Supplementary Material of "Deep Transferring Quantization"

Zheng Xie[1,2*], Zhiquan Wen[1*], Jing Liu[1*], Zhiqiang Liu[1,2], Xixian Wu[3], and Mingkui Tan[1†]

[1] South China University of Technology, Guangzhou, China
{sexiezheng,sewenzhiquan,seliujing,sezhiqiangliu}@mail.scut.edu.cn
mingkuitan@scut.edu.cn
[2] PengCheng Laboratory, Shenzhen, China
[3] HuNan Gmax Intelligent Technology, Changsha, China
wuxixian@gmax-ai.com

A Experiments on Face Recognition

A.1 Source and Target Data sets

To further evaluate the proposed deep transferring quantization (DTQ), we conduct experiments on the face recognition task. We use the visible light (VIS) face data set CASIA-WebFace [7] as the source data set, and the PolyU near-infrared ray (NIR) face data set (PolyU-NIRFD) [8] as the target data set. Specifically, CASIA-WebFace contains 494,414 VIS face images from 10,575 different individuals. PolyU-NIRFD contains 38,430 NIR face images from 335 identities. We randomly select 80 identities of PolyU-NIRFD as the validation set and the others as training data. In total, we sample 29,158 images for training and 9,272 images for validation.

A.2 More Implementation Details

We adopt LResNet18E-IR and LResNet34E-IR [1, 2] as the base models. Following the settings in [1, 2], we use the model pre-trained on CASIA-WebFace as the full-precision source model. We use SGD with a mini-batch size of 64, where the momentum term is set to 0.9. The initial learning rate is set to 0.01. We train low-precision models for 9k iterations, and the learning rate is divided by 10 at the 6k-th iteration. Following [4], we also take feature maps from four intermediate layers for attentive feature alignment. Since the task of transferring quantization has not received enough attention from the community, we fail to find very related baselines for comparison. Therefore, we construct three methods for comparison, including L^2-Q, L^2-SP-Q [5] and DELTA-Q [4]. We use true acceptance rate (TAR) at different levels of false acceptance rate (FAR) to measure the performance of the low-precision models for face recognition.

* Authors contributed equally.
† Corresponding author.

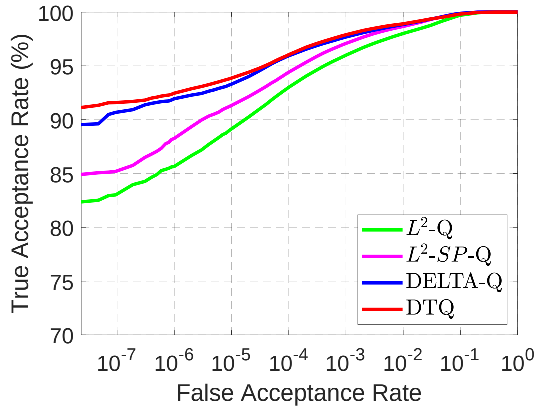

Fig. 1. ROC [3, 6] curves of 4-bit LResNet34E-IR on the PolyU-NIRFD data set

Table 1. Performance comparisons of different methods on the PolyU-NIRFD data set. "FAR" denotes the false acceptance rate

LResNet34E-IR (4 bit)	True Acceptance Rate (%)		
	FAR=1e-5	FAR=1e-6	FAR=1e-7
L^2-Q	89.2	85.7	83.0
L^2-SP-Q	91.3	88.3	85.2
DELTA-Q	93.3	91.9	90.7
DTQ	**93.9**	**92.5**	**91.6**

A.3 More Results on Face Recognition

We quantize LResNet34E-IR to 4-bit and report the results in Table 1 and Fig. 1. Specifically, Fig. 1 presents the receiver operating characteristic (ROC) [3, 6] curves of different methods. From these results, we make the following observations. First, L^2-SP-Q, DELTA-Q and our DTQ achieve much better performance than L^2-Q, which demonstrates the necessity of transfer learning for the transferring quantization task. Specifically, our DTQ outperforms L^2-Q by 8.6% in the TAR at FAR=1e-7. Second, compared with L^2-Q, L^2-SP-Q and DELTA-Q, our DTQ achieves the best performance. Specifically, our DTQ surpasses DELTA-Q by 0.9% in the TAR at FAR=1e-7. These results demonstrate the effectiveness of the proposed DTQ on face recognition.

B More Visualization Results

To further investigate the effect of the losses in the proposed DTQ, we visualize more feature maps of the penultimate layer of 5-bit MobileNetV2 on Caltech 256-30. From Fig. 2, when we only use the cross-entropy (CE) loss, the model fails to focus on the target object. When we add the attentive feature alignment

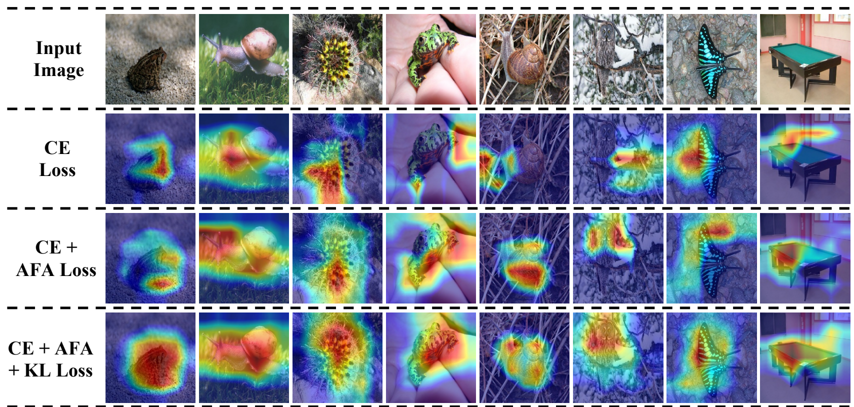

Fig. 2. Visualization of features from models with different losses. CE loss is the cross-entropy loss, AFA loss is the attentive feature alignment loss and KL loss is the Kullback–Leibler divergence loss. Samples are taken from the features of the penultimate layer of 5-bit MobileNetV2 on Caltech 256-30

(AFA) loss, the model achieves significantly better performance. Furthermore, the model equipped with all three losses shows a better concentration on the target object than that equipped with two losses. These visualization results further demonstrate the effectiveness of the proposed losses in our DTQ.

References

1. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: Additive angular margin loss for deep face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 4690–4699 (2019)
2. Deng, J., Guo, J., Zafeiriou, S.: ArcFace: Additive angular margin loss for deep face recognition. arXiv preprint arXiv:1801.07698v1 (2018)
3. Fawcett, T.: An introduction to ROC analysis. Pattern Recognition Letters **27**(8), 861–874 (2006)
4. Li, X., Xiong, H., et al.: DELTA: Deep learning transfer using feature map with attention for convolutional networks. In: International Conference on Learning Representations (2019)
5. Li, X., Grandvalet, Y., Davoine, F.: Explicit inductive bias for transfer learning with convolutional networks. In: International Conference on Machine Learning. pp. 2830–2839 (2018)
6. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: DeepFace: Closing the gap to human-level performance in face verification. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1701–1708. IEEE Computer Society (2014)
7. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv preprint arXiv:1411.7923 (2014)
8. Zhang, B., Zhang, L., Zhang, D., Shen, L.: Directional binary code with application to polyu near-infrared face database. Pattern Recognition Letters (2010)