캡스톤 디자인 I 최종결과 보고서

프로젝트 제목(국문): 안티포렌식에 강인한 딥페이크 탐지 모델 개발
프로젝트 제목(영문): Development of robust deepfake detector against anti-forensics

프로젝트 팀(원): 학번: 20191769 이름: 김지수
프로젝트 팀(원): 학번: 20191767 이름: 김민지
프로젝트 팀(원): 학번: 20191730 이름: 민지민

1. 중간보고서의 검토결과 심사위원의 '수정 및 개선 의견'과 그러한 검토의견을 반영하여 개선한 부분을 명시하시오.

없음

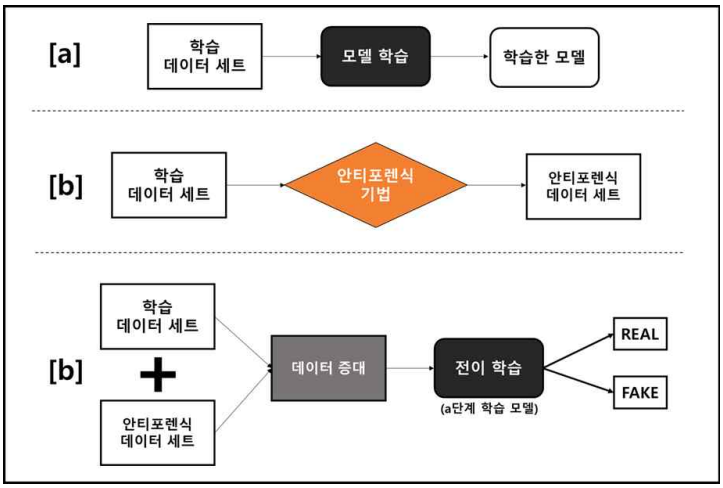
2. 기능, 성능 및 품질 요구사항을 충족하기 위해 본 개발 프로젝트에서 적용한 주요 알고리즘, 설계방법 등을 기술하시오.

이미지 편집 도구를 이용한 안티 포렌식 기법을 원본 데이터셋에 적용하여 안티 포렌식 데이터셋을 생성한다. 생성한 안티 포렌식 데이터셋을 기존 학습 데이터셋에 추가하여 적대적 학습을 수행함으로써 안티포렌식에 강인한 딥페이크 탐지 모델을 개발한다.

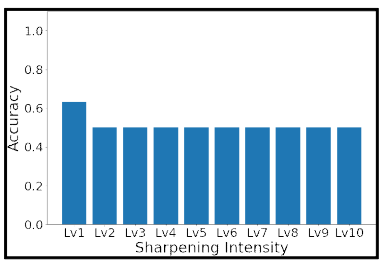
우선, 원본 데이터셋만을 사용해 딥페이크 탐지기를 학습한 모델을 도출한다. 각 안티포렌식 기법의 파라미터 수치를 조절하여 10단계의 강도로 적용해 안티 포렌식 데이터셋을 생성한다. 도출된 모델의 학습 데이터셋으로 원본 데이터셋과 생성한 안티 포렌식 데이터셋을 함께 포함하는 데이터 증대를 적용한 적대적 학습을 통하여 안티 포렌식에 강인한 모델을 개발한다.

3. 요구사항 정의서에 명세된 기능 및 품질 요구사항에 대하여 최종 완료된 결과를 기술하시오.

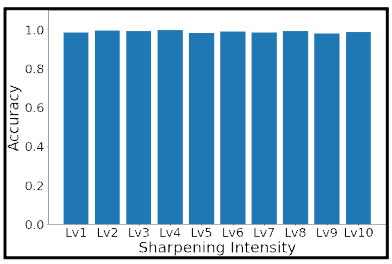
전체 시스템 구성도

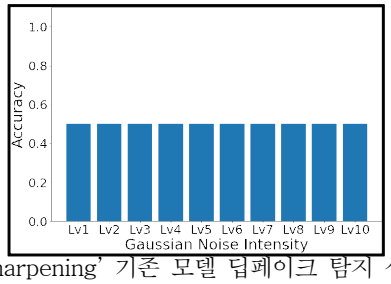


<'Gaussian Noise' 기존 모델 딥페이크 탐지 성능>

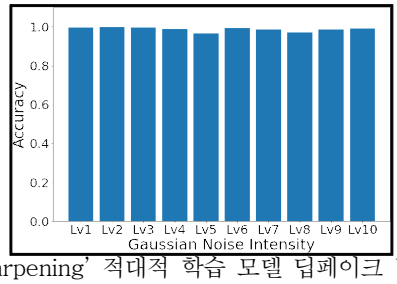


<'Gaussian Noise' 적대적 학습 모델 딥페이크 탐지 성능 >

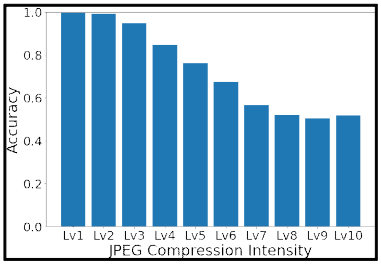




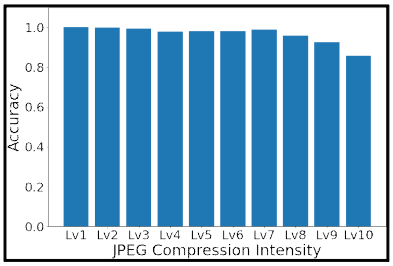
<'Sharpening' 기존 모델 딥페이크 탐지 성능>



<'Sharpening' 적대적 학습 모델 딥페이크 탐지 성능>



<'JPEG Compression' 기존 모델 딥페이크 탐지 성능>



<'JPEG Compression' 적대적 학습 모델 딥페이크 탐지 성능>

간단한 이미지 변형에도 취약했던 기존 딥페이크 탐지 모델에 적대적 학습 기법을 적용함으로써 높은 탐지 강인성을 획득하였다.

4. 우수성 입증 자료

2022 정보처리학회(ASK) 춘계 학술대회 '학부생 논문경진대회'에서 동상을 수상하였다.

• ASK 2022(춘계) 학부생논문경진대회 수상자 (31명)

포상명	수상자	소 속	공동저자
대 상 (1)	김동원	성균관대학교	김동원, 신유정
금 상 (2)	마상균	영남대학교	마상균, 박제원, 서정석
	강민주	숙명여자대학교	강민주, 전우진, 윤용익
은 상 (5)	오은비	서울여자대학교	강제희, 오은비, 이승연, 이원경, 김상욱
	최욱철	성균관대학교	최욱철, 구자환, 김동오
	우성미	중앙대학교	최경희, 우성미, 이대현
	한제상	숭실대학교	한제상, 장준서, 권재연, 허원제, 김영종,
	송성호	경기대학교	송성호, 김인철
동 상 (8)	장준모	한국외국어대학교	장준모
	이계혁	홍익대학교	이계혁, 황민제, 구영민, 변동섭, 유동엽
	양수빈	상명대학교	양수빈, 김민태, 권수빈, 우나현, 김학제, 정태경, 이성주
	민지민	한양대학교	민지민, 김지수, 김민지, 장현덕
	전기범	숭실대학교	전기범, 이유진, 안민하, 김동규, 김영종
	정승균	대구가톨릭대학교	정승균, 김규동, 김별찬
	박정현	호서대학교	박정현, 송민석, 백찬영, 홍우섭, 김민정, 문남미
	문지원	대구가톨릭대학교	문지원, 김장산, 김영은, 강다은, 이제민, 황정민, 이채원, 김동주
장려상 (15)	한성민	국립한경대학교	한성민, 박병숙, 김상훈
	이연아	홍익대학교	이연아, 정유진, 하지혜, 이수연, 유동엽
	Edward Dwijayanto Cahyadi	세명대학교	Edward Dwijayanto Cahyadi, 송미화
	정해린	동덕여자대학교	정해린, 김도영, 정원경, 김성경, 김현희
	박선호	숭실대학교	박선호, 박근형, 김태현, 김유림, 김영종
	박병수	숭실대학교	박병수, 박진혁, 임성현, 유준형, 김영종
	장환근	숭실대학교	장환근, 박성철, 나상우, 김 민, 이영제, 김영종
	김나연	동덕여자대학교	김나연, 김도영, 김미래, 정지영, 김현희
	신준석	서원대학교	신준석, 이덕규
	채수지	광운대학교	권나은, 이규민, 이지윤, 채수지, 박규동, 이상민
	박규환	서원대학교	박규환, 이덕규
	이창현	한양대학교 ERICA	이창현, 이진현, 차정원, 서승현
	박영서	백석대학교	박영서, 강 혁, 이근호
	이홍환	세명대학교	이홍환, 곽희홍, 박기수, 송미화
	차건화	숭실대학교	차건화, 김원승, 한성국, 김영종

4. 구현하지 못한 기능 요구사항이 있다면 그 이유와 해결방안을 기술하시오,
(작성요령: 전부 구현한 경우는 “이유”란에 “해당사항 없음”이라고 기재하고, 만약 요구사항대비 구현하지 못한 기능이 있다면 “이유”란에 그 사유를 기재함)

최초 요구사항	구현 여부(미구현, 수정, 삭제 등)	이유(일정부족, 프로젝트 관리미비, 팀원변동, 기술적 문제 등)
		해당사항 없음

5. 요구사항을 충족시키지 못한 성능, 품질 요구사항이 있다면 그 이유와 해결방안을 기술하시오.
(작성요령: 요구사항을 충족한 경우는 “이유”란에 “해당사항 없음”이라고 기재하고, 만약 요구사항대비 구현하지 못한 기능이 있다면 “이유”란에 그 사유를 기재함)

분류(성능, 속도 등) 및 최초 요구사항	충족 여부(현재 측정결과 제시)	이유(일정부족, 프로젝트 관리미비, 팀원변동, 기술적 문제 등)
		해당사항 없음

6. 최종 완성된 프로젝트 결과물(소프트웨어, 하드웨어 등)을 설치하여 사용하기 위한 사용자 매뉴얼을 작성하시오.
(작성요령: 여기에서 작성하는 사용자 매뉴얼은 개발한 시스템(환경)을 설치하여 사용할 수 있을 정도로 상세히 기술합니다)
- 사용자 매뉴얼과 실행파일

저희가 만든 모델을 다운받아서 실행하고자 하는 코드에 load하면 바로 사용할 수 있다.

7. 캡스톤디자인 결과의 활용방안
- 캡스톤 디자인을 통하여 완성된 프로젝트가 미치는 사회적/기술적/경제적 파급효과, 기대효과 등을 자유롭게 기술함

딥페이크 기술은 ‘개인 정보 침해’, ‘사기’ 등 다양한 범죄에 이용되고 있다. 저희가 만든 모델을 이용하면, 직접적으로 범죄가 발생하지 않도록 할 수는 없지만 저희가 만든 모델을 이용하면 피해자가 피해 본 사실을 밝힐 수 있다. 즉, 딥페이크로 합성된 이미지인지 아닌지 판별해냄으로써 관련 범죄 발생률을 조금이나마 줄일 수 있다.

현재 사용되고 있는 딥페이크 탐지기가 초보자도 쉽게 할 수 있는 이미지 편집기법 공격으로 무력화되고 있는데, 저희가 만든 모델은 이 점에 초점을 두어 안티포렌식으로 조작된 이미지나 영상을 좀 더 민감하게 판별함으로써 딥페이크 기술이 악용되지 않도록 도울 수 있다.