

Image to Image Translation 기반의 안저 의료영상의 병변 위변조에 관한 연구

박기훈, 김용석*

건양대학교, *건양대학교

ghpark86@gmail.com, *YongSuk@konyang.ac.kr

A Study on the Disease modulation of Retinal Medical Image Data based on Image to Image Translation

Park Gi Hun, Kim Yong Suk*

konyang Univ., *konyang Univ.

요 약

현재 'Image-to-Image Translation'을 기반으로 한 이미지 변환 기술이 빠르게 발전하고 있지만 그에 따라 많은 문제가 발생하고 있다. 특히 딥페이크는 가짜뉴스 또는 동영상으로 많은 피해를 입힐 수 있기 때문에 이를 분류하기 위한 연구가 진행되고 있다. 해당 문제는 다른 분야에서도 발생 할 수 있으며, 의료 영상 분야에서도 발생할 수 있다. 의료 영상 분야에서 인공지능에 대한 관심이 높아지면서 인공지능을 이용한 병변 결정에 대한 연구가 많이 진행됐지만 이런 가짜 영상이 판별될 수 있을지는 확인되지 않았다. 따라서 본 연구에서는 딥페이크(Deepfake)에서 주로 사용되었던 Image-to-Image Translation을 사용하여 안저 의료 영상 데이터의 병변 추가 및 제거 연구를 실시한다. 그런 다음 기존에 녹내장을 판별하는 데 사용되었던 인공지능망을 이용하여 조작한 영상을 검증하고 기존의 인공지능망의 정확도를 확인한다.

I. 서 론

현재 딥러닝에 관련된 기술이 빠른 발전을 이루고 있으며 그 중 image-to-image translation은 GAN(Generative Adversarial Nets)을 이용하여 입력되는 이미지를 바탕으로 전혀 새로운 이미지를 만들거나, 특성 추가(Attributing), 변화하여 이미지를 바꾸는 분야이다. image-to-image translation은 손상된 이미지 복원, 사진의 이미지화(e.g. 위성사진을 바탕으로 지도 생성), Deep Learning에서 부족한 학습데이터를 보충하는 등 여러 긍정적인 방법으로 사용될 수 있다.

하지만 그와 반대로 부정적인 측면으로도 사용될 수 있다. 최근 image-to-image translation기반의 기술은 사람의 얼굴을 학습하여 새로운 사람의 얼굴을 생성하거나 특징 추가, 교체, 표정변화 등을 할 수 있게 되었다. 이 기술은 이미지뿐만 아니라 영상에도 적용이 가능했으며, 이는 가짜 뉴스 및 가짜 동영상 등 부정적인 결과를 야기하였다. 이러한 문제는 단순히 사람의 얼굴에서 끝나지 않고, 미래에는 여러 분야에서도 발생 할 수 있는 일이다. 특히 의료분야에서도 이와 같은 일은 심각한 피해를 초래 할 수 있다. 예를 들어 사람의 흉체 이미지를 학습하여 흉체인식 보안을 뚫을 수도 있으며, 어떠한 사람의 병변 여부를 임의로 변경할 수도 있다. 따라서 본 연구에서는 Image-to-Image Translation 기반의 의료 영상 데이터의 병변 위변조 연구를 진행하고 이를 기존에 사용된 병변 분류 딥러닝 모델을 통해 검증한다[1]. image-to-image translation 기술은 Deepfake와 관련되어 많은 연구가 진행된 상태이며[2], 이 중 CycleGAN을 활용하여 본 연구를 진행한다. cycleGAN은 데이터(xi)와 레이블(yi)가 1:1 매칭이 아니라 데이터(Xn)그룹과 레이블(Ym)그룹의 n:m 매칭으로도 학습할 수 있게 구성된 GAN기법이다. 기존의 딥러닝에서는 데이터 xi가 있으면 이 데이터가 학습되어 얻어지는 레이블 yi가 1:1로 매칭되어야하기

때문에 데이터 수집에 제약이 있었다. 하지만 CycleGAN은 이러한 문제점을 해결하기 위해 데이터 Xn 전체와 레이블 Ym 전체를 n:m으로 매칭하여 학습이 가능하도록 구성하였다[3][4]. 본 연구에서는 정상 안저 의료 영상 데이터와 녹내장 안저 의료 영상 데이터를 사용하여 연구를 진행하며, 안저 영상은 사람에 따라 모두 다르기 때문에 CycleGAN과 같은 n:m 매칭이 가능한 모델을 사용하여 연구를 진행한다.

II. 본론

본 연구의 전체 구조는 그림 1과 같다. 정상 안저 영상 데이터(L1:Level1,Normal)과 녹내장 안저 영상 데이터(L2:Level2,Glaucoma)를 검증에 위한 기존 모델 InceptionV3과 변조를 위한 CycleGAN에 각각 학습 시킨다. 그리고 CycleGAN으로부터 생성된 위조된 안저 의료 영상 데이터를 검증 모델로 분류하며, 위조 전의 값과 비교하여 정확도를 확인한다.

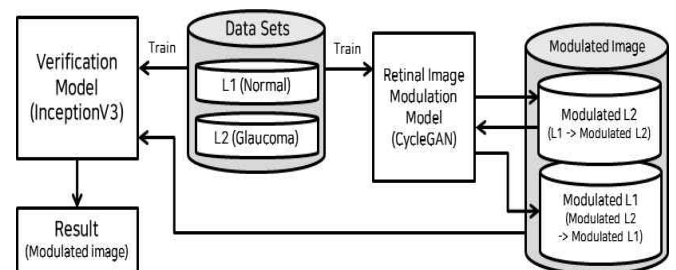


그림 1 연구 진행 구조도

안저 의료 영상 데이터의 위변조에 사용되는 CycleGAN은 그림 2와 같이 구성하였다. CycleGAN은 L1을 입력받아 위조된 L2를 생성한다. 그리고 위조된 L2에 대한 검증 모델의 결과를 확인하고, 위조된 L2를 다시 CycleGAN에 입력하여 위조된 L1을 생성한 뒤 동일하게 검증모델로 결과를 확인한다.

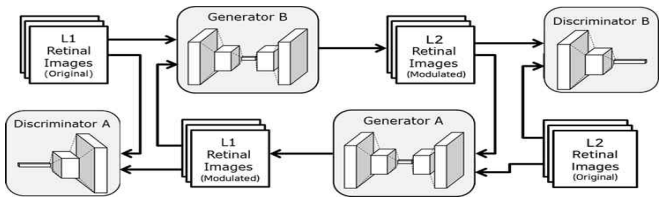
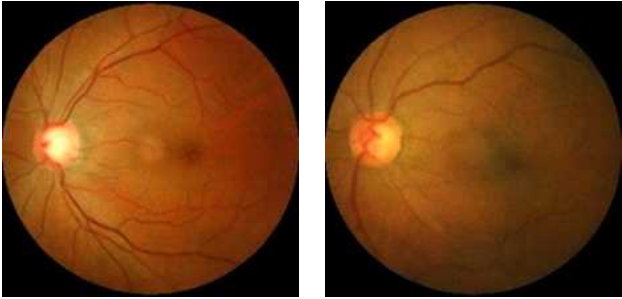
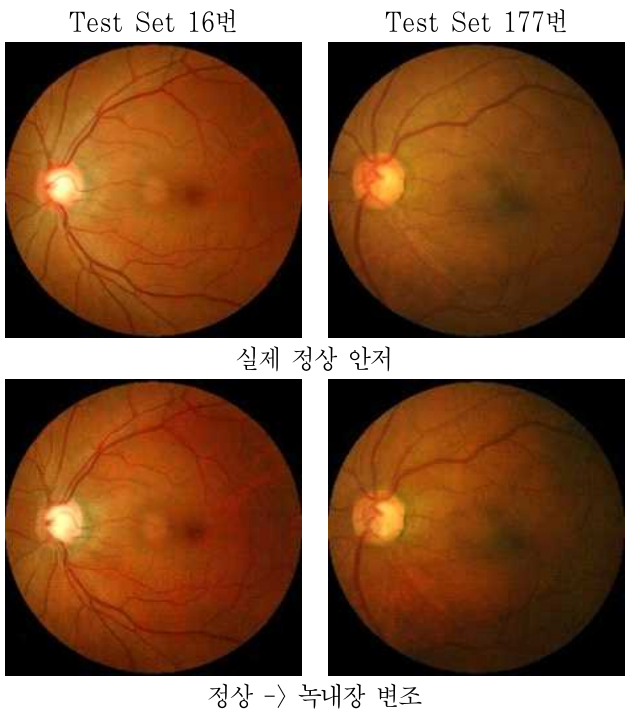


그림 2 CycleGAN 구조

실제 안저 의료 영상 데이터 및 위조된 안저 의료 영상 데이터에 대한 검증모델의 정확도는 표 1과 같으며 L1과 L2 각각 200개씩 총 400개의 데이터로 검증을 진행했다. Test Set L1에 대한 검증 모델의 정확도는 92%이며, Test Set L1으로 위조된 데이터 ML2(Modulated L2)를 검증 모델로 예측한 결과 76%를 L2로 예측하였으며, L1은 기존 대비 68%감소한 24%의 결과를 얻었다. 또한 위조된 ML2를 다시 ML1으로 변환했을 때 L1에 대해 91%의 정확도를 얻었다.

표 1 Test Set을 이용한 검증 모델 정확도 확인

Test Set Prediction		Real Data		Modulated Data	
		L1	L2	ML1	ML2
Inception V3 Prediction	L1	0.92	0.10	0.91	0.24
	L2	0.08	0.90	0.10	0.76



정상-> 녹내장-> 정상 변조

그림 3 임의의 Test Set 2개에 대한 변조 결과 이미지

III. 결론

본 연구에서는 image-to-image translation 기술에 대한 문제점이 의료 분야에서도 발생 할 수 있을 가능성에 대해서 주목하여 실제로 image-to-image translation 기술이 의료분야에서도 적용이 가능한지에 대해 확인하기 위한 연구를 진행하였다. 결과, CycleGAN을 이용하여 정상 안저 의료 영상 데이터를 녹내장 안저 의료 영상 데이터로 변조했을 때, 병변 여부를 분류하기 위해 정상적인 데이터로 학습을 진행한 검증모델(InceptionV3)의 정확도는 기존 92%에서 24%로 약 70%가량 감소된 것을 확인하였다. 또한, 위조된 녹내장 안저 의료 영상 데이터를 다시 정상 안저 의료 영상 데이터로 변환 하였때 91%의 정확도를 보인 것으로 보아 검증 모델은 데이터 변조에 대해서 예측 못 하는 것을 알 수 있다. 이를 바탕으로 향후 연구에서는 의료이미지의 위조, 변조를 판단 할 수 있는 연구를 진행 할 계획이다. 추가적으로 GAN에서 사용된 Discriminator 모델을 다른 연구에서 진행된 안저 병변 예측 딥러닝 모델을 사용하여 Generator의 성능 변화에 대한 연구를 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부(미래창조과학부)의 재원으로 한국연구재단 - 현장맞춤형 이공계 인재양성지원사업의 지원을 받아 수행된 연구임 (No. 2020082500).

참 고 문 헌

[1] GULSHAN, Varun, et al. "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs.", *Jama*, 2016, vol.316.22, p.2402-2410.

[2] Tolosana, Ruben, et al. "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection." *arXiv preprint arXiv:2001.00179* 2020.

[3] ZHU, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks.", *Proceedings of the IEEE international conference on computer vision*. 2017, p.2223-2232.

[4] CHOI, Yunje, et al. "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation.", *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, p.8789-8797.