A. Brömme, C. Busch, N. Damer, A. Dantcheva, M. Gomez-Barrero, K. Raja,
C. Rathgeb, A. Sequeira, and A. Uhl (Eds.): BIOSIG 2021,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021

QualFace: Adapting Deep Learning Face Recognition for ID and Travel Documents with Quality Assessment

João Tremoço[1] Iurii Medvedev[2] Nuno Gonçalves[3]

Abstract:

Modern face recognition biometrics widely rely on deep neural networks that are usually trained on large collections of wild face images of celebrities. This choice of the data is related with its public availability in a situation when existing ID document compliant face image datasets (usually stored by national institutions) are hardly accessible due to continuously increasing privacy restrictions. However this may lead to a leak in performance in systems developed specifically for ID document compliant images. In this work we proposed a novel face recognition approach for mitigating that problem. To adapt deep face recognition network for document security purposes, we propose to regularise the training process with specific sample mining strategy which penalises the samples by their estimated quality, where the quality metric is proposed by our work and is related to the specific case of face images for ID documents. We perform extensive experiments and demonstrate the efficiency of proposed approach for ID document compliant face images.

Keywords: face recognition, biometric template, document security

1 Introduction

Security border control applications widely embed biometrics recognition where the face image is one of the most popular biometric source for such applications. The standard approach to the face recognition nowadays implies learning deep face features that are combined into a biometric template. This template is further utilised for distinguishing identities with relatively simple similarity metric and may be stored in a secured database or even embedded to the document itself for performing verification in the match-on-document scenario [MGC20].

The features of the template may be learned explicitly by contrastive methods (i.e. by the contrast between match/non-match pairs [SKP15]) or implicitly in the multiclass (identities) classification manner [De19a]. The deep networks, which are used for extracting the biometric template, usually have complex architectures of stacked convolutional layers. These networks are usually trained on big collections of labelled face images of celebrities [Ca18, Gu16].

Face recognition for document security applications possesses specificities. Official identification documents (i.e. biometric passports, national ID cards) adopt only the frontal face

[1] University of Coimbra, Institute of Systems and Robotics - Coimbra, Portugal, joao.tremoco@isr.uc.pt
[2] University of Coimbra, Institute of Systems and Robotics - Coimbra, Portugal, iurii.medvedev@isr.uc.pt
[3] University of Coimbra, Institute of Systems and Robotics - Coimbra, Portugal; Portuguese Mint and Official Printing Office (INCM), Lisbon, Portugal, nunogon@deec.uc.pt

images compliant to ICAO standards [IS18, IS19]. In comparison with unconstrained face recognition systems, which adapts to variations in illumination, pose, occlusion, facial expressions, document security solutions deal with more regular conditions especially in a situation when biometric enrolment tends to become more controlled [Eu18].

At the same time, the collections of ICAO compliant enrolled images, which are usually stored by national institutions, are hardly available for the research and development due to privacy issues. As an example, European GDPR (General Data Protection Regulation) categorise face images as sensitive personal data which results in many constraints for their collecting and distributing [Eu16]. Recently, following this trend, many of the face datasets (even public wild datasets of celebrities) were withdrawn and usually available only in a form of redistribution.

That is why there is a challenge for face recognition in document security when for efficient training of the face recognition algorithms one require large ICAO compliant face image datasets which remain private, and the ones that are public available are of insufficient size. In this situation the most effective approach is to follow training on available wild datasets and then apply some optional measures (like fine-tuning) for achieving better performance in the deploy scenario [SJ18].

In this work we address the problem of this inconsistency between the training and deploy data and introduce a novel approach to mitigate this issue. We propose to emphasise the face features which are more characteristic for ID document compliant images by designing a sophisticated sample mining strategy which regularises the training process. The developed strategy penalises the samples by their quality score (estimated by several metrics). Our approach allows to learn facial biometric template which better suits the document security applications.

2 Related Work

Loss function design have been in a focus of many recent investigations of deep learning face recognition. The general trend of these works was directed onto the increasing the discriminative power of learned features. Most of the current state of the art methods follow the approach of multi-class classification with use of softmax based loss functions. To increase intra-class compactness and inter-class dispersion, several marginal modifications of softmax were proposed. For instance SphereFace, CosFace and ArcFace introduced the margin (in different manner) to the feature logits in the angular domain [Li17, Wa18, De19a]. These methods demonstrated clear geometric interpretation at the same time having relatively simple implementation. Although these loss functions allowed to achieve state of the art performance in several benchmarks, they do not account the hardness and variability of each sample.

Hard sample mining strategies allowed to improve the face recognition performance in several approaches. For instance, MV-Softmax [Wa20] treats miss-classified samples as hard samples increasing their weights in the training process. CurricularFace [Hu20] also uses miss-classification for indicating hard samples and adapts the curricular learning strategy to the face recognition. Hard samples are emphasised increasingly over the training duration with an additional hyper-parameter. NPCFace [Ze20] makes the important distinction between hard positive and hard negative samples and show that for large datasets hard positives will usually be hard negatives for another class as well. The form of the negative logit is defined with use of a binary mask that indicates whether a sample is hard or not. Following the ArcFace approach, the NPCFace also utilises a margin for the positive logits, which is controlled by the hardness of the sample.

These methods try to optimise their performance towards hard samples, however we propose that for the document security applications emphasising higher quality samples during training better suits the target scenario. Unlike the previous works mentioned, MagFace [Me21] includes the quality of the samples in the training process in a way that pulls easy (high quality) samples closer to the class centre and pushes harder (lower quality) samples away. The authors follow a formulation similar to ArcFace where the margin parameter varies for each sample with accordance to its quality. In MagFace, the quality of each sample is defined by magnitude of the feature vector. This approach shares several conceptual similarities to our approach, however we shift our attention to adapting the quality sampling to the document security images scenario.

Document security specific face recognition investigation is reported in several works. DocFace [SJ18] present a method for matching Identification Document (ID) photos to live photos. The authors use a pair of trained sibling networks and fine-tune them on a small private ID-Selfie dataset. The method achieves better performance over general methods, however the dataset used for benchmarking is private. Several improvements on the ID-Selfie dataset and the loss function for fine-tuning were introduced in the DocFace+ [SJ19].

Face Image Quality Assessment (FIQA) inherits aspects from general image QA also considering several other attributes such as pose, illumination, face occlusion or facial expressions. A survey on this topic was done recently by Schlett et. al [Sc20]. Blur is good baseline indicator for the quality of any image. The blur of an image can be extracted by convolving the image with a Laplacian filter and then calculating the variance of the result [BRC16]. BRISQUE [MMB12] is a no-reference generic image quality assessment method. Through the use of scene statistics this method is able to quantify the "naturalness" and quality of an image. Regarding face specific attributes, several works have been recently developed to extract face specific meta-information from images. The pose of a face in an image can be characterised as a rotation in three dimensions, the yaw pitch and roll. Estimating these angles is helpful to understand a datasets pose distribution. Ruiz et. al [RCR18] use a Convolutional Neural Network (CNN) to estimate these three angles. The quality of facial illumination is also a useful indicator of the quality of a facial image. Zhang et. al [ZZL17] use a CNN, which is trained on the FIIQD dataset to score the quality of illumination. FaceQnet [He19] is a face image QA CNN based method. It used a third party framework to calculate ICAO compliance scores used as ground-truth values to train the network. The authors also show high correlation between the resulting scores and face biometric verification performance for a variety of off-the-shelf biometric recognition systems.

Some recently developed methods of face image quality assessment were developed in such a way to remove human perception from the quality estimation process. SER-FIQA [Ou21] is a quality estimation method based on the use of dropout during the training of a model. The quality of a sample is defined with respect with the robustness of its embeddings in different sub-networks. The closer the outputs are for different sub-networks, the higher the quality of the sample is. Shi and Jain introduced the concept of Probabilistic Face Embedding (PFE) [SJK19]. This work shows that poor image quality affects the similarity scores of genuine and impostor pairs in such a way that higher degradation of an image leads to higher probability of false reject or false accept of these pairs (named Feature Ambiguity Dilemma). As such, instead of the normal deterministic face embedding, the authors propose to encode the uncertainty in the representation of the face with two different output vectors one representing the Gaussian mean and the other for the Gaussian variance. The authors also introduce a method for matching the PFEs that penalises high levels of uncertainty (variance). SDD-FIQA also bases its quality classification on the recognition performance of the sample in question. This is done by mapping the inter-class and intra-class similarity scores to quality pseudo-labels through the use of a distribution distance metric. Afterwards, these quality values are used to train a network to predict quality scores.

3 Methodology

Deep learning classification approaches usually utilise softmax loss function, which now serves as basis for most of recently developed loss functions in the field of face recognition. It is usually formulated as follows:

$$L_{softmax} = \frac{1}{N}\sum_{i} -\log\left(\frac{e^{f_{y_i}}}{\sum_j^C e^{f_{y_j}}}\right) \tag{1}$$

where C is the number of classes in the classification problem, y_i is the index of the class of the $i-th$ sample, N is the number of samples in a batch and f_{y_j} is the $y_j - th$ component of the final layer's logits $\mathbf{f}$. If l2 normalisation of the weights $\mathbf{w_j}$ and biometric feature set $\mathbf{x_i}$ is performed, then f_{y_j} can be represented as: $f_{y_j} = w_j^T x_i = \cos(\theta_j)$. The normalised features are constrained on the hyper sphere in $\mathbb{R}^d$ space (where d is the size of $\mathbf{f}$), which leads to the angular similarity metric between samples. By reformulating softmax with this normalisation and adding an angular margin parameter m to the positive logit we obtain the ArcFace loss:

$$L_{arcface} = \frac{1}{N}\sum_{i} -\log\left(\frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j\neq y_i} e^{s\cos\theta_j}}\right) \tag{2}$$

QualFace. Basing on the cooperative margin presented in NPCFace [Ze20], we introduce the concept of adaptive margin with regard to image quality. Our approach, unlike others previously mentioned, implies developing the sample mining strategy, which enhance the impact of higher quality samples instead of harder samples. In this case deep feature distribution is characterised by the concentration of the qualitative samples closer to the class feature centre (see Fig. 1). With this approach, higher impact means higher loss value for samples with better quality. This is done by increasing the margin parameter in the ArcFace loss in an adaptive way, which results in the following formulation:

$$L_q = \frac{1}{N}\sum_i -\log\left(\frac{e^{s\cos(\theta_{y_i}+m_i)}}{e^{s\cos(\theta_{y_i}+m_i)}+\sum_{j\neq y_i} e^{s\cos\theta_j}}\right) \tag{3}$$

where the adaptive margin parameter m_i is defined as a baseline value plus an added constant dependent on the quality of the image:

$$m_i = m_0 + \sum_j^Q w_j q_{ij} m_1 \tag{4}$$

Here, m_0 and m_1 are hyper-parameters, q_{ij} represents the normalised $j-th$ quality score value for the sample i. Q is the total number of quality attributes and w_j is the weight of each score. For travel document photos, we consider high quality samples as samples that have high ICAO standards compliance [IS18]. For instance, images with frontal poses, clear background, frontal face lighting, no face occlusion, no facial expressions, etc. In our work we use five different indicators of quality that are inspired by ICAO recommendations for portrait photographs: Blur [BRC16], FaceQNet scores [He19], BRISQUE scores [MMB12], Face Illumination quality [ZZL17] and a pose score [RCR18]. The pose scores used were calculated as the average of absolute values of the yaw, pitch and roll angles. QualFace strengthens the supervision on higher quality samples through the use of external quality indicators. The following section will show the advantages of QualFace on document security applications.

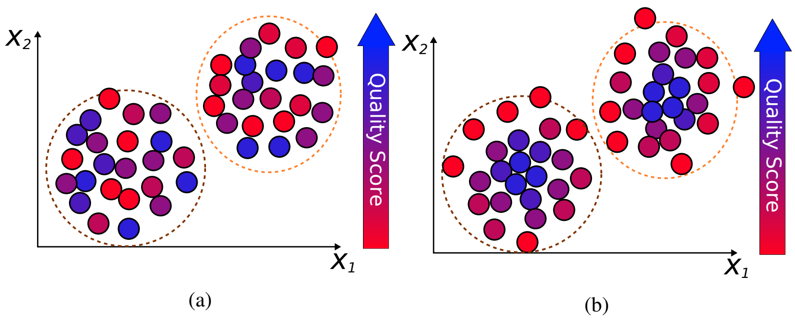


Fig. 1: The spatial distribution of two high level features; a) default feature distribution; b) desired distribution in our method.

4 Experiments and Results

We perform extensive training experiments with QualFace and several baseline loss functions and benchmark the result models in a following way.

Training. As a training data we used the subset of public VGGFace2_train dataset [Ca18], selecting classes with more than 400 images per identity. The resulting dataset has a total of 1.34M images and 2842 identities. Face detection and alignment to 299×299 is performed with use of RetinaFace method [De19b]. Each image channel is normalised by subtracting the mean of the training dataset. The scores (FaceQNet, BRISQUE, pose score) were extracted from the aligned images. They are normalised and fed to the model as additional input.

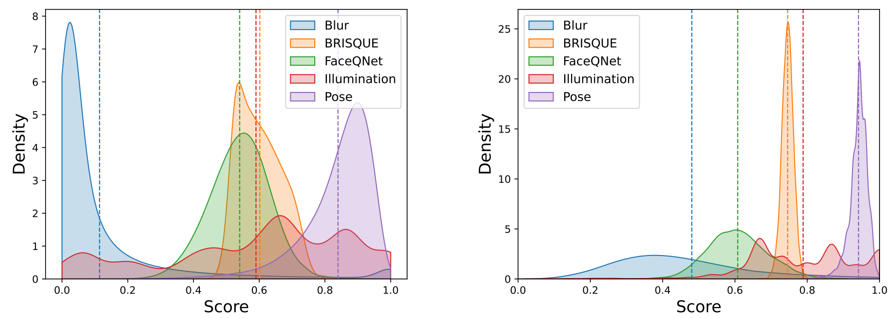
As a backbone CNN architecture we choose the ResNet50V2 [He16], adding the fully connected feature layer with 512 nodes. We initialise all models with the imagenet weights before training. The training was performed on a NVIDIA RTX 3090 GPU. We limit the batch size with 24 images and decay the learning rate with cosine annealing scheduler from $5e-3$ in the beginning to $1e-5$ in the end. The model is trained with SGD optimiser for $6-th$ epochs with a momentum parameter of 0.5 and weight decay of 0.0005.

Benchmarking. In order to demonstrate the effect of our method, and its superiority for ID document compliant images, we designed two different benchmarks datasets. The first one includes ”wild” images, and the second one is comprised of images that are compliant to ICAO standards (we call it ”strict”). The wild benchmark dataset was created basing on a subset of VGGFace2_test part and include identities disjoint from the training set. It contains 31k face images of 147 identities. The strict dataset was created with images from the Face Recognition Grand Challenge V2 (FRGC_V2) dataset [Ph05]. Since its default version includes wild images, we performed its filtering in a semi-automatic way choosing only ICAO compliant images. The final strict dataset contains 11.7k images from 565 identities. For each dataset we generated the protocols for 1-1 for verification by random selecting of comparison pairs. Each protocol contains around 110K pairs for match comparison and 220K pairs for non-match comparison. [3]

To demonstrate the relative difference of distributions across two benchmark datasets we performed min-max normalisation with respect to the minimum and maximum scores values for the VGGFace2_train. One can see that the designed strict benchmark (see Fig. 2b) has better image quality with respect to the five scores presented. The wild benchmark dataset distributions, as expected, turned out to be identical to the train dataset distributions (see Fig. 2a).

Results Discussion. We performed intensive experiments training deep networks with QualFace and observed that the strong applied adaptation usually lead to a problem with the convergence. However, applying regular and careful adaptation, we could attain the superiority of our method. We achieved the best results in two following configurations: $m_0 = 0.4$ with $m_1 = 0.1$ and $m_1 = 0.2$. For each of those we trained five different models using a single score: Blur, BRISQUE, FaceQNet, Illumination and Pose. The Receiver

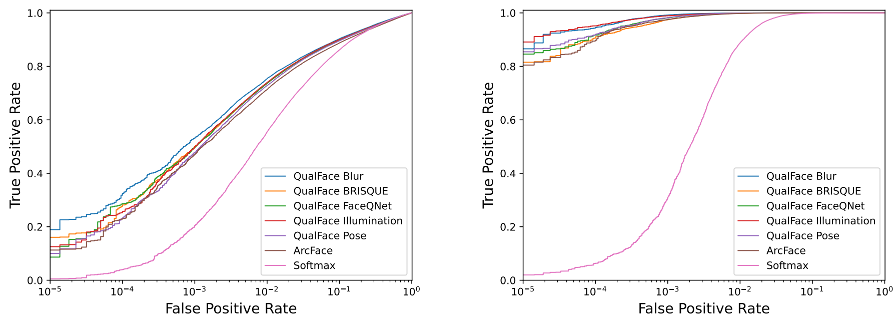
[3] https://github.com/visteam-isr-uc/QualFace



(a) (b)

Fig. 2: Normalised quality scores distributions across the datasets; a) VGGFace2_train dataset (identical to VGGFace2_test); b) FRGC_V2 test strict dataset.

Operating Characteristic (ROC) curves of the trained QualFace models (with $m_0 = 0.4$ and $m_1 = 0.1$) are represented in Fig. 3 as well as ArcFace and Softmax models for comparison.



(a) (b)

Fig. 3: ROC curves; a) Wild Benchmark; b) Strict ICAO compliance Benchmark.

From the ROC curves one can see that most of the QualFace models have better operation curves than ArcFace and Softmax. For the strict benchmark, the illumination score QualFace model exhibits the best results while for the wild benchmark the blur scores is the best performing. We estimate the performance by several metrics: False Non-Match Rate at False Match Rate (FNMR@FMR) and Area Under Curve (AUC) of ROC (see Table 1).

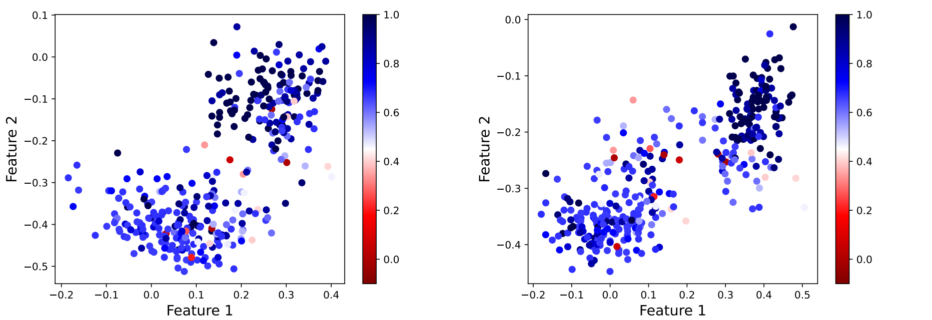
From the results obtained, we conclude that QualFace significantly enhances the biometric verification performance in ICAO compliant face images when compared to a simple margin based loss function like ArcFace. This statement can be verified for most of the models trained in both configurations, however the models with $m_1 = 0.1$ clearly show superior results. Considering wild benchmarks, our approach performs on par with the baseline models. However, most of QualFace experiment results still slightly outperform

Tab. 1: FNMR@FMR thresholds and AUC scores for two benchmarks.

Method		Wild			Strict			
		1e-2	1e-3	AUC	1e-3	1e-4	1e-5	AUC
Softmax		0.44502	0.79633	0.944118	0.69017	0.93655	0.98027	0,995333
ArcFace		0.28680	0.52938	0.951181	0.02486	0.10205	0.19507	0,999871
QualFace (m_0=0.4, m_1=0.1)	Blur	**0.24600**	**0.46806**	**0.959089**	**0.00793**	0.05453	0.13429	**0.999957**
	BRISQUE	0.26185	0.49934	0.954925	0.02556	0.08950	0.18444	0.999878
	FaceQNet	0.26383	0.50290	0.956515	0.01874	0.08284	0.15398	0.999910
	Illumination	0.26037	0.50076	0.956797	0.01066	**0.04835**	**0.10878**	0.999936
	Pose	0.27177	0.52186	0.954926	0.01805	0.08011	0.14550	0.999917
QualFace (m_0=0.4, m_1=0.2)	Blur	0.27460	0.52942	0.956314	0.04183	0.13423	0.19618	0.999813
	BRISQUE	0.28253	0.57363	0.953276	0.04329	0.21139	0.30880	0.999772
	FaceQNet	0.26524	0.54649	0.956252	0.03185	0.05963	0.19958	0.999944
	Illumination	0.29351	0.56006	0.952535	0.04046	0.14946	0.21644	0.999792
	Pose	0.28781	0.51604	0.951155	0.01146	0.11058	0.19958	0.999838

ArcFace. We conclude that our method allows to regularise the training process in deeper manner (not just adapting to qualitative samples) but generally learns better (more qualitative/discriminative) face features. From that point of view, our approach inherently shares conceptual similarities with the curriculum learning strategy.

Feature distribution. To better understand the QualFace impact to the learning process we analysed the real feature distribution for several particular identities in the benchmark datasets. To constrain the analysis in the 2D case we extract two principal components from the 512 dimensional embeddings with PCA (Principal Component Analysis). We represent the resulting feature distributions for two identities from the FRGC_V2 Dataset Fig. 4.



(a) (b)

Fig. 4: Features distribution of 2 different identities (04430 and 02463) from the FRGC_V2 dataset with Illumination scores represented in colour; a) ArcFace Model; b) Illumination Score QualFace Model with $m_0 = 0.4$ and $m_1 = 0.1$.

Basing on our results we make two observations. First, the separation between identities, which is commonly seen in margin based methods can be confirmed both in ArcFace and QualFace cases. Second, while ArcFace does not take into account image quality, Qual-

Face pulls high quality samples towards the class centre and compacts their distribution, while the low quality samples are pushed away as theoretically hypothesised in Fig. 1b.

Combined scores experiments After the experiments with sampling by a single score we intuitively investigated several scores averaging techniques. Namely, we utilised straight forward mean value, weighted mean and several median value implementations. The median implementations used three scores each. The *Median Lower* model averaged the three lower scores, the *Median* model - the three centre scores and the *Median Higher* averaged the three highest scores, for each image. We also made experiments with uniforming scores distributions in the range $[0,1]$ before averaging for equalising their impact. The ROC curves of the combined models are represented on Fig. 5

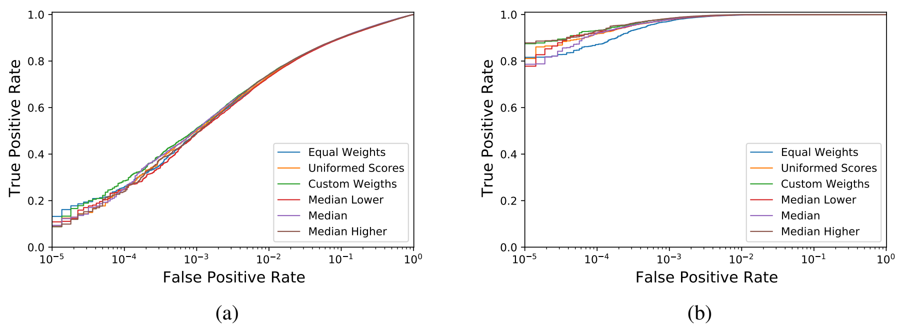


Fig. 5: Combined model ROC curves; a) Wild Benchmark; b) Strict ICAO compliance Benchmark.

In our experiments the models with *Custom Weights* and *Median Higher* weights averaging demonstrated the best performance in benchmarks. This can also be confirmed from the AUC and FNMR@FMR metrics, which are represented in table 2.

Tab. 2: FNMR@FMR thresholds and AUC scores for two benchmarks using all five scores QualFace models with m_0=0.4, m_1=0.1.

Models	Wild			Strict			
	1e-2	1e-3	AUC	1e-3	1e-4	1e-5	AUC
Equal Weights	0.26495	0.50912	0.956074	0.02869	0.12879	0.18398	0.999850
Uniformed Scores	0.26393	0.50244	0.957280	0.02171	0.08221	0.18881	0.999897
Custom Weights	**0.25735**	**0.48964**	0.956706	0.01834	0.06875	0.12521	0.999907
Median Lower	0.26914	0.51016	0.955681	0.02195	0.07807	0.22204	0.999891
Median	0.25829	0.49400	**0.958679**	0.02087	0.07853	0.21371	0.999905
Median Higher	0.25877	0.51080	0.957604	**0.01629**	**0.07027**	**0.12184**	**0.999929**

We made several observations regarding the usage of combined scores. Scores uniforming indeed allowed better regularise the training and achieve better performance results. Scores weighing demonstrated its importance and the best performance was achieved when the weights were selected according to the results of single score models (Blur - 0.3, BRISQUE - 0.1, FaceQnet - 0.15, Illumination - 0.3, Pose - 0.15 in our experiments).

In the list of models with median averaging, *Median Higher* case gave the most promising result, which means that the QualFace sampling strategy should be good score biased. In other words, it is better to treat a sample by its best scores rather than consider it as a bad sample even if it has some lower scores.

The use of combined scores did not demonstrate the superiority in any particular benchmark. However, it allowed to achieve more regular results across the two utilised benchmarks (strict and wild) making the face representation more universal in applications with unspecified scenario. This can be verified when comparing the *Custom Weight* and *Median Higher* model with the single score blur and illumination models.

We conclude that sampling of face images with single generic illumination and blur quality metrics allow to learn better face representation when applying the QualFace technique. Particularly, illumination quality is better suitable in application to the document security scenario, while blur score better shifts the performance towards wild face recognition scenario.

5 Conclusions

In this work we proposed a novel approach of adapting deep learning face recognition methods for document security applications. We introduced a sophisticated sample mining strategy that regularises the training process by careful emphasising the impact of samples which are better suitable for document security. The method allows to effectively train face recognition networks on big wild datasets and at the same time reduce the effect of ”wildness” of these datasets. The extensive experiments with the selected marginal loss function (ArcFace) proved the superiority of adapted models against the default ones in tests with ID compliant images. The introduced strategy can also be applied to other loss functions. Our future work will focus on the study of additional image quality metrics more specific to concrete ICAO requirements. Experiments with different loss functions and finding better normalisation for the quality scores are also part of our future work plan.

6 Acknowledgements

The authors would like to thank the Portuguese Mint and Official Printing Office (INCM) and the Institute of Systems and Robotics - University of Coimbra for the support of the project Facing. This work has been supported by Fundação para a Ciência e a Tecnologia (FCT) under the project UIDB/00048/2020.

References

[BRC16] Bansal, Raghav; Raj, Gaurav; Choudhury, Tanupriya: Blur image detection using Laplacian operator and Open-CV. In: 2016 International Conference System Modeling Advancement in Research Trends (SMART). pp. 63–67, 2016.

[Ca18] Cao, Q.; Shen, L.; Xie, W.; Parkhi, O. M.; Zisserman, A.: VGGFace2: A Dataset for Recognising Faces across Pose and Age. In: 2018 13th IEEE International Conference on AFGR. pp. 67–74, 2018.

[De19a] Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S.: ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In: 2019 IEEE Conference on Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4685–4694, 2019.

[De19b] Deng, J.; Guo, J.; Zhou, Y.; Yu, J.; Kotsia, I.; Zafeiriou, S.: RetinaFace: Single-stage Dense Face Localisation in the Wild. CoRR, abs/1905.00641, 2019.

[Eu16] General Data Protection Regulation. Official Journal of the European Union.

[Eu18] European Enrolment Guide for Biometric ID Documents. CEN/TC 224.

[Gu16] Guo, Y.; Zhang, L.; Hu, Y.; He, X.; Gao, J.: MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition. In: 2016 ECCV. volume 9907, pp. 87–102, 10 2016.

[He16] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian: Identity Mappings in Deep Residual Networks. CoRR, abs/1603.05027, 2016.

[He19] Hernandez-Ortega, J.; Galbally, J.; Fierrez, J.; Haraksim, R.; Beslay, L.: FaceQnet: Quality assessment for face recognition based on deep learning. arXiv, 2019.

[Hu20] Huang, Y.; Wang, Y.; Tai, Y.; Liu, X.; Shen, P.; Li, S.; Li, J.; Huang, Feiyue: CurricularFace: Adaptive Curriculum Learning Loss for Deep Face Recognition. In: 2020 IEEE Conference on Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5900–5909, 06 2020.

[IS18] Portrait Quality - Reference Facial Images For MRTD. https://www.icao.int/Security/FAL/TRIP/Documents/TR - Portrait Quality v1.0.pdf.

[IS19] Information technology — Extensible biometric data interchange formats — Part 5: Face image data. ISO/IEC 39794-5:2019.

[Li17] Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L.: SphereFace: Deep Hypersphere Embedding for Face Recognition. In: IEEE Conference on Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6738–6746, 2017.

[Me21] Meng, Qiang; Zhao, Shichao; Huang, Zhida; Zhou, Feng: MagFace: A Universal Representation for Face Recognition and Quality Assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14225–14234, June 2021.

[MGC20] Medvedev, I.; Gonçalves, N.; Cruz, L.: Biometric System for Mobile Validation of ID And Travel Documents. In: 2020 International Conference of the BIOSIG. pp. 1–5, 2020.

[MMB12] Mittal, A.; Moorthy, A. K.; Bovik, A. C.: No-Reference Image Quality Assessment in the Spatial Domain. IEEE Transactions on Image Processing, 21(12):4695–4708, 2012.

[Ou21] Ou, Fu-Zhao; Chen, Xingyu; Zhang, Ruixin; Huang, Yuge; Li, Shaoxin; Li, Jilin; Li, Yong; Cao, Liujuan; Wang, Yuan-Gen: SDD-FIQA: Unsupervised Face Image Quality Assessment With Similarity Distribution Distance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7670–7679, June 2021.

[Ph05] Phillips, P.J.; Flynn, P.J.; Scruggs, T.; Bowyer, K.W.; Chang, Jin; Hoffman, K.; Marques, J.; Min, Jaesik; Worek, W.: Overview of the face recognition grand challenge. In: 2005 IEEE Computer Society Conference on Conference on Computer Vision and Pattern Recognition (CVPR). volume 1, pp. 947–954, 2005.

[RCR18] Ruiz, N.; Chong, E.; Rehg, J.: Fine-Grained Head Pose Estimation Without Keypoints. In: The IEEE Conference on Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 06 2018.

[Sc20] Schlett, Torsten; Rathgeb, Christian; Henniger, Olaf; Galbally, Javier; Fierrez, Julian; Busch, Christoph: , Face Image Quality Assessment: A Literature Survey, 09 2020.

[SJ18] Shi, Yichun; Jain, Anil K.: DocFace: Matching ID Document Photos to Selfies*. In: 2018 IEEE 9th International Conference on BTAS. pp. 1–8, 2018.

[SJ19] Shi, Yichun; Jain, Anil K.: DocFace+: ID Document to Selfie Matching. IEEE Transactions on Biometrics, Behavior, and Identity Science, 1(1):56–67, 2019.

[SJK19] Shi, Yichun; Jain, Anil K.; Kalka, N.: Probabilistic Face Embeddings. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6901–6910, 2019.

[SKP15] Schroff, F.; Kalenichenko, D.; Philbin, J.: FaceNet: A unified embedding for face recognition and clustering. In: 2015 IEEE Conference Conference on Computer Vision and Pattern Recognition (CVPR). pp. 815–823, 2015.

[Wa18] Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; Liu, W.: CosFace: Large Margin Cosine Loss for Deep Face Recognition. In: 2018 IEEE/CVF Conference on Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5265–5274, 2018.

[Wa20] Wang, X.; Zhang, S.; Wang, S.; Fu, T.; Shi, H.; Mei, T.: Mis-Classified Vector Guided Softmax Loss for Face Recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 34:12241–12248, 04 2020.

[Ze20] Zeng, D.; Shi, H.; Du, H.; Wang, J.; Lei, Z.; Mei, T.: NPCFace: A Negative-Positive Cooperation Supervision for Training Large-scale Face Recognition. CoRR, abs/2007.10172, 2020.

[ZZL17] Zhang, L.; Zhang, L.; Li, L.: Illumination Quality Assessment for Face Images: A Benchmark and a Convolutional Neural Networks Based Model. Lecture Notes in Computer Science, 10636 LNCS:583–593, 2017.