

레터논문 (Letter Paper)

방송공학회논문지 제28권 제2호, 2023년 3월 (JBE Vol.28, No.2, March 2023)

<https://doi.org/10.5909/JBE.2023.28.2.238>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

특징 분해를 이용한 얼굴 나이 분류 및 합성

김 찬 호^{a)}, 박 인 규^{a)†}

Facial Age Classification and Synthesis using Feature Decomposition

Chanho Kim^{a)} and In Kyu Park^{a)†}

요 약

최근 인공지능 모델을 이용한 얼굴인식, 얼굴 수정 등 다양한 얼굴 작업들이 실생활에도 광범위하게 사용되고 있다. 그러나 모델의 학습에 사용되는 대부분의 얼굴 데이터셋은 사회활동이 활발한 특정 나이에 편중되고, 어린아이나 노인의 데이터가 적은 경향이 있다. 이와 같은 데이터셋 불균형 문제는 모델의 학습에도 좋지 않은 영향을 끼쳐, 아이나 노인같이 데이터가 적은 나이의 사람이 인공지능 모델을 사용할 때 사회활동이 활발한 나이의 사람이 사용할 때보다 성능이 떨어질 수 있고, 이들의 인공지능 모델 사용을 어렵게 할 가능성이 높다. 이를 개선하기 위해 본 논문은 특징 분해를 활용하여 얼굴 영상으로부터 나이를 분류하고 목표 나이로 합성하는 기법을 제안한다. 제안하는 기법은 FFHQ-Aging 데이터셋을 이용한 정량적, 정성적 평가를 통해 기존의 방법보다 더 나은 성능을 보인다.

Abstract

Recently deep learning models are widely used for various tasks such as facial recognition and face editing. Their training process often involves a dataset with imbalanced age distribution. It is because some age groups (teenagers and middle age) are more socially active and tends to have more data compared to the less socially active age groups (children and elderly). This imbalanced age distribution may negatively impact the deep learning training process or the model performance when tested against those age groups with less data. To this end, we propose an age-controllable face synthesis technique using a feature decomposition to classify age from facial images which can be utilized to synthesize novel data to balance out the age distribution. We perform extensive qualitative and quantitative evaluation on our proposed technique using the FFHQ dataset and we show that our method has better performance than existing method.

Keyword : Face synthesis, feature decomposition, age classification

a) 인하대학교 전기컴퓨터공학과(Inha University, Department of Electrical and Computer Engineering)

† Corresponding Author : 박인규(In Kyu Park)

E-mail: pik@inha.ac.kr

Tel: +82-32-860-9190

ORCID: <https://orcid.org/0000-0003-4774-7841>

※ This work was partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.2022R1A4A1033549 and No.NRF-2019R1A2C1006706). This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS-2022-00155915, Artificial Intelligence Convergence Innovation Human Resources Development (Inha University)).

· Manuscript February 20, 2023; Revised March 17, 2023; Accepted March 17, 2023.

Copyright © 2023 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

얼굴 나이 분류는 얼굴 데이터로부터 나이를 분류하는 작업이며, 분류한 나이를 바탕으로 나이를 활용해야 하는 작업에 사용된다. 이와 더불어 얼굴을 목표 나이로 합성하는 연구도^[1,2,3] 활발히 진행되고 있는데, 이를 이용하면 기존 얼굴 관련 데이터셋에 존재하는 데이터 다양성에 대한 한계점을 개선할 수 있다. 기존 데이터셋에는 사회활동이 활발한 젊은 층에 대한 데이터는 다수 존재하지만, 활동이 적은 어린아이나 노인의 얼굴 데이터는 적다. 이는 얼굴 데이터를 사용하는 다양한 작업에 영향을 주며, 인공지능 모델에 아이나 노인 얼굴을 이용할 때 성능 하락의 원인이 되므로 중요한 부분이다.

이전 연구^[1]는 중간 나이 합성의 경우 좋은 결과를 보여주지만, 아이나 노인의 합성 결과는 그 나이에 맞지 않아 보이는 결과를 합성한다. 이후의 연구^[2]에서는 아이나 노인 합성에서 더 좋은 결과를 합성하지만, 여전히 나이 특성이 적게 반영되었다. 최근에는 나이 불변 얼굴 인식과 얼굴 합성을 다중 작업 학습하여 얼굴 합성과 인식 등을 한 번에 수행하는 연구^[3]도 제안되었다. 하지만 이는 얼굴 합성에 집중하지는 않으며, 큰 나이 그룹을 사용하여 세밀한 나이 조절이 어렵다.

본 논문은 이전 나이 수정 네트워크^[2]를 기반으로 특징 분해를 이용하여 얼굴 수정과 나이 분류를 동시에 학습할 수 있는 네트워크를 구성한다. 이는 입력된 얼굴의 나이를 예측하고 입력 얼굴의 나이를 수정하여 다양한 나이로 만들어, 얼굴 나이 데이터의 특정 나이 편중 문제를 개선할 수 있으며, 숫자로 나이 조절이 가능하고 아이나 노인 합성에 있어 기존 기법보다 더 좋은 결과를 보여준다.

II. 제안하는 기법

1. 네트워크 구조

제안하는 기법은 오토 인코더 구조로 구성되며 인코더와 디코더, 특징 분해 모듈(feature decomposition module)과 나이 특징 CNN(age feature CNN)으로 이루어진다. 인코더에서는 입력된 얼굴 영상을 통과시켜 특징을 추출한다. 추출한 특징은 이전 연구^[3]에서 영감을 받아, 특징 분해 모듈에서 나이 관련 특징과 나이 관련하지 않은 특징을 분리한다. 분리하는 주의 기반 방법을 이용하며 자기주의(self-attention)로 집중되는 정도에 따라 인코더에서 추출된 특징에서 집중된 특징과 되지 않은 특징을 분리하는 방식을 사용한다. 분리된 나이 관련 특징은 완전 연결 레이어(fc)를 통과해 나이를 분류하며 교차 엔트로피를 이용하여 나이 관련 특징을 학습할 수 있도록 한다. 나이 관련하지 않은 특징은 디코더에 직접 전달됨과 동시에, 합성하고자 하는 목표 나이와 합쳐져 조건부 배치 정규화가 포함된 나이 특징 CNN을 통해 디코더의 각 레이어에 전달된다. 이 방식으로 나이 관련 특징이 분리된 특징에서 최적의 나이 인식 특징을 학습한다. 디코더에서 출력된 목표 나이로 수정한 얼굴은 입력된 영상과 비교하여 손실 함수를 얻는다.

2. 학습방법

제안하는 네트워크의 학습은 이전 연구^[2]의 방법을 따르며, 나이 분류와 얼굴 합성이 한 번에 진행된다. 손실 함수는 사실적인 합성에 초점을 맞추도록 구성하였다.

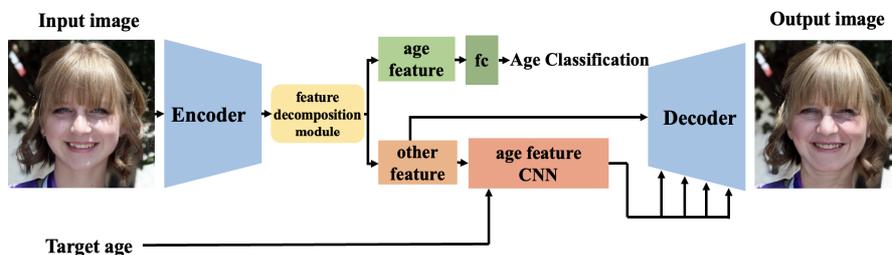


그림 1. 제안하는 네트워크 구조
 Fig. 1. Proposed network architecture

$$L_{Total} = \lambda_{rec}L_{rec} + \lambda_{cyc}L_{cyc} + \lambda_{adv}L_{adv} + \lambda_{age}L_{age} + \lambda_{vgg}L_{vgg} \quad (1)$$

첫 번째와 두 번째 항은 재구성 손실과 순환 손실로, 평균 절대 오차를 기본으로 하며 재구성 손실은 모델에 얼굴을 합성할 목표 나이로 원래의 나이를 입력해 준 다음, 출력된 영상과 입력된 영상의 차이를 줄이는 방식으로 계산한다. 이는 목표 나이로 원래의 나이를 입력하면 입력 영상과 합성 영상이 같아야 한다는 것을 이용한 것이다. 순환 손실은 입력된 영상을 무작위의 나이로 합성한 다음, 이를 다시 원래의 나이로 합성하여 출력된 영상을 입력 영상과 비교하는 식으로 계산한다. 세 번째 항은 적대적 손실로 사실적인 결과를 위하여 여러 합성 연구에서 많이 사용되는 손실이며 본 기법에서도 사용한다. 네 번째 항은 나이 손실로 입력 영상에 대한 나이 분류를 교차 엔트로피 손실로 계산한다. 마지막 항은 합성된 영상의 세부사항을 살리기 위하여 많이 사용되는 지각 손실 함수인 VGG 손실이다. VGG19 네트워크를 통과한 특징으로 계산하며, 재구성 손실을 계산하기 위해 합성된 영상을 이용하여 계산한다. 순환 손실을 계산하기

위해 모델을 두 번 통과하여 합성된 영상을 지각 손실 계산에 사용하면 학습에 악영향을 주는 것이 확인되어 사용하지 않는다. 하이퍼 파라미터는 $\lambda_{rec} = 10, \lambda_{cyc} = 1, \lambda_{adv} = 1, \lambda_{age} = 0.01, \lambda_{vgg} = 0.01$ 을 사용하며 이는 경험적으로 설정되었다.

III. 실험 결과 및 분석

실험은 FFHQ 데이터셋 기반에 나이 라벨과 메타 데이터가 추가된 FFHQ-Aging^[4]을 사용한다. 학습은 RTX A6000 GPU로 16배치 크기, 100 에포크로 학습하였다.

1. 정성적 결과

본 논문에서 제안하는 기법을 이용하여, 목표 나이로 입력 얼굴 영상을 합성하는 작업을 평가한다. 평가는 이전 연구인 HRFAE^[1], RAGAN^[2] 논문과 비교하며 HRFAE는 공개된 가중치를 이용, RAGAN은 제안하는 기법과 조건을

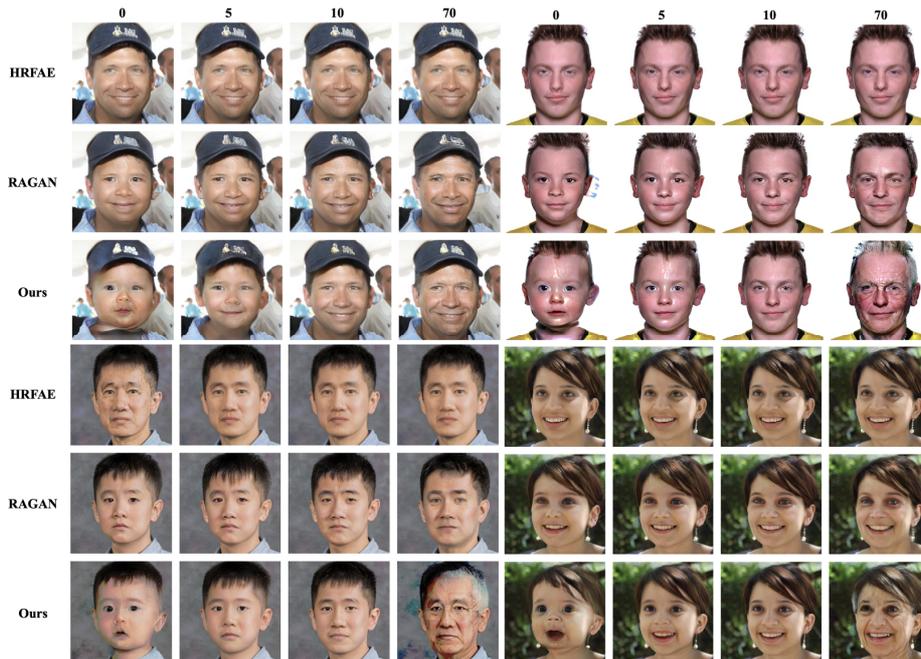


그림 2. 정성적 결과 비교

Fig. 2. Qualitative comparison between our proposed methods and prior works

같이 학습하여 비교한다. 그 예시를 그림 2에 보인다.

중간 나이 (20-60)의 결과는 각 논문들의 결과가 큰 차이가 없어 그림에 보이지 않았다. 하지만 목표 나이가 아이나 노인일 때 합성된 결과에 대해서는 차이를 보인다. HRFAE의 결과는 입력 영상과 크게 바뀌지 않는 것을 볼 수 있으며, RAGAN 논문의 결과는 대체로 이전보다 나이 특징이 더 고려되어 아이나 노인에서 더 적합해 보이는 영상을 합성하지만, 목표 나이로 보이지 않는 경우도 있었다.

제안하는 기법의 결과에서는 아이나 노인에서 더 적합해 보이는 영상을 합성하는 것을 확인할 수 있다.

2. 정량적 결과

먼저, 나이 분류 결과는 각 나이 클래스로 분류하였을 때 약 60%의 정확도를 보이며 인접한 클래스로 분류하는 비율이 95% 이상으로 확인되었다. 따라서 나이 관련 특징이 충분히 분리되고 있음을 확인할 수 있다.

표 1의 결과는 모델의 입력으로 목표 나이를 원래의 나이로 입력 후 합성한 결과 비교이다. 차이는 합성 모델에서 많이 사용되는 L1과 LPIPS를 이용하여 계산하며, 합성한 영상이 입력 영상과 차이가 작을수록 좋은 결과이다.

표 1. 원래 나이를 입력으로 합성한 영상의 정량적 결과 비교
 Table 1. Quantitative comparison with previous work for synthesizing an image using its original age

	L1(↓)	LPIPS(↓)
RAGAN[2]	0.033	0.182
Ours	0.028	0.072

표 2는 입력 영상의 나이를 무작위로 수정한 영상을 다시 원래의 나이로 합성한 결과이다. 제안하는 기법이 RAGAN보다 차이가 적다는 것을 확인할 수 있다.

표 2. 영상의 나이를 무작위로 수정 후 원래의 나이로 합성한 결과
 Table 2. Quantitative comparison for cycle generation evaluation. synthesize a novel image with random age, and then synthesize the original image and compare

	L1(↓)	LPIPS(↓)
RAGAN[2]	0.042	0.308
Ours	0.040	0.155

표 3은 합성된 얼굴 영상이 다른 데이터셋에서 학습된 나이 분류 네트워크를 통해 목표 나이로 분류되는지를 확인한다. 분류가 잘 될수록 성능이 좋은 것이며, 제안하는 방법이 더 좋은 결과를 보여주는 것을 확인할 수 있다.

표 3. 합성된 영상이 나이 분류 모델에서 분류되는 정확도
 Table 3. Comparison of age classification accuracy for age-controlled synthesized images between our method and prior works

Target age	0-14	15-19	20-39	40-49	50+
HRFAE ^[1]	52.2	55.3	58.1	57.3	59.4
RAGAN ^[2]	55.1	57.0	60.1	63.1	65.4
Ours	57.3	57.5	63.1	64.2	68.3

IV. 결론

본 논문에서는 특징 분해를 이용하여 나이 관련 특징을 전체 얼굴 특징에서 분리해 나이 수정과 분류에 활용한다. 제안하는 기법은 나이 분류와 합성이 동시에 가능하며, 아이와 노인 얼굴 합성에서 기존의 방법보다 더 좋은 결과를 보였다. 제안하는 기법을 통하여 얼굴 데이터셋의 아이와 노인 데이터가 적은 문제를 개선할 수 있을 것으로 기대되며, 향후 연구에서는 여러 얼굴 작업에서 얼굴 데이터셋을 사용할 때 증강을 이용한 개선 효과를 검증해 보고자 한다.

참고 문헌 (References)

- [1] X. Yao, G. Puy, A. Newson, Y. Gousseau, P. Hellier, "High resolution face age editing," Proc. International Conference on Pattern Recognition, January 2021.
doi: <https://doi.org/10.1109/icpr48806.2021.9412383>
- [2] F. Makhmudkhujjev, S. Hong, and I. K. Park, "Re-Aging GAN: Toward personalized face age transformation," Proc. IEEE/CVF International Conference on Computer Vision, October 2021.

doi: <https://doi.org/10.1109/iccv48922.2021.00388>

- [3] Z. Huang, J. Zhang, and H. Shan, "When age-invariant face recognition meets face age synthesis: A multi-task learning framework," Proc. IEEE/CVF Computer Vision and Pattern Recognition, June 2021.
doi: <https://doi.org/10.1109/cvpr46437.2021.00720>
- [4] R. Or-El, S. Sengupta, O. Fried, E. Shechtman, I. Kemelmacher-Shlizerman, "Lifespan age transformation synthesis," Proc. European Conference on Computer Vision, August 2020.
doi: https://doi.org/10.1007/978-3-030-58539-6_44