실용적인 경량 네트워크 얼굴 검증 모델 분석

Laudwika Ambardi, 박인규, 홍성은
인하대학교 정보통신공학과
{laudwika@gmail.com, pik@inha.ac.kr, csehong@inha.ac.kr}

Analysis on Practical Face Verification Models with Lightweight Networks

Laudwika Ambardi, In Kyu Park, Sungeun Hong
Department of Information and Communication Engineering, Inha University

요 약

얼굴 검증 기술은 출입통제 시스템이나 모바일 기기에서의 열람 또는 금융 서비스 등 보안이 요구되는 다양한 분야에서 널리 활용되고 있다. 최근 얼굴 검증 분야에서 높은 성능 향상을 보인 대부분의 검증 모델은 깊은 네트워크를 사용하므로 상대적으로 매우 큰 컴퓨팅 파워를 요구한다. 따라서 해당 모델들을 실환경에 적용하기 위해서는 모델 경량화 기술에 대한 고려가 반드시 필요하다. 얼굴 검증 연구에서 경량화 기술의 중요성에도 불구하고 해당 연구는 이제까지 잘 다뤄지지 않았다. 본 논문은 주요 얼굴 검증 모델에 대해서 지식 증류 기술을 수행하고, 이에 따른 실험 결과를 비교 분석하여 제시함으로써 경량화 기술 적용에 대한 방향성을 제시한다.

1. Introduction

Deep learning methods have taken over the digital world by storm. Entertainment, self-driving cars, and even medical applications have started to implement deep learning. With all these advancements in deep learning, security in biometrics is also progressing by leveraging deep learning techniques, e.g., fingerprint recognition, iris recognition, face recognition, and other biometric-based recognition. Although fingerprint recognition used to be very popular, recognition [1] is more commonly used now.

Face recognition has many applications in all sorts of forms, such as face verification to unlock your phone or even face identification to track down a criminal. It is common for everyday users to have a face verification lock on their devices. While previous handcrafted or machine learning techniques have worked, it takes quite some time for inference to be performed. Because of this, deep learning research in face recognition has risen in the past few years, reaching high accuracy in the process.

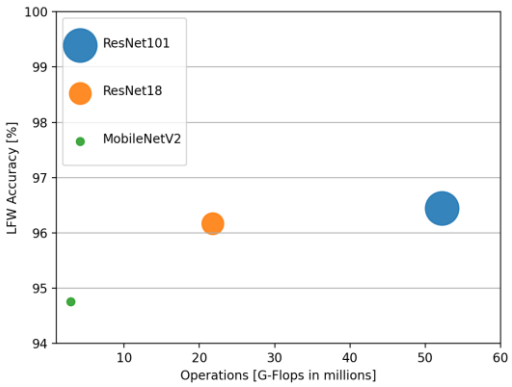

Figure 1: Total number of operations compared to the LFW verification Accuracy.

Table 1: Verification Accuracy (%)

Metric / Model	LFW			CasiaWebFace		
	Vanilla	SL KD	ADV KD	Vanilla	SL KD	ADV KD
ResNet101	96.4	-	-	91.3	-	
ResNet18	96.0	95.7	94.92	88.7	87.3	85.09
MobileNet	94.6	-	95.0	85.1	-	85.8

Table 2: Evaluation on practical metrics

Metric / Model	Inference Time (ms)	Memory Usage (GB VRAM)	Model Size ma(MB)
ResNet101	48.0 ± 1.7	1.4	204
ResNet18	12.0 ± 0.9	1.2	85
MobileNet	12.3 ± 1.33	1.1	14

Although face verification in research prioritize accuracy, which is great to drive progress, they cannot be easily implemented in everyday devices. The reason being these models are encumbered by the many layers of the deep networks. Causing great performance at the cost of computational power and time. However, when we transfer the knowledge of these models to a lightweight model, we can have a smaller and more efficient model without the heavy cost of computational power and time.

Existing face verification models used in research are encumbered as they are deep networks with many layers. These models were made to be the best in terms of accuracy with the cost of computational power. as mentioned in by Zhou. Et al. [2], while other papers use lightweight CNN, they rarely mention the architecture of the lightweight models, making a comparison unfair in the process. In turn, while the mainstream face verification techniques report high accuracy, the lightweight models are not specified enough to perform simple inference on, making practical uses difficult to perform for both researchers and commercial use.

In this paper, we discuss about using a few knowledge distillation techniques to train multiple lightweight networks from a teacher network. We then evaluate each network in terms of accuracy inference time, computational power, and model size.

2. Proposed Approach

As a first step, we use the soft-target knowledge distillation [3] and the adversarial discriminative knowledge distillation [4] to train our students networks. The soft-target knowledge distillation works by comparing

the final fully connected layer to penalize the student target to be the same as the teacher model; this forces the feature extractor to get the same results as the teacher network.

While the adversarial knowledge distillation utilizes a discriminator to compare the feature maps between the teacher and student model. This punishes the feature extractor layers instead of the final fully connected layer. By utilizing this method, the student model is forced to learn to be as close as possible to the teacher model.

3. Experiments

3.1. Implementation Details

Face verification backbone. We train all our models to be a feature extractor outputting a feature vector of 512. All our networks utilize ArcFace [5] as the loss of our network, as it can be implemented as the final layer of the models and can be easily removed to use the feature vector for inference

Teacher Network. As our method utilizes the offline method of knowledge distillation, we first train a ResNet101 network to be used as the teacher model for the lightweight networks

Student Networks. We train as normal then knowledge distill to evaluate on two lightweight networks which are ResNet18 and MobileNet **Datasets.** For our training phase we use MS1M-Ibug for all the networks. As for our evaluation, we use LFW[6] and CasiaWebface [7] for accuracy. Then finally we test on a precollected video dataset to evaluate computational power and average inference time.

Evaluation. We evaluate each network by measuring the overall verification accuracy and the verification false acceptance rate of each dataset as well as the inference

time, computational power, and model size to find a more appropriate network for practical use.

3.2. Evaluation results

In figure 1 we can see that the accuracy of the models compared to the number of operations. ResNet101 with the larger network outperforms the lightweight models. But while MobileNet has 30x less operations than ResNet101, it still has comparable results.

In Table 1, we first see the overall accuracy of verification performance. The teacher model outperforms all the student models. And the ResNet18 model decreases in accuracy while MobileNet increases after applying knowledge distillation. Although MobileNet with the adversarial knowledge distillation increased in accuracy, it failed to converge when using a soft label knowledge distillation method, making a comparison unfair.

We then analyze using such models for practical use. (see Table 2). Although ResNet18 needs more power and has a larger model size, it performs slightly faster than MobileNet. While these two models are quite similar in terms of speed, we can see that compared to the teacher model it has a significant increase making lightweight models much more suitable for practical uses.

4. Conclusion

While the overall verification accuracy may be high, it should not be enough to evaluate for practical uses. Evaluation on the overall practical metrics is needed to see which models are suitable to use. We can conclude that although the MobileNet model has fewer operations, smaller model size and Vram usage, the overall accuracy is comparable to ResNet101 and ResNet18, while adding knowledge distillation methods help increase this model's performance. In the future we may extend our work to compare more lightweight models and model compression techniques to give a better standard for lightweight verification models.

Acknowledgements

This work was supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government (MSIT) (2017-0-00142, Development of Acceleration SW Platform Technology for Ondevice Intelligent Information Processing in Smart Devices and 2020-0-01389, Artificial Intelligence Convergence Research Center (Inha University))

References

[1] L. Weiyang, Y, Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition" Proc. of the IEEE conference on computer vision and pattern recognition, 2017, pp. 6738-6746.

[2] H. Zhou, J. Liu, Z. Liu, Y. Liu, X. Wang, "Rotate-and-Render: Unsupervised Photorealistic Face Rotation from Single-View Images", In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2020.

[3] G. Hinton, O. Vinyals, and JJ. Dean, "Distilling the Knowledge in a Neural Network", NIPS Deep Learning and Representation Learning Workshop, 2015.

[4] I. Chung, S. Park, J. Kim, and N. Kwak, "Feature-map-level Online Adversarial Knowledge Distillation" In Proc. of the International Conference on Machine Learning, 2020.

[5] J. Deng, J. Guo, N. Xue, and S. Zefeiriou, "Arcface: Additive angular margin loss for deep face recognition" In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 4690-4699. Computer Vision and Pattern Recognition, 2018, pp. 4510-4520

[6] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.", Technical Report, 2007, 07-49.

[7] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning Face Representation from Scratch", arXiv:1411.7923, 2014.