인간 및 AI 실험 기반 비식별화 기법 성능 평가

유예지[1], 이정연[1], 이지연[2], 양소윤[3], 최현우[4]
[1]성신여자대학교 융합보안공학과 학부생
[2]성신여자대학교 컴퓨터공학과 학부생
[3]성신여자대학교 수리통계데이터사이언스학부생
[4]성신여자대학교 융합보안공학과 교수
20200936@sungshin.ac.kr, 20200948@sungshin.ac.kr, 20201006@sungshin.ac.kr, 20230875@sungshin.ac.kr,
zemisolsol@sungshin.ac.kr

Performance Evaluation of Human and AI Experiment-Based Anonymization Techniques

Ye-Ji Yu[1], Jung-Yeon Lee[1], Ji-Yeon Lee[2], So-Yun Yang[3], Hyun-Woo Choi[4]
[1]Dept. of Convergence Security Engineering, Sungshin Women's University
[2]Dept. of Computer Engineering, Sungshin Women's University
[3]Dept. of Mathematical Statistics and Data Science, Sungshin Women's University
[4]Dept. of Convergence Security Engineering, Sungshin Women's University

요 약

본 연구는 마스킹, 가우시안 블러, 모자이크 비식별화 기법의 성능을 평가했다. 인간/AI 비식별률, 객체 탐지 정확도, 이미지 품질을 분석한 결과, 마스킹은 높은 비식별화 성능을 보였으나 데이터 유용성이 낮았다. 가우시안 블러와 모자이크는 강도에 따라 비식별화와 데이터 유용성 간 균형을 제공했다. AI 비식별률이 인간 비식별률보다 높아, 현재 기법들이 AI 기반 인식에 더 효과적임을 확인했다. 이는 자율주행차 학습 등 AI 응용 분야에서의 활용 가능성을 시사한다.

1. 서론

데이터 기반 의사결정과 AI 기술의 발전으로 개인정보의 보호와 활용 사이의 균형이 중요한 과제로 대두되고 있다[1]. 비식별화는 이러한 균형을 위한 핵심 기술이지만, 다양한 기법의 효과성에 대한 체계적인 평가가 부족한 실정이다. 본 연구는 주요 비식별화 기법의 성능을 인간과 AI의 재식별 시도, 객체 탐지 정확도, 이미지 품질 측면에서 종합적으로 분석함으로써, 데이터의 안전한 공유와 활용을 위한 최적의 전략을 제시하고자 한다.

2. 실험 설계 및 평가 방법

2.1 실험 시스템 및 데이터셋

React.js 프론트엔드와 Django REST Framework 백엔드로 구성된 웹 플랫폼을 개발하여 실험을 진행했다. 10000장의 사람 얼굴 이미지로 학습시킨 YOLO 모델로 얼굴을 탐지하고 OpenCV로 이미지를 처리했다.[2][3] 데이터셋으로는 6명의 한국 유명인(김지원, 서강준, 하지원, 엄정화, 권상우, 이이경) 얼굴 이미지를 각 11장씩, 총 66장을 사용하였다.

2.2 비식별화 기법 및 강도 설정

마스킹(대상 영역을 완전히 가림), 가우시안 블러(가우시안 함수로 이미지를 흐리게 함), 모자이크(이미지를 큰 픽셀로 재구성)를 사용했다. 가우시안 블러와 모자이크의 강도는 50-90 범위에서 5단계(50, 60, 70, 80, 90)로 설정했으며, 이는 비식별화 처리의 정도를 나타낸다. 강도 I에 따른 실제 적용 비율 R은 다음 수식으로 계산된다: R = Rmin + (I / 100) * (Rmax - Rmin) 여기서 Rmin과 Rmax는 각각 최소, 최대 적용 비율이다. 가우시안 블러의 경우 Rmin = 0.02, Rmax = 0.2이며, 모자이크의 경우 Rmin = 0.05, Rmax = 0.3이고 (100 - I)를 사용한다.

2.3 평가 방법

(1) 인간 평가자 실험: 직접 제작한 웹 페이지를 통해 자원한 실험자 20명에게 "못 알아볼 것 같은 사진"을 선택하게 하고, 비식별 선택률(HR)을 계산했다. HR = (선택된 횟수 / 총 노출 횟수) * 100

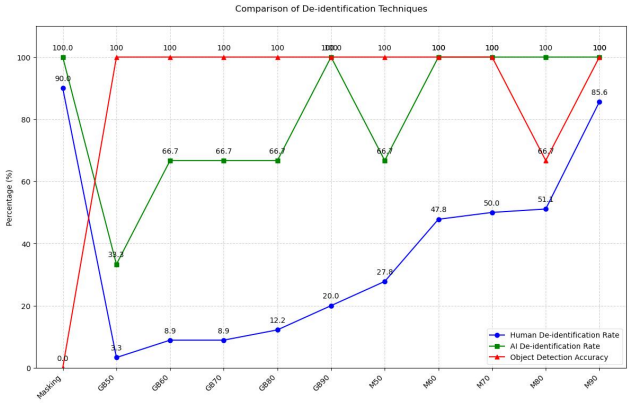
(2) AI 재식별 실험: AI 동일인물 비교 사이트(https://facecomparison.toolpie.com)를 사용하여 원본과 비식별화 이미지의 유사도를 비교했다.[4] 유사도 80% 이상을 동일인물로 판단하고, 재식별률(RR)을 계산했다. RR = (동일인물로 판단된 경우의 수 / 총 시도 횟수) * 100

(3) 객체 탐지 정확도: OpenCV의 Face Recognition 모듈을 사용하여 얼굴 영역 탐지를 수행하고, 탐지 정확도(DA)를 계산했다. DA = (정확히 탐지된 얼굴 수 / 총 이미지 수) * 100

(4) 이미지 선명도 평가: PSNR(Peak Signal-to-Noise Ratio)과 SSIM(Structural Similarity Index)을 사용했

다.[5] PSNR은 높을수록, SSIM은 1에 가까울수록 원본 이미지와 유사함을 의미한다.

3. 실험 결과 및 분석



(그림 1) 비식별화 기법별 인간/AI 비식별률 및 객체 탐지 정확도 비교

Technique	Human De-identification Rate	AI De-identification Rate	Object Detection Accuracy	PSNR (dB)	SSIM
Masking	**90.0%**	**100.00%**	**0.00%**	**9.71 dB**	**0.6312**
Gaussian Blur (50)	**3.3%**	**33.3%**	**100%**	**31.27 dB**	**0.9543**
Gaussian Blur (60)	**8.9%**	**66.7%**	**100%**	**30.94 dB**	**0.9359**
Gaussian Blur (70)	**8.9%**	**66.7%**	**100%**	**28.80 dB**	**0.9258**
Gaussian Blur (80)	**12.2%**	**66.7%**	**100%**	**29.17 dB**	**0.9200**
Gaussian Blur (90)	**20.0%**	**100.0%**	**100%**	**27.27 dB**	**0.8901**
Mosaic (50)	**27.8%**	**66.7%**	**100%**	**29.53 dB**	**0.9057**
Mosaic (60)	**47.8%**	**100%**	**100%**	**27.60 dB**	**0.8983**
Mosaic (70)	**50.0%**	**100%**	**100%**	**24.82 dB**	**0.8419**
Mosaic (80)	**51.1%**	**100%**	**66.7%**	**23.67 dB**	**0.8183**
Mosaic (90)	**85.6%**	**100%**	**100%**	**24.49 dB**	**0.8486**

<표 1> 비식별화 기법별 성능 지표

마스킹은 가장 높은 인간 비식별률(90.0%)과 완벽한 AI 비식별화를 보였으나, 객체 탐지가 불가능했다. 가우시안 블러와 모자이크는 강도 증가에 따라 인간 비식별률과 AI 비식별률이 향상되었다.

AI 비식별률이 인간 비식별률보다 전반적으로 높게 나타났으며, 기존의 비식별화 기술들이 AI를 활용한 식별 방식에 대해 더욱 효과적으로 작용함을 볼 수 있었다. 또한 마스킹 기법 외의 대다수의 비식별화 방법들에서 객체 탐지의 정확도가 상당히 높은 수준을 유지하는 것으로 관찰되었다.

높은 객체 탐지 정확도와 AI 비식별률은 데이터 활용 면에서 큰 가치를 가진다. 이는 개인을 식별하지 않으면서도 사람의 존재를 감지해야 하는 자율주행차 학습 등의 AI 응용 분야에서 효과적으로 활용될 수 있다. PSNR과 SSIM 결과는 비식별화 강도와 이미지 품질 사이의 균형에 관한 관계를 보여준다. 강도 증가에 따라 이 값들이 감소하지만, 가우시안 블러와 모자이크는 중간 강도에서 상대적으로 높은 값을 유지하였다. 이는 개인정보보호와 데이터의 유용성, 이미지 품질 사이의 상호 보완과 균형을 맞출 수 있음을 시사하고 있다.

4. 결론 및 향후 연구

본 연구 결과, 비식별화 기법 선택 시 데이터의 특성과 활용 목적을 고려해야 함을 확인했다.[6] 고위험 개인정보에는 마스킹이 적합하나, 데이터 유용성이 중요한 경우 가우시안 블러나 모자이크를 적절한 강도로 적용할 것을 권장한다. 이 과정에서 AI의 재식별 시도에 대한 저항력을 중점적으로 고려해야 한다.[7] 특히, 본 연구에서 최초로 시행한 AI 대 인간의 비식별화 인지 실험을 통해 두 주체 간 인지 차이를 정량적으로 분석하고, 객관적인 비식별화 기준점을 도출한 점에 큰 의의가 있다.

향후 연구에서는 다양한 비정형 데이터 유형에 대한 실험과 함께, 하이브리드 비식별화 기법 개발 및 AI 재식별 방지 기술 연구를 진행할 예정이다. 또한, 본 연구에서 도출된 정량화된 기준을 토대로 AI와 인간의 인지 차이를 고려한 맞춤형 비식별화 전략을 수립하고자 한다. 이를 통해 본 연구는 개인정보 보호와 데이터 활용 사이의 균형을 찾는 기술 발전에 상당한 기여를 할 것으로 기대된다.

※ 본 논문은 과학기술정보통신부
대학디지털교육역량강화 사업의 지원을 통해 수행한
ICT멘토링 프로젝트 결과물입니다.

참고문헌

[1] Sweeney, L. (2002). k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05), 557-570.

[2] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[3] Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools, 25(11), 120-125.

[4] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).

[5] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.

[6] Mun-ho, Ju. (2023). [Personal Information Protection Column-5] Domestic and International Trends in Personal Information De-identification Systems. Security News. Retrieved September 11, 2024, from https://www.boannews.com/media/view.asp?idx=119998

[7] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. In 2017 IEEE Symposium on Security and Privacy (SP) (pp. 3-18). IEEE.