합성곱 신경망(CNN)을 이용한 U-Net 기반의 인공지능 안면 정면화 모델1)

이상민*, 손원호**, 진창균***, 김지현*, 김지윤+, 박나은++, 김가은*, 권진영2),
이혜리++, 김종완†, 오덕신§
*삼육대학교 인공지능·빅데이터 소프트웨어코드 연구소
**충북대학교 컴퓨터공학과
삼육대학교 {***컴퓨터 메카트로닉스 공학부, +식품영양학과, † 스미스학부대학,
§경영정보학과}, ++성신여자대학교 융합보안공학과, 2)건국대학교
정보통신경영학과
sangmin010203@gmail.com

Face Frontalization Model with A.I. Based on U-Net using Convolutional Neural Network

Sangmin Lee*, Wonho Son**, ChangGyun Jin***, Ji-Hyun Kim*, JiYun Kim+,
Naeun Park++, Gaeun Kim*, Jin young Kwon*, Hye Yi Lee*, Jongwan Kim†,
Dukshin Oh§

요 약

안면 인식은 Face ID를 비롯하여 미아 찾기, 범죄자 추적 등의 분야에 도입되고 있다. 안면 인식은 최근 딥러닝을 통해 인식률이 향상되었으나, 측면에서의 인식률은 정면에 비해 특징 추출이 어려우므로 비교적 낮다. 이런 문제는 해당 인물의 정면이 없고 측면만 존재할 경우 안면 인식을 통한 신원확인이 어려워 단점으로 작용될 수 있다. 본 논문에서는 측면 이미지를 바탕으로 정면을 생성함으로써 안면 인식을 적용할 수 있는 상황을 확장하는 인공지능 기반의 안면 정면화 모델을 구현한다. 모델의 안면 특징 추출을 위해 VGG-Face를 사용하며 특징 추출에서 생길 수 있는 정보 손실을 막기 위해 U-Net 구조를 사용한다.

1. 서론

안면 인식은 신원확인을 가능하게 하는 대표적인 생체인식 기술 중 하나로써 [1] 컴퓨터 비전 분야에서 활발히 연구되고 있다 [2]. 안면 인식은 보안이나 미아 찾기, 범죄자 추적, 등에 사용되지만, 특징점을 추출하기 어려운 측면에서의 인식률이 비교적 낮아 [3] 안면 인식이 가능한 상황이 제한된다. 기존에는 새로운 손실 함수나 Maxpooling 연산을 통해 각도 변화에 강인한 안면 인식 모델 생성에 관한 연구가 진행되었다 [3, 4]. 하지만 안면의 각도와 인식률은 비례하기 때문에 [5] 이를 해결하고자 안면을 정면화하는 방법이 제시되었다.

안면 정면화는 상대적으로 신원확인이 어려운 측면의 안면을 정면으로 변환하는 작업이다. 안면 인식은 측면보다 정면의 안면이 인식률이 높은 특징을 갖고 있기 때문에 안면 정면화는 측면을 정면으로 변환하여 측면 인식이 어려운 문제를 해결한다.

1) 이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2018R1D1A1B07045642, NRF-2017R1D1A1B03035884).
§ 교신저자 (ohds@syu.ac.kr)

본 논문은 VGG-Face [6]와 U-Net [7]을 사용하여 안면 정면화 신경망을 구축한다. 안면 정면화는 특징을 추출하는 신경망과 정면을 생성하는 신경망으로 나뉜다. 특징 추출은 ±60°의 안면 각도에서도 평균 91%의 안면 인식률을 보여준 VGG-Face 모델의 안면 특징 신경망을 사용하며 [8], 정면 생성은 질감 정보와 중요 정보 유지에 뛰어난 [7] U-Net 구조를 사용한다.

본 논문의 이후 구성은 다음과 같다. 2절에서는 안면 정면화와 관련된 연구를 살펴보고, 3절에서는 본 연구에서 구성한 신경망을 설명한다. 4절은 실험을 통해 신경망의 성능을 검증하고 5장에서 결론을 서술한다.

2. 관련 연구

2.1 안면 정면화

안면 정면화는 측면에서 추출된 제한적인 특징으로 정면을 생성해야 하므로 해결하기 어려운 문제에 속한다 [5]. 초기의 정면화 기법으로는, 안면 이미지를 3차원 모델링하여 정면화 하는 기법 [9]과, 통계를 기반으로 한 기법 [10]이 존재한다. 이후, 딥러닝의 등장으로 인해, 더 큰 각도의 측면에서도 정면화

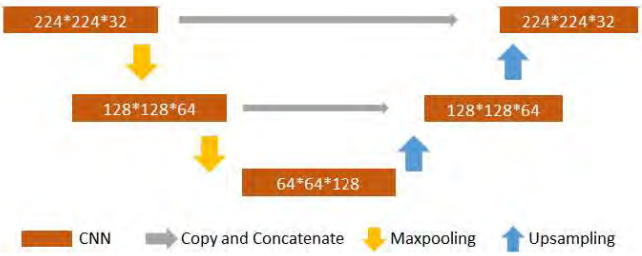
가 가능하게 되어 활발한 연구가 진행되었다. 딥러닝을 사용한 기법으로는, 안면 각도에 따라 다른 방식으로 특징을 추출하여 정면을 생성하는 conditional-CNN 기법 [11]이 제시되었으며, 초기의 3차원 모델링 기법과 실사 같은 이미지를 생성하는 생산적 적대 신경망 (Generative Adversarial Network)구조를 병합하여 정면을 생성하는 기법 [12]도 연구되었다.

2.2 VGG-Face

CNN이 등장한 이후, Labeled Faces in the Wild (LFW) 벤치마크 [13]에서 완벽에 가까운 안면 인식이 가능하게 되었다 [14]. 하지만 조명이나 배경, 안면 각도가 다양한 환경에서는 인식률이 떨어지기 때문에 이 문제를 해결하는 연구가 진행되었다 [3, 4]. VGG-Face는 image-net.org에서 준우승을 거둔 VGG16 신경망 [15]에 안면 데이터를 학습시킨 모델이다. VGG-Face는 깊은 층의 CNN을 구성하여 조명이나 배경의 변화에도 강인하며 [16], ±60°의 안면 각도에도 91%의 인식률을 보여준 안면 인식 모델이다. VGG-Face 내부에 존재하는 분류 층을 제거하면 안면 특징 추출 신경망으로 사용할 수 있어 안면 이미지를 사용하는 연구 [17, 18]에 사용된다.

2.3 U-Net

신경망 구조에서 입력 데이터를 압축하는 인코더와 다시 입력 데이터와 같은 차원의 데이터를 출력하는 디코더로 구성된 신경망을 오토인코더 (Autoencoder)라고 한다. U-Net은 오토인코더의 변형으로, 이미지 데이터를 다룬다. U-Net 구조는 디코더에서 이미지를 팽창시킬 때마다 하위층에서 추출된 특징을 연결 (Concatenate)하여 추가적인 정보를 예측에 사용하는 구조다 (그림 1). 이 방법은 스킵 연결 (Skip connection)이라고 불리며 인코더에서 데이터 압축 중에 손실될 수 있는 특징을 유지해주는 특징이 있다.

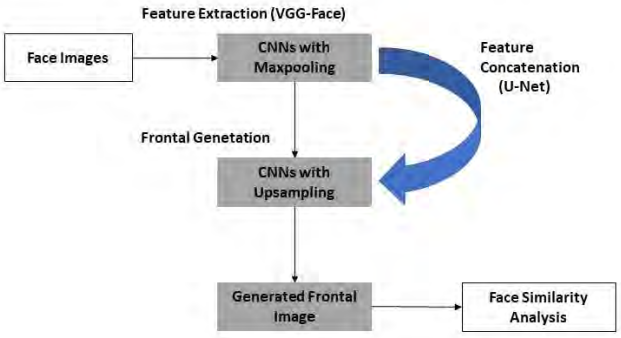


(그림 1) U-Net 구조를 가진 신경망의 예시.

U-Net은 입력 데이터의 질감을 유지하는 데에 있어 효과적이며, 데이터에서 중요한 정보가 무엇인지 파악하기 수월하게 한다. U-Net 구조의 이러한 장점은 오토인코더 구조로 이미지를 다루는 연구에 자주 사용되며, U-Net 구조를 사용하여 안면 정면화를 시도한 연구 또한 존재한다 [19].

3. 안면 정면화 신경망

측면은 안면의 특징점이 적고 추출하기 어렵기 때문에 효과적인 신경망을 구성하는 것이 중요하다. 본 연구의 안면 정면화 신경망은 특징 추출에 뛰어난 VGG-Face를 사용하는 특징추출 신경망과 이미지의 정보 손실을 막기 위한 U-Net을 사용한 정면 생성 신경망으로 구성된다 (그림 2).



(그림 2) 본 연구의 안면 정면화 순서도.

3.1 안면 특징 추출 신경망

본 연구에서는 비 제한적인 환경에서도 뛰어난 성능을 보여준 VGG-Face 사전학습 모델을 기반으로 특징 추출 신경망을 생성한다. 사전학습 모델을 사용하는 것은 새로운 신경망을 학습시키는 것보다 시간과 성능의 면에서 효과적이므로 [20] 신경망의 완성도를 높일 수 있다.

VGG-Face는 13개의 합성곱 층과 5개의 Maxpooling 연산으로 구성된다. VGG-Face의 입력으로는 128*128*3 크기의 안면 이미지가 들어가며, 4*4*512 크기의 특징을 추출하는 Maxpooling 연산이 마지막 층이다. 추가로, 인코더의 마지막 층을 기준으로 대칭을 이루는 U-Net 구조를 사용하기 위해 1024 크기의 특징을 추출하는 합성곱 층 하나를 추가하여 결과적으로 4*4*1024 크기의 특징이 추출된다.

3.2 정면 생성 신경망

인코더가 안면의 핵심적인 특징을 추출하는 것에 성공하더라도, 정면 생성 신경망의 성능이 떨어지면 좋은 결과를 기대하기 어렵다. 따라서 본 논문의 디코더는 이미지 처리에 뛰어난 합성곱 신경망과 특징 추출 중 손실될 수 있는 특징을 보존하는 U-Net 구조를 사용한다.

디코더는 5개의 합성곱 층과 5개의 Upsampling 연산으로 구성되며, Upsampling 이후 같은 크기를 갖는 인코더의 층에서 특징을 연결하여 안면 이미지

의 질감과 중요 정보를 유지한다.

인코더와 디코더를 거쳐 생성된 결과는 실제 정면과의 유사도 비교를 통해 완성도가 평가되며 유사도 측정 기술은 COMTRUE의 Face Similarity Analysis [21]를 사용한다.

4. 실험

4.1 학습 및 검증 데이터

본 실험에 사용된 데이터는 한국인 안면 데이터 [22]를 사용하였다. 제공된 데이터에는 영상 해상도, 액세서리, 조명 세기, 표정, 안면 각도, 인물 ID와 같이 데이터에 변화를 줄 수 있는 항목들이 존재한다. 본 논문에서는 한국인 300명의 상하좌우의 안면 각도 20개, 조명 세기 및 방향 30개, 표정 세 종류, 안경과 안경이 없는 상태를 포함한 총 1,080,000개의 안면 이미지를 데이터로 사용한다.

데이터에는 배경과 해당 인물의 옷, 등 부가적인 정보가 포함되어 있어 정면화에 불필요한 정보들이 포함될 수 있으므로, 이미지에서 안면만 잘라내는 전처리 과정을 거친다. 데이터의 300명 중 280명의 데이터를 학습 데이터로 사용하였으며 나머지 20명을 모델의 검증 데이터로 사용하였다.

4.2 실험 결과

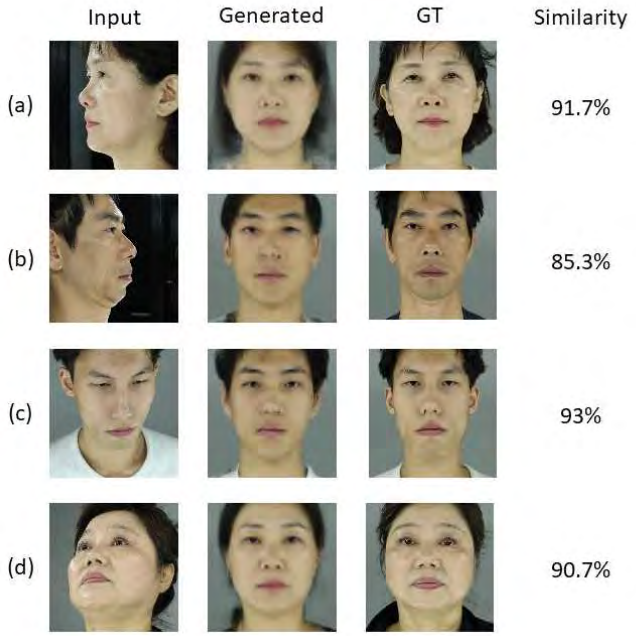
안면 정면화는 다양한 안면 각도에도 신원 정보가 손실되지 않은 정면을 생성하는 것이 중요하다. 본 실험에서는 수평 각도에서는 0°를 제외하고 ±75°까지 총 10개의 안면 각도에서의 정면화 결과를 실험하며, 수직으로도 -15°부터 30°까지의 안면 각도를 다룬다.

VGG-Face와 U-Net을 사용한 본 연구의 정면화 결과는 그림 3에서 확인 할 수 있다. 그림 3(a)와 그림 3(b)는 수직으로는 각도 변화가 없는 측면을, 그림 3(c)와 그림 3(d)는 수직과 수평의 각도를 모두 포함한 측면의 예측 결과를 나타낸다.

표 1은 검증에 사용된 20명의 측면들을 정면화 하였을 때 생성된 정면과 실제 정면과의 유사도를 입력된 안면의 각도에 따라 세분화하여 나타낸다. 모든 결과를 평균 내었을 때 유사도가 78.26% 였으며, 입력 이미지가 정면에 가까울수록 유사도가 상승하였다. 안면에 대한 수직 각도가 0이면서 수평 각도가 ±15°일 때 가장 높은 85.32%의 유사도가 확인되었으며, 수직 각도와 수평 각도가 각각 30°, ±15°일 때 유사도가 가장 낮았다.

<표 1> 안면의 각도에 따른 유사도

수직 \ 수평	±15°	±30°	±45°	±60°	±75°
0°	85.32%	82.9%	82%	80.1%	76.6%
-15°	-	71.23%	-	-	-
30°	69.68%	-	-	-	-



(그림 3) 안면 정면화 결과.

5. 결론

안면 정면화는 측면 이미지를 정면으로 변환하는 작업으로, 안면 인식의 측면 인식 문제를 해결해주는 중요한 역할을 한다. 본 논문에서는 측면에서도 높은 인식률을 보여준 VGG-Face를 사용하여 안면의 특징을 추출하며, U-Net 구조를 통해 추출 과정 중에 생길 수 있는 정보 손실을 막는다. 앞서 언급된 기술을 사용하여 생성된 정면은 실제 정면과 비교했을 때 평균적으로 78.26% 유사했다. 따라서 본 연구는 안면 인식률이 낮은 측면의 이미지를 정면으로 변환하여 신원 확인을 수월하게 한다. 이를 통해 본 논문의 정면화 모델은 안면 인식 시스템이 측면에서도 사용될 수 있도록 확장 할 수 있는 가능성을 확인하였다.

참고문헌

[1] I. Masi, Y. Wu, T. Hassner and P. Natarajan, “Deep Face Recognition: A Survey.”, 31st SIBGRAPI Conference on Graphics, Patterns and Images, Parana, 2018, pp. 471-478.

[2] P. Sinha, B. Balas, Y. Ostrovsky and R. Russell, "Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About.", in Proceedings

of the IEEE, vol. 94, no. 11, pp. 1948-1962, 2006.
[3] M. Kan, S. Shan and X. Chen, “Multi-view Deep Network for Cross-View Classification.”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 4847-4855.
[4] F. Schroff, D. Kalenichenko and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering.”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 815-823.
[5] R. Huang, S. Zhang, T. Li and R. He, “Beyond Face Rotation: Global and Local Perception GAN for Photorealistic and Identity Preserving Frontal View Synthesis.”, IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 2458-2467.
[6] O. M. Parkhi, A. Vedaldi and A. Zisserman, “Deep Face Recognition.”, Proceedings of the British Machine Vision Conference (BMVC), Swansea, UK, 2015, pp. 1-12.
[7] O. Ronneberger, P. Fischer and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation.”, Medical Image Computing and Computer-Assisted Intervention, MICCAI, 2015, pp. 234-241.
[8] X. Xu, H. A. Le, P. Dou, Y. Wu and I. A. Kakadiaris, “Evaluation of a 3D-aided pose invariant 2D face recognition system.” IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, 2017, pp. 446-455.
[9] T. Hassner, S. Harel, E. Paz, R. Enbar, “Effective Face Frontalization in Unconstrained Images.”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4295-4304.
[10] C. Sagonas, Y. Panagakis, S. Zafeiriou and M. Pantic, “Robust Statistical Face Frontalization.” IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 3871-3879.
[11] C. Xiong, X. Zhao, D. Tang, K. Jayashree, S. Yan and T. Kim, “Conditional Convolutional Neural Network for Modality-Aware Face Recognition.″ IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 3667-3675.
[12] X. Yin, X. Yu, K. Sohn, X. Liu and M. Chandraker, “Towards Large-Pose Face Frontalization in the Wild”, Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 3990-3999.
[13] G. B. Huang, M. Mattar, Ta.Berg and E. Learned-Miller. “Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments.” Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Marseille, France, 2008.
[14] Y. Sun, D.Liang, X. Wang and X. Tang, “DeepID3: Face Recognition with Very Deep Neural Networks.”, arxiv, 2020.
[15] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, arxiv, 2014.
[16] P. J. Phillips, “A Cross Benchmark Assessment of a Deep Convolutional Neural Network for Face Recognition.”, 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, 2017, pp. 705-710.
[17] Z. Qawaqneh, A. A. Mallouh and B. D. Barkana, “Deep Convolutional Neural Network for Age Estimation based on VGG-Face Model”, arxiv, 2017.
[18] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li and A. Hadid, “An original face anti-spoofing approach using partial convolutional neural network.” Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, 2016, pp. 1-6.
[19] Y. Yin, S. Jiang, J. P. Robinson and Y. Fu, “Dual-Attention GAN for Large-Pose Face Frontalization.”, arxiv, 2020.
[20] D. Marmanis, M. Datcu, T. Esch and U. Stilla, ″Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks,″ in IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 1, pp. 105-109, 2016.
[21] 유사도 측정, http://www.comtrue.com.
[22] 한국인 안면 데이터, https://aihub.or.kr/aidata/73.