

Real2Animation:애니메이션 제작지원을 위한 딥페이크 기술 활용 연구

신동주¹, 최봉준^{2*}

¹동서대학교 소프트웨어학과, ²동서대학교 소프트웨어융합대학

Real2Animation: A Study on the application of deepfake technology to support animation production

Dongju Shin¹, Bongjun Choi^{2*}

¹Department of Software, Dongseo University

²College of Software Convergence, Dongseo University

요약 최근 인공지능, 빅데이터, IoT 등의 다양한 컴퓨팅 기술이 발달하고 있다. 특히 콘텐츠 및 의료 산업 등 여러 분야에서 인공지능 기반의 딥페이크(Deepfake) 기술이 다양하게 활용되고 있다. 딥페이크 기술이란 딥러닝과 fake의 합성어로, AI의 핵심기술인 딥러닝을 통해 사람의 얼굴이나 신체를 합성하여 억양, 목소리 등을 따라 하게 만드는 기술이다. 본 논문은 딥페이크 기술을 활용하여 애니메이션 모델과 실제 인물사진의 합성을 통한 가상 캐릭터생성을 연구한다. 이를 통해 애니메이션 제작과정에서 일어나는 여러 가지 비용 손실을 최소화하고 작가들의 작업을 지원할 수 있다. 또한, 딥페이크 오픈소스가 인터넷에 퍼짐에 따라 많은 문제들이 나타나면서 딥페이크 기술을 악용한 범죄가 성행하고 있다. 본 연구를 통해서 딥페이크 기술을 성인물이 아닌 아동물에 적용하여 이 기술에 대한 새로운 관점을 제시한다.

• 주제어 : 딥페이크, 인공지능, 애니메이션, 딥러닝, GAN

Abstract Recently, various computing technologies such as artificial intelligence, big data, and IoT are developing. In particular, artificial intelligence-based deepfake technology is being used in various fields such as the content and medical industry. Deepfake technology is a combination of deep learning and fake, and is a technology that synthesizes a person's face or body through deep learning, which is a core technology of AI, to imitate accents and voices. This paper uses deepfake technology to study the creation of virtual characters through the synthesis of animation models and real person photos. Through this, it is possible to minimize various cost losses occurring in the animation production process and support writers' work. In addition, as deepfake open source spreads on the Internet, many problems emerge, and crimes that abuse deepfake technology are prevalent. Through this study, we propose a new perspective on this technology by applying the deepfake technology to children's material rather than adult material.

• Key Words : Deepfake, Artificial Intelligence, Animation, Deep Learning, GAN

Received 13 September 2022, Revised 28 September 2022, Accepted 30 September 2022

* Corresponding Author Bongjun. Choi, Colloge of Software Convergence, Dongseo University, 47, Jurye-ro, Sasang-gu, Busan, Korea.
E-mail: bongjun.choi@dongseo.ac.kr

I. 서론

최근 콘텐츠, 교육, 의료 등 다양한 분야에서 인공지능을 활용한 많은 연구가 진행되고 있다[1]. 특히 인공지능 기술인 딥페이크 기술을 영화산업과 OTT, 디지털 플랫폼에서 사용한 영상제작물이 늘어나면서 YouTube 및 구독 기반의 플랫폼 시장이 빠르게 성장하고 있다[2].

애니메이션 작가들이 캐릭터를 만들거나 디자인할 때 본인의 생활반경 안에 있는 사람들을 참고해서 많이 제작한다고 한다. 예를 들면 어떤 애니메이션의 주인공이 작가 가족의 얼굴을 모티브로 해서 만들 수 있을 것이다[3]. 새로운 캐릭터를 만드는 것은 쉽지 않은 작업이기에 실제로 많은 작가들이 성격, 이름, 외형, 일화 등을 주변 지인을 대상 또는 본인으로 한 경우가 많다[4]. 하지만 애니메이션 캐릭터를 제작할 때 누군가의 얼굴을 모티브로 한다고 해서 캐릭터의 특징을 표현하기 어렵고 많은 시간과 재정적인 비용이 발생한다[5]. 본 논문은 애니메이션 제작과정에서 비용 손실을 최소화하고 작가들의 작업을 지원하기 위해 딥페이크 기술을 활용하여 애니메이션 모델과 실제 인물사진을 합성한다. 합성된 캐릭터는 애니메이션의 새로운 등장인물 탐색 시 활용될 수 있다.

딥페이크 기술이란 딥러닝과 fake의 합성어로 인공지능 기술을 이용하여 진짜인지 가짜인지 구별하기 힘든 이미지나 비디오를 생성하는 기술을 뜻한다[6]. 딥페이크 단어가 등장한 시기는 2017년으로, 미국 온라인 커뮤니티 Reddit의 한 회원이 기존 영상에 유명인의 얼굴을 입혀 가짜 콘텐츠를 게재한 데서 유래되었다. 이후 온라인 커뮤니티와 SNS를 중심으로 급속히 확산된 딥페이크 콘텐츠는 최근 DeepFaceLab, Faceswap 등 오픈 소스 형태의 영상 합성 제작 프로그램이 배포되면서 더욱 성행하고 있다. 이에 본 연구는 딥페이크 기술을 활용하여 영화산업 중 애니메이션 영화 제작에 있어서 캐릭터의 얼굴에 가까운 가족이나 친구들 지인들을 포함해 실제 사람의 얼굴을 동화에 접목시켜 새로운 결과물을 생성하고자 한다. 딥페이크 기술로 새롭게 생성된 애니메이션을 통해 폭넓은 레퍼런스 선택과정과 작업 과정 생략은 콘텐츠 제작 프로세스 전반에 긍정적인 효과를 기대할 수 있을 것이다. 또한, 현재 딥페이크에 관한 연구가 활발히 이루어지

고 있는 가운데 딥페이크 오픈 소스가 인터넷에 퍼짐에 따라 많은 문제들이 나타나면서 딥페이크 기술을 악용한 범죄가 성행하고 있다. 온라인에 쉽게 공개되어있는 유명인, 연예인들의 사진이나 영상을 합성하여 음란물을 만드는가 하면 정치적으로도 악용되어 이용된 사례가 있다. 이뿐만 아니라 일반인들까지도 SNS의 발달로 인해 SNS에 게시된 사진, 영상을 통해 피해를 입어 일반인들조차 보호받지 못하는 경우가 있다. 본 연구는 애니메이션 제작 지원을 위한 딥페이크 연구를 기술하고 영상 애니메이션 제작 환경에 공헌할 뿐만 아니라 딥페이크 기술을 성인물이 아닌 아동물에 적용하여 이 기술에 대한 새로운 관점을 제시한다.

II. 관련 연구

2.1 딥페이크 영상 탐지 기술

딥페이크는 딥러닝 기술을 이용하여 동영상을 조작하는 것으로, 크게 추출(extraction) → 학습(learning) → 생성(generation) 단계로 수행된다. 먼저 추출은 바꾸거나 바뀔 얼굴 이미지의 특징을 식별하여 얼굴 이미지의 특징을 확보하는 단계로서 기계가 얼굴 특징을 배울 수 있도록 해 주는 것이다. 특정인의 얼굴 사진을 입력하고 얼굴 인식 기술을 통해 얼굴 부분을 감지한 후 그 부분만 잘라내서 얼굴의 위치가 맞도록 정렬한다. 추출이 완료되면 학습을 진행한다. 이 단계는 딥페이크의 핵심 단계로 앞에서 추출한 얼굴 특징을 바꾸고 싶은 얼굴에 재생성하도록 하는 과정이다. 딥페이크에서 학습이 이뤄지게 하는 장치는 머신러닝의 한 종류인 오토인코더(Autoencoder)로 인풋레이어-인코더-디코더-아웃풋레이어로 구성되어 있으며 입력값을 받아들이고 인코더가 압축한 후 디코더가 출력값이 동일하도록 생성시키는 알고리즘을 가지고 있다. 오토인코더는 인코딩 과정에서 데이터(얼굴 특징)의 불필요(노이즈)를 제거하여 중요한 특징만 뽑아낸다. 훈련 과정이 끝나면, 마지막으로 재구축된 얼굴을 원본 이미지에 병합하고 생성하는 단계로 넘어간다. 이렇게 만들어진 딥페이크 영상을 보면 맨눈으로 구별하기 어렵게 얼굴이 바뀌는 경우도 있으나 뭔가 이상한 것이 느껴지는 것들도 있다. 이것은 딥페이크가 가지는 한계 때문이다. 인코더는 피부나 다른 장면 정보를 인식하지 못한

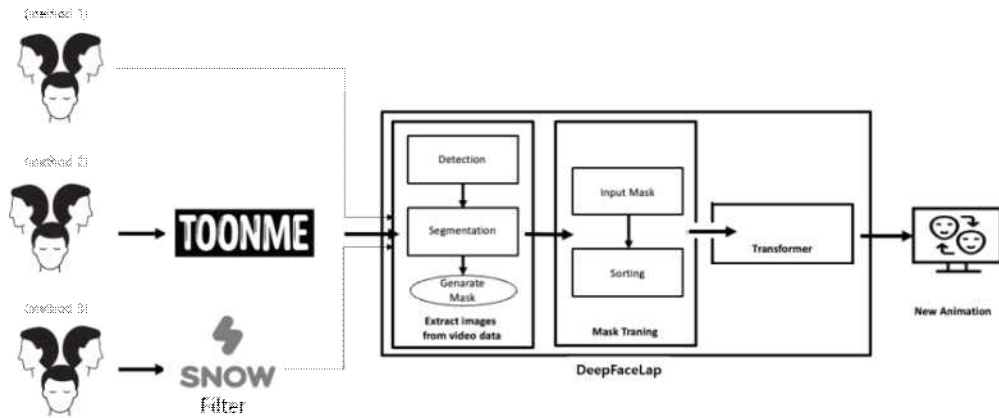


Fig. 1. Overall Process of Our Study.

다. 삽입하려는 얼굴과 원래 얼굴 사이에서 경계효과를 가지는 것이 일반적사례이다. 또한, 조명이 일관되지 않아 얼굴이 번쩍거리는 현상(대부분의 가짜 영상들에서 공통적인 얼굴 부분이 깜빡이는 현상)이 발생하기도 한다. 이에 따라 딥페이크 동영상에 대응하고 탐지하려는 많은 연구 및 시도들이 진행되고 있다[7]. 얼굴이 변환된 영상에서 발생하는 이러한 문제점들은 실제 인물 얼굴을 애니메이션 캐릭터와 변환하였을 때 더 큰 문제가 발생한다. 이러한 문제점을 해결하기 위해 본 연구에서는 영상 필터를 활용한 자연스러운 얼굴변환기법 기술을 제안한다.

2.2 컨볼루션 신경망을 이용한 페이스 스왑

포즈, 표정 및 조명을 유지하면서 입력 ID가 대상 ID로 변환되는 이미지의 얼굴 스와핑 문제를 고려한다. 이 매핑을 수행하기 위해 구조화되지 않은 사진 컬렉션에서 대상 아이덴티티의 모양을 캡처하도록 훈련된 컨볼루션 신경망을 사용한다.

이 접근 방식은 다른 스타일의 이미지를 렌더링하는 것이 목표인 스타일 전송 측면에서 얼굴 스와핑 문제를 프레임화함으로써 가능한 것이다. 이 분야의 최근 발전을 바탕으로 네트워크가 매우 사실적인 결과를 생성할 수 있도록 하는 새로운 손실 함수를 고안했다. 심층 신경망을 사용하여 거의 실시간으로 완전히 자동화된 얼굴 스왑에 대한 개념 증명을 제공했다. 새로운 목표를 도입하고 신경망을 사용한 스타일 전달이 사람 얼굴의 사실적인 이미지를 생성할 수 있음을 보여주었

다. 제안하는 방법은 특정 유형의 안면 교체를 다룬다. 여기서 가장 큰 어려움은 원본 포즈, 표정 및 조명을 변경하지 않고 아이덴티티를 변경하는 것이었다. 이 특정 문제는 이전에 해결되지 않았고 분명히 극복해야 할 몇 가지 문제가 있지만 신경망 기반 얼굴 스와핑이라는 어려운 문제에 대해 상당한 진전을 이루었다[8]. 하지만 본 연구에서는 실제 얼굴을 애니메이션 캐릭터와 얼굴 스와핑을 진행하기 때문에 이 연구의 방법으로는 본 연구의 문제점을 해결할 수 없는 문제점이 있다.

III. 시스템 연구

본 연구에서는 딥페이크 기술로 DeepFaceLab 라이브러리를 통해 실제 사진의 캐릭터화를 통한 애니메이션 딥페이크 작업에 중점을 두어 영상 품질을 높이는 것을 목표로 연구가 진행되었다. 연구 과정에서 실제 인물의 영상과 애니메이션 영상의 딥페이크 훈련 시 소스 비디오(src)에서 추출하고 훈련하여, 대상 비디오(dst)의 인물을 변환하는 딥페이크의 특성상 애니메이션캐릭터와 실제 사진 간의 극명한 특징적 차이로 인해 같은 스타일의 딥페이크 영상물 제작 시보다 낮은 훈련 성능을 보임을 확인한다.

이처럼 딥페이크 영상 제작에 있어 애니메이션 대상 비디오 생성에 대한 문제가 존재하며 연구가 활발히 이루어지지 않았다. 또한 딥페이크 기술에 있어서 딥페이크의 악용으로 인한 무분별하고 악의적인 영상물 탐지에 관한 연구들만 다수이고 딥페이크 기술의

활용방안에 관한 연구는 거의 이루어지지 않았다. 따라서 본 논문에서는 DeepFaceLab을 통한 실제인물 영상과 애니메이션 영상 인물의 딥페이크를 통해 위와 같은 문제를 해결하기 위한 방법론적 연구 방향을 제시한다. 이는 다양한 딥페이크 방법론 제시를 통해 실제인물 영상과 애니메이션 영상 간 딥페이크 성능을 강화하여 영상 제작에서 딥페이크 기술의 활용방안과 딥페이크 자체의 인식 개선을 제시했다는 점에서 연구적 의의가 있다.

본 연구가 진행한 연구의 방법은 다음과 같다(Fig. 1): (method 1) 먼저 실제 인물 영상과 캐릭터 영상을 학습한다; (method 2) TOONME라는 인물 사진을 캐릭터 사진으로 변환해주는 프로그램을 활용하여 실제인물 영상의 인물을 캐릭터화하여 학습한다; (method 3) 인물 사진을 필터링을 통해 캐릭터화하여 영상으로 변환해주는 SNOW 프로그램을 사용하여 캐릭터화된 영상과 애니메이션캐릭터 영상을 학습한다. 각 순서에 대한 자세한 연구 방법은 다음과 같다.

3.1 method 1 : Learning real person images

본 방법에서는 사용자의 실제 얼굴 사진을 여러 방향으로 촬영한 후 사진을 영상으로 변환하여 소스 비디오(src)의 데이터로 제작하였다(Fig. 2) 제작된 데이터를 DeepFaceLab을 통해 애니메이션 캐릭터 영상과 학습을 진행한다.

학습된 데이터를 통해 대상 비디오(dst)의 인물변환시 대상비디오(dst)의 인물변환은 잘 진행되었지만, 현실의 실제 인물 이미지와 애니메이션 이미지 영상의 텍스처 차이로 인해서 의도치 않은 결과물을 볼 수 있다(Fig. 2).



Fig. 2. Method 1 Result

3.2 method 2 : Learning Cartoonization images

본 방법에서는 실제인물 사진을 애니메이션화 하여 소스 비디오(src)의 데이터로 input 하는 방법을 선택한다. 애니메이션화에는 인물 사진을 캐릭터 사진으로 변환해주는 프로그램 중 가장 대중적인 TOONME라는 ai 변환 라이브러리를 활용하여 캐릭터화하였다. 그 후 캐릭터화된 사진들을 이어 붙여 영상으로 변환 후 애니메이션캐릭터 영상과 학습을 진행한다. 변환 결과 단순히 실제 인물 사진을 애니메이션화 한 뒤 영상 파일 형태로 만들었기 때문에 대상비디오(dst)의 인물변환은 method 1 방법에 비해 캐릭터가 결합된 부분이나 얼굴의 움직임 등이 자연스러워진 것을 확인할 수 있다. 하지만 단순히 정적인 사진 파일 여러 장을 연결하여 영상화하였기 때문에 영상이 재생될 때 캐릭터의 입 모양이 벌어지지 않거나 말을 더듬는 듯한 현상 등의 섬세한 움직임이 나타나지 않는 문제가 여전히 남아 있다(Fig. 3).



Fig. 3. Method 2 Result

3.3 method 3 : Learning cartoon video filter

본 방법에서는 캐릭터 영상 필터를 사용하여 학습 영상데이터로 사용하였다. 영상 필터로는 SNOW 애플리케이션을 활용하여 촬영한다.

영상 필터를 소스 비디오(src)의 데이터로 학습하여 대상 비디오(dst)의 캐릭터를 변환한 결과 영상 필터를 사용한 소스 비디오의 다양한 얼굴 데이터를 통해 대상 비디오 캐릭터의 섬세한 움직임이 나타나며 변환이 잘 진행된 자연스러운 결과물을 생성할 수 있었다. 정적인 사진을 바로 적용한 경우(method 1), 정적인 사진을 캐릭터화 후 적용한 경우(method 2)와 비교했을 때 얼굴이 결합된 부분이나 얼굴의 움직임뿐만 아니라 캐

릭터의 표정이나 말하는 입 모양이 자연스러워진 것을 확인할 수 있다.



Fig. 4. Method 3 Result

IV. 결과 및 분석

본 연구의 실험 결과 method 2, 3 방법에서와 같이 애니메이션 캐릭터와 실제 인물의 딥페이크 결과가 자연스럽게 나오는 것을 확인할 수 있다.

method 1의 경우 앞의 연구 부분에서 언급했듯이 애니메이션과 실제 사진 간의 텍스처와 다른 얼굴 구조로 인해 합성 부분의 경계선 부분에 극심한 차이가 발생함을 볼 수 있었다. 또한 정적인 이미지 사진을 촬영하여 영상 형태로 만들었기에 한 인물에 대한 입 모양이나 눈 모양의 자연스러운 학습이 되지 않아 영상 변환 시에 대상 인물이 대사할 때 입이 벌어지지 않거나 눈이 깜빡이지 않는 등의 문제가 발생하며 어색한 모습을 보여준다.

method 2에서는 자연스러운 텍스처 합성을 위해 실제 인물 이미지를 카툰화 라이브러리를 통해 변환하였지만, 학습데이터 역시 method 1과 같이 카툰 느낌으로 변환된 이미지의 영상 형태로 만들었다는 점에서 같은 맥락이기에 인물의 얼굴 움직임에 제대로 학습되지 않은 어색한 모습을 보인다.

마지막 method 3에서는 method 1과 method 2의 사진 데이터를 영상으로 변환한 것과 달리 카툰 영상 필터 소스를 통해 바로 애니메이션 느낌의 영상데이터를 만들었다(Fig. 5). 이를 통해 자연스러운 입의 움직임과 얼굴 근육의 움직임이 학습된 것을 확인할 수 있다.

본 논문에서 제안한 3가지 method를 확인해본 결과 실제 인물의 사진 또는 영상의 데이터만으로 간단하게 새로운 애니메이션 캐릭터를 생성하고 영상에 적용할

수 있음을 확인할 수 있고 각 method 들의 결과를 함께 보면 method 1에 비해 method 2와 3은 품질이 많이 개선된 것을 확인할 수 있다(Fig. 6).



Fig. 5. Cartoon video filter Learning



Fig. 6. Compare all method

하지만 애니메이션 비디오 필터의 학습 과정에서 중간중간 애니메이션 필터의 깜빡임으로 인한 실제 얼굴이 함께 학습되어 안정성이 떨어져 오랜 시간 학습할수록 몇몇 구간 강한 노이즈가 생겨 알아보지 못하는 문제가 발생한다(Fig. 7).



Fig. 7. Video noise

V. 결론 및 향후 개선 방향

본 연구는 딥페이크 기술 중 DeepFaceLab 라이브러리를 활용하여 애니메이션 제작 지원을 위한 딥페이크 활용 기술을 연구하였다. 연구 결과들을 바탕으로 method 1과 같이 캐릭터화하지 않은 일반영상에 대해서는 딥페이크 변환이 잘 이루어지지 않음을 확인하였다. 이를 바탕으로 method 2와 3의 연구 방법을 통해 실제인물과 애니메이션 인물 간의 딥페이크 변환 시 결과물의 완성도를 높여 나갈 수 있었다.

향후 연구에서는 실제인물과 애니메이션 인물 간의 차이를 좁혀나가며 애니메이션화에 사용되는 라이브러리의 튜닝을 통해 실제인물과 가장 비슷한 애니메이션 얼굴로 변환하여 가장 자연스러운 학습 결과물 도출하는 것을 목표로 하며 동시에 학습 시간을 줄이면서 데이터가 깨지지 않는 최대의 품질을 얻을 수 있는 설정값을 찾기 위해 연구할 계획이다.

ACKNOWLEDGMENTS

본 연구는 2022년 학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 연구결과로 수행되었음 (2019-0-01817).

REFERENCES

[1] M. S. Choi and B. J. Choi, "A Study on the PBL-based AI Education for Computational Thinking", KICSP, Vol. 22, No. 3, 2021, pp 110-115.

[2] S. W. Lee, S. M. Lee, H. M. Joo and Y. J. Nam, "Examining Factors Influencing Early Paid Over-The-Top Video Streaming Market Growth: A Cross-Country Empirical Study", Sustainability, 13(10), 5702, 2021.

[3] "The sound of my heart, Ae-bong and Jo-seok. GFRIEND's motif character eyes", MBN's daily broadcast, 2014.07. Available: <https://star.mbn.co.kr/view.php?no=995789&year=2014>

[4] B. R. Yi. "A Study on Kiduit Character Consuming Trend' s Creation process and Characteristics : A Focus on Webtoon The Sound of Mind", The Korean Journal of animation, Vol 12, No. 3, 2016

[5] T. H. Cho, J. P. Jeong and S. M. Choi, "3D Emotional Avatar Creation and Animation using Facial Expression Recognition", JKMS, Vol. 17. No. 9, 2014, pp.1076-1083.

[6] J. H. Kim, J. J. Ahn, B. S. Yang, J. U. Jung And S. S. Woo, "Investigation of the Latest Technology Trends in Data-Based Deepfake Detection Techniques", Journal of the Society for Information Protection, Vol. 30, No. 5, pp.11212, 2020. 10.

[7] J. Park and Y. H. Cho, "Technology Trends Related to Deepfake Video Detection", 2019 Korea Software Conference Paper Collection, 724-726

[8] I. Korshunova, W. Shi, J. Dambre and L. Theis. "Fast face-swap using convolutional neural networks." Proceedings of the IEEE international conference on computer vision. 2017.

저자소개

신 동 주 (Dongju Shin)



2017년 3월~현재 : 동서대학교
소프트웨어학과
관심분야 : 인공지능, 딥러닝

최 봉 준 (Bongjun Choi)



2013년 2월 : 동서대학교
컴퓨터정보공학부(공학사)
2015년 2월 : 연세대학교
컴퓨터과학과(공학석사)
2014년 12월 ~ 2019년 2월 :
LG 전자
2019년 2월 ~ 2021년 1월 :
Aalborg University Operation

Research Lab

2021년 3월 ~ : 동서대학교 소프트웨어융합대학 조교수
관심분야 : 데이터 분석, 데이터 생성, GAN, 자율주행, 중등SW교육