캡스톤디자인 II 중간보고서(표지)

프로젝트명 : 다양한 유형의 안티포렌식에 강인한 탐지 모델 개발

캡스톤 디자인II, 중간보고서

Version 1.0

개발 팀원 명(김지수):
민지민
김민지
대표 연락처:010-7132-6024
e-mail: 20191769@edu.hanbat.ac.kr

캡스톤 디자인 II 중간보고서 내용

1. 요구사항 정의서에 명세된 기능에 대하여 현재까지 진척된 결과 및 그 내용을 기술하시오.

데이터셋을 수집하여 안티포렌식 기법(median blur, gaussian noise, sharpening, JPEG compression, googlenet universal perturbation, poisson noise, impulse noise, unsharp mask filter, resnet universal perturbation)을 원본 데이터셋에 적용한 안티포렌식 데이터셋을 생성하였다. 발생할 수 있는 다양한 안티포렌식 공격을 우회하기 위해 각각의 공격들을 총 10단계의 강도로 적용하였다. 원본 데이터셋만을 사용해 딥페이크 탐지기를 학습한 Xception 기반 모델을 도출하였다. 여기서 도출한 모델로 강도별 안티포렌식 데이터셋에 대한 탐지 정확도를 측청해 보았고, 낮은 탐지 정확도를 확인하였다. 안티포렌식 기법(median blur, gaussian noise, sharpening, JPEG compression, googlenet universal perturbation)이 원본 데이터셋에 랜덤한 순서와 강도로 적용된 안티포렌식 데이터셋을 생성하였다. 기존 딥페이크 탐지 모델의 학습 데이터셋으로 원본 데이터셋과 다양한 유형의 공격들의 강도 및 공격순서가 랜덤하게 적용된 안티포렌식 데이터셋을 사용하는 적대적 학습으로 최종 모델을 도출해냈다. 여기서 도출한 모델로 적대적 학습에 반영한 공격과 반영하지 않은 공격에 대한 탐지 정확도를 측정해 보았고, 학습에 반영한 공격뿐만 아니라 학습에 반영하지 않은 다양한 종류의 공격들에 대해서도 높은 판별 정확도를 확인하였다.

2. 프로젝트 수행을 위해 적용된 추진전략, 수행 방법의 결과를 작성하고, 만일 적용과정에서 문제점이 도출되었다면 그 문제를 분석하고 해결방안을 기술하시오.

- 추진 전략 : 매주 담당 교수님에게 프로젝트 진행 사항을 발표하며, 교수님의 피드백과 팀원들과 토의함으로써 수월하게 진행함
- 수행 방법 : 각자 장비 환경에 맞추어 효율적으로 업무를 분담하고, 또한 '구글의 공유 플랫폼'을 적극적으로 사용하여 효율성을 높일 것
- 추진 전략 및 수행 방법의 결과 : 업무 효율이 높았으며, 역할 분담을 함으로써 모든 팀원의 참여도가 높았음
- 팀원의 책임 및 역할 수행에 대한 결과 : 업무를 균등하게 분담함으로써 원활하게 프로젝트 수행하여 큰 문제는 없었음
- 프로젝트 일정계획에 맞추지 못한 경우의 문제점과 해결 방안 :
- 요구사항 변경관리 : 변경 사항 없음

[별첨1]

프로젝트명 : 다양한 유형의 안티포렌식에 강인한 탐지 모델 개발

소프트웨어 요구사항 정의서

Version 1.0

개발 팀원 명(김지수):
민지민
김민지
대표 연락처: 010-7132-6024
e-mail: 20191769@edu.hanbat.ac.kr

목차

요구사항 정의서에 사용되는 양식 설명

요구사항 고유번호(ID): 제안요청서에 정의된 요구사항에 대해 계약, 사업수행, 사업완료 및 검수까지 변경, 삭제, 수정 여부에 대한 추적관리를 위해 고유의 번호를 부여하도록 한다.

요구사항 구분 및 ID부여 규칙

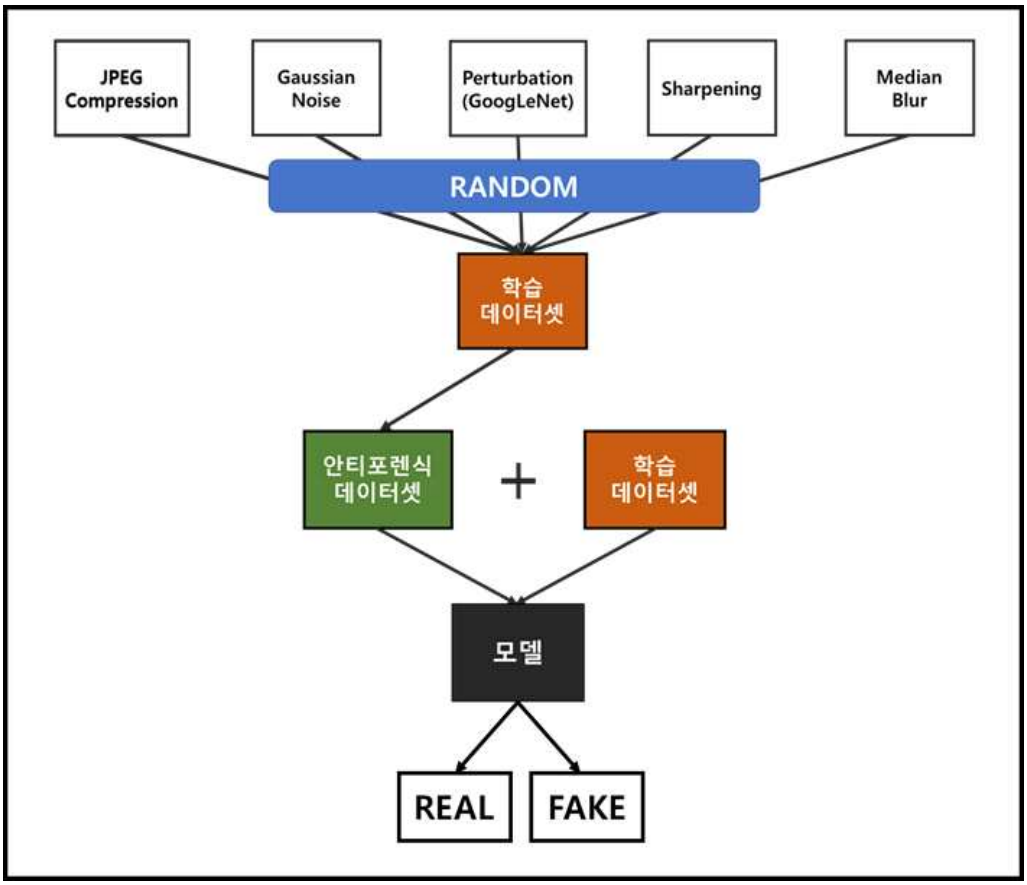
요구사항 구분		ID 부여 규칙
시스템 장비 구성 요구사항	Equipment Composition Requirement	ECR-000
기능 요구사항	System Function Requirement	SFR-000
성능 요구사항	Performance Requirement	PER-000
인터페이스 요구사항	System Interface Requirement	SIR-000
데이터 요구사항	Data Requirement	DAR-000
테스트 요구사항	Test Requirement	TER-000
보안 요구사항	Security Requirement	SER-000
품질요구사항	Quality Requirement	QUR-000
제약사항	Constraint Requirement	COR-000
프로젝트 관리 요구사항	Project Mgmt. Requirement	PMR-000
프로젝트 지원 요구사항	Project Support Requirement	PSR-000

요구사항 세부내용 작성표 양식 및 항목설명

요구사항 고유번호		(설명) 요구사항 추적관리를 위해 독립적인 고유번호(ID) 부여
요구사항 명칭		(설명) 요구사항 명칭을 작성함
요구사항 분류		(설명) 요구사항 분류기준에 따른 분류를 기입
요구사항 상세 설명	**정의**	(설명) 요구사항 정의
	세부 내용	(설명) 요구사항 구체적인 세부 내용을 설명
산출정보		(설명) 해당기능을 통해 산출되는 결과물 혹은 정보를 표기
관련 요구사항		(설명) 정의된 요구사항과 관련된 요구사항에 대해 기술
요구사항 출처		(설명) 기능 도출내용에 대한 출처(source) 표기

1. 시스템 개요

- 모델 학습 절차



2. 시스템 장비 구성요구사항

요구사항 고유번호		ECR-001		
요구사항 명칭		장비 요구사항		
요구사항 분류		시스템 장비구성 요구사항	응락수준	필 수
요구사항 상세 설명	정의	모델 학습 장비		
	세부 내용	- 장비 품목 : GPU (NVIDIA GeForce RTX 3060) - 장비 수량 : 1개 - 장비 기능 : 모델의 연산 속도를 높인다. - 장비 성능 및 특징 : 메모리 12GB - 시간 제약사항 : 해당 사항 없음 - 자원 제약사항 : 다양한 실험의 필요성으로 로컬 GPU만을 사용하지 않고, Google Colab GPU를 사용하였다. - 장애 처리 : 해당 사항 없음		

3. 기능 요구사항

요구사항 고유번호		SFR-001
요구사항 명칭		AI 모델 개발
요구사항 분류		기능
요구사항 상세 설명	정의	딥페이크 탐지 AI 모델 개발
	세부 내용	• Xception 기반의 AI 모델 개발 • 여러 가지 안티포렌식 기법을 적용한 데이터 증대(Data Augmentation)을 사용하여 발생 가능한 공격상황에 강인하도록 할 것 • 여러 가지의 공격의 순서와 강도를 달리하여 무작위로 적용한 안티포렌식 데이터셋을 생성할 것 • 위 단계에서 생성한 데이터셋을 학습함으로써 학습하지 않은 공격에 대해서도 강인하도록 할 것 • 다양한 공격에 강인한지 보기 위하여 학습에 반영한 공격과 학습에 반영하지 않은 공격 모두에 대한 테스트를 진행할 것

4. 성능 요구사항

요구사항 고유번호		PER-001		
요구사항 명칭		처리 속도 및 시간		
요구사항 분류		성능 요구사항	응락수준	필 수
요구사항 상세 설명	정의	처리 속도 및 시간		
	세부 내용	모델이 딥페이크 detection 하는 시간을 의미함 사진이 real인지 fake인지 판별하는 시간을 의미함		

5. 인터페이스 요구사항

해당사항 없음

6. 데이터 요구사항

해당 사항 없음

7. 테스트 요구사항

요구사항 고유번호		TER-001		
요구사항 명칭		성능 테스트		
요구사항 분류		테스트 요구사항	응락수준	필 수
요구사항 상세 설명	정의	성능 테스트		
	세부 내용	- 구축된 모델이 제대로 탐지하는 지를 테스트하고 점검하기 위한 평가 기준으로 loss와 acc로 평가함 - real, fake 이미지를 제대로 판단하는지에 대한 테스트		

요구사항 고유번호		TER-002		
요구사항 명칭		성능 테스트		
요구사항 분류		테스트 요구사항	응락수준	필 수
요구사항 상세 설명	정의	성능 테스트		
	세부 내용	- 원본 이미지와 안티 포렌식 이미지 사이의 왜곡 정도를 의미함 - 데이터 셋에 노이즈를 추가할 때, psnr 28이상인 데이터 셋만 다룸		

8. 보안 요구사항

요구사항 고유번호	SER-001
요구사항 명칭	보안지침 준수
요구사항 분류	보안
요구사항 세부내용	• 사용한 데이터 세트를 배포한 AIHUB의 보안 및 저작권 관련 데이터 이용정책에 따라 개발이 수행되어야 함

9. 품질 요구사항

요구사항 고유번호		QUR - 001
요구사항 명칭		구축데이터 품질관리
요구사항 분류		품질
요구사항 상세 설명	정의	품질관리(기술 관점)
	세부 내용	- 학습데이터는 민간 개방 시 사용에 제약이 없도록 개인정보 사용에 따른 동의를 확보해야 한다. - 데이터 수집 시 데이터 이용을 위해 사전허가가 필요한 경우에는 반드시 해당 기관의 사전허가를 득해야 한다. - 원천데이터는 중복이 없어야 한다. - 정제단계에서 데이터 중복을 확인해서 제외해야 한다. - 데이터는 학습에 유용한 수량이어야 한다. - 데이터의 카테고리 별 인스턴스 수량의 균일성과 적정한 비율을 확보하여야 한다. - 카테고리 라벨은 반드시 분류체계에 따라 명확하게 정의된 라벨을 사용해야 하고 카테고리 간 모호성이 없어야 한다. - 안티 포렌식 기법을 적용한 이미지의 psnr이 28이상이어야 한다.
산출정보		높은 탐지 정확도

10. 제약 사항

요구사항 고유번호	COR-001
요구사항 명칭	시스템 개발과 설계 및 구현 제약사항
요구사항 분류	제약사항
요구사항 세부내용	• 현재 보유하여 활용 가능한 H/W, S/W를 최대한 활용함 • 대부분의 인공지능 모델 개발에 사용되는 Python(언어), PyTorch(프레임워크)를 사용함 • 딥페이크 탐지에 경험적으로 좋은 성능을 보이는 네트워크를 기반으로하여 모델 설계가 이루어져야 함

요구사항 고유번호	COR-002
요구사항 명칭	데이터 제약사항
요구사항 분류	제약사항
요구사항 세부내용	• 사람의 얼굴 영상을 사용하는 프로젝트이기 때문에 모델 학습에 사용하는 영상의 초상권 관련 이슈가 발생하지 않도록 함 • 사용한 데이터 세트를 배포한 AIHUB의 보안 및 저작권 관련 데이터 이용정책에 따라 개발이 수행되어야 함

11. 프로젝트 관리 요구사항

요구사항 고유번호		PMR-001
요구사항 명칭		프로젝트 관리
요구사항 분류		프로젝트 관리
요구사항 상세 설명	세부 내용	- 세부 작업 분할구조 : 1. 분석 안티 포렌식 기법 조사 2. 데이터수집 및 생성 데이터 수집, 안티 포렌식 데이터셋 생성 3. 시스템 설계 원본 데이터셋만을 이용한 딥페이크 탐지 모델 도출, 적대적 학습 기반의 딥페이크 탐지 모델 도출 4. 실험 원본 데이터셋만을 이용한 탐지 모델을 이용하여 안티 포렌식 데이터셋에 대한 탐지 정확도 측정, 적대적 학습 기반의 딥페이크 탐지 모델을 이용하여 안티 포렌식 데이터셋에 대한 탐지 정확도 측정, 학습에 반영하지 않은 안티 포렌식 기법에 대한 정확도 측정 5. GUI 구현 딥페이크 탐지결과를 보여줄 GUI 구현

		- 프로젝트 수행조직에 대한 구성, 역할 김지수 : 실험 데이터셋 생성 김민지 : 학습 모델 생성 민지민 : 학습 모델로 성능 확인