Quality measures in Face Biometrics

Jingwen Yang, Peter Szabo

Abstract— **In this paper, we present a Quality Assessment approach of predicting face images' performance in face recognition task, which is called FaceQnet. The FaceQnet model is a Convolutional Neural Network fine-tuned on the ResNet-50 architecture. The model can be used to obtain a prediction of the suitability of a specific image for face recognition, i.e. a performance-based quality measure. The training process was done using the SCFace and BioSecure database.**

First we used the BioLab-ICAO framework for labeling the SCFace and BioSecure training images with quality information related to their ICAO compliance level. Then, we obtained groundtruth quality measures, for each subject, by calculating the L2 distances between the best quality image(gallery image) chosen by the BioLab-ICAO framework, and the rest of the images(probe images) of the same subject. The generated verification scores are the quality measure for each training image. Finally, we used the groundtruth data for fine-tuning the FaceQnet model.

We tested our model against 2 different test databases: VGGFace2 and LFW. We perform face verification comparing each test image with all the images belonging to the same subject. Next, we compute the metrics describing the correlation between the quality measures and the verification accuracy. We checked the accuracy of our quality measure for predicting the performance in face recognition. As a result, we show that the predictions from FaceQnet are highly correlated with the face recognition accuracy.

FaceQnet is the newly proposed method in the literature based on deep learning massively scalable without human intervention. Also, it could possibly be used on different domain besides face recognition.

I. INTRODUCTION

Biometric reconstruction is a tool of recognizing the identity of a person based on its physical or behavioural attributes associated to them. These attributes can be fingerprint, iris, face, signature or the fusion of them. Among the biometric traits, fingerprint, iris and retina are proved to be the most accurate ones. However, they are harder to be acquired. Facial recognition on the other hand, provides easier acquisition and is widely acceptable among user.

In the recent years, we have witnessed the explosion in terms of computational hardware capability. Also, algorithms became more complex and sophisticated especially with the birth of convolutional neural network. Due to these factors, the accuracy of the facial recognition system improved remarkably. However, they are still behind the more accurate biometric traits. Good accuracy can only be achieved under well-controlled environments, with the collaboration of the user, and hardly applicable in unconstrained conditions such as bad visual conditions, presence of occlusion, illumination, inadequate resolution and blurred image. In addition, the face itself shows broad range of variability in terms of different facial expressions, facial hair, different hairstyles or glasses.

So that facial recognition system could be improved, and possibly overpass the performance of fingerprint biometrics if these challenges can be overcame. They all prevail an effect on the quality of the images, which brings us the desire to a method which is capable to deal with these effects.

Such a tool could have an essential role in many field of application. It could be used to enhance the accuracy of facial recognition based security system widely used both public or private sector. Such systems play a crucial role at border control, at criminology, or identification. In addition, that tool could be used, to predict the accuracy of the facial recognition system in advance, enhancing the reduction of false acceptance or false rejection or any kind of recognition error, that comes from the bad quality of the image. In cases, where this tool perceives images with precarious quality, it could alarm the system to ask for a new acquisition.

Numerous works endeavours to solves this issue, and in this work we will present, and implement the work of [20]. In their work they present a novel open-source face quality assessment tool, based on deep learning, that aims to predict the quality of facial images.

The structure of the report will be organized as the following: Section 2 will introduce the tools, that were used during our work, also the data sets that were crucial both for training our model and evaluating it. Section 3 will present the result of [20], which we aim to achieve on a different data set. Section 4 will describe the experimental schemes and our solution towards out goal. In Section 5 we present and evaluate our results, and in the VI. Sections we draw our conclusions.

II. PRELIMINARIES AND TOOLS

In this section we define both the theoretical and practical tools we used in our experiment. In the theoretical section we clarify the most important concepts related to biometric facial recognition.

A. Definitions

In the quality assessment task, our goal is to infer the biometric quality of the given image. Biometric quality is a measure, which describes the quality of a given biometric trait of a given input. The quality is related to *utility* of the trait. It refers to the condition of the trait, how much accuracy we may expect from it. *Character* determines the expected distinctiveness, and *fidelity* refers to the trustworthiness compared to the real sample.

The following definitions are crucial to be specified in order to be able to fully understand the quality measures for face recognition:

- **Groudtruth:** Two type of ground truth can be differentiated: the human perceived, and performance based. The former defines the human based approach, the latter specifies the expected performance of the biometric trait in the recognition task
- **Input types** can differentiated based on how much initial information they require. This spectrum streches from Full-Reference, requiring a good image library, though Reduced-Reference, needing only partial information, towards No-Reference system, requiring no initial information.
- **Features:** hand-crafted or deep learning features could be extracted.
- **Output:** the systems can produce different outputs, based on the task they perform. In a regression problem they produce a real, numeric vale, in a classification can produce discrete values or in verification only a binary decision.

In [20], the face recognition is based on deep learning, which predicts a performance based numerical ground truth without any reference image. In this report we have the same ambition.

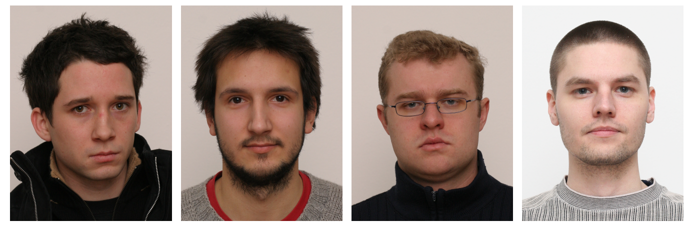
B. Tools

During our work we were working with python, using tensorflow framework for deep learning.
For the automatic generation of the ICAO compliance scores we used the BioLab framework, described in [6]. ICAO scores are face quality standards, **BioLab** uses the most relevant: ISO/IEC 19794-5. It provides 23 test for us among the position of the eye.

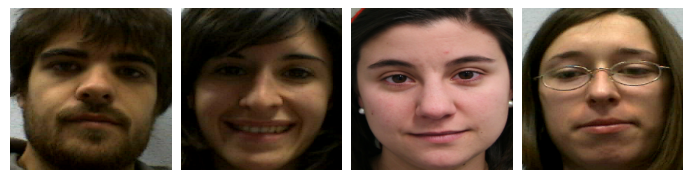
FaceQNet is a modified version of **ResNet50** described in [12]. ResNet50 is a deep convolutional neural network, with 50 layers. Residual means, that it has skip connections among the layers. It allows to learn identity function, which enables the building of deeper neural network, offsetting the vanishing gradient (through back-propagation the gradient would be too small in deeper networks). The ResnNet50 weights were trained on ImageNet, and it is able to classify images into 1000 objects. It is beleived that the network has learned a rich feature representation for a wide range of images.

C. Dataset

In this section we describe the characteristics of the databases used in the experiments of the paper: SCFace, Biosecure,VGGFace2, and LFW. A sample of images with different qualities from each database can be seen in FigureII-C.

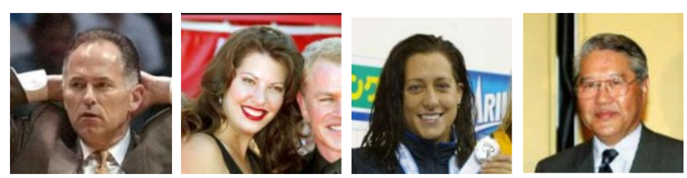


(a) Tuning set 1: SCFace



(b) Tuning set 2: BioSecure

(c) Testing set 1: VGGFACE2



(d) Testing set 2: LFW

Fig. 1. **A few samples of Datasets**

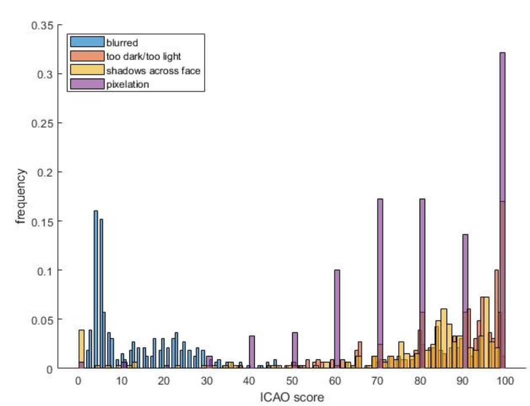


Fig. 2. Quality of images in SCFace dataset based on a few selected features ICAO features

1) SCFace Database: In this work, we use Surveillance Cameras Face Database(SCFace)[3] for fine-tuning our qual-

ity assessment network FaceQnet. SCface is a database of static images containing 4160 static images (in visible and infrared spectrum) of 130 subjects. Images were taken in uncontrolled indoor environment using five video surveillance cameras of various qualities and one IR mug shots camera. However, in the naming system, the cam6 indicates the cam1 in night vision mode, the cam7 indicates the cam5 in night vision mode and the cam8 indicates the camera for taking IR mug shots. All of them are fixed to same positions and were not moved during the whole capturing process. For every subject, 21 images are respectively taken by cam1-cam7 at distances of 4.20m, 2.60m and 1.00m; then 9 images are taken from different angles with equal steps of 22.5 degrees and total range of 180 degrees; in the end, two images will be taken by IR frontal mug shot and visible light mug shot. Therefore, overall 32 images are taken for each subject, among which the visible light mug shot ensures there is at least a good quality picture can be used as the gallery image. The other images from different quality cameras mimic the real-world conditions, which is very helpful in the case of fine-tuning of our quality assessment network.
Figure II-C clearly visualises with the separate peaks of pixelation bins, the different distances the images were made. The higher the distance is, the smaller the resolution of the image will be. In addition, we can see, that the images were blurry, but apart from that feature, in terms of the rest of the features the database shows a have a decent quality.

(a) Images with their quality scores (First two on the left, are from SCFace, the two on the right are from BioSecure)

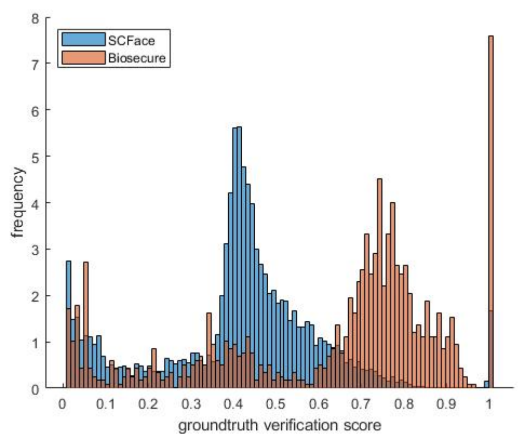


Fig. 3. Histogram of the quality scores of SCFace and BioSecure dataset, after the normalization and filtering of low quality images.

2) BioSecure Database: In the present work we have used BioSecure Multimodal Database[2] together with the SCFace database for fine-tuning purpose. It consists of around 1 000 subjects whose biometric samples were acquired in three different scenarios. Images for the first scenario were obtained remotely using a webcam, the second is a more controlled mugshot-type scenario using a high quality camera with homogeneus background, and the third scenario is uncontrolled, captured with mobile cameras both indoors and outdoors. We employed around 1500 images of 145 subjects from the second and third scenarios for obtaining quality measures with FaceQnet. Since this dataset is not as clear as the SCFace dataset that we can easily distinguish the best quality picture from the rest, it needs to be processed by ICAO framework to get a label of quality score, which will be later fed to FaceQnet to learn.

3) VGGFace2 Database: In this work we use a subset of the VGGFace2[18] to evaluate our quality metric. The full database contains 3.31 million images of 9131 different identities, with an average of 362.6 images for each subject. All the images in the database were obtained from Google Images, and they correspond to well known celebrities such as actors/actresses, politicians, etc. The images have been acquired under unconstrained conditions and present large variations in pose, age, illumination, etc. These variations mean different levels of quality. Furthermore, the pre-trained model we used for the quality assessment network ResNet-50[12] is trained on VGGFace2 database. It is shown that ResNet-50 has great performance in challenging face recognition benchmarks such as IJB-C[19], QUIS-CAMPI[11] OR PaSC[7].

4) LFW Database: The Labeled Faces in the Wild (LFW)[1] database has been also employed to test the performance of the FaceQnet. The database consists of 13233 images of 5749 different subjects, having 1680 of them two or more different images. The images are labeled with the identity of each correspondent subject. This database has been widely used in the recent years for studying face recognition under unconstrained conditions, and having a performance-based quality measure for each image can help to boost the accuracy of the state-of-the-art face recognition systems that use this dataset for their benchmarks.

III. Related work

There are several face quality standards have been proposed so far, being the most relevant and extended ones the ICAO 9303 and the ISO/IEC 19794-5. They includes a series of guidelines for the acquisition of good quality images. Researchers have developed relative systems to automatically evaluate if an image satisfies the guidelines given in these standards.The systems output either a binary vector or a vector with scores indicating if the image pass/not-pass the standard or its numerical scores of fulfilling certain standards.
In TableIII we include a compilation of some relevant related works in quality assessment for face recognition. Algorithms are classified according to the different characteristics.
In this work,in the groundtruth generation step, we made use of the BioLab-ICAO framework[6], an evaluation tool for automatic ICAO compliance checking, to choose a best quality image as the gallery image for the BioSecure dataset. Since the SCFace dataset already has the best quality image, it doesn't need to be processed by BioLab-ICAO framework.

BioLab-ICAO framework performs 30 different individual tests for each input image, and gives an output consists of a numerical score for each test, going from 0-100. Those 30 individual scores are not necessarily integrated into a final unified quality metric.
The use of Convolutional neural network in quality assessment for face recognition started in 2017. A pretrained CNN based on VGGFace is employed to extract features from the images.Then those features have been used to train other classifiers, which means that they successfully transferred learning from face recognition to the quality prediction task. They have the common limitations of: 1) a high amount of human effort is required to label the database with human perceived quality; 2) a manual selection of a high quality image is needed for each subject to obtain the machine accuracy prediction, which also involves human effort and human bias. The improvement in recent years' study in this area, and also in our present work, is a step forward in making fully automatic the generation of groundtruth quality labels.
In this work, we are basically reproducing the results they got in the paper [20]. The difference is we are fine-tuning the ResNet-50 on SCFace and BioSecure while they are training on VGGFace2. The test datasets for us are a subset of VGGFace2 and LFW while they tested against a subset of VGGFace that not overlapping with the training set. For the evaluation part, we didn't split into three quality subsets, which we may do further verification in the next semester.

IV. Proposed approach

This chapter contains the experimental schemes, and our solutions. It consists of two parts. In the first part, we describe the generation of ground truth, using SCFace and BioSecure databases. This results a ground truth data set, containing images with their quality score. This dataset is used in the next part.
In the second part, the construction and the training of FaceQNet is detailed.

A. Construction of training database

In this chapter we introduce the way we created our ground truth dataset, later used for training the FaceQNet.
Our training database consisted of 2 parts: SCFace and BioSecure. In SCFace dataset, the highest quality images were available, separated from the lower quality ones. On the other hand in the case of BioSecure database, they were not provided. In order get them, we utilized BioLab ICAO framework. The framework provided us 23 tests, representing different features for each images. Based on intuition we decided to consider the: blur, visual conditions, the pixelation, the variability of the background, the roll/pitch/yaw size (should be greater than a predefined threshold), the presence of shadow across the face, the presence of hat cap or vail, or other toys too close to the face. Inspecting those features we choose the best quality for each user (each user, that had at least one image hitting a predefined threshold).
In order to reduce the variability as a source of error all images were cropped and aligned.
To make the network learn to tell the quality of an image is good or not, we need to feed it with both good images and bad images, which are, the gallery and the probe.
In the present work we made the assumption that a perfectly compliant ICAO image represents perfect quality. Therefore, if we match such a perfect ICAO image to another image from the same subject and the comparison score is low, it is safe to assume that the second image is of low quality. This way, by comparing an image B with a perfect ICAO gallery image A of the same subject, we can use the resulting comparison score as the groundtruth quality measure for the image B, and then use both image B and its groundtruth quality to train FaceQnet.
To achieve this target, we need an ICAO compliance score for each image in the training database. We will use these scores for selecting the gallery images, i.e. the ones with the highest ICAO score. To obtain those ICAO compliance scores we used the BioLab framework from [6]. This framework outputs a score between 0 and 100 for each individual ICAO compliance test. In this work a subset of tests are chosen as shwon in table IV-A. As one of the training set for our quality assessment metric, we selected a subset of 145 subjects from the bioSecure database. For each user we selected as gallery image the one that obtained the highest ICAO compliance score. This way, the training set was divided into gallery images (the one with the highest ICAO compliance score for each subject), and probe images.
After getting the gallery image set and probe image set, an existing CNN pretrained for face recognition, the FaceNet model from [16], is employed as a feature extractor to get embeddings for all the images in the database. FaceNet was developed in Keras with Tensorflow as its backend. This CNN was trained with the CASIA-WebFace database. It has shown to obtain high accuracy levels in face recognition.

Initially, 128-dimensional feature vectors are extracted from the last fully-connected layer of FaceNet. Calculation the euclidian distance between between the embedding of the hight quality image and the rest of the image we get a dissimilarity measure between the users. This measure will be normalized with the normalization function: 2/(1+exp(x)), where x is the distance between the embeddings, this function projects the distances into the [0,1] range. After this process we remove the worse quality images. This will result N-1 scores for each users, assuming we have N image for a given user. Lower value represents lower quality compared (higher distance from be high ICAO quality image), and value closer to one stands for better quality.
These N-1 scores will be mated with their corresponding image and they will consturct the training dataset. As we can see on Figure 3 the quality of SCFcae is lower, since it contains many low resolution images (images that were captured from higher distance).

TABLE I

SUMMARY OF QUALITY ASSESSMENT WORKS FOR FACE RECOGNITION, CLASSIFIED BY: 1) THE GROUNDTRUTH GENERATION PROCESS; 2) THE TYPE OF INPUT; 3) THE FEATURES EXTRACTED; 4) THE TYPE OF OUTPUT PRODUCED

Ref	Year	Groundtruth Definition	Type of Input	Features Extracted	Output
[4]	2011	Human-based	No-Reference	Face features,image features	Score:presence of each factor
[5]	2012	Performance-based	Reduced-Reference	Contrast, brightness, focus, sharpness and illumination	FQI(Face Quality Index):0-1
[6]	2012	Human-based	No-Reference	20 ICAO compliance features	Score from each individual test
[8]	2013	Performance-based	Reduced-Reference	Image features, matcher features, sensor features	Low/high quality label
[9]	2014	Performance-based	No-Reference	Texture features	FQI(Face Quality Index):0-1
[10]	2015	Performance-based	No-Reference	2 face features: pose, illumination	Predicted FMR/FNMR
[13]	2016	Human-based	No-Reference	Image features	FQI(Face Quality Index):0-1
[16]	2017	Performance-based	No-Reference	Regression from images	Predicted verification rate
[15]	2017	Human-based	No-Reference	Image features	Numerical quality metric
[14]	2017	Human-based	Full-Reference	CNN features	Similarities between test and reference images
[17]	2018	Human-based&Performance-based	No-Reference	CNN features	MQV(Machine-based Q.), and HQV(Human-based Q.)
[20]	2019	Human-based&Performance-based	No-Reference	CNN features	Numerical quality metric:0-1

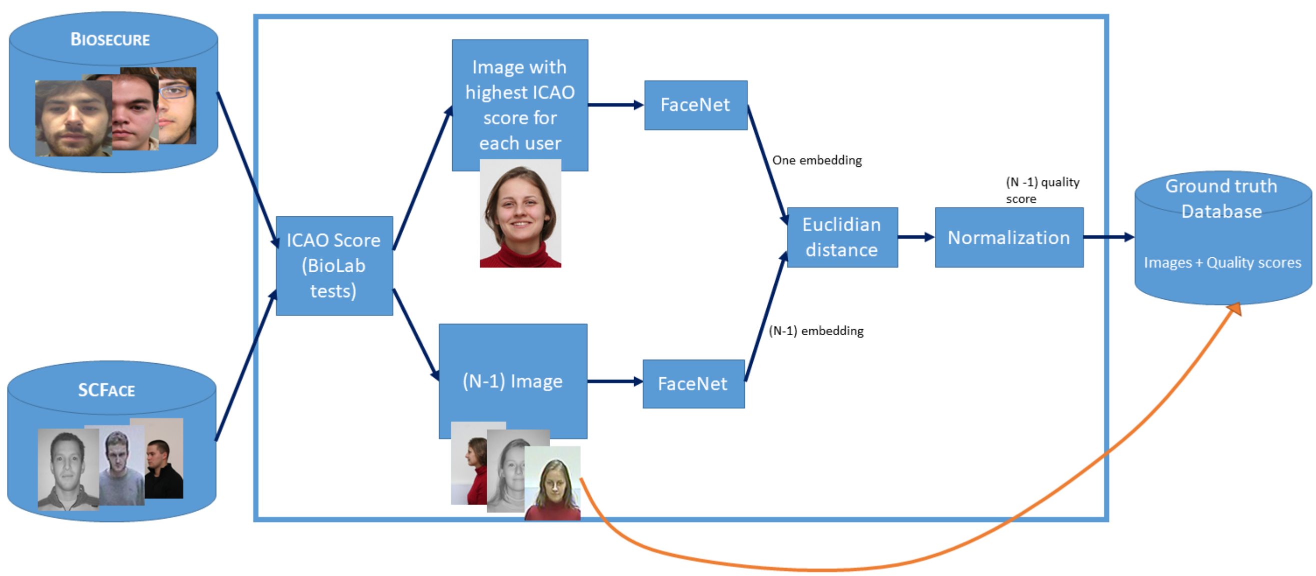


Fig. 4. **Generation of quality ground truth** we select 4292 images from SCFace, and 1452 images from BioSecure. Based on the ICAO compliance score, the image with the highest quality will be selected for each user. Utilizing FaceNet, we extract their embedding vector is compared with the corresponding FaceNet embedding of the rest of the images of the same user. This normalized euclidean distance will result the quality score. The quality scores paired with the related images will provide the training database for the FaceQNet.

TABLE II

TESTS DEFINED TO EVALUATE IMAGES FOR ICAO COMPLIANCE CHECK

Number	Description of the test
1	Blurred
2	Too dark/Light
3	Pixelation
4	Varied background
5	Roll/Pitch/Yaw greater than a predefined threshold
6	Shadows across face
7	Dark tinted lenses
8	Hat/Cap
9	Veil over face
10	Presence of other faces or toys too close to face

B. *Training of FaceQNet*

The goal of FaceQnet is to be able to predict the quality score of an image. To overcome the issue of the lack of sufficient data, we followed the transfer learning approach. We assumed, that he features, that are important in face recognition, are likely useful at image quality assessment, since information about identity contains information about the quality. It means, that via knowledge transfer it is possible to gain quality features from identity feature vectors. Based on that, we used ResNet50 with pretrained weights from [12]. This network was trained for face recognition, so according to our hypothesis, it is likely to be able to provide quality information thanks to its rich feature space acquired for facial recognition. The last layer of the ResNet is removed, and replaced with with our own Fully Connected (FC) layer, followed by the output. The FC layer combines the embedding of the body of the ResNet into a vector of 32, and the final layer performs the regression, the inference of quality score. The replacement of the last layer is explained by the assumption, that the earlier layer of a neural network extracts more general features, and the deeper we go, the more finer the features are, making only the last layers to be task specific. By replacing the last layer we can solve a similar task without modifying all the weights of the neural network. That is the reason why during the training the

pretrained weights of the ResNet50 were frozen, and only the final layers were learning. The inputs of the network are 224x224x3 sized images. The images are assumed to be cropped and aligned.
Once we have the trained network, we can use it as a black box. We give it the properly preprocessed image, and it produces a corresponding score between 0 and 1, which will be related of to approximated accuracy of facial recognition systems.

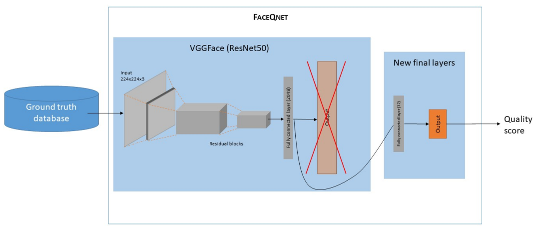


Fig. 5. **FaceQNet** The ground truth data set is used to trained the modified ResNet architecture. It consists of the body of the ResNet, and the replaced the upper Fully Connected layer and the output. ResNet body loads the pretrained weights of [12]. During training these weights are frozen, and only the weights of the last two layer are trained. After the successful training the network performs a regression task, outputting the qulity score of the given image.

V. Experimental study

This section we describe the method we use to evaluate the goodness of our results and using that technique we discuss our results.

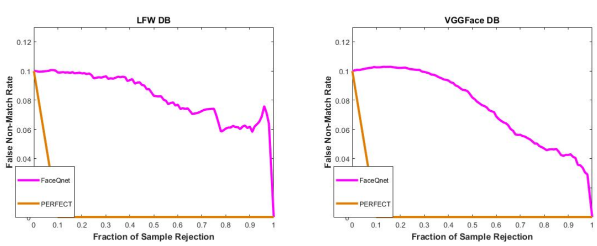


Fig. 6. **ERC curves** on LFW and VGGFace2 dataset using FaceNet comparator. The PERFECT sample was obtained using max(FNMR - Fracion of Sample Rejection),0). Our goal is to get as close as possible to the PERFECT line, meaning better accuracy.

A. Experiment description

During the evaluation we tested our FaceQNet on 2 different databases: Labeled Faces in the Wild (LFW), and VGGFace2 database. Their detailed description can be found in the II. Section, under C subsection. These databases were captured in real time, complex environment, having high variability among conditions, and containing challenging challenging scenarios.
In our experiments we calculated the Error vs Reject Curves compared with a compactor, called FaceNet. FaceNet outputs a value between [0,1], higher value meaning better score. This approach allowed us to analyse how well our version of FaceQNet correlated with other quality systems.
ERC plots are calculated by measuring the False Non Match Rate (FNMR), and iteratively discarding increasing percentage of the lowest quality measures, and checking the behaviour of FNMR. The ERC curves finally show the relationship between FNMR and rejection rates. Our expectation is, that as we discard the worse quality images, the FNMR is supposed to decrease, as only the better and more reliable images are remaining in the dataset. If our network work properly it is able to predict which are those low quality images that should be discarded first and his presents a correlation between the accuracy our quality measurement system and the accuracy of a face recognition system.

B. Discussion

As we can see on Figure 6, we can clearly see the correlation between our FaceQNet and the comparator. This can be seen via the decrease of the ERC curve. It is also worth to notice that the our QA system proves to be more accurate on VGGFace dataset, but it can be still applied in numerous scenarios.

VI. Conclusions

In this paper, we proposed a Quality Assessment (QA) system for face recognition. We clarified certain aspects that our QA system is taking into consideration when it assess an image's potential accuracy in face recognition task. Our solution consists in employing a CNN previously trained for face recognition(ResNet-50), and fine-tuning the last two layers with an automatically generated quality groundtruth, which is based on the comparison scores between many unconstrained images and a perfectly compliant ICAO quality measures image, for the SCFace and BioSecure databases.
Based on the results of our experiments, these quality assessment is shown to be a reliable estimation of the face recognition accuracy. It is supposed to provide a reference for choosing images with certain quality level so that it guarantees certain performance in a further face recognition task.
One of the possible further improvements can be to quantitize the correspondence between certain quality level we ensure on the input images and certain performance it can achieve in face recognition.
Besides that, the combination of the 10 features we chose in the ICAO Compliance Check in the process of groundtruth generation is debatable. Other combinations with different features can also be implemented and compared between each other. Furthermore, relying in ICAO conformance for this task is not the only way to obtain those high quality templates. Other data-driven and groundtruth-free quality definitions could be also investigated to avoid system dependence.
Further preprocessing of the database could be also helpful for obtaining better results, since the network has shown to be sensitive to outliers in the input data.
Last but not least, faceQnet has been developed specifically

to be applied to the face recognition task, but the methodology designed in this paper for building a fully automatic quality estimator based on deep learning could be adapted to other problems as well.

References

[1] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," 2008.

[2] J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M. R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J.-L. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, *et al.*, "The multiscenario multienvironment biosecure multimodal database (bmdb)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1097–1111, 2009.

[3] M. Grgic, K. Delac, and S. Grgic, "Scface–surveillance cameras face database," *Multimedia tools and applications*, vol. 51, no. 3, pp. 863–879, 2011.

[4] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *CVPR 2011 WORKSHOPS*, IEEE, 2011, pp. 74–81.

[5] A. Abaza, M. A. Harrison, and T. Bourlai, "Quality metrics for practical face recognition," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, IEEE, 2012, pp. 3103–3107.

[6] M. Ferrara, A. Franco, D. Maio, and D. Maltoni, "Face image conformance to iso/icao standards in machine readable travel documents," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 4, pp. 1204–1213, 2012.

[7] J. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer, *et al.*, "The challenge of face recognition from digital point-and-shoot cameras," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, IEEE, 2013, pp. 1–8.

[8] P. J. Phillips, J. R. Beveridge, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, S. Cheng, M. N. Teli, and H. Zhang, "On the existence of face quality measures," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, IEEE, 2013, pp. 1–8.

[9] R. Raghavendra, K. B. Raja, B. Yang, and C. Busch, "Automatic face quality assessment from video using gray level co-occurrence matrix: An empirical study on automatic border control system," in *2014 22nd International Conference on Pattern Recognition*, IEEE, 2014, pp. 438–443.

[10] A. Dutta, R. Veldhuis, and L. Spreeuwers, "Predicting face recognition performance using image quality," *arXiv preprint arXiv:1510.07119*, 2015.

[11] J. C. Neves, G. Santos, S. Filipe, E. Grancho, S. Barra, F. Narducci, and H. Proença, "Quis-campi: Extending in the wild biometric recognition to surveillance environments," in *International Conference on Image Analysis and Processing*, Springer, 2015, pp. 59–68.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[13] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Signal Processing: Image Communication*, vol. 40, pp. 1–15, 2016.

[14] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu, "Deepsim: Deep similarity for image quality assessment," *Neurocomputing*, vol. 257, pp. 104–114, 2017.

[15] Y. Liu, J. Yan, and W. Ouyang, "Quality aware network for set to set recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5790–5799.

[16] J. P. Phillips, N. A. Yates, R. J. Beveridge, and H. G. Givens, "Predicting face recognition performance in unconstrained environments," *CVPR Workshops*, pp. 557–565, 2017.

[17] L. Best-Rowden and A. K. Jain, "Learning face image quality from human assessments," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 12, pp. 3064–3077, 2018.

[18] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, IEEE, 2018, pp. 67–74.

[19] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney, *et al.*, "Iarpa janus benchmark-c: Face dataset and protocol," in *2018 International Conference on Biometrics (ICB)*, IEEE, 2018, pp. 158–165.

[20] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, and L. Beslay, "Faceqnet: Quality assessment for face recognition based on deep learning," *arXiv preprint arXiv:1904.01740*, 2019.