

웹 서비스를 위한 도커 컨테이너 기반 얼굴 변형 학습 모델 연동 기술 연구

조수화*, 김채규**, 송복득***

*부경대학교, IT 융합응용공학과

**부경대학교, 인공지능 연구소

***한국전자통신연구원

e-mail: kyu0707@pknu.ac.kr

Face Deformation Learning Model Using Docker Container for Web Services

Su-Hwa Jo*, Chae-Kyu Kim**, Bok-Deuk Song***

*Dept of IT Convergence and Application Engineering,
Pukyong University

**Artificial Intelligence Laboratory, Pukyong University

***Electronics and Telecommunications Research Institute (ETRI).

요 약

본 논문은 가상화 기술 중 도커를 활용하여 단일 서비스 시스템에서 자원을 공유하여 이종의 시스템인 인터랙티브 미디어 창작 서비스 플랫폼과 얼굴 변형 학습 모델을 연동에 관한 연구를 수행하였다. 이는 다양한 딥러닝 프레임워크를 별도의 시스템 구축 없이 동일 자원에서 이종 시스템을 구축하여 자원 소모를 줄이고 추가 딥러닝 프레임워크와의 연동을 원활하게 할 수 있는 장점을 제공한다.

I. 서 론

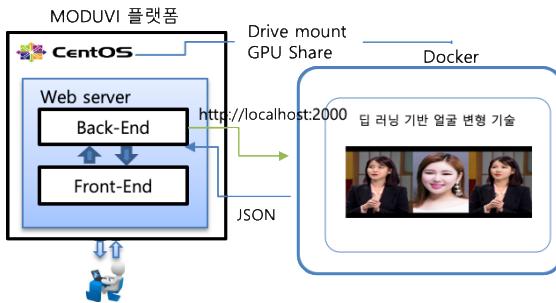
오늘날 인공 지능 관련 기술이 발전하면서 다양한 딥러닝 프레임워크가 제공되고 있다. 특히 얼굴 변형을 위한 최근의 딥러닝 모델은 정면 얼굴 포즈 영상 변환에 적합한 Adaptive 2-D faceswap 모델[1][2], 신체 움직임이 활발한 얼굴 포즈 영상 변환에 적합한 Adaptive OpenPose 모델[3], 좌우 15 도 정도의 얼굴 움직임에 적합한 Adaptive Original High Resolution 모델[4], 그리고 GAN 을 이용한 얼굴 변형 기능의 탐색 모델[5][6] 등이 다양한 프레임워크를 이용하여 개발되고 있다. 이러한 다양한 학습 모델을 적용하기 위해서는 각각의 딥러닝 프레임워크를 별도로 시스템 환경 구축 해야 하는 번거로움이 발생한다.

한편으로 이러한 번거로움을 개선하기 위해 컨테이너 기반 가상도구인 Docker 가 개발되어 있다. 도커는 2014년 1.0 버전을 발표한 이후, 현재까지 최신 버전으로 발전하고 있다. 도커는 컨테이너 기반의 오픈소스 가상화 플랫폼으로 OS 를 가상화 시켜 별도의 환경을 컴퓨터들이 갖추지 않아도 되는 장점을 가진다.

본 논문은 가상화 기술 중 도커[7]를 활용하여 딥러닝 프레임워크를 사용하는 얼굴 변형 학습 모델을 도커 컨테이너로 구축하여 웹 서비스 환경인 리눅스 환경의 인터랙티브 미디어 창작 서비스 플랫폼과의 연동에 관해 연구하였다.

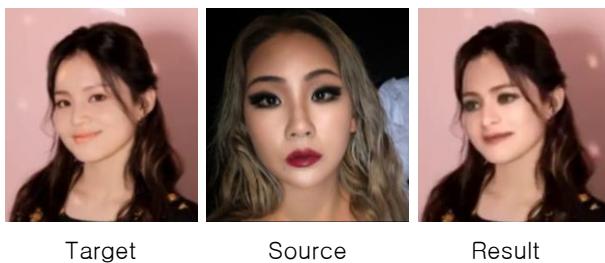
II. 본 론

<그림 1>은 딥러닝 프레임워크를 사용하는 얼굴변형 학습 모델을 도커 컨테이너로 구축하여 웹서비스와 연동하는 과정이다. <그림 1>과 같이 단일 서비스 시스템에서 GPU 를 공유하면서 웹서비스 시스템과 딥러닝 기반 얼굴 변형 학습 모델을 연동함으로써 별도 시스템을 구축할 필요 없이 사용자의 인터랙션에 따른 얼굴 변형 기술을 적용할 수 있다. 인터랙티브 미디어 창작 서비스 플랫폼인 MODUVI 에서 얼굴 변형 기술을 사용하기 위해 변형하고자 하는 얼굴 정보를 입력 받는다. 또한 사용자의 선택에 따른 얼굴 변형 기술을 적용하기 위해 사용자의 인터랙션을 UI 를 통해서 입력 받는다. 입력된 얼굴 변형 정보를 PHP 로 구현된 Back-End 모듈에서 REST API 로 도커 컨테이너로 실행 중인 얼굴 변형 모듈을 호출한다. MODUVI 플랫폼에서는 변형하고자 하는 타켓 영상 정보, 변환 이미지, 변형하고자 하는 얼굴 영역 정보를 얼굴 변형 모듈에 전송한다. 얼굴 변형 모듈에서는 입력된 타켓 영상과 얼굴 영역 정보를 이용하여 얼굴 변형 기능을 실행한다. 실행이 완료되면 MODUVI 플랫폼에 결과 정보를 JSON 포맷으로 전송한다. MODUVI 플랫폼은 전송된 JSON 내용에서 실행결과 파일 정보를 읽어와서 재생한다.



<그림 1> 도커 컨테이너 기반 얼굴 변형 모델 연동 과정

본 논문에서 사용된 얼굴 변형 모듈은 Talking-Head[8]의 구조를 빌려 얼굴 재현 네트워크를 설계하였고 Meta-train 을 수행하였다. Meta-train 은 VoxCeleb2[9] 비디오 데이터 셋으로 6,112 인물의 145,569 개의 비디오에서 각각 8 장의 프레임을 추출하여 사용하였다. 기존 논문에서는 256 사이즈의 텐서를 사용하였지만, 데이터 셋의 해상도가 낮았고, 224 사이즈의 이미지에서 노이즈가 적게 발견되었기 때문에 224 사이즈 이미지를 사용하였다. 설계한 얼굴 재현 네트워크를 이용한 결과는 <그림 2>와 같다.



<그림 2> Talking-Head 기반 얼굴 재생성 결과 화면

얼굴 변형 모델의 환경은 다음과 같다.

```
python 3.6
torch 1.1.0
face-alignment 1.0.0
numpy 1.16.4
opencv-python 4.1.0.25
matplotlib 3.1.1
```

<그림 3>은 MODUVI 플랫폼에서 얼굴 변형 학습 모델과 연동한 결과 화면으로 왼쪽은 원본 영상이고 오른쪽은 얼굴 변형 모듈을 통해서 출력된 결과 화면이다. 사용자가 변형하고자 하는 이미지를 활용하여 다양한 형태의 얼굴 변형 영상을 생성할 수 있다.



<그림 3> 얼굴 변형 모델 연동 결과 화면

III. 결 론

인터랙티브 미디어 창작 서비스 플랫폼은 웹 서비스 환경으로 구축되어 있고 얼굴 변형 학습 모델은 딥러닝 프레임워크로 구축되어 있다. 이종의 시스템을 개별 시스템으로 구축하지 않고 가상화 기술인 도커 컨테이너를 이용하여 사용자의 선택에 따라 다양한 얼굴 변형 영상 결과를 제공받을 수 있다. 향후 다른 형태의 딥러닝 프레임워크의 얼굴 변형 모듈과 병행하여 사용할 수 있을 것으로 예상된다.

또한 다중사용자 환경에서 다수의 요청을 병렬처리하기 위해서 딥러닝 API 서버의 스케일을 조정하여 보다 효율적인 자원사용 및 빠른 응답 속도를 기대할 수 있다.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [20AR1100, Development of programmable interactive media creation service platform based on open scenario].

참 고 문 헌

- [1] D.E. King, "Dlib-ml: A Machine Learning Toolkit," *Journal of Machine Learning Research*, Vol. 10, Issue 7, pp. 1755-1758, 2009.
- [2] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with An Ensemble of Regression Trees," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867-1874, 2014.
- [3] Z. Cao, T. Simon, S.E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *arXiv. eprint arXiv:1812.08008*, 2018.
- [4] Deepfakes-faceswap, GitHub, <https://github.com/deepfakes/faceswap> (accessed April, 09, 2019).
- [5] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair Bengio, et al., "Generative Adversarial Nets," *NIPS'14 Proceedings of the 27th International Conference on Neural Information Processing Systems*. Vol. 2. pp. 2672-2680, 2014.
- [6] J.Y. Zhu, T. Park, P. Isola, and A.A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," *Proceedings of IEEE International Conference on Computer Vision*, pp. 2242-2251, 2017.
- [7] Merkel D. Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux J.* 2014;2014(239).
- [8] Zakharov, Egor, et al. "Few-Shot Adversarial Learning of Realistic Neural Talking Head Models.". *arXiv preprint arXiv:1905.08233* (2019).
- [9] Nagrani, A., Chung, J. S., & Zisserman, A. (2017). Voxceleb: a large-scale speaker identification dataset. *arXiv preprint arXiv:1706.08612*.