

VGG-Kface : VGG-Face를 이용한 한국 인 얼굴 인식에 관한 최적화 연구

¹이성찬, ¹김승한, ¹김민경, ¹조민진, ¹고범석, ³유용만
 중앙대학교 전자전기공학부 학부생, ¹인하대학교 전자공학과
 학부생 ¹숙명여자대학교 컴퓨터과학과 학부생,
¹이화여자대학교 전자전기공학과 학부생
¹동국대학교 인공지능학과 석사과정, ³서울아산병원

lsc429@cau.ac.kr, 12191453@inha.edu, wendy5619@sookmyung.ac.kr, minjin9384@ewhain.net,
 roy7001@dgu.ac.kr, yongman@amc.seoul.kr

VGG-Kface : An Optimization Study on Korean Face Recognition Using VGG-Face

¹Seong-Chan Lee, ¹Seung-Han Kim, ¹Min-Gyeong Kim,
¹Min-jin Cho, ¹Beom-Seok Ko, ³Yong-man Yu
¹Dept. of Electrical Electronic Engineering, Chung-Ang University,
¹Dept. of Electronic Engineering, Inha University,
¹Dept. of Computer Science, Sookmyung Women's University,
¹Dept. of Electronic Engineering, Ewha Women's University,
¹Dept. of AI, Dongguk University, ³Asan medical center

요 약

얼굴인식 모델이 서양인 얼굴에 맞춰져 있어 한국인 얼굴에 대한 인식 성능 향상이 필요하다. 본 논문에서는 얼굴인식 모델에 AIHub에서 제공하는 한국인 얼굴 데이터 셋을 추가하고, 서양인 비교되는 한국인의 특징을 추가하여 얼굴인식을 진행하였다. contrastive learning의 image pair 쌍의 적합한 비율 평가를 계층적으로 진행하여 한국인 인식 성능을 높인 VGG-Kface를 제안한다.

1. 서론

최근 아이폰의 얼굴 인식과 같은 인공지능 기반 얼굴인식 보안 적용 사례는 증가하고 있다. 인공지능 분야에서 얼굴 인식 능력과 인식 속도에 초점을 맞추어 FaceNet[1] 과 ArcFace와 같은 모델들이 요구되고 있다. 하지만 모델의 학습 과정에 들어간 데이터셋은 대부분 서양인들의 얼굴로 구성된 CelebA 나 LFW이다. 한국인들의 얼굴 특징은 서양인들과 다르고 이에 따른 얼굴 인식 능력도 비교적 저조하여 한국인들을 위한 모델의 필요성을 인식하였다.

본 논문에서는 서양인의 이미지로 사전 훈련된 얼굴인식 모델 VGGFace에 한국인 이미지[2]를 추가로 학습시키는 방법을 제안하고, 모델을 최적화시키기 위해 계층적인 평가를 통하여 성능을 증진시켜 모델이 한국인 얼굴에 대해 견고해짐을 보인다.

2. 모델 환경 설정

VggFace 모델은 두 얼굴 이미지가 같은 사람의 얼굴인지 여부를 판단하는 모델이다. VggFace 모델의 마지막 컨볼루션 레이어를 제외한 나머지 레이어

들의 파라미터는 업데이트 되지 않도록 동결해주었고 마지막 레이어의 활성화 함수를 tanh로 설정하였다. 두 얼굴 특징 벡터 간의 거리는 유클리디안 거리를 이용하였다. 손실 함수로는 binary cross-entropy를 사용하였다. 학습 데이터는 resizing과 color adjustment의 전처리를 적용한 AIHub의 '한국인 안면 이미지' 데이터셋 중 400명의 사람에 대하여 조도와 액세서리(안경, 모자 등) 유무가 다른 5장의 정면 사진으로만 구성하였다. 그리고 얼굴 영역만 dlib로 crop하여 이용하였고 배치 사이즈는 32로 설정하여 200 epoch까지 모델을 훈련시켰다. 최적화 알고리즘은 Adam을 사용하였고 초기 학습률은 0.01로 설정하였다. 또한 코사인 스케줄러를 사용하여 0.01 ~ 0.001의 범위 내에서 전체 epoch에 걸쳐 decaying 하도록 학습률을 조정하였다. 모델 학습 환경은 colab 환경에서 GPU (V100 혹은 A100)을 이용하였다.

A. Image Pair 비율에 따른 성능 평가

VGGFace에서는 손실 함수로 triplet loss를 채택하므로 두 개의 이미지 쌍이 네트워크에 입력된다. 두

개의 이미지가 서로 다른 인물인 negative pair와, 같은 인물인 positive pair이다. negative pair와 positive pair는 실험 진행 시 각각 0과 1로 labeling 된다. 또한 모든 가능한 이미지 쌍을 만들어 학습하는 것이 아닌, 적절한 비율로 샘플링을 거쳐 학습을 진행하도록 하였다.

1.1 비율 조정

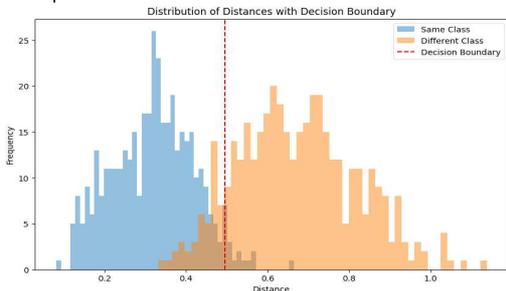
학습 데이터의 클래스 분포의 균일성이 학습 성능에 영향을 줄 것으로 판단했다. 이에 negative pair와 positive pair의 비율에 따른 성능을 실험했다

n:p	accuracy	loss
1:9	0.96	0.15
2:8	0.92	0.23
3:7	0.92	0.21
4:6	0.94	0.17
5:5	0.94	0.17
6:4	0.94	0.16
7:3	0.94	0.17
8:2	0.95	0.18
9:1	0.97	0.08

1.2 경계기준 설정

위 결과들의 주요 지표는 accuracy, loss로 이들은 단순히 모델이 label을 잘 예측했는지를 나타내는 지표이며, 클래스에 따른 특징 벡터 간 거리를 얼마나 뚜렷하게 학습하느냐에 관한 지표도 제시될 필요가 있다. 따라서 same class와 different class 각각에 해당하는 데이터를 400개씩 샘플링하여 클래스별 벡터 거리의 분포를 확인하고 모델의 성능 지표로 두 분포가 겹치는 영역의 넓이를 새로 제시했다.

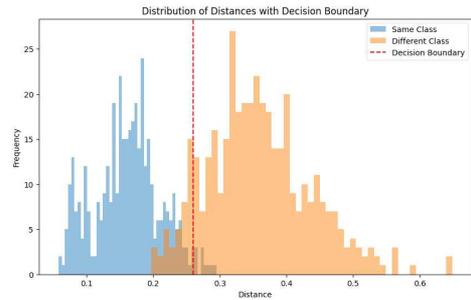
n:p	겹치는 영역	threshold
1:9	0.48	0.20
2:8	0.55	0.26
3:7	0.51	0.24
4:6	0.44	0.24
5:5	0.46	0.26
6:4	0.45	0.27
7:3	0.61	0.30
8:2	0.45	0.36
9:1	0.59	0.49



<그래프 1> n:p=4:6일 때의 분포도

B. 사람당 구성되는 이미지 수에 따른 성능 평가

한 사람당 몇 장의 사진을 학습시켰을 때 최적화된 모델이 나오지 실험했다. n:p=4:6에 대해, 사람 400명에 대하여 각 5장씩 학습시킨 결과와 400명에 대하여 각 10장씩 학습시킨 결과를 비교하였다. 사진의 구성을 달리 해 본 결과, 성능 저하의 주된 원인은 조명의 다양성이었다. 10장의 사진을 학습할 때 다른 조명 조건으로 데이터셋을 구성하면서 성능이 더 나빠지는 경향이 있었다.



<그래프 2> 400명에 대하여 10장에서 분포도

	accuracy	loss	겹치는 영역
5장	0.94	0.17	0.44
10장	0.85	0.36	0.66

3. 결론

한국인 얼굴 특징을 잘 추출할 수 있는 VGG-Kface를 제안하며, 성능을 더 높일 수 있는 실험을 몇 가지 진행하였다. 첫 번째로는, 모델에 입력되는 positive pair와 negative pair의 비율에 따른 모델의 성능 변화에 대해 실험해보았고 accuracy 지표로는 n:p가 9:1로 했을 때가 성능이 제일 좋았으며, 겹치는 영역을 지표로 봤을 때는 n:p가 4:6일 때 성능이 제일 우수함을 확인할 수 있다. 두 번째로는, 한 사람당 구성되는 이미지 수에 따른 성능 평가를 실험해보았고, 인물 당 10장의 이미지로 구성했을 때보다 5장의 이미지를 구성했을 때가 성능이 더 좋은 것을 확인할 수 있다. VGG-Kface는 그 구조가 간단하고 직관적이며, 실시간 추론 속도가 빠르고 성능 또한 준수해 edge에서도 쉽게 사용 가능하다. 또한 기업, 군 시설 등 보안을 필요로 하는 시설의 출입 시스템에 활용될 수 있음을 시사한다.

※ 본 논문은 해양수산부 실무형 해상물류 일자리 지원사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

참고문헌

[1] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition.
 [2] CHOI Yeji et al. K-face: A large-scale kist face database in consideration with unconstrained environments. arXiv preprint arXiv:2103.02211, 2021.