부산대학교 제로트러스트 클라우드 보안 연구센터

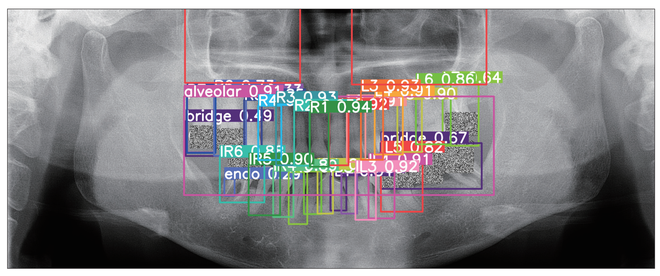
제로트러스트 클라우드 보안 연구센터 ZeroTrust Cloud Security Research Center

# Differential Privacy(DP) Patch for Privacy Preserving Deep Learning

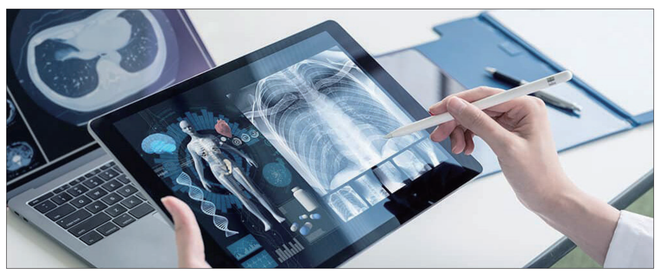
## 민감한 정보를 중심으로 한 ROI 기반 프라이버시 보존 이미지 생성 방법 및 장치

**적용 분야·제품**

AI 보안 | 프라이버시 | 빅데이터 분석 | 이미지 분석



의료 이미지 데이터 분석



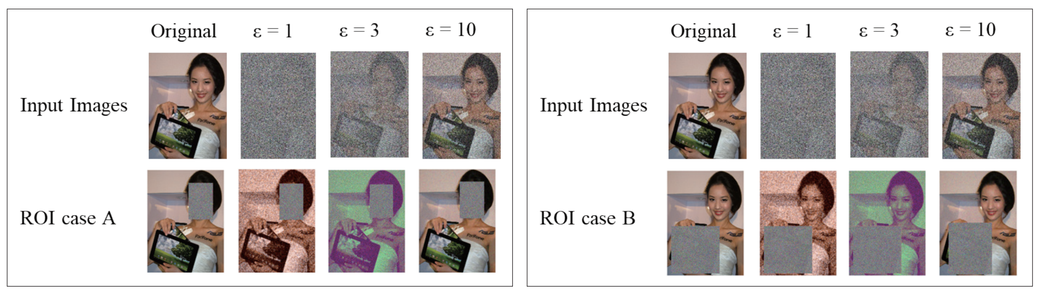
헬스케어 데이터 분석


핀테크 데이터 분석

**연구 목적**

- 기존 방식: 이미지 전체가 민감 영역으로 간주됨으로써 민감하지 않은 영역에도 노이즈가 추가되어 데이터의 활용 영역이 제한될 수 있음
- 제안 하는 방식: Differential Private 데이터(정형, 비정형) 생성 시 Utility 및 Quality 저하 문제가 발생하며 이를 해결하기 위한 Differential Privacy(DP) Patch 방안을 제안함

→ 전체 이미지의 왜곡 없이 특정의 민감한 객체만을 (ROI 기반) 정밀하게 숨기는 프라이버시 보존 이미지를 효율적으로 생성




<다양한 민감한 정보(ROI, Region of Interest)으로부터 생성된 프라이버시 보존 이미지 예시: 좌-얼굴, 우-태블릿>

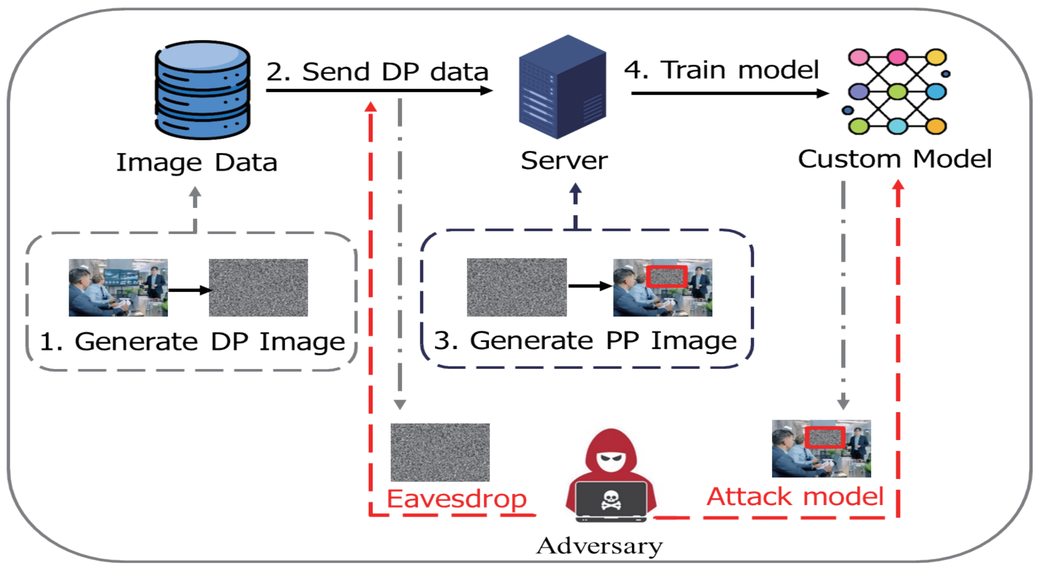
**연구 내용**

① 안전한 전송을 위한 Differential Private 이미지 생성
- 가우시안 메커니즘을 통해 노이즈 추가

② 네트워크를 통한 Differential Private 이미지 전송

③ 안전한 학습을 위한 Differential Private 이미지 기반 프라이버시 보존 이미지 생성
- 각 ε 값에 따라 노이즈 분포에 대해 학습된 디노이징 모델을 사용하여 디노이징 작업 진행
- 민감한 객체 목록을 기반으로 ROI 식별
- ROI에 Laplace Differential Private 패치 노이즈 추가

④ 맞춤형 딥러닝 모델 훈련




< 제안된 DP Patch의 워크플로우 및 프라이버시 노출 방지 사례>

**기술 경쟁력**

**기술 차별성**

기존기술

- 차분 프라이버시는 기능이 통합된 생성모델(GAN, VAE, 확사 모델 등)을 활용해 새로운 이미지를 생성
- 이미지 픽셀에 노이즈를 추가하는 이미지 섭동 기법 활용

**기술적 한계**

- 섭동 또는 생성된 이미지를 기반으로 하는 컴퓨터 비전 작업의 성능을 저하 시킬 수 있음
- 전체 이미지가 민감한 객체로 판단되는 경우, 이미지의 모든 픽셀에 대해 섭동이 이루어짐

대상기술

- 데이터 소유자로 하여금 보호하고자 하는 구체적이고 중요한 민감 정보들을 지정할 수 있음
- DP 노이즈가 넓은 영역으로 포함되더라도, 이미지를 정확하게 분류 가능

**기술적 우위**

- 민감한 객체들을 학습 및 예측 과정에서 제외
- 이미지의 민감한 객체를 재구성하는 것을 목표로 한 모델의 Model Inversion 공격에 대한 내성 강화

**문의처 |** 제로트러스트 클라우드 보안 연구센터 E. csrc@pusan.ac.kr T. 051. 510. 7212