MixFaceNets: Extremely Efficient Face Recognition Networks

Fadi Boutros\textsuperscript{1,2}, Naser Damer\textsuperscript{1,2}, Meiling Fang\textsuperscript{1,2}, Florian Kirchbuchner\textsuperscript{1}, Arjan Kuijper\textsuperscript{1,2}
\textsuperscript{1}Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany
\textsuperscript{2}Mathematical and Applied Visual Computing, TU Darmstadt, Darmstadt, Germany
Email: fadi.boutros@igd.fraunhofer.de

Abstract

In this paper, we present a set of extremely efficient and high throughput models for accurate face verification, MixFaceNets which are inspired by Mixed Depthwise Convolutional Kernels. Extensive experiment evaluations on Label Face in the Wild (LFW), Age-DB, MegaFace, and IARPA Janus Benchmarks IJB-B and IJB-C datasets have shown the effectiveness of our MixFaceNets for applications requiring extremely low computational complexity. Under the same level of computation complexity (\( \leq 500\text{M FLOPs} \)), our MixFaceNets outperform MobileFaceNets on all the evaluated datasets, achieving 99.60\% accuracy on LFW, 97.05\% accuracy on AgeDB-30, 93.60 TAR (at FAR1e-6) on MegaFace, 90.94 TAR (at FAR1e-4) on IJB-B and 93.08 TAR (at FAR1e-4) on IJB-C. With computational complexity between 500M and 1G FLOPs, our MixFaceNets achieved results comparable to the top-ranked models, while using significantly fewer FLOPs and less computation overhead, which proves the practical value of our proposed MixFaceNets. All training codes, pre-trained models, and training logs have been made available https://github.com/fdbtrs/mixfacenets.

1. Introduction

Recent advancements in the development of deep convolutional neural networks and the novelty of margin-based Softmax loss functions have significantly improved the State-Of-the-Art (SOTA) performance of face recognition to an extraordinary level [6], even under challenging conditions [5, 2]. However, deploying deep learning-based face recognition models in the embedded domains and other use-cases constrained by the computational capabilities and high throughput requirements is still challenging [19, 7].

Recently, a great progress in designing efficient face recognition solutions has been achieved by utilizing lightweight deep learning model architecture designed for common computer version tasks such as MobileNetV2 [24], ShuffleNet [16], VarGNet [31] for face recognition [4, 15, 18, 29]. MobileFaceNet [4] was one of the earliest works that proposed an efficient face recognition model with around 1M of parameters and 439M FLOPs. MobileFaceNet architecture is inspired by MobileNetV2 [24]. The architecture of MobileNetV2 is based on an inverted residual structure and the depthwise separable convolution [10]. AirFace [15], ShuffleFaceNet [18] and VarGFaceNet [29] model architectures are built from MobileNetV2 [24], ShuffleNetV2 [16] and VarGNet [31], respectively, reaching high levels of accuracy using compact models with around 1G FLOPs computation complexity. ShuffleNetV2 utilizes channel shuffle operation proposed by ShuffleNetV1, achieving an acceptable trade-off between accuracy and computational efficiency. VarGFaceNet [31] proposed to fix the number of input channels in each group convolution instead of fixing the total group numbers in an effort to balance the computational intensity inside the convolutional block. VarGFaceNet was the winner of the DeepGlint-Light track of ICCV Lightweight Face Recognition (LFR) Challenge (2019) [7]. The deepglink-light track of the LFR challenge targets face recognition for environment constraint by the computational complexity of 1G FLOPs and memory footprint of 20M (around 5M of trainable parameters). Very recently, MixNets proposed new convolution building block (MixConv) by extending vanilla depthwise convolution [10] with multiple kernel sizes. Using neural architecture search, MixNets developed highly efficient networks. MixNets outperformed previous mobile models including MobileNets [24, 10], ShuffleNetV2 [16] on image classification and object detection tasks.

In this work, we propose a set of extremely efficient architectures for accurate face verification and identification, namely the MixFaceNets. We opt to use MixNets as a baseline network structure to develop our MixFaceNets. We carefully designed tailored head and embedding settings that are suitable for face recognition. We also extend the MixConv block with a channel shuffle operation aiming at increasing the discriminative ability of MixFaceNets. With computation complexity of 451M FLOPs, our MixFaceNet-
S and ShuffleMixFaceNet-S achieved 99.60 and 99.56% accuracies on Labeled Faces in the Wild (LFW) [12] and 92.23 and 93.60 TAR (at FAR1e-6) on MegaFace [14] which are significantly higher than the ones achieved by MobileFaceNets [4] with a comparable level of computational complexity (99.55% accuracy on LFW and 90.16% TAR (at FAR1e-6) on MegaFace). Also, our MixFaceNets achieve comparable results to the SOTA solutions that have computation complexity of thousands of MFLOPs.

2. Related Works

Common deep learning-based face recognition models are typically computationally expensive for deployment on low computational power devices and high throughput processes. This challenge has received increased attention in the literature in the last few years [19, 7]. In this section, we list out and discuss the recent efforts on designing an efficient deep learning model for face recognition. While there are a wide range of evaluation datasets for face recognition and each of the listed previous works reported the performance on a different set of evaluation datasets, we present in this section the reported accuracy by these previous works on the LFW [12] as it is the most commonly reported benchmark. We also account for the computational complexity of the list of works in terms of the number of trainable parameters and the FLOPs when it is feasible.

The Light CNN [28] was one of the earliest works that presented 3 network architectures for learning compact representation on a large-scale database. The proposed architectures- Light CNN-4, Light CNN-9, and Light CNN-29, contain 4, 5.5, and 12.6m trainable parameters and FLOPs of 1.5G, 1, and 3.9G, respectively. The best performance among these three architectures on the LFW dataset (99.33%) is achieved by the Light CNN-29. Compared to the recent efficient face recognition models [19], Light CNN architecture is considered computational expensive. ShiftNet [27] proposed shift-based modules as an alternative to spatial convolutions and it is adopted for the face recognition task. The presented face model contains 0.78m parameters and it achieved 96% accuracy on the LFW dataset. MobileFaceNets [4] is based on MobileNetV2 [24] and it achieved a high accuracy on LFW dataset with around one million of trainable parameter and 443M FLOPs. MobileFaceNets model architecture is based on the residual bottlenecks proposed by MobileNetV2 [24] and depth-wise separable convolutions layer, which allows building CNN with a smaller set of parameters compared to standard CNNs. Different from MobileNetV2 architecture, MobileFaceNet uses Parametric Rectified Linear Unit (PReLU) [9] as the non-linearity in all convolutional layers and replaces the last global average pooling with linear global depth-wise convolution layer as a feature output layer. The best-reported accuracy on LFW by MobileFaceNets (1m parameters and 433M FLOPs) was 99.55%.

Several recent works utilized lightweight models designed for a common computer vision task to learn face representation [18, 29]. ShuffleFaceNet [18] is a compact face recognition model based on ShuffleNet [16]. Similar to MobileFaceNet [4], ShuffleFaceNet replaces the last global average pooling layer with a global depth-wise convolution layer and a Rectified Linear Unit (ReLU) with PReLU. ShuffleFaceNet presented four models with different complexity levels. The best-reported accuracy on LFW was 99.67% (2.6m parameters and FLOPs of 577.5M). VarGFaceNet [29] deployed variable group convolutional network proposed by VarGNet [31] to design a compact face recognition model with 5m trainable parameters and 1G FLOPS. VarGFaceNet adds squeeze and excitation (SE) block on the VarGNet block, replaces ReLU with PReLU, and uses variable group convolution along with pointwise convolution as the feature output layer. VarGFaceNet achieved 99.683% accuracy on LFW. This accuracy is increased to 99.85% by training VarGFaceNet with recursive knowledge distillation. However, the computational cost of VarGFaceNet is higher than the ones of ShuffelFaceNet and MobileFaceNet. AirFace [15] proposed to increase the MobileFaceNet network width and the depth and adding attention module, achieving 99.27% accuracy on LFW. The work also presented a loss function named Li-ArcFace which is based on ArcFace. Li-ArcFace demonstrates better converging and performance than ArcFace loss on low dimensional features embedding. The proposed model by AirFac has a computational cost of 1G FLOPS.

In a recent survey by Martinez-Diaz et al. [19], the computational requirements and the verification performance of five lightweight model architectures are analyzed and evaluated. The evaluated models are MobileFaceNet (0.9G FLOPs and 2.0m parameters), VarGFaceNet [29] (1G FLOPs and 5m parameters), ShufleFaceNet [18] (577.5M FLOPs and 2.6m parameters), MobileFaceNetV1 (1.1G FLOPs and 3.4m parameters), and ProxylessFaceNAS (0.9G FLOPs and 3.2m parameters). MobileFaceNetV1, ProxylessFaceNAS and MobileFaceNet are extended versions of MobileNetV1 [10], ProxylessNAS [3], and MobileFaceNets [4], respectively. The evaluated models in [19]- ShufleFaceNet, VarGFaceNet, MobileFaceNet achieved very close accuracy on LFW (99.7%) and other evaluated datasets, while MobileFaceNetV1 and ProxylessFaceNAS achieved slightly lower accuracy.

Among the previous listed works, MobileFaceNets [4] is the only architecture that achieved high accuracy with less than 500M FLOPs. With the almost same number of FLOPs as in MobileFaceNets, our MixFaceNets outperform MobileFaceNets and achieved competitive results to other models with fewer FLOPs using extremely efficient architecture as shown in Tables 1 and 2.
3. Approach

This section presents the architecture of our efficient MixFaceNets designed for accurate face verification. Figure 1 illustrates the architecture of the MixFaceNet, partially inspired by MixNets [25]. To improve the accuracy and the discriminative ability of MixNet, we (a) implement different head settings, (b) introduce channel shuffle operation to the MixConv block, and (c) propose different embedding settings. We discuss in this section the MixConv as an inspiration to our work. Then, we present the detailed architecture of our MixFaceNet.

3.1. Mixed Depthwise Convolutional Kernels

Depthwise Convolution is one of the most popular building block for mobile models [10, 24, 16]. Depthwise Convolutional applies a single convolution filter over each channel of input, thus it reduces the number of parameters and achieves computational efficiency while maintaining the discriminative ability of the convolution [10]. Mixed Depthwise Convolutional Kernels (MixConv) [25] extends vanilla depthwise convolution by using multiple kernel sizes in a single convolution. MixConv depends on mixing up multiple kernel sizes in a single convolution by splitting up the input of convolution into groups and applying different kernel sizes to each group. Different from vanilla depthwise convolution, MixConv can capture different patterns from convolution input at various resolutions. Also, it requires fewer parameters and it is more computationally efficient than using a single kernel e.g. using multiple kernels of size $[(3, 3), (5, 5), (7, 7)]$ is more computationally efficient than using a single kernel of size $7 \times 7$. For example, given a convolution input of size $w \times h \times c$ and multiple kernels of size $[(3, 3), (5, 5), (7, 7)]$, MixConv split the input into 3 groups, each of them has a dimension of $w \times h \times c/3$. Then it uses different kernels for each of these groups. Finally, the three outputs are concatenated to produce the final convolution output. An example of MixConv and downsample-MicConv blocks are shown as part of Figure 1. Unlike manually designed mobile models [16, 18, 24], MixConv utilized neural architecture search to develop new series of MixConv-based networks, namely MixNets (MixNet-S, MixNet-M, and MixNet-L). MixNet-
S and MixNet-M are developed using neural architecture search, while MixNet-L is obtained by scaling up the number of channels in each block by a factor of 1.3. For details about the network structure and search space, we refer to the original work [25].

3.2. MixFaceNet Architecture

We deploy MixNets [25] as baseline network structure to develop our proposed MixFaceNets. Figure 1 illustrates the network architecture for MixFaceNet-S. For the network head, we apply fast down-sampling in the first $3 \times 3$ convolution (stride=2) followed by batch normalization [13] and PReLU non-linearity [9]. Then, we use one residual block as shown in Figure 1.b. For the MixFaceNet-S network, we use the same global structure as MixNet-S. However, different from MixNet-S, we did not apply down-sampling at the first convolution after the head stage to reserve as much as possible information at the earliest stage of the network. We propose to mix up both channels and kernels to increase the discriminative ability of MixFaceNet and improve the model accuracy. We achieve that by introducing shuffle operation to the MixConv block. The channel shuffle operation is proposed by [18] to enabling information flowing between different groups of channels. We apply a channel shuffle operation with a group value of 2 after each MixConv block. Thus, MixFaceNet can capture high and low-resolution patterns at different scales and it also enables information communication between various groups of channels. Figures 1.c and 1.d show the detailed structure of MixConv and down-sampling MixConv blocks with the channel shuffle operation. All MiXConv blocks uses swish as an activation function [23] followed by batch normalization. MiXConv includes also squeeze-and-excitation (SEBlock) [11] at the end of each block. Finally, to obtain the feature embedding of the input face image, we replace the last global average pooling layer with global depthwise convolution. Specifically, we first add $1 \times 1$ convolutional layer (Conv1) with stride=1 and zero paddings followed by batch normalization [13] and PReLU none-linearity [9]. In Conv1, we expand the channel from 200 to 1024. Then, we use $7 \times 7$ convolution layer (stride=1, padding=0 and grouping=1024) followed by batch normalization. Finally, we use $1 \times 1$ convolution with 512 output channels followed by batch normalization to obtain the final feature embedding which is of size $512 - d$, as shown in Figure 1.e.

4. Experimental Setup

Dataset: We use the MS1MV2 dataset [6] to train our MixFaceNet models. The MS1MV2 is a refined version of the MS-Celeb-1M [8] by [6] and it contains 5.8M images of 85K identities. The Multi-task Cascaded Convolutional Networks (MTCNN) solution [30] is used to detect and align face images. The MixFaceNet models process aligned and cropped face images of the size $112 \times 112 \times 3$ to produce $512 - d$ feature embeddings. We evaluate our MixFaceNets on the widely used LFW [12] and on the AgeDB-30 [21] datasets. Also, we report the performance of the MixFaceNets on large scale evaluation datasets including MegaFace [14], IJB-B [26] and IJB-C [20].

MixFaceNets Training Setup: The proposed models in this paper are implemented using Pytorch. All models are trained using ArcFace loss [6]. We set the margin value of ArcFace loss to 0.5 and the feature scale to 64. We set the batch size to 512 and train our model using dis-
Table 1: MixFaceNets verification accuracies on LFW and AgeDB-30 datasets. The first row of the table shows the achieved result by the current SOTA ReNet100 models. The table is divided into two parts. The first part of the table shows the achieved result by models that have computational complexity between 500 and 1000M FLOPs. The second part of the table shows the achieved by models that have computational complexity less than 500M FLOPs. The number of decimal points is reported as in the related works.

<table>
<thead>
<tr>
<th>Method</th>
<th>FLOPs (M)</th>
<th># Params. (M)</th>
<th>LFW (%)</th>
<th>AgeDB-30 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArcFace (LResNet100E-IR) [6]</td>
<td>24211</td>
<td>65.2</td>
<td>99.83</td>
<td>98.15</td>
</tr>
<tr>
<td>AirFace [15]</td>
<td>1000</td>
<td>-</td>
<td>99.27</td>
<td>-</td>
</tr>
<tr>
<td>ShuffleFaceNet 2x [18]</td>
<td>1050</td>
<td>4.5</td>
<td>99.62</td>
<td>97.28</td>
</tr>
<tr>
<td>ShuffleFaceNet 1.5× [18]</td>
<td>577.5</td>
<td>2.6</td>
<td>99.67</td>
<td>97.32</td>
</tr>
<tr>
<td>VarGFaceNet [29]</td>
<td>1022</td>
<td>5</td>
<td>99.68</td>
<td>98.10</td>
</tr>
<tr>
<td>MobileFaceNet [19]</td>
<td>933</td>
<td>2.0</td>
<td>99.7</td>
<td>97.6</td>
</tr>
<tr>
<td>MobileFaceNetV1 [19]</td>
<td>1100</td>
<td>3.4</td>
<td>99.4</td>
<td>96.4</td>
</tr>
<tr>
<td>ProxylessFaceNAS [19]</td>
<td>900</td>
<td>3.2</td>
<td>99.2</td>
<td>94.4</td>
</tr>
<tr>
<td>MixFaceNet-M (ours)</td>
<td>626.1</td>
<td>3.95</td>
<td>99.68</td>
<td>97.05</td>
</tr>
<tr>
<td>ShuffleMixFaceNet-M (ours)</td>
<td>626.1</td>
<td>3.95</td>
<td>99.60</td>
<td>96.98</td>
</tr>
<tr>
<td>MobileFaceNets [4]</td>
<td>439.8</td>
<td>0.99</td>
<td>99.55</td>
<td>96.07</td>
</tr>
<tr>
<td>ShuffleFaceNet 0.5× [18]</td>
<td>66.9</td>
<td>0.5</td>
<td>99.23</td>
<td>93.22</td>
</tr>
<tr>
<td>MixFaceNet-S (ours)</td>
<td>451.7</td>
<td>3.07</td>
<td>99.60</td>
<td>96.63</td>
</tr>
<tr>
<td>ShuffleMixFaceNet-S (ours)</td>
<td>451.7</td>
<td>3.07</td>
<td>99.58</td>
<td>97.05</td>
</tr>
<tr>
<td>MixFaceNet-XS (ours)</td>
<td>161.9</td>
<td>1.04</td>
<td>99.60</td>
<td>95.85</td>
</tr>
<tr>
<td>ShuffleMixFaceNet-XS (ours)</td>
<td>161.9</td>
<td>1.04</td>
<td>99.53</td>
<td>95.62</td>
</tr>
</tbody>
</table>

Distributed Partial-FC algorithm [1] on one machine with 4 Nvidia GeForce RTX 6000 GPUs to enable faster training on a single node. All models are trained with Stochastic Gradient Descent (SGD) optimizer with an initial learning rate of 1e-1. We set the momentum to 0.9 and the weight decay to 5e-4. The learning rate is divided by 10 at 80k, 140k, 210k, and 280k training iterations. During the training, we evaluate the model on LFW and AgeDB after each 5650 training iterations. The training is stopped after 300k iterations. During the testing, the classification layer is removed and the feature is extracted from the last layer, which is of the size $512 - d$. We used euclidean distance between feature vectors in all experiments for comparison.

5. Results

This section presents the achieved result by the MixFaceNets on different benchmarks. We acknowledge the evaluation metrics in the ISO/IEC 19795-1 [17] standard. However, for the sake of comparability and reproducibility, we follow the evaluation metrics used in the utilized benchmarks and the previous works reporting on them.

5.1. Result on LFW and AgeDB-30

LFW [12] is one of the widely used datasets for unconstrained face verification. The dataset contains 13,233 images of 5749 different identities. The result on LFW is reported as verification accuracy (as defined in [12]) following the unrestricted with labeled outside data protocol using the standard 6000 comparison pairs defined in [12]. AgeDB [21] is common used in-the-wild dataset for evaluating age-invariant face verification. It contains 16,488 images of 568 different identities. We report the performance as verification accuracy for AgeDB-30 (years gap 30) as it is the most challenging subset of AgeDB. Also, it is the commonly reported set of AgeDB by the recent SOTA face recognition models. Similar to the LFW, we followed the standard protocol provided by AgeDB to evaluate our models on AgeDB-30 [21]. Table 1 shows the achieved result on LFW and AgeDB-30. We first report the result for one of the top-ranked face recognition models, ArcFace (LResNet100E-IR) [6], to give an indication of the current SOTA performance on LFW (99.83 %) and AgeDB-30 (98.15 %). Although, the ArcFace (LResNet100E-IR) model [6] is far from being considered an efficient model, in comparison to lightweight models, with 24211M FLOPs and 65.2m parameters. Then, the second section of Table 1 presents the achieved result by the recent lightweight models that have computational complexity less than 500M FLOPs. The best-reported result on LFW (99.70% accuracy) is achieved by the MobileFaceNet [19] (933M FLOPs). Our MixFaceNet-M achieved a competitive result on LFW (99.68% accuracy) using 38% fewer FLOPs (626M). A similar result has been achieved on AgeDB-30. Our MixFaceNets achieved very close accuracy to the current SOTA models using a more efficient model architecture with almost the same number of parameters. Among all models that have computational complexity less than 500M FLOPs, our MixFaceNet models outperform all listed mod-
Table 2: The achieved results on large-scale evaluation datasets- MegaFace, IJB-B, and IJB-C. The results on MegaFace and MegaFace (R) [6] using FaceScrub as probe set are reported as face identification (Rank-1 %) and verification (TAR at FAR=1e–6) for different lightweight models. The last two columns of the table show 1:1 verification TAR (at FAR=1e–4) on IJB-B and IJB-C. The first row reports the evaluation result using SOTA face recognition model- ArcFace (LResNet100E-IR) which contains 65.2M parameters and 24211M FLOPs. The rest of the table is organized in two parts- models with computational complexity between 500 and 1000M FLOPs and models with less than 500M FLOPs. The number of decimal points is reported as in the related works. Considering the computation complexity, our MixFaceNet models are evaluated as ones of top-ranked models.

![Table 2: The achieved results on large-scale evaluation datasets- MegaFace, IJB-B, and IJB-C. The results on MegaFace and MegaFace (R) [6] using FaceScrub as probe set are reported as face identification (Rank-1 %) and verification (TAR at FAR=1e–6) for different lightweight models. The last two columns of the table show 1:1 verification TAR (at FAR=1e–4) on IJB-B and IJB-C. The first row reports the evaluation result using SOTA face recognition model- ArcFace (LResNet100E-IR) which contains 65.2M parameters and 24211M FLOPs. The rest of the table is organized in two parts- models with computational complexity between 500 and 1000M FLOPs and models with less than 500M FLOPs. The number of decimal points is reported as in the related works. Considering the computation complexity, our MixFaceNet models are evaluated as ones of top-ranked models.](image-url)

5.2. Result on IJB-B and IJB-C

The IARPA Janus Benchmark-B ((IJB-B) face dataset consists of 1,845 subjects of 21,798 still images and 55,026 frames from 7,011 videos [26]. The IJB-B verification protocol provides a list of 10,270 genuine comparisons and 8M impostor comparisons. The IARPA Janus Benchmark-C (IJB-C) face dataset is an extension of IJB-B by increasing the database variability and size with additional 1,661 new subjects [20]. The IJB-C consists of 31,334 still images and 117,542 frames from 11,779 videos of 3531 subjects. The IJB-C verification protocol provides a list of 19,557 genuine comparisons and 15,638,932 impostor comparisons. The result on IJB-C and IJB-B is reported in terms of true accepted rates (TAR) at false accepted rates (FAR) (as defined in [26]) equal to 1e-4 to provide a comparable result with the previous works evaluated on these datasets. The achieved verification performances on IJB-B and IJB-C by our MixFaceNet models are reported as part of the Table 2. Our MixFaceNet-M and ShuffleMixFaceNet-M models achieved close results to the top-ranked models using significantly fewer FLOPs.

5.3. Result on MegaFace

The evaluation protocol of MegaFace includes gallery (1m images from Flickr) and probe (FaceScrub and FGNe) sets. In this paper, we use MegaFace [14] as a gallery set and FaceScrub [22] as the probe set to provide a comparable result with the previous works evaluated on this dataset. The MegaFace [14] contains 1m images of 690K different identities and the FaceScrub contains 100K images of 530 identities [22]. The result on MegaFace is reported as identification (Rank-1) and verification (TAR at FAR=1e–6) to be compatible with the previous works evaluated on this dataset [19]. Also, we report the result on the refined version of the MegaFace presented in [6]. The face verification and identification results on the MegaFace and the refined version of MegaFace (noted as MegaFace (R)) are presented in Table 2. For all evaluated models that have computational complexity between 500 and 1000M FLOPs, our MixFaceNet-M outperformed ProxylessFaceNAS [19], VarGFaceNet [19], MobileFaceNetV1 [19], and ShuffleFaceNet [19, 18]. And it achieved very close verification and identification results to the top-ranked models- AirFace [15] and MobileFaceNet [19] using less than half the number of FLOPs. Also, when the considered computational cost is less than 500M FLOPs, in the third section of Table 2, our ShuffleMixFaceNet-S achieved the highest verification and identification performances.

5.4. Performance vs. Computational Complexity

To present the achieved results in terms of the trade-off between the verification performance and the computational complexity (represented by the number of FLOPs), we plot the number of FLOPs vs. the verification performance of our MixFaceNets and the SOTA solutions. The plots for the comparisons on the LFW, AgeDB-30, MegaFace, IJB-B,
IJB-C and MegaFace(R) benchmarks are presented in Figures 2 (a), (b), (c), (d), (e) and (f). Each of the reported solutions is presented by an indicator on the plot, where an ideal solution will tend to be placed on the top left corner (high performance and low complexity). In most ranges of the number of FLOPs, and on the 6 benchmarks, different versions of our MixFaceNets achieved the highest verification performance. Similar conclusions can be made by analyzing the presented values in Table 2.

6. Conclusion

We presented in this paper accurate and extremely efficient face recognition models, MixFaceNets. We conducted extensive experiments on popular, publicly available, datasets including LFW, AgeDB-30, MegaFace, IJB-B, and IJB-C. The overall evaluation results demonstrate the effectiveness of our proposed MixFaceNets for applications associated with low computational complexity requirements. Our MixFaceNet-S and ShuffleMixFaceNet-S outperformed MobileFaceNets [4] under the same level of computation complexity (≤500M FLOPs). Also, our MixFaceNet-M is showed to be one of the top-ranked performing models, while using significantly fewer FLOPs than the SOTA models.

Acknowledgment: This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References


