

2020년 인공지능 학습용 데이터 교육영상 “딥페이크 변조영상” 소개 및 활용

2021.03

(주) 머니브레인
연구원 권용재



KoDF: A Large-scale Korean DeepFake Detection Dataset

MoneyBrain, Inc.

데이터 소개
Introduction

01

데이터 구조
Structure

02

데이터 활용
Application

03

TABLE OF CONTENTS

01

데이터 소개

KoDF: A Large-scale Korean DeepFake Detection Dataset





Face Swap - Photo Face Swap
VysionApps Photography ★★★★★ 44,576

Contains Ads
⚠️ You don't have any devices.

Add to Wishlist [Install](#)

This is a screenshot of an app store listing for 'Face Swap - Photo Face Swap'. The app is developed by VysionApps and is categorized under Photography. It has a 5-star rating from 44,576 reviews. The listing includes a circular icon showing a face swap, a 'Contains Ads' warning, and an 'Install' button.



RefaceAI
Face swap technology demo

The image features the RefaceAI logo, which consists of two overlapping green circles with a white double-headed arrow between them. Below the logo is the text 'RefaceAI' in a large, white, sans-serif font, and 'Face swap technology demo' in a smaller, white, sans-serif font below that.



ANDROID / LIFESTYLE / FUNNY / ZAO

ZAO
1.71
MOMO

★ 4 | 📄 10 | ⬇️ 105.8k | 🌐

Add your face to any video and make surprising deepfakes

This is a screenshot of an app store listing for 'ZAO'. The app is categorized under ANDROID, LIFESTYLE, FUNNY, and ZAO. It has a 4-star rating from 10 reviews and has been downloaded 105.8k times. The listing includes the ZAO logo, the text 'ZAO 1.71 MOMO', and a description: 'Add your face to any video and make surprising deepfakes'.

여성신문 PINK 2021.01.13 내미비뉴스

아이돌 불법촬영 '딤펀이' 영상 유포...소속사 "법적 대응"

역성 연예인 딤펀이로 강력 처벌 청와대 국민청원도 하루 세 16만명 동참 최근 걸그룹 '모모랜드' 멤버와 관련된 불법촬영물과 딤펀이로 영상어 은근리장식 회피자, 소속사가 강력 ...



“여성연인 영상 불법물 딤펀이로 처벌” 하루... 종합일보 2021.01.13 내미비뉴스

“여성연인 ‘딤펀이’로 강력 처벌” 후 청원 하루... 노컷뉴스 2021.01.13 내미비뉴스

음반물해 여성연인 영상... ‘딤펀이’로 처벌... 동아일보 2021.01.13 내미비뉴스

“여성연인 불법영상 ‘딤펀이’로 처벌” 취청... 스포츠서울 2021.01.13 내미비뉴스

종합뉴스 24건 전체보기 >

새들경제 PNC 7일 전 내미비뉴스

미성년자 대상 '딤펀스-딤펀이' 논란...“제2의 n번방 사태”

언리 딤펀이(왕성 편집물)를 포함한 당사자 동의 없이 유포한 음란 영상물까지 디지털 성범죄물로 다... 또 “딤펀이”는 이미 디지털 성범죄 범정부 공동추진단에서 단속하고 ...



“미루다가 보아올린 남자 결혼...알래스 딤펀이로 논란” 뉴스1 7일 전 내미비뉴스

[종합뉴스] 어부다 후폭풍? 딤펀이로 알래스 ‘성희롱 논란’ 참... 뉴스투데이 7일 전

뉴스 전체 > 최근 1.49 / 10,500건 2020년 1월 전체 >

최근전제 > 언론사전제 > 후기를 > X

알래스-딤펀이로 논란...제2의 n번방 우려 4시간전 여대생이 다들뉴스
논란이 연다 글을 보면 보이고 있다. 몇몇 남성들의 여성을 대상으로한 딤펀이로(왕성)의 불법물 모호노...
정신 유해 공격성이 강하다는 지적이 등장하자 알래스...
- 수호신 2월13일 09:46에 연다 후자로 연다 딤펀이... 3시간전 뉴스엔 다들뉴스

딤펀이로? 딤펀이로? 음란 영상물, 딤펀이로 2021.01.13 07:47 다들뉴스
여해 차로 올라간 여성 연예인을 딤펀이로 처벌 관련 청원 동영상과 하루 만에 40만명을 얻었습니다. 모 년...
간 여자 연예인은 물론 여성 상층, 일반인까지 성폭행성폭행...

딤펀이로 국민청원 이틀만 30만 8천여명 동의 1시간전 스포츠서울 다들뉴스
[스포츠서울 김수현 기자] 딤펀이로 국민청원 이틀만 30만 8천여명 동의 딤펀이로(Deppent) 강력 처벌 청원...
이 30만 8천여명 동의를 기록했다. 14일 오후 4시 48분 현재 집계...

알래스-딤펀이로, 2년엔 수사시대(수호신) 4시간전 스포츠서울 다들뉴스
[스포츠서울] 알래스-딤펀이로 등 주로 연예인을 대상으로한 불법 촬영물들에 대한 논란이 연이어 불거지...
고 있다. 사건 정파에 청원게시물 연이연 법적 대응의 순번...

알래스-딤펀이로, 또 네티즌 성폭행(스포츠서울) 2021.01.13 스포츠서울 다들뉴스
[스포츠서울 종합기자]알래스와 딤펀이로가 유선 귀로 올라왔다. 알래스가 지난 10일 청와대 국민청원에...
모르며 공격했지만 딤펀이로 역시 12일 국민청원에 올라...
- 도박 사이트 연이연 불법(왕성)과 딤펀이로 간담 불... 23시간전 연일뉴스 다들뉴스
- 알래스부터 딤펀이로까지 성적 대상화 처벌 촉구... 23시간전 엑스뉴스로... 다들뉴스...

왕성/네티즌 딤펀이로 '속임수'인사'...성폭행 범지는 '알래스'... 2021.01.13 매일경제 다들뉴스
알래스(Real Person Beauty)에 대해 연이연의 얼굴을 편집물에 등장하는 '딤펀이로(Deppent)'를 처벌해야하는 국민청...
원자 시작 하루만에 26만 동의를 얻으면서 연이연...

알래스-금성/딤펀이로 처벌, 남녀공동 무대 뒤 흥행 2021.01.13 매일경제 다들뉴스
왕 에로 카라니다에서 알래스는 실제 성범죄로 연이연까지 있는 반면 여성 연예인은 딤펀이로(왕성)와 같은...
다 단속한 취재를 듣고 있는 수감자 나오고 있다. 특히 네티즌...
- 알래스-금성/딤펀이로 처벌, 흥행해서 남녀... 2021.01.13 조선일보 다들뉴스
- ‘딤펀이로’ 처벌 ‘알래스’도 금성...청와대 청원... 23시간전 스포츠서울 다들뉴스
- 왕성 공동 무대 뒤 청원에 제치는 알래스 논란... 2021.01.13 매일신문

연이연 “알래스-딤펀이로” 실사한 부처, 아이돌 100명(왕성)
2021.01.13 스포츠서울 다들뉴스
네티즌 아이돌 등을 성적 대상화 하는 알래스(Real Person Beauty)에 딤펀이로(Deppent)가 사용자...
이수로 피로감을 가면서, 연이연 관계자들이 분해해 심각한...

이연물 성적 대상화 무도...알래스(수호신)로 제2의 n번방(왕성) 2시간전 스포츠서울 다들뉴스
알래스(수호신)가 기술은 얼굴을 교묘하게 편집물에 편집, 편집한 영상으로 유포하지만 딤펀이로(왕성)가 여러 차례들고...
를 연이연물 제작자로 연이연고 있다는 지파이다. 원 데미트는 지난...
- 알래스-딤펀이로(왕성)로 제2의 n번방(왕성)... 23시간전 스포츠서울 다들뉴스

알래스...딤펀이로 “아이돌물 성적 대상화” 연이연(왕성) 2021.01.13 5:45 다들뉴스
제하고 있다. 12일 가두게에 따르면, 여성 차치들에 대한 성적폭행 다들뉴스는 딤펀이로(Deppent), 보정 여...
객들을 성적대상화 하는 알래스(Real Person Beauty)...





딥페이크 탐지의 다음 스텝

<https://www.aihub.or.kr/aidata/8005>



Differences



한국인

기존 딥페이크 탐지 데이터셋
(e.g. DFDC, FF++)의 동양인 비율
부족

스크립트, 배경,
딥페이크 모델, 방해 소음,
참여인 연령, 성별 등을 다양화

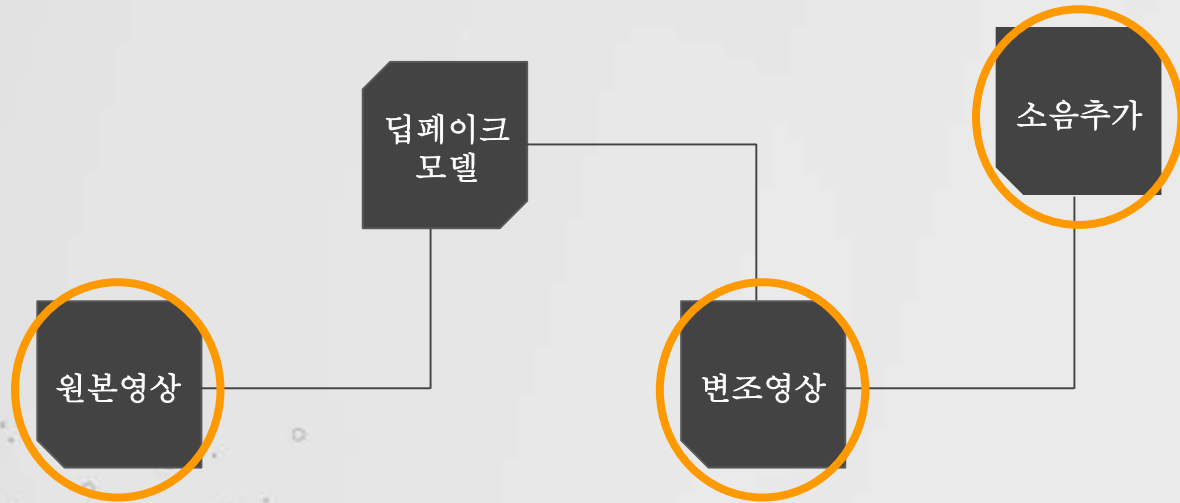
다양성



일관성

영상포맷, 길이,
촬영 각도, 조명,
철저한 메타데이터 등
데이터 품질을 정규화

OVERVIEW OF KoDF



OUR NUMBERS

400명

촬영자 수

88.5일

데이터 분량

6종

딥페이크 모델



Comparison

	변조영 상수	촬영자 수	전체영 상수	변조모 델수	전수검 수	직접촬 영
FF++	4,000	-	5,000	4	✗	✗
DFDC	104,500	960	128,154	2 (+ 6)	✓	✗
KoDF	175,776	400	237,942	6	✓	✓



02

데이터 구조

KoDF: A Large-scale Korean DeepFake Detection Dataset

데이터 수집 및 가공 과정

원본 데이터 수집

- (사전에 동의한) 400명의 참가자들을 다양한 환경에서 촬영
- 주어진 지문을 읽는 스크립트 세션, 질문에 자유롭게 답하는 시나리오 세션으로 분리

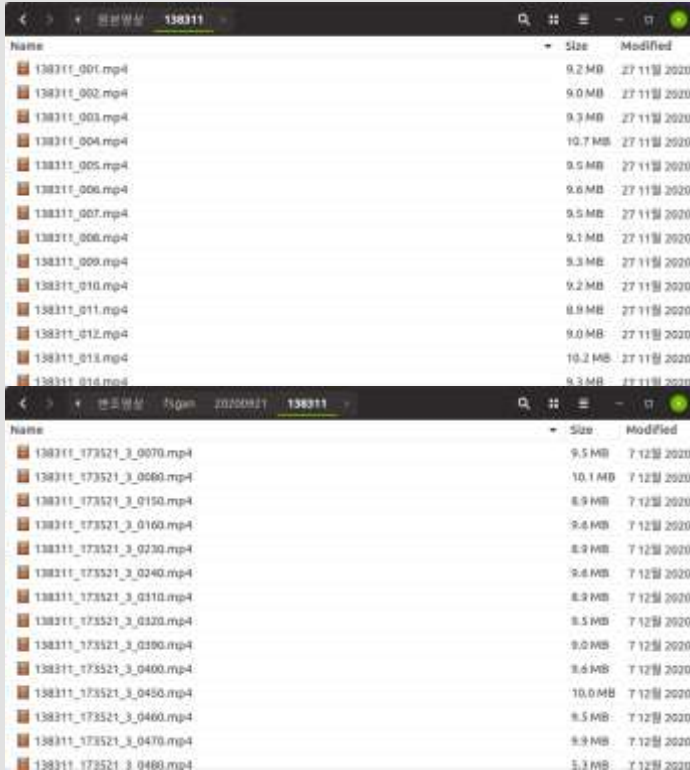


변조 데이터 가공

- Face Swap 방법론 : 한 사람의 얼굴을 다른 사람의 것으로 바꾸는 방법론
 - 모델 : DeepFaceLab, FaceSwap, FSGAN
- Face Reenactment 방법론 : 한 사람의 얼굴을 유지한 채로 표정을 바꾸는 방법론
 - 모델 : First Order, 3DMem, Wav2Lip



데이터 적재 구조



The image shows two screenshots of a file explorer window. The top screenshot shows a folder named '138311' containing 15 video files named '138311_001.mp4' through '138311_015.mp4'. The bottom screenshot shows a folder named '138311' containing 15 video files named '138311_173521_3_0070.mp4' through '138311_173521_3_0460.mp4'. Both screenshots show columns for Name, Size, and Modified.

Name	Size	Modified
138311_001.mp4	9.2 MB	27 11월 2020
138311_002.mp4	9.0 MB	27 11월 2020
138311_003.mp4	9.3 MB	27 11월 2020
138311_004.mp4	10.7 MB	27 11월 2020
138311_005.mp4	9.5 MB	27 11월 2020
138311_006.mp4	9.6 MB	27 11월 2020
138311_007.mp4	9.5 MB	27 11월 2020
138311_008.mp4	9.1 MB	27 11월 2020
138311_009.mp4	9.3 MB	27 11월 2020
138311_010.mp4	9.2 MB	27 11월 2020
138311_011.mp4	8.9 MB	27 11월 2020
138311_012.mp4	9.0 MB	27 11월 2020
138311_013.mp4	10.2 MB	27 11월 2020
138311_014.mp4	9.2 MB	27 11월 2020
138311_173521_3_0070.mp4	9.5 MB	7 12월 2020
138311_173521_3_0080.mp4	10.1 MB	7 12월 2020
138311_173521_3_0150.mp4	8.9 MB	7 12월 2020
138311_173521_3_0160.mp4	9.4 MB	7 12월 2020
138311_173521_3_0230.mp4	8.9 MB	7 12월 2020
138311_173521_3_0240.mp4	9.4 MB	7 12월 2020
138311_173521_3_0310.mp4	8.9 MB	7 12월 2020
138311_173521_3_0320.mp4	8.5 MB	7 12월 2020
138311_173521_3_0390.mp4	9.0 MB	7 12월 2020
138311_173521_3_0400.mp4	8.6 MB	7 12월 2020
138311_173521_3_0450.mp4	10.0 MB	7 12월 2020
138311_173521_3_0460.mp4	8.5 MB	7 12월 2020
138311_173521_3_0470.mp4	8.9 MB	7 12월 2020
138311_173521_3_0480.mp4	5.3 MB	7 12월 2020

- 익명성을 위해, 각 참가자들의 본명이 아닌 식별 ID를 사용함
- 원본 데이터 폴더 구조:
 - [참가자의 식별 ID] / [동영상 파일]
- 변조 데이터 폴더 구조:
 - [변조에 사용한 모델 ID]
 - [변조에 사용한 원본영상의 촬영 일자]
 - [참가자의 식별 ID] / [동영상 파일]



03

데이터 활용


KoDF: A Large-scale Korean DeepFake Detection Dataset

Hackathon

딥페이크 번조 영상 탐지 AI 경진대회
 시뮬레이션 | 영상 | GAN | 분해 | Accuracy

📍 장소: LAD2관
 📅 2020.10.19 - 2020.11.19 17:50 [신청하기](#)

👤 371명 🗳️ 투표



● WINNER ● 2% ● 4% ● 10%

순위	팀명	팀 멤버	지원 횟수	재출구	등록일
1	WeAreApp	👤👤👤👤	3,39478	60	2일 전
2	원리유치	👤👤👤	3,39329	34	4일 전
2	디얼지니	👤👤👤	3,35430	46	2일 전
4	ISPL	👤👤👤	3,36152	25	2일 전
5	ReLU의 #양재 박진호	👤👤👤👤	3,349	14	2일 전
6	김영재	👤👤👤👤	0,33802	37	2일 전



데이터 논문

KoDF: A Large-scale Korean DeepFake Detection Dataset

arXiv:1901.06971v3 [cs.CV] 26 Aug 2019

FaceForensics++: Learning to Detect Manipulated Face Images

Andreas Blattner¹, Davide Cozzolino¹, Emma Vecchioli¹, Christian Hoi²,
Alexei A. Efros¹, Marco Nanni¹

¹University of Applied Sciences, ²University of Salzburg, Austria



Figure 1. Architecture of a dataset of face images that include alterations to train deep-learning based approaches to a generalization problem. The dataset contains approximately equal mix of deep fakes, real faces, and real faces with different manipulations.

Abstract

The rapid progress in synthetic image generation and manipulation has now even given the creation of deep fakes a name for its, the machine generated face. As this field has a lot of open research, we consider a scientific evaluation for generating face manipulations in this paper. We propose a dataset that includes a mix of deep fakes and real faces with different manipulations and their corresponding real faces.

In this paper, we evaluate the detection methods, we propose an automatic benchmark for deep face manipulation. We provide a list of deep fakes and real faces with different manipulations. We provide a list of deep fakes and real faces with different manipulations.

1. Introduction

Manufacture of valid content has become ubiquitous, and one of the most common types is deep fakes. For instance, Facebook’s Live shows how real-time face swaps can be used to deceive people by replacing their face for the face of a different person. From our set of open research, we present a scientific evaluation of various methods. Using the methodology and marking of human faces, a well-annotated face to computer vision [1]. This is the foundation of face-swap applications. Recently, this has led to a great deal of interest in face-swap applications, and the face of a person can be swapped with another face [2]. This is the foundation of face-swap applications.

Common face manipulation methods can be categorized into two categories: facial expression manipulation and face-to-face manipulation. The first category is face-to-face manipulation, which is the most common. The second category is facial expression manipulation, which is the most common. The third category is face-to-face manipulation, which is the most common. The fourth category is facial expression manipulation, which is the most common.

© 2019 IEEE. This article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC 4.0 International).

Cats-DF: A Large-scale Challenging Dataset for DeepFake Forensics

Tomasz I. Kozłowski¹, Paweł Hryciuk² and Bartek Lysy¹

¹University of Applied Sciences, ²University of Wrocław, Poland

Abstract

A common face-manipulation method is known as DeepFakes, an emerging problem concerning the manipulation of video sequences. The need to detect and analyze DeepFakes has become ubiquitous, not only in the media but also in the legal system. However, current DeepFakes datasets are limited in terms of the number of manipulated faces and the quality of the manipulations. We consider a comprehensive evaluation of the detection of face manipulations in this paper. We propose a dataset that includes a mix of deep fakes and real faces with different manipulations.

5. Introduction

A rapid rise in the dissemination of video information in digital video content by 3D technologies, in particular, deep neural networks (DNNs), although beneficial and advantageous in many ways, has also led to the development of deep fakes. Deep fakes are a type of deep fakes that use deep learning to generate a realistic face from a target face. This category of deep fakes is the most common. The second category is face-to-face manipulation, which is the most common. The third category is facial expression manipulation, which is the most common. The fourth category is face-to-face manipulation, which is the most common.

The DeepFake Detection Challenge (DFDC) Dataset

Shuaqun Dong, Yueshan Chen, Yan Peng, Hui Lu, Bao Jiang, Mengjie Wang, Jintao Liang, Yifan Li

Beihai University



Abstract

DeepFake can be viewed as the deep transformation from the original video to the manipulated video. This is a challenging task because of the high similarity between the original and the manipulated video. In this paper, we propose a dataset that includes a mix of deep fakes and real faces with different manipulations.

1. Introduction

DeepFake can be viewed as the deep transformation from the original video to the manipulated video. This is a challenging task because of the high similarity between the original and the manipulated video. In this paper, we propose a dataset that includes a mix of deep fakes and real faces with different manipulations.

KoDF: A Large-scale Korean DeepFake Detection Dataset

Seungyeon Lee, Seungyeon Lee, Seungyeon Lee



Abstract

DeepFake can be viewed as the deep transformation from the original video to the manipulated video. This is a challenging task because of the high similarity between the original and the manipulated video. In this paper, we propose a dataset that includes a mix of deep fakes and real faces with different manipulations.

1. Introduction

DeepFake can be viewed as the deep transformation from the original video to the manipulated video. This is a challenging task because of the high similarity between the original and the manipulated video. In this paper, we propose a dataset that includes a mix of deep fakes and real faces with different manipulations.

1. Introduction

DeepFake can be viewed as the deep transformation from the original video to the manipulated video. This is a challenging task because of the high similarity between the original and the manipulated video. In this paper, we propose a dataset that includes a mix of deep fakes and real faces with different manipulations.



THANKS

Does anyone have any questions?

patrick@moneybrain.ai
+82-10-6337-6283
<http://moneybrain.ai>