Exploring the Effectiveness of Lightweight Architectures for Face Anti-Spoofing

Yoanna Martínez-Díaz, Heydi Méndez-Vázquez
Advanced Technologies Application Center (CENATAV)
7A ♯21406 Siboney, Playa, P.C.12200, Havana, Cuba
{ymartinez,hmendez}@cenatav.co.cu

Luis S. Luevano, Miguel Gonzalez-Mendoza
Tecnológico de Monterrey,
Campus Estado de México, Estado de México, México
{luis.s.luevano, mgonza}@tec.mx

Abstract

Detecting spoof faces is crucial in ensuring the robustness of face-based identity recognition and access control systems, as faces can be captured easily without the user's cooperation in uncontrolled environments. Several deep models have been proposed for this task, achieving high levels of accuracy but at a high computational cost. Considering the very good results obtained by lightweight deep networks on different computer vision tasks, in this work we explore the effectiveness of this kind of architectures for face anti-spoofing. Specifically, we asses the performance of three lightweight face models on two challenging benchmark databases. The conducted experiments indicate that face anti-spoofing solutions based on lightweight face models are able to achieve comparable accuracy results to those obtained by state-of-the-art very deep models, with a significantly lower computational complexity.

1. Introduction

Face recognition is one of the most used biometric techniques, with many practical and commercial applications including access control, forensics, and human-computer interactions [2]. Some systems require an interactive capture of the user's face, for example in check-in and mobile payment applications. However, if an impostor steals or replicates the registered user's face image, the system's security may be compromised. For those reasons, face anti-spoofing (FAS) has been in the focus of both academy and industry [5], aiming at detecting fake faces created using various spoofing attack methods, such as 3D masks, printed photographs, or video.

Several FAS methods have been proposed in the literature. The traditional ones used handcrafted features, such as texture, color, or shape. These methods are prone to fail in various scenarios, such as when the spoof video is of high quality, or the attacker creates a mask with realistic texture and shape. Recently, deep learning-based methods have emerged as a robust solution to these issues. However, most of the Convolutional Neural Networks (CNNs) designed for face anti-spoofing demand significant computational resources. A well-designed FAS network architecture is vital for real-time applications, especially for those intended for use in mobile and wearable devices [33].

Some studies evaluate the performance of different CNNs for face anti-spoofing [34, 53], but to the best of our knowledge, none of them have included lightweight models on their analysis, nor the computational complexity of the evaluated models is considered. Moreover, different competitions and databases have been released to boost research progress on face anti-spoofing [21, 24, 41, 50]. These challenges have focused on specific problems that affect the accuracy of the methods such as multi-modal information [22], cross-ethnicity [20], 3D mask attack [23] and surveillance scenarios [9], however less constrains have been imposed to model's compactness and efficiency.

In this paper, we aim at studying the effectiveness of lightweight deep network architectures to enable the future development of accurate and efficient face anti-spoofing solutions in embedded and mobile devices. Specifically, we selected three state-of-the-art lightweight models that have showed to be well-suitable for different computer vision tasks resulting in very low computational cost. The main contributions of this work are summarized below:

- We investigate the effectiveness of lightweight deep networks for face anti-spoofing scenarios.
- We provide an extensive experimental evaluation on challenging unseen benchmarks and cross-dataset con-

ditions, which serves as a baseline for future research directions towards solving the efficient deployment of FAS solutions.

- We show the advantages of using lightweight architectures for efficient FAS deployment by comparing their computational complexity with state-of-the-art FAS methods.

The remainder of this paper is organized as follows. Section 2 takes an overview of recent FAS methods as well as existing lightweight deep networks. In Section 3, we explain the methodology used for applying lightweight architectures on FAS scenarios. Section 4 presents an extensive experimental evaluation with the selected lightweight models on FAS datasets. Finally, Section 5 summarizes the conclusions of this work.

2. Related work

In this section, we provide an overview on the state-of-the-art Face Anti-Spoofing (FAS) approaches and Lightweight Deep Networks focused on improving efficiency run-time performance.

2.1. Face anti-Spoofing

There is a large number of FAS methods described in literature. They can be broadly categorized into two types: those that rely on specific hardware/sensors and those that use only RGB cameras. The approaches that incorporate specific hardware may use structured-light 3D sensors, Time of Flight (ToF) sensors, Near-infrared (NIR) sensors, thermal sensors, or other aides that considerably enhance the accuracy of this task [3]. 3D sensors, for example, can distinguish a genuine 3D face from a 2D photo by analyzing depth maps [3]. NIR sensors, on the other hand, can detect video replay attacks as electronic displays appear uniformly dark when exposed to NIR illumination. While these techniques offer great accuracy, they are not commonly used in practical applications because they typically require costly sensors or are integrated into other devices [33].

Earlier approaches that do not require any specific hardware, were mainly designed for printed photo attacks and were based on handcrafted features such as speeded-up robust features (SURF), histogram of oriented gradients (HoG), local binary patterns (LBPs), among others to extract discriminative features that were used by a trained binary classifier to distinguish between live faces and spoof faces [47]. Robust input spaces like color and frequency spectrum were also explored, as well as the inclusion of temporal cues from eyes or lips, to improve spoof detection performance [42], or the use of remote PhotoPlethysmoGraphy (rPPG) for measuring facial micro-intensity changes corresponding to blood pulse [25].

In recent years, deep learning-based FAS methods have become the most common choice as feature extractors, achieving state-of-the-art performance in most of large-scale face-anti-spoofing benchmarks [40]. The first attempt to use Convolutional Neural Networks (CNNs) to detect spoofing attacks was proposed in [45]. The method utilized a one-path AlexNet to learn distinctive features between genuine and fake faces. Later in 2016, the first end-to-end framework for face anti-spoofing was proposed, also based on one-path AlexNet [36]. Since then, various architectures have been developed for detecting photo and video replay attacks [34, 43]. One of the most prevalent CNN architectures used in face anti-spoofing is ResNet, which can learn powerful feature representations of facial images [34]. However, designing neural networks in advance is still quite challenging. Hence, there is a recent trend towards automatically designing neural networks [49]. Unfortunately, most of these methods use computationally expensive models that are unsuitable for real-time FAS applications.

Few works consider lightweight CNNs for efficient deployment of FAS models. In [48], the authors propose a method called Auto-FAS, intending to discover well-suitable lightweight networks for mobile-level face anti-spoofing. In Auto-FAS, a special search space is designed to restrict the model's size, and pixel-wise binary supervision is used to improve the model's performance. Thus, the development of FAS solutions based on lightweight network architectures still deserves more attention.

2.2. Lightweight deep networks

To overcome the challenge of computational complexity, many works have proposed efficient and lightweight architectures for various computer vision tasks. Common solutions involve quantizing weights and/or activations of a baseline CNN model into lower-bit representations [17], or pruning unimportant filters based on FLOPs [10]. Other methods directly hand-craft more efficient mobile architectures. One such example is SqueezeNet [16], which employs lower-cost 1×1 convolutions to reduce the number of parameters and filter sizes. MobileNet [14] heavily utilizes depthwise separable convolution to minimize computation density. ShuffleNets [28, 51] use low-cost group convolution and channel shuffling. More recently, MobileNetV2 [**?**] set a new benchmark for lightweight models in image classification by introducing inverted residuals and linear bottlenecks, while MicroNet [19] is one of the most compact and energy-efficient convolutional networks developed thus far by using Micro-Factorized convolutions. However, designing hand-crafted models requires significant human effort due to the potentially vast design space.

Reinforcement learning has been one of the most used approach to automate architecture design and efficiently search for competitive accuracy. Due to the exponential

growth of a fully configurable search space, initial efforts focused on cell level structure search with the same cell being reused in all layers [37, 55]. More recent methods attempt to optimize multiple objectives, such as model size and accuracy, while searching for light CNNs [8, 15, 39, 54]. Among these methods, MnasNet [39], was built upon the MobileNetV2 structure by introducing lightweight attention modules based on squeeze and excitation into the bottleneck structure. Then, a combination of these layers was used as building blocks on MobileNetv3, in order to build a most effective model [13].

Several of the above lightweight mobile architectures have been modified in order to enhance their ability to discriminate and generalize for face recognition purposes [6, 29, 44]. These specific models have been explored on different face recognition scenarios, such as image and video FR [32], cross-pose FR [4], low-resolution FR [30] and masked FR [31]. However, very few works consider lightweight CNNs for face anti-spoofing as an option for maintaining accuracy performance while improving efficiency in real-world scenarios.

3. Methodology

In this section, we present the lightweight architectures selected in this study, as well as some implementation details for applying them in FAS scenarios.

3.1. Baseline architectures

We have implemented three state-of-the-art lightweight deep models for face anti-spoofing: ShuffleNetv2 [28], MobileNetv3 [13] and MicroNet [19].

ShuffleNetv2 [28] is a popular hand-craft efficient mobile architecture, whose design is guided by the evaluation of a direct metric (e.g., speed) instead of indirect ones (e.g., FLOPs) on the target platform. It is inspired by ShuffleNetv1 [51], which utilised pointwise group convolutions, bottleneck-like structures, and a channel shuffle operation. Based on some practical guidelines, a simple operator called channel split was introduced by the authors, allowing to maintain a large number and equally wide channels with neither dense convolution nor too many groups. Similar to ShuffleNetv1, the number of channels in each block is scaled to generate ShuffleNetv2 networks for four different levels of complexities marked as 0.5×, 1×, 1.5× and 2×.

MobileNetv3 [13] is a convolutional neural network specifically designed for mobile phone CPUs. It combines platform-aware network architecture search and the NetAdapt algorithm [46] to effectively find optimized models for a given hardware platform. In addition, the network design includes the incorporation of an squeeze-and-excitation block into the core architecture and the redesign of some of the expensive layers. MobileNetV3 is defined as two models: MobileNetV3-Large and MobileNetV3-Small for different multipliers like 0.35, 0.5, 0.75, 1.0 and 1.25. These models are targeted at high and low resource use cases, respectively.

MicroNet [19] is an efficient lightweight CNN architecture manually designed for improving accuracy at the extremely low FLOPs. It is based on Micro-Factorized convolution, which factorizes a convolution matrix into low rank matrices, to integrate sparse connectivity into convolution. Four handcrafted models (M0, M1, M2, M3) of different computational cost (4M, 6M, 12M, 21M MAdds) were proposed without using network architecture search. All these models consist of three Micro-Blocks that combine Micro-Factorized pointwise and depthwise convolutions in different ways but following the same pattern (stem layer $\rightarrow$ Micro-Block-A $\rightarrow$ Micro-Block-B $\rightarrow$ Micro-Block-C) from low to high layers. In addition, a new dynamic activation function, named Dynamic Shift Max (DY-Shift-Max), is proposed to improve the non-linearities of the network by fusing channels with dynamic coefficients.

3.2. Implementation details

All the lightweight models take face images as the input with a resolution size of 224×224. To avoid the contribution of the context information from the liveness background area such as paper edges or mobile phone borders, we extract the bounding box of face regions as preprocessing stage. First, face detection and landmark localization are performed by RetinaFace algorithm [7]. The detected faces are aligned by using five landmark points and then, we used square bounding boxes from the aligned crop faces in order to maintain the aspect ratio of the image, and preserving any significant features. These square bounding boxes are obtained by expanding the bounding boxes returned by RetinaFace, ensuring the size of resulting square face regions do not grow more than 20% of RetinaFace region sizes. Thus, any forgery boundaries from spoof face samples are included in the obtained square face regions. Finally, in order to fulfill with the input resolution size of lightweight networks, the square bounding boxes are resized to 224×224.

For the experiments, we use ShuffleNetv2 with a complexity level of 2×, MobileNetv3-Large with multiplier of 1.25 and MicroNet-M3. For all models, we replace the last layer with a new fully-connected layer making it suitable for binary classification.

4. Experimental evaluation

In this section, we show the overall advantages of using lightweight CNNs for face anti-spoofing and compare the

obtained results with state-of-the-art methods on different datasets. In addition, we include ResNet model [12] with 50 layers (ResNet50) as a baseline in all of the experiments.

4.1. Databases

We conduct extensive experiments on two challenging and large-scale face anti-spoofing datasets: the CelebA-Spoof [52] and the InsightFace Wild Face Anti-Spoofing (InsightFace WFAS) [41].

CelebA-Spoof [52] is a large-scale face anti-spoofing dataset, which contains 625,537 images from 10,177 subjects. Live images are selected from the CelebA dataset [27], while spoof images are captured from 8 scenes with more than 10 sensors including phones, pads and personal computers equipped with different resolutions. Each image in CelebA-Spoof is annotated with 43 rich attributes on face, illumination, environment and spoof types. There are four major categories for the spoof type including Print, Paper Cut, Replay and 3D. Figure 1 shows some examples of the original live and spoof images from the CelebA-Spoof database.

Figure 1. Examples of original live and spoof type images from CelebA-Spoof database.

The CelebA-Spoof dataset is splited into training, validation, and test sets with a ratio of 8:1:1, ensuring that for all three sets there is no overlap on subjects. This means that, there is no case of a live image of one certain subject in the training set while its counterpart spoof image in the test set.

InsightFace WFAS [41] is a large-scale in-the-wild dataset recently released at the fourth edition of the Face Anti-Spoofing Workshop and Challenge@CVPR 2023. It contains 529,571 live images of 148,169 identities and 853,729 spoof images of 321,751 identities. In the spoof photos, there are three major categories including 2D Print, 2D Display and 3D PAS Type, and 17 subcategories. In Figure 2, we show some examples of original images and spoof images from the three major categories.

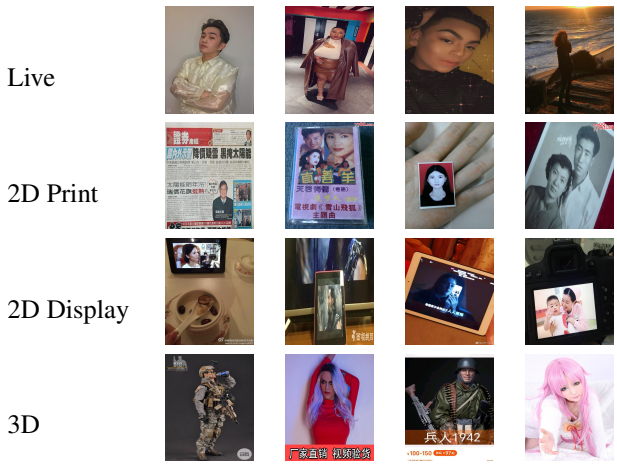


Figure 2. Examples of original live and spoof type images from InsightFace WFAS database.

For the experiments, the InsightFace WFAS dataset is divided in train, development and test subsets. For training there are 205,146 live images and 282,917 spoof images, while the development and test sets consist of 51,299/88,759 and 273,126/482,053 live/spoof images, respectively.

4.2. Training setting

We trained the models for 70 epochs on two NVIDIA-A6000 GPUs using PyTorch. We used the Stochastic Gradient Descent optimizer, with a learning rate value of 0.1 multiplied by 0.2 every at epochs 20 and 40. We also employed a weight decay of 0.0005 with a momentum of 0.9. The batch size was adjusted per our memory limitations, usually set for 1024 batches.

For data augmentation, we performed random horizontal flipping, ISO noise, brightness and contrast, and motion blur transformations, with probabilities of 0.5, 0.2, 0.3, and 0.2, respectively. The networks were supervised with a softmax cross-entropy loss (CE) function.

For training in the CelebA-Spoof database, the pretrained weights on ImageNet were used to initialize the networks, while for InsightFace WFAS dataset, the networks are trained from scratch, per the guidelines of the FAS@CVPR2023 Wild track challenge.

4.3. Performance metrics

For the intra-dataset testing, we selected the standardized ISO/IEC 30107-3:2017 metrics [1]: Attack Presentation Classification Error Rate (APCER), Normal/Bona Fide Presentation Classification Error Rate (NPCER/BPCER) and Average Classification Error Rate (ACER). APCER and BPCER/NPCER are used to measure the classification error rates on presentation attacks and live samples, respectively:

$$APCER = \frac{FP}{TN + FP} \tag{1}$$

$$BPCER = \frac{FN}{FN + TP} \tag{2}$$

The ACER value is calculated as the mean of BPCER and APCER, evaluating the reliability of intra-dataset performance.

For cross-test evaluation between CelebA-Spoof and InsightFace WFAS, we adopt Half Total Error Rate (HTER), which is defined in terms of two error rates, False Acceptance Rate (FAR) and False Rejection Rate (FRR) as follows:

$$HTER = \frac{FAR + FRR}{2} \tag{3}$$

4.4. Intra-dataset evaluation

In this section, we follow the widely used intra-dataset evaluation in order to analyze the discrimination ability of models for face anti-spoofing detection under scenarios with slight domain shift. As the training and testing data are from the same datasets, they share similar domain distribution in terms of the recording environment, subject behavior, etc.

4.4.1 Results on CelebA-Spoof

In Table 1 we report the performance of FAS models trained and evaluated on the whole training set and testing set of CelebA-Spoof, respectively. We follow the general "intra" protocol, in which different spoof types, environments and illumination conditions are used for both training and testing. Table 1 shows the APCER, BPCER, and ACER metrics for the selected lightweight models (ShuffleNetv2, MobileNetv3, and MicroNet), the baseline (ResNet50) and four models reported in [52] (Auxiliary [26], BASN [18], AENet, and AENet$_{C,S,G}$).

We can observed that the best results are obtained by AENet$_{C,S,G}$, however, this is somewhat expected since this method uses the semantic and geometric information from the annotations in CelebA-Spoof dataset [52]. The results reported without considering the complementary information (AENet) are very similar to the ones obtained by the lightweight models. In particular, MobileNetv3, outperforms this AENet model and the ResNet50 baseline.

Method	APCER(%)	BPCER(%)	ACER(%)
ResNet50	11.64	0.48	6.06
Auxiliary [52]	5.71	1.41	3.56
BASN [52]	4.00	1.10	2.60
AENet [52]	6.10	1.60	3.80
AENet$_{C,S,G}$ [52]	**2.29**	**0.96**	**1.63**
ShuffleNetv2	7.79	1.25	4.52
MobileNetv3	5.46	0.96	3.21
MicroNet	8.31	1.49	4.90

Table 1. Performance comparison on CelebA-Spoof dataset.

4.4.2 Results on InsightFace WFAS

Table 2 presents the performance of lightweight models trained and tested on InsightFace WFAS dataset. Development set is used to select the threshold that is used in the testing phase. We also include the results of ResNet18 provided by the authors of this dataset [41]. However, we must highlight that these results do not provide a fair comparison since they are based on a different version of the test set that was not further released. Although ResNet50 achieves the best results, the lightweight models, and specifically MobileNetv3, reaches comparable results.

Method	APCER(%)	BPCER(%)	ACER(%)
ResNet18	6.15	**8.87**	7.51
ResNet50	**4.47**	10.05	**7.26**
ShuffleNev2	7.66	9.61	8.64
MobileNetv3	6.20	10.14	8.17
MicroNet	9.31	12.84	11.08

Table 2. Results on Test Set from InsightFace WFAS database.

In Table 3 we show more detailed results for the different spoof types existing on the InsightFace WFAS dataset. It can be seen that the highest errors for all models are obtained for 3D attacks. The highest gap between the lightweight models and the baseline is obtained in 2D Print attacks, and in particular the larger difference is in the Scan-Photo class.

4.5. Ablation study

To further analyze the performance and explore some details of selected lightweight models, we conduct various ablation studies.

	2D Print			2D Display			3D	
	PictureBook	ScanPhoto	Packaging	Cloth	TV	Computer	Doll	Wax
ResNet50	2.50	5.51	2.03	0.00	1.23	1.87	8.05	9.00
ShuffleNetv2	6.04	9.75	4.61	2.70	2.20	3.12	10.19	14.92
MobileNetv3	3.85	11.30	2.71	1.35	2.24	2.52	16.71	16.40
MicroNet	7.03	8.84	4.33	0.45	2.36	2.86	15.86	18.54

Table 3. APCER results (%) on the InsightFace WFAS Test Set by spoof category.

4.5.1 Impact of non-linearities

In Table 4, we compare the performance of lightweight models using three different activation functions ReLU [35], PReLU [11] and DY-Shift-Max [19] on the InsightFace WFAS dataset. We can see that, for ShuffleNetv2 and MobileNetv3, ReLU and PReLU exhibit very similar performance, being PReLU a little better. In the case of MicroNet, DY-Shift-Max reaches the best ACER value, which corroborates the effectiveness of authors' proposal of using this activation function with their model [19].

Method	Activation	ACER(%)
ShuffleNetv2	ReLU	8.72
	PReLU	**8.64**
	DY-Shift-Max	11.42
MobileNetv3	ReLU	8.41
	PReLU	**8.17**
	DY-Shift-Max	10.33
MicroNet	ReLU	13.86
	PReLU	11.67
	DY-Shift-Max	**11.08**

Table 4. Impact of non-linearities on the lightweight FAS models in the InsightFace WFAS dataset.

4.5.2 Impact of fine-tuning

Table 5 presents the performance of lightweight face models with and without fine-tuning (FT) on the CelebA-Spoof dataset. It can be clearly seen that, retraining models from scratch degrades considerably the performance of models, especially for the spoof images classification. In contrast, fine-tuning offers a more accurate option which also requires less computational efforts.

4.6. Cross-dataset evaluation

In order to measure the cross-dataset level domain generalization ability, models trained on one dataset (source domain) are then tested on an unseen dataset (shifted target domain). Specifically, the cross-dataset testing is conducted between CelebA-Spoof and InsightFace WFAS databases. Thus, two evaluation protocols are implemented. In the first one, the models are trained on CelebA-Spoof and then, they are tested on InsightFace WFAS, while the second one consists of training on InsightFace WFAS and testing on CelebA-Spoof.

Method	APCER(%)	BPCER(%)	ACER(%)
ShuffleNetv2	17.77	**1.13**	9.45
ShuffleNetv2-FT	**7.79**	1.25	**4.52**
MobileNetv3	15.24	**0.91**	8.08
MobileNetv3-FT	**5.46**	0.96	**3.21**
MicroNet	13.71	3.29	8.50
MicroNet-FT	**8.31**	**1.49**	**4.90**

Table 5. Performance comparison of using models trained from scratch and fine-tuning (FT) models on CelebA-Spoof dataset.

Table 6 shows the results from cross-dataset testing. It can be seen that, for all models, the performance degrades with respect to those obtained in the intra-dataset evaluations. Nevertheless, lightweight FAS models improve the HTER results achieved by ResNet50 on the first protocol, being the MicroNet the best performing one. On the second protocol, both ResNet50 and lightweight models obtain comparable results.

Protocol	Model	HTER(%)
Trained on CelebA-Spoof	ResNet50	44.5
Tested on InsightFace WFAS	ShuffleNetv2	43.5
	MobileNetv3	38.1
	MicroNet	**36.2**
Trained on InsightFace WFAS	ResNet50	**29.1**
Tested on CelebA-Spoof	ShuffleNetv2	**29.1**
	MobileNetv3	30.1
	MicroNet	29.2

Table 6. HTER (%) results of cross-dataset testing between CelebA-Spoof and InsightFace WFAS datasets.

4.7. Visualization

In order to gain a deeper understanding of the characteristics that impact the classification decisions, we generated relevancy maps for some live and spoof faces. We adopt the Gradient-weighted Class Activation Mapping (Grad-CAM) method [38] that uses the gradient information flowing into the last convolutional layer of the CNN to assign importance values to each neuron for a particular decision of interest, without any modification in the network architecture. In Figures 3 and 4, we show the activated regions for some sample images of CelebA-Spoof and InsightFace WFAS databases, respectively. As it can be seen, there are subtle differences between the relevance regions on spoof and live samples.

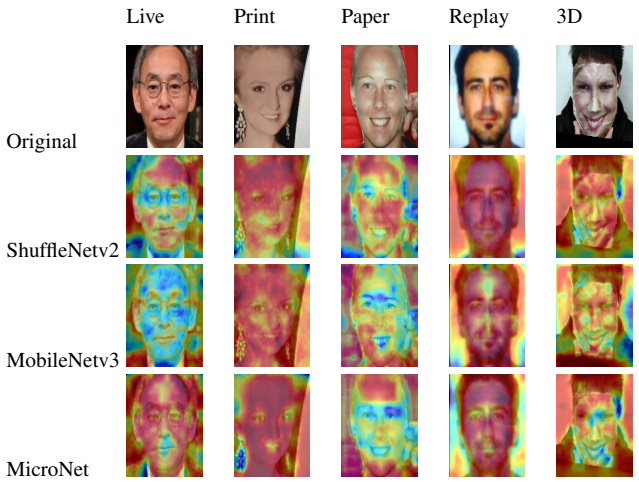


Figure 3. Grad-CAM for images from the CelebA-Spoof database.

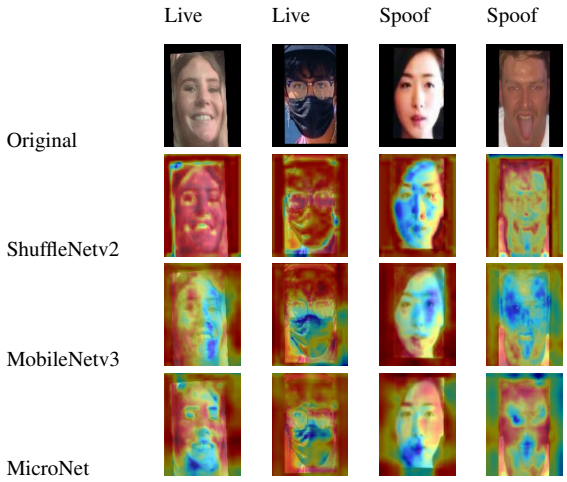


Figure 4. Grad-CAM for images from the InsightFace WFAS database.

Figure 5 and Figure 6 show the t-SNE plots of ShuffleNetv2, MobileNetv3 and MicroNet embeddings obtained from pretrained models on the InsightFace WFAS and CelebA-Spoof datasets, respectively. The plots were computed using the predicted labels on the InsightFace WFAS plot and the provided spoof type labels on the CelebA-Spoof plot, for a subset from the test sets of both databases. It can be seen that in general, there is an acceptable separability between spoof and live samples for the three models. MobileNetV3 shows more complex embedding regions per class, capturing non-linearities more accurately.

4.8. Computational complexity

In this section, we assess the computational complexity of the used lightweight models and compare them with popular FAS deep models. Specifically, we analyze the Floating Point Operations Per Second (FLOPs) and model parameters. From Table 7, we can see that the selected lightweight architectures exhibit significant computation improvements, allowing us to reduce the complexity of FAS solutions. In particular, ShuffleNetv2 has the least number of parameters, followed by MicroNet which presents an extremely low GFLOPs value, outperforming the remaining models by a large margin.

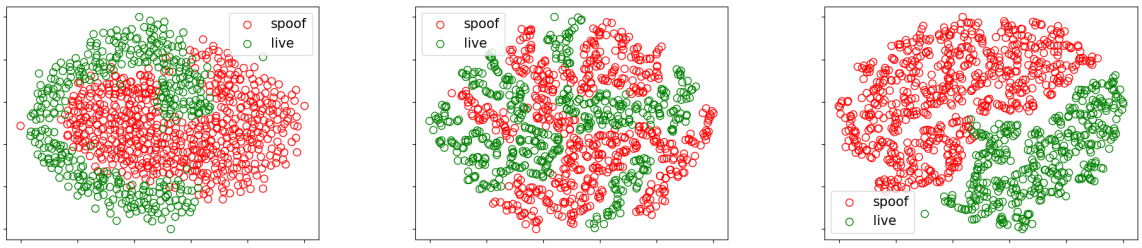
Method	GFLOPs	Params (M)
ResNet18	1.82	11.18
ResNet50	4.14	25.64
AENet [52]	3.64	42.70
Auxiliary [26]	8.50	2.20
Auto-FAS [48]	0.53	2.70
ShuffleNetv2	0.15	**1.57**
MobileNetv3	0.38	4.75
MicroNet	**0.03**	1.82

Table 7. Computational complexity comparison.

4.9. Discussion

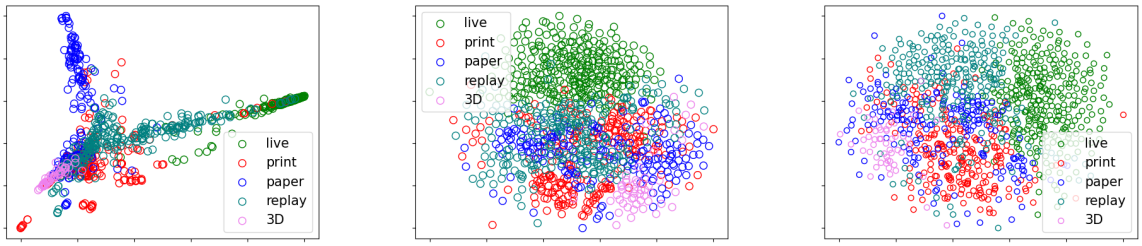
From the experimental results obtained on both CelebA-Spoof and InsightFace WFAS datasets, it can be seen that FAS methods based on lightweight networks are able to achieve an accuracy as good as state-of-the-art FAS methods based on heavier deep models, with a lower computational complexity. In particular, lightweight models were able to improve the performance of ResNet50 in most of evaluated FAS scenarios.

On the other hand, it was shown that PReLU activation function is the best choice for ShuffleNetv2 and MobileNetv3, while DY-Shift-Max achieves the best performance in the case of MicroNet. In addition, we corroborate that fine-tuning already pre-trained models offers signifi-



(a) (b) (c)

Figure 5. T-SNE plots corresponding to the embedding distribution from (a) ShuffleNetv2, (b) MobileNetv3 and (c) MicroNet on the InsightFace WFAS dataset.



(a) (b) (c)

Figure 6. T-SNE plots corresponding to the embedding distribution from (a) ShuffleNetv2, (b) MobileNetv3 and (c) MicroNet on the CelebA-Spoof dataset.

cantly better results than retraining the models from scratch. Even though lightweight FAS models work well on intra-dataset scenario, cross-dataset performance needs further improvements.

Taking into account that, just training or fine-tuning lightweight models for the FAS task was sufficient to achieve competitive results in challenging benchmarks, future improvements can be focus on using this kind of architectures as backbone of more complex FAS approaches including domain generalization or zero/few-shot learning techniques. Improvements in training methodologies are also a viable research direction. Exploiting rich attribute information can lower the error of specific spoof classes; however, it does not guarantee an improvement on the Average Classification Error rate.

All the results show that, using lightweight architectures is promising and reliable to be deployed FAS applications in mobile devices. In general, MobileNetv3 model offers the best trade-off between accuracy and efficiency, indicating its effectiveness in the design of real-time FAS.

5. Conclusion

In this paper, we have shown the effectiveness of lightweight deep networks for the face anti-spoofing task. Specifically, assess the performance of three state-of-the-art lightweight models: ShuffleNetv2, MobileNetv3 and MicroNet, in the challenge and large-scale CelebA-Spoof and InsightFace WFAS datasets. Experimental results on both unseen attacks and domains show that, using lightweight models achieved very competitive results compared with heavy deep models such as ResNet50. Moreover, we analyzed the impact on the accuracy of the lightweight models when applying different activation functions and using fine-tuned pretrained models instead of trained from scratch. Regarding the computational complexity of the lightweight models, they prove to be very suitable in dealing with the current limitations of FAS methods for efficient deployment. Among the tested lightweight networks, MobileNetv3 exhibited the best overall performance for FAS scenarios. We hope that, this work serve as baseline for further research related to the development of accurate and efficient face anti-spoofing solutions in mobile devices.

Acknowledgments

The authors would like to thank the financial support from Tecnológico de Monterrey through the “Challenge-Based Research Funding Program 2022”. Project ID # E120-EIC-GI06-B-T3-D.

References

[1] Information technology- biometric presentation attack detection- part 3: Testing and reporting, international organization for standardization, iso/iec dis 30107-3:2017, 2017. 5

[2] Insaf Adjabi, Abdeldjalil Ouahabi, Amir Benzaoui, and Abdelmalik Taleb-Ahmed. Past, present, and future of face recognition: A review. *Electronics*, 9(8):1188, 2020. 1

[3] Ghazel Albakri and Sharifa Alghowinem. The effectiveness of depth data in liveness face authentication using 3d sensor cameras. *Sensors*, 19(8):1928, 2019. 2

[4] Fernando Alonso-Fernandez, Javier Barrachina, Kevin Hernandez-Diaz, and Josef Bigun. Squeezefaceposenet: Lightweight face verification across different poses for mobile platforms. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part VIII*, pages 139–153. Springer, 2021. 3

[5] Peter Anthony, Betul Ay, and Galip Aydin. A review of face anti-spoofing methods for face recognition systems. In *2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–9. IEEE, 2021. 1

[6] Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In *Biometric Recognition: 13th Chinese Conference, CCBR 2018, Urumqi, China, August 11-12, 2018, Proceedings 13*, pages 428–438. Springer, 2018. 3

[7] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5203–5212, 2020. 3

[8] Jin-Dong Dong, An-Chieh Cheng, Da-Cheng Juan, Wei Wei, and Min Sun. Dpp-net: Device-aware progressive search for pareto-optimal neural architectures. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 517–531, 2018. 3

[9] Hao Fang, Ajian Liu, Jun Wan, Sergio Escalera, Chenxu Zhao, Xu Zhang, Stan Z Li, and Zhen Lei. Surveillance face anti-spoofing. *arXiv preprint arXiv:2301.00975*, 2023. 1

[10] Ariel Gordon, Elad Eban, Ofir Nachum, Bo Chen, Hao Wu, Tien-Ju Yang, and Edward Choi. Morphnet: Fast & simple resource-constrained structure learning of deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1586–1595, 2018. 2

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015. 6

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 4

[13] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019. 3

[14] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. 2

[15] Chi-Hung Hsu, Shu-Huan Chang, Jhao-Hong Liang, Hsin-Ping Chou, Chun-Hao Liu, Shih-Chieh Chang, Jia-Yu Pan, Yu-Ting Chen, Wei Wei, and Da-Cheng Juan. Monas: Multi-objective neural architecture search using reinforcement learning. *arXiv preprint arXiv:1806.10332*, 2018. 3

[16] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016. 2

[17] Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew Howard, Hartwig Adam, and Dmitry Kalenichenko. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2704–2713, 2018. 2

[18] Taewook Kim, YongHyun Kim, Inhan Kim, and Daijin Kim. Basn: Enriching feature representation using bipartite auxiliary supervisions for face anti-spoofing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019. 5

[19] Yunsheng Li, Yinpeng Chen, Xiyang Dai, Dongdong Chen, Mengchen Liu, Lu Yuan, Zicheng Liu, Lei Zhang, and Nuno Vasconcelos. Micronet: Improving image recognition with extremely low flops. In *Proceedings of the IEEE/CVF International conference on computer vision*, pages 468–477, 2021. 2, 3, 6

[20] Ajian Liu, Xuan Li, Jun Wan, Yanyan Liang, Sergio Escalera, Hugo Jair Escalante, Meysam Madadi, Yi Jin, Zhuoyuan Wu, Xiaogang Yu, et al. Cross-ethnicity face anti-spoofing recognition challenge: A review. *IET Biometrics*, 10(1):24–43, 2021. 1

[21] Ajian Liu, Zichang Tan, Jun Wan, Sergio Escalera, Guodong Guo, and Stan Z Li. Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1179–1187, 2021. 1

[22] Ajian Liu, Jun Wan, Sergio Escalera, Hugo Jair Escalante, Zichang Tan, Qi Yuan, Kai Wang, Chi Lin, Guodong Guo, Isabelle Guyon, et al. Multi-modal face anti-spoofing attack detection challenge at cvpr2019. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 1

[23] Ajian Liu, Chenxu Zhao, Zitong Yu, Anyang Su, Xing Liu, Zijian Kong, Jun Wan, Sergio Escalera, Hugo Jair Escalante, Zhen Lei, et al. 3d high-fidelity mask face presentation attack detection challenge. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 814–823, 2021. 1

[24] Ajian Liu, Chenxu Zhao, Zitong Yu, Jun Wan, Anyang Su, Xing Liu, Zichang Tan, Sergio Escalera, Junliang Xing, Yanyan Liang, et al. Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection. *IEEE Transactions on Information Forensics and Security*, 17:2497–2507, 2022. 1

[25] Siqi Liu, Pong C Yuen, Shengping Zhang, and Guoying Zhao. 3d mask face anti-spoofing with remote photoplethysmography. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14*, pages 85–100. Springer, 2016. 2

[26] Yaojie Liu, Amin Jourabloo, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 389–398, 2018. 5, 7

[27] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015. 4

[28] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European conference on computer vision (ECCV)*, pages 116–131, 2018. 2, 3

[29] Yoanna Martinez-Diaz, Luis S Luevano, Heydi Mendez-Vazquez, Miguel Nicolas-Diaz, Leonardo Chang, and Miguel Gonzalez-Mendoza. Shufflefacenet: A lightweight face architecture for efficient and highly-accurate face recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019. 3

[30] Yoanna Martínez-Díaz, Heydi Méndez-Vázquez, Luis S Luevano, Leonardo Chang, and Miguel Gonzalez-Mendoza. Lightweight low-resolution face recognition for surveillance applications. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 5421–5428. IEEE, 2021. 3

[31] Yoanna Martínez-Díaz, Heydi Méndez-Vázquez, Luis S Luevano, Miguel Nicolás-Díaz, Leonardo Chang, and Miguel González-Mendoza. Towards accurate and lightweight masked face recognition: an experimental evaluation. *IEEE Access*, 10:7341–7353, 2021. 3

[32] Yoanna Martinez-Diaz, Miguel Nicolas-Diaz, Heydi Mendez-Vazquez, Luis S Luevano, Leonardo Chang, Miguel Gonzalez-Mendoza, and Luis Enrique Sucar. Benchmarking lightweight face architectures on specific face recognition scenarios. *Artificial Intelligence Review*, pages 1–44, 2021. 3

[33] Zuheng Ming, Muriel Visani, Muhammad Muzzamil Luqman, and Jean-Christophe Burie. A survey on anti-spoofing methods for facial recognition with rgb cameras of generic consumer devices. *Journal of Imaging*, 6(12):139, 2020. 1, 2

[34] Chaitanya Nagpal and Shiv Ram Dubey. A performance evaluation of convolutional neural networks for face anti spoofing. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019. 1, 2

[35] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010. 6

[36] Keyurkumar Patel, Hu Han, and Anil K Jain. Cross-database face antispoofing with robust feature representation. In *Biometric Recognition: 11th Chinese Conference, CCBR 2016, Chengdu, China, October 14-16, 2016, Proceedings 11*, pages 611–619. Springer, 2016. 2

[37] Hieu Pham, Melody Guan, Barret Zoph, Quoc Le, and Jeff Dean. Efficient neural architecture search via parameters sharing. In *International conference on machine learning*, pages 4095–4104. PMLR, 2018. 3

[38] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 7

[39] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, Mark Sandler, Andrew Howard, and Quoc V Le. Mnasnet: Platform-aware neural architecture search for mobile. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2820–2828, 2019. 3

[40] Denis Timoshenko, Konstantin Simonchik, Vitaly Shutov, Polina Zhelezneva, and Valery Grishkin. Large crowdcollected facial anti-spoofing dataset. *2019 Computer Science and Information Technologies (CSIT)*, pages 123–126, 2019. 2

[41] Dong Wang, Qiqi Shao, Haochi He, Zhian Chen, Jia Guo, Jiankang Deng, Jun Wan, Ajian Liu, Sergio Escalera, Hugo Jair Escalante, Isabelle Guyon, Zhen Lei, Chenxu Zhao, and Shaopeng Tang. Insightface wild anti-spoofing dataset. https://github.com/deepinsight/insightface/tree/master/challenges/cvpr23-fas-wild, 2023. 1, 4, 5

[42] Zezheng Wang, Zitong Yu, Chenxu Zhao, Xiangyu Zhu, Yunxiao Qin, Qiusheng Zhou, Feng Zhou, and Zhen Lei. Deep spatial gradient and temporal depth learning for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5042–5051, 2020. 2

[43] Hao Xue, Jing Ma, and Xiaoyu Guo. A hierarchical multimodal cross-attention model for face anti-spoofing. *Available at SSRN 4327758*. 2

[44] Mengjia Yan, Mengao Zhao, Zining Xu, Qian Zhang, Guoli Wang, and Zhizhong Su. Vargfacenet: An efficient variable group convolutional neural network for lightweight face recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019. 3

[45] Jianwei Yang, Zhen Lei, and Stan Z Li. Learn convolutional neural network for face anti-spoofing. *arXiv preprint arXiv:1408.5601*, 2014. 2

[46] Tien-Ju Yang, Andrew Howard, Bo Chen, Xiao Zhang, Alec Go, Mark Sandler, Vivienne Sze, and Hartwig Adam. Netadapt: Platform-aware neural network adaptation for mobile applications. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 285–300, 2018. 3

[47] Zitong Yu, Xiaobai Li, Jingang Shi, Zhaoqiang Xia, and Guoying Zhao. Revisiting pixel-wise supervision for face anti-spoofing. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(3):285–295, 2021. 2

[48] Zitong Yu, Yunxiao Qin, Xiaqing Xu, Chenxu Zhao, Zezheng Wang, Zhen Lei, and Guoying Zhao. Auto-fas: Searching lightweight networks for face anti-spoofing. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 996–1000, 2020. 2, 7

[49] Zitong Yu, Chenxu Zhao, Zezheng Wang, Yunxiao Qin, Zhuo Su, Xiaobai Li, Feng Zhou, and Guoying Zhao. Searching central difference convolutional networks for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5295–5305, 2020. 2

[50] Shifeng Zhang, Ajian Liu, Jun Wan, Yanyan Liang, Guodong Guo, Sergio Escalera, Hugo Jair Escalante, and Stan Z Li. Casia-surf: A large-scale multi-modal benchmark for face anti-spoofing. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(2):182–193, 2020. 1

[51] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6848–6856, 2018. 2, 3

[52] Yuanhan Zhang, ZhenFei Yin, Yidong Li, Guojun Yin, Junjie Yan, Jing Shao, and Ziwei Liu. Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*, pages 70–85. Springer, 2020. 4, 5, 7

[53] Yuanhan Zhang, Zhenfei Yin, Jing Shao, Ziwei Liu, Shuo Yang, Yuanjun Xiong, Wei Xia, Yan Xu, Man Luo, Jian Liu, et al. Celeba-spoof challenge 2020 on face anti-spoofing: Methods and results. *arXiv preprint arXiv:2102.12642*, 2021. 1

[54] Yanqi Zhou, Siavash Ebrahimi, Sercan Ö Arık, Haonan Yu, Hairong Liu, and Greg Diamos. Resource-efficient neural architect. *arXiv preprint arXiv:1806.07912*, 2018. 3

[55] Barret Zoph and Quoc V Le. Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*, 2016. 3