테크니컬 리포트

2020 년 1 차
인공지능
학습용
데이터 구축

기타 영역

딥페이크 변조영상

개요: 딥페이크 변조영상 데이터셋이란?

딥페이크 변조영상 데이터셋이란, 다양한 변조영상 생성 알고리즘을 통해 생성된 변조영상(딥페이크)을 탐지·검출하는 AI 기술 개발을 위한 학습용 변조영상 데이터이다.

인공지능 컴퓨터 비전(Computer Vision)에서 최근 많은 주목을 받고 있는 딥페이크 탐지기술을 개발하는 용도로 활용할 수 있는 학습 데이터셋으로 (주) 머니브레인에서 구축했으며, 한국인 중심의 얼굴에 집중하여 6 만건 이상의 원시영상, 15 만건 이상의 변조영상으로 구성되어 있다. 수집된 데이터의 일부를 활용하여 해커톤을 진행하고, 이를 통한 우수한 탐지 모델 확보까지가 프로젝트의 목표이다.

딥페이크와 딥페이크 탐지의 예시에 대해서는 아래를 참고할 수 있다.

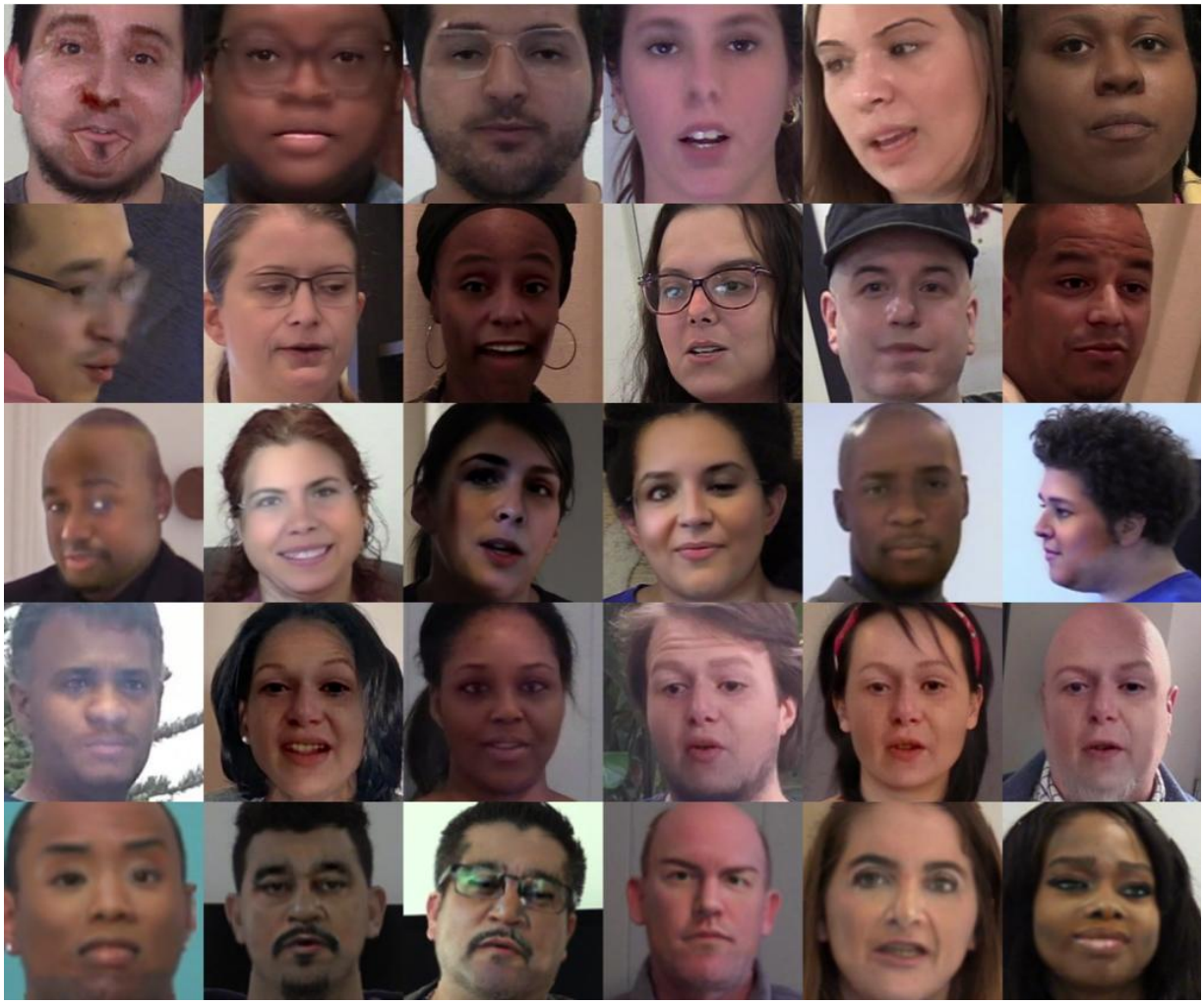

그림 1. 딥페이크의 예시

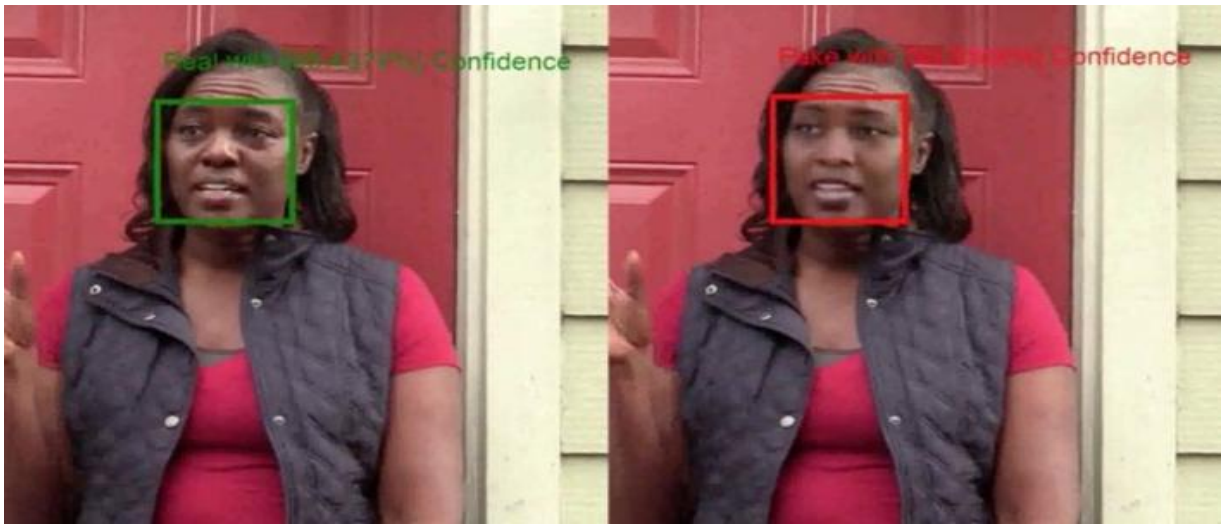

그림 2. 딥페이크의 탐지 예시

데이터셋의 구성

본 데이터셋은 400 명 이상의 참여자로부터 크라우드소싱 방식을 통해서 얻은 각 90 초 이상의 (초당 30 프레임) FHD (1920 X 1080) 영상 6 만 건 이상 (참여자 1 인당 150 건 이상), 그리고 이러한 원시 영상을 사용해 6 가지의 변조모델을 학습하여 생성한 각 15 초 이상의 FHD 영상 15 만 건 이상으로 구성된다.

동영상 기준 21 만 건(프레임 기준 약 2 억 3 천만 건)은 연구자가 일반적으로 연구를 진행하기에 충분한 양이며, 최신 유명 해외 연구 사례(예. FaceForensics++, DeepFakeChallengeDataset)와 비교해서도 사이즈와 품질, 데이터의 균등한 분포 면에서 비교우위를 갖는다.

데이터 종류	포함 내용	제공 방식
원시 영상 (크라우드소싱)	가변배경의 원시영상 52,500 건 이상 (350 명 이상 참여자의 인당 150 건 이상 영상)	mp4 포맷 동영상 파일 (fhd, fps 30)
원시 영상 (전문 스튜디오)	크로마키 배경 원시영상 7,500 건 이상 (50 명 참여자의 인당 150 건 이상 영상)	mp4 포맷 동영상 파일 (fhd, fps 30)
변조 영상	6 종의 변조모델을 사용하여 생성한 변조영상 150,000 건 이상	mp4 포맷 동영상 파일 (fhd, fps 30)
탐지 방해 영상	21,000 건 가량의 탐지방해 소음이	mp4 포맷 동영상 파일 (fhd,

	추가된 영상 (탐지방해소음 추가를 통한 데이터 증강)	fps 30)
메타데이터	상기 데이터의 영상단위 세부사항에 대한 메타데이터셋	csv 파일

데이터셋의 설계 기준과 분포

데이터셋을 설계할 때 가장 중요하게 고려했던 점은 분류별 분포와 품질, 그리고 다양성이다. 인물 다양성 확보를 위해 참여자 섭외시 성비를 균등하게 조정하고, 나이와 체형이 최대한 자연적 분포를 따르도록 하였다.

배경 다양성을 위해서 복수의 촬영 공간을 섭외하고 촬영 공간내 복수의 지점에서 촬영을 진행했으며, 촬영시에는 소품 등을 다르게 배치하고 뒷면에 전지를 붙이는 등의 기법을 통해 다양성을 모색하였다. 또한 전문 스튜디오에서는 차후 배경에 대한 합성까지 예상하여 크로마키를 배경으로 채택하였다.

참여자가 전문 방송인이 아닌 점을 고려하여, 자연스런 발화를 유도하고 그 내용의 다양성을 강화하기 위해서 촬영자가 정해진 대본을 읽는 스크립트 세션과 본인이 원하는 질문에 대답하는 시나리오 세션으로 분류를 나누었다. 또한 상기 대본과 질문은 감정 분류, 문장형태 등의 분류에 따라 내용을 구성했다. 스크립트의 경우 표준 국어 대사전에 등재된 모든 예문과 정의를 수집/정제 후 평서문, 의문문, 감탄문으로 분류했고, 군산대 감정사전에 기반해서 각 문장에 대한 감정을 기계태깅하여, 그 분류에 맞도록 문장 비율을 조정했다. 시나리오의 경우, 흥미로운 질문들을 크라우드소싱하여 복수의 어노테이터가 점검 후 420 개의 질문을 추려 정성적으로 감정 분류를 실시했다.

변조모델의 경우, 두 얼굴을 바꾸는 Face Swapping 방법과 오디오나 영상 신호를 기반으로 얼굴을 재건하는 Face Reenactment 방법 간의 균형을 이루면서 현 시대 가장 대중적이고 효과적인 모델들 중 라이센스 문제가 없는 모델들을 우선적으로 선택했다.

모델명	설명	수량
DeepFaceLab (이하 DFL)	가장 대중적인 모델이며, 평균적인 완성도가 가장 높은 Face Swapping 모델.	53,816 개 생성 (목표 수량 37,500 개)
DeepFakes/FaceSwap (이하 DFFS)	최초의 딥페이크 모델들 중 하나이자 결과물이 안정적인 Face Swapping 모델	52,209 개 생성 (목표 수량 37,500 개)
Face Swapping GAN	비교적 최근에 공개된 생성적 적대	53,816 개 생성

(이하 FSGAN)	신경망 기반 Face Swapping / Reenactment 모델	(목표 수량 37,500 개)
First Order Model (이하 FO)	영상 움직임 키포인트 기반 Face Reenactment 모델	72,298 개 생성 (목표 수량 37,500 개)
Audio-driven Talking Face Video Generation with Learning-based Personalized Head Pose (이하 Audio-Driven)	비교적 최근에 공개된 음성입력 및 3D 모델링을 통해 얼굴영상을 생성하는 Face Reenactment 모델 (음성을 사용하는 공통점 때문에 아래 모델과 목표수량을 공유하도록 함)	21,731 개 생성 (목표 수량 3,000 개)
Wav2Lip (이하 Audio-driven)	가장 최근에 공개된 음성입력 기반 Face Reenactment 모델	

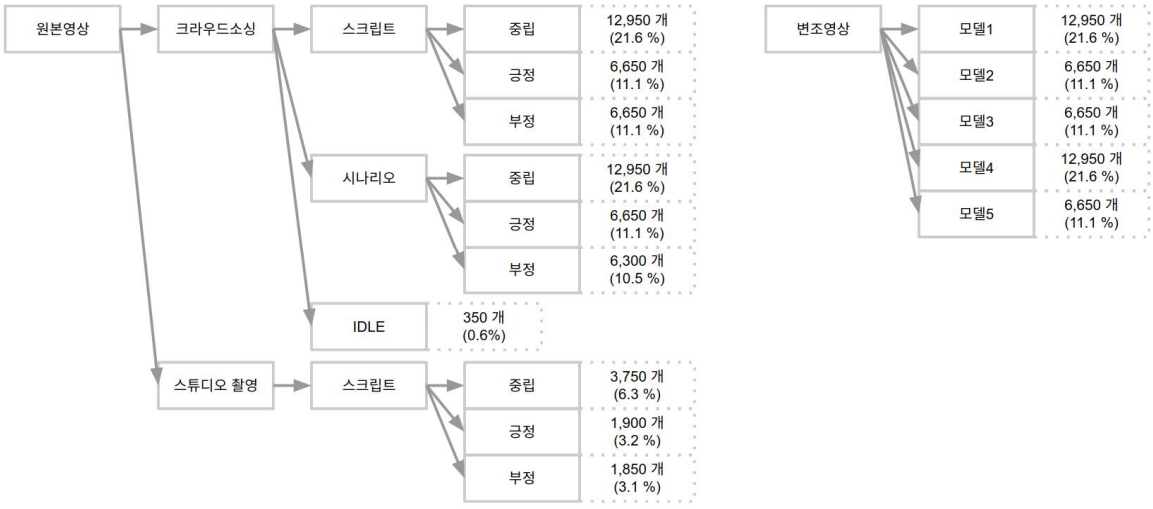

그림 3 데이터셋 구성 개요

데이터 구조

분류	원시영상 데이터셋	변조영상 데이터셋
내용	90 초이상의 영상	15 초 이상의 영상
수량	6 만	15 만
항목	**포함여부**	
영상 ID	Y	Y

타겟 UUID	Y	Y
촬영일자	Y	Y
촬영장소	Y	N
촬영기기	Y	N
스크립트파일	Y	N
시나리오번호	Y	N
원영상 ID	N	Y
소스 UUID	N	Y
변조 모델	N	Y

데이터 예시

원시영상 데이터셋 예시		변조영상 데이터셋 예시	
영상 ID	179032_027.mp4	영상 ID	179032_175277_2_0270.mp4
타겟 UUID	179032	타겟 UUID	179032
촬영일자	2020-10-17	촬영일자	2020-10-17
촬영장소	코지모임공간	원영상 ID	179032_027.mp4
촬영기기	아이폰 XR	소스 UUID	175277
스크립트파일	16901	변조 모델	dffs
시나리오번호			

데이터 구축 과정

데이터 구축은, 2020 년 6 월과 7 월 양월간 딥페이크 탐지데이터의 해외연구 사례를 참고 및 개선하여 데이터 구성의 설계를 완성했고, 이와 동시에 변조모델에 대한 선행연구 서베이를 통해 6 종의 최신 변조모델을 선정하였다.

2020 년 7 월부터 11 월까지 전문 스튜디오와 3 곳 이상의 촬영 장소를 섭외하여 400 명 촬영자에 대한 원시영상 촬영을 완료하였고, 검수가 이루어진 원시영상을 확보함과 동시에 학습을 진행하여 변조영상 생성을 12 월까지 마무리지었다.

문제소지가 있는 원본영상과 저품질의 변조영상을을 제거하는 필터링을 거쳐 원시영상 6 만 건 이상, 변조영상 15 만 건 이상을 기준으로 분야별 분류(성별, 시나리오/스크립트)와 변조모델별 목표 분포를 이루어지게 했다.

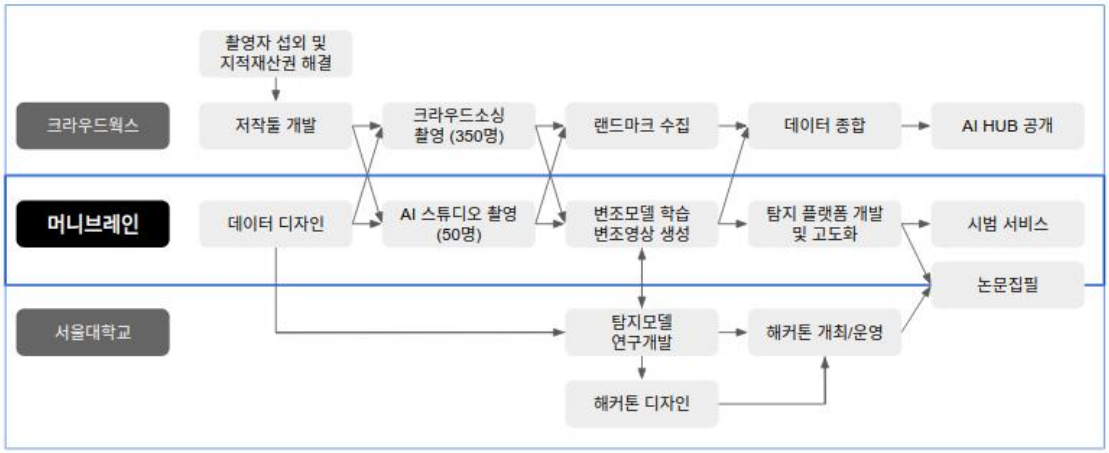

그림 4 데이터 구축 프로세스

검수와 품질 확보

원시영상의 과반수 이상을 고품질의 촬영 결과를 산출하기 어려운 크라우드소싱 방식으로 생성했기 때문에, 원시영상들과 이를 바탕으로 생성된 변조영상들의 검수과정은 해당 과제에 있어 매우 중요한 절차이다.

원시영상의 경우, 크라우드소싱 결과물들이 변조영상 생성 가이드라인에 맞는지 (얼굴의 위치 및 각도, 조명 등)을 크라우드웍스측 검수자들이 확인해서 통과된 결과물들만 머니브레인에 전달되었다.

변조영상의 경우, 한 영상당 2 명의 머니브레인측 검수자들이 영상의 품질을 확인하고 특이사항을 기록하였다. 이 과정에서 2 명 다 통과시킨 영상들만이 최종 변조영상 데이터셋에 등록되었다. 또한 정량 적 평가지표로 Structural Similarity Index Metrics 와 Average Keypoint Distance 를 사용해 원본 이미지와 만들어진 이미지 사이의 구조적 유사성을 측정했으며, 그 수치에 기반한 품질이 기존 가장 널리 연구에 사용된 해외 데이터인 FaceForensics++를

상회토록 하였다(AKD < .75, SSIM > .75 목표에서 각각 .31 과 .78 달성).

$$SSIM(r,g) = \frac{(2\mu_r\mu_g + c_1)(2\sigma_{rg} + c_2)}{(\mu_r^2 + \mu_g^2 + c_1)(\sigma_r^2 + \sigma_g^2 + c_2)}$$

수식 1

$$AKD = \frac{1}{F}\sum_{f}^{F}\sqrt{\sum_{p}^{P}(r_p^f - g_p^f)^2}$$

수식 2

	A	B	C	D	E	F
1	변조영상명	변조영상 위치	검수자 A	인물이 진짜라고	영상의 전반적인	특이사항
2	output_sohee_001_1	sohee > dfl	Jay	예	예	전반적으로 얼굴이 매우
3	output_sohee_001_2	sohee > dfl	Jay	예	예	
4	output_sohee_001_3	sohee > dfl	Jay	예	예	
5	output_sohee_001_4	sohee > dfl	Jay	예	예	
6	output_sohee_001_5	sohee > dfl	Jay	예	예	
7	output_sohee_001_6	sohee > dfl	Jay	예	예	
8	output_sohee_002_1	sohee > dfl	Jay	예	예	
9	output_sohee_002_2	sohee > dfl	Jay	예	예	
10	output_sohee_002_3	sohee > dfl	Jay	예	예	
11	output_sohee_002_4	sohee > dfl	Jay	예	예	
12	output_sohee_002_5	sohee > dfl	Jay	예	예	
13	output_sohee_002_6	sohee > dfl	Jay	예	예	
14	output_sohee_003_1	sohee > dfl	Jay	예	예	
15	output_sohee_003_2	sohee > dfl	Jay	예	예	
16	output_sohee_003_3	sohee > dfl	Jay	예	예	
17	output_sohee_003_4	sohee > dfl	Jay	예	예	
18	output_sohee_003_5	sohee > dfl	Jay	예	예	
19	output_sohee_003_6	sohee > dfl	Jay	예	예	
20	output_sohee_004_1	sohee > dfl	Jay	예	예	
21	output_sohee_004_2	sohee > dfl	Jay	예	예	
22	output_sohee_004_3	sohee > dfl	Jay	예	예	
23	output_sohee_004_4	sohee > dfl	Jay	예	예	
24	output_sohee_004_5	sohee > dfl	Jay	예	예	
25	output_sohee_004_6	sohee > dfl	Jay	예	예	
26	output_sohee_005_1	sohee > dfl	Jay	예	예	
27	output_sohee_005_2	sohee > dfl	Jay	예	예	
28	output_sohee_005_3	sohee > dfl	Jay	예	예	
29	output_sohee_005_4	sohee > dfl	Jay	예	예	
30	output_sohee_005_5	sohee > dfl	Jay	예	예	
31	output_sohee_005_6	sohee > dfl	Jay	예	아니오	5초이상 끊김

그림 5. 변조영상 검수과정 예시

데이터 구축 담당자

수행기관(주관) : ㈜머니브레인 (전화: +82-2-858-5683), 이메일: com2best@moneybrain.ai