LightQNet: Lightweight Deep Face Quality Assessment for Risk-Controlled Face Recognition

Kai Chen, Taihe Yi, and Qi Lv

***Abstract*—End-to-end face quality assessment based on deep learning can directly predict the overall quantitative score of face quality, thus helping to control the risk of face recognition system. Thanks to the development of automatic quality pseudo-label generation, most recent methods can use large-scale face datasets to learn the quality model. However, existing methods use regression models to fit the pseudo-labels, which lack attention to samples that are easy to be misidentified, and require large models for training. The paper treats the quality assessment as a classification problem, focusing on difficult samples near the classification boundary. Specifically, pairwise binary quality pseudo-label is generated based on the face similarity score without additional manual annotation. An identification quality loss is used to decouple the pairwise network training. In addition, a lightweight quality network is trained by performing knowledge distillation on the quality prediction branch of the face recognition network. Experiments show that the proposed quality network achieves state-of-the-art results with only 0.45M parameters and 77M FLOPs.**

***Index Terms*—Face quality assessment, deep learning, face recognition, data uncertainty estimation.**

I. Introduction

IN REAL-WORLD scenes, the face recognition system is susceptible to the noise of the input data, such as low illumination, large pose, large expression, occlusion and blur, and so on, making the face image easy to be misidentified [1], [2]. Checking the quality of the input face images is one way to improve the security and reliability of the face recognition system. In complex recognition scenarios, we typically expect that the recognition system will be able to control its own recognition risk. The system should be able to reject the image when the it is uncertain about the input image. The uncertainty of face recognition is affected by the ability of the recognition model (model uncertainty) and the quality of the input image (data uncertainty) [2]. Quantitative assessment of face image quality can help face recognition models to reduce the data uncertainty. As shown in Fig. 1, faces that do not meet the recognition conditions can be filtered out using the predictive quality, thereby minimizing mis-recognition caused by low-quality faces.

The challenge of face quality assessment is how to determine a good quality metric that can be used as an "oracle" that are highly correlated to recognition performance [3].

Manuscript received July XX, 2021; revised XX XX, 2021. This work was supported by the National Natural Science Foundation of China (NSFC) under Grants 61906207 and 61803376. *(Corresponding author: Qi Lv.)*

K. Chen and T. Yi are with the College of Systems Engineering, National University of Defense Technology, Changsha, China (e-mail: chenkai@nudt.edu.cn; yitaihe@nudt.edu.cn).

Q. Lv is with the College of Meteorology and Oceanography, National University of Defense Technology, Changsha, China (e-mail: lvqi@nudt.edu.cn).

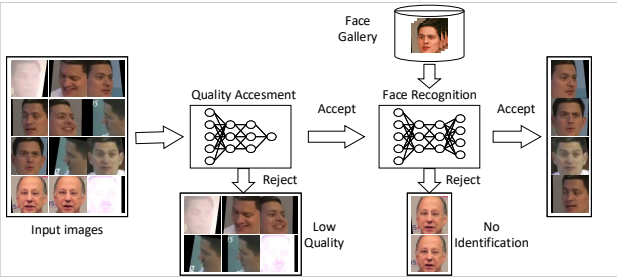


Fig. 1. Face quality assessment can improve the performance of face recognition by rejecting low-quality faces.

In early work, face quality is measured using a variety of analytical metrics, such as posture, expression, illuminance, and occlusion. These metrics are usually predicted through image analysis [4], [5], machine learning [6]–[8] and deep learning [9]–[11] methods. However, since humans may not know the best factors influencing the face recognition, these methods may only provide sub-optimal results.

Recently, end-to-end deep learning provides a method of directly predicting face quality score, thereby improving the overall results of the face recognition system [12]. Building a labeled dataset that can train the neural networks is time-consuming and labor-intensive. So that the automatic quality label generation methods are proposed to save the annotation work [13], [14]. These methods assume that a low similarity of two face images will be caused by the lower quality image [15]. This assumption is successfully used to design pseudo-labels in FaceQNet [14] and PCNet [16], as shown in Fig. 2(a) and (b). Both methods used regression to fit the pseudo-labels, with each sample having the same impact on training. However, images of different quality have different effects on face recognition. In practice, we must specify a quality threshold to reject low-quality images. Therefore, images near to the threshold have a stronger impact on face recognition. In addition, existing methods usually use large models such as ResNet [14], [16], [17], or directly use the existing face recognition model [2], [18], resulting in the lack of flexibility.

In this paper, we treat the quality assessment as a binary classification task, as shown in Fig. 2(c), so that the samples at the classification boundary (near the threshold) will get more attention. We propose a binary pseudo-label generation method and the corresponding quality loss function for training. In order to reduce the complexity of quality network, we propose a branch-based quality distillation method, which can improve the accuracy of lightweight model, as shown in Fig. 3. Experiments show that the proposed method can achieve state-of-the-art performance with small model size and low

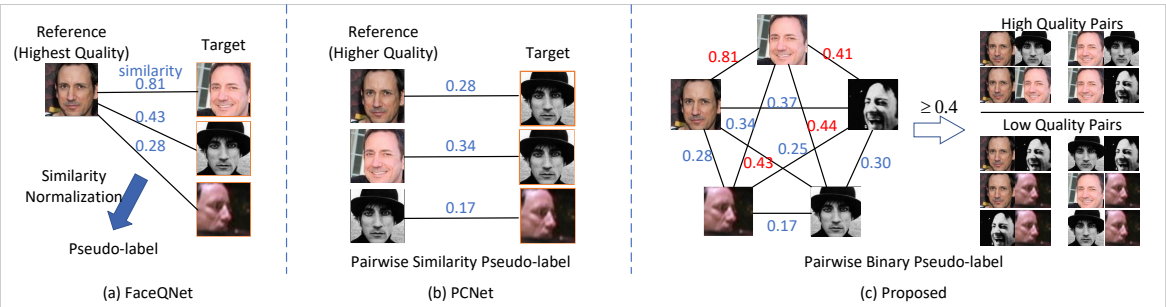


Fig. 2. Similarity-based pseudo-label generation. (a) The highest quality face image is selected as the reference to generate pseudo-labels of the targets. (b) Pairwise similarity is used as the pseudo-labels of the lower-quality faces. (c) Proposed pairwise binary pseudo-label with the threshold $m = 0.4$.

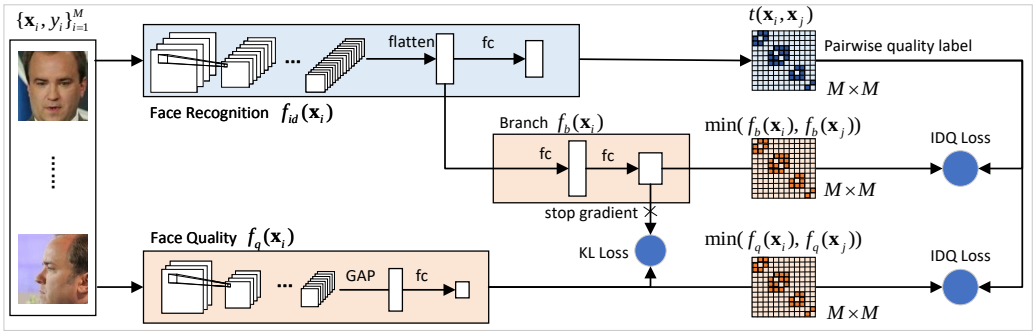


Fig. 3. Architecture of proposed method. The weights of the face recognition network are fixed during training. Pairwise binary pseudo-labels are generated based on the similarities of M faces in the mini-batch. IDQ loss is used to train the branch network and quality network. KL loss is used to perform knowledge distillation from branch network to face quality network. The highlighted elements in $M \times M$ matrices indicate that the sample pair belongs to the same class.

computation complexity. The codes have been released[1].

II. Proposed Methods

A. Identification Quality Loss

To improve the face recognition performance, we should focus on the samples that easily cause face recognition errors due to low-quality. We convert the prediction of the face quality score into a binary classification problem. So that images far from the classification boundary become simple samples, which usually have higher or lower quality. Images near the classification boundary are more difficult to distinguish and will get more attention during the training process.

For image pairs $\mathbf{x}_i$ and $\mathbf{x}_j$ with the same identification label $y_i = y_j$, we define the pairwise binary pseudo-label as follow:

$$t_{ij}^{hard} = I\{\cos(f_{id}(\mathbf{x}_i), f_{id}(\mathbf{x}_j)) > m\} \tag{1}$$

where m is the quality threshold parameter, $f_{id}(\mathbf{x})$ is the face recognition model, and $I\{\cdot\}$ is the indicator function that is 1 when the cosine score of the two face images is greater than the threshold m and 0 otherwise. The label of Equation (1) is either 0 or 1, so we call it hard label. In order to make full use of the information of face similarity score, we define the following soft label:

$$t_{ij}^{soft} = \frac{1}{1 + \exp(-s \cdot (\cos(f_{id}(\mathbf{x}_i), f_{id}(\mathbf{x}_j)) - m))} \tag{2}$$

[1]https://github.com/KaenChan/lightqnet

where s is scale factor that controls the smoothness of the label. When s is large, the distribution of t is more concentrated to 0 or 1. When s is small, the distribution of t is more scattered. Soft label is a method of output regularization, which has been used in label smoothing [19] and distillation [20]. The use of soft labels can prevent over-fitting of the quality network.

Similar to PCNet [16], we use the "loser take all" scheme to decouple the pairwise scores and train the network for a single face. Specifically, assuming that the face similarity is determined by the image with the lower quality, we can take the minimum of the two quality predictions as the output of the pair of images. Then, the identification quality (IDQ) loss can be obtained through cross-entropy as follow:

$$\begin{aligned} L_{IDQ}^{q} &= \frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} -t_{ij} \log \hat{f}_{ij} - (1 - t_{ij}) \log(1 - \hat{f}_{ij}) \\ \hat{f}_{ij} &= \min(f_q(\mathbf{x}_i), f_q(\mathbf{x}_j)) \end{aligned} \tag{3}$$

where $\mathcal{P}$ is the set of pairs satisfying $y_i = y_j$ in the mini-batch and $|\mathcal{P}|$ is the number of the pairs. $f_q(\mathbf{x})$ is the quality network. During training, the min operation will select sample with low predictive quality for gradient back-propagation.

B. Branch-based Quality Distillation

Knowledge distillation is a powerful method to improve the performance of lightweight models, which uses the dark knowledge of the large network to teach the training process

of the small network [20]. As shown in Fig. 3, we use the branch network $f_b(\mathbf{x})$ based on the last convolutional layer of the face recognition model as the teacher, which can be trained using the identification quality loss L^b_{IDQ}, by replacing the $f_q(\mathbf{x})$ with $f_b(\mathbf{x})$ in Equation (3). Branch-based network is a commonly used technique in the tasks of probabilistic face embeddings [2], learning confidence [21], selective classification [22], etc. KL loss is used to apply the knowledge distillation from teacher $f_b(\mathbf{x})$ to student $f_q(\mathbf{x})$. In order to prevent the student network from interfering with the training of the teacher network, we prohibit passing the gradient of the distillation loss back to the teacher network. To this end, we define ${f'}_b(\mathbf{x})$ to be a non-gradient version of $f_b(\mathbf{x})$. Then the distillation loss function is defined as follows:

$$L_{Distill} = \sum_{i=1}^{M} D_{KL}({f'}_b(\mathbf{x}_i)||f_q(\mathbf{x}_i)) \tag{4}$$

Combining the above losses, we can write the total loss as follow:

$$L = L^q_{IDQ} + L^b_{IDQ} + \lambda L_{Distill} \tag{5}$$

where λ are the weight of distillation loss.

C. Network Architectures

The proposed lightweight network architecture, LightQNet, is based on the lite version of MobileNetv3-small [23]. Compared with the original version, the lite version of the network removes the SENet structure and changes the activation function to ReLU, making it easier to deploy the network to various platforms. The input image size is 96×96. The stride of the first layer is set to 1 in order to prevent the feature map from being too small. The last layer of the network uses the Sigmoid activation function to output a face quality score of 0 to 1. The number of parameters in LightQNet is only 0.45M, and the amount of calculation is 77M FLOPs.

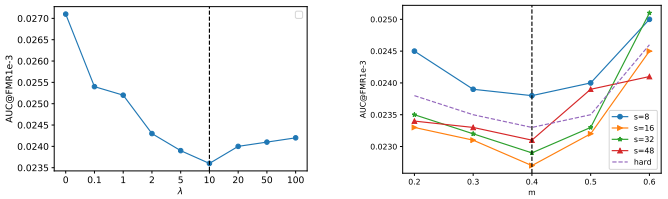
III. Experiments

A. Experiments Setup

Datasets: We use MS-Celeb-1M-v2 dataset [24] as the training dataset, which is a clean version of MS-Celeb-1M dataset [25]. There are 5 testing sets, including LFW [26], CPLFW [27], CFP-FP [28], Adience [29] and IJB-B [30]. The image preprocessing is similar to ArcFace [24]. All training and test images are aligned by affine transformation according to the key points of the face, and resized to 96×96.

Implementation Details: The pre-trained face recognition model is trained on MS-Celeb-1M-v2 dataset using the ArcFace loss [24] based on ResNet64 with embedding size 256. The quality modules are trained by SGD optimizer with momentum of 0.9 and weight decay of $1e-5$ on one GTX1080Ti GPU. The batch size is 128. In each batch, 8 persons are sampled, each with 16 images. The learning rate starts at 0.01, then decreases to 0.001 and 0.0001 in steps of 32K and 48K, respectively, and ends with a step of 64K.

Evaluation Protocol: Following [31], [32], we use Area-Under-Curve (AUC) of the error-versus-reject curve (ERC) as the evaluation metric, $AUC = \int ERC - A_{best}$, where A_{best} is the area under the best case where the error value decrease equals the rejected fraction percentage. AUC at 0.001 FMR (False Match Rate) is reported for LFW, CPLFW, CFP-FP and Adience. For the set-to-set dataset IJBB, the quality score is used to perform a weighted average of the face features in the frame set as [16] and the TAR@FAR1e-4 is reported.



(a) Different λ with fixed $s = 16$ and $m = 0.4$

(b) Different s and m with fixed $\lambda = 10$

Fig. 4. Parameters sensitivity results on the CPLFW dataset.

B. Ablation Experiments

Fig. 4 analyzes the parameters sensitivity of our approach on CFP-FP dataset. Distillation can improve the performance as λ increases and get the best result when $\lambda = 10$, as shown in Fig. 4(a). For pairwise binary label in Fig. 4(b), soft label with $s = 16$ and $m = 0.4$ gets the best AUC@FMR=1e-3 result. In subsequent experiments, the hyper-parameter λ is set to 10, and s and m are set to 16 and 0.4 respectively.

Table I reports the influences of different pairwise labels and losses, with or without distillation. We can see that the result of classification is significantly better than regression. Soft label can further improve performance due to the output regularization. Furthermore, branch-based quality distillation can effectively improve the performance of small LightQNet model.

TABLE I
ABLATION RESULTS (AUC@FMR=1E3) OF PROPOSED METHOD. CE DENOTES CROSS ENTROPY.

Network	Pairwise label	Loss	CPLFW	CFP-FP	Adience
Branch	Cosine	L2	0.0377	0.0074	0.0238
	Binary(hard)	CE	0.0238	0.0061	0.0231
	Binary(soft)	CE	**0.0236**	**0.0056**	**0.0227**
LightQNet w/o distill	Cosine	L2	0.0556	0.0199	0.0327
	Binary(hard)	CE	0.0341	0.0143	0.0288
	Binary(soft)	CE	**0.0337**	**0.0136**	**0.0279**
LightQNet w/ distill	Cosine	L2	0.0306	0.0103	0.0285
	Binary(hard)	CE	0.0259	0.0101	0.0243
	Binary(soft)	CE	**0.0257**	**0.0098**	**0.0238**

C. Comparison with State-Of-The-Art

In Table II, we use the same recognition model to generate quality pseudo-labels and to extract features, both using ResNet64. Our method is compared with related quality assessment methods. Among them, FaceQNetv1 is the v1 version model of FaceQNet[2]. SER-FIQ [18] uses the ResNet64 model

[2]https://github.com/uam-biometrics/FaceQnet

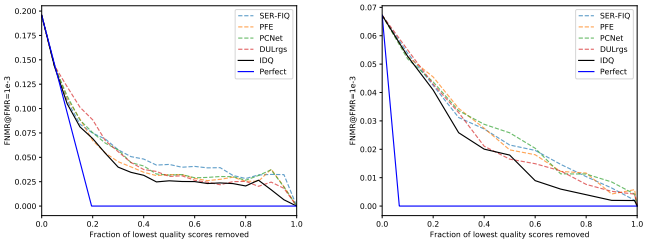
to perform test-time dropout. The losses of PFE [2], PCNet [16], DUL-rgs [33] and soft-label IDQ are used to train the branch network and LightQNet network with distillation. It can be seen from the experimental results: 1) IDQ Loss can get the best results on most of the test items;2) After using distillation, LightQNet The performance can be comparable to the SOTA method; 3) For the LFW dataset, the results of the methods have little difference, because the face image quality of this dataset is better.

TABLE II
RESULTS ON THE SAME RECOGNITION MODEL, RESNET64. TAR IS REPORTED AT FAR=1E-4.

Methods	AUC@FMR1e-3 ↓ LFW	CPLFW	CFP-FP	Adience	TAR ↑ IJB-B
Baseline	-	-	-	-	93.89
FaceQNetv1	0.0040	0.1578	0.0585	0.0533	93.89
SER-FIQ(R64)	0.0023	0.0369	0.0257	0.0239	94.23
R64 Branch Network					
PFE	**0.0014**	0.0377	0.0074	0.0240	94.00
PCNet	0.0018	0.0252	0.0078	0.0238	94.20
DUL-rgs	0.0016	0.0323	0.0069	0.0235	94.25
IDQ	0.0015	**0.0236**	**0.0056**	**0.0227**	**94.33**
LightQNet + Branch-based Distillation					
PFE	0.0015	0.0306	0.0103	0.0285	94.23
PCNet	**0.0014**	0.0304	0.0127	0.0275	94.22
DUL-rgs	0.0016	0.0290	0.0124	0.0243	94.17
IDQ	**0.0014**	**0.0257**	**0.0098**	**0.0238**	**94.32**

In Table III, we use ResNet64 model to generate quality pseudo-labels and use LResNet100E-IR[3] [24] to extract features. It can be seen that the proposed IDQ loss can get the best results, indicating that it has good generalization capabilities across different face recognition networks. The results of LightQNet on LFW, Adience and IJB-B are similar to those of branch-based quality network, indicating that the proposed approach is very effective in training lightweight quality assessment networks.

Fig. 5 shows the Error-versus-reject curve on CPLFW and Adience using the same recognition model. The "Perfect" line is generated using $\max(FNMR - x, 0)$, which means it can perfectly predict which image pairs are implicated in false non-matches [15]. Please refer to the appendix for more results.



(a) CPLFW (b) Adience

Fig. 5. Error-versus-reject curve on CPLFW and Adience.

[3]https://github.com/deepinsight/insightface

TABLE III
RESULTS ON CROSS RECOGNITION MODEL, LRESNET100E-IR. TAR IS REPORTED AT FAR=1E-4.

Methods	AUC@FMR1e-3 ↓ LFW	CPLFW	CFP-FP	Adience	TAR ↑ IJB-B
Baseline	-	-	-	-	94.67
FaceQNetv1	0.0040	0.1312	0.0412	0.0529	94.68
SER-FIQ(R64)	0.0023	0.0265	0.0133	0.0305	94.79
R64 Branch Network					
PFE	0.0018	0.0258	0.0051	0.0226	94.66
PCNet	0.0021	0.0166	0.0052	0.0232	94.83
DUL-rgs	0.0018	0.0254	0.0053	0.0234	94.86
IDQ	**0.0016**	**0.0162**	**0.0040**	**0.0220**	**94.90**
LightQNet + Branch-based Distillation					
PFE	0.0017	0.0231	0.0072	0.2710	94.81
PCNet	0.0018	0.0246	0.0076	0.0286	94.80
DUL-rgs	0.0016	0.0294	0.0131	0.0257	94.84
IDQ	**0.0013**	**0.0195**	**0.0064**	**0.0245**	**94.88**

D. Run-Time Comparison

Table IV reports the results of run-time comparison on two types of processors. The Intel i7-8700 processor has 8 core and runs at 3.7 GHz. The Kirin 985 has 8 ARM Cortex cores and a Mali-G77 GPU. we test the models using TensorFlow 1.12.0 and Python 3.6 in i7-8700. In Kirin 985, TensorFlow Lite, a framework for on-device inference, is used to deploy the models. SER-FIQ is the most time-consuming since the convolutional layers are run once, but the fully connected layer is run 100 times. The input size of FaceQNet(ResNet50) is 224x224, so it takes longer than PFE(ResNet64) with an input size of 96x96. When the model is large, the ARM CPU has the slowest speed, followed by the Mail-G77 GPU, and the i7-8700 has the fastest speed. For our small model, the run-time can be controlled within a few milliseconds. Among them, ARM CPU has the best performance.

TABLE IV
THE RESULTS (MS) OF RUNTIME COMPARISON.

Methods	i7-8700	Kirin 985 (ARM) #Thread (CPU) 1	2	4	Mali GPU
SER-FIQ(R64)	450.72	-	-	-	-
FaceQNet(R50)	107.31	-	-	-	-
PFE(R64)	48.27	421.32	255.85	178.73	59.37
PCNet(R18)	12.27	95.90	58.44	42.62	16.51
LightQNet	3.98	4.02	3.30	2.57	5.57

IV. CONCLUSION

This paper proposes a lightweight face quality assessment approach based on the pairwise binary quality pseudo-label generated by the face similarity. Identification quality loss with soft-label is used to effectively train the neural network. At the same time, the branch-based quality distillation is used to guide the training of the lightweight model. Experiments demonstrate that LightQNet can outperform state-of-the-art baselines with small model size and low computation complexity.

REFERENCES

[1] S. Bharadwaj, M. Vatsa, and R. Singh, "Biometric quality: a review of fingerprint, iris, and face," *EURASIP journal on Image and Video Processing*, vol. 2014, no. 1, pp. 1–28, 2014.

[2] Y. Shi and A. K. Jain, "Probabilistic face embeddings," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 6902–6911.

[3] 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.

[4] D. Rizo-Rodríguez, H. Méndez-Vázquez, and E. García-Reyes, "An illumination quality measure for face recognition," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 1477–1480.

[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] A. Abaza, M. A. Harrison, T. Bourlai, and A. Ross, "Design and evaluation of photometric image quality measures for effective face recognition," *IET Biometrics*, vol. 3, no. 4, pp. 314–324, 2014.

[7] 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.

[8] H.-I. Kim, S. H. Lee, and Y. M. Ro, "Face image assessment learned with objective and relative face image qualities for improved face recognition," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 4027–4031.

[9] X. Qi, C. Liu, and S. Schuckers, "Boosting face in video recognition via cnn based key frame extraction," in *2018 International conference on biometrics (ICB)*. IEEE, 2018, pp. 132–139.

[10] Z. Lijun, S. Xiaohu, Y. Fei, D. Pingling, Z. Xiangdong, and S. Yu, "Multi-branch face quality assessment for face recognition," in *2019 IEEE 19th International Conference on Communication Technology (ICCT)*. IEEE, 2019, pp. 1659–1664.

[11] J. Rose and T. Bourlai, "On designing a forensic toolkit for rapid detection of factors that impact face recognition performance when processing large scale face datasets," in *Securing Social Identity in Mobile Platforms*. Springer, 2020, pp. 61–76.

[12] G. Aggarwal, S. Biswas, P. J. Flynn, and K. W. Bowyer, "Predicting performance of face recognition systems: An image characterization approach," in *CVPR 2011 WORKSHOPS*. IEEE, 2011, pp. 52–59.

[13] J. Chen, Y. Deng, G. Bai, and G. Su, "Face image quality assessment based on learning to rank," *IEEE signal processing letters*, vol. 22, no. 1, pp. 90–94, 2014.

[14] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, and L. Beslay, "Faceqnet: Quality assessment for face recognition based on deep learning," in *2019 International Conference on Biometrics (ICB)*. IEEE, 2019, pp. 1–8.

[15] P. Grother, A. Hom, M. Ngan, and K. Hanaoka, "Ongoing face recognition vendor test (frvt)–part 5: Face image quality assessment," *Draft NIST Interagency Report*, 2020.

[16] W. Xie, J. Byrne, and A. Zisserman, "Inducing predictive uncertainty estimation for face verification," in *31st British Machine Vision Conference 2020, BMVC 2020, Virtual Event, UK, September 7-10, 2020*. BMVA Press, 2020. [Online]. Available: https://www.bmvc2020-conference.com/assets/papers/0149.pdf

[17] J. Hernandez-Ortega, J. Galbally, J. Fierrez, and L. Beslay, "Biometric quality: Review and application to face recognition with faceqnet," *arXiv preprint arXiv:2006.03298v3*, 2021.

[18] P. Terhorst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, "Serfiq: Unsupervised estimation of face image quality based on stochastic embedding robustness," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5651–5660.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[20] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[21] T. DeVries and G. W. Taylor, "Learning confidence for out-of-distribution detection in neural networks," *arXiv preprint arXiv:1802.04865*, 2018.

[22] Y. Geifman and R. El-Yaniv, "Selectivenet: A deep neural network with an integrated reject option," in *International Conference on Machine Learning*. PMLR, 2019, pp. 2151–2159.

[23] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1314–1324.

[24] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4690–4699.

[25] 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*. Springer, 2016, pp. 87–102.

[26] 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.

[27] T. Zheng and W. Deng, "Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments," *Beijing University of Posts and Telecommunications, Tech. Rep*, vol. 5, 2018.

[28] S. Sengupta, J.-C. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, "Frontal to profile face verification in the wild," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2016, pp. 1–9.

[29] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014.

[30] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A. K. Jain, J. A. Duncan, K. Allen *et al.*, "Iarpa janus benchmark-b face dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 90–98.

[31] M. A. Olsen, V. Šmida, and C. Busch, "Finger image quality assessment features–definitions and evaluation," *IET Biometrics*, vol. 5, no. 2, pp. 47–64, 2016.

[32] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch, "Face image quality assessment: A literature survey," *arXiv preprint arXiv:2009.01103v2*, 2021.

[33] J. Chang, Z. Lan, C. Cheng, and Y. Wei, "Data uncertainty learning in face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5710–5719.

APPENDIX

A. Derivation Analysis of Identification Quality Loss

Assuming $h_q(\mathbf{x})$ is the output of last layer before the Sigmoid activation, $f_q(\mathbf{x}) = Sigmoid(h_q(\mathbf{x}))$, we can take the derivative of loss function in Equation (3):

$$\frac{\partial L^q_{IDQ}}{\partial \theta} = \frac{2}{|\mathcal{P}|}\sum_{i=1}^{M} g_i \frac{\partial h_q(\mathbf{x}_i)}{\partial \theta} + \frac{1}{|\mathcal{P}|}\sum_{i=1}^{M} e_i \frac{\partial h_q(\mathbf{x}_i)}{\partial \theta}$$

$$g_i = \sum_{j=1}^{M} I_{y_i=y_j} I_{f_q(\mathbf{x}_i)<f_q(\mathbf{x}_j)} f_q(\mathbf{x}_i) - \sum_{j=1}^{M} I_{y_i=y_j} I_{f_q(\mathbf{x}_i)<f_q(\mathbf{x}_j)} t_{ij}$$

$$e_i = \sum_{j=1}^{M} I_{y_i=y_j} I_{f_q(\mathbf{x}_i)=f_q(\mathbf{x}_j)} f_q(\mathbf{x}_i) - \sum_{j=1}^{M} I_{y_i=y_j} I_{f_q(\mathbf{x}_i)=f_q(\mathbf{x}_j)} t_{ij}$$

where $I_{f_q(\mathbf{x}_i)<f_q(\mathbf{x}_j)}$ is used to select the network with lower predicted quality and M is the batch size. The second items of g_i and e_i are the sum of all the pairwise quality labels that satisfy the condition. Therefore, in order to make the gradient value of each $\mathbf{x}_i$ more stable, we sample multiple images for each class during training. Table A1 shows that better results can be obtained when the number of sampled images is greater than 4.

TABLE A1
ANALYSIS OF THE NUMBER OF IMAGES SAMPLED FOR EACH CLASS (AUC@FMR=1E3). THE BATCH SIZE IS 128.

#Image Per Class	2	4	8	16	32
CPLFW	0.0266	0.0259	**0.0257**	0.0258	0.0259
CFP-FP	0.0109	0.0104	**0.0098**	**0.0098**	0.0099

B. Additional Experiments and Visualization

Table A2, Table A3 and Fig. A2 compare the results of LightQNet with state-of-the-art methods on more datasets. Table A3 reports the results on IJBB using image-to-image protocal as [16]. The results of PFE, PCNet, DUL-rgs and IDQ are all based on LightQNet with distillation. which once again demonstrates the effectiveness of our method.

The visualization results in Fig. A3 shows that the output score of LightQNet is consistent with the quality of human subjective assessment.

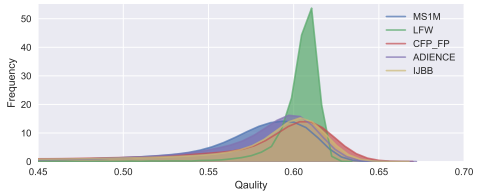


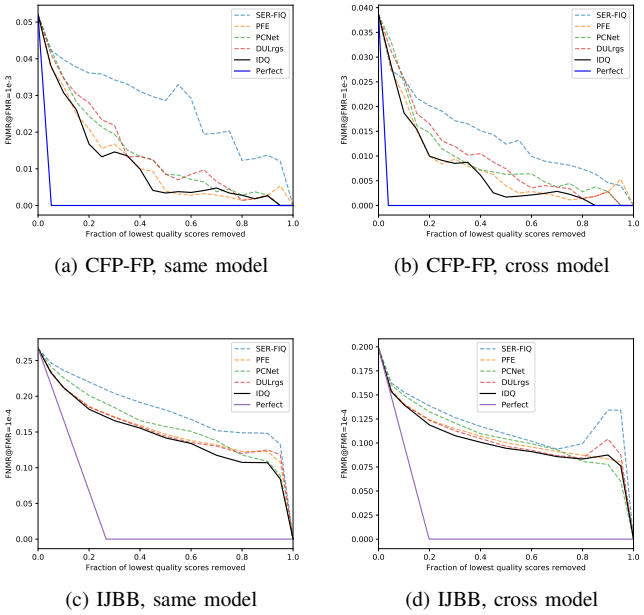
Fig. A1. Distribution of the Quality Score of LightQNet.

TABLE A2
RESULTS (AUC@FMR=1E3) ON MORE DATASETS.

Feats	Methods	SLLFW	CALFW	CFP-FF	Vgg2FP	AgeDB
R64 Same	FaceQNetv1	0.0057	0.0738	0.0045	0.1825	0.0674
	SER-FIQ	0.0046	0.0599	0.0017	0.0800	0.0218
	PFE	0.0030	0.0654	0.0013	0.0714	0.0253
	PCNet	**0.0027**	0.0602	0.0012	0.0738	0.0219
	DUL-rgs	0.0029	0.0610	0.0013	0.0719	0.0229
	IDQ	**0.0027**	**0.0592**	**0.0007**	**0.0710**	**0.0196**
IR100 Cross	FaceQNetv1	0.0053	0.0729	0.0047	0.1118	0.0647
	SER-FIQ	0.0039	0.0602	0.0018	0.0613	0.0159
	PFE	0.0025	0.0659	0.0014	0.0574	0.0196
	PCNet	0.0027	0.0635	0.0011	0.0571	0.0217
	DUL-rgs	0.0024	0.0625	0.0014	0.0596	0.0227
	IDQ	**0.0023**	**0.0604**	**0.0007**	**0.0567**	**0.0142**

TABLE A3
RESULTS ON IJBB USING IMAGE-TO-IMAGE PROTOCAL.

Methods	AUC on same model		AUC on cross model	
	FMR1e-5	FMR1e-4	FMR1e-5	FMR1e-4
SER-FIQ	0.3439	0.1453	0.3532	0.1177
PFE	0.2729	0.1187	0.2718	0.0859
PCNet	0.2237	0.1242	0.2727	0.0873
DULrgs	0.2539	0.1191	0.2615	0.0857
IDQ	**0.1837**	**0.1118**	**0.1678**	**0.0820**



(a) CFP-FP, same model (b) CFP-FP, cross model

(c) IJBB, same model (d) IJBB, cross model

Fig. A2. Error-versus-reject curve on CFP-FP and IJB-B.

Fig. A3. Visualization of IJB-B dataset. The faces are ranked by the quality scores that are predicted by LightQNet.