

Research on a Lightweight Deep Learning Model Suitable for Face Recognition for Mobile Devices

Sang-Hun Lee*, Il-Yong Chun**, Yeung-Hak Lee***, and Sang-Hee Park****

This work was supported by Korea Planning & Evaluation Institute of Industrial Technology funded by the Ministry of the Interior and Safety (MOIS, Korea) (grant number 20025207)

Abstract

Recently, research on lightweight deep learning has been applied to various fields due to issues such as cost reduction, security, and power consumption due to decentralization. The lightweight deep learning model provides distributed processing of data and various services through a mobile environment. In this study, we compare two lightweight facial recognition deep learning models suitable for the mobile environment and propose a more suitable model. MobileFaceNet is a model optimized for deployment in an embedding environment, and we sought to find a more suitable model by comparing it with the ResNet model that has been recently studied. WebFace42M was used as the dataset, and landmarks were extracted using RetinaFace as a face alignment technique, and faces were aligned using opencv's affine transformation. As a result of applying the two models, ResNet-100 showed better performance in the same embedding environment.

요약

최근 경량 딥러닝에 대한 연구는 탈중앙화에 따른 비용 절감, 보안, 전력 소모 등의 이슈로 인해 다양한 분야에 적용되고 있다. 경량화된 딥러닝 모델은 모바일 환경을 통해 데이터의 분산처리와 다양한 서비스를 제공한다. 본 연구에서는 모바일 환경에 적합한 두 가지 경량 얼굴 인식 딥러닝 모델을 비교하고 보다 적합한 모델을 찾고자 한다. MobileFaceNet은 이를 임베딩 환경에 배포하는데 최적화된 모델로 최근 연구되고 있는 ResNet모델과 비교하여 좀 더 적합한 모델을 찾고자 하였다. 데이터셋은 WebFace42M를 사용하였으며, 얼굴 정렬 기법으로는 RetinaFace를 이용하여 랜드마크를 추출하고, opencv의 affine 변환을 이용하여 얼굴을 정렬하였다. 두 모델을 적용한 결과, 동일한 임베딩 환경에서 ResNet-100이 더 나은 성능을 보였다.

Keywords

embedding environment, lightweight deep learning, face recognition, MobileFaceNet, ResNet

* Principal Researcher, Gumi Electronics & Information Technology Research Institute

- ORCID: <https://orcid.org/0000-0002-2195-4415>

** Executive Director, Metsakuur Company

- ORCID: <https://orcid.org/0009-0002-4861-0134>

*** Professor, SW Convergence Education Centre, Andong National University

- ORCID: <https://orcid.org/0000-0002-9037-2646>

**** Professor, Dept. of Speech & Language Pathology, Daegu Cyber University

- ORCID: <https://orcid.org/0000-0002-0083-238X>

· Received: May 17, 2024, Revised: Dec. 03, 2024, Accepted: Dec. 06, 2024

· Corresponding Author: Sang-Hee Park

Dept. of Speech & Language Pathology, Daegu Cyber University Korea

Tel.: +82-53-859-7451, Email: psh4292@dcu.ac.kr

I. Introduction

Lightweight deep learning research is an essential technology for embedding existing cloud-based learning models into lightweight devices, providing various benefits such as reduced latency, protection of sensitive personal information, and reduced network traffic[1]. Because centralized cloud-based GPU-based computing resources have high costs and power consumption, decentralization allows efficient use of computing resources while performing desired functions by developing and applying lightweight deep learning models in an embedded environment capable of distributed processing. In addition, in the field of facial recognition security, non-invasive biometric methods such as facial recognition reduce the risk and difficulty of processing confidential biometric data. It also simplifies the task of providing security while ensuring accurate results. Integrating biometric systems with edge computing and deep learning reduces latency and bandwidth usage, making the system more robust and dynamic[2].

In this study, we compare two lightweight deep learning models for face recognition suitable for the mobile environment and propose a more suitable model.

MobileFaceNet is a model optimized to deploy it in embedding environment and is used by many companies as a commercial face recognition model[3].

On the other hand, ResNet, which is being compared in this paper, is known as a network that shows high performance by sequentially stacking residual blocks through skip connection. In addition, ResNet-100 was used to simultaneously consider face recognition performance, speed, and memory through learning and evaluation.

In this paper, we compare the performance between the two models and propose a lightweight deep learning model suitable for the mobile environment with improved recognition accuracy for face recognition and security.

This paper is organized as follows. In Section 2,

we describe the dataset and preprocessing to be used in the experiment. In Section 3, we examine the characteristics of the two models compared in this paper. Finally, in Section 4, we present the experimental results between the two models and suggest a model that is more suitable for the mobile environment.

II. Data Preprocessing

In this chapter, we will explain the learning data set selection and preprocessing methods, which are important for the performance of face recognition models.

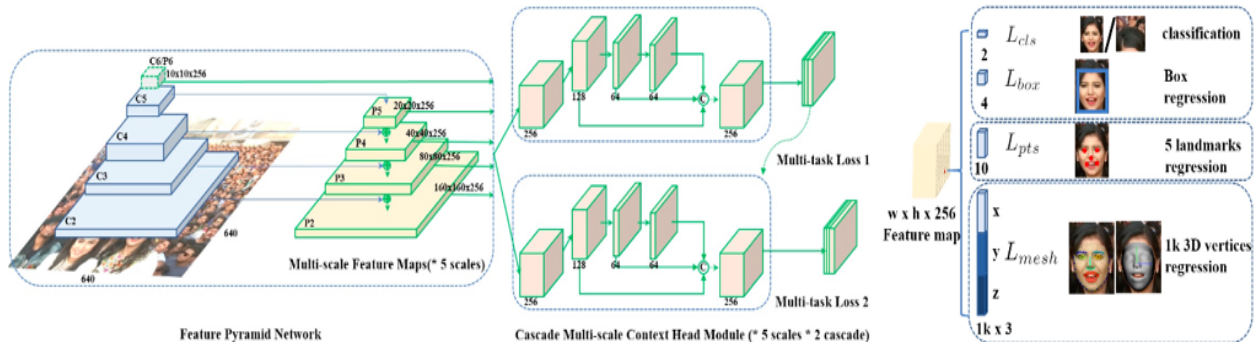
2.1 Dataset

Selection of a learning data set to train a face recognition model plays an essential role in analyzing face recognition performance.

Popular evaluations of face recognition, including the LFW family, CFP, AgeDB, RFW, MegaFace, and IJB families, mainly aim at pursuing accuracy, which has recently reached almost saturation[4].

In real application systems, facial recognition is always limited by inference time, such as unlocking a phone with a seamless experience. The lightweight face recognition challenge takes a step toward this goal, but ignores the time cost of detection and alignment.

In this paper, WebFace42M was used as the learning data set. WebFace42M is a refined dataset of WebFace260M and is the largest public facial recognition training set. WebFace42M is equipped with a distributed training framework to perform efficient optimization. It is also actively maintained and updated, as well as consisting of a test set with rich properties capable of activating the Face Recognition Under Inference Time Constraint (FRUITS) protocol, including evaluating different ages, genders, races, and scenarios[4].



(a) Network structure of RetinaFace (b) Multi-task loss of RetinaFace
 Fig. 1. RetinaFace structure used for five landmarks[5]

2.2 Data preprocessing

Various methods for face detection and alignment have been proposed for preprocessing of learning data.

As shown in Fig. 1., An overview of the single-shot multi-level face localisation approach(a). Detailed illustration of RetinaFace design(b). RetinaFace is designed based on the feature pyramids with five scales. For each scale of the feature maps, there is a deformable context module. Following the context modules, we calculate a joint loss (face classification, face box regression[5], five facial landmarks regression and 1k 3D vertices regression) for each positive anchor. To minimise the residual of localisation, we employ cascade regression[3]. five landmarks (eye centers, nose tip, mouth corners) were extracted through RetinaFace[5], and the faces were aligned using the affine conversion function provided by opencv[6].

learning complex patterns or transformations.

The structure of MobileFaceNet is as follows, where t represents the expanded size within the inverted residual block, c represents the value of the output channel, n represents the number of repetitions, and s represents the stride value.

As shown in Fig. 2, The first layer, conv2d, consists of a general 3x3 convolution, and the layers in between have a bottleneck structure called an inverted residual block. For the last layer, conv2d goes through 1x1 and avgpool and ends in linear form[3].

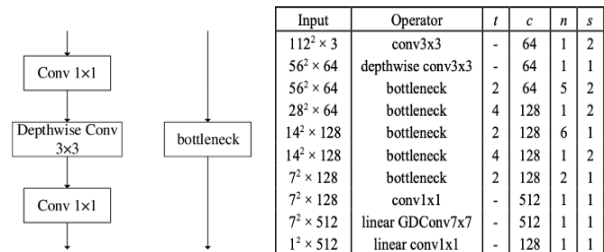


Fig. 2. Structure of MobileFaceNet

III. Network Comparison

3.1 MobileFaceNet

MobileFaceNet[7] is a lightweight neural network optimized for mobile and embedded systems. It is based on the MobileNet architecture and uses depthwise separable convolutions to reduce the number of parameters and the amount of computation to provide fast inference speed. However, it is limited in

3.2 ResNet

ResNet[7] features residual connections, which were introduced to solve the vanishing gradient problem that can occur in deep neural networks, and has an excellent ability to learn complex features through deep layers. Because it includes more layers and parameters, it is possible to recognize more elaborate patterns and transformations, but it has the disadvantage of being relatively slow.

ResNet has a much deeper structure than MobileFaceNet, which allows it to learn more complex features and achieve higher performance in face recognition.

In addition, ResNet is a network that shows high performance by sequentially stacking residual blocks through skip connection. High performance is achieved by sequentially stacking blocks, but as the amount of computation and number of parameters increases, there is a trade-off relationship between performance and speed/memory. Therefore, SE-LResNet100E-IR was applied as the backbone by simultaneously considering face recognition performance and speed/memory through learning and evaluation.

SE-LResNet100E-IR uses IR Block and SE Block instead of Bottleneck Block of existing ResNet in ResNet-100, the input layer is conv3x3 stride 1, and the output layer applies Residual Block to BN-Dropout-FC-BN(Fig. 3)[8].

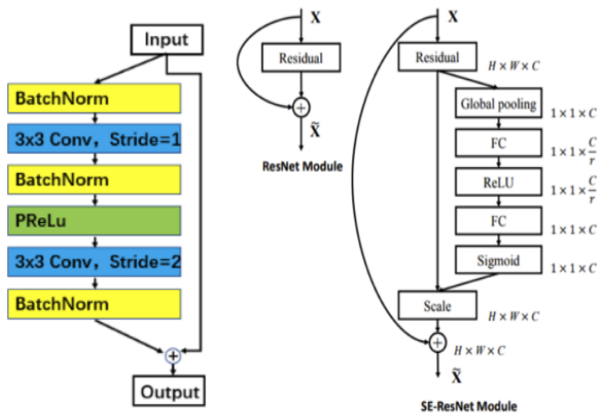


Fig. 3. Schema of residual unit(left) and the SE-ResNet module(right).

IV. Result

This paper used the WebFace42M dataset, which is the largest training set for face recognition. For face detection and alignment, five landmarks were extracted through RetinaFace and the faces were aligned using the affine conversion function provided by openCV.

In experiment result, as shown in Table 1, the performance of the ResNet-100(SE-LResNet100E-IR) model is slightly better. The biggest difference was shown in 'Recall rate', which is the part that accurately recognizes the existing face value in facial recognition security and is the most important part in biometric information analysis. 'F-1 score' and 'Accuracy' showed the next difference, and 'Precision' and 'Specificity' showed a slight difference in the performance for misrecognition.

V. Conclusion

There has been a lot of development in the technology of face recognition. In particular, improving the accuracy of Edge AI is an important part of facial recognition security, and it can simplify the task of providing security through accurate results[2]. In this paper, the performance comparison results between the two models show that the ResNet-100 (SE-LResNet100E-IR) model shows slightly better performance in Accuracy, Precision, Recallrate, Specificity, and F-1 score. This is expected to be more meaningful in limited embedded environments.

Table 1. Performance comparison between two models

Backbone	Accuracy(%)	Precision(%)	Recallrate(%)	Specificity(%)	F-1 score(%)
MobileFaceNet	99.6833	99.8996	99.4667	99.9000	99.6827
ResNet-100 (SE-LResNet100E-IR)	99.8500	100.0000	99.7000	100.0000	99.8498

We did not consider computational efficiency through model optimization in the lightweight deep learning model.

There is no data that directly compares the hardware usage environment, such as computational resources and memory usage, between the two models presented in this paper, and this requires additional research. It is also necessary to study the optimization of the model structure by considering the number of parameters of the lightweight deep learning model, FLOPs (Floating point Operations Per second), etc. [9], so in the future, we plan to study a model that balances computational efficiency and accuracy well through model structure optimization by considering various embedded environments.

References

- [1] Recent R&D Trends for Lightweight Deep Learning, https://ettrends.etri.re.kr/ettrends/176/0905176005/34-2_40-50.pdf [accessed : Feb. 09. 2023]
- [2] C. G. Raju, B. U. Matha, V. Canamedi, S. Sarkar and Radhika K. R, "Facial Recognition Using Edge-Driven Biometric System", 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, pp. 1107-1113, May 2022. <https://doi.org/10.1109/ICAAIC53929.2022.9792659>.
- [3] S. Chen, Y. Liu, G. Xiang, and H. Zhen, "MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices", Biometric Recognition (CCBR 2018), Vol 10996, pp. 428-438, Aug. 2018. https://doi.org/10.1007/978-3-319-97909-0_46.
- [4] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, and X. Chen, "WebFace260M: A Benchmark Unveiling the Power of Million-Scale Deep Face Recognition", 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp. 10487-10497, Jun. 2021. <https://doi.org/10.1109/CVPR46437.2021.01035>.
- [5] J. Deng, J. Guo, E. Ververas, Il Kotsia, and S. Zafeiriou, "RetinaFace: Single-shot Multi-level Face Localisation in the Wild", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 5203-5212, Jun. 2020. <https://doi.org/10.1109/CVPR42600.2020.00525>.
- [6] OpenCV Affine Transformations, https://docs.opencv.org/4.x/d4/d61/tutorial_warp_affine.html [accessed : May 13. 2024]
- [7] J. N. Kolf, F. Boutros, J. Elliesen, M. Theuerkauf, N. Damer, and M. Alansari, "EFaR 2023: Efficient Face Recognition Competition", 2023 IEEE International Joint Conference on Biometrics (IJCB), Ljubljana, Slovenia, pp. 1-12, Sep. 2023. <https://doi.org/10.1109/IJCB57857.2023.10448917>.
- [8] J. Deng, J. Guo, N. Xue and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition", 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp. 4685-4694, Jun. 2019. <https://doi.org/10.1109/CVPR.2019.00482>.
- [9] F. Boutros, N. Damer, M. Fang, F. Kirchbuchner and A. Kuijper, "MixFaceNets: Extremely Efficient Face Recognition Networks", 2021 IEEE International Joint Conference on Biometrics (IJCB), Shenzhen, China, pp. 1-8, Aug. 2021. <https://doi.org/10.1109/IJCB52358.2021.9484374>.

Authors

Sang-Hun Lee



2011. 3 : PhD degrees,
Department of Electronic
Engineering, Yeungnam
University
2012. 5 ~ Present : Principal
Researcher, Gumi Electronics &
Information Technology Research

Institute

Research interests : Medical Artificial Intelligence and
Lightweight Deep Learning

Il-Yong Chun



1993. 3 : BL degrees, Department
of Law, Yonsei University
2022. 7 ~ Present : Executive
Director, Metsakuur Company
Research interests : Application
architecture of Vision AI,
Software Architecture.

Yeung-Hak Lee



2003. 8 : PhD degrees,
Department of Electronic
Engineering, Yeungnam
University
2019. 7 ~ Present : Professor, SW
Convergence Education Centre,
Andong National University

Research interests : Image Processing, Pattern
Recognition, Artificial Intelligence, Robot Vision

Sang-Hee Park



2003. 3 : PhD degrees,
Department of Rehabilitation
Science, Daegu University
2006. 3 ~ Present : Professor,
Department of Speech &
Language Pathology, Daegu Cyber
University

Research interests : Rehabilitation of Cochlear implant
& Articulation Disorder, Rehabilitation using AI
systems