

MobileFaceNet 기반 Label2label 모델 경량화 및 인종 편향 완화

권은아, 김현탁, 이다은, 정채영, 함우주
성균관대학교

rnsdmsdk1@skku.edu, gusxkr1706@gmail.com, dlekdms7931@skku.edu, pcc1016@skku.edu,
hahmwj@gmail.com

MobileFaceNet-Based Label2label: Compression Approach and Racial Bias Mitigation

Kwon Eun A, Kim Hyeon Tak, Lee Da Eun, Jeong Chae Young, Ham Woo Joo
Sungkyunkwan Univ.

요약

얼굴 속성 분류(Facial Attribute Classification)는 얼굴 이미지에서 다양한 속성을 자동으로 식별하는 기술이다. 이러한 기술은 주로 엣지 디바이스 환경에서 활용되는데 기존 Label2label 모델은 파라미터 수와 연산량이 많아 적합하지 않다. 따라서 엣지 디바이스 환경에 적용하기 위해서는 경량화가 필요하다. 또한 기존 얼굴 속성 분류에 사용되는 데이터셋은 백인에 편향된 경우가 많아서 동양인에 적용했을 때 성능이 저하될 수 있다. 본 논문에서는 MobileFaceNet 으로 Label2label 로 경량화를 수행하고 한국인 데이터셋을 추가적으로 학습해서 인종에 대한 편향을 완화했다.

I. 서론

얼굴 속성 분류 (Facial Attribute Classification)는 얼굴 이미지에서 나이, 성별 등 다양한 속성을 분류하는 기술로 얼굴 식별, 얼굴 인식, 얼굴 검색 등에 사용된다. 특히 모바일, IoT (Internet of Things) 와 같은 엣지 디바이스 (Edge devices) 환경이 가정되는 경우가 많아 메모리와 연산 자원이 제한되는 상황에서도 효율적인 성능을 발휘해야 한다[1].

Label2label[2]은 각 속성을 단어로 취급하여 이미지에 의미 정보를 도입한 최초의 다중 속성 분류 모델로, 언어 모델링을 이용하여 State-of-the-art 를 달성하였다. 그러나 Label2label 은 Resnet-50 을 백본으로 사용하기 때문에 연산량과 파라미터 수가 많아 엣지 디바이스 환경에서 사용하기에 부적합하다. 이에 본 논문은 ResNet-50 에 비해 파라미터 수와 연산량이 적은 MobileFaceNet[4]을 백본으로 사용하여 모델 경량화를 진행했다.

또한, 이전 얼굴 속성 분류 연구에서 많이 사용되던 데이터셋은 대부분 백인으로 구성되어 있어 모델이 편향되는 문제가 발생한다[3]. 이러한 편향은 인종에 따른 모델의 일반화 성능을 저해한다. 따라서 한국인 안면 데이터셋을 추가적으로 활용하여 모델을 파인 튜닝(Fine-tuning)해 동서양인에 대해 성능을 향상시켰다. 본 연구는 얼굴 다중 속성 분류 작업에서 효율성과 일반화 성능을 동시에 향상시키는 데 기여하고자 한다.

II. 본론

2.1 데이터셋 구성

기존 안면 분류 모델이 사용하는 CelebA[5], LFWA[5] 데이터셋은 백인 인종이 많이 포함되어 있어 백인에 대한 편향이 생기기 쉽다[3]. 이러한 편향을 줄이기 위해 CelebA 와 AI Hub¹에서 제공하는 한국인 안면 이미지를 추가하여 데이터셋을 보완한다. 두 데이터셋의 분포를 맞추기 위해 정면을 보고 있는 이미지와, 빛의 강도와 각도를 유사한 한국인 안면 데이터를 활용했다. 본 논문에서는 CelebA 와 한국인 안면 이미지를 합쳐 데이터셋을 재정의한다.

2.2 MobileFaceNet

백본 모델로 사용한 MobileFaceNet[4]은 모바일, 임베디드 디바이스에서 실시간 얼굴 식별을 위한 경량화 모델이다. 이 모델은 Face feature embedding CNN 의 마지막 합성곱 층 이후에 전역 풀링 층이나 완전 연결 층 대신에 전역 깊이별 합성곱 층을 사용하여 파라미터 수를 효과적으로 줄였다. 표 1 은 MobileFaceNet 과 Resnet-50 이 단일 이미지 인식을 수행할 때 파라미터 수와 연산량 비교를 나타낸 것이다[6].

모델	파라미터 수	FLOPs
Resnet-50	40.29M	2.19G
MobileFaceNet	1.0M	0.45G

표 1. ResNet-50 과 MobileFaceNet 비교

¹ 이 연구는 과학기술정보통신부의 재원으로 한국지능정보사회진흥원의 지원을 받아 구축된 "데이터명"을 활용하여 수행된 연구입니다. 본 연구에 활용된 데이터는 AI 허브(aihub.or.kr)에서 다운로드 받으실 수 있습니다.

2.3 Label2label

Label2label[2]는 얼굴의 다중 속성간 복잡한 상관 관계를 모델링하기 위해 설계되었다. 각 속성 레이블의 샘플을 설명하는 “단어”로 취급한다. 각 샘플은 여러 속성 레이블로 주석이 달려있으므로, 이러한 “단어”는 순서는 없지만 의미 있는 “문장”을 형성한다. 예를 들어 속성 레이블인 “attractive”, “no eyeglasses”가 있을 때, 이를 “attractive, no eyeglasses, ...” 와 같은 문장으로, 얼굴을 묘사하기 위해 사용한다. 이 모델은 레이블 “문장”에서 일부 “단어”를 무작위로 마스크하고, 마스크된 “문장”과 이미지 특징에 의한 컨텍스트를 기반으로 이를 복원한다. 이러한 방식으로 다중 속성에 대해 효과적으로 학습할 수 있다.

기존 Label2label 은 ResNet-50 을 백본으로 활용하였으나, 파라미터 수와 연산량이 많아 엣지 디바이스 환경에서 적절하지 않다. 이에 엣지 디바이스 환경에 적절한 경량화 모델인 MobileFaceNet 을 활용했다. 그림 1 은 모델 아키텍처를 나타낸다.

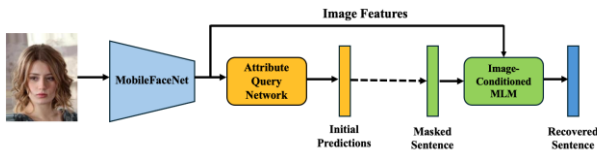


그림 1. 경량화 Label2label 아키텍처

III. 실험 결과

3.1 실험환경

3.2 에서는 Label2label 의 백본 모델을 MobileFaceNet 으로 대체하고 CelebA 데이터에 대해 학습과 테스트를 진행한다. 3.3 에서는 CelebA 로 Pre-trained 된 Label2label 을 한국인 안면 데이터셋 8 개의 얼굴 다중 속성 레이블 (안경, 모자, 블러, 웃음, 마스크, 젊음, 수염, 금발, 갈발, 흑발, 회색머리)에 대해 실험을 수행하고 한국인 안면 데이터셋으로 파인튜닝을 수행한다.

3.2 경량화

표 2 는 Label2label 을 변경한 것에 대한 결과이다. 모델의 파라미터 수가 크게 감소하였음에도 불구하고 정확도는 유사하다. 이러한 결과에서 Resnet-50 에서 MobileFaceNet 으로 교체하여 큰 성능 저하 없이 경량화하였다.

Backbone	Accuracy
ResNet-50	0.8813
MobileFaceNet	0.8810

표 2. Label2label 백본에 따른 정확도

3.3 한국인 안면 데이터 파인 튜닝

표 3 은 기존 Label2label 을 얼굴 인식 연구에서 주로 쓰이는 데이터인 CelebA 에 대해 Pre-train 시킨 뒤 한국인 데이터셋에 대하여 테스트를 진행한 결과이다. 실험 결과 일반화 성능이 감소하는 것을 확인했다.

Backbone	Fine-tuning	accuracy
MobileFaceNet	X	0.8552

표 3.경량화 Label2label 을 한국인 데이터에 테스트

따라서 CelebA 에 Pre-train 된 경량화된 Label2label 을 한국인 안면 데이터셋에 대해 파인 튜닝을 진행한다. 표 4 는 이에 대한 결과이다. 파인 튜닝을 진행하기 전인 표 3 보다 일반화 성능이 향상된 걸 확인할 수 있다.

Backbone	Fine-tuning	accuracy
MobileFaceNet	O	0.8820

표 4.경량화 Label2label 한국인 데이터셋에 파인튜닝

IV. 결론

본 논문에서는 엣지 디바이스 환경에서 얼굴 다중 속성 분류의 효율성을 높이기 위해 MobileFaceNet 을 기반으로 Label2label 모델을 경량화했다. 이 과정에서 백인 편향을 완화하기 위해 한국인 안면 데이터를 함께 활용했다. 실험 결과, 경량화된 모델은 성능을 유지하면서도 파라미터 수를 크게 절감했다.

본 연구는 MobileFaceNet 을 활용한 접근법이 엣지 디바이스에서 실시간 얼굴 속성 분류를 가능하게 하는 유망한 해결책을 입증했다. 또한 데이터셋 재정의로 인종에 대한 일반화를 향상시켰다. 향후에는 더 다양한 인종 데이터를 바탕으로 모델의 일반화 성능을 향상시켜 엣지 디바이스에서의 얼굴 다중 분류 연구에 기여될 수 있다.

ACKNOWLEDGMENT

이 논문은 2024 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2019-0-00421, 인공지능대학원지원(성균관대학교))

참고 문헌

- [1] He, Yonghao, et al. "Lffd: A light and fast face detector for edge devices." *arXiv preprint arXiv:1904.10633* (2019).
- [2] Li, Wanhua, et al. "Label2label: A language modeling framework for multi-attribute learning." *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022.
- [3] Karkkainen, Kimmo, and Jungseock Joo. "Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation." *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2021.
- [4] Chen, Sheng, et al. "Mobilefacenet: Efficient cnns for accurate real-time face verification on mobile devices." *Biometric Recognition: 13th Chinese Conference, CCBIR 2018, Urumqi, China, August 11-12, 2018, Proceedings 13*. Springer International Publishing, 2018.
- [5] Liu, Ziwei, et al. "Deep learning face attributes in the wild." *Proceedings of the IEEE international conference on computer vision*. 2015.
- [6] Zhao, Feng, et al. "UnifiedFace: A uniform margin loss function for face recognition." *Applied Sciences* 13.4 (2023): 2350.