졸업논문
소개

# 확산 모델 기반 얼굴 교체를 위한 속성 보존 의사 라벨링 연구

강지원 / KAIST 김재철AI대학원

얼굴 교체(Face Swapping)는 소스(source) 얼굴의 신원(identity)을 타깃(target) 얼굴로 옮기면서 타깃이 가진 자세, 표정, 피부색, 조명, 시선, 화장과 같은 고유한 속성을 보존하는 것을 목표로 한다. 영화 및 디지털 콘텐츠 제작, 프라이버시 보호 등 다양한 분야에 활용될 수 있지만, 실제로 동일한 장면과 속성을 유지한 채 사람의 신원만 변경한 정답 영상은 존재하지 않기 때문에 직접적인 지도 학습이 어렵다는 근본적인 문제가 있다. 따라서 높은 신원 유사성과 자연스러운 영상 품질을 동시에 확보하면서도 타깃의 세부 속성을 정확하게 유지하는 것이 얼굴 교체 연구의 핵심 과제이다.

초기의 얼굴 교체 연구는 주로 생성적 적대 신경망(GAN)을 이용하였으며, 소스의 신원을 전달하기 위한 신원 손실과 타깃의 모습을 유지하기 위한 복원 손실을 동시에 최적화하였다. 그러나 서로 상충하는 목적 함수를 조정해야 하기 때문에 학습이 불안정하고, 결과 영상에서 얼굴을 단순히 붙여 넣은 듯한 왜곡이나 부자연스러운 질감이 나타나는 한계가 있다. 최근에는 높은 생성 품질과 안정적인 학습이 가능한 확산 모델(diffusion model)을 얼굴 교체에 활용하는 연구가 활발하게 이루어지고 있다. 기존 확산 모델 기반 방법들은 주로 타깃 얼굴 영역을 가린 뒤 이를 다시 생성하는 조건부 인페인팅(conditional inpainting) 문제로 얼굴 교체를 정의한다. 그러나 얼굴 영역을 마스킹하면 타깃의 신원 정보뿐 아니라 피부색, 조명, 화장 및 미세한 표정과 같이 보존되어야 할 시각적 단서까지 함께 제거된다.

본 연구에서는 이러한 문제를 해결하기 위해 APPLE (Attribute-Preserving Pseudo-Labeling)을 제안한다. APPLE은 고품질의 속성 보존 의사 라벨(pseudo-label)을 생성하는 Teacher와 이를 활용해 학습하는 Student로 구성된 완전한 확산 모델 기반의 Teacher-Student 프레임워크이다. 핵심 아이디어는 손상된 타깃 영상을 직접 조건으로 사용하는 기존 방법과 달리, Teacher가 먼저 소스의 신원으로 교체되면서도 타깃의 시각적 속성이 최대한 유지된 영상을 생성하고, 이 영상을 Student 학습을 위한 깨끗한 조건 영상으로 활용하는 것이다. 이를 통해 Student는 얼굴 교체를 복잡한 복원 문제가 아닌 직접적인 영상 편집 문제로 학습할 수 있다.

먼저 속성 보존에 특화된 Teacher를 학습하기 위해 기존의 조건부 인페인팅 대신 조건부 디블러링(conditional

deblurring)을 도입한다. 타깃 얼굴 전체를 제거하는 대신 얼굴 영역을 저해상도로 축소한 후 다시 확대하여 흐리게 만들면, 신원을 구분하는 고주파 세부 정보는 제거하면서 피부색, 조명, 자세, 표정과 같은 저주파 속성 정보는 유지할 수 있다. 구체적으로 타깃 얼굴을 8×8 해상도로 축소한 뒤 원래 해상도로 복원하고, 얼굴 파싱을 통해 얻은 마스크 내부에만 블러를 적용한다. 또한 3D 얼굴 랜드마크뿐 아니라 시선 정보를 나타내는 눈 랜드마크와 안경 영역의 분할 정보 등을 조건에 추가하여 자세, 시선 및 액세서리와 같은 구조적 속성도보다 안정적으로 보존하도록 한다.

하지만 블러된 입력만으로는 화장이나 액세서리처럼 고주파 영역에 존재하는 미세한 속성을 완전히 복원하기 어렵다. 이를 해결하기 위해 본 연구에서는 속성 인지형 역변환(attribute-aware inversion)을 추가로 제안한다. 일반적인 확산 모델의 역변환을 통해 얻은 노이즈에는 입력 영상의 의미 정보가 일부 남아 있는데, 기존 연구에서는 이러한 정보가 영상 편집을 방해한다고 보고 제거하려는 경우가 많았다. 반면 얼굴 교체에서는 신원만 변경하고 나머지 속성을 유지해야 하므로, 이러한 잔여 정보를 적극적으로 활용할 수 있다. 이에 타깃의 신원 정보는 제외하고 속성 조건만을 이용해 타깃 영상을 역변환함으로써, 신원 교체 가능성을 유지하면서 화장과 같은 미세한 시각적 특성이 포함된 초기 노이즈를 생성한다. 실험에서도 속성 조건만 사용한 역변환이 신원 정보가 포함된 역변환보다 인공적인 왜곡을 줄이면서 세부 속성을 효과적으로 보존하는 것으로 나타났다.

이렇게 학습된 Teacher를 이용해 서로 다른 두 사람의 얼굴로 얼굴 교체를 수행하고, 생성된 결과를 의사 라벨로 사용하여 pseudo-triplet을 구성한다. Student는 소스 영상으로부터 신원 정보를 입력받고, Teacher가 생성한 깨끗한 의사 라벨 영상으로부터 타깃의 속성을 입력받아 원래 타깃 영상을 복원하도록 학습된다. 기존 방법에서 사용하는 마스크 또는 블러 영상과 달리 Student는 완전한 영상 형태의 조건을 사용할 수 있기 때문에 타깃의 속성을 직접적으로 학습할 수 있다. 또한 추론 시에는 Teacher에서 필요했던 얼굴 랜드마크나 블러링 등의 복잡한 전처리 과정 없이 타깃 영상 자체를 조건으로 사용할 수 있다는 장점이 있다.

제안 기법은 FLUX.1-Krea [dev] 기반의 확산 모델을 사용하여 학습하였으며, VGGFace2-HQ 데이터셋을 통해 Teacher와 Student를 학습하고 FFHQ의 1,000개 소스-타깃 쌍을 이용하여 평가하였다. 영상 품질은 FID, 신원 전달 성능은 신원 유사도와 retrieval 정확도, 타깃 속성 보존은 자세 및 표정 오차를 이용해 측정하였다. 실험 결과 APPLE Student는 FID 2.18, 신원 유사도 0.54, Top-1 신원 검색 정확도 90.50%, 자세 오차 1.85, 표정 오차 0.64를 기록하였다. 특히 비교한 기존 GAN 및 확산 모델 기반 방법 중 가장 낮은 FID와 자세·표정 오차를 보이면서도 경쟁력 있는 신원 전달 성능을 유지하였다. 또한 조건부 디블러링과 속성 인지형 역변환을 순차적으로 적용할수록 Teacher의 속성 보존 성능이 향상되었으며, APPLE Teacher가 생성한 의사 라벨을 이용한 Student가 기존 GAN 기반 Teacher의 의사 라벨을 이용한 경우보다 높은 속성 보존 성능을 보였다.

본 연구에서는 정답 얼굴 교체 데이터가 존재하지 않는 환경에서 고품질 의사 라벨의 생성과 활용을 통해 신원 전달과 속성 보존 사이의 문제를 해결하는 APPLE을 제안하였다. 조건부 디블러링을 이용하여 타깃의 전역적인 시각 속성을 보존하고, 속성 인지형 역변환을 통해 미세한 속성을 추가로 유지하며, 이를 통해 생성된 속성 정렬 의사 라벨로 Student를 학습하였다. 결과적으로 Student는 Teacher뿐 아니라 기존 얼굴 교체 방법보다 세밀한 타깃 속성을 충실히 유지하면서 높은 영상 품질과 신원 전달 성능을 달성하였다. 이는 고품질의 의사 데이터를 활용하는 Teacher-Student 학습이 확산 모델 기반 얼굴 교체의 속성 보존 성능을 효과적으로 향상시킬 수 있음을 보여준다.

졸업논문
소개



## 강 지 원

- 2025년 2월 : 고려대학교 컴퓨터학과 학사
- 현재 : KAIST 김재철AI대학원 석사과정
- 주관심분야 : 이미지/비디오 확산 생성 모델