얼굴 인식 시스템의 재식별 실패 구간에 대한 실험적 분석

황승연*, 이상홍**, 장석우*
*안양대학교 소프트웨어학과
**안양대학교 컴퓨터공학과
e-mail:swjang7285@gmail.com

An Experimental Analysis of Re-identification Failure Regions in Face Recognition Systems

Seung-Yeon Hwang*, Sang-Hong Lee**, Seok-Woo Jang*
*Dept. of Software, Anyang University
**Dept. of Computer Engineering, Anyang University

요 약

최근 얼굴 인식 기술의 발전으로 다양한 환경에서 개인 식별이 가능해졌으나, 동시에 입력 데이터의 변형에 대한 모델의 강건성 문제도 중요한 연구 주제로 부각되고 있다. 특히 생성형 모델을 이용한 얼굴 변환은 시각적으로 자연스러운 결과를 유지하면서도 특징 공간에 변화를 유도할 수 있어서 얼굴 인식 모델의 취약성을 분석하는 데 유용한 도구로 활용될 수 있다. 본 연구에서는 디퓨전 기반 얼굴 변환을 입력 교란 방식으로 활용하여 얼굴 인식 모델의 재식별 성능 변화를 분석하였다. 변환 강도를 단계적으로 증가시키며 동일 인물 판정 비율을 측정한 결과, 재식별 성능은 점진적으로 감소하지 않고 특정 구간에서 급격히 붕괴하는 비선형적 특성을 보였다. 특히 중간 강도 구간에서 인식 성능이 급격히 저하되며 이후 구간에서는 거의 무작위 수준으로 수렴하는 경향이 나타났다. 이러한 결과는 얼굴 인식 모델이 특정 수준 이상의 표현 변화에 대해 안정적으로 동작하지 않음을 시사하며 특징 공간의 구조적 붕괴 가능성을 보여준다.

1. 서론

얼굴 인식 기술은 보안, 인증, 감시 등 다양한 분야에서 핵심적인 역할을 수행하고 있으며 최근 딥러닝 기반 모델의 발전으로 높은 수준의 인식 정확도를 달성하고 있다[1]. 그러나 이러한 성능은 입력 데이터의 분포가 학습 환경과 유사하다는 가정에 크게 의존하며 입력이 변형되거나 교란되는 경우 성능 저하가 발생할 수 있다[2].

기존의 연구들은 주로 적대적 공격이나 노이즈 기반 변형을 통해 모델의 취약성을 분석했지만, 이러한 방식은 시각적으로 자연스럽지 않은 결과를 생성하는 경우가 많고 실제 환경을 충분히 반영하지 못하는 한계가 있다[3]. 반면, 최근 발전한 생성형 모델은 입력 이미지의 전반적인 구조와 시각적 일관성을 유지하면서도, 내부의 세부 특징이나 표현을 유의미하게 변화시키는 특징을 갖는다[4]. 이러한 특성은 단순한 노이즈 추가나 왜곡 기반 변형과 달리 실제 환경에서 발생할 수 있는 자연스러운 변형을 모사할 수 있게 하며, 이에 따라 현실적인 조건에서 얼굴 인식 모델의 강건성을 평가할 수 있다[5]. 특히 생성 과정에서 변화의 정도를 단계적으로 제어할 수 있다는 점은 입력 변화에 따른 모델의 반응을 체계적으로 분석하는 데 있어 중요하다.

본 연구에서는 디퓨전 기반 얼굴 변환을 활용하여 얼굴 인식 모델의 재식별 성능 변화를 분석하고 특히, 인식 성능이 급격히 변화하는 구간의 존재 여부를 실험적으로 확인하고자 한다.

2. 연구 방법

본 연구에서는 얼굴 인식 모델의 입력에 대해 디퓨전 기반 변환을 적용하고, 변환 강도에 따른 재식별 성능 변화를 측정한다. 먼저, 입력 이미지에서 얼굴 영역 추출 및 정규화를 수행하고 디퓨전 모델을 이용하여 얼굴 변환을 수행한다. 그리고 변환된 이미지와 원본 이미지 간의 유사도를 계산한다. 얼굴 인식 모델은 입력 이미지를 임베딩 벡터로 변환하며 벡터 간 유사도를 기반으로 동일 인물 여부를 판단한다.

디퓨전 기반 변환 과정에서 노이즈 수준을 조절하여 얼굴의 표현 변화를 단계적으로 유도한다. 낮은 강도에서는 원본 특징이 비교적 유지되며 강도가 증가함에 따라 특징 표현의 변화가 커지는 것으로 가정한다.

재식별 성능은 동일 인물로 판정된 비율로 정의하였다. 각 강도

조건에서 원본 이미지와 변환 이미지 간의 유사도를 계산하고 사전에 설정된 기준값을 통해 동일 인물 여부를 판단한다.

3. 실험 결과

3.1 재식별 성능 변화

변환 강도가 증가함에 따라 얼굴 인식 성능은 단순한 선형 감소 양상을 보이지 않고 특정 구간에서 급격히 변화하는 특성이 관찰되었다. 낮은 강도 구간에서는 원본 얼굴의 주요 특징이 비교적 잘 유지되기 때문에 높은 재식별 성능이 지속되었으며 얼굴 인식 모델 또한 안정적인 판단을 수행하는 것으로 나타났다. 그러나 강도가 증가함에 따라 특징 표현이 점진적으로 변형되기 시작하며 특히, 중간 강도 구간에 도달하면서 재식별 성능이 급격히 감소하는 현상이 확인되었다. 이후 일정 수준 이상의 강도에서는 재식별 성능이 거의 유지되지 않는 상태로 수렴하였으며 동일 인물로 판정되는 비율이 매우 낮은 수준으로 감소하였다. 이는 단순히 특징 정보가 약화되는 수준을 넘어 얼굴 인식 모델이 활용하는 핵심 표현 자체가 더 이상 유효하지 않게 되는 상태에 도달했음을 의미한다. 이러한 결과는 얼굴 인식 모델이 입력 변화에 대해 점진적으로 반응하기보다는 특정 수준 이상의 변화에서 급격한 인식 실패를 보이는 비선형적 붕괴 특성을 가진다는 것을 시사한다. 즉, 입력 변화의 정도와 재식별 성능 간의 관계는 선형적이기보다 임계점을 중심으로 급격히 변하는 구조를 가지는 것으로 해석할 수 있다.

3.2 실패 구간 분석

중간 강도 구간에서 관찰된 급격한 성능 저하는 일종의 전이 현상으로 해석할 수 있다. 해당 구간에서는 입력 이미지의 시각적 구조는 유지되지만, 얼굴 인식 모델이 활용하는 특징 표현이 급격히 변형되면서 동일 인물로 인식되지 않게 된다.
이는 얼굴 인식 모델의 특징 공간이 일정 수준 이상의 변화에 대해 안정적으로 유지되지 않음을 시사한다.

4. 결론

본 연구에서는 디퓨전 기반 얼굴 변환을 활용하여 얼굴 인식 모델의 재식별 성능 변화를 분석하고 특정 구간에서 인식 성능이 급격히 붕괴하는 현상을 실험적으로 확인하였다. 특히, 변환 강도에 따른 재식별 성능이 단순히 점진적으로 감소하는 것이 아니라 특정 구간에서 급격한 전이를 보인다는 점을 확인하였으며 이를 통해 얼굴 인식 모델의 비선형적 반응 특성을 확인할 수 있었다. 이러한 결과는 얼굴 인식 모델이 일정 수준 이상의 입력 변화에 대해 취약할 수 있음을 보여주며 특히, 실제 환경에서 발생할 수 있는 자연스러운 변형에도 불구하고 안정적인 인식을 보장하기 위해서는 보다 강건한 특징 표현이 필요함을 시사한다.
향후 연구에서는 다양한 얼굴 인식 모델을 대상으로 동일한 분석을 수행하여 이러한 현상이 일반적으로 나타나는지 검증할 필요가 있으며 입력 변화에 대해 안정적으로 동작할 수 있는 강건한 모델 설계 및 학습 방법에 관한 연구가 추가적으로 요구된다. 또한 변환 강도와 인식 성능 간의 관계를 정밀하게 모델링하여 특정 응용 환경에 적합한 안정적인 동작 범위를 정의하는 연구도 필요할 것으로 판단된다.

참고문헌

[1] 이은진, 성정환, “얼굴 인식 기술을 활용한 실감형 인터랙티브 콘텐츠의 구현 – (르네마그리트 특별전) AR포토존을 중심으로”, 한국게임학회 논문지, 제 20권 5호, pp. 13-20, 2020년.
[2] 김영국, 임채현, 손민지, 김명호, “마스크를 착용한 얼굴 인식을 위한 방법 연구”, 한국IT정책경영학회 논문지, 제 12권 4호, pp. 1939-1944, 2020년.
[3] 신종환, 박찬미, 이헌주, 이성현, 이재광, “얼굴인식 기술을 적용한 실종자 식별 시스템 설계 및 구현”, 한국컴퓨터정보학회 논문지, 제 26권 2호, pp. 19-25, 2021년.
[4] 이소정, 최영우, “CycleGAN을 이용한 Unpaired 근적외선 얼굴 이미지 생성 및 평가”, 한국정보처리학회 논문지, 제 21권 3호, pp. 593-600, 2019년.
[5] 허석렬, 김강민, 이완직, “딥러닝 기반 얼굴인식 방문자 출입관리 시스템 설계”, 한국디지털정책학회, 제 19권 2호, pp. 245-251, 2021년.