applied sciences

MDPI

Article

Deep Learning for Facial Recognition on Single Sample per Person Scenarios with Varied Capturing Conditions

Belén Ríos-Sánchez †,*, David Costa-da-Silva †, Natalia Martín-Yuste † and Carmen Sánchez-Ávila †

Group of Biometrics, Biosignals, Security and Smart Mobility, Universidad Politécnica de Madrid, 28040 Madrid, Spain; dcosta@cedint.upm.es (D.C.-d.-S.); nmartin@cedint.upm.es (N.M.-Y.); csa@cedint.upm.es (C.S.-Á.)

* Correspondence: brios@cedint.upm.es; Tel.: +34-91-067-9630

† Current address: Edif. CeDInt-UPM, Campus de Montegancedo, Pozuelo de Alarcón, 28223 Madrid, Spain.

Received: 22 October 2019; Accepted: 5 December 2019; Published: 13 December 2019

Abstract: Single sample per person verification has received considerable attention because of its relevance in security, surveillance and border crossing applications. Nowadays, e-voting and bank of the future solutions also join this scenario, opening this field of research to mobile and low resources devices. These scenarios are characterised by the availability of a single image during the enrolment of the users into the system, so they require a solution able to extract knowledge from previous experiences and similar environments. In this study, two deep learning models for face recognition, which were specially designed for applications on mobile devices and resources saving environments, were described and evaluated together with two publicly available models. This evaluation aimed not only to provide a fair comparison between the models but also to measure to what extent a progressive reduction of the model size influences the obtained results.The models were assessed in terms of accuracy and size with the aim of providing a detailed evaluation which covers as many environmental conditions and application requirements as possible. To this end, a well-defined evaluation protocol and a great number of varied databases, public and private, were used.

Keywords: biometrics; face; deep learning; single sample per person; knowledge generalisation

1. Introduction

Facial biometrics is widely extended nowadays because of its great number of applications, the maturity of the technology and the users acceptance. Typical uses include data, devices or facilities access control, surveillance, border crossing, entertainment and human–computer interaction. As might be expected, the variety of these application scenarios implies different needs regarding security, storage of information, computing capacity or samples collection. For instance, certain applications cannot store user's templates, are require to verify the identity of the user against a smart card (ID card, driving license or passport) or present difficulties to collect multiple samples of each user.

Focussing on the availability of one or more samples during the training or the enrolment of the users into the system, applications can be divided in two groups, which are known as single sample per person (SSPP) and multiple samples per person (MSPP) problems. SSPP problem has received considerable attention during the last decades because of the relevance of its applications, particularly those related to security, surveillance and border crossing. Nowadays, it is increasingly popular between e-voting and bank of the future service providers. These applications need to guarantee the security of transactions but also to offer an attractive, useful and comfortable user experience, including the access from mobile devices, so they typically avoid to store any information about the user and compare a user's photograph and the image in his/her ID card. SSPP scenarios present lower costs of collecting, storing and processing samples but also add new challenges to typical face recognition difficulties (sensitivity to capturing conditions, facial expressions and changes in the appearance of the users among others). The reduction of the number of images implies a severe reduction on the recognition accuracy of most of the methods developed up to the moment, which strongly rely on the number of samples and their quality to generate a good facial model that generalises inter- and intra-person variability. Accordingly, a solution able to extract and generalise knowledge from previous experiences and similar environments is required.

On the other hand, deep learning solutions for face recognition have received great attention during the last years. In fact, the introduction of convolutional neural networks (CNNs) for facial features extraction marked a turning point when Deep Face [1] and DeepID [2] were presented in 2014, as can be seen in the survey presented by Wang and Deng [3]. At the beginning, CNNs were mostly applied in MSPP scenarios, although they were recently also applied for SSPP problems achieving very promising results. Table 1 summarises the latest works on deep learning for face recognition on SSPP scenarios. Information about the dataset (name and main variations), the number of people and images used during the test, whether an additional dataset is used for training and the match rate are provided for each method to describe their performance. Most of these works provide an evaluation of the methods in terms of accuracy for identification scenarios. However, identity verification is a widely extended scenario. In addition, the specific requirements of the varied applications make necessary a complimentary evaluation of the architectures and the size of the models similarly to the work presented by Hu et al. [4], which quantitatively compare the architectures of CNNs and evaluate the effect of different implementation choices using the public database LFW to train the models. Moreover, the use of public datasets for testing is crucial and the direct comparison against publicly available models seems to be necessary to state a base line.

For these reasons, in this work, two publicly available models, FaceNet [5] and OpenFace [6], were evaluated for SSPP verification scenarios together with two small-size proprietary models, which were specially designed for mobile devices applications and resources saving environments. This evaluation aimed not only to provide a fair comparison between the models but also to measure to what extent a progressive reduction of the model size influences the obtained results. The models were assessed in terms of accuracy and size with the aim of evaluating their applicability to scenarios with different environmental conditions and requirements, including a test where the image of the ID Card is compared against a usual facial image. Many varied databases, public and private, were used with the aim of covering as many scenarios as possible. A distance based classifier was used for matching purposes due to the SSPP nature of the scenarios and to guarantee the straightforward portability of the solution to mobile devices.

The remainder of this article is organised as follows. First, the methods involved in the evaluation are presented in Section 2. Then, the evaluation protocol, the involved databases and the obtained results are presented in Section 3. Finally, conclusions are provided in Section 4.

Table 1. Summary of different works relating to deep learning for face recognition from a single image. ATD column shows if an additional dataset is used for training. E, I, T, LT, O, S, B, V, A and P in Variations column stand for Expression, Illumination, Time, Long Time, Orientation, Scale, Blurring, View, Accessories and Pose, respectively.

Method	Database	Main Variations	# People	# Images	ATD	Match Rate (%)
Face Identity Preserving Features + DNN ([7])	Multi-PIE	P, I, E, T	337	128,940	No	90.46
Face Recovery + FCDN ([8])	LFW	P, I, S, A, E, T	5749	13,233	CelebFaces	97.27
Deep Supervised Autoencoders ([9])	Ext. Yale B	I	38	2432	No	82.22
	AR	I, A	100	2600	No	85.21
	PIE	I, P, E	68	5508	No	72.36
	Multi-PIE	LT, I, V	337	20,209	No	76.12
Virtual image synthesis + DDAN ([10])	EK-LFH	I, B, P, V, S	30	15,930	No	72.08
	LFW	E, I, LT, O, V, A, P	158	>1580	No	97.91
JCR-ACF ([11])	AR	I, E, A	100	2600	CASIA-WF	96.10
	Multi-PIE	P, I, E	249	14,940	CASIA-WF	70.40
	LFW	E, I, LT, O, V, A, P	158	>1580	CASIA-WF	86.00
	CASIA-WF	I, P, E, A	9175	406,423	No	15.00
TLFL framework ([12])	Ext. Yale B	E	38	2404	No	34.86
	AR	E, I	100	1400	No	55.50
	PIE	P, I, E	68	11,630	No	55.89
	LFW	E, I, LT, O, V, A, P	158	>1422	No	15.21
	CAS-PEAL	E	284	1420	No	72.95
	JAFFE	E	10	213	No	89.47
Mean search + LSH + DNN ([13])	Msceleb	-	>10,000	>100,000	No	95.00
	CASIA-WF	I, P, E, A	10,408	>492,744	No	52.43
DCNN ([14])	AR	I, E, A, T	100	2600	CASIA-WF	99.76
	Ext. Yale B	P, I	38	2432	CASIA-WF	88.30
	FERET	P, E, I	200	1400	CASIA-WF	93.90
	LFW	E, I, LT, O, V, A, P	50	>500	CASIA-WF	74.00
NDRDF ([15])	AR	E, I, A, T	116	3016	No	98.00
Transfer Learning + KCFT ([16])	ORL	E, I, A, T	40	400	CASIA-WF	98.14
	FERET	P, E, I	200	1400	CASIA-WF	93.04
	LFW	E, I, LT, O, V, A, P	50	>500	No	97.49

2. Methods

2.1. Face Detection and Alignment

Before extracting the facial biometric features, it is required to locate the face area within the image. It is recommendable that the face area does not contain face surroundings, such as hair, clothes or background. Given the tendency towards deep neural networks to solve object recognition tasks and its good performance, a detector based on cascaded convolutional networks was used [17]. It is composed of three carefully designed deep convolutional networks fed an image pyramid to estimate face position and landmark locations in a coarse-to-fine manner, and exploits the inherent correlation between detection and alignment to increase their performance. In particular, the implementation provided by FaceNet [5] was employed. When multiple faces are present in the image, the biggest face is selected assuming that the user whose identity is validated is closer to the camera.

Since deep learning models are sensitive to the position of the face elements within the images, once the face is detected, it is aligned to a common reference framework. To this end, some reference points belonging to eyes, nose and mouth are detected and transformed to a fixed position.

Finally, face images must have the same size to improve the comparison performance, so all the images are resized to a common size of 160 × 160 px.

2.2. Feature Extraction

As commented above, face recognition schemes changed since the presentation of Deep Face [1] and DeepID [2] in 2014. Convolutional neural networks learn a mapping from facial images to a compact space where distances directly correspond to a measure of face similarity.

In this study, four models for facial features extraction based on the GoogleNet [18] architecture were compared: FaceNet [5], OpenFace [6], gb2s_Model1 and gb2s_Model2. These models were trained using a deep convolutional network that directly optimises the embedding itself using a triplet-based loss function based on large margin nearest neighbour [19]. These triplets consist of two matching and a non-matching roughly aligned face patches and the loss aims to separate the positive pair from the negative by a squared L2 distance margin. This way, faces of the same person have small distances and faces of different people have large distances.

Given the relevance of triplets selection, FaceNet presents a novel online negative exemplar mining strategy, which ensures consistently increasing difficulty of triplets as the network trains, and explores hard-positive mining techniquess which encourage spherical clusters for the embeddings of a single person to improve clustering accuracy. As a result, FaceNet surpassed state-of-the-art face recognition performance using only 128 bytes per face, getting a classification accuracy of 99.63% ± 0.09 on the Labeled Faces in the Wild (LFW) dataset [20]. FaceNet was trained with a private dataset containing 100M–200M images and the model size is 90 MB. Details about the methodology, system architecture and network structure can be found in [5]. The TensorFlow implementation provided by Sandberg [21] was used in this comparative.

The good results achieved by FaceNet contributed to increase the accuracy gap between state-of-the-art publicly available and private face recognition systems. Aimed to bridge this gap, a general-purpose face recognition library oriented to mobile scenarios was developed: OpenFace [6]. This real-time face recognition system was specially designed to provide high accuracy with low training and prediction times and adapts depending on the context. The model was trained using a modified version of FaceNet's nn4 network, nn4.small2, which reduces the number of parameters and the number of training images. In this case, 500 k images coming from CASIA-WebFace [22] and FaceScrub [23] datasets were used to train the model after a preprocessing stage where they were aligned. Details about the system architecture, network structure and implementation can be found in [6]. Among the four models offered by OpenFace, the nn4.small2 was selected for this comparative as it gets the best performance (92.92% ± 1.34% on the LFW dataset).

OpenFace is oriented to mobile applications and considerably reduces the size of the model (30 MB) compared to FaceNet's (90 MB). However, it could still be too large for being embedded in some mobile applications or devices with limited resources. Thus, in this work, two smaller models were trained, viz. gb2s_Model1 and gb2s_Model2, following a similar reduction process than OpenFace. These models were trained using a network inspired by FaceNet but composed of a smaller number of layers and filters. As the idea is to measure the influence of a progressive reduction of the model size on the results, the same structure of layers as OpenFace was used as an starting point but varying the number of filters to reduce the final number of parameters of the model. The architectures of the proposed models are presented in Tables 2 and 3. In addition, the training process followed the same triplet-loss architecture as FaceNet and OpenFace, thus it also provided an embedding on the unit hypersphere and Euclidean distance represented similarity. A subset of the LFW database was used to train the network and the size of the resulting gb2s_Model1 and gb2s_Model2 are 22 MB and 12.5 MB, respectively.

Table 2. gb2s_Model1 network architecture.

Type	Output Size	#1×1	#3×3 Reduce	#3×3	#5×5 Reduce	#5×5	Pool Proj
conv1 (7×7×3,2)	48×48×64						
max pool + norm	24×24×64						m 3×3,2
inception (2)	24×24×192		64	192			
norm + max pool	12×12×192						m3×3,2
inception (3a)	12×12×256	64	96	128	16	32	m, 32p
inception (3b)	12×12×320	64	96	128	32	64	ℓ_2, 64p
inception (3c)	6×6×640		128	256,2	32	64,2	m 3×3,2
inception (4a)	6×6×640	256	96	192	32	64	ℓ_2, 128p
inception (4e)	3×3×1024		160	256,2	64	128,2	m 3×3,2
inception (5a)	3×3×384	128	64	192			ℓ_2, 64p
inception (5b)	3×3×384	128	64	192			m, 64p
avg pool	384						
linear	128						
ℓ_2 normalisation	128						

Table 3. gb2s_Model2 network architecture.

Type	Output Size	#1×1	#3×3 Reduce	#3×3	#5×5 Reduce	#5×5	Pool Proj
conv1 (7×7×3,2)	48×48×64						
max pool + norm	24×24×64						m 3×3,2
inception (2)	24×24×192		64	192			
norm + max pool	12×12×192						m3×3,2
inception (3a)	12×12×256	64	96	128	16	32	m, 32p
inception (3b)	12×12×320	64	96	128	32	64	ℓ_2, 64p
inception (3c)	6×6×640		128	256,2	32	64,2	m 3×3,2
inception (4a)	6×6×312	128	48	96	16	32	ℓ_2,56p
inception (4e)	6×6×504		80	128,2	32	64,2	m 3×3,2
inception (5a)	3×3×256	96	32	128			ℓ_2, 32p
inception (5b)	3×3×256	96	32	128			m, 32p
avg pool	256						
linear	128						
ℓ_2 normalisation	128						

2.3. Matching

A distance based classifier was used to compare the biometric features given the lack of multiple samples during the enrolment in single sample per person scenarios, which are necessary to train more complex classifiers. Its simplicity and low computational requirements also guarantee a straightforward portability of the solution to any environment, including mobile devices. This classifier provides a numeric value as a result, which represents the difference between two feature vectors. Accordingly, the decision policy established in the system is to consider the compared vectors as belonging to the same person if the computed distance is lower than a previously established threshold. Since deep learning approaches evaluated in this study map the images to a compact Euclidean space, Euclidean distance was applied in this study.

3. Evaluation

3.1. Databases

Many images coming from different databases both, public and private, were used in this evaluation. The choice of databases aimed to cover as many of the most realistic cases of use as possible. Accordingly, the images present different degrees of difficulty regarding to pose, scale, background, lighting conditions, appearance, accessories and expressions. In particular, BioID ([24]), EUCFI ([25]), ORL ([26]), Extended Yale B ([27]), Print-Attack ([28]) and gb2sμMOD_Face_Dataset ([29]) were used together with three proprietary datasets:

- gb2sTablet: A set of 250 frames coming from the gb2sTablet_Face_Dataset which is part of the proprietary gb2sTablet_Database. It contains images from 60 people captured in an indoor environment using artificial lighting and simple but uncontrolled backgrounds. Images were recorded in a frontal position but without restrictions about the distance to the camera, appearance or accessories. The size of the images is 320 × 240 px.
- gb2s_Selfies: A proprietary database oriented to emulate real daily-life scenarios. Accordingly, this dataset shows different illumination conditions and backgrounds, possible presence and absence of glasses, hear variations, or expressions. Twenty-six individuals participated in the database creation and 10 images per person were recorded in 10 different sessions, one image per day. Each person captured images of his/her face using the frontal camera of his/her own mobile phone the more frontal as possible (Selfie position).
- gb2s_IDCards: A private database oriented to evaluate security applications which require from an official document to verify the identity of the users. In this case, the same 26 individuals participated in the database creation and 10 images per person were recorded in 10 different sessions, one image per day. Each person captured images of his/her ID card using the back camera of his/her own mobile phone.

In general, feature extraction models able to generalise knowledge for recognising new people are required. It is even more necessary in SSPP scenarios, thus feature extraction models were previously trained using images coming from totally different datasets. Since this training is separated from the evaluation of the whole system, training datasets are not referenced at this point.

Table 4 summarises the databases involved in the evaluation.

Table 4. Databases overview. A, B, E, L, P, S and T stand for Accessories, Background, Expressions, Lighting, Pose, Scale and Time, respectively.

Database	Access	#Users	#Images	Image Size	Color	Variations
Ext. Yale B	Public	28	16,128	640×480	gray	L, P
ORL	Public	40	400	92×112	gray	L, T, E, A
BioID	Public	23	1521	384×286	gray	B, L, S
EUCFI	Public	395	7900	180×200	color	B, S, E, A
PrintAttack	Public	38	1400	320×240	color	L
gb2sμMOD	Public	60	4220	640×480	color	B, S, A
gb2sTablet	Private	60	16,593	320×240	color	B, S, A
gb2s_Selfies	Private	26	262	-	color	B, L, T, E, A
gb2s_IDCards	Private	26	261	-	gray	L, T

3.2. Evaluation Protocol

An evaluation protocol based on the definitions suggested by the ISO/IEC 19795 standard ([30,31]) was applied to quantify the performance of the different methods described above, ensuring fair comparison between them and hopefully future research. It is composed of three parts:

1. *Dataset Organisation*. First, each evaluation dataset was divided into validation and test subsets, which contain 70% and 30% of the samples respectively. Next, validation subset is in turn separated into enrolment and access samples (also 70% and 30%, respectively), to generate the biometric profile of the users and to simulate accesses into the system. This way, methods were validated and the acceptance threshold was adjusted. Then, new accesses into the already configured system were made using the test samples, allowing for the calculation of more realistic performance rates.

2. *Computation of Scores*. A list of genuine and zero-effort impostor scores was generated at this stage. To this end, the biometric template of each user was created using only one enrolment sample, and access samples were divided into genuine and impostor, corresponding to authentic and forger users, respectively. Then, genuine scores were computed by comparing the access samples against the reference template of the same user, and impostor scores were obtained by comparing each access sample against the biometric templates of the other users. The same process was repeated for each enrolment sample.
3. *Metrics calculation*. Finally, certain metrics about the performance of the system were obtained from genuine and impostor scores. Concretely, validation results are offered in terms of Equal Error Rate (EER) and test results wer measured in terms of False Match Rate (FMR) and False Non-Match Rate (FNMR).The threshold associated to the EER computed in the validation stage was used in the test stage.

3.3. Results

An evaluation of the complete system in a SSPP scenario with identity verification purposes was carried out for each database separately, allowing for the comparability of the results. In addition, a test where an image coming from the gb2s_Selfies dataset was compared against an image of the same user coming from the gb2s_IDCards dataset was included to evaluate this relevant and complicated scenario. Tables 5 and 6 gather the results obtained during de validation and test stages respectively. Results are also illustrated in Figures 1 and 2.

It can be seen that capturing conditions have a great influence on the results. In fact, the results achieved on databases recorded under quite controlled capturing conditions, which usually present lower variability between images, are quite good for every model. However, as soon as the complexity of the datasets increases, the influence of the model become more evident. It can be observed that the most influencing conditions are lighting, in particular low light levels such as those present on Extended Yale B database, expressions and appearance of the users. The image size and the number of images per person does not seem to be very relevant. As can be expected, stronger differences between enrolment and access samples lead to worse results. It is clearly shown in the Selfies-IDCards scenario.

Focusing on the feature extraction models, it can be seen that the progressive model reduction from 9 MB to 30 MB, 22 MB and 12 MB increases the averaged validation EER in 8.37%, 10.31% and 9.79%, respectively, while the averaged test FMR raises are 8.27%, 10.11% and 9.80%, respectively, and the averaged test FNMR increments are 8.81%, 10.74% and 10.94%, respectively. It is evident that a big and well trained model is required to deal with the most complicated scenarios. Accordingly, FaceNet is able to deal even with the Selfies-IDCards, providing quite acceptable results. Applications associated to this scenario typically have strong security requirements, and the use of a bigger model could be justified. However, more research efforts are required to determine a competitive solution in terms of model size and accuracy for this particular scenario. On the other hand, applications where resources saving is a priority need smaller models such as OpenFace and gb2s_model(s). The results achieved by these models do not present a clear pattern and they strongly depend on the dataset used during the evaluation. In one half of the cases, OpenFace surpasses gb2s_models, but, on the other half, the results are similar or even worse. Differences between gb2s Model1 and gb2s Model2 are not relevant, being a little bit better the gb2s Model2 in many cases, which is the smallest. Obtained results highlight that the use of a reduced model for face recognition is viable when they are trained with a sufficiently big dataset containing enough representativeness.

Comparing validation and test results, a great consistency can be observed. The biometric template of each user is composed just by one single sample and the feature extraction models were trained with separated datasets, thus the learnt features were not directly related with the images used in the experiments and their capturing conditions. Therefore, validation and test results obtained for each dataset are pretty similar.

Finally, even when comparison against the state of the art is difficult because different authors follow different evaluation methodologies, it can be seen that the presented results generally match or exceed the state-of-the-art solutions. Focusing on those works which use the same databases for evaluation purposes, it can be observed that FaceNet and OpenFace clearly surpasses the actual state of the art. This is not the case of gb2s_Model(s), which in some cases surpasses the state of the art but in many cases does not. However, it can be observed that most state-of-the-art solutions use the same databases for training and testing purposes, which could influence the results. Thus, it is not possible to provide a fair comparison against the state of the art, and a new study following the same evaluation protocol and using the same databases, which should be different from those involved in the training process, must be carried out to this end.

Table 5. Facial recognition using deep learning and DBC on single sample per person scenarios: validation results (EER (%)).

Feature Extraction	BioID	EUCFI	ORL	Ext. Yale B	Print Attack	gb2s Tablet	gb2s µMOD	gb2s Selfies	gb2s IDCards	Selfies-IDCards
Face Net	1.47	1.16	0.25	3.60	0.05	0.05	4.28	1.91	0.55	4.66
Open Face	6.77	2.39	3.54	9.35	3.46	2.21	10.44	16.04	1.43	46.05
gb2s_Model1	16.85	2.65	6.65	27.76	4.09	0.76	18.94	11.98	2.56	28.86
gb2s_Model2	13.22	2.63	6.25	29.02	6.30	0.74	19.72	9.56	1.79	26.61

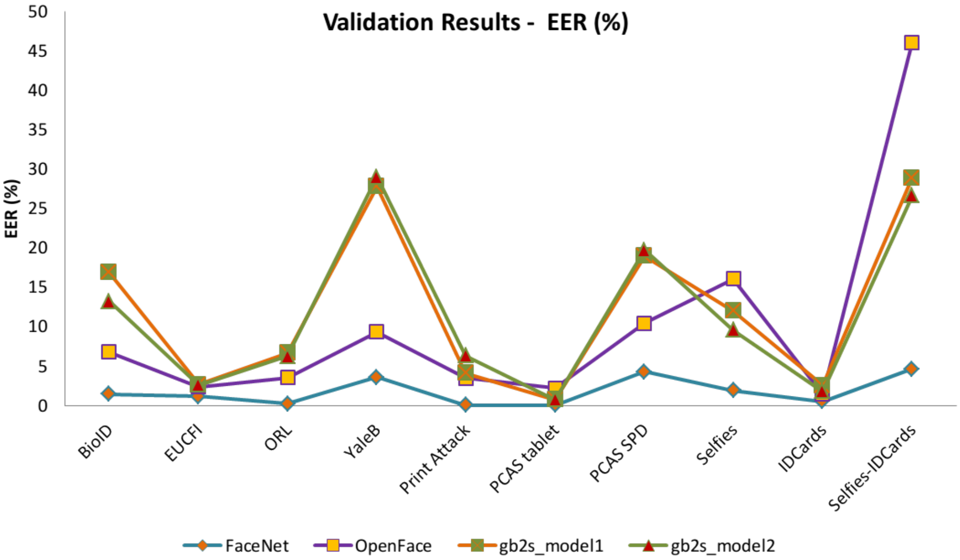

Figure 1. SSPP validation results.

Table 6. Facial recognition using deep learning and DBC on single sample per person scenarios: test results (%).

Database	FaceNet		OpenFace		gb2s_Model1		gb2s_Model2	
	FMR	FNMR	FMR	FNMR	FMR	FNMR		
BioID	1.34	0.07	6.99	7.73	16.17	17.72	13.70	14.30
EUCFI	1.21	1.85	2.44	2.16	2.55	3.71	2.79	3.30
ORL	0.35	1.25	2.62	5.60	3.29	7.25	6.35	8.25
Ext. Yale B	3.85	3.96	9.39	9.63	30.19	27.91	31.68	29.62
Print Attack	0.02	0.09	3.60	2.94	4.15	6.64	6.07	8.61
gb2sTablet	0.05	0.07	2.20	2.19	0.72	1.21	0.68	1.07
gb2sµMOD	3.89	4.77	10.67	9.76	16.18	24.87	16.53	26.91
gb2s_Selfies	1.66	0.00	13.91	17.90	10.60	11.83	8.67	11.45
gb2s_IDCards	0.70	0.76	1.86	3.33	2.47	3.05	1.51	2.29
Selfies - IDCards	4.89	2.29	47.00	42.00	29.77	18.32	28.00	18.70

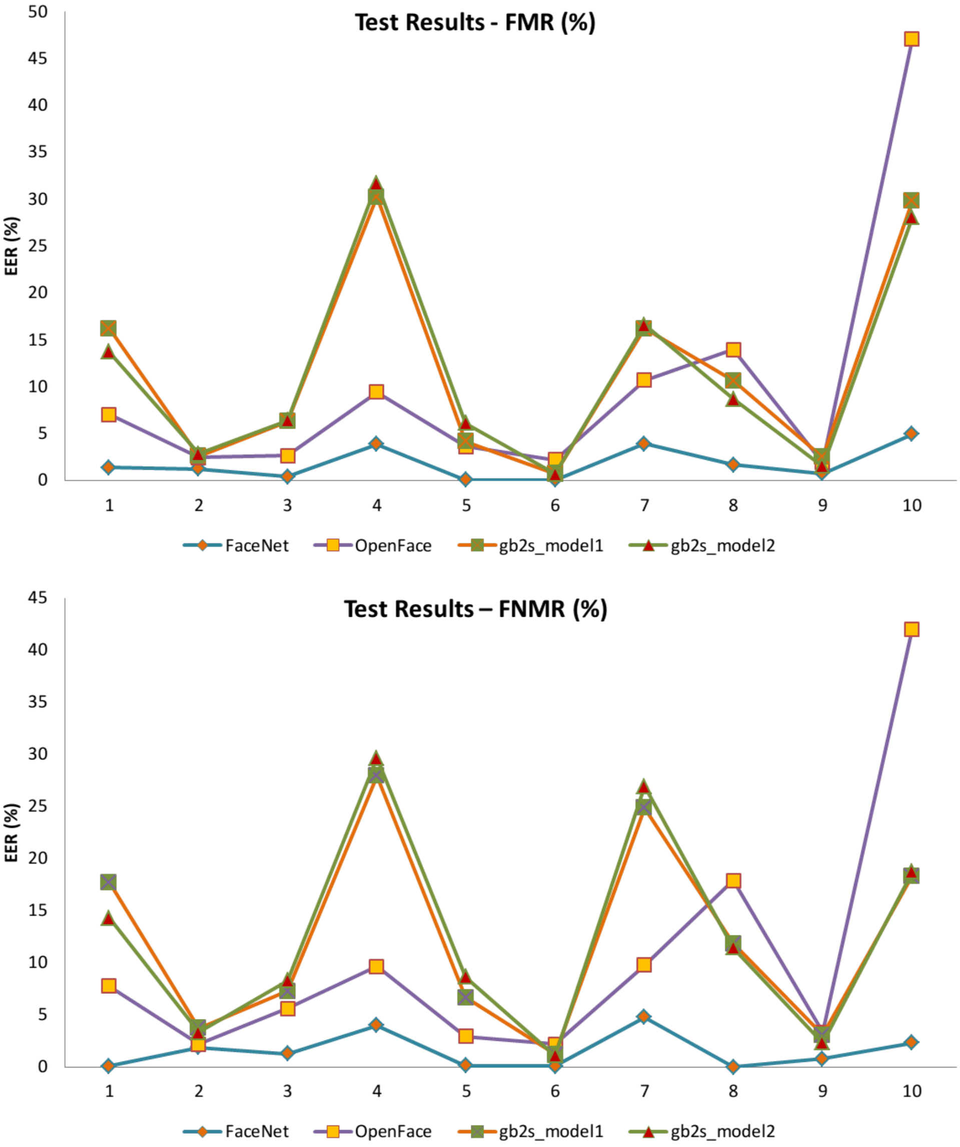

Figure 2. SSPP Validation Results.

4. Conclusions

In this study, two deep learning models for face recognition, which were specially designed for applications on mobile devices and resources saving environments, were described and evaluated together with two publicly available models (FaceNet and OpenFace) for identity verification at single sample per person scenarios. This evaluation aimed not only to provide a fair comparison between the models but also to measure to what extent a progressive reduction of the model size influences the obtained results. The models were assessed in terms of accuracy and size with the aim of evaluating their applicability to scenarios with different environmental conditions and requirements. To this end, a great number of varied databases, public and private, was used. Given that SSPP scenarios imply that there is only one sample per person during the enrolment into the system, a distance based classifier was used to compare the biometric features.

The results show that the influence of the feature extraction model become more evident as the complexity of the images increases. In fact, a big and well trained model is required to deal with really complicated scenarios that present high variability between images used to enrol users into the system and posterior accesses. However, for those scenarios where resources saving is a priority, smaller models are viable if they are trained with a sufficiently big dataset containing enough representativeness.

Author Contributions: B.R.-S. and C.S.-Á. conceived the research and the methodology. B.R.-S., D.C.-d.-S. and N.M.-Y. implemented the software and carried out the experiments. B.R.-S., D.C.-d.-S., N.M.Y. and C.S.-Á. analysed the results. B.R.-S. wrote and edited the paper and C.S.-Á. reviewed it. C.S.-Á. was in consultation at every moment and provided valuable suggestions and criticisms. C.S.-Á. and B.R.-S. acquired the fundings and administered the projects.

Funding: This work was partially funded by the project "Awesome Possum: Data driven authentication", lead by Telenor Digital AS and financed by The Research Council of Norway under project number 256581, and a private project lead by Indra Sistemas S.A.

Acknowledgments: The authors would also like to thank BioID® for the use of BioID Face Database, Libor Spacek for the use of Essex University's Collection of Facial Images (Face Recognition Data), AT&T Laboratories Cambridge for the use of the ORL Database, UCSD for the use of Extended Yale B database and IDIAP for the use of Print-Attack database, as well as all the contributors who have participated in the creation of our proprietary databases for their patience and cooperation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708.
2. Sun, Y.; Wang, X.; Tang, X. Deep Learning Face Representation from Predicting 10,000 Classes. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; IEEE Computer Society: Washington, DC, USA, 2014; pp. 1891–1898.
3. Wang, M.; Deng, W. Deep Face Recognition: A Survey. *arXiv* **2018**, arXiv:1804.06655 .
4. Hu, G.; Yang, Y.; Yi, D.; Kittler, J.; Christmas, W.; Li, S.Z.; Hospedales, T. When Face Recognition Meets with Deep Learning: An Evaluation of Convolutional Neural Networks for Face Recognition. In Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Tampa, FL, USA, 5–8 December 2015; pp. 384–392.
5. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A Unified Embedding for Face Recognition and Clustering. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015, Boston, MA, USA, 7–12 June 2015.
6. Amos, B.; Bartosz, L.; Satyanarayanan, M. *OpenFace: A General-Purpose Face Recognition Library with Mobile Applications*; Technical Report, CMU-CS-16-118; CMU School of Computer Science: Pittsburgh, PA, USA, 2016.
7. Zhu, Z.; Luo, P.; Wang, X.; Tang, X. Deep Learning Identity-Preserving Face Space. In Proceedings of the 2013 IEEE International Conference on Computer Vision and Pattern Recognition, Sydney, Australia, 1–8 December 2013; IEEE Computer Society: Washington, DC, USA, 2013; pp. 113–120.
8. Zhu, Z.; Luo, P.; Wang, X.; Tang, X. Recover Canonical-View Faces in the Wild with Deep Neural Networks. *arXiv* **2014**, arXiv:1404.3543.
9. Gao, S.; Zhang, Y.; Jia, K.; Lu, J.; Zhang, Y. Single Sample Face Recognition via Learning Deep Supervised Autoencoders. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 2108–2118. [CrossRef]
10. Hong, S.; Im, W.; Ryu, J.; Yang, H.S. SSPP-DAN: Deep domain adaptation network for face recognition with single sample per person. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 825–829.
11. Yang, M.; Wang, X.; Zeng, G.; Shen, L. Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person. *Pattern Recognit.* **2017**, *66*, 117–128. [CrossRef]
12. Guo, Y.; Jiao, L.; Wang, S.; Wang, S.; Liu, F. Fuzzy Sparse Autoencoder Framework for Single Image per Person Face Recognition. *IEEE Trans. Cybern.* **2018**, *48*, 2402–2415. [CrossRef] [PubMed]
13. Xihua, L. Improving Precision and Recall of Face Recognition in Sipp with Combination of Modified Mean Search and Lsh. Ph.D. Thesis, Beihang University, Beijing, China, 2018.

14. Zeng, J.; Zhao, X.; Gan, J.; Mai, C.; Zhai, Y.; Wang, F. Deep Convolutional Neural Network Used in Single Sample per Person Face Recognition. *Comput. Intell. Neurosci.* **2018**, *2018*, 3803627. [CrossRef] [PubMed]
15. Ouanan, H.; Ouanan, M.; Aksasse, B. Non-linear dictionary representation of deep features for face recognition from a single sample per person. *Procedia Comput. Sci.* **2018**, *127*, 114–122. [CrossRef]
16. Min, R.; Xu, S.; Cui, Z. Single-Sample Face Recognition Based on Feature Expansion. *IEEE Access* **2019**, *7*, 45219–45229. [CrossRef]
17. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
18. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
19. Weinberger, K.Q.; Blitzer, J.; Saul, L.K. Distance Metric Learning for Large Margin Nearest Neighbor Classification. In *Advances in Neural Information Processing Systems 18*; Weiss, Y., Schölkopf, B., Platt, J.C., Eds.; MIT Press: Cambridge, MA, USA, 2006; pp. 1473–1480.
20. Huang, G.B.; Mattar, M.; Berg, T.; Learned-Miller, E. Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments. In Proceedings of the Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Marseille, France, 17 October 2008.
21. Sandberg, D. Face Recognition Using Tensorflow. 2017. Available online: https://github.com/davidsandberg/facenet (accessed on 8 March 2019).
22. Yi, D.; Lei, Z.; Liao, S.; Li, S. Learning Face Representation from Scratch. *arXiv* **2014**, arXiv:1411.7923.
23. Ng, H.W.; Winkler, S. A data-driven approach to cleaning large face datasets. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; IEEE Xplore Digital Library: Piscataway, NJ, USA, 2014; pp. 343–347.
24. BioID. BioID Face Database. Available online: https://www.bioid.com/facedb/ (accessed on 8 March 2019).
25. EUCFI. The Essex University Collection of Face Images. Available online: https://cswww.essex.ac.uk/mv/allfaces/index.html (accessed on 8 March 2019).
26. ORL. The Database of Faces. Available online: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html (accessed on 8 March 2019).
27. YaleB. The Extended Yale Face Database B. Available online: http://vision.ucsd.edu/~iskwak/ExtYaleDatabase/ExtYaleB.html (accessed on 8 March 2019).
28. Anjos, A.; Marcel, S. Counter-Measures to Photo Attacks in Face Recognition: A public database and a baseline. In Proceedings of the International Joint Conference on Biometrics 2011, Washington, DC, USA, 11–13 October 2011.
29. Ríos-Sánchez, B.; Arriaga-Gómez, M.; Guerra-Casanova, J.; de Santos-Sierra, D.; de Mendizábal-Vázquez, I.; Bailador, G.; Sánchez-Ávila, C. gb2sμMOD: A MUltiMODal biometric video database using visible and IR light. *Inf. Fusion* **2016**, *32*, 64–79. [CrossRef]
30. ISO. *ISO/IEC 19795-1:2007: Information Technology—Biometric Performance Testing and Reporting—Part 1: Principles and Framework*; International Organization for Standardization (ISO): Geneva, Switzerland; 2007.
31. ISO. *ISO/IEC 19795-2:2007: Information Technology—Biometric Performance Testing and Reporting—Part 2: Testing Methodologies for Technology and Scenario Evaluation*; International Organization for Standardization (ISO): Geneva, Switzerland; 2007.