Check for updates

Journal of KIIT. Vol. 23, No. 4, pp. 11-19, Apr. 30, 2025. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2025.23.4.11

EfficientNet 기반 얼굴 영상 나이 인식 방법

최성은*, 김익재**

Age Estimation Method of Facial Image based on EfficientNet

Sung Eun Choi*, Ig-Jae Kim**

요 약

딥러닝 기술의 발전에 따라 얼굴영상의 나이인식 분야에서도 딥러닝 기반의 연구가 활발히 진행되고 있다. 특히 영상인식 분야에서 우수한 성능을 보이는 CNN(Convolutional Neural Networks) 기반의 딥러닝 기술이 얼굴영상의 나이인식 분야에서 많이 활용되고 있다. 본 논문에서는 기존 CNN의 성능을 결정하는 주요요소인 모델의 깊이(Depth), 너비(Width) 및 해상도(Resolution)를 체계적으로 설정하여 성능을 향상시킨 EfficientNet을 기반으로 한 얼굴영상의 나이인식 방법을 제안하였고, 얼굴의 나이 정보를 보다 잘 인식하기 위해 평균-분산 손실함수를 이용하여 EfficientNet을 학습하였다. 실험 결과, 본 논문에서 제안하는 EfficientNet 기반의 얼굴 나이 인식 성능은 MAE 2.85로 기존 CNN 모델 ResNet34의 성능 MAE 3.02에 비해 우수함을 확인하였다.

Abstract

With the advancement of deep learning technology, research on facial age estimation using deep learning has been actively conducted. In particular, Convolutional Neural Networks(CNNs), which have demonstrated outstanding performance in image recognition, are widely applied in this field. In this paper, we propose a facial age estimation method leveraging EfficientNet, which optimally enhances CNN performance. EfficientNet improves CNN effectiveness through a compound scaling method that systematically adjusts model depth, width, and resolution. Additionally, we trained EfficientNet using a mean-variance loss function to further enhance its performance. Experimental results confirm that the proposed EfficientNet-based approach outperforms existing CNN models in facial age estimation.

Keywords
facial age estimation, CNN, EfficientNet, mean-variance loss

* 한양여자대학교 빅데이터과 교수(교신저자)
- ORCID: https://orcid.org/0009-0001-6702-9719
** 한국과학기술연구원 AI.로봇연구소
- ORCID: https://orcid.org/0000-0002-2741-7047

· Received: Oct. 10, 2024, Revised: Apr. 20, 2025, Accepted: Apr. 23, 2025
· Corresponding Author: Sung Eun Choi
Dept. of Big Data, Hanyang Women's University
200, Salgoji-gil, Seongdong-gu, Seoul, 04763, Korea
Tel.: +82-2-2290-2386, Email: choise28@gmail.com

Ⅰ. 서 론

얼굴 영상을 기반으로 나이를 인식하는 기술이 다양한 분야에서 활용이 증가됨에 따라 그에 대한 연구가 활발히 진행되고 있다. 얼굴 영상 기반의 나이 인식 기술은 고객 정보 관리(ECRM)에서 연령대별 고객 정보를 자동으로 분석하거나, 얼굴 영상의 과거 또는 미래의 모습을 생성하는 나이 변환 분야에 적용되어 장기 실종자 또는 미아 찾기에 활용되고 있다. 또한 얼굴 영상으로 개인의 신원을 인식하는 얼굴 인식 기술에 적용되어 얼굴 인식 성능을 향상시키는 역할을 수행하고 있다[1].

딥러닝 기술의 발전에 따라 얼굴 영상의 나이 인식 분야에서도 딥러닝 모델의 활용이 증가하고 있다. 특히 딥러닝 모델 중 영상 데이터에서 우수한 성능을 보여주는 AlexNet[2], VGG[3], ResNet[4] 등의 CNN 기반 모델이 얼굴 영상 나이 인식 분야에도 많이 활용되었다. 영상 인식 분야에서 활발히 사용되고 있는 CNN의 성능을 향상시키기 위한 많은 연구도 함께 수행되고 있다. 기존의 많은 연구에서 CNN의 영상 인식 성능을 향상시키기 위해 성능을 결정하는 주요 요소인 모델의 깊이, 너비, 해상도를 조정한 다양한 모델들을 제안하였다[4]-[7]. 이러한 방법 중 EfficientNet[7]은 CNN 모델의 성능을 결정하는 주요 요소들을 체계적이며 복합적인 스케일링 방법(Compound scaling method)으로 설정하여 한정된 연산량을 사용하며 성능을 효율적으로 향상시켜 영상 인식 분야에서 우수한 성능을 보여주고 있다. 기존 얼굴 영상의 나이 인식에서는 AlexNet[2], VGG[3], ResNet[4] 등의 CNN 모델이 주로 활용되었고, EfficientNet을 활용한 연구가 수행되지 않았다. 나이 인식 성능을 보다 향상시키기 위해 영상 인식 분야에서 우수한 성능을 보여주는 EfficientNet을 얼굴 영상의 나이 인식 분야에 적용하고 기존 CNN 방법과 비교하는 연구가 필요하다. 따라서 본 논문에서는 EfficientNet을 기반으로 얼굴 영상의 나이를 인식하는 방법을 제안하였고, EfficientNet이 얼굴 영상의 나이 특징 분포를 보다 잘 학습할 수 있도록 소프트맥스 손실함수와 평균-분산 손실함수(Mean-variance loss)[8]를 함께 사용하여 학습하였다. 또한 제안 방법의 성능을 기존 다양한 CNN 기반 얼굴 나이 인식 방법의 성능과 비교하는 연구를 수행하였다.

Ⅱ. 관련 연구

2.1 딥러닝 기반 얼굴 나이 인식 기술

딥러닝 모델이 영상 인식분야에서 뛰어난 성능을 보여준 이후로 얼굴 영상의 나이 인식 연구에서도 딥러닝 모델이 활발히 활용되고 있으며, 특히 딥러닝 기술 중 영상 인식 분야에서 우수한 성능을 보여주는 CNN 모델 기반의 연구가 활발히 진행되고 있다. CNN을 활용하여 얼굴 영상으로부터 나이 특징을 추출하고 추출된 특징을 이용하여 나이를 인식하는 과정을 통째학습(End-to-end learning)함으로써 나이 인식 성능이 크게 향상되었다. W. Shen et al.[9]은 CNN과 Regression forest를 결합한 DRFs(Deep Regression Forests) 기반 얼굴 나이 인식 방법을 제안하여, 얼굴 나이 인식 문제를 비선형 회귀(Non-linear regression)문제로 정의하고 해결하였다. R. Rothe et al.[10][11]는 CNN 모델 중 VGG16[3] 모델을 기반으로 하는 DEX 방법을 제안하였다. X. Liu et al.[12]는 얼굴 나이 인식 모델의 경량화에 중점을 둔 연구를 진행하였다. 경량화 모델인 SuffleNetV2과 어텐션(Attention) 기법을 결합한 MA-SFV2(Mixed Attention-ShuffleNetV2)의 얼굴 나이 인식 모델을 제안하였다. C. Zhang et al.[13]도 모델의 경량화를 통해 모바일 환경에서도 동작할 수 있는 C3AE라는 얼굴 나이 인식 방법을 제안하였다. C3AE는 나이 분포(Age label distribution)와 나이 정보(Age label)를 캐스케이드(Cascade) 방식으로 인식하여 나이 인식 성능을 개선하였다. H. Pan et al.[8]는 CNN 모델로 얼굴 나이 인식을 수행할 때 나이 분포(Age distribution)를 보다 잘 학습할 수 있도록 소프트맥스(Softmax)와 함께 평균-분산 손실함수를 사용하는 방법을 제안하였다. 일반적으로 분류에 사용하는 소프트맥스 로스만을 사용하는 것에 비해 평균-분산 로스를 함께 사용하여 CNN을 학습할 때 나이 인식 성능이 더 향상되는 결과를 보여주었다.

I. Dagher et al.,[14]는 사전 학습된 CNN 모델을 기반으로 계층적으로 나이를 인식하는 모델을 제안하였고, G. George et al.,[15]은 기존 CNN 모델에서 변형된 모델을 사용하여 얼굴 영상의 연령대를 예측하는 모델을 제안하였다. B. Abirami et al.,[16]는 CNN과 Haar Cascades, caffenet을 이용하여 실시간으로 얼굴 영상에서 얼굴을 검출하고 나이와 성별을 인식하는 모델을 제안하였다. S. E. Choi[17]는 다양한 CNN 모델을 활용한 얼굴 나이 인식 방법들의 성능을 비교하는 연구를 수행하였다. S. E. Bekhouche et al.,[18]는 기존의 얼굴 나이 인식 방법에서 사용하는 제한된 특징 추출 및 비효율적인 계층적 표현 학습을 개선한 MSDNN(Multi-Stage Deep Neural Network)라는 모델을 제안하였다. Q. Zhao et al.,[19]는 연령대 분류와 정확한 나이 인식을 위한 계층적 딥러닝 모델을 설계하여 나이를 인식하는 방법을 제안하였다. 이와 같이 기존의 연구에서 다양한 CNN 모델을 기반으로 얼굴 영상의 나이를 인식하는 연구들이 수행되었다.

이와 같이 얼굴 나이 인식을 위한 다양한 연구가 활발히 이루어지고 있음에도 불구하고, 보안 등 실제 응용 분야에서의 활용을 확대하기 위해서는 기존 연구보다 더 높은 성능과 정확도를 제공할 수 있는 기술에 대한 연구가 필요하다.

2.2 EfficientNet

EfficientNet은 기존 CNN 모델의 성능에 영향을 주는 깊이, 너비, 해상도를 동시에 조정하는 복합 스케일링 방법을 사용하여 최적의 성능을 얻기 위해 제안된 모델이다. 이전의 연구들에서는 CNN의 성능 향상을 위해 깊이, 너비, 해상도를 개별적으로 조정하는 방식을 사용했으나, EfficientNet은 복합 스케일링 방법을 통해 세 요소를 효율적으로 조정하여 CNN의 성능을 최적화하였다.

CNN의 성능을 결정하는 깊이, 너비 및 해상도를 그림 1에서 보여준다. 모델의 깊이는 네트워크의 깊이를 의미하여, 네트워크를 구성하는 층(Layer)의 수를 나타낸다. 일반적으로 네트워크의 깊이가 깊어지도록 층을 늘리는 방법으로 설계하여 CNN의 성능을 향상시킬 수 있다. 이와 같이 네트워크의 깊이를 조정하여 성능을 향상시킨 대표적인 방법이 앞에서 언급한 ResNet이다. 네트워크의 깊이가 깊을수록 더 풍부하고 복잡한 특징을 추출할 수 있기 때문에 네트워크의 성능이 향상되지만, 깊이가 너무 깊어지면 기울기 소실(Gradient vanishing) 문제가 발생하여 네트워크를 학습하기 어려운 문제가 있다.

CNN 성능에 영향을 주는 요소인 너비는 네트워크의 채널(Channel) 수를 의미한다. 너비를 넓히는 것은 네트워크의 채널수를 증가시키는 것이고, 채널수가 증가하면 모델의 성능이 향상될 수 있다. 일반적으로 너비가 넓은 네트워크에서는 보다 세밀한 특징(Fine-grained feature)을 잘 찾아내고 학습할 수 있지만, 네트워크의 깊이가 증가하지 않고 너비만 증가하면 깊은 층에서 학습된 특징(High level feature)을 얻기 어려워지는 문제가 있다. 또 다른 요소인 해상도는 CNN의 입력 영상의 해상도를 의미하며, 일반적으로 입력 영상의 해상도가 클수록 영상의 세밀한 패턴(Fine-grained pattern)을 활용할 수 있다.

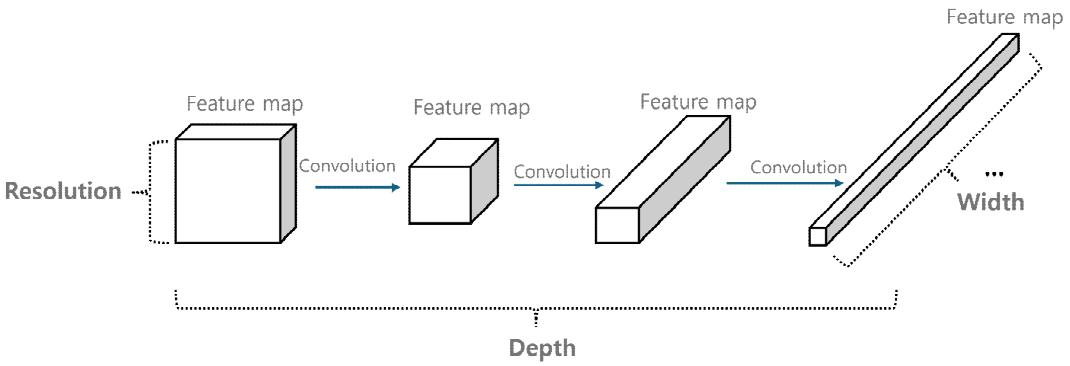

그림 1. CNN 모델의 깊이(depth), 너비(width), 해상도(resolution)
Fig. 1. Depth, width, resolution of CNN

그러나 영상의 해상도 증가에 비례하여 네트워크의 성능이 증가하지는 않는다.

이와 같이 네트워크의 깊이, 너비, 해상도의 관계를 고려하지 않고 개별적으로 조정하여 CNN의 성능을 향상시키는 방법은 한계가 있다. 따라서 EfficientNet에서는 세 요소의 상호관계를 고려하여 최적의 성능을 얻을 수 있는 요소의 조합을 찾는 복합 스케일링 방법을 제안하였다. 복합 스케일링 방법은 동일한 연산량 조건에서 최적의 깊이, 너비, 해상도 변수의 조합을 찾는 방법이다. 일반적인 컨볼루션(Convolution) 연산의 연산량은 모델의 깊이가 α배 증가하면 α배 증가하고, 너비가 β배 증가하면 β^2배 증가하며, 해상도가 γ배 증가하면 γ^2배 증가한다. 이러한 특징을 기반으로 복합 스케일링 방법을 다음과 같은 식 (1)로 표현할 수 있다.

$$\begin{aligned} & d = \alpha^{\phi}, w = \beta^{\phi}, r = \gamma^{\phi} \\ & s.t. \quad \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\ & \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \end{aligned} \tag{1}$$

위 수식의 d는 깊이, w는 너비, r은 해상도를 의미하며, Φ는 복합 계수(Compound coefficients)로 그 값이 커질수록 모델의 크기가 커지게 되는 사용자가 설정하는 변수이다. α, β, γ는 상수값으로 작은 그리드 서치(Small grid search)에 의해 결정된다. 앞 식에 의해, 모델의 깊이, 너비, 해상도 변수는 각각 가중치 α, β, γ에 비례하며, $\alpha \cdot \beta^2 \cdot \gamma^2$값이 일정하게 유지되도록 설정함으로써 동일한 연산량을 유지하며 변수를 조정할 수 있다.

EfficientNet의 기본 모델(Baseline model)을 NAS(Neural Architecture Search)[20]를 이용하여 설계하였다. NAS는 최적의 딥러닝 모델을 찾기 위해 사용하는 방법으로 성능과 연산량이 최적인 모델을 찾아준다. NAS를 기반으로 설계된 기본 모델의 깊이, 너비, 해상도 변수를 작은 그리드 서치(Small grid search) 방법으로 찾은 모델을 EfficientNet-B0이라고 부른다. EfficientNet-B0를 기반으로 복합 계수(Φ)를 변경하며 모델 스케일링을 수행하여 EfficientNet-B1 ~ B7을 생성한다.

본 논문에서는 얼굴 나이 인식 성능을 향상시키기 위해, CNN의 성능을 최적화하기 위해 제안된 EfficientNet 모델을 활용하였다.

III. 얼굴 영상의 나이 인식

3.1 제안 방법

본 논문에서는 EfficientNet을 기반으로 얼굴 영상의 나이를 인식하는 방법을 제안하였으며, 제안된 네트워크를 보다 효율적으로 학습하기 위해 소프트맥스와 평균-분산 손실함수[8]를 함께 사용하였다. 본 논문에서 제안하는 얼굴 나이 인식 방법은 그림 2에서 보여준다.

3.2 EfficientNet 기반 얼굴영상 나이 인식

본 논문에서는 얼굴 영상의 나이 인식을 위해 EfficientNet을 활용하였다. EfficientNet에서 제안한 기본 모델인 EfficientNet-B0의 복합 계수(Φ)를 변경하며 모델 스케일링을 수행하면 EfficientNet-B1 ~ B7을 생성할 수 있다. 본 논문에서는 얼굴 나이 인식을 위해 기본 모델인 EfficientNet-B0와 모델 스케일링이 적용된 EfficientNet-B1, EfficientNet-B7 모델을 사용하였다.

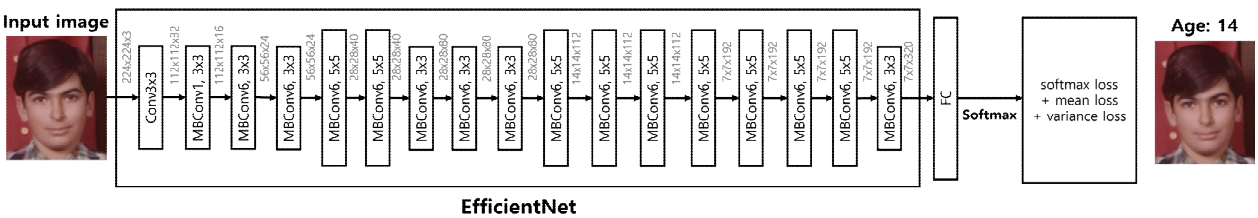

그림 2. EfficientNet과 평균-분산 손실함수를 이용한 얼굴 영상의 나이인식
Fig. 2. Proposed method of facial age estimation using EfficientNet and mean-variance loss

3.3 평균-분산 손실 함수

EfficientNet 기반의 얼굴 나이 인식 모델이 보다 나이 정보를 잘 인식할 수 있도록 평균-분산 손실함수를 사용하여 학습하였다. Hongyu Pan[8]은 일반적으로 딥러닝 모델에서 다중 클래스 분류를 위해 주로 사용되는 소프트맥스 손실(Softmax loss)과 평균-분산 손실(Mean-variance loss)을 함께 사용하는 나이 인식 방법을 제안하였다. 실험 결과, 얼굴 나이 인식에서 평균-분산 손실을 함께 사용하면 얼굴 나이 인식 성능이 향상되는 것을 확인하였다. 본 논문에서도 나이 인식 성능을 보다 향상시키기 위해 기존 방법[9]에서 제안한 평균-분산 손실과 소프트맥스 손실을 함께 사용하여 EfficientNet을 학습하였다.

평균-분산 손실을 구하기 위해 H. Pan et al.[8]은 CNN 모델의 마지막 완전연결층(Fully-connected layer)에서의 출력과 소프트맥스 확률(Softmax probability)을 다음 식 (2)과 같이 표현한다.

$$z = f(x_i)\theta^T, p_{i,j} = \frac{e^{z_{i,j}}}{\sum_{k=1}^{K} e^{z_{i,k}}} \tag{2}$$

z는 마지막 완전연결층의 출력, x_i는 i번째 샘플의 특징 벡터(Feature vector), θ는 마지막 완전연결층의 파라미터, $f(x_i)$는 완전연결층 이전 층의 출력, $p_{i,j}$는 부류 j에 속하는 샘플 i의 확률을 나타낸다. 식 (1)을 기반으로 모든 부류 K에 대해 샘플 i의 인식된 나이(Estimated age label)의 분포를 나타내는 p_i의 평균(m_i)과 분산(v_i)을 식 (3)과 같이 정의한다.

$$m_i = \sum_{j=1}^{K} j * p_{i,j},\ v_i = \sum_{j=1}^{K} p_{i,j} * (j - m_i)^2 \tag{3}$$

평균 손실(L_m)은 인식된 나이 분포(Estimated age distribution)의 평균(m_i)과 실제 나이(y_i)의 차이로 구하며, 다음과 같은 식 (4)로 계산된다.

$$L_m = \frac{1}{2N}\sum_{i=1}^{N}(m_i - y_i)^2 = \frac{1}{2N}\sum_{i=1}^{N}(\sum_{j=1}^{K} j * p_{i,j} - y_i)^2 \tag{4}$$

식 (4)에서 N은 배치사이즈(Batch size)를 의미한다. 일반적으로 소프트맥스 손실이 분류(Classification) 관점에서의 오차를 표현하고, 평균 손실은 회귀(Regression) 관점에서의 오차를 표현하기 때문에 두 손실을 상호 보완적으로 사용할 수 있다. 분산 손실(L_v)은 인식된 나이 분포(Estimated age distribution)에 제한을 두는 역할을 하며, 다음과 같은 식 (5)로 표현한다.

$$L_v = \frac{1}{N}\sum_{i=1}^{N} v_i = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} p_{i,j} * (j - \sum_{k=1}^{K} k * p_{i,k})^2 \tag{5}$$

기존 연구[8]에서 얼굴 나이 인식 모델을 학습할 때 소프트맥스 손실과 평균-분산 손실함수를 함께 사용하면 소프트맥스 손실만을 사용하는 것에 비해 성능이 향상된다는 것을 확인하였다. 본 논문에서도 EfficientNet기반의 얼굴 나이 인식 모델을 학습하기 위해 소프트맥스 손실과 평균-분산 손실함수를 함께 사용하였으며, 다양한 CNN 기반 모델과의 성능을 비교할 때도 동일한 방법을 사용하였다.

Ⅳ. 실 험

4.1 실험 설정

본 논문에서 제안하는 방법의 성능을 평가하기 위해 얼굴 나이 인식 분야에서 많이 사용되는 FG-Net[21], UTKFace[22], IMDB-WIKI[23] 데이터셋을 이용하였다. FG-Net[21]은 얼굴 나이에 관한 연구에서 많이 사용되는 데이터셋으로 82명의 유럽인을 대상으로 수집되었다. FG-Net은 한 사람당 연령대가 다른 여러 장의 사진으로 구성되어 있으며, 연령대는 0~69세 사이로 분포한다. 또한 포즈, 표정, 사이즈, 배경, 조명 등의 변화가 다양한 특징이 있다. UTKFace[22]는 다양한 인종으로 구성된 데이터셋으로 연령대는 1~116세 사이로 분포하며, 2000장 이상의 영상으로 구성되어 있다. IMDB-WIKI[23]는 IMDb와 Wikipedia에서 수집된 유명인의 얼굴 영상으로 구성되어 있다. IMDB-WIKI[23]는 IMDb에서

20,284명으로부터 수집한 460,723장의 영상과 Wikipedia에서 수집된 62,328장의 얼굴 영상을 포함하여 총 523,051장의 얼굴 영상으로 구성되어 있으며, 스크래핑을 통해 취득된 나이 정보를 함께 제공하고 있다.

각 데이터셋에서 제공하는 얼굴 영상은 포즈, 크기, 해상도 등이 다양한 특성이 있기 때문에, 나이 인식 정확도를 향상시키기 위해서 얼굴 영상을 정규화하는 과정이 필요하다. 본 논문에서는 얼굴의 주요 특징점을 기준으로 정규화를 수행하기 위해 FAN[24]을 이용하여 68개의 주요 특징점을 추출하였고, 이를 기반으로 얼굴의 크기, 기울기, 위치 등을 정규화하는 과정을 수행하였다. 그림3에서 FAN에 의해 추출된 68개의 얼굴 특징점을 보여준다. 이와 같은 방법으로 FG-Net, UTKFace, IMDB-WIKI 데이터셋의 얼굴 영상을 정규화하여 학습 및 테스트에 활용하였다.

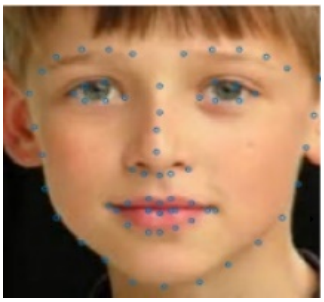

그림 3. FAN에 의해 추출된 68개 얼굴 특징점
Fig. 3. 68 facial landmarks extracted by using FAN

본 논문에서 제안한 모델은 먼저 IMDB-WIKI 523,051개 영상과 UTKFace 데이터셋 23,612개 영상을 이용하여 사전 학습하였고, 사전 학습된 모델을 FG-Net 데이터셋으로 미세조정(Fine-tuning)하였다. 또한 FG-Net 데이터셋을 평가데이터로 활용하여 나이 인식 성능을 측정하였다. FG-Net 데이터셋으로 평가를 수행할 때 동일인이 학습과 테스트 셋에 동시에 포함되지 않도록 LOPO(Leave-One-Person-Out) 방법[1]을 사용하였다. LOPO 방법은 FG-Net 데이터셋에 포함된 82명 중, 1명에 해당하는 10여장의 사진을 테스트 데이터셋으로 구성하고 나머지 81명의 사진을 학습 데이터셋으로 구성하는 방법이다. 실험을 위해 총 82-fold를 구성하였고, 각 fold 성능의 평균을 최종 성능으로 측정하였다.

성능 평가 지표로 식 (6), (7), (8)로 표현되는 MAE(Mean Absolute Error), RMSE(Root Mean Squared Error), CS(Cumulative Score)를 사용하였다.

$$MAE = \sum_{k=1}^{N} \left| \hat{a}_k - a_k \right| / N \tag{6}$$

$$RMSE = \sqrt{\sum_{k=1}^{N} (\hat{a}_k - a_k)^2 / N} \tag{7}$$

MSE, RMSE를 계산하는 식 (6), (7)의 N은 테스트 데이터의 수, a_k는 실제 나이, $\hat{a}_k$는 인식된 나이를 의미한다. CS는 식 (8)에 의해 계산되며, 식 (8)의 $N_{e \le th}$는 인식 오류(e)가 임계값(th)보다 작은 테스트 데이터의 수를 의미한다.

$$CS(th) = N_{e \le th} / N \times 100(\%) \tag{8}$$

제안된 모델의 학습에 사용된 하이퍼파라미터로 배치사이즈(Batch size) 8, 세대수(Epoch) 100, 학습률(Learning rate) 0.001을 사용하였다.

4.2 실험 결과

본 논문에서 제안한 EfficinetNet을 이용한 얼굴 영상의 나이 인식 성능을 표 1에서 보여준다. EfficientNet의 기본 모델인 EfficientNet-B0와 기본 모델에 스케일링을 적용한 EfficientNet-B1, EfficientNet-B7 모델을 사용한 경우의 성능을 비교하였다. 기본 모델인 EfficientNet-B0를 사용한 경우 MAE 3.06, 모델 스케일링을 적용한 EfficientNet-B1의 경우 MAE 3.09, EfficientNet-B7의 경우 MAE 2.85의 성능을 보였다. EfficientNet의 모델 스케일을 확장한 EfficientNet-B7에서 얼굴 영상의 나이 인식 성능이 가장 우수한 결과를 보였다.

표 1. EfficientNet 기반의 얼굴 나이인식 성능
Table 1. Facial age estimation performance using EfficientNet

Model	MAE
EfficientNet-B0	3.06
EfficientNet-B1	3.09
EfficientNet-B7	**2.85**

본 논문에서 제안한 나이 인식 방법과 기존 CNN을 사용한 나이 인식 방법의 성능을 비교[17]한 결과를 표 2에서 볼 수 있다. 기존 CNN 모델 중 가장 성능이 좋은 모델은 ResNet-34로, MAE 3.02의 성능을 얻었다. EfficientNet-B7을 사용한 경우 MAE 2.85로 ResNet-34 모델에 비해 성능이 보다 향상된 결과를 얻었다. 이 실험 결과를 통해 복합 스케일링 방법으로 모델의 스케일을 효율적이고 체계적으로 조정하는 EfficientNet이 얼굴 영상 나이 인식 분야에서 우수한 성능을 보임을 확인할 수 있다.

표 2. 다양한 CNN모델에 따른 얼굴 나이인식 성능 (MAE)
Table 2. Facial age estimation performance using various CNN models and EfficinetNet (MAE)

Model	MAE
AlexNet	4.47
VGG-16	3.13
VGG-19	3.83
ResNet-18	3.09
ResNet-34	**3.02**
ResNet-50	3.50
ResNet-101	3.49
ResNet-152	3.55
EfficientNet-B0	3.06
EfficientNet-B1	3.09
EfficientNet-B7	**2.85**

얼굴 나이 인식 성능을 RMSE(Root Mean Squared Error)로 비교한 결과를 표 3에서 보여준다. 표 3에서 볼 수 있듯이, CNN 모델들의 성능을 RMSE로 비교한 결과 EfficientNet-B7이 다른 CNN 모델과 비교하여 성능이 가장 우수한 것을 확인하였다.

표 3. CNN 모델에 따른 얼굴 나이인식 성능 (RMSE)
Table 3. Facial age estimation performance using various CNN models (RMSE)

Model	RMSE
AlexNet	6.00
VGG-16	4.36
ResNet-34	4.26
EfficientNet-B7	**4.03**

다음은 CS를 시각화하여 기존 CNN모델과 EfficientNet 기반 모델의 나이 인식 성능을 비교하였고, 그 결과를 그림 4에서 보여준다. 그림 4의 x축은 나이 인식 오차를 나타내며, y축은 CS를 나타낸다. CS를 시각화한 그래프가 위에 그려질수록 성능이 우수한 것을 나타낸다. 그림 4에서 볼 수 있듯이, CS 그래프에서도 EfficientNet-B7을 사용한 경우 성능이 가장 우수한 것을 확인하였다.

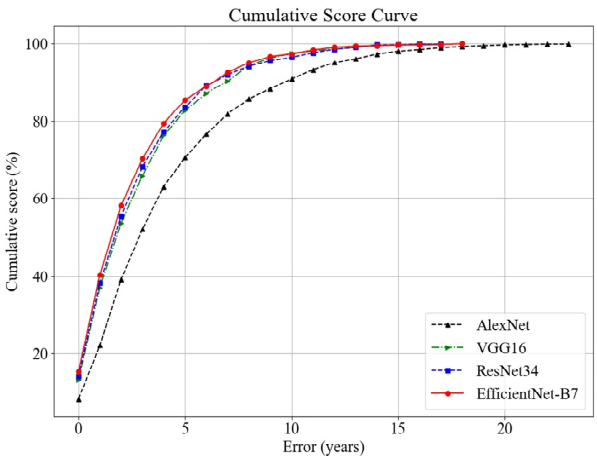

그림 4. EfficientNet과 기존 CNN 모델의 얼굴 나이 인식 CS 그래프 비교
Fig. 4. CS comparison between CNNs and EfficientNet face age estimation

V. 결론 및 향후 과제

본 논문에서는 EfficientNet을 기반으로 얼굴 영상의 나이를 인식하는 모델을 제안하였고, 제안된 모델을 학습하기 위해 평균-분산 손실을 소프트맥스 손실과 함께 사용하였다. 제안된 방법으로 얼굴 영상의 나이 인식을 수행한 결과, 기존 CNN 기반 방법 중 가장 우수한 성능을 보인 ResNet-34를 사용한 경우 MAE 3.02, EfficientNet-B7을 사용한 경우 MAE 2.85의 성능을 얻었다. RMSE, CS를 이용하여 성능을 비교한 결과에서도 EfficientNet-B7이 기존 CNN 방법에 비해 우수한 성능을 보이는 것을 확인하였다. 이를 통해 제안 방법의 얼굴 나이 인식 성능이 기존 CNN 방법에 비해 우수한 것을 확인하였다. 향후에는 Transformer 기반의 다양한 모델을 얼굴 나이 인식에 적용하고, 기존 CNN 기반 모델과의 성능 및 효율성을 비교·분석하는 연구를 진행할 예정이다.

References

[1] R. Angulu, J. R. Tapamo, and A. O. Adewumi, "Age estimation via face images: a survey", EURASIP Journal on Image and Video Processing, Vol. 2018, No. 1, pp. 1-35, Jun. 2018. https://doi.org/10.1186/s13640-018-0278-6.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Communications of the ACM, Vol. 60, No. 6, pp. 84-90, May 2017. https://doi.org/10.1145/3065386.

[3] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for large-scale image recognition", 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 2015. https://doi.org/10.48550/arXiv.1409.1556.

[4] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016. https://doi.org/10.1109/CVPR.2016.90.

[5] S. Zagoruyko and N. Komodakis, "Wide residual networks", The British Machine Vision Conference (BMVC), York, UK, Sep. 2016, https://doi.org/10.48550/arXiv.1605.07146.

[6] Y. Huang, et al., "Gpipe: Efficient training of giant neural networks using pipeline parallelism", Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, Canada, Dec. 2019. https://doi.org/10.48550arXiv.1811.06965

[7] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks", In International Conference on Machine Learning (ICML 2019), Long Beach, Jun. 2019. http://proceedings.mlr.press/v97/tan19a.html.

[8] H. Pan, H. Han, S. Shan, and X. Chen, "Mean-Variance Loss for Deep Age Estimation from a Face", IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, Jun. 2018. https://doi.org/10.1109/CVPR.2018.00554.

[9] W. Shen, Y. Guo, Y. Wang, K. Zhao, B. Wang, and A. Yuille, "Deep Regression Forests for Age Estimation", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, Jun. 2018. https://doi.org/10.1109/CVPR.2018.00245

[10] R. Rothe, R. Timofte, and L. V. Gool, "Deep Expectation of Real and Apparent Age from a Single Image without Facial Landmarks", International Journal of Computer Vision, Vol. 126, pp. 144-157, Apr. 2018. https://doi.org/10.1007/s11263-016-0940-3.

[11] R. Rothe, R. Timofte, and L. V. Gool, "DEX: Deep EXpectation of apparent age from a single image", 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, Chile, Dec. 2015. https://doi.org/10.1109/ICCVW.2015.41.

[12] X. Liu, Y. Zou, H. Kuang, and X. Ma, "Face Image Age Estimation Based on Data Augmentation and Lightweight Convolutional Neural Network", Symmetry, Vol. 12, No. 1, Jan. 2020. https://doi.org/10.3390/sym12010146.

[13] C. Zhang, S. Liu, X. Xu, and C. Zhu, "C3AE: Exploring the Limits of Compact Model for Age Estimation", The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, Jun. 2019. https://doi.org/10.48550/arXiv.1904.05059.

[14] I. Dagher and D. Barbara, "Facial age estimation using pre-trained CNN and transfer learning", Multimedia Tools and Applications, Vol. 80, No. 13, May 2021. https://doi.org/10.1007/s11042-021-10739-w.

[15] G. George, S. Adeshina, and M. M. Boukar,

"Age Estimation from Facial Images Using Custom Convolutional Neural Network (CNN)", Proceeding Book of 1st International Conference on Frontiers in Academic Research (ICFAR), Turkey, Vol. 1, Feb. 2023

[16] B. Abirami, T. S. Subashini and V. Mahavaishnavi, "Gender and age prediction from real time facial images using CNN", Materials today Proceedings, Vol. 33, No. 3, pp. 4708-4712, Sep. 2020. https://doi.org/10.1016/j.matpr.2020.08.350.

[17] S. E. Choi, "A study on age estimation of facial images using various CNNs (Convolutional Neural Networks)", Journal of Platform Technology, Vol. 11, No. 5, pp. 16-22, Oct. 2023

[18] S. E. Bekhouche, A. Benlamoudi, F. Dornaika, H. Telli, and Y. Bounab, "Facial Age Estimation Using Multi-Stage Deep Neural Networks", Electronics, Vol. 13, No. 16, Aug. 2024. https://doi.org/10.3390/electronics13163259.

[19] Q. Zhao, J. Liu, and W. Wei, "Mixture of deep networks for facial age estimation", Information Sciences, Vol. 679, Sep. 2024. https://doi.org/10.1016/j.ins.2024.121086.

[20] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and Q. V. Le, "MnasNet: Platform-aware neural architecture search for mobile", The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, Jun. 2019. https://doi.org/10.1109/CVPR41558.2019.

[21] The FG-NET Aging Database, https://kaggle.com/datasets/aiolapo/fgnet-dataset. [accessed: Oct. 30, 2024]

[22] UTKFace, https://susanqq.github.io/UTKFace/. [accessed: Oct. 30, 2024]

[23] IMDB-WIKI, https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/. [accessed: Oct. 30, 2024]

[24] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks)", International conference on computer vision (ICCV), Venice, Italy, Oct. 2017. https://doi.org/10.1109/ICCV.2017.116.

저자소개

최 성 은 (Sung Eun Choi)

2004년 2월 : 이화여자대학교 정보통신학과(공학사)
2010년 2월 : 연세대학교 생체인식협동과정(공학석사)
2015년 2월 : 연세대학교 전기전자공학과(공학박사)
2015년 3월 ~ 2020년 2월 : KIST 영상미디어연구단 박사후연구원
2020년 3월 ~ 현재 : 한양여자대학교 빅데이터과 조교수
관심분야 : 컴퓨터비전, 인공지능, 머신러닝, 딥러닝, 영상처리

김 익 재 (Ig-Jae Kim)

1996년 2월 : 연세대학교 공과대학 전기공학과(공학사)
1998년 2월 : 연세대학교 공과대학 전기공학과(공학석사)
2009년 2월 : 서울대학교 공과대학 전기·컴퓨터공학부(공학박사)
2009년 2월 ~ 2010년 8월 : MIT 미디어랩 박사후연구원
1998년 2월 ~ 2017년 2월 : KIST 영상미디어연구단 책임/선임/연구원
2017년 2월 ~ 2020년 8월 : KIST 영상미디어연구단 단장
2009년 2월 ~ 현재 : UST AI-로봇전공 전임교수
2020년 2월 ~ 현재 : 고려대학교 공과대학 겸임교수
2020년 2월 ~ 현재 : KIST AI.로봇연구소 소장
관심분야 : 컴퓨터비전, 패턴인식, 컴퓨터그래픽스, 딥러닝