

인간 및 AI 실험기반 비식별화 기법 성능 평가

ACK 2024

유예지, 이정연, 이지연, 양소윤, 최현우(지도교수)

2024.10.16

목차 페이지

1 서론

연구 배경과 문제점
데이터 보호와 활용의 균형 중요성

2 주요 목표

비식별화 기법의 성능 평가
인간과 AI 간 인식 차이 분석.

3 비식별화 기술 소개

마스킹 기법 설명
가우시안 블러 기법 설명
모자이크 기법 설명

5 평가 방법

인간 비식별 평가 방법
AI 재식별 평가 방법
객체 탐지와 이미지 품질 평가

7 결론

비식별화 기법의 적합성 제언
데이터 보호와 유용성의 균형 중요성

4 실험 구성

사용된 데이터셋 설명
비식별화 강도 설정
평가 방법론 소개

6 실험 결과

인간과 AI 비식별률 비교
객체 탐지 정확도 분석
이미지 품질 변화 분석

8 향후 연구 방향

서론



배경

- AI와 데이터 분석 기술의 발전으로 개인 프라이버시 침해 우려가 증가
 - 개인정보보호위원회에서 비정형데이터의 가명처리 가이드라인을 지킬 것을 권고

문제 제기

- 기존 비식별화 기법이 인간과 AI 모두에게 효과적인지에 대한 체계적인 평가가 부족

목적

- 주요 비식별화 기법의 성능을 인간과 AI 인식에 대해 평가하고, 데이터 보호와 활용의 균형을 맞추는 방법 제시.

목표

주요 목표

- 마스킹, 가우시안 블러, 모자이크 방법의 비식별화 성능을 평가

세부 목표

- 인간 평가자에 대한 비식별률 측정
- AI 인식 시스템에 대한 비식별률 측정
- 물체 감지 정확도에 미치는 영향을 평가
- PSNR 및 SSIM 메트릭을 통해 이미지 품질 저하 분석

중요성

- 개인정보 요구와 데이터 활용도에 따라 적절한 비식별화 방법을 선택하는 데 필요한 통찰력 제공

비식별화 기술 소개

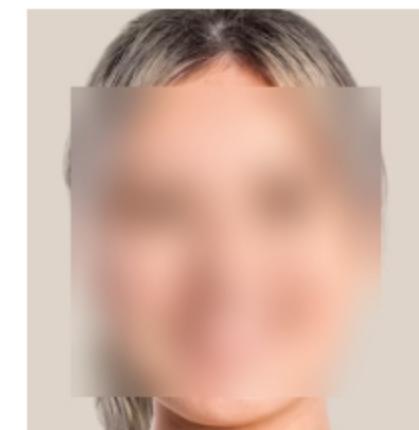
마스킹

- 이미지에서 얼굴 영역을 단색 또는 패턴으로 완전히 덮는 것
- 얼굴 위에 단색 모양을 오버레이하여 모든 얼굴 특징을 제거



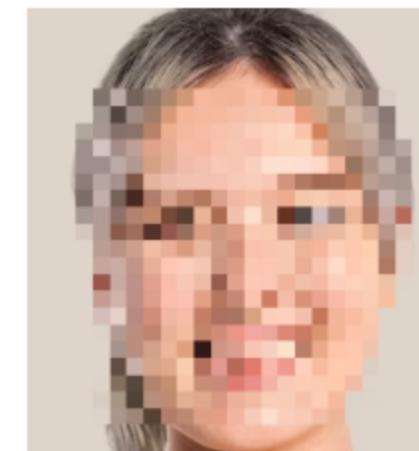
가우시안 블러

- 얼굴 영역에 흐림 필터를 적용하여 가우시안 함수를 기반으로 픽셀 값을 평균화하여 이미지를 흐리게 처리
- 미세한 디테일을 가리는 부드러운 효과가 나타남



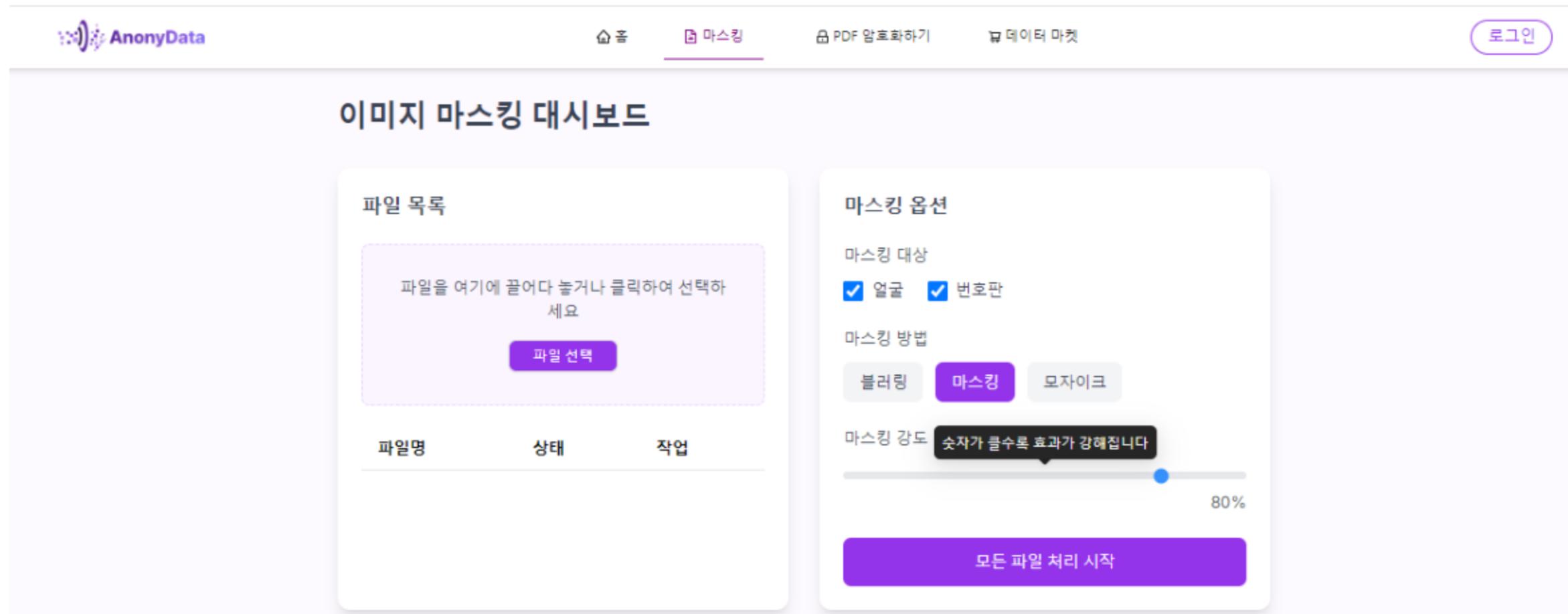
모자이크(픽셀화)

- 픽셀의 크기를 확대하여 얼굴 영역 내의 이미지 해상도를 낮춤
- 얼굴이 블록 형태의 디테일이 낮은 이미지로 표현



강도 설정

- 50에서 90까지 10(5단계) 단위로 조정
- 강도가 높을수록 비식별화 효과가 더 강해짐
- 직접 개발한 비정형데이터 비식별화 처리 웹 플랫폼에서 강도를 조절하며 비식별화 처리 진행



강도 설정 적용

가우시안 블러

- 공식: $\text{kernel_size} = \text{int}(\min(\text{height}, \text{width}) * (0.02 + 0.18 * (\text{intensity} / 100)))$
- 강도가 높을수록 흐림 정도 증가

$$R = R_{\min} + \frac{I}{100} \times (R_{\max} - R_{\min})$$

여기서 $R_{\min} = 0.02$, $R_{\max} = 0.2$.

모자이크

- 공식: $\text{pixel_size} = \text{int}(\min(\text{height}, \text{width}) * (0.05 + 0.25 * ((100 - \text{intensity}) / 100)))$
- 강도가 높을수록 더 큰 픽셀로 구성

$$R = R_{\min} + \frac{I}{100} \times (R_{\max} - R_{\min})$$

여기서 $R_{\min} = 0.05$, $R_{\max} = 0.3$.

모자이크는 이미지의 축소율이 클수록 효과가 높기 때문에, 강도에 반비례 관계(100-intensity) 관계를 적용

데이터셋 설명

선정 기준

- 대중 인식도가 높은 6명의 한국 유명인 선정.

이미지 수집

- 총 66장의 이미지 사용 (각 인물별 11장)
- 원본 이미지 1장
- 마스킹 1장, 가우시안 블러 5장, 모자이크 5장

사진 처리

- 플랫폼 내 기능을 활용하여, YOLO 모델을 사용해 얼굴 영역 탐지 후 비식별화 기법 적용



평가 방법

항목	기법 설명	지표 계산
인간 평가자 실험	<ul style="list-style-type: none">- 직접 제작한 웹 페이지를 통해 20명의 자원자가 비식별화된 이미지 평가- 참가자들은 각 이미지를 보고, "못 알아볼 것 같은 사진"을 선택- "인물 식별 불가"로 선택된 횟수를 기록	<ul style="list-style-type: none">- 비식별 선택률(HR)- $HR = (\text{선택된 횟수} / \text{총 노출 횟수}) * 100$
AI 재식별 실험	<ul style="list-style-type: none">- AI 동일인물 비교 사이트(https://facecomparison.toolpie.com)를 사용하여, 원본 이미지와 비식별화된 이미지 간의 유사도를 비교- 유사도가 80% 이상일 경우 동일인물로 판단	<ul style="list-style-type: none">- 재식별률 (RR)은 AI가 동일인물로 인식한 비율- $RR = (\text{동일인물로 판단된 경우의 수} / \text{총 시도 횟수}) * 100$
객체 탐지 정확도	<ul style="list-style-type: none">- OpenCV의 Face Recognition 모듈을 사용하여 비식별화 처리 후에도 얼굴 영역이 정확히 탐지되는지를 평가	<ul style="list-style-type: none">- 탐지 정확도 (DA)는 얼굴이 정확히 탐지된 이미지- $DA = (\text{정확히 탐지된 얼굴 수} / \text{총 이미지 수}) * 100$
이미지 선명도 평가	<ul style="list-style-type: none">- PSNR은 비식별화된 이미지와 원본 이미지 간의 차이를 평가하는 지표- SSIM은 이미지의 구조적 유사성을 평가하는 지표	<ul style="list-style-type: none">- PSNR은 값이 높을수록 두 이미지의 차이가 적다는 것을 의미- SSIM은 1에 가까울수록 원본 이미지와 유사함을 나타냄.

인간 평가자 실험 세부사항

참가자

- 연령과 배경이 다양한 20명의 참가자들이 실험에 참여

절차

- 각 참가자는 웹 플랫폼에서 총 66장의 비식별화된 이미지를 무작위 순서로 노출
- 참가자들에게는 "인물을 알아볼 수 없는" 이미지를 선택하도록 요청

데이터 수집

- 각 이미지가 "인식 불가"로 선택된 횟수를 기록

비식별 선택률 (HDR) 계산

- $\text{HDR} = (\text{선택된비식별이미지수}/\text{총노출횟수}) \times 100$

이미지 비식별화 효과 실험

Stage 1: 못 알아볼 것 같은 사진 선택
못 알아볼 것 같은 사진을 모두 골라주세요.



테스트 결과

전체 정확도: 75.00%

비식별화 방법별 정확도:

blur: 66.67%
mask: 66.67%
pixel: 100.00%

비식별화 방법 및 강도별 정확도:

blur (50%): 0.00%
blur (75%): 66.67%
mask: 66.67%
pixel (50%): 100.00%
pixel (75%): 0.00%

인물별 정확도:

P1: 0.00%
P2: 100.00%
P3: 100.00%
P4: 100.00%
P5: 0.00%
P6: 100.00%
P7: 100.00%
P8: 100.00%

개별 테스트 결과:

테스트 1 (blur 75%): 선택 P3 (정답: P1) - 오답
테스트 2 (mask): 선택 P2 (정답: P2) - 정답
테스트 3 (pixel 50%): 선택 P3 (정답: P3) - 정답
테스트 4 (blur 75%): 선택 P4 (정답: P4) - 정답
테스트 5 (mask): 선택 P8 (정답: P5) - 오답
테스트 6 (pixel 50%): 선택 P6 (정답: P6) - 정답
테스트 7 (blur 75%): 선택 P7 (정답: P7) - 정답
테스트 8 (mask): 선택 P8 (정답: P8) - 정답

AI 재식별 실험 세부사항

사용된 AI 툴

- AI 얼굴 비교 사이트를 사용하여 99% 이상의 정확도로 재식별 실험을 진행

절차

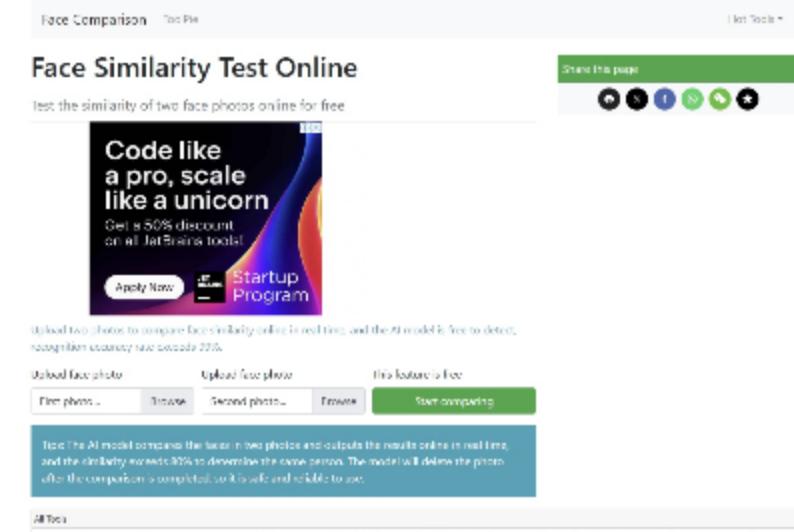
- 비식별화된 이미지를 원본 이미지와 비교하여 유사도 점수를 계산

재식별 기준

- 유사도 점수가 80% 이상이면 AI가 인물을 인식했다고 판단

AI 비식별화율 (ADR) 계산

- $ADR = (\text{인식실패횟수} / \text{총비교횟수}) \times 100$



객체 탐지 정확도 평가

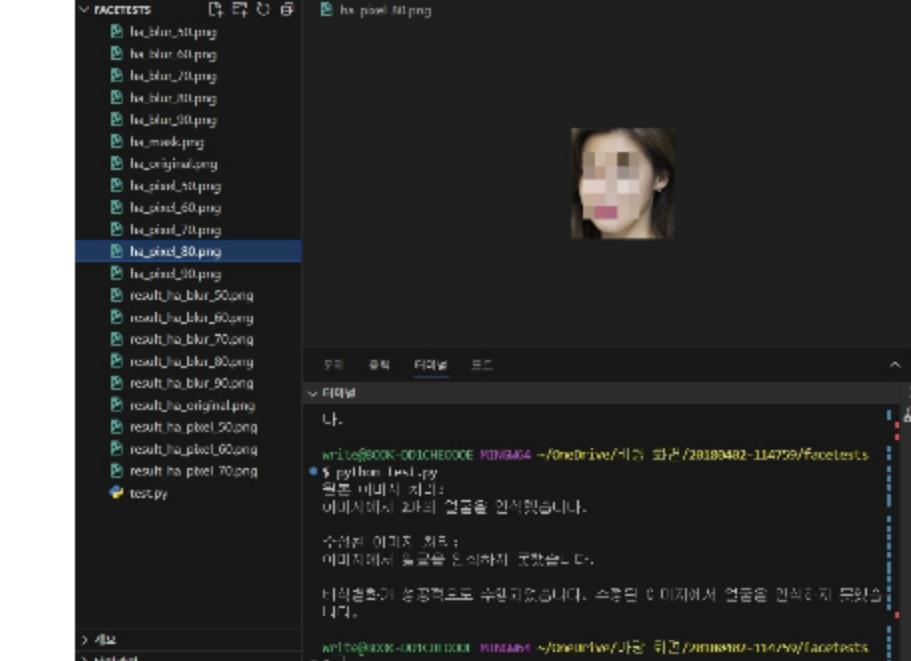
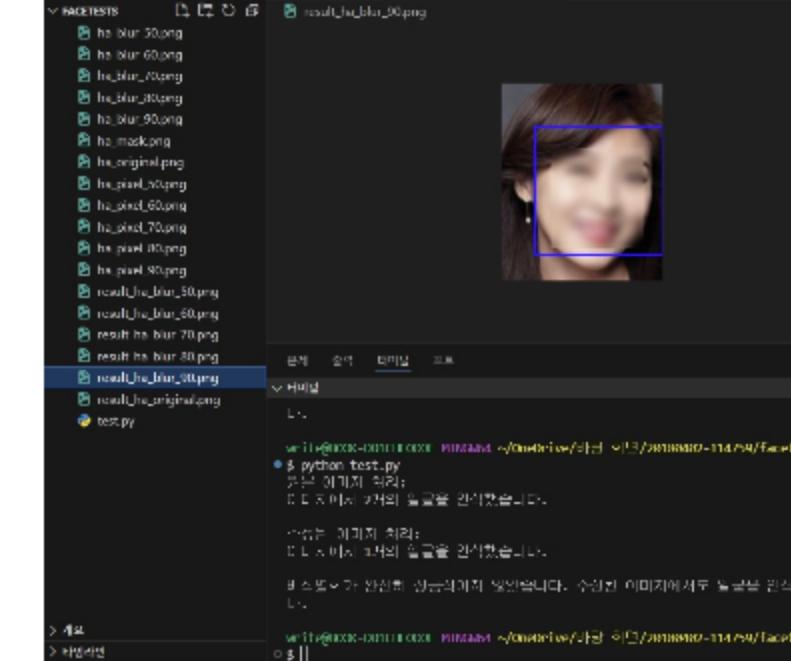
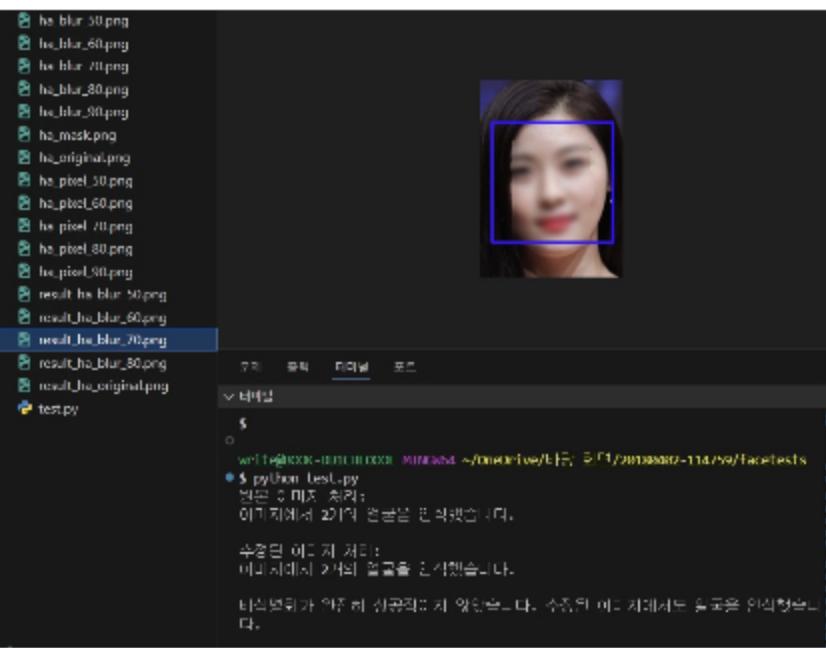
평가 방법

목적

- OpenCV의 얼굴 탐지 모듈을 사용하여 비식별화된 이미지에서 얼굴 탐지가 가능 한지 확인
- 비식별화 후에도 얼굴 탐지가 가능한지, 데이터 유용성을 유지할 수 있는지 평가

탐지 정확도 (DA) 계산

- DA=(정확히 탐지된 얼굴수/총이미지수)×100



이미지 품질 지표

PSNR (피크 신호 대 잡음비)

- 픽셀 차이를 측정하며, 값이 높을수록 원본과 유사

SSIM (구조적 유사성 지수)

- 구조적 유사성을 평가하며, 값이 1에 가까울수록 이미지가 원본과 유사

목적

- 비식별화로 인해 이미지 품질이 얼마나 저하되는지를 정량화

실험 결과

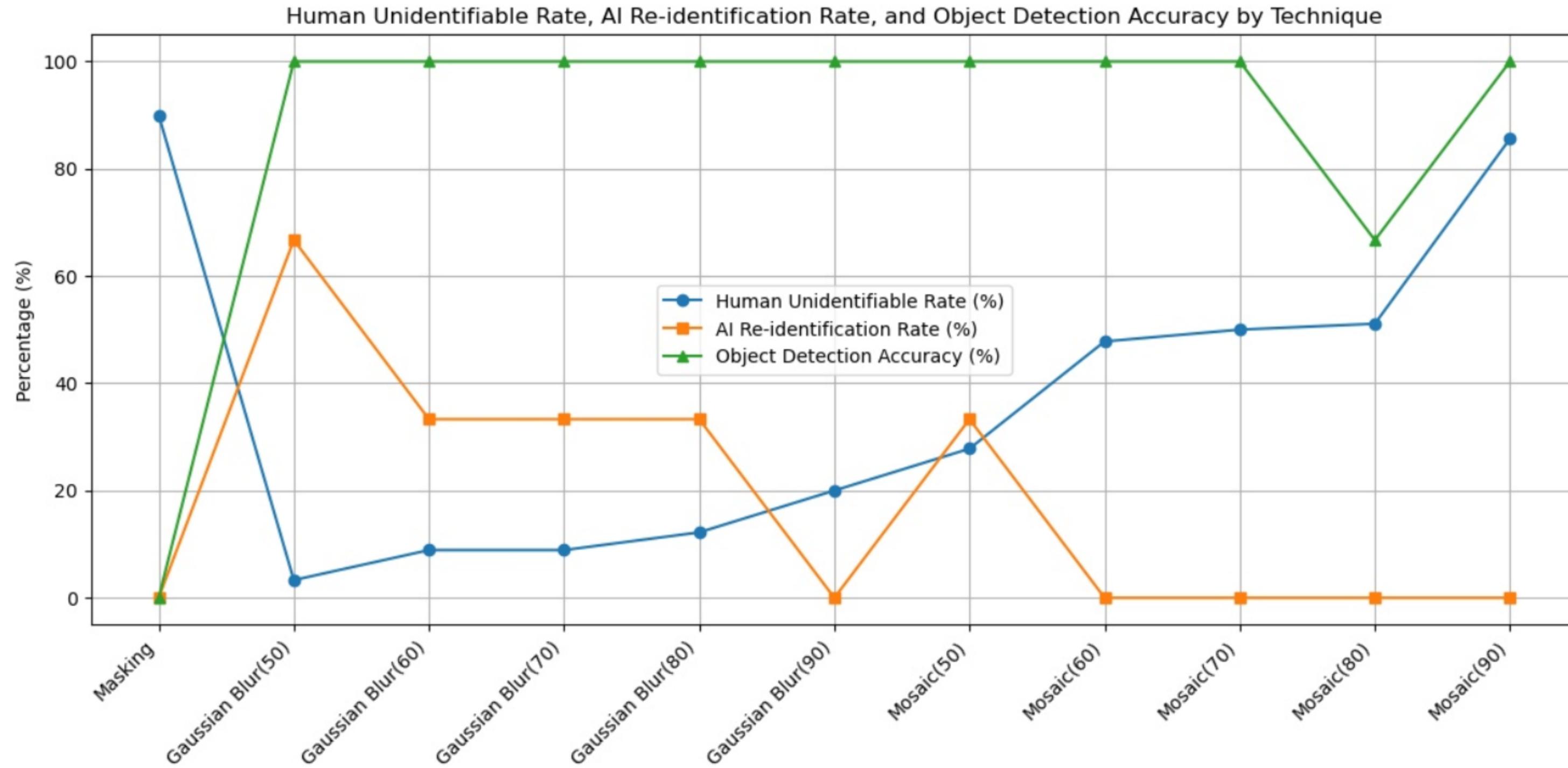
기법	인간 비식별률	AI 비식별률	객체 탐지 정확도	PSNR (dB)	SSIM
마스킹	90.0%	100.00%	0.00%	9.71 dB	0.6312
가우시안 블러(50)	3.3%	33.3%	100%	31.27 dB	0.9543
가우시안 블러(60)	8.9%	66.7%	100%	30.94 dB	0.9359
가우시안 블러(70)	8.9%	66.7%	100%	28.80 dB	0.9258
가우시안 블러(80)	12.2%	66.7%	100%	29.17 dB	0.9200
가우시안 블러(90)	20.0%	100.0%	100%	27.27 dB	0.8901
모자이크(50)	27.8%	66.7%	100%	29.53 dB	0.9057
모자이크(60)	47.8%	100%	100%	27.60 dB	0.8983
모자이크(70)	50.0%	100%	100%	24.82 dB	0.8419
모자이크(80)	51.1%	100%	66.7%	23.67 dB	0.8183
모자이크(90)	85.6%	100%	100%	24.49 dB	0.8486

HDR, ADR, DA, PSNR, SSIM 등 각 기법 및 강도별 주요 결과

결과 분석

- 마스킹: HDR과 ADR이 가장 높았으나, DA는 가장 낮음.
- 가우시안 블러 및 모자이크: 강도에 따라 효과가 다르게 나타남.
- 강도가 높을수록 이미지 품질이 감소

실험 결과



세부 분석 - 마스킹

인간 비식별화율 (HDR)

- 90%의 높은 비식별화 성능을 보임

AI 비식별화율 (ADR)

- AI 재식별에 대해 100%의 완벽한 비식별화 성능을 보임

객체 탐지 정확도 (DA)

- 얼굴 탐지가 불가능하여 0%로 나타남

이미지 품질

- PSNR: 9.71 dB
- SSIM: 0.6312

세부 분석 - 가우시안 블러

강도 증가에 따른 경향

- 인간 비식별화율 (HDR): 3.3%에서 20%로 증가.
- AI 비식별화율 (ADR): 강도 90에서만 100% 달성

객체 탐지 정확도 (DA)

- 모든 강도에서 100%로 유지됨

이미지 품질

- PSNR: 31.27 dB에서 27.27 dB로 감소
- SSIM: 0.9543에서 0.8901로 감소

세부 분석 - 모자이크

강도 증가에 따른 경향

- 인간 비식별화율 (HDR): 27.8%에서 85.6%로 급격히 증가.
- AI 비식별화율 (ADR): 강도 60부터 100% 달성

객체 탐지 정확도 (DA)

- 대부분의 강도에서 100% 유지되었으나, 강도 80에서 66.7%로 일시적 감소

이미지 품질

- PSNR: 29.53 dB에서 24.49 dB로 감소
- SSIM: 0.9057에서 0.8486으로 감소

기법 비교 분석

인간 vs. AI 비식별화율

- AI는 더 높은 강도가 필요하며, 인간보다 더 어렵게 비식별화되는 경향을 보임

트레이드오프

- 강도가 높아질수록 프라이버시 보호는 향상되지만, 이미지 품질은 저하됨
- 객체 탐지는 마스킹을 제외한 모든 기법에서 가능

의미

- 기법과 강도의 선택은 프라이버시 보호와 데이터 유용성에 대한 요구 사항에 따라 달라져야 함

결론

핵심 요점

- **마스킹:** 프라이버시 보호에 가장 효과적이지만, 데이터 유용성은 크게 감소.
- **가우시안 블러:** 프라이버시와 데이터 유용성의 균형을 제공하며, 적절한 강도가 필요.
- **모자이크:** 높은 강도에서 프라이버시와 데이터 유용성 간의 최적 균형을 제공

추천 사항

- 고위험 데이터의 경우, 마스킹을 선택.
- 균형이 필요한 응용에는 적절한 강도의 모자이크 기법 사용.
- AI 재식별 가능성을 고려해 기법을 선택

향후 연구

다음 단계

- 새로운 비식별화 기법 및 하이브리드 기법 탐구.
- 더 크고 다양한 데이터셋을 활용한 추가 실험 진행.
- 고급 AI 재식별 기법을 연구하여 방어 전략 강화



장기 목표

- 인간과 AI 인식 능력을 모두 고려한 맞춤형 비식별화 전략 개발

감사합니다