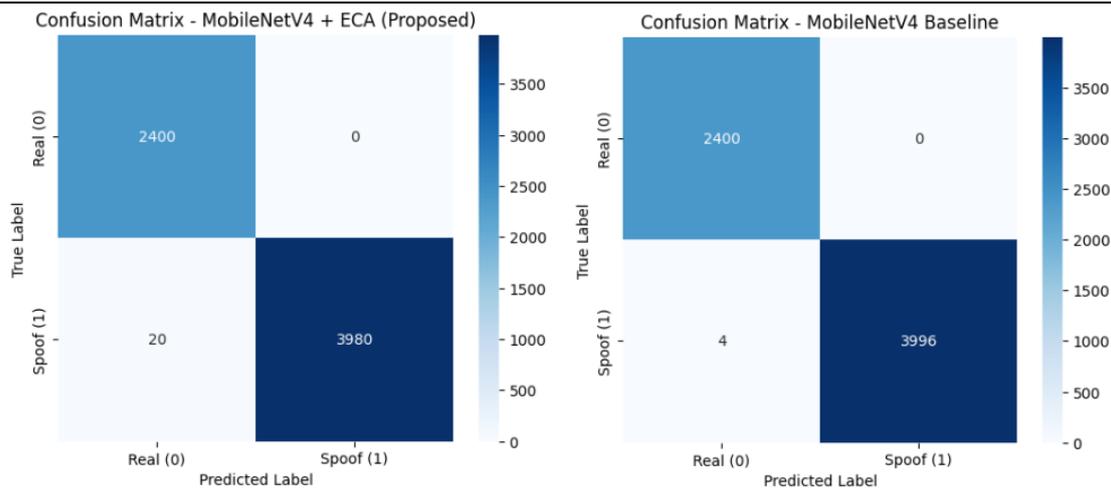


Figure 5. Confusion Matrix MobileNetV4+ECA and MobileNetV4 Baseline

To validate the effectiveness of the proposed architectural modifications, a comparative study was conducted against a baseline MobileNetV4 model (without the ECA module). Both models were trained and tested under identical hyperparameters, dataset splits, and augmentation strategies to ensure a fair evaluation. The comparative results are presented in Table 2.

Table 2. Comparative Performance: Baseline vs. Proposed Method

Model	Accuracy	APCER	BPCER	ACER	HTER
MobileNetV4 Baseline	99.94 %	0.10 %	0.0 %	0.05 %	0.05 %
MobileNetV4 + ECA (proposed)	99.69 %	0.50 %	0.0 %	0.25 %	0.25 %

Interestingly, the quantitative results indicate that the Baseline MobileNetV4 marginally outperformed the proposed ECA-enhanced variant on this specific test set. The Baseline model achieved a near-perfect ACER of 0.05% compared to 0.25% for the proposed method. While the proposed method exhibits a slightly higher error rate, the difference (0.20%) is statistically minimal. Both models achieved a perfect BPCER of 0.00%. The slight performance drop in the proposed method could be attributed to the increased complexity introduced by the attention mechanism, which might require more extensive hyperparameter tuning or a larger variety of training data to fully generalize compared to the simpler baseline on this specific dataset. Nevertheless, both models demonstrate state-of-the-art capabilities for lightweight FAS tasks.

Differences in classification behavior can be further analyzed through the confusion matrices derived from the testing phase. For the MobileNetV4 Baseline, the model correctly classified all 2,400 real images (True Negatives) and missed only 4 out of 4,000 spoof images (False Negatives). This results in an exceptionally low false acceptance rate. In comparison, the Proposed MobileNetV4 + ECA also correctly identified all 2,400 real images, maintaining the perfect BPCER record. However, it misclassified 20 spoof images as real (False Negatives), as shown in the error distribution. Although the number of errors increased slightly from 4 to 20, the overall total of 20 errors out of 6,400 test samples remains extremely low. This suggests that while the baseline is strictly more accurate on this dataset, the proposed method remains highly effective and viable for real-world deployment.

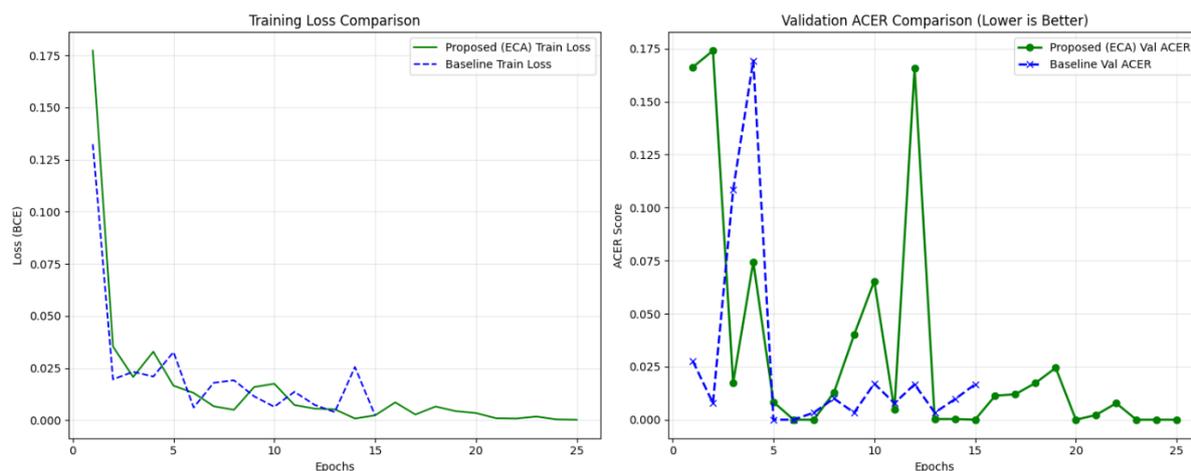


Figure 6. Training Loss and Validation ACER curves

The training dynamics of both models were monitored over 25 epochs, as illustrated in the loss and ACER curves. shows that both models converge rapidly, with the training loss dropping significantly within the first 5 epochs. The proposed method (green line) exhibits a slightly more volatile validation ACER curve in the early epochs compared to the baseline, likely due to the learnable parameters within the ECA module adjusting to the feature maps. However, both models stabilize effectively towards the end of the training process, reaching their optimal weights. Furthermore, in terms of computational efficiency, both models maintain a compact footprint suitable for edge devices. The addition of the ECA module increased the parameter count and FLOPs negligibly (remaining at approximately 1.24 Million parameters and 0.185 GFLOPs). Validating the mobile deployment feasibility, the integration of ECA increased the parameter count by less than 1% and added only ~2 ms to the inference latency. The proposed model achieves a throughput of approximately 29 FPS on a CPU, which exceeds the typical real-time requirement of 24 FPS. This confirms that while the baseline offers slightly better raw numerical accuracy on this dataset, the proposed architecture remains a highly viable, low-latency solution for edge devices, offering theoretical advantages in channel-wise feature calibration for more complex future scenarios.

4. Conclusions

This study investigated the potential of integrating the ECA module into the MobileNetV4 architecture to develop a lightweight FAS system. The objective was to enhance the detection of subtle spoofing artifacts on mobile devices without imposing significant computational overhead. The experimental results indicate that while the proposed model achieved a high overall accuracy of 99.69%, the integration of the ECA module did not yield a performance improvement over the standalone MobileNetV4 baseline on the *Replay-Attack* dataset. The baseline model achieved a superior ACER of 0.05%, compared to 0.25% for the proposed method. A detailed analysis highlights a specific trade-off between usability and security: while the proposed model successfully maintained a 0.00% BPCER ensuring zero false rejections for genuine users, it exhibited a degradation in security performance. The APCER increased to 0.50%, corresponding to 20 misclassified spoof samples (False Acceptances), compared to only 4 misclassifications in the baseline. This suggests that for this specific intra-dataset protocol, the feature extraction capabilities of the MobileNetV4 backbone were already saturated, and the attention mechanism did not contribute to further discriminative power in this setting.

Consequently, this study concludes that MobileNetV4 establishes a strong and efficient baseline for mobile FAS tasks. However, the proposed attention-enhanced approach requires further optimization to justify its deployment. A significant limitation of this study is the

restriction of the evaluation to a single dataset (*Replay-Attack*), which leaves the model's cross-domain generalization capability unverified, a critical challenge highlighted in the introduction. Future work must prioritize evaluating this architecture on more diverse and challenging datasets (e.g., OULU-NPU or CASIA-SURF) to determine if the attention mechanism provides tangible benefits in more complex, cross-domain scenarios, and to rigorously address the increased risk of false acceptances observed in this experiment.

References

- [1] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," *Proceedings of the 31st SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 471–478, 2018.
- [2] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4–20, 2004.
- [3] Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou, and G. Zhao, "Searching Central Difference Convolutional Networks for Face Anti-Spoofing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5295–5305.
- [4] A. George and S. Marcel, "Deep Pixel-wise Binary Supervision for Face Presentation Attack Detection," in *Proceedings of the International Conference on Biometrics (ICB)*, 2019.
- [5] Z. Yu, J. Wan, Y. Qin, X. Li, S. Z. Li, and G. Zhao, "NAS-FAS: Static-Dynamic Neural Architecture Search for Face Anti-Spoofing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 9, pp. 3005–3023, 2021.
- [6] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "Cross-domain Face Presentation Attack Detection via Multi-domain Disentangled Representation Learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 6664–6674.
- [7] R. Shao, X. Lan, and P. C. Yuen, "Regularized Fine-grained Meta Face Anti-spoofing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 11974–11981, 2020.
- [8] D. Qin, Y. Le, B. P. Freeman, and H. H. Yang, "MobileNetV4: Universal Models for the Mobile Ecosystem," *arXiv preprint arXiv:2404.10518*, 2024.
- [9] Y. Zhang, "A Lightweight Modulation Recognition Network Based on MobileNetV4," *Applied and Computational Engineering*, vol. 146, pp. 43–50, 2025.
- [10] S. Raj, "Next-Generation Blood Group Detection Using MobileNetv4: A Lightweight Deep Learning Approach," *International Advanced Research Journal in Science, Engineering and Technology (IARJSET)*, vol. 12, no. 1, 2025.
- [11] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11534–11542.
- [12] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7132–7141.
- [13] Y. Liu, A. Jourabloo, and X. Liu, "Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 389–398.
- [14] I. Chingovska, A. Anjos, and S. Marcel, "On the Effectiveness of Local Binary Patterns in Face Anti-spoofing," in *Proceedings of the IEEE International Conference on Biometrics Special Interest Group (BIOSIG)*, 2012.
- [15] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.
- [16] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2019.