JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 25 No. 5, pp. 1-9, May 2020
https://doi.org/10.9708/jksci.2020.25.05.001

Improved STGAN for Facial Attribute Editing by Utilizing Mask Information

Hyeon Seok Yang*, Jeong Hoon Han*, Young Shik Moon*

*Student, Dept. of Computer Science and Engineering, Hanyang University, Ansan, Korea
*Student, Dept. of Computer Science and Engineering, Hanyang University, Ansan, Korea
*Professor, Dept. of Computer Science and Engineering, Hanyang University, Ansan, Korea

[Abstract]

In this paper, we propose a model that performs more natural facial attribute editing by utilizing mask information in the hair and hat region. STGAN, one of state-of-the-art research of facial attribute editing, has shown results of naturally editing multiple facial attributes. However, editing hair-related attributes can produce unnatural results. The key idea of the proposed method is to additionally utilize information on the face regions that was lacking in the existing model. To do this, we apply three ideas. First, hair information is supplemented by adding hair ratio attributes through masks. Second, unnecessary changes in the image are suppressed by adding cycle consistency loss. Third, a hat segmentation network is added to prevent hat region distortion. Through qualitative evaluation, the effectiveness of the proposed method is evaluated and analyzed. The method proposed in the experimental results generated hair and face regions more naturally and successfully prevented the distortion of the hat region.

▸ **Key words**: Facial attribute editing, GAN, Deep learning, Mask, STGAN

[요 약]

본 논문에서는 머리카락과 모자 영역의 마스크 정보를 활용하여 더 자연스러운 얼굴 속성 편집(facial attribute editing)을 수행하는 모델을 제안한다. 최신 얼굴 속성 편집 연구인 STGAN은 다중 얼굴 속성을 자연스럽게 편집하는 성과를 보였다. 그러나 머리카락과 관련된 속성을 편집할 때 부자연스러운 결과를 생성할 수 있다. 제안하는 방법의 핵심 아이디어는 기존 모델에서 부족했던 얼굴 영역의 정보를 모델에 추가로 반영하는 것이다. 이를 위해 세 가지 아이디어를 적용한다. 첫째로 마스크를 통해 머리카락 면적 속성을 추가하여 머리카락 정보를 보완한다. 둘째로 순환 일관성 손실(cycle consistency loss)을 추가하여 영상의 불필요한 변화를 억제한다. 셋째로 모자 분할 신경망을 추가하여 모자 영역 왜곡을 방지한다. 정성적 평가를 통해 제안하는 방법 적용 여부에 따른 유효성을 평가 및 분석한다. 실험 결과에서 제안하는 방법이 머리카락 및 얼굴 영역을 더 자연스럽게 생성하고, 모자 영역의 왜곡을 성공적으로 방지했다.

▸ **주제어**: 얼굴 속성 편집, GAN, 딥러닝, 마스크, STGAN

• First Author: Hyeon Seok Yang, Corresponding Author: Young Shik Moon
*Hyeon Seok Yang (hsyang@visionlab.or.kr), Dept. of Computer Science and Engineering, Hanyang University
*Jeong Hoon Han (bghan@visionlab.or.kr), Dept. of Computer Science and Engineering, Hanyang University
*Young Shik Moon (ysmoon@hanyang.ac.kr), Dept. of Computer Science and Engineering, Hanyang University
• Received: 2020. 04. 20, Revised: 2020. 05. 14, Accepted: 2020. 05. 14.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

얼굴 속성 편집(facial attribute editing)은 얼굴 영상의 속성을 원하는 대로 변경하는 방법에 대한 연구 분야이다. 이 분야는 GAN(Generative Adversarial Net)[1,2]을 포함한 딥러닝(deep learning) 모델의 연구 성과에 힘입어 점차 자연스러운 결과 생성에 성공하고 있다. 최근에는 주로 GAN과 인코더-디코더(encoder-decoder) 모델[3,4]에 기반한 모델들이 제안되고 있다[5-10].

최근 얼굴 속성 편집 모델 중 하나인 AttGAN[5]은 다수의 속성을 하나의 모델로 학습하고, 원하는 속성 이외의 속성은 변화하지 않도록 세 가지 규제를 사용하여 자연스러운 속성 편집을 수행하였다. 또한 STGAN[6]은 AttGAN의 인코더-디코더 구조에서 인코더의 특징이 변경하고자 하는 속성에 따라 선택적으로 디코더로 전달될 수 있게 하는 STU(Selective Transfer Unit)를 추가하여 품질 개선을 이루어냈다. 그러나 머리카락과 연관성이 있는 속성의 편집 할 때에 종종 부자연스러운 편집 결과가 생성된다. 또한 모자가 포함된 영상에서 머리카락과 관련된 속성을 변경하면 모자 영역이 변형된 부자연스러운 영상이 생성된다.

본 연구에서는 기존 STGAN의 한계점을 극복하기 위하여 세 가지 접근을 사용하였다. 첫째로 머리카락 영역의 마스크를 이용하여 머리카락 비율 속성(hair ratio attribute)을 추가함으로써 머리카락과 관련된 속성에서의 품질을 향상시킨다. 둘째로 순환 일관성 손실(cycle consistency loss)[12]을 추가하여 얼굴 속성 편집 과정에서의 과도한 변경 등의 부자연스러운 결과를 억제하였다. 셋째로 생성자(generator)에 모자 영역을 식별하는 모자 분할 신경망(hat segmentation network)을 추가하여 모자 영역의 왜곡을 줄였다.

본 논문의 구성은 다음과 같다. II 장에서는 GAN과 얼굴 속성 편집 등의 관련 분야를 소개한다. III 장에서는 제안하는 방법의 모델과 접근을 설명한다. IV 장에서는 제안하는 방법의 각 접근을 기존 STGAN과 비교하여 품질 개선을 확인한다. V 장에서는 연구 결과를 정리하고 결론을 맺는다.

II. Related Works

1. GANs

최근 얼굴 속성 편집은 GAN[1,2]과 인코더-디코더 구조를 활용한 모델들이 좋은 성과를 보인다[5-10]. GAN은 생성자와 식별자(discriminator)가 서로 경쟁적인 방식으로 학습을 하는 모델로, 생성자는 무작위 벡터를 입력받아서 사실적인 영상을 생성하는 것을 목표로 하고, 식별자는 생성자가 생성한 가짜 영상과 데이터 세트의 진짜 영상을 식별해내는 것을 목표로 한다. 이 과정을 통해 생성자는 점차 진짜 영상과 구별이 되지 않는 가짜 영상을 생성할 수 있게 된다.

2. Image-to-image translation

얼굴 속성 편집은 얼굴 도메인(domain)에서 입력된 영상을 원하는 속성을 조건으로 새로운 영상으로 변환하는 영상 변환(image-to-image translation)[12-14]의 일종으로 볼 수 있다. pix2pix[13]는 영상 변환을 위한 범용적 모델로 제안되었다. pix2pix는 인코더-디코더 형태의 모델로 다양한 종류의 영상 간 변환이 가능하다. 다만 학습할 때 입력 영상에 대응하는 정답 영상이 요구되는데 많은 영상 변환 문제에서는 입력과 대응되는 정답 영상을 획득하기 어렵다는 한계가 있다.

CycleGAN[12]는 pix2pix가 학습을 위해 입력 영상과 대응되는 정답 쌍이 있어야 하는 한계를 순환 일관성 손실을 적용하여 해결하였다. 순환 일관성 손실은 A 도메인의 영상을 B 도메인의 영상처럼 변환하는 함수를 학습하고, 반대로 B 도메인의 영상을 A 도메인의 영상처럼 변환하는 함수를 학습하여 A 도메인에서 B 도메인으로 변환 후, 다시 A 도메인으로 되돌렸을 때 영상이 원래 영상과 유사하도록 강제하는 손실 함수(loss function)이다.

3. Facial attribute editing

몇 가지 모델들은 얼굴 속성 편집이라는 문제에 집중하여 연구가 진행되었다. 특히 얼굴 속성 편집을 할 때의 왜곡을 방지하기 위하여 몇 가지 방법이 제안되었다.

ResGAN[7]은 전체 얼굴 영상을 수정하는 대신에 수정 전과 수정 후의 영상의 차이인 잔차 영상(residual image)을 학습하고 L1 노름 규제(L1 norm regularization)를 적용하여 제한된 영역에 변경이 이루어지도록 한다. SaGAN[8]은 공간적 주의 기제를 활용하여 두 개의 신경망으로 구성된 생성자를 통해 제한된 영역에 변경이 이루어지도록 한다. ResGAN과 SaGAN은 서로 다른 방식으로 원하지 않는 변경을 억제하였으나 하나의 모델로 하나의 속성만을 변경할 수 있다.

AttGAN[5]은 얼굴 속성 편집할 때에 변경하고자 하는 속성 이외의 속성이 변화하는 현상을 억제하는 것을 목표로 한 모델이다. AttGAN에서는 이를 위해 두 접근법을 사용한다. 첫째로 식별자로 생성된 영상의 속성이 원하는 속성을 갖도록 규제하는 속성 분류 제약(attribute classification

constraint)을 적용한다. 둘째로 입력 영상을 원래의 속성으로 영상을 생성했을 때 입력 영상에 가깝게 복원되도록 하는 복원 손실(reconstruction loss)을 사용한다. 이 두 접근을 통해 원하는 속성을 가지면서도 다른 속성이 변화하지 않게 억제하였다. 또한 하나의 모델로 동시에 다수의 속성을 학습하고, 다수의 속성 변환이 가능하다.

STGAN[6]은 기존 AttGAN에서 병목 구간에서 영상의 품질이 저하되는 현상을 개선한 연구이다. STGAN은 속성 벡터 값을 그대로 사용하는 대신에 변경할 속성 벡터와 기존 속성 벡터의 차이인 차속성 벡터(difference attribute vector)를 사용한다. 또한 변경하는 속성에 따라 선택적으로 인코더의 특징을 디코더로 전달할 수 있는 STU를 추가하여 품질을 개선하였다.

그러나 AttGAN과 STGAN은 여전히 배경이나 모자 등의 영역에서 일부 왜곡이 발생한다. 일부 연구[10,11]는 AttGAN에서 여전히 원하지 않는 변경이 발생하는 것을 방지하기 위하여 별도의 영상 분할 신경망으로 획득한 마스크를 통해 원하는 마스크 영역만을 수정하는 모델을 제안하였다. 이를 통해 배경 영역의 변경을 방지할 수 있음이 확인되었다. 본 연구에서는 STGAN에서 모자가 있는 영상에 적용되었을 때 발생하는 왜곡의 방지에 유사한 접근을 활용한다.

4. Dataset

얼굴 속성 편집을 위해서는 얼굴 영상별로 속성 정보를 제공하는 데이터 세트가 요구된다. 얼굴 속성 편집 분야에서 대표적으로 많이 사용되는 데이터 세트는 CelebA(large-scale Celebfaces Attributes)와 LFW(Labeled Faces in the Wild)다[15,16]. CelebA는 10,177명의 유명인의 얼굴 영상으로 구성된 총 202,559장의 데이터 세트이다. 각 영상마다 40개의 이진 속성(binary attribute)과 5개의 표지점(landmark)을 제공하고 있다. LFW는 5,749명으로 구성된 총 13,233장 규모의 데이터 세트로 다양한 환경에서의 얼굴 영상을 제공한다. 40개에서 73개의 이진 속성을 제공한다.

CelebA와 LFW는 여러 얼굴 영상과 속성 정보로 구성된 데이터 세트를 제공하지만, 얼굴 영역의 의미적 영상 분할 정보는 제공하지 않는다. 최근에 공개된 CelebAMask-HQ [17]는 CelebA 데이터 세트의 고해상도 영상 30,000장을 선별하여 얼굴 영역의 의미적 영상 분할 마스크 19종(모자, 머리카락, 피부, 눈 등)을 함께 제공하고 있다. 본 연구에서는 CelebAMask-HQ 데이터 세트를 활용하여 연구를 수행한다.

III. The Proposed Scheme

1. Goal and approach

제안하는 방법은 마스크 정보를 활용하여 STGAN의 결과를 개선하는 것을 목표한다. 제안하는 3가지 접근은 다음과 같다. 첫 번째로 머리카락의 면적을 컨트롤하기 위하여 머리카락 영역의 마스크를 활용해 머리카락 면적 속성을 추가한다. 두 번째로 순환 일관성 손실을 추가하여 편집 과정에서의 과도한 변형이나 왜곡을 억제한다. 세 번째로 모자를 쓴 얼굴의 머리카락 관련 속성 편집할 때의 왜곡을 해결하기 위하여 모자 영역을 의미적 영상 분할로 검출하여 마스킹한다.

2. Network structure

제안하는 방법의 신경망 구성은 Fig. 1과 같다. 구성은 크게 생성자 G와 식별자 D로 나누어진다. 생성자 G는 얼굴 영상을 원하는 속성을 반영하도록 편집하는 역할을 하고, 식별자 D는 편집된 영상이 사실적인지와 원하는 속성이 반영되었는지를 판정하는 역할을 하여, 서로 경쟁적인 방식으로 학습한다.

생성자 G는 얼굴 영상 x_s와 변경할 차 속성 att_{diff}를 입력받아서 편집된 얼굴 영상 $\hat{x_t}$를 출력한다. 차속성 att_{diff}은 원천 속성 att_s에서 목표 속성 att_t로의 변환을 차이로 표현한다. 아래 수식 (1)은 차속성 att_{diff}을 정의하고, 수식 (2)는 생성자를 정의한다.

$$att_{diff} = att_t - att_s, \tag{1}$$

$$\hat{x_t} = G(x_s, att_{diff}). \tag{2}$$

생성자 G는 내부적으로는 두 개의 신경망을 갖는다. 첫 번째 신경망은 얼굴 편집 신경망(face editing network) G_e이고, 두 번째 신경망은 모자 분할 신경망(hat segmentation network) G_m이다. 두 신경망은 마지막 계층을 제외하고 동일한 구조를 갖지만, 가중치는 공유하지 않는다. 모자 분할 신경망 G_m의 구조는 다른 의미적 영상 분할이 가능한 모델로 대체해도 무방할 것으로 추측되며 실험 편의상 얼굴 편집 신경망 G_e와 동일한 구조를 사용하였다. 얼굴 편집 신경망 G_e는 얼굴 속성을 편집하는 용도로 얼굴 영상 x_s와 변경할 속성 값 att_{diff}을 입력받아 편집된 얼굴 영상 x_e을 출력한다. 아래 수식 (3)은 얼굴 편집 신경망 G_e를 나타낸다.

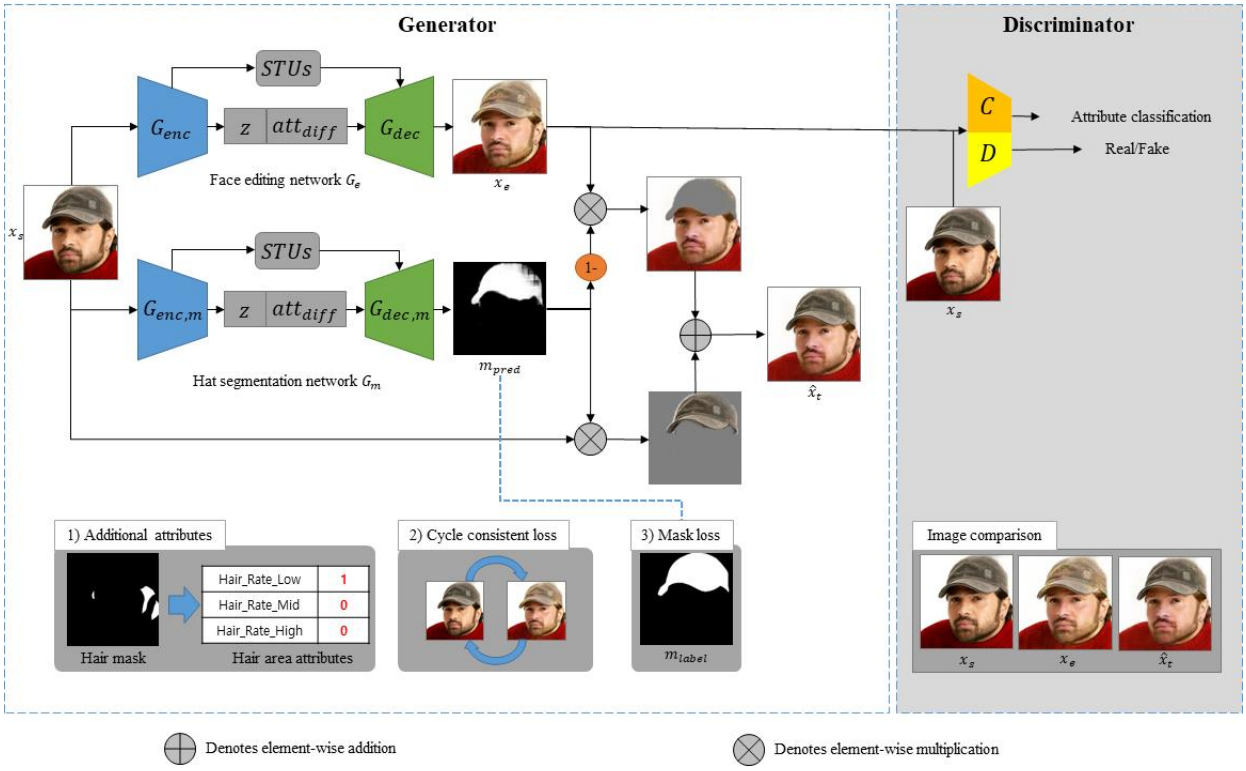


Fig. 1. Overview of the proposed method

$$x_e = G_e(x_s, att_{diff}). \quad (3)$$

모자 분할 신경망 G_m은 의미적 영상 분할 방법으로 모자 영역을 검출하기 위한 용도로 얼굴 영상 x_s와 변경할 속성 값 att_{diff}을 입력받아 예측한 모자 마스크 m_{pred}을 생성한다. 모자 분할 신경망 G_m의 수식 (4)는 다음과 같다.

$$m_{pred} = G_m(x_s, att_{diff}). \quad (4)$$

모자 마스크 m_{pred}은 모자로 판정된 영역은 1에 가까운 값을 갖고 그 외의 영역은 0에 가까운 값을 갖도록 학습한다. 모자 마스크 m_{pred}을 이용하여 편집된 얼굴 영상 x_e의 모자 이외의 영역과 입력 영상 x_s의 모자 영역을 합쳐서 최종 편집된 얼굴 영상 $\hat{x}_t$를 만든다. 최종 편집된 얼굴 영상 $\hat{x}_t$의 생성 과정의 수식 (5)는 다음과 같다.

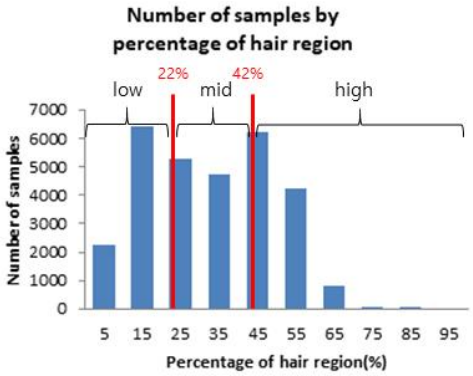
$$\hat{x}_t = x_e * (1 - m_{pred}) + x_s * m_{pred} \quad (5)$$

식별자 D는 영상이 사실적인지를 식별하는 D_{adv} 신경망과 영상이 어떤 얼굴 속성을 갖는지를 분류하는 D_{att} 신경망으로 구성된다. 두 신경망은 생성자로 편집된 얼굴 영상 x_e와 데이터 세트의 얼굴 영상 x_s를 입력 받으며, 마지막 완전 연결층을 제외하고는 가중치를 공유한다. 식별자의 입력으로 생성자의 최종 편집된 얼굴 영상 $\hat{x}_t$가 아니라 편집된 얼굴 영상 x_e를 사용하는 이유는 최종 편집된 얼굴 영상 $\hat{x}_t$를 사용하는 경우에 금발 속성의 적용이 모자로 마스킹 될 경우, 과도한 변경이 발생해 얼굴색이 심하게 왜곡되는 현상이 발생했기 때문이다. 과도한 변경의 원인은 금발 영역이 모자 마스크로 가려지면서 금발속성이 제대로 반영되지 않은 것으로 식별자가 판별하기 때문인 것으로 보인다. 때문에 모자 마스크의 적용을 별도로 수행함으로써 과도한 변경을 억제한다.

3. Attributes of hair region

얼굴 속성 편집에서 머리카락 영역은 다양한 속성과 연관성을 갖는다. 머리카락은 다양한 색상, 형태, 면적을 가질 수 있고 다른 영역(얼굴, 어깨, 배경 등)과 겹쳐질 수 있어서 편집이 어렵다. 만약 머리카락 마스크의 정보를 추가로 학습에 활용할 수 있다면 얼굴 속성 편집의 품질이 개선될 것으로 예상할 수 있다.

제안하는 방법에서는 데이터 세트를 영상 내 머리카락 면적의 크기에 따라 세 그룹으로 나누고, 속하는 그룹을 식별하는 세 속성(Hair_Rate_Low, Hair_Rate_Mid, Hair_Rate_High)을 추가하였다. 각 속성은 순서대로 머



(a) Distribution of 3 groups according to the percentage of the hair region

Hair_Rate_Low	Hair_Rate_Mid	Hair_Rate_High
0	0	1

(b) An example of adding attributes based on the percentage of the hair region

Fig. 2. The process of making hair region attributes

리카락 면적률의 적음, 보통, 많음을 의미한다. Fig. 2는 머리카락 비율 속성을 만드는 과정을 나타낸다. 먼저 전체 데이터 세트 30,000장을 머리카락 영역의 면적률로 정렬하고, 면적률에 따라 약 10,000장씩 3개의 그룹에 할당되도록 임계값을 정했다. 임계값은 22%와 42%로 결정되었다. 각 그룹은 대응하는 속성은 1로 설정하고, 나머지 두 속성은 0으로 설정한다. 예를 들어, Fig. 2의 샘플은 머리카락 면적률이 49%로 42%보다 높아서 머리카락 면적률이 높은 그룹에 속하므로 Hair_Rate_High는 1로 설정하고, Hair_Rate_ Low와 Hair_Rate_Mid는 0으로 설정한다.

또한 다른 속성과의 모순이 발생하지 않도록 속성 적용할 때에 연관된 속성을 변경하는 규칙을 추가한다. 세 속성(Hair_Rate_Low, Hair_Rate_Mid, Hair_Rate_High)은 한 번에 하나의 속성만 1이 되고, 나머지 속성은 0이 된다. Bald 속성을 1로 설정하는 경우에는 Hair_Rate_Low를 1로 설정한다. Bald 속성을 0으로 설정하는 경우에는 hair_Rate_Mid를 1로 설정한다.

4. Objective function

제안하는 방법에서는 기존 STGAN[5]의 목적 함수에 자연스러운 편집 결과의 생성을 위하여 순환 일관성 손실[10]과 모자 영상 분할을 위한 마스크 손실을 추가한다. 순환 일관성 손실 L_{cyc}은 원본 영상을 다른 속성의 편집 영상으로 변환한 뒤, 다시 원래 속성으로 되돌렸을 때 원본과의 차이를 측정한다. 아래 수식 (6)은 원본 영상을 다른 속성의 편집 영상으로 변환하는 과정으로, 수식 (3)과 동일하다. 수식 (7)은 순환 일관성 손실 L_{cyc}이다.

$$x_e = G_e(x_s, att_{diff}), \tag{6}$$

$$L_{cyc} = E[\|x_s - G_e(x_e, att_s - att_t)\|_1], \tag{7}$$

위 수식에서 $x_s \sim p_{data}$, $att_s, att_t \sim p_{att}$이며, p_{data}는 실제 영상의 분포, p_{att}은 속성의 분포를 나타낸다. 마스크 손실 L_{mask}은 정답 모자 마스크와의 차이를 측정하여 최소화하도록 학습시킨다. 아래 수식 (8)은 마스크 손실 L_{mask}이다.

$$L_{mask} = E[\|m_{pred} - m_{label}\|_2], \tag{8}$$

m_{label}은 x_s에 대응되는 정답 모자 마스크이며, 마스크 손실 함수는 예측한 m_{pred}와의 차이를 L2로 측정한다.

아래 수식 (9)는 생성자의 목적 함수이고, 수식 (10)은 식별자의 목적 함수이다.

$$\min_G L_G = -L_{G_{adv}} + \lambda_1 L_{G_{att}} + \lambda_2 L_{rec} + \lambda_3 L_{cyc} + \lambda_4 L_{mask}, \tag{9}$$

$$\min_D L_D = -L_{D_{adv}} + \lambda_5 L_{D_{att}}, \tag{10}$$

$\lambda_1, \ldots, \lambda_5$는 각각 손실 함수의 가중치를 뜻한다.

IV. Experimental Results

1. Experimental settings

제안하는 방법의 효과를 검증하기 위하여 기존 방법인 STGAN[6]과 제안하는 세 가지 접근의 조합을 비교 실험한다. 첫 번째 접근은 머리카락 면적 속성 추가, 두 번째 접근은 순환 일관성 손실 추가, 세 번째 접근은 모자 마스킹이다. 동등한 상황에서 비교하기 위하여 각 모델을 같은 데이터 세트에서 같은 에폭(epoch)만큼 학습하였다.

제안하는 방법의 실험 환경은 다음과 같다. 사용한 데이터 세트는 CelebAMask-HQ[17]이다. 학습은 데이터 세트에서 제공하는 학습 세트 20,000장에 대해 수행하였다. 테스트 또한 데이터 세트에서 제공하는 테스트 세트 5,000장에 대해 수행하였다. 사용한 모든 영상은 128×128 해상도로 정규화되었다. 학습에 사용한 속성은 16 가지(Bald, Bangs, Black_Hair, Blond_Hair, Brown_Hair, Bushy_Eyebrows, Eyeglasses, Male, Mouth_Slightly_Open, Mustache, No_Beard, Pale_Skin, Young, Hair_Rate_Low, Hair_Rate_Mid, Hair_Rate_High)이고, 이 중 세 가지(Hair_Rate_Low, Hair_Rate_Mid, Hair_Rate_High)는 제안하는 방법으로 추가한 속성이다.

Table 1은 제안하는 방법의 모자 분할 신경망 G_m의 구조이다. 나머지 신경망의 구조는 STGAN와 동일하다. 제안하는 방법의 파라미터 세팅은 배치 크기는 16이고 학습률은 2×10^{-4}로 설정하여 변경 없이 유지했다. 실험에 사용한 모든 모델들은 150 에폭씩 학습했다. 목적 함수의 가중치는 $\lambda_1=10, \lambda_2=100, \lambda_3=10, \lambda_4=100,$ 그리고 $\lambda_5=1$이다. 실험에 사용한 PC는 Windows 10, Intel core i7 6700 3.40GHz, NVIDIA GeForce RTX 2080 Ti, DDR 4 16Gb 메모리, 데이터 세트의 저장장치는 SSD를 사용하였다. 또한 아나콘다(Anaconda) 가상환경의 파이썬(Python) 3.6.10, 텐서플로(TensorFlow) 1.15.0 버전에서 테스트하였다.

Table 1. Architecture of hat segmentation network G_m

L	$G^l{}_{enc,m}$	$G^l{}_{dec,m}$
1	Conv(d,4,2), BN, Leaky ReLU	DeConv(1,4,2), Sigmoid
2	Conv(d*2,4,2), BN, Leaky ReLU	DeConv(d*2,4,2), BN, ReLU
3	Conv(d*4,4,2), BN, Leaky ReLU	DeConv(d*4,4,2), BN, ReLU
4	Conv(d*8,4,2), BN, Leaky ReLU	DeConv(d*8,4,2), BN, ReLU
5	Conv(d*16,4,2), BN, Leaky ReLU	DeConv(d*16,4,2), BN, ReLU

2. Evaluation of experiments

Fig. 3은 모자가 없는 샘플에 대해 머리카락 면적 속성 추가(A)와 순환 일관성 손실(C)의 조합을 실험한 경우이다. 가장 두드러지는 차이는 대머리(Bald) 속성을 추가한 경우(빨간 사각형)이다. STGAN은 눈 부분이 흐리게 변화하는 품질 저하가 일어났고, 두상도 다소 부자연스럽다. 제안하는 방법에서 속성 추가(A)와 순환 복원 손실(C)를 각각 적용한 경우에는 눈이 약간 흐려졌으나 STGAN보다는 뚜렷한 걸 확인할 수 있다. 그러나 두상은 부자연스럽다. 반면 두 접근을 함께 적용한 결과(A+C)에서는 눈이 흐려지지 않았으며

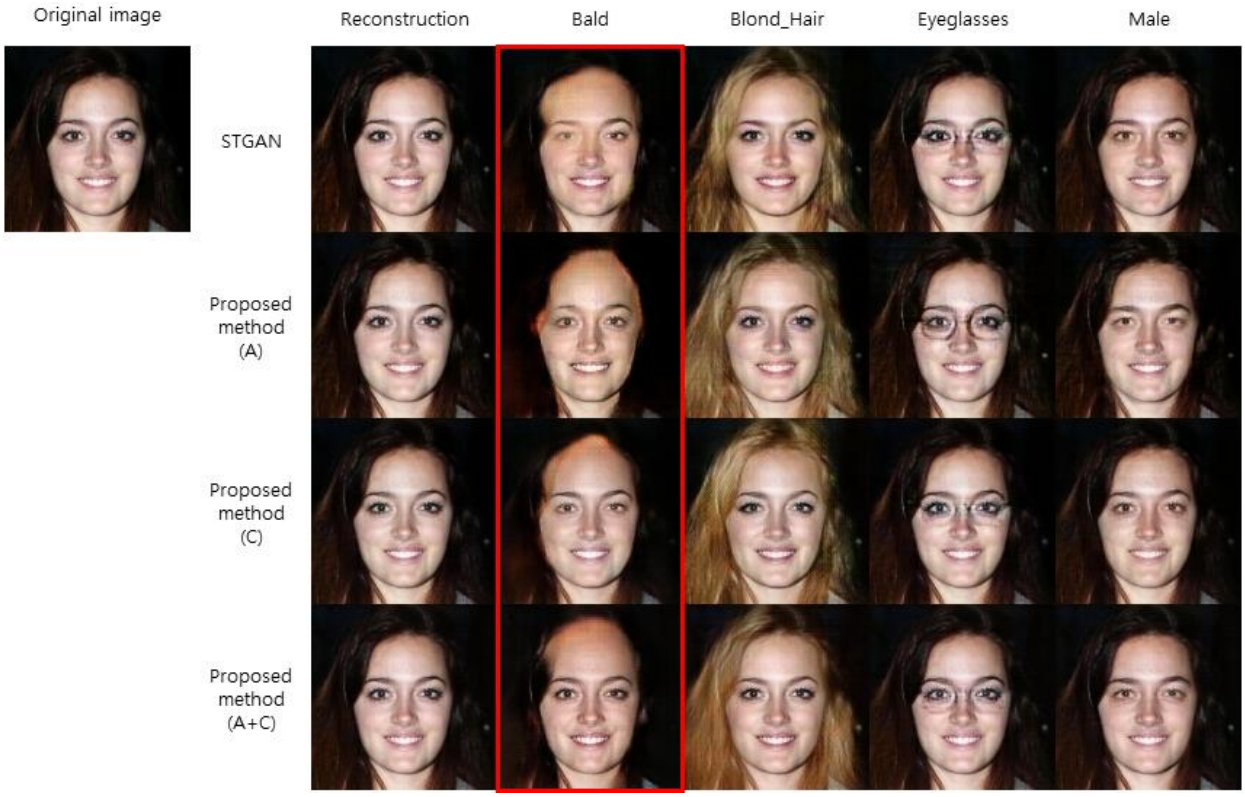


Fig. 3. Experimental results of face images without hat. It shows that the proposed method can achieve clearer and more natural results when applying Bald attribute(red square). 'A' means adding the hair ratio attributes. 'C' means the addition of cycle consistency loss.

두상도 자연스럽게 편집된 것을 확인할 수 있다. 금발(Blond_Hair) 속성은 기존 STGAN에서 일부 변경하지 못한 머리카락 영역을 제안하는 방법이 약간 더 수정하였다. 안경(Eyeglasses)은 큰 차이는 없으나 속성추가(A)의 경우가 좀 더 선명하게 표현되었다. 나머지 복원(Reconstruction)과 남성(Male) 속성은 거의 차이를 식별할 수 없다.

Fig. 4는 모자를 포함한 영상에 대하여 머리카락 면적 속성 추가(A), 순환 일관성 손실(B), 모자 마스크(C)의 조

Fig. 4. Experimental results of a face image with a hat. It shows that when applying the Bald or Blond_Hair attributes, the hat region can be masked to obtain a more natural result(red square). 'A' means adding the hair ratio attributes. 'C' means the addition of cycle consistency loss. 'M' means hat masking.

합을 실험한 경우이다. STGAN과 모자 마스크(C)를 적용하지 않은 경우들은 대머리(Bald) 속성과 금발(Blond_Hair) 속성의 적용할 때에 모자 부분이 왜곡된 것을 확인할 수 있다. 모자는 변경하고자 하는 속성과 무관하며 결과가 부자연스럽다. 반면 모자 마스크(M)을 적용한 방법들은 모자 영역이 대체로 보존되어 더 자연스러운 결과를 얻었다. 안경(Eyeglasses) 속성은 차이는 있으나 품질 차이는 크지 않았다. 여성(Female) 속성은 유사하되 입과 수염 부분에서 다소 차이가 있다. 머리카락 면적 속성 추가(A)할 때에는 입 주변 수염이 약간 흐려진 것을 볼 수 있다. 특히 머리카락 면적 속성 추가와 순환 일관성 손실을 함께 적용한 경우(A+C)에는 수염이 다른 접근보다 흐려진 것을 확인할 수 있다. 수염이 흐려진 이유는 여성 속성이 수염이 없음과 상관 관계가 있기 때문으로 보인다. 복원(Reconstruction)은 차이를 확인하기 어렵다.

V. Conclusions

본 논문에서는 머리카락 및 모자 영역의 마스크를 활용하여 더 자연스러운 얼굴 속성 편집을 수행하는 모델을 제안하였다. 제안하는 방법은 기존 최첨단(state-of-the-art) 모델인 STGAN을 개선하기 위한 세 가지 방법을 제안하였다. 첫째로 머리카락 마스크의 통계에 바탕한 머리카락 비율 속성을 추가하였고, 둘째로 순환 일관성 손실을 추가하였으며, 셋째로 모자 분할 신경망을 추가하여 모자 영역의 왜곡을 방지하였다. 기존 방법에 제안한 방법들의 적용 여부에 따른 결과를 확인하여 기존 모델에서 대머리 속성 적용할 때 얼굴 영역이 불명료하게 나타나거나 모자 영역이 왜곡되는 부자연스러운 결과를 제안하는 방법의 적용을 통해 개선할 수 있음을 보였다.

향후 연구로는 머리카락의 길이나 스타일을 변경하는 연구, 머리카락 제거할 때 가려진 영역을 자연스럽게 생성하는 연구, 모자나 의상 등을 변경하는 연구 등이 필요하다.

REFERENCES

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," Advances in neural information processing systems, pp. 2672-2680, Dec. 2014.

[2] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," arXiv preprint arXiv:1411.1784, pp. 1-7, Nov. 2014.

[3] G. Perarnau, J. V. D. Weijer, B. Raduanu, and J. M. Álvarez, "Invertible Conditional GANs for Image Editing," NIPS 2016 Workshop on Adversarial Training, pp. 1-9, Dec. 2016.

[4] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding Beyond Pixels using a Learned Similarity Metric," Proceedings of the 33rd International Conference on Machine Learning, Vol. 48, pp. 1-9, Feb. 2016.

[5] Z. He, W. Zuo, and S. Shan, "AttGAN: Facial Attribute Editing by Only Changing What You Want." IEEE Transactions on Image Processing, Vol. 28, No. 11, pp. 5464-5478, May. 2019. DOI: 10.1109/TIP.2019.2916751

[6] M. Liu, Y. Ding, M. Xia, X. Liu, E. Ding, W. Zuo, and S. Wen, "STGAN: A Unified Selective Transfer Network for Arbitrary Image Attribute Editing," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3673-3682, Jun. 2019.

[7] W. Shen and R. Liu, "Learning Residual Images for Face Attribute Manipulation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4030-4038, Jul. 2017. DOI: 10.1109/CVPR.2017.135

[8] G. Zhang, M. Kan, S. Shan, and X. Chen, "Generative Adversarial Network with Spatial Attention for Face Attribute Editing," Proceedings of the European Conference on Computer Vision (ECCV), pp. 417-432, Sep. 2018. DOI: 10.1007/978-3-030-01231-1_26

[9] H. S. Yang, J. H. Han, Y. C. Cho, H. G. Lee, Y. Park, and Y. S. Moon, "Study on Performance Improvement of SAGAN using Mask," Proceeding of 2019 Korea Signal Processing Conference, pp. 2557-2560, Sep. 2019.

[10] H. S. Yang and Y. S. Moon, "Face Attribute Editing using AttGAN and Guide Mask," 2019 International Conference on Electronics, Information, and Communication (ICEIC), pp. 1-3, Jan. 2019. DOI: 10.23919/ELINFOCOM.2019. 8706471

[11] P. Chen, Q. Xiao, J. Xu, X. Dong, and L. Sun, "Facial Attribute Editing using Semantic Segmentation," 2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), pp. 97-103, May 2019. DOI: 10.1109/HPBDIS.2019.8735455

[12] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks," Proceedings of the IEEE international conference on computer vision, pp. 2223-2232, Oct. 2017. DOI: 10.1109/ICCV.2017.244

[13] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efos, "Image-to-Image Translation with Conditional Adversarial Networks," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134, Jul. 2017. DOI: 10.1109/CVPR.2017.632

[14] Y. Choi, M. Choi, M. Kim, J. W. Ha, S. Kim, and J. Choo, "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8789-8797, Jun. 2018. DOI: 10.1109/CVPR.2018.00916

[15] X. Zheng, Y. Guo, H. Huang, Y. Li, and R. He, "A Survey to Deep Facial Attribute Analysis," International Journal of Computer Vision, pp. 1-33, Mar. 2020. DOI: 10.1007/s11263-020-01308-z

[16] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep Learning Face Attributes in the Wild," Proceedings of the IEEE International Conference on Computer Vision, pp. 3730-3738, Dec. 2015. DOI: 10.1109/ICCV.2015.425

[17] C. H. Lee, Z. Liu, L. Wu, and P. Luo, "MaskGAN: Towards Diverse and Interactive Facial Image Manipulation," arXiv preprint arXiv:1907.11922v2, pp. 1-20, Apr. 2020.

Authors

Hyeon Seok Yang received his B.S. degree in the Department of Electronics and Information Engineering from Yeungnam University, Korea, in 2010. He received the M.S. degrees in the Department of Computer Science & Engineering from Hanyang University, Korea, in 2012. He is studying for his PhD. degree in the Department of Computer Science & Engineering from Hanyang University, Korea. His research interests include computer vision, pattern recognition, and deep learning.

Jeong Hoon Han received his B.S. degree in the Department of Computer Science and Engineering from Hallym University, Korea, in 2016. He is currently working towards PhD. Degree at the Department of Computer Science and Engineering from Hanyang University, Korea, From 2016. His research interests include computer vision and machine learning.

Young Shik Moon received the B.S. and M.S. degrees in Electronics Engineering from Seoul National University and Korea Advanced Institute of Science and Technology, Korea, in 1980 and 1982, respectively, and PhD. degree in Electrical and Computer Engineering from the University of California at Irvine, CA, in 1990. From 1982 to 1985, he had been a researcher at the Electronics and Telecommunication Research Institute, Daejon, Korea. In 1992, he joined the Department of Computer Science and Engineering at Hanyang University, Korea, as an Assistant Professor, and is currently a Professor. Dr. Moon served as General Chair of 2014 IEEE International Symposium on Consumer Electronics, and worked as the President of the Institute of Electronics and Information Engineer, Korea.