

저화질 미디어의 감정인식 문제 해결을 위한 고품질 변환 기술 현황과 성능분석

김나연*, 조숙희**, 배병준*, **

*과학기술연합대학원대학교, **한국전자통신연구원

{boboss, shee, 1080i}@etri.re.kr

Status and Performance of the High-Quality Conversion Technology for the Problem of Facial Expression Recognition in Low-Quality Media

Nayeon Kim, Sukhee Cho, Byungjun Bae

*Univ. of Sci. and Tech., **Electronics and Telecommunications Research Institute

요 약

본 논문에서는 지하철 CCTV 영상과 같은 저화질 미디어에서 사람의 표정으로 감정을 인식할 때 성능을 향상시킬 수 있는 딥러닝 기반 미디어 고품질 변환 기술에 대해 알아보고, 성능 평가 결과를 제공하여 범죄 예방에 활용할 수 있는 기초자료를 제공하는 것을 목표로 한다. 성능 평가는 공개된 감정인식 데이터셋을 일반적인 쌍선형 보간법으로 해상도를 8 배한 이미지와 GAN 모델을 기반으로 8 배한 이미지를 ResNet 으로 각각 학습하여 수행되었으며, 학습, 검증, 평가 실험 모두에서 GAN 기반 초해상도 이미지로 학습한 모델이 더 높은 감정인식 정확도가 나타난 것을 확인하였다.

I. 서 론

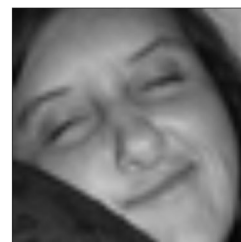
표정은 감정을 인식하는 데 중요한 역할을 한다. 특히, 저화질 미디어에서의 표정인식은 범죄취약지역의 방범용 CCTV 나 지하철 등의 저화질 CCTV 처럼 얼굴 식별이 어려운 미디어에 적용하여 범죄 예방에 활용될 수 있다.

표정을 통한 감정 인식은 컴퓨터 비전 분야에서 오랫동안 연구되어 왔으나, 딥러닝의 발전 이후에는 딥러닝을 기반으로 표정을 인식하는 연구들이 활발하게 진행되고 있다. 이러한 이미지 작업의 성능은 이미지의 해상도에 큰 영향을 받기 때문에, 해상도가 낮고 저품질인 얼굴 이미지에서 감정인식 성능의 저하를 방지하기 위한 다양한 방법들이 함께 제시되었으며[1][2], 최근에는 새로운 데이터 생성에 사용되는 딥러닝 모델 GAN(Generative Adversarial Network)을 통해 이미지를 고품질화하여 감정인식 성능을 높이는 연구가 시도되고 있다[3].

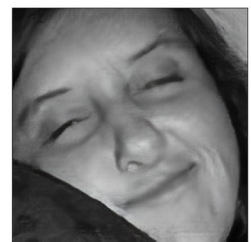
이에 본 논문에서는 감정인식 성능을 높이는 데 활용할 수 있는 GAN 기반 고품질 변환 기술에 대해 파악하고, 기본적인 이미지 보간 방법인 쌍선형 보간법(ilinear)과 GAN 기반 얼굴 이미지 생성 모델[8] 사이의 감정인식 성능을 비교 및 분석하여 저화질 미디어에서의 감정인식 성능 향상을 위한 고품질 변환 기술 설계 및 개발의 기초자료로 제공하고 사회 문제 해결에 기여하고자 한다.

II. GAN 기반 얼굴 이미지 고품질 변환 기술

얼굴 이미지 고품질 변환은 단일 이미지에서 임의의 장면 대신 얼굴 이미지만 고려하기 때문에 랜드마크나 히트맵과 같은 얼굴의 지역 특징들이 강력한 사전



(a) ilinear



(b) DIC

그림 1. FER 데이터셋(48x48)의 쌍선형 보간법과 DIC 적용 결과(384x384)

지식으로써 활용되고 있다[4]-[7]. 이러한 사전 정보를 충분히 활용하기 위해 고품질 변환의 지역화 및 정렬 과정에서 왜곡없이 랜드마크 및 얼굴 구성 요소 맵을 정확하게 추정하기 위한 방법 또한 제안되었다[8].

그림 1 은 감정인식 데이터셋 FER2013 에 쌍선형 보간법과 [8]에서 제안된 GAN 기반 초해상도 기술을 적용한 결과를 나타낸다.

III. 고품질 변환 이미지의 감정인식 성능 평가

성능 평가는 감정인식 데이터셋인 FER2013 에서 잘못된 감정 라벨을 다시 지정한 FER plus 를 사용하여 수행되었다. FER 데이터셋에 쌍선형 보간법과 고품질 변환을 적용하여 ResNet[9]을 각각 학습시킨 후 감정인식 정확도를 평가했고, 고품질 변환에 사용한 얼굴 초해상도 기술로는 [8]에서 제안된 얼굴 이미지 복구

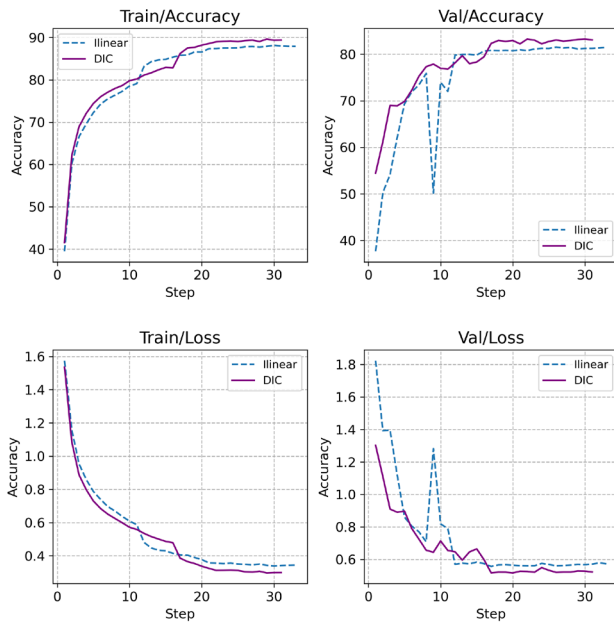


그림 2. 쌍선형 보간법과 DIC 를 적용한 감정인식 학습 및 검증 실험의 정확도와 Loss 그래프

네트워크와 정확한 랜드마크 추정 네트워크가 반복 협업하여 이미지를 생성하는 DIC (Deep Iterative Collaboration)를 사용하였다.

그림 2 성능 평가 결과 그래프로, 학습 및 검증 시 정확도 변화와 Loss 변화를 나타낸다. 쌍선형 보간법을 적용한 실험의 경우 학습 정확도 87.84%, 검증 정확도 81.52%, 평가 정확도 83.45%가 가장 높은 정확도로 나타났으며, DIC 실험은 학습 정확도 88.99%, 검증 정확도 83.24%, 평가 정확도 83.52%가 가장 높은 정확도로 나타났다. 세 가지 분리된 데이터셋 모두에서 DIC 를 적용한 실험이 더 높은 정확도를 보였다.

IV. 결론

본 논문에서는 저화질 미디어에서의 표정을 통한 감정 인식 기술들을 살펴보고, 그 중 저화질 미디어에 초해상도 기술을 적용했을 때 감정인식 성능을 평가 및 비교하였다. 저화질 미디어의 감정인식 기술은 얼굴 식별이 어려운 미디어에 적용하여 범죄취약지역의 범죄 혹은 디지털 성범죄 예방에 활용될 수 있다. 본 논문에서 수행한 평가 결과가 이러한 사회 문제 해결을 위한 기술 개발의 기초자료로 활용될 것을 기대한다. 그러나, 본 논문에서 사용된 FER2013 데이터셋은 구글 이미지 검색 API 를 사용하여 감정, 성별, 나이 등의 키워드로 검색된 결과 이미지들로 구성되어 있기 때문에 실제 환경에서 감정인식 성능을 보기 위해서는 조명, 각도, 가림 등 현실의 다양한 상황을 반영한 데이터셋을 이용하는 성능 평가 또한 필요하다.

ACKNOWLEDGMENT

“이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2021-0-00087, SD/HD 급 저화질 미디어의 고품질 변환 기술 개발)”

참 고 문 헌

- [1] CHENG, Bowen, et al. Robust emotion recognition from low quality and low bit rate video: A deep learning approach. In: 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, 2017. p. 65-70.
- [2] H. Jung, S. Lee, J. Yim, S. Park and J. Kim, "Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 2983-2991
- [3] HU, Xiaobin, et al. Face super-resolution guided by 3d facial priors. In: European Conference on Computer Vision. Springer, Cham, 2020. p. 763-780.
- [4] Yu, Xin, et al. "Super-resolving very low-resolution face images with supplementary attributes." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
- [5] KIM, Deokyun, et al. Progressive face super-resolution via attention to facial landmark. arXiv preprint arXiv:1908.08239, 2019.
- [6] Xin Yu, Basura Fernando, Bernard Ghanem, Fatih Porikli and Richard Hartley, "Face super-resolution guided by facial component heatmaps", ECCV, pp. 217-233, 2018.
- [7] Yu Chen, Ying Tai, Xiaoming Liu, Chunhua Shen and Jian Yang, "Fsnet: End-to-end learning face super-resolution with facial priors", CVPR, pp. 2492-2501, 2018.
- [8] C. Ma, Z. Jiang, Y. Rao, J. Lu and J. Zhou, "Deep Face Super-Resolution With Iterative Collaboration Between Attentive Recovery and Landmark Estimation," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5568-5577
- [9] He, Kaiming, et al. "Identity mappings in deep residual networks." European conference on computer vision. Springer, Cham, 2016.