1차 회의록

2025/09/12

프로젝트 목표: 딥페이크 방지 필터 웹페이지 제작

주요 이용자: SNS에 사진을 업로드 하는 일반 사용자들

웹사이트 서비스 기능

1. 사진을 업로드
2. 사진에 사람이 인식하기 어려운 노이즈를 추가
3. 노이즈가 추가된 사진을 반환
4. 필터가 씌인 사진을 딥페이크 앱/웹에 넣었을 경우 딥페이크 성능 하락

(나중에 필터 강도를 약함, 중간, 강함 3가지로 사용자가 선택하는 기능을 넣어도 좋을 듯)

고려사항

1. 필터 기능이 유효하되, SNS 업로드 하는 것에 불편함이 없어야 함
2. 사진의 용량 부담을 최소화 해야함
3. 어느정도의 사진 품질을 유지해야 함

필터 성공 목표: 딥페이크 생성 앱이나 웹 이용자 top 10을 대상으로 성공률 80%이상  
-성공률은 사진의 품질에 따라 변경 가능

-성공의 기준은 아직 명확하게 정하지 않음(이것도 기준을 찾아봐야함)

사용할 기술

-FGSM: 이미지 적대적 예제를 생성하는 기술 중 하나, 노이즈를 추가하기 위해 사용됨

다음 회의까지 해야할 것

1. FGSM에 대한 이해(공부)
2. FGSM을 이용한 필터가 여러 딥페이크 모델에도 공격이 가능한지
3. 사진 용량을 줄이기 위한 방식을 찾기(얼굴 인식, 경계선 인식 등)