Exploring Factors for Improving Low Resolution Face Recognition

Omid Abdollahi Aghdam[1] Behzad Bozorgtabar[2]
Hazım Kemal Ekenel[1] Jean-Philippe Thiran[2]
[1]SiMiT Lab, ITU, Turkey [2] LTS5, EPFL, Switzerland
abdollahi15@itu.edu.tr behzad.bozorgtabar@epfl.ch
ekenel@itu.edu.tr jean-philippe.thiran@epfl.com

Abstract

State-of-the-art deep face recognition approaches report near perfect performance on popular benchmarks, e.g., Labeled Faces in the Wild. However, their performance deteriorates significantly when they are applied on low quality images, such as those acquired by surveillance cameras. A further challenge for low resolution face recognition for surveillance applications is the matching of recorded low resolution probe face images with high resolution reference images, which could be the case in watchlist scenarios. In this paper, we have addressed these problems and investigated the factors that would contribute to the identification performance of the state-of-the-art deep face recognition models when they are applied to low resolution face recognition under mismatched conditions. We have observed that the following factors affect performance in a positive way: appearance variety and resolution distribution of the training dataset, resolution matching between the gallery and probe images, and the amount of information included in the probe images. By leveraging this information, we have utilized deep face models trained on MS-Celeb-1M and fine-tuned on VGGFace2 dataset and achieved state-of-the-art accuracies on the SCFace and ICB-RW benchmarks, even without using any training data from the datasets of these benchmarks.

1. Introduction

Face recognition systems are now very common, from applications in our smartphones to security gates in the airports. These systems work flawlessly, when the training and test images are of high quality, have similar distributions, and do not vary much. However, in the surveillance scenarios, in which the training and test images do not have the same distribution, face recognition systems' performance deteriorates. Figure 1 illustrates the face identification scenario addressed in this paper to explore this problem. The scenario resembles a watchlist one, in which we have high quality gallery face images of the individuals recorded at indoor studio settings and low quality probe face images recorded by indoor, as in the SCFace [5], or outdoor, as in the ICB-RW [15], surveillance cameras.

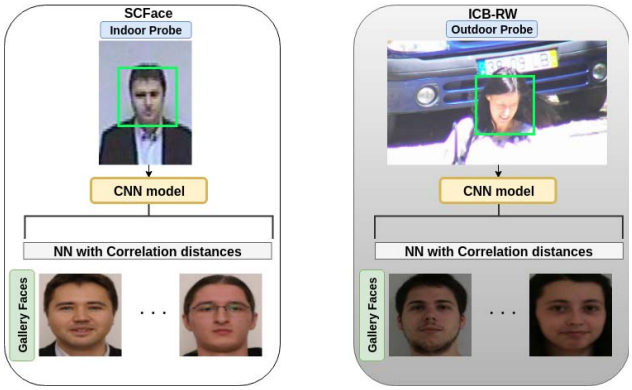

Figure 1. Face identification scenario addressed in this paper. The scenario resembles a watchlist one, in which we have high quality gallery face images of the individuals recorded at indoor studio settings and low quality probe face images recorded by indoor, as in the SCFace benchmark [5] (left), or outdoor, as in the ICB-RW benchmark [15] (right), surveillance cameras.

The recent breakthroughs in deep learning architectures [8, 7, 20, 18] and availability of large-scale training databases, e.g. CASIA Webface [25], MS-Celeb-1M [6], VGGFace2 [1], have aided the research in face recognition (FR). The advancements have been significant on the benchmarks that have relatively high resolution face images in gallery and probe sets, e.g. Labeled Faces in the Wild (LFW) [9] and YouTube Faces (YTF) [23].

In low resolution face recognition under surveillance scenarios on the other hand, for example in a watchlist application, there is a single high resolution frontal face image per subject in the gallery set, whereas, there are low resolution face images captured with surveillance cameras in the probe set, which contain appearance variations due to changes in illumination, expression, pose, motion-blur,

978-1-7281-2506-0/19/$31.00 ©2019 IEEE
DOI 10.1109/CVPRW.2019.00290

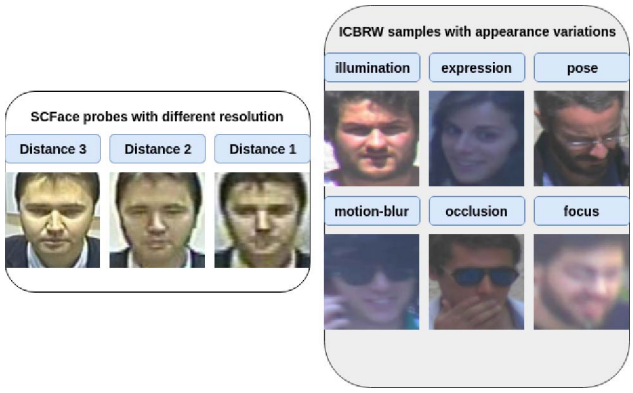

Figure 2. Sample probe images from the SCFace and the ICB-RW datasets. The probe set of SCFace contains face images with three different resolutions depending on the distance between the subject and the cameras (left), probe set of ICB-RW includes face images recorded outdoors and contain challenging appearance variations (right).

occlusion, focus, and varying resolutions as can be seen in Figure 2 for the ICB-RW [15] and the SCFace [5] benchmarks. Probe face images' quality problems and the quality mismatch between the gallery and probe images are the main causes of the performance drop in deep face recognition models when they are tested under such conditions.

To address the challenges posed by low resolution face recognition, in this work, we explore the factors that would improve Low Resolution Face Recognition (LRFR) performance. We investigate the factors, such as, appearance variety and resolution distribution of the training database, resolution matching between the gallery and probe images, and the amount of information included in the probe images. We observe that all of these factors improve the performance. We test the robustness of four state-of-the-art deep convolutional neural network (CNN) models, namely, ResNet-50 [7], SENet-50 [8], LResNet50E-IR [3], LResNet100E-IR [3] and utilize two large scale face databases, VGGFace2 [1] and MS-Celeb-1M [6], to train and fine-tune them. We present that appearance variety and resolution distribution of the training database is of paramount importance. We also analyze the impact of the resolution matching between the gallery and probe images. In contrast to a previous work [22], instead of super-resolving low resolution face images to match the resolution of the gallery and probe images, we down-sample the high resolution gallery images. We observe that matching the gallery and probe face images' resolution increases the performance significantly in the cases where the probe face images' quality is very low. Finally, we experiment different face crop sizes in order to assess the impact of information included in the face images. Experimental results indicate that cropping a larger region of the face images improves the performance.

By leveraging these factors, we achieve state-of-the-art results on the SCFace [5] and ICB-RW [15] benchmarks, even without using any data from these benchmarks to train or fine-tune the employed deep CNN models. To compare our results on the SCFace [5] benchmark with [12], we conduct 10 Repeated Random Sub-Sampling Validation (RRSSV) experiment on 80 subjects out of 130 subjects and report the mean and standard deviation of Rank-1 Identification Rate (IR). We achieve $78.5\% \pm 1.67$, $98.38\% \pm 0.48$, and $99.75\% \pm 0.16$ Rank-1 IR for distance 1, 2, and 3 (d1, d2, and d3) of SCFace [5], respectively. Our approach outperforms the state-of-the-art results presented in Deep Coupled-ResNet (DCR) [12] by the large margins of 5.2%, 4.88%, and 1.75% for d1, d2, and d3, respectively. In contrast to DCR [12], we do not exploit target dataset for fine-tuning. Furthermore, we evaluate the proposed factors on ICB-RW benchmark [15] and outperform the results reported in Ghaleb et al. [4], the best performing system in the ICB-RW 2016 challenge [15], by a significant margin of 12.52% for Rank-1 IR.

The remainder of this paper is organized as follows. In section 2, we provide an overview of related work. In section 3, face detection, feature extraction, and face identification steps are explained. Experimental results are presented and discussed in section 4. Finally, in section 5 conclusions of this work are summarized.

2. Related Work

The related works for face recognition can be grouped into Low Resolution Face Recognition (LRFR) and High Resolution Face Recognition (HRFR). The reported results on the HRFR benchmarks are nearly perfect. In FaceNet [17], a Deep Convolutional Neural Network (DCNN) architecture with Inception [20] modules is trained on a very large-scale database of 260 M images. After that, the features are L2 normalized and triplet loss is proposed to learn deep face representations. The proposed method achieved 99.63% face IR on the LFW benchmark [9] and 95.12% face IR on the YTF benchmark [23]. Sun, et al. [19] included Inception modules [20] into two VGG architectures [18], and concatenated extracted features from 25 different crops of each face per network. Afterwards, a joint Bayesian model is learned for face recognition. The proposed method achieved 99.54% verification accuracy on the LFW [9]. In SphereFace [11] the Angular-Softmax loss is introduced and adopted ResNet architecture [7] to learn face embeddings in training phase. They applied nearest neighbor classifier with cosine similarity for face identification. The applied method achieved 99.42% verification accuracy on the LFW [9] and 95.0% on YTF [23] datasets, respectively. ArcFace [3] leveraged ResNet [7] architecture and train the face identification model with additive angular margin loss. Their reported best verification accuracy are

99.83% on the LFW [9] and 98.02% on the YTF [23].

In contrast to HRFR, the performance of deep CNN models degrade significantly in LRFR. Lee et al. [10], extracted local color vector binary patterns and nearest neighbor classifier with euclidean distance metric are carried out for face identification. Average Rank-1 Identification Rate (IR) of 67.68% is reported for distance 1 and 2 (4.20m, 2.60m, respectively) of SCFace [5]. De Marsico et al. [2] applied pose and illumination normalization on faces and localized spatial correlation index for face matching. They reported 89% Rank-1 IR for distance 3 (1.0m) of SCFace [5]. A Patch Based Cascaded Local Walsh Transform (PCLWT) followed by whitened principal component analysis is employed in [21] for feature extraction. They reported 64.76%, 80.8%, and 74.92% Rank-1 IR for d1, d2, and d3 respectively. Yang et al. [24] proposed the Local-Consistency-Preserved Discriminative Multidimensional Scaling (LDMDS) approach to learn compact intra-class features and maximize inter-class distance. They selected 50 subjects, out of 130 subjects available in SCFace [5], for training and calculated Rank-1 IR for the remaining 80 subjects. They reported 62.7%, 70.7%, and 65.5% Rank-1 IR for distance 1, 2, and 3, respectively. Following [24], in Deep Coupled-ResNet (DCR) [12] a two-step multi-scale training strategy is performed to train a trunk and two branches (HR and LR branches). They trained the trunk network with three different image resolutions (112×96, 40×40, and 6×6) pre-processed from CASIA Webface database [25]. In the second step, they fixed the weights of the trunk network and trained the HR and LR branches. For that, they trained HR branch with 112×96 pixel resolution and LR branch with 112×96, 30×30, and 20×20 pixel resolutions based on the image resolutions of the distances 1, 2, and 3 respectively. After that, they fine-tuned the HR and LR branches with 50 randomly selected subjects of the SCFace dataset [5]. Deep face embedding of the gallery and probe faces of SCFace dataset [5] are extracted with HR branch and LR branches, respectively. They evaluated the proposed method on 80 remaining subjects of SCFace dataset [5] and achieved 73.3%, 93.5%, and 98.00% Rank-1 IR for distance 1 (4.20m), distance 2 (2.60m), and distance 3 (1.0m), respectively. As it can be noticed from these results, performance of the proposed methods deteriorate significantly when the resolution of the probe faces decreases. In GenLR-Net [14] authors employed VGGFace [16] pre-trained model to construct two branches network to overcome performance degradation in LR face recognition. Their proposed method significantly improved the results of HR-LR verification task on modified fold 1 of LFW benchmark [9] from 69.16% using original VGGFace [16] model to 90.00%. There are also deep learning based super-resolution methods to deal with low resolution faces, however, these methods are not optimized for LRFR [26] and yield modest performance improvement [22].

3. Methodology

In the following sections, we present the building blocks of the system, which are employed face detector [27], utilized training databases [6, 1] and the deep CNN models [1, 3], proposed strategy to match the resolution of the gallery and probe images, the crop ratios to adjust the amount of information included in the face images, and finally the similarity measurement and the evaluation metric.

3.1. Face Detection

The bounding boxes of the faces in the gallery and probe sets are detected using the Multi-Task Cascaded Convolutional Neural Networks (MTCNN) [27] model. The faces are cropped and resized to 224×224 or 112×112 pixel resolutions depending on the input size of the deep learning models.

3.2. Feature Extraction

We employ four state-of-the-art deep CNNs, namely ResNet-50 [7], SENet-50 [8], LResNet50E-IR [3], and LResNet100E-IR [3]. The deep models are trained or fine-tuned on VGGFace2 [1] and MS-Celeb-1M databases [6] to learn the face embedding of the gallery and probe face images in the SCFace [5] and ICB-RW [15] benchmarks. Please note that we do not take advantage of these benchmarks for fine-tuning.

3.2.1 Deep face models

The deep face models that are utilized in this study are listed in Table 1 and named as model *a, b, c, ..., h*. Off-the-shelf models described in VGGFace2 [1] and ArcFace [3] are used for models *a, b, c, d* and *e, g*, respectively. Furthermore, models *e* and *g* are fine-tuned on the VGGFace2 database [1] to learn models *f* and *h*, respectively.

3.2.2 Fine-tuning

The detected face images of the VGGFace2 database [1] are aligned with respect to the positions of the center of the eyes, tip of the nose, and the corners of the mouth. The aligned faces are then resized to 112×112 pixel resolution and finally pixel intensity values are normalized by subtracting 127.5 and dividing by 128. These pre-processed face images are then provided for fine-tuning.

Model f: model *e* is fine-tuned on the VGGFace2 database [1] using additive angular margin loss [3] with $m = 0.5$ and $s = 64.0$. Stochastic gradient descent with momentum 0.9 and learning rate of 0.01 are used to fine-tune the network with the batch size of 64. The learning

Models	CNNs	Trained on	Fine-tuned on	Input size	Embedding size
a	ResNet-50 [7]	VGGFace2 [1]	n/a	224×224	2048
b	ResNet-50 [7]	MS-Celeb-1M [6]	VGGFace2 [1]	224×224	2048
c	SENet-50 [8]	VGGFace2 [1]	n/a	224×224	2048
d	SENet-50 [8]	MS-Celeb-1M [6]	VGGFace2 [1]	224×224	2048
e	LResNet50E-IR [3]	MS-Celeb-1M [6]	n/a	112×112	512
f	LResNet50E-IR [3]	MS-Celeb-1M [6]	VGGFace2 [1]	112×112	512
g	LResNet100E-IR [3]	MS-Celeb-1M [6]	n/a	112×112	512
h	LResNet100E-IR [3]	MS-Celeb-1M [6]	VGGFace2 [1]	112×112	512

Table 1. The eight combinations resulting from the different deep CNN architectures and training databases that are used for feature extraction in this study.

rate is divided by 10 at 20K, 28K iterations and the training process is stopped at 32K iterations as in ArcFace [3]. The obtained verification accuracy of the validation set, LFW dataset [9], is 99.6%.

Model h: model *g* is fine-tuned on the VGGFace2 database [1] with the same setting as in the model *f*, however, the learning rate is set to 0.001. The achieved verification accuracy on the LFW dataset [9] is 99.7%.

3.3. Amount of information

To adjust the amount of information to be included in the face images, we extend the face bonding boxes. In a previous work [13], it has been shown that this has a significant effect on the performance. In our study, we also expect this adjustment to contribute positively to the performance of LRFR due to two main reasons. The first reason is that due to low resolution, the face images contain limited information, extending face bounding boxes would allow to include more information, for example about the shape of the face, etc. The second one is related to the upsampling factor. Since input size of the face images to the deep learning models are relatively high, in our case 224×224 or 112×112 pixels, this requires upsampling of the low resolution face images with a large scaling factor. A larger crop of the face region would decrease the scaling factor, thus, less degradation would occur due to upsampling. In this work, we control the amount of information to be included in the face images with six different crop ratios (1.0, 1.1, 1.2, 1.3, 1.35, 1.40) as shown in Figure 3.

3.4. Matching the resolution

An important challenge in LRFR is that features extracted from very low resolution faces in the probe set and high resolution images in the gallery set can potentially have higher intra-class distance than inter-class distance. We hypothesize that if we could make the appearance of the gallery face images similar to the probe face images, intuitively, we would minimize the intra-class distance. Therefore, to imitate low resolution we downsampled the gallery images. That is the gallery face images are downsampled

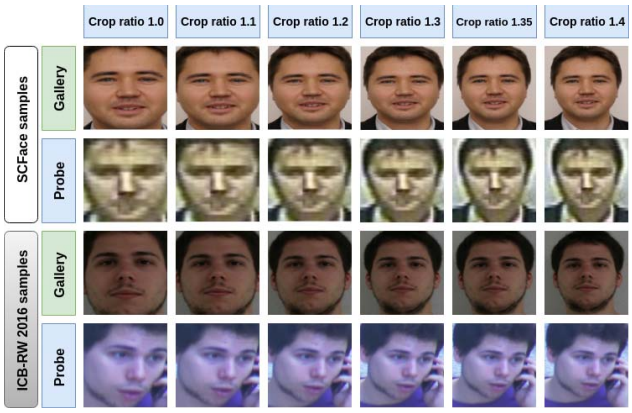

Figure 3. Gallery and probe faces of a subject from SCFace and ICB-RW benchmarks cropped with six different crop ratios.

Figure 4. Gallery faces of a subject from SCFace benchmark cropped with 1.3 extension factor and matched resolution of them with five different pixel resolutions (24×24, 32×32, 40×40, 48×48, 64×64) are shown here.

and this way their resolution is matched with the resolution of probe face images. For this purpose we picked five different resolutions (24×24, 32×32, 40×40, 48×48, 64×64) and select 32×32, 48×48, and 64×64, which are closest to the resolution of d1, d2, and d3 probe face images in SCFace [5], respectively. We take the original resolution of gallery face images in ICB-RW [15] experiments, which matches the resolution of the probe face images. In Figure 4, the first column shows the gallery face of a subject from SCFace [5] cropped with 1.3 extension ratio, whereas, the other columns show downsampled gallery face images at five different resolutions to make their image quality similar to the probe face images in the SCFace [5] benchmark.

3.5. Face Identification

The face embedding of the gallery and probe sets are extracted using eight deep face models described in Table 1. The identification task for probe faces are carried out by nearest neighbor classification method with the correlation distance metric (eq. 1) as similarity measurement:

$$Corr.distance(u,v) = 1 - \frac{(u-\bar{u}).(v-\bar{v})}{\|(u-\bar{u})\|_2\|(v-\bar{v})\|_2} \quad (1)$$

where *u, v* are the face feature vectors and $\bar{u}, \bar{v}$ are mean of the face feature vectors. Rank-1 IR is reported as the evaluation metric.

4. Experimental Results

We conduct our experiments in three steps on the SCFace and ICB-RW benchmarks. Firstly, we crop faces with bounding boxes detected by MTCNN [27] before feature extraction. Secondly, larger crops are used for feature extraction, and finally, the gallery faces' pixel resolution are matched with the resolution of probe face images before extracting the face embedding. In this section, we provide the experimental results for these steps.

4.1. Datasets

We evaluate the proposed methods on the SCFace [5] and ICB-RW [15] benchmarks.

There are 130 subjects in SCFace dataset [5], one frontal image (gallery set) and 15 LR images per subject (probe set). The gallery faces are captured in controlled conditions, whereas, the probe faces are captured with five indoor surveillance cameras located at three different distances, d1, d2, and d3 (4.20, 2.60, and 1.00 meters, respectively) resulting in the probe images with varying image quality. Please note that in this study we do not fine-tuned our models with target dataset and we report the Rank-1 IR for 130 subjects of SCFace [5]. However, in order to be able to compare our results with previous works, we report the mean and standard deviation of Rank-1 IR of 10 RRSSV experiments for 80 subjects out of 130 in model *h** (Table 4).

ICB-RW benchmark [15] contains 90 subjects, each having one high quality gallery image and 5 probe images, recorded outdoors, containing variations in illumination, expression, pose, motion-blur, occlusion, and focus. Figure 2 illustrates the aforementioned probe image quality problems in SCFace [5] and ICB-RW [15] benchmarks.

4.2. Baseline experiments

The faces are detected using the MTCNN [27] and cropped according to the face detection output. The face embeddings are extracted with eight deep CNN models, as presented in Table 1. Thereupon, face embeddings are fed into the nearest neighbor classifier with correlation distance metric as the similarity measurement. The Rank-1 IR results on the SCFace [5] and ICB-RW [15] benchmarks are reported in Table 2. It can be seen from the results that the performance of the state-of-the-art deep CNN models plummet at d1, which contains very low resolution probe face images. We fine-tune models *e* and *g* using VGGFace2 database [1] to learn models *f* and *h*, respectively. After that, a significant improvement in performance of models *f* and *h* for d1 of SCFace [5] (see Table 2) are observed. The improvement can be described to the fact that the VGGFace2 database [1] contains approximately 20% of the face images with pixel resolution lower than 50 pixel, which allow the model to learn better feature representation for low resolution face images.

	SCFace			ICB-IRW
Model	d1	d2	d3	probe
a	40.15	**91.38**	**98.15**	79.11
b	**41.85**	89.54	97.69	77.56
c	33.08	86.92	96.62	**81.33**
d	35.69	86.00	97.23	79.56
e	13.85	59.54	86.31	40.44
f	20.46	71.54	85.38	48.00
g	25.38	84.00	**98.15**	68.22
h	37.54	87.69	96.00	69.33

Table 2. The Rank-1 IR results (%) of eight deep models are reported for d1 (4.2 m), d2 (2.6 m), and d3 (1.0 m) probe faces of SCFace and ICB-RW in which we detected the faces with MTCNN model and cropped them with 1.0 ratio.

	SCFace			ICB-IRW
Model	d1	d2	d3	probe
a	53.54	94.92	99.38	80.67
b	52.15	93.85	98.00	81.56
c	49.08	93.54	98.92	82.00
d	50.77	94.00	99.08	**82.67**
e	23.23	77.54	93.23	58.22
f	47.69	87.23	93.38	60.00
g	50.46	**96.31**	**99.69**	82.00
h	**60.62**	96.15	99.38	78.67

Table 3. The Rank-1 IR (%) of deep CNN models using 1.30 crop ratio are reported for d1, d2, and d3 in SCFace and probe faces of ICB-RW.

4.3. Effect of increasing the amount of information

As we discussed in section 3.3, we control the amount of information to be included in the gallery and probe face images by using six different crop ratios. Empirical results show a compelling improvement on the performance of eight deep CNN models. We plot the Rank-1 IR of deep

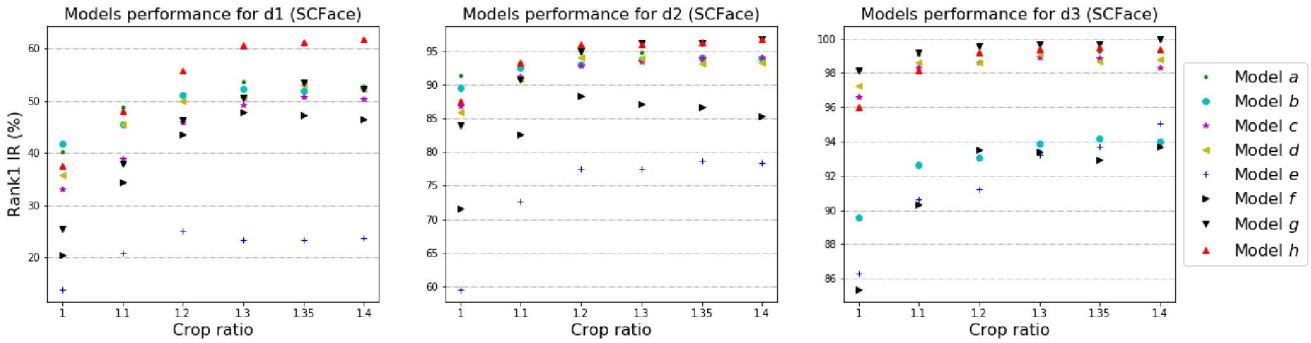

Figure 5. The Rank-1 IR (%) of deep CNN models on probe faces of SCFace benchmark for six different crop ratios.

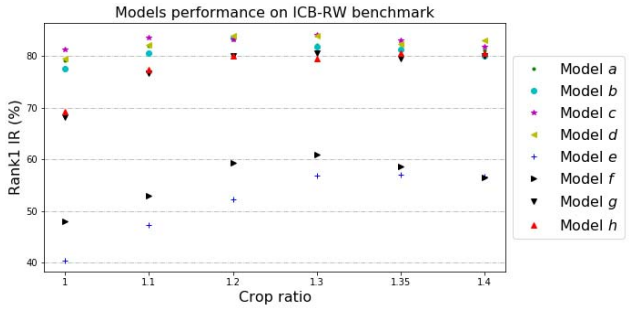

Figure 6. The Rank-1 IR (%) of deep CNN models on probe faces of ICB-RW benchmark for six different crop ratios.

	SCFace			ICB-RW
Model	d1	d2	d3	probe
a	56.72	95.23	99.23	82.22
b	59.38	96.00	98.00	82.00
c	54.15	94.77	98.92	**84.22**
d	60.62	94.46	99.23	84.00
e	33.38	80.62	95.23	58.67
f	55.38	89.69	93.85	60.89
g	67.08	97.23	**100**	81.78
h	75.08	97.69	99.69	79.78
*h**	**78.5**	**98.38**	99.75	n/a
DCR [12]	73.3	93.5	98.0	n/a
LDMDS [24]	62.7	70.7	65.5	n/a
PCLWT [21]	64.76	80.8	74.92	n/a
Ghaleb et al. [4]	n/a	n/a	n/a	71.7

Table 4. The results achieved with 1.3 crop ratio are reported for DCNN models. * denotes that model *h** results are mean of 10 RRSSV for 80 subjects out of 130 in SCFace [5]. The presented mean face identification rates for d1, d2, and d3 have 1.67, 0.48, and 0.16 standard deviation, respectively.

CNN models for each of six crop ratios as illustrated in Figure 5 for SCFace [5], and Figure 6 for ICB-RW [15] benchmarks. Table 3 summarizes the Rank-1 IR results achieved by 1.30 crop ratios for the eight deep models. These results show the impact of the increased information in the significant improvement of the models' performance, especially, for the probe face images that have lower resolution. Our results also validate the results in [13], which presented the performance improvement in face recognition using extended bounding boxes.

4.4. Effect of matching the resolution

As it is mentioned in section 3.4, we conduct experiments on SCFace [5] and ICB-RW [15] benchmarks using eight deep CNN models to test the contribution of matching the resolution at performance improvement. We observe that Rank-1 IR improves significantly for the low resolution probe faces as in SCFace [5], however, there is not much improvement in the higher resolutions probe images as in ICB-RW [15] which already have a matching resolution with the gallery face images. Table 4 shows the Rank-1 IR achieved by DCNN models on SCFace [5] and ICB-RW benchmark [15]. The models with 224×224 input size (*a*, *b*, *c*, *d*) achieve higher Rank-1 IR for ICB-RW benchmark [15], which can be described to the fact that the probe images of ICB-RW [15] have higher resolution. The presented Rank-1 IR on SCFace benchmark [5] are achieved with 32×32, 48×48, and 64×64 downsampled gallery face images which are close to the average resolution of d1, d2, and d3 in SCFace benchmark [5], respectively. Please note that in DCR [12] and LDMDS [24] randomly selected 50 subjects out of 130 subjects in SCFace [5] are used for fine-tuning and the results are reported on 80 remaining subjects. To compare our results we also report the mean and standard deviation of 10 RRSSV experiments on 80 randomly selected subjects (model *h**). As can be seen from Table 4, on d1 and d2 subsets around 5% and on d3 subset 2% absolute performance improvement has been achieved compared to the DCR [12] leading to the state-of-the-art results for the SCFace dataset [5]. Similarly, the proposed approach enhances the state-of-the-art accuracy on the ICB-RW benchmark [15] from 71.7% to 84.22%.

5. Conclusion

In this paper, we explore the factors that would contribute to improve identification accuracy of low resolution face recognition under mismatched conditions. We observe that models f and h fine-tuned on the VGGFace2 dataset significantly improve Rank-1 IR for very low resolution probe face images (d1 of SCFace) compared to off-the-shelf models (models e and g), which are trained on MS-Celeb-1M dataset [6]. This can be explained to the fact that VGGFace2 [1] has about 20% of the face images with resolution lower than 50 pixels, which helps the model to learn robust features for low resolution faces. The experimental results show that including more information in the cropped faces and matching the resolution between gallery and probe sets enhance the Rank-1 IR significantly. Our model h achieves state-of-the-art Rank-1 IR results on 130 subjects of SCFace benchmark [5] which are 75.08%, 97.69%, and 99.69% Rank-1 IR for d1, d2, and d3 respectively. We also significantly improve the Rank-1 IR on ICB-RW benchmark with model c that achieves 84.22% Rank-1 IR outperforming the validation results reported in Ghaleb et al [4] by 12.52 margin.

References

[1] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *International Conference on Automatic Face & Gesture Recognition*, pages 67–74, 2018. 1, 2, 3, 4, 5, 7

[2] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler. Robust face recognition for uncontrolled pose and illumination changes. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43(1):149–163, 2013. 3

[3] J. Deng, J. Guo, X. Niannan, and S. Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *Conference on Computer Vision and Pattern Recognition*, 2019. 2, 3, 4

[4] E. Ghaleb, G. Ozbulak, H. Gao, and H. K. Ekenel. Deep representation and score normalization for face recognition under mismatched conditions. *IEEE Intelligent Systems*, 33(3):43–46, 2018. 2, 6, 7

[5] M. Grgic, K. Delac, and S. Grgic. SCface – surveillance cameras face database. *Multimedia Tools and Applications*, 51(3):863–879, 2011. 1, 2, 3, 4, 5, 6, 7

[6] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao. Ms-Celeb-1M: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision*, pages 87–102, 2016. 1, 2, 3, 4, 7

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 1, 2, 3, 4

[8] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018. 1, 2, 3, 4

[9] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *European Conference on Computer Vision Workshop on faces in Real-Life Images: Detection, Alignment, and Recognition*, 2008. 1, 2, 3, 4

[10] S. H. Lee, J. Y. Choi, Y. M. Ro, and K. N. Plataniotis. Local color vector binary patterns from multichannel face images for face recognition. *IEEE Transactions on Image Processing*, 21(4):2347–2353, 2012. 3

[11] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. SphereFace: Deep hypersphere embedding for face recognition. In *Conference on Computer Vision and Pattern Recognition*, pages 212–220, 2017. 2

[12] Z. Lu, X. Jiang, and A. C. Kot. Deep coupled resnet for low-resolution face recognition. *IEEE Signal Processing Letters*, 25(4):526–530, 2018. 2, 3, 6

[13] M. Mehdipour Ghazi and H. Kemal Ekenel. A comprehensive analysis of deep learning based representation for face recognition. In *Conference on Computer Vision and Pattern Recognition Workshop on Biometrics*, pages 34–41, 2016. 4, 6

[14] S. P. Mudunuri, S. Sanyal, and S. Biswas. GenLR-Net: Deep framework for very low resolution face and object recognition with generalization to unseen categories. In *Conference on Computer Vision and Pattern Recognition Workshop on Biometrics*, pages 602–60209, 2018. 3

[15] J. Neves and H. Proença. ICB-RW 2016: International challenge on biometric recognition in the wild. In *International Conference on Biometrics*, pages 1–6, 2016. 1, 2, 3, 4, 5, 6

[16] O. M. Parkhi, A. Vedaldi, A. Zisserman, et al. Deep face recognition. In *British Machine Vision Conference*, volume 1, pages 41.1–41.12, 2015. 3

[17] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015. 2

[18] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015. 1, 2

[19] Y. Sun, D. Liang, X. Wang, and X. Tang. DeepID3: Face recognition with very deep neural networks. *CoRR*, abs/1502.00873, 2015. 2

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015. 1, 2

[21] M. Uzun-Per and M. Gökmen. Face recognition with patch-based local walsh transform. *Signal Processing: Image Communication*, 61:85–96, 2018. 3, 6

[22] Z. Wang, S. Chang, Y. Yang, D. Liu, and T. S. Huang. Studying very low resolution recognition using deep networks. In *Conference on Computer Vision and Pattern Recognition*, pages 4792–4800, 2016. 2, 3

[23] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In *Conference on Computer Vision and Pattern Recognition*, pages 529–534, 2011. 1, 2, 3

[24] F. Yang, W. Yang, R. Gao, and Q. Liao. Discriminative multidimensional scaling for low-resolution face recogni-

tion. *IEEE Signal Processing Letters*, 25(3):388–392, 2018. 3, 6

[25] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *CoRR*, abs/1411.7923, 2014. 1, 3

[26] X. Yu, B. Fernando, R. Hartley, and F. Porikli. Super-resolving very low-resolution face images with supplementary attributes. In *Conference on Computer Vision and Pattern Recognition*, pages 908–917, 2018. 3

[27] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016. 3, 5