

텍스트 조건 인페인팅을 이용한 얼굴 익명화

박소정*, 장석우*

*안양대학교 소프트웨어학과

e-mail: swjang7285@gmail.com

Face Anonymization Using Text-Conditioned Inpainting

So-Jeong Park*, Seok-Woo Jang*

*Department of Software, Anyang University

요약

본 논문은 Stable Diffusion Inpainting을 활용하여 텍스트 조건을 기반으로, 얼굴 영역만 새로운 신원으로 치환하되 핵심 속성과 장면 맥락을 유지하는 속성 유지형 얼굴 익명화 방식을 제안한다. 제안된 워크플로우는 첫째, 전처리 과정으로 MTCNN을 이용한 다중 얼굴 검출과 메타데이터화, DeepFace로 핵심 속성 추출 및 정규화, MediaPipe Face Mesh로 인페인팅용 정밀 마스크를 생성한다. 둘째, 조건 기반 생성 과정으로 속성 메타데이터 기반 긍정 및 부정 프롬프트를 구성하고, 파라미터와 랜덤 변형을 적용해 두 장의 후보 이미지를 생성한다. 마지막으로 두 장의 후보 중 임베딩 유사도, 속성 일치, 경계 품질 기준으로 최종 한 장을 선택해 좌표 기반으로 합성한다. 단일 인물 이미지인 CelebA와 다중, 군중 이미지인 WIDER FACE로 평가한 결과, 본 방식은 식별률 3.59%, 코사인 거리 0.89, 평균 속성 유사율 87%를 달성했으며, 기존 StarGAN v2 기반 방법 대비 유의미한 개선을 보였다. 제안된 방법은 참조 얼굴 없이 텍스트 조건만으로 제어 가능하고, 단일, 그룹 이미지에 동일 절차로 적용되어 재현성과 적용성이 높아, SNS 게시 이미지와 CCTV, 스트리밍 프레임의 외부 공유 등 실제 환경에서 개인정보 보호와 시각적 유용성 간 균형 있는 대안을 제공한다.

1. 서론

개인 촬영과 방송의 증가, 온라인 플랫폼의 확산으로 얼굴 이미지, 영상이 대규모로 축적 및 유통되고 있다. 동시에 얼굴 인식과 재식별 기술의 성능이 향상되면서, 데이터 분석 및 이미지 공유 과정에서 개인 신원이 노출될 위험이 증가한다. 전통적인 비식별화 기법인 모자이크와 블러는 실시간 영상에도 빠르고 쉽게 적용되지만, 얼굴에서 얻을 수 있는 의미 있는 정보를 모두 제거해 데이터 활용도를 크게 낮추고, 복원 공격에 취약하다는 한계가 있다[1]. 이런 배경에서 본 연구는 개인정보 보호와 시각적 유용성의 균형을 목표로, 얼굴 영역만 새롭게 생성 및 교체하여 다른 신원으로 치환하되, 성별, 나이, 인종, 표정 4가지 핵심 속성과 얼굴 영역 밖의 장면 맥락은 유지하는 텍스트 조건 기반 속성 유지형 얼굴 비식별 방식을 제안한다. 본 연구의 핵심 아이디어는 두 가지다. 첫째, 생성 모델의 부분 편집 능력으로 정체성 제거와 생성 얼굴의 자연스러움의 균형을 달성한다. 둘째, 속성 메타데이터를 조건으로 활용하여 익명화와 그 이후 분석, 학습에 필요한 통계적 속성을 유지한다.

본 연구는 Stable Diffusion Inpainting을 생성 모델로 사용한다[2]. 이 모델은 잠재 확산 구조로, 텍스트 임베딩을 U-Net 의

크로스 어텐션에 주입하고, 입력 이미지를 VAE 인코더로 잠재 공간에 투영한 뒤, 이진 마스크로 지정한 내부 영역만 반복적으로 갱신한다. 이때 마스크 외부 영역은 고정되어 장면 일관성이 유지되고, 마스크 내부는 텍스트 조건과 확산 역과정을 통해 새로운 얼굴로 치환된다[3]. 생성 시 가이던스 스케일과 스텝 수 조절로 품질과 다양성을 충족할 수 있어, 재현성과 제어 가능성을 보장할 수 있다. 본 논문에서는 Stable Diffusion Inpainting을 기반으로 속성 기반 프롬프트 설계 및 정밀 마스크 생성, 후보 선택을 통한 익명화 얼굴 생성 파이프라인을 제안한다.

2. 익명화 얼굴 생성

제안된 워크플로우는 크게 전처리, 조건 기반 생성, 선택 및 교체 세 가지 과정으로 나뉜다. 먼저 전처리 과정에서는 MTCNN으로 이미지 내 모든 얼굴을 검출 및 추출하고, 각 얼굴에 image id - face id를 부여한다. 그 뒤 원본 및 크롭 경로와 이미지 내 얼굴의 위치 좌표를 포함한 정보들을 CSV/JSON 메타데이터로 표준화하여 저장한다. 그룹 이미지의 확장성을 고려해 최소 얼굴 크기와 임계값은 회수율 중심으로 완화하고, 면적 비율 하한으로 얼굴 인식 및 선별이 힘든 소형 얼굴은 배제한다. 속성은

DeepFace로 성별, 나이, 인종, 표정 4가지 정보를 추출 및 정규화하여 메타데이터에 반영한다. 라벨 표기와 값 범위를 일치시켜 프롬프트 생성 및 평가에서 혼선을 줄이고자 하였다. 마스크는 MediaPipe Face Mesh로 윤곽을 기반으로 소폭 확장, 페더링하여 경계 이음새를 줄이고, 필요시 극성 자동 보정으로 파이프라인 요구와 일치시킨다. 좌표 클리핑과 고정 패딩을 병행하여 합성 단계의 위치 오차를 최소화한다.

조건 기반 생성 단계에서는 속성 메타데이터를 문장 템플릿으로 변환한 구조화 프롬프트를 구성한다. 긍정 프롬프트에는 핵심 속성과 촬영 품질 조건을 명시하고, 부정 프롬프트에는 배경, 헤어 등 부가 조건 변경, 과도한 보정 및 만화, 비현실적 스타일을 넣어 실사 이미지 및 마스크 외부 보존을 강화한다. 또한 결과의 다양성을 위해 얼굴의 정체성 기여 부위에 랜덤 변형을 소수 적용한다[4]. 각 face_id 당 두 후보를 생성한 뒤 임베딩 기반 유사도와 속성 일치율 기준으로 최종 이미지를 선택한다.

마지막으로 합성 단계에서는 좌표 기반으로 원래 얼굴을 제거한 뒤 새로 생성된 얼굴을 합성하고, 조절한 파라미터, 프롬프트, 시드 등을 로그로 저장한다. 이 방식으로 워크플로우가 단일 인물, 다중 인물 이미지에 모두 동일하게 적용되며, 추후 자동화 및 운영으로 확장될 수 있다.

사용 데이터셋은 단일 인물, 그룹으로 구분해 적용하였다. CelebA는 주로 단일 인물의 정면 얼굴 이미지로 구성되어, 속성 기반 프롬프트의 효과와 실험 결과 평가에 사용하였다. WIDER FACE는 다중 인물, 군중 장면에서 작거나 겹치는 얼굴이 많은 특성을 갖기 때문에, 검출, 마스크, 합성 파이프라인의 다수 얼굴 확장성 및 강건성을 검증하는 데 활용하였다.

3. 실험 결과

익명화 결과는 얼굴 유사도와 속성 유사도로 판단한다. 얼굴 유사도는 두 가지 지표로 해석한다. 첫째 식별률은 원본과 익명쌍이 동일인으로 판단된 비율로 값이 낮을수록 정체성 제거가 강함을 나타낸다[5]. 둘째, 코사인 거리는 얼굴 임베딩 방향 차이를 수치화한 값으로, 클수록 서로 다른 정체성으로 본다. 속성 유사도는 익명화 전후의 성별, 나이, 인종, 표정, 일치율로 산출하며, 나이는 실제 적용성을 고려해 5살 이내 오차를 허용한다. 제안된 파이프라인의 결과는 식별률 3.59%, 코사인 거리 0.89, 평균 속성 유지율 87%를 보였고, 특히 인종·표정 보존에서 기존 방법 대비 높은 개선을 확인했다. 시각적 평가에서도 배경, 헤어, 의복은 안정적으로 유지되면서 얼굴만 자연스럽게 치환되는 경향을 관찰하였다.

4. 결론

본 연구는 텍스트 조건 인페인팅을 이용하여 속성을 유지한 채 얼굴 영역만 새 신원으로 교체하는 얼굴 익명화 파이프라인을 제안하였다. 제안된 방법은 잠재 확산 기반 선택 영역 치환으로 장면 맥락을 보존하고, 속성 메타데이터로 프롬프트를 제어하며, 경계 품질을 고려한 마스크와 후보 생성($k=2$) 및 선택 규칙을 결합해 낮은 식별률과 높은 속성 유지율을 동시에 달성하였다. 또한 image_id-face_id와 메타데이터 관리로 그룹 이미지에 확장 적용되도록 설계했으며, SNS 게시 이미지나 CCTV, 스트리밍 추출 프레임 등 실제 활용 시나리오에서 적용 가능하다.

향후 연구로는 동영상으로 확장해 프레임 간 일관성, 추적을 결합하고, 실제 서비스 환경에서의 적용성을 더욱 높일 계획이다.

참고문헌

- [1] R. McPherson, R. Shokri, and V. Shmatikov, "Defeating Image Obfuscation with Deep Learning," arXiv, September 2016.
- [2] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 10684-10695.
- [3] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "RePaint: Inpainting using Denoising Diffusion Probabilistic Models," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 11461-11471.
- [4] O. Avrahami, D. Lischinski, and O. Fried, "Blended Diffusion for Text-driven Editing of Natural Images," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 18208-18218.
- [5] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019, pp. 4690-4699.