

Sparse CNN을 이용한 위조 안저영상 포렌식 모델에 관한 연구

송호중, 한주혁, 김웅식*

건양대학교, *건양대학교

20806608@konyang.ac.kr, 21856503@konyang.ac.kr, *wskim@konyang.ac.kr

A Study on the Forensic Model of Forged Fundus Image Using Sparse CNN

Song Ho-Jung, Han Ju-Hyuck, Kim Woong-Sik*

Konyang Univ., *Konyang Univ.

요약

본 논문은 최근 대두되는 의료이미지를 활용한 이미지 재생성 기술의 악용을 방지하기 위한 연구로 의료인들이 구분하기 어려운 정도의 재생성 이미지를 인공지능 모델로 구분해냈다. 우리는 의료이미지 중, 안저 이미지를 기반으로 연구했다. 데이터는 3개의 Kaggle 안저 이미지 데이터 셋을 사용했으며, 이를 전처리하여 위변조 데이터 셋을 구성하기 위한 위변조 모델에 입력으로 사용했다. 위변조 모델은 U-Net과 Cycle GAN 모델을 사용하여 구성했으며, 생성된 데이터는 3명의 안과전문의와 2명의 일반의들로부터 검증을 받았다. 본 연구에서는 4개의 병변(정상, 녹내장, 당뇨병 망막병증, 황반변성)을 대상으로 원본 안저 이미지와 조작된 안저 이미지를 Sparse CNN에 기반으로 하는 포렌식 모델에 입력하여 위변조 유무를 판단한다. 본 논문의 모델 성능은 동일한 테스트 데이터를 5명의 의료진과 포렌식 모델이 그룹별 오피서 테스트로 진행했다. 이는 포렌식 모델의 성능을 의료진과 비교하기 위한 방법이며, 그 결과 포렌식 모델의 성능이 전문의 그룹의 테스트 결과보다 전체적으로 약 10% 더 높은 검출률을 보였다. 따라서 포렌식 모델은 환자의 정보, 임상실현, 사회보험제도, 등 의료 데이터가 필요한 여러 영역에서 위변조 이미지로 인해 발생하는 문제를 사전에 검출 할 수 있다.

I. 서론

최근 인공지능 기술은 이미지 분야에서 많은 발전을 이루고 있다. 특히 이미지 재생성 기술은 실재와 구분하기 어려울 정도의 유사한 이미지를 생성할 수 있다. 이는 뛰어난 기술적 발전이지만, 악용 시에 큰 피해가 예상된다. 특히 의료분야에서는 조작된 이미지가 환자의 생명에 직접적인 영향을 끼칠 수 있다. 또한 임상연구 조작, 사회보험 조작과 같이 사회적인 2차 피해를 발생시킬 수 있어, 그 피해가 치명적이다.[1] 따라서 본 논문에서는 여러 의료이미지 중, 안저 이미지를 대상으로 위조된 이미지를 검출해 낼 수 있는 인공지능 모델을 연구했다.

II. 본론

본 연구에서는 인공지능 기반의 안저 이미지 포렌식 모델을 구성하기 위해, 실제 데이터와 이를 기반으로 조작된 위변조 데이터를 Kaggle에서 수집 및 위변조 하였다. 본 연구의 데이터 셋은 Kaggle의 Shanggong Medical Technology Co, Ltd.에서 수집한 Ocular Disease Recognition 데이터 셋을 사용했다.[2] 이는 5,000명의 환자를 대상으로 수집한 안저 이미지 데이터로 이미지 정보 및 진단 Label이 포함된다. 본 연구에서는 데이터 셋에서 수집된 약 6,000개(환자 1인당, 중복수집) 데이터에서 우측 안구 이미지만을 사용했으며, 이미지 품질을 고려하여 데이터를 분류하여 진행했다. 본 연구에 사용된 Shanggong 데이터 셋은 총 356개의 이미지이며, 정상(Normal) 114개, 녹내장(Glaucoma) 67개, 당뇨병 망막병증(Diabetic Retinopathy) 78개, 황반변성(Macular Degeneration) 97개로 구성된다. 또한 다양한 소스와 환경을 포괄하는 포렌식 모델을 구성하기 위하여, Kaggle의 MESSIDOR-2 DR Grades와 Glaucoma Detection 데이터 셋을 함께 사용했다.[3][4] 이 데이터 셋은 당뇨병 망막병증과 녹내장의 안저 이미지와 진단 Label을 포함한다.

본 연구에서는 데이터 위변조를 위하여, Cycle GAN과 U-Net을 사용하여 위변조 모델을 구성했다. 위변조 모델에 입력되는 데이터는 앞서 이기한 데이터 셋에 이미지 크기 조정 및 전처리 과정을 진행하여 사용했다. 전처리 과정은 원본 데이터에 Sharp Filter를 적용했으며, 필터는 Gaussian Filter로 구성했다. 이는 원본 이미지의 크기를 축소하는 과정에서 발생하는 혈관 및 병변의 뭉개짐(Blur) 현상을 극복하기 위하여 사용했다. 또한 이미지의 크기 축소는 데이터 셋마다 상이한 크기를 맞추고 위변조 모델의 효율성을 증가시키는 역할을 한다.

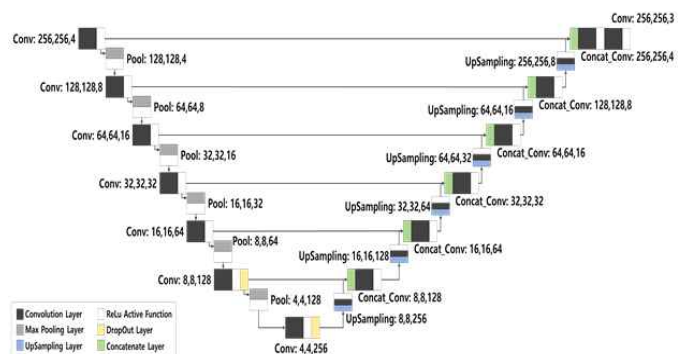


그림 1. 위변조 모델의 U-Net 구성

그림 1은 위변조 모델 중 U-Net 형태로 구성된 모델의 구성도이다. 본 모델은 원본 데이터를 위변조 시키기 위한 모델이며, CNN(Convolutional Neural Network)를 중심으로 구성되어 있다. 모델의 입력은 전처리된 데이터를 랜덤 샘플링하여 사용했으며, 입력의 채널은 4채널로 기본적인 3채널(RGB)과 혈관 Segmentation 채널을 흑백으로 구성했다.

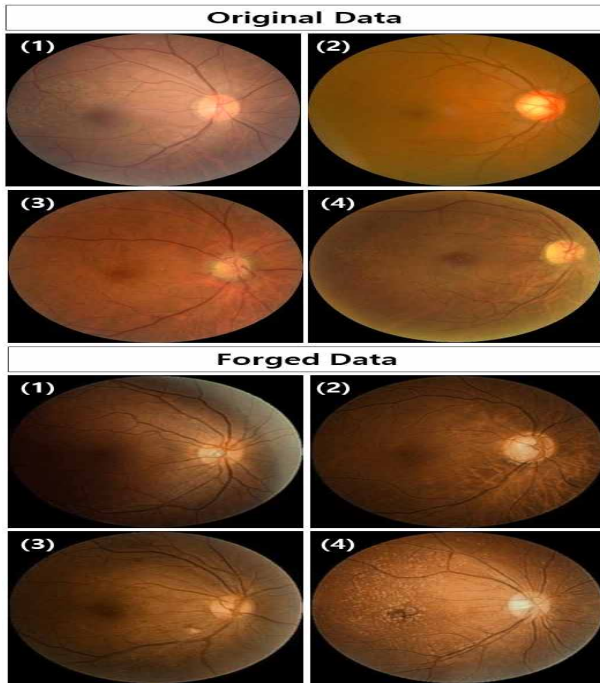


그림 2. 원본 및 위변조 이미지 비교

그림 2는 상단의 원본 이미지와 하단의 위변조 이미지를 비교한 그림이다. 위변조 이미지는 상단의 위조 이미지를 기반으로 만들어졌으며, 각 번호는 (1) 정상, (2) 녹내장, (3) 당뇨병성 망막병증, (4) 황반변성이다. 본 연구에서는 위변조된 데이터와 실제 데이터의 비교를 위하여 위변조 데이터를 생성했으며, 생성된 데이터를 검증하기 위해 안과전문의 3명(Specialist Group)과 일반의 2명(Doctor Group)에게 검증을 요청했다. 또한 포렌식 모델과의 성능비교를 진행했다.

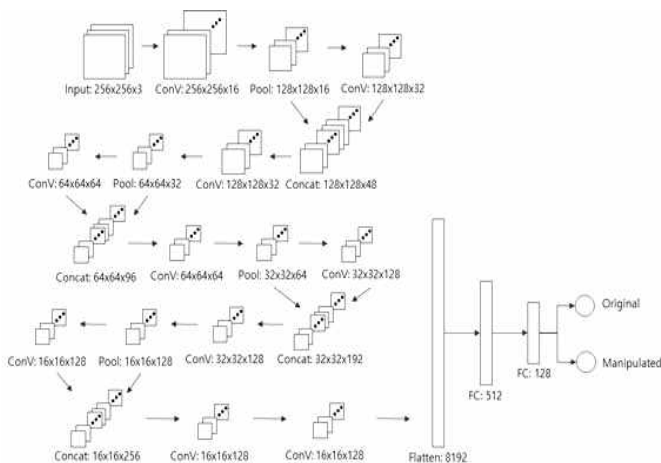


그림 3. 포렌식 모델 구성

그림 3은 실제 이미지와 위변조 이미지를 판단하기 위한 인공지능 모델이며, 모델 구성은 Sparse CNN의 형태로 구성했다.[5] 이는 최초 입력 데이터의 특징을 추출한 Feature Map이 CNN이 계산됨에 따라서 초기 값에 대한 가중치를 계산하지 못하는 문제를 해결하기 위해 사용했다. 또한 포렌식 모델은 2진 분류로 원본과 위변조를 판단하기 위해서, 병변별로 동일한 모델이 학습, 테스트 되었다.

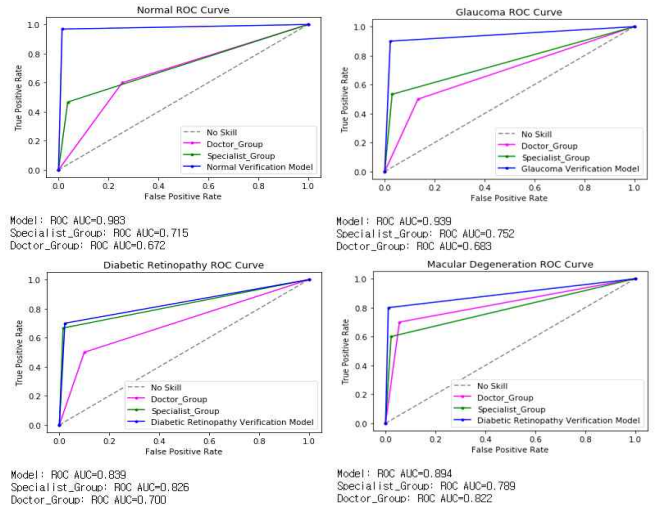


그림 4. 포렌식 모델 및 검증 그룹의 성능비교

본 연구의 결과는 그림 4와 같이 나타나며, 이는 본 포렌식 모델이 안과 전문의 및 일반의와 비교했을 때, 원본과 위변조된 이미지에 대한 판단 능력이 약 10.42% 더 높았다.

III. 결론

본 논문에서는 안저 이미지를 대상으로 위변조된 이미지와 실제 이미지를 구분해낼 수 있는 포렌식 모델을 연구했다. 재생성 기술로 조작된 위변조 이미지는 의료현장의 전문의조차 구분하기 어려운 수준으로 재생성되어 환자의 정보, 임상실험, 사회보험제도, 등 여러 의료적, 사회적 문제를 발생시킬 수 있다. 따라서 본 논문의 Sparse CNN을 바탕으로한 포렌식 모델은 이러한 문제를 사전에 방지할 수 있는 기회를 제공한다. 향후연구로는 다양한 의료 이미지를 활용하여, 포렌식 모델을 연구할 예정이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 ‘바이오융복합기술 전문인력 양성사업’의 지원을 받아 수행된 연구임(No. P0017805).

참 고 문 헌

- [1] Kim, H-J., Jung, D-C., and Choi, B-W. "Exploiting the vulnerability of deep learning-based artificial intelligence models in medical imaging: adversarial attacks.," Journal of the Korean Society of Radiology, pp. 259-273, 2019.
- [2] Kaggle, "Ocular Disease Recognition: Right and left eye fundus photographs of 5000 patients.," (<https://www.kaggle.com/andrewmvd/ocular-disease-recognition-dir5k>).
- [3] Kaggle, "MESSIDOR-2 DR Grades: Adjudicated DR Severity, DME, and Gradability for the MESSIDOR-2 fundus dataset.," (<https://www.kaggle.com/google-brain/messidor2-dr-grades>).
- [4] Kaggle, "Glaucoma Detection: OCT Scans," (<https://www.kaggle.com/sshikamaru/glaucoma-detection>).
- [5] Liu, B., Wang, M., Foroosh, H., Tappen, Marshall., and Pensky, M. "Sparse convolutional neural networks.," Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.