캡스톤디자인I 중간보고서

<table>
<tr><td rowspan="2">제 목</td><td colspan="2">국문</td><td colspan="3">안티포렌식에 강인한 딥페이크 탐지 모델 개발</td></tr>
<tr><td colspan="2">영문</td><td colspan="3">Development of robust deepfake detector against anti-forensics</td></tr>
<tr><td rowspan="3">진 행
상 황</td><td>중요마일스톤</td><td colspan="4"><주요 기능 : 모델 학습>
1. 딥페이크 데이터셋 구축
2. 데이터셋 라벨링
3. Xception기반 탐지 모델 생성
4. 안티포렌식 데이터셋 구축
5. 안티포렌식 데이터셋에 대한 탐지 성능 측정
6. 적대적 학습으로 안티포렌식에 강인한 탐지 모델 생성</td></tr>
<tr><td>진행상황</td><td colspan="4">딥페이크 관련 논문조사 완료, 안티포렌식 기법 조사 완료, 데이터 수집 완료, 안티포렌식 데이터셋 생성 완료, 원본 데이터셋만을 이용한 딥페이크 탐지 모델 도출 완료, 원본 데이터셋만을 이용한 탐지 모델을 이용하여 안티포렌식 데이터셋에 대한 탐지 정확도 측정 완료, 적대적 학습 기반의 딥페이크 탐지 모델 도출 완료, 적대적 학습 기반의 딥페이크 탐지 모델을 이용하여 안티포렌식 데이터셋에 대한 탐지 정확도 측정 완료</td></tr>
<tr><td colspan="2">산출물</td><td colspan="4">요구사항 정의서(별첨 1), 중간보고서(별첨 2)</td></tr>
<tr><td rowspan="4">팀
구성원</td><td>학년</td><td>학 번</td><td>이 름</td><td colspan="2">연락처(전화번호/이메일)</td></tr>
<tr><td>4</td><td>20191730</td><td>민지민</td><td colspan="2">20191730@edu.hanbat.ac.kr</td></tr>
<tr><td>4</td><td>20191769</td><td>김지수</td><td colspan="2">20191769@edu.hanbat.ac.kr</td></tr>
<tr><td>4</td><td>20191767</td><td>김민지</td><td colspan="2">20191767@edu.hanbat.ac.kr</td></tr>
<tr><td colspan="6">컴퓨터공학과의 프로젝트 관리규정에 따라 다음과 같이 요구사항 정의서와 중간보고서를 제출합니다

2022 년 4월 29 일

책임자 : 김지수 (인)
지도교수 : 장한얼 (인)</td></tr>
</table>

[별첨1]

프로젝트명 : 안티포렌식에 강인한 탐지 모델 개발

소프트웨어 요구사항 정의서

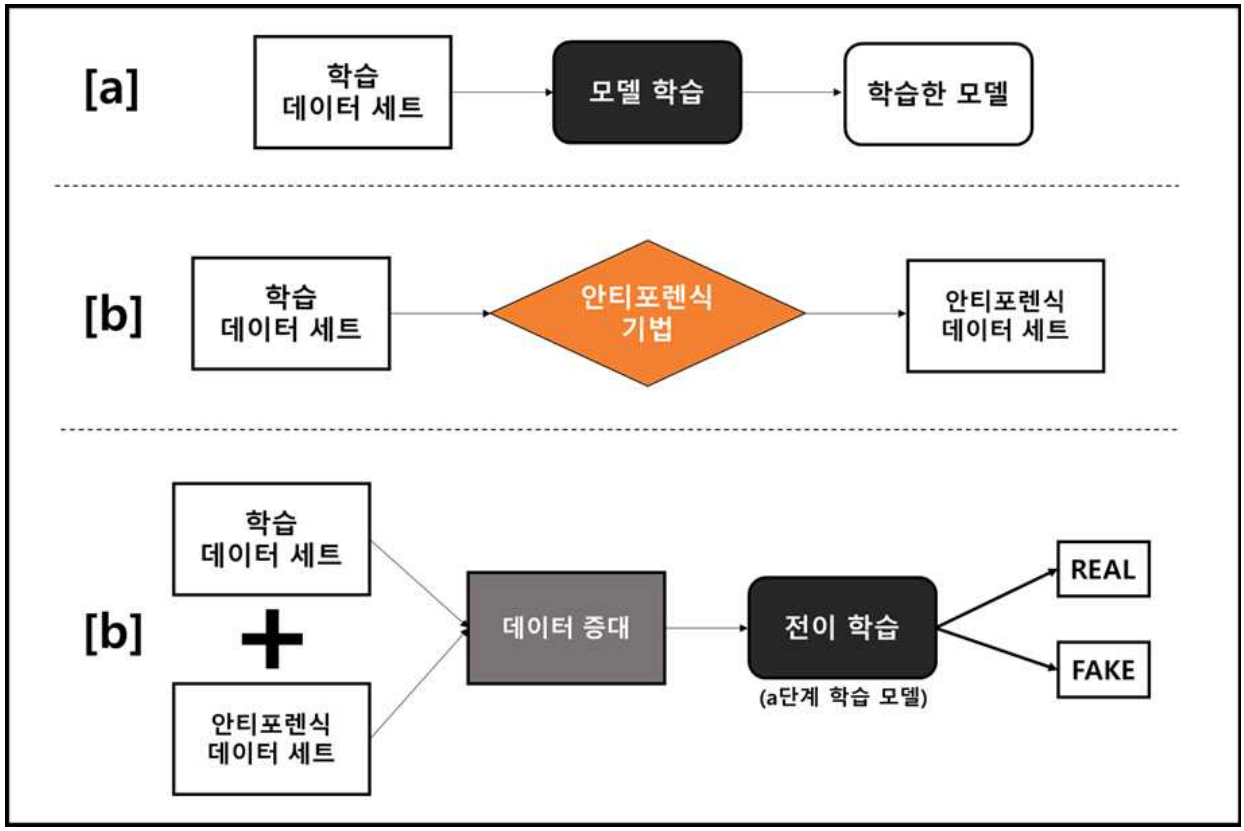
Version 1.0

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

목차

1. 시스템 개요

- 모델 학습 절차



2. 시스템 장비 구성 요구사항

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 • 적대적 학습 기법을 통해 학습하지 않은 공격에 대해서도 강인하도록 할 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 사용자가 요청한 사진을 판별하는 시간을 의미함		

5. 인터페이스 요구사항

해당 사항 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 detection 하는 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 30이상인 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 30이상이어야 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	• 사람의 얼굴 영상을 사용하는 프로젝트이기 때문에 모델 학습에

세부내용	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 딥페이크 관련 논문 조사, 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 - 프로젝트 수행조직에 대한 구성, 역할 김지수 : 논문 및 자료조사, 데이터셋 생성, 모델링 김민지 : 논문 및 자료조사, 데이터셋 생성, 데이터 분석 및 전처리 민지민 : 논문 및 자료조사, 데이터셋 생성, 모델실험

[별첨2]

중간보고서

1. 요구사항 정의서에 명세된 기능에 대하여 현재까지 분석, 설계, 구현(소스코드 작성) 및 테스팅한 내용을 기술하시오.

데이터셋을 수집하여 안티포렌식 기법(Gaussian Noise, Sharpening, JPEG Compression)을 원본 데이터셋에 적용하여 안티포렌식 데이터셋을 생성하였다. 발생할 수 있는 다양한 안티포렌식 공격을 우회하기 위해 'Gaussian Noise', 'Sharpening', 'JPEG Compression'을 총 10단계의 강도로 적용하였다. 원본 데이터셋만을 이용해 Xception 기반 딥페이크 탐지 모델을 도출하였다. 여기서 도출한 모델로 강도별 안티포렌식 데이터셋에 대한 탐지 정확도를 측정해 보았고, 낮은 탐지 정확도를 확인하였다. 기존 딥페이크 탐지 모델의 학습 데이터셋으로 원본 데이터셋과 생성한 안티포렌식 데이터셋을 함께 포함하는 데이터 증대를 적용한 적대적 학습을 통하여 안티포렌식에 강인한 모델을 도출했다. 여기서 도출한 모델로 안티포렌식 공격이 가해진 데이터셋에 대한 탐지 정확도를 측정해 보았고, 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 : 모델 실험에 관련해서 환경 설정 문제에 봉착해 일정계획에 차질이 생겼지만, 원래 계획에 맞추기 위해 교수님과 추가 미팅함으로써 문제해결 방안을 알게 됐고 모든 팀원이 실험을 다시 진행하여 다음 일정계획에 차질이 없도록 함
- 요구사항 변경관리 : 변경 사항 없음

캡스톤 디자인 I 중간보고서 채점표

평가도구	평 가 항 목	평 가 점 수				
		1	2	3	4	5
중간 보고서 및 실행 결과	1. 요구사항 정의서(기능, 성능, 인터페이스 등)가 구체적으로 작성되었는가?					
	2. 요구분석, 설계 산출물(모델, 프로토타입 등)의 내용이 충실한가?					
	3. 설계 및 구현 문제를 위해 적용한 이론, 문제해결 방법이 제시되었으며 그 적용이 적합한가?					
	4. 구현된 소프트웨어(또는 이와 동등한 하드웨어 시스템)가 버그 없이 실행되었는가?					
	5. 구현된 소프트웨어(또는 이와 동등한 하드웨어 시스템)의 성능 요구사항은 충족되었는가?					
도구활용	6. 설계 및 구현을 위해 도구가 적절히 활용되었는가?					
	7. 도구의 활용수준(능숙도)은 프로젝트 수행에 적합한가?					
팀원의 업무 및 역할	8. 팀원의 업무분담에 따른 역할 및 협력이 충실히 이루어졌는가? (평가자에 의한 질의)					
	9. 프로젝트 중간 진척상황에 대해 팀원이 충분히 인지하고 있는가?(평가자에 의한 질의)					
합계						

***검토 의견(최종완료 때까지 보완해야할 점에 대해 작성 요망)**

심사위원(소속): **(이름)** **(인)**