생성적 적대 신경망을 이용한 얼굴의 가려진 영역 복원

이동규, 한동석*
경북대학교

jasmindoe@knu.ac.kr, *dshan@knu.ac.kr

Restoration of Occluded Facial Area using Generative Adversarial Networks

Dong Gyu Lee, Dong Seog Han*
Kyungpook National Univ.

요 약

이미지에서 객체를 찾거나 추적하는 과정에서 다른 요소로 인해 객체가 가려지는 경우가 발생할 수 있다. 이러한 요소들은 객체 탐색 및 추적 과정에서 성능 저하나 오류를 유발할 수 있다. 본 논문은 얼굴이 가려진 상황에서 해당 영역을 복원하는 방법에 대해 다룬다. 일반적으로 가려진 영역을 복원하기 위해 마스크, 손, 안경 등 특정 객체나 무작위 형태의 객체를 학습하는 방식이 사용된다. 그러나 이러한 방법은 학습되지 않은 방해물에 대해 인식이 어렵거나 성능 저하를 초래할 수 있다. 본 논문에서는 이러한 한계를 극복하기 위해 객체 인식 기반이 아닌, 얼굴 특징점 인식과 생성적 적대 신경망(GAN)을 활용한 가려진 얼굴 복원 모델을 제안한다. 제안 모델은 기존 모델에 비해 적은 양의 데이터로도 얼굴을 가리는 무작위 형태의 객체를 검출하여 얼굴 요소를 복원한다.

I. 서론

얼굴 복원은 영상 및 얼굴 인식 시스템 등에서 지난 몇 년 동안 연구되어왔다[1], [2], [5]. 얼굴 인식 분야에서 안경, 마스크, 손 등에 의한 가림은 방해요소가 된다. 얼굴의 가림은 정보의 손실을 야기하여 얼굴 인식과 같은 알고리즘의 성능에 영향을 준다. 가려지는 객체의 유형, 모양, 위치에 따라 얼굴의 가려진 영역을 복원하는 것은 매운 어려운 문제이다.

이 문제를 해결하기 위하여 다양한 방법의 연구들이 진행되고 있다. 이들 중 딥러닝 기반 학습은 좋은 성능을 보여주고 있다. 최근에는 적대적 이미지 생성 네트워크(generative adversarial network, GAN)를 사용하여 가려진 영역을 복원하는 연구가 진행되고 있다[3], [4], [7], [8]. NovelGAN [7] 모델은 이중 구조의 GAN을 사용하여 첫 단계에서 무작위 객체를 인식하여 추출하고 두번째 단계에서 해당 객체를 제거하여 지워진 부분을 복원한다. 무작위 객체를 인식하기 위해서 인식 및 추출을 위한 복원 모델을 사용하게 되는데 이를 위해 가려진 객체에 대한 학습을 요구한다. 따라서 학습되지 않은 전혀 새로운 유형의 방해물에 대해서는 인식 성능이 저하되는 문제점이 존재한다.

본 논문은 가려진 얼굴의 요소를 복원하는 것이다. 이를 위하여 얼굴 특징점을 파악하여 무작위 객체로 가려진 부분을 인식하여 복원하고자 한다.

II. 본론

제안하는 모델은 세개의 모듈로 구성되며 방해물 인식, 가려진 영역 복원, 복원 영역 판단으로 나뉜다. 방해물 인식은 입력 이미지에서 방해물이 있는 영역 또는 방해물 객체를 인식한다. 기존에는 단순히 생성자 모델을 통한 인코더-디코더(encoder-decoder) 방식의 네트워크를 사용한다[9]. 특정한 타입의 방해물 객체나 무작위 모양의 노이즈를 해당 모델에 학습시켜 인식한다. 그러나 이 방식은 학습되지 못한 가리는 영역을 인식을 하지 못하거나 잘못된 영역을 인식한다.

본 논문은 무작위 객체로 인해 얼굴이 가려진 영역 인식하기 위하여 객체 인식이 아닌 얼굴 특징점을 찾는 딥러닝 모델을 사용한다. 해당 모델은 CNN 기반의 ResNet [10]을 뼈대로 사용하고 입력된 이미지에서 얼굴의 주요 이목구비의 형태와 위치를 파악한다. 총 64개의 키 포인트가 얼굴 형태, 눈, 코, 입의 모양을 나타낸다. 가려진 이미지에서 인식된 얼굴 특징점은 미리 학습된 어텐션 모듈로 입력이 된다. 이 모듈은 이미지에서 가려진 이목구비 위치를 파악한다. 방해물로 인해 가려진 이목구비 정보는 기존의 얼굴 정보와 함께 입력되어 얼굴특징 분포도를 비교하게 되고 유사도를 측정한다. 측정된 유사도를 기반으로 가려진 이목구비 부위를 추정한다.

추정된 정보는 생성자의 입력으로 사용되기 위해 인코더-디코더 모델을 통해 가려진 영역을 강조한 바이너리 이미지가 된다. 가려진 얼굴 이미지와 방해물 바이너리 이미지가 같이 생성자 네트워크의 입력으로 사용된다. 복원 네트워크는 인코더-디코더 형태를 가지고 있고 U-Net 구조를 가진다[11]. 추출된 특징 블록은 디코더를 통해 복원 이미지가 생성된다. 디코더는 인코더와 대칭되는 구조를 가지고 있으며 노드마다 인코더에 연결되어 있다.

생성자 모델들은 방해물 영역 인식과 영역 복원을 위한 손실함수로 2가지를 사용한다. 재구축 L1 손실함수와 SSIM(structural similarity index measure)

손실함수를 사용한다. 재구축 L1 손실 함수는 생성된 이미지 결과물과 정답 이미지 사이의 픽셀 차이의 평균을 오차로 사용한다. SSIM은 생성된 결과물과 정답 이미지 사이의 구조적 유사성을 측정한다. 구조적 유사성으로는 생성된 이미지와 정답 이미지의 색상, 대조, 광원을 비교하여 오차로 사용한다.

생성자가 만든 결과물의 품질을 높이기 위해 추가적인 구별자를 사용한다. 1개의 생성자에 대한 2개의 구별자를 사용하고 구별자들은 지역 구별자와 전역 구별자로 사용된다. 전역 구별자는 기존의 구별자처럼 생성자가 만든 전체 이미지 영역을 정답 이미지와 비교한다. 반면 지역 구별자는 바이너리 이미지를 통해 추출된 영역과 생성자가 복원한 부분을 비교한다. 해당 영역의 정답 이미지와 생성 이미지를 비교하여 생성자에게 피드백을 전달한다. 해당 피드백은 생성자가 가려진 영역을 세밀하게 재현할 수 있도록 한다.

지각(perceptual) 네트워크는 미리 학습된 VGG-19 모델을 사용한다[13]. 지각 네트워크는 글로벌 구별자와 동일한 입력을 받는다. 네트워크는 결과 이미지와 정답 이미지 사이의 지각 거리를 계산하게 되고 특징 분포도를 정답 이미지에 근접한다.

모델 학습을 위해 Celeb-A [12] 데이터 세트와 FFHQ 데이터 세트를 사용했다. 본 논문에서는 이들 중 각각 50,000장과 25,000장의 이미지를 무작위로 추출하여 학습에 사용하였다. 가려진 얼굴 이미지를 만들기 위해 이들 데이터에 무작위 형상의 영역을 생성하였다.

제안 모델은 2개의 판별기를 사용하여 학습 불균형 문제가 발생할 가능성이 있다. 이를 방지하기 위하여 학습 중 반복 단계(epoch)의 20%까지 전역 판별기만을 사용해서 생성자를 학습한다. 이후 학습에서는 지역 판별기를 추가하여 학습한다. 이 때 지역 판별기의 피드백 값에 가중치를 주어 학습 시 전역 판별기보다 더 많은 피드백을 전달하게 한다. 그리고 학습이 진행됨에 따라 지역 판별기의 피드백 가중치를 점진적으로 감소하고 최종적으로는 지역 판별기와 전역 판별기의 피드백 값 비중을 6대 4가 되게 학습한다. 얼굴 특징점 모델은 미리 학습된 ResNet 19에 Flickr 데이터세트를 사용하여 학습하였다.

제안 모델과 비교 모델의 성능을 표1에서 나타내고, 제안 모델이 모든 지표에서 뛰어난 성능을 보이고 있다.

표. 1. 모델 성능 비교 결과

	U-net 생성기의 GAN			
성능지표	patchGAN	전체-로컬 GAN	2Stage-GAN	제안하는 모델
FID	15.46	18.31	14.84	14.2
SSIM	35.65	27.06	38.13	39.51
PSNR	0.746	0.536	0.784	0.798

Ⅲ. 결론

본 논문에서는 얼굴이 가려진 영역을 복원하기 위해 기존 연구에서 가졌던 한계점을 개선하였다. 얼굴 가림에서 학습하지 못한 객체에 대해서 성능이 저하되는 문제점을 해결하고자 얼굴 특징점 모듈을 융합하였다. 그리고 얼굴 가림 인식 모델로 다양한 가림 객체를 추가적으로 학습 없이 이 영역 인식할 수 있게 하였다. 생성기는 방해물 이미지와 가려진 얼굴 이미지로 복원된 이미지를 생성하고 판별기는 두 개의 모델로 구성된다. 제안한 방식은 기존의 연구 방식 모델들에 비해 FID 점수, SSIM 및 PSNR 측면에서 더 나은 성능을 보여주었다.

ACKNOWLEDGMENT

이 논문은 2023년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원(P0024162, 2023년 지역혁신클러스터육성)과 정부(과학기술정보통신부)의 재원으로 과학기술사업화진흥원의 지원을 받아 수행된 연구임('학연협력플랫폼구축 시범사업' RS-2023-00304695).

참 고 문 헌

[1] Li, Si-Qi, Yue Gao, and Qiong-Hai Dai. "Image De-occlusion via Event-enhanced Multi-modal Fusion Hybrid Network." *Machine Intelligence Research* 19.4, pp. 307-318, 2022.

[2] Zhan, Xiaohang, et al. "Self-supervised scene de-occlusion." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

[3] Dong, Jiayuan, et al. "Occlusion-aware gan for face de-occlusion in the wild." *2020 IEEE International conference on multimedia and expo (ICME)*. IEEE, 2020.

[4] Yuan, Xiaowei, and In Kyu Park. "Face de-occlusion using 3d morphable model and generative adversarial network." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.

[5] Zhang, Ni, et al. "Face de-occlusion with deep cascade guidance learning." *IEEE Transactions on Multimedia,* 2022.

[6] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[7] Din, Nizam Ud, et al. "A novel GAN-based network for unmasking of masked face." *IEEE Access* 8, pp. 44276-44287, 2020.

[8] Creswell, Antonia, et al. "Generative adversarial networks: An overview." *IEEE signal processing magazine* 35.1 ,pp. 53-65, 2018.

[9] Cho, Kyunghyun. "Learning Phrase Representations using RNN Encoder– Decoder for Statistical Machine Translation." *arXiv preprint arXiv:1406.1078,* 2014.

[10] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[11] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015.

[12] Liu, Ziwei, et al. "Large-scale celebfaces attributes (celeba) dataset." Retrieved August 15. 11, 2018.

[13] Simonyan, Karen. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556, 2014.