딥페이크 (1)

양유진

한성대학교 HANSUNG UNIVERSITY

유튜브 주소: https://youtu.be/kq45C05q_i8

CryptoCraft LAB
https://crypto.modoo.at

Contents

CryptoCraft LAB

딥페이크란? (정의)

Deep(인공지능 딥러닝) + Fake(가짜)

- 인공 지능(AI) 기반으로 만든 가짜 이미지나 영상물
- 육안으로는 진위 여부를 판단하기 힘듦.

*딥페이크 악용기술

나쁜 의도를 가지고 딥러닝기술(GAN, AutoEncoder)을
이용하여 조작된 음성, 영상, 이미지를 만드는 방법

딥페이크 문제점

1. 기술이 공개되어 있음 → 누구나 쉽게 악용 가능

2. 부정적인 의도로 악용되는 경우가 많음
   (딥페이크 영상 中 96%가 불법 음란 동영상)
   (가짜뉴스와 연결됨)

오바마 딥페이크 영상
https://youtu.be/cQ54GDm1eL0

딥페이크 영상 생성 기술

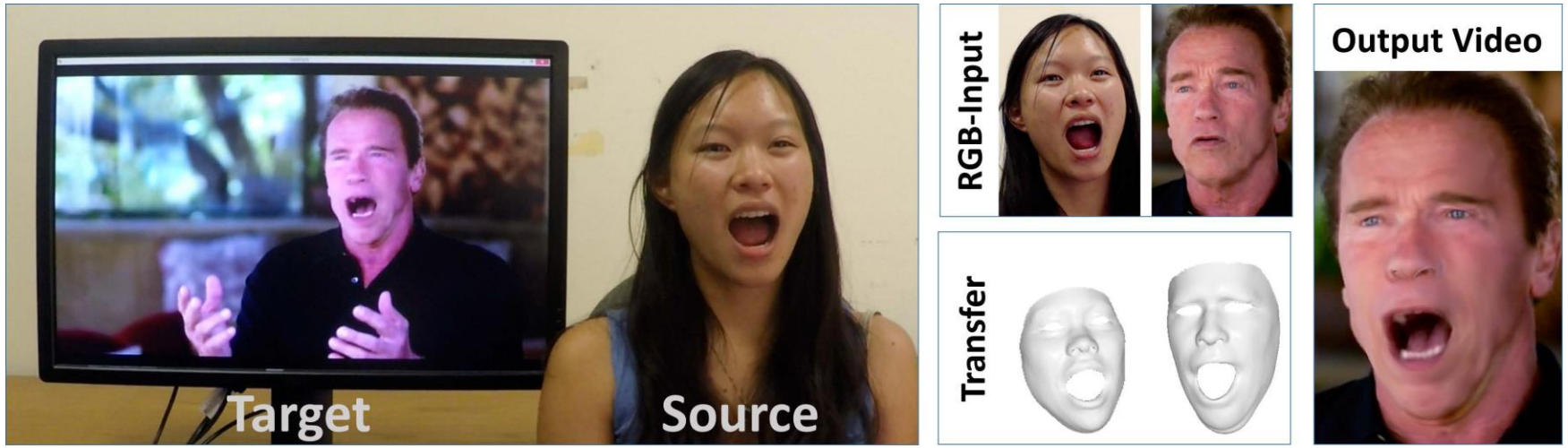
딥페이크 영상 생성 기술
- 컴퓨터 그래픽스 기반 방법 — Face2Face, FaceSwap
- 학습 기반 방법 — DeepFakes, NeuralTextures

가짜 영상 범주
- 얼굴 재연 방법 – 표정/감정 전달
- 얼굴 교체 방법 – 얼굴을 완전히 바꿈

컴퓨터 그래픽스 기반 방법(1) Face2Face

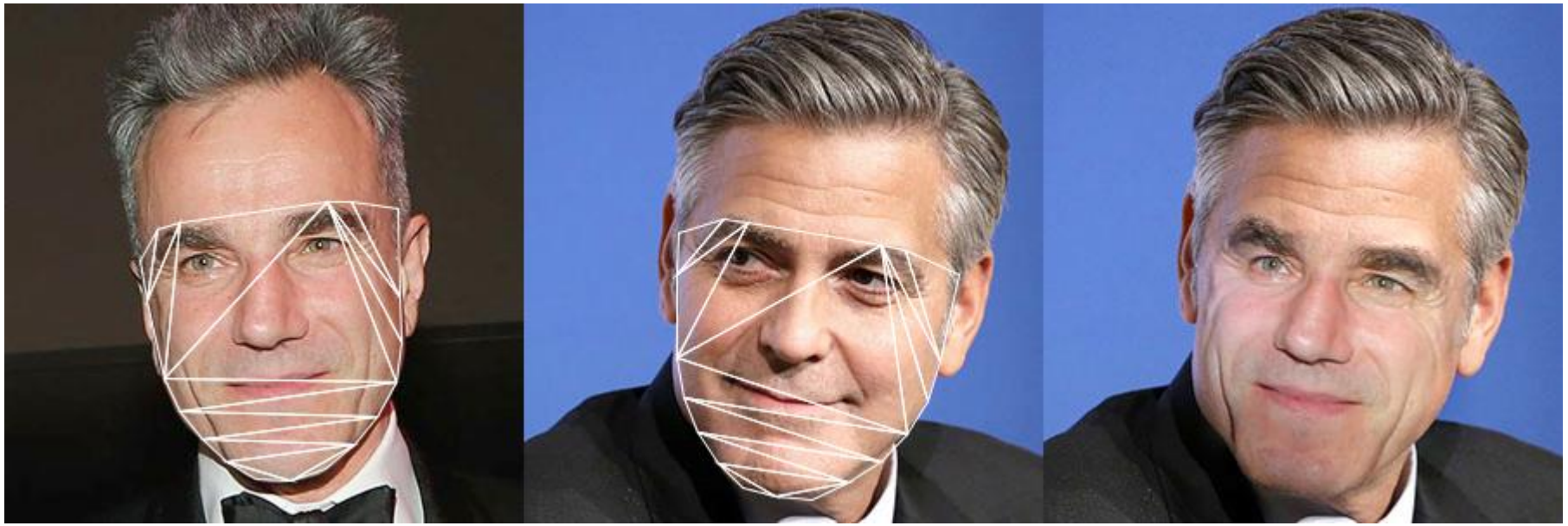
- 얼굴 재연 방법 (실시간으로 구현 가능)

: 목표 영상의 얼굴은 그대로 유지하고 원본 영상의 표정만 목표 영상에 전송하는 방법



컴퓨터 그래픽스 기반 방법(2) FaceSwap

- 얼굴 부위를 원본 → 목표 영상으로 전송하는 방법



원본 목표

컴퓨터 그래픽스 기반 방법(2) FaceSwap

1) 얼굴의 랜드마크 탐지 → 얼굴지역(눈,코,입 등) 추출

2) 탐지한 랜드마크로 3D 템플릿 모델 만듦

투영된 모양-랜드마크 간 차이
최소화시켜 대상 이미지에 투영됨.

3) Blendshapes(새로운 얼굴 표정 합성하는 기술) → 랜드마크 적용

4) 렌더링된 모델에 이미지 혼합 & 색상보정 적용 → 자연스럽게 생성

학습기반 방법(1) DeepFakes

- (딥러닝기반) 얼굴교체 방법

: 목표 영상 얼굴을 원본 영상/이미지 얼굴로 바꾸는 방법

- 2개의 오토인코더(원본 재구성용1/목표 재구성용1)

오토인코더

- 인코더: 입력데이터(이미지)의 특징들을 임의의 형태로 압축된 공간에 변환시켜 저장
- 디코더: 압축된 공간 정보를 다시 출력데이터(이미지)로 변환시킴

학습기반 방법(1) DeepFakes

(가짜 영상 생성)

각각 압축된 공간을 서로 바꿔 출력

→ 원본영상에서 학습된 인코더&디코더를 목표 얼굴에 적용

- 오토인코더 출력을 Poisson Image Editing에 넣어 나머지 부분 처리함

→ 이미지 자연스럽게 합성시켜줌

학습기반 방법(2) NeuralTextures

- 얼굴 재연 방법

: 목표 영상의 얼굴은 그대로 유지하고 원본 영상의 표정만 목표 영상에 전송하는 방법

- 목표 얼굴의 neural texture을 측광재구성손실&적대적손실 구성으로 학습함.

Photometric reconstruction loss & adversarial loss

신경텍스처

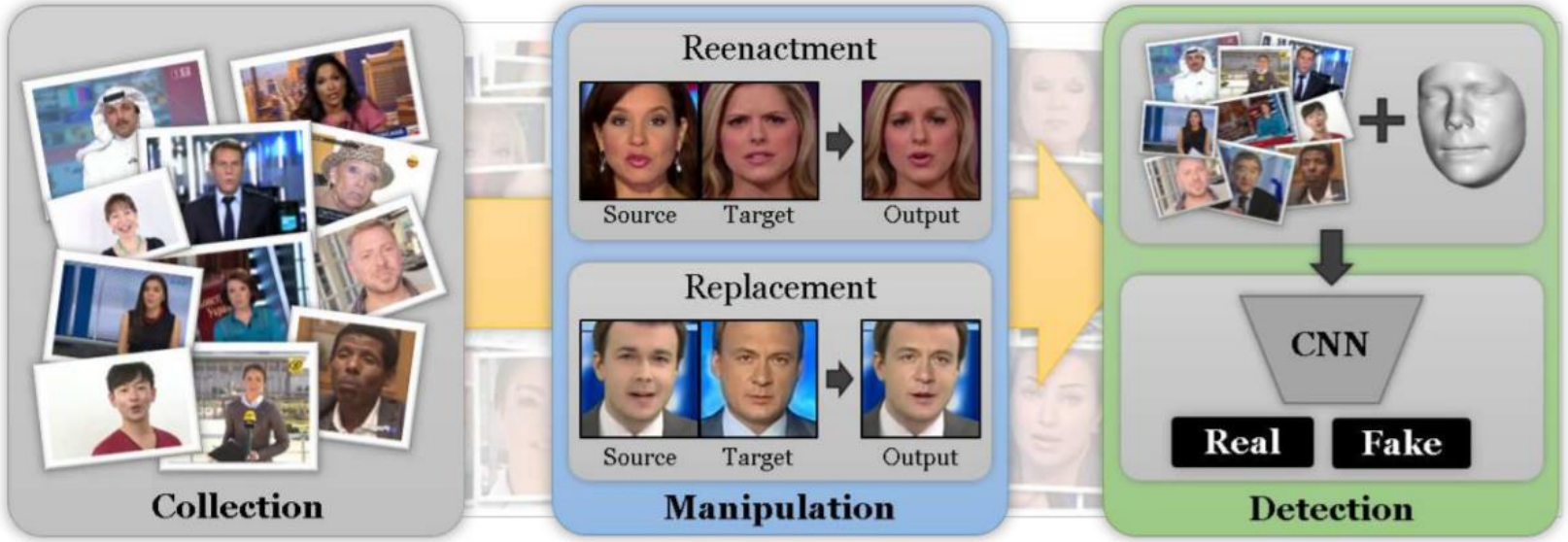
:그래픽의 기본 요소로 임의의 차원을 가질 수 있고 텍셀(텍스처 화소)당 학습된 고차원의 특징벡터를 저장할 수 있음.

딥페이크 데이터셋 – FaceForensics++

(1) SNS를 통해 동영상 수집

(2) 가짜 동영상 제작 및 배포

(3) 가짜 영상 탐지



FaceForensics++ 프로세스 흐름도

딥페이크 데이터셋 – FaceForensics++

한계

1) 대부분 서양인 얼굴 데이터셋 기반으로 제작되었음
→ 동양인 얼굴 데이터셋이 적음.

2) 딥페이크 생성 기술은 빠르게 발전 중이나 데이터셋이
이 기술의 발전 속도를 쫓아가지 못함.
→ 새로운 기술 포함x

감사합니다