

PGGAN synthetic data를 활용한 Class간 데이터분포의 불균형 완화가 X-ray 질병 진단 정확도에 미치는 영향 연구

정행복*, 최새연*, 김현수*

A study on the impact of relieving the imbalance in data distribution between classes using PGGAN synthetic medical X-ray data on X-ray disease diagnostic classification accuracy.

Haengbok Chung*, SaeYoun Choi*, and Hieonn Kim*

요 약

의료 영역의 인공지능 진단 모델 구축 시, 모델 학습을 위해 균등하고 풍부한 X-ray데이터를 얻는 것은 여전히 어려운 과제이다. 하지만 X-ray데이터의 특성상 세밀한 차이가 중요하기 때문에 GAN을 이용한 데이터가 모델의 성능에 오히려 악영향을 줄 것이라는 견해도 존재한다. PGGAN을 이용한 합성 이미지를 생성하고, 이를 통해 Class Data Imbalance를 완화한 뒤 이러한 합성 데이터가 질병 Classification 모델의 정확도에 어떤 영향을 미치는지 연구하고자 하였다. 그 결과 증강된 이미지를 사용하지 않았을 경우 Test Binary Accuracy가 96.9%인 반면 증강된 이미지를 사용하여 질병 Classification을 수행한 결과 97.1%로 분류 정확도가 향상되었다. 또한 모든 질병에서 AUROC가 상승하여 전체 질병의 평균 AUROC의 경우 0.7848에서 0.8412로 크게 개선되었다.

Abstract

When building an artificial intelligence diagnostic model in the medical field, obtaining even and abundant X-ray data for model learning is still a challenging task. However, since detailed subtle differences are important due to the nature of X-ray data, there is also an opinion that data using GAN will adversely affect the performance of the model. Therefore, we tried to create a composite image using PGGAN, alleviate Class Imbalance through this, and study how Class Imbalance mitigation through such composite data affects the accuracy of the disease Classification model. As a result, the test binary Accuracy was 96.9% when the augmented image was not used, while the classification accuracy was improved to 97.1% as a result of performing disease classification using the augmented image. In addition, it was confirmed that AUROC increased in all diseases, significantly improving from 0.7848 to 0.8412 in the average AUROC of all diseases.

Key words

Synthetic data, X-ray, Medical image diagnosis, GAN, PGGAN

* 이화여자대학교 소프트웨어학부 컴퓨터공학과, rhtr9088@ewhain.net, 0613sy@ewhain.net, hieonn@ewha.ac.kr

I. 서 론

의료 영역의 인공지능 진단 모델 구축 시, 모델 학습을 위해 균등하고 풍부한 X-ray데이터를 얻는 것은 어려운 과제이다. 다시 말해 의료데이터의 경우 데이터셋 내에서 Class간의 발병율에 따라 불균형이 필연적으로 발생하게 된다. 따라서 의료 이미지를 이용해서 분류모델을 구축하는 경우에는 이와 같은 문제를 해결하기 위해 모델 학습 시 개수가 적은 class의 data를 over-sampling하는 방법을 주로 사용해왔다[1]. 하지만 이미지 분류 성능 향상을 위해 GAN[2]으로 증강된 이미지를 사용하는 것 또한 분류 성능을 높이는 잘 알려진 방법 중 하나이다 [3]. 의료 이미지를 분류하여 환자의 질병을 진단하고자 하는 시도에도 이러한 방법을 적용하기 위해 이전부터 GAN을 이용하여 의료 이미지를 생성하려는 시도가 존재해 왔다[4]. 물론 의료데이터의 경우 작은 픽셀 단위의 차이도 큰 의미를 내포할 수 있기 때문에 GAN을 이용해서 생성한 데이터의 실용성에 대한 의문이 제기되어왔고 따라서 GAN으로 생성한 의료 이미지 데이터의 현실성, 즉 사용가능성을 증명하고자 하는 시도 또한 존재해왔다[5]. 이러한 사실들을 바탕으로 본 논문에서는 PGGAN을 이용하여 X-ray데이터를 생성한 후 원래의 학습 데이터셋에 PGGAN을 이용하여 생성된 X-ray데이터들을 추가하여 Class Imbalance를 완화한 후 이것이 분류AUC에 미치는 영향에 대해 연구하고자 한다.

II. 본 론

2.1 실험 데이터

실험을 위해 사용한 X-ray데이터는 NIH[6] 제공 chest x-ray image 데이터를 이용하였다. 데이터는 Cardiomegaly, Emphysema, Effusion, Hernia, Infiltration, Mass, Nodule, Atelectasis, Pneumothorax, Pleural Thickening, Pneumonia, Fibrosis, Edema, Consolidation의 총 14가지 폐질환으로 구성된다.

증강된 데이터를 Train Data에 추가하기 전 데이터셋의 질병별 분포는 아래와 같다. 각 질병 Class별로 불균등이 심한 것을 확인할 수 있다. 아래의

X-ray데이터 분포의 레이블은 차례대로 Cardiomegaly, Emphysema, Effusion, Hernia, Infiltration, Mass, Nodule, Atelectasis, Pneumothorax, Pleural Thickening, Pneumonia, Fibrosis, Edema, Consolidation이다.

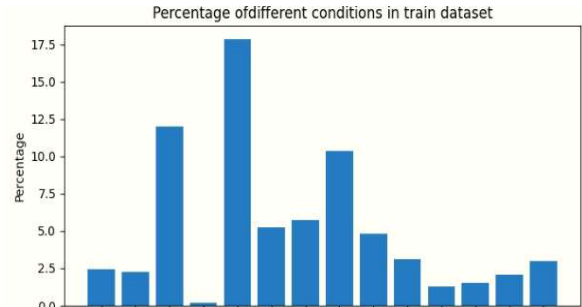


그림 1. 증강 전 클래스별 질병 분포도
Fig. 1. Data distribution before using PGGAN

데이터 증강을 위해 Dataset을 중 적은 비율을 차지하는 6개 질병에 해당하는 X-ray데이터를 PGGAN의 목표 데이터로 사용했다.

표1. 증강을 위해 사용한 데이터셋
Table 1. Experimental data sets for PGGAN

질병	개수
Cardiomegaly	1093
Consolidation	1310
Edema	628
Emphysema	892
Fibrosis	727
Pneumonia	322

2.2 데이터 증강

PGGAN을 이용하여 증강된 데이터를 증강한 후 이를 학습 데이터에 추가하여, Class Imbalance를 완화한 데이터셋의 질병 별 데이터 분포는 아래와 같다.

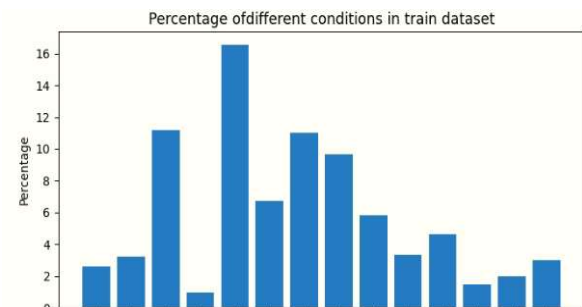


그림 2. 증강 후 클래스별 질병 분포도
Fig. 2. Data distribution after using PGGAN

이때, Class간의 Imbalance를 완전히 제거하지 않은 이유는 합성되지 않은 본래의 데이터 수가 합성으로 생성된 데이터의 과반을 초과하지 않도록 하기 위함이었다.

이렇게 분류된 데이터셋을 재료로 사용, PGGAN을 이용하여 9265개의 훈련용 X-ray이미지를 생성하였다.

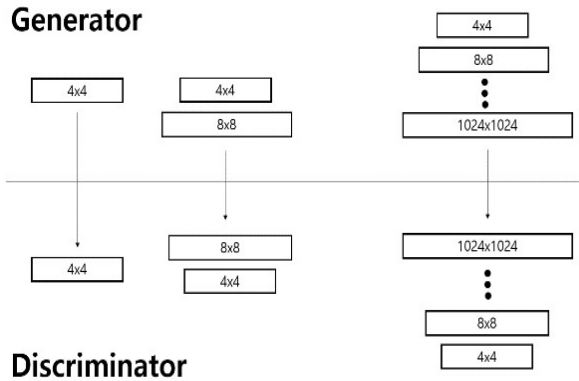


그림 3. PGGAN의 학습 구조
Fig. 3. Learning structure of PGGAN

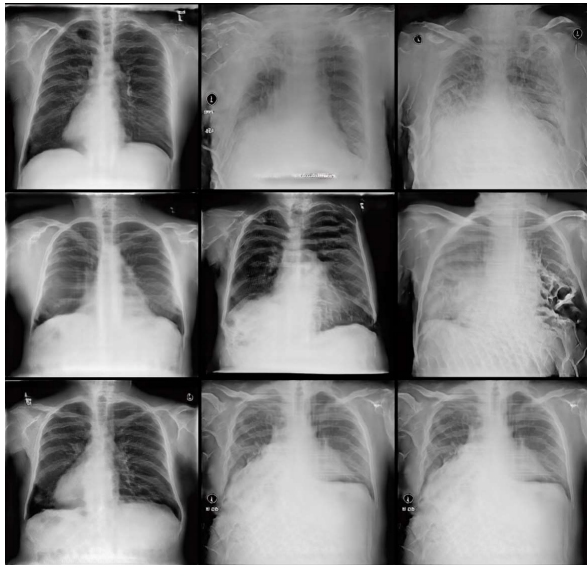


그림 4. PGGAN을 이용해서 생성한 X-ray 데이터
Fig. 4. Generated X-ray data using PGGAN

2.3 생성된 X-ray데이터를 이용한 분류 AUC 측정

PGGAN을 이용하여 생성된 데이터를 분류 모델의 학습데이터셋에 추가한 후 그것이 분류 AUC에

미치는 영향을 연구하는 과정은 다음과 같다. 우선 실험에 사용할 분류 모델을 구축한다.

표 2. 분류 모델 구조

Table 2. Structure of classification model

EfficientNetB1
GlobalAveragePooling2D
Dropout(0.5)
Dense(512)
Dropout(0.5)
Dense(14)

본 논문의 연구목적은 PGGAN을 이용하여 생성한 데이터를 이용하여 Class Imbalance를 완화하는 것이 질병 분류 AUC에 미치는 영향, 즉 PGGAN을 이용하여 생성된 X-ray데이터를 학습데이터에 추가하기 전과 후의 차이를 알아보는 것이므로 생성된 데이터를 학습 데이터에 추가하기 전과 후의 Backbone모델과 학습 방식을 변경하지 않는 선에서 실험을 진행하였다. 분류 모델을 구축한 후에는 각 질병 별 data의 분포를 확인하고 상대적으로 수가 적은 질병의 데이터를 따로 분류한다. 그런 다음 각 데이터 세트를 증강하고 이를 학습데이터셋에 추가하여 모델을 학습한다. 그 다음, 합성데이터가 포함되지 않은 테스트 데이터를 이용하여 모델의 최종 성능을 평가한다.증강된 데이터를 분류하는 경우에 Effusion, Infiltration, Mass, Nodule, Atelectasis, 로 판별되는 데이터는 학습 데이터셋에 추가하지 않았는데 위의 질병의 경우 이미 데이터 분포에서 많은 부분을 차지하고 있으므로 이 데이터를 추가하는 경우에 Class Imbalance가 오히려 심화될 수 있기 때문이었다.

2.4 실험결과

증강된 데이터를 학습에 활용하기 전과 후의 분류 모델의 Test Accuracy는 아래와 같다.

표 3. 증강 전후의 질병 분류 정확도

Table 3. Test Accuracy before and after using PGGAN data

	Test Binary Accuracy
증강전	96.9%
증강후	97.1%

증강된 데이터를 학습에 활용하기 전과 후 각 질병별 AUC의 변화는 다음과 같다.

표 4. 증강 전후의 질병 AUC

Table 4. AUC before and after using PGGAN data

Disease	증강전 AUC	증강후 AUC
Cardiomegaly	0.838	0.9
Emphysema	0.87	0.92
Effusion	0.867	0.895
Hernia	0.65	0.877
Infiltration	0.702	0.723
Mass	0.804	0.861
Nodule	0.691	0.774
Atelectasis	0.784	0.826
Pneumothorax	0.851	0.903
Pleural Thickening	0.757	0.812
Pneumonia	0.691	0.736
Fibrosis	0.764	0.828
Edema	0.899	0.897
Consolidation	0.819	0.825
Average	0.7848	0.8412

III. 결 론

X-ray 데이터를 이용한 질병 진단에 있어 필연적으로 발생하는 데이터 간의 Class불균형을 Synthetic Medical X-ray 데이터를 이용해서 해소하는 것이 모델의 Test Accuracy를 높이는 것을 확인할 수 있었다. 또한 AUROC의 경우 전체 질병에서 개선되어 증강된 데이터를 이용하여 학습하는 것이 전체 질병의 진단에 기여한 것으로 확인하였다. 모든 질병의 AUROC의 평균 값을 살펴보면 0.7848에서 0.8412로 큰 폭으로 향상되었는데 이는 증강된 데이터가 X-ray를 통한 질병 Classification의 정확도에 큰 기여를 했음을 확인할 수 있는 결과이다. 이러한 결과를 통해 PGGAN을 이용하여 X-ray데이터를 생성한 후 이를 분류모델의 학습데이터에 추가하여 Class Imbalance를 완화하는 것이 질병 분류 AUC에 긍정적인 영향을 미친다고 볼 수 있다.

본 논문에서는 본래의 훈련 데이터의 특징을 훼손하기 않기 위해 총 훈련 데이터셋에서 생성된 X-ray데이터의 비율이 실제 X-ray데이터보다 적은 비율을 차지하도록 하였는데 후속 연구를 통하여 훈련 데이터 셋에서 생성된 X-ray데이터의 비율을

증가시켜 Class Imbalance를 완전히 제거한 경우의 정확도 또한 연구해 나갈 계획이다.

참 고 문 헌

- [1] Li, Der-Chiang, Chiao-Wen Liu, and Susan C. Hu. "A learning method for the class imbalance problem with medical data sets." Computers in biology and medicine 40.5, pp. 509-518, 2010.
- [2] Creswell, Antonia, et al. "Generative adversarial networks: An overview." IEEE Signal Processing Magazine 35.1, pp. 53-65m 2018.
- [3] Perez, Luis, and Jason Wang. "The effectiveness of data augmentation in image classification using deep learning." arXiv preprint arXiv:1712.04621, 2017.
- [4] Beers, Andrew, et al. "High-resolution medical image synthesis using progressively grown generative adversarial networks." arXiv preprint arXiv:1805.03144, 2018.
- [5] Segal, Bradley, et al. "Evaluating the Clinical Realism of Synthetic Chest X-Rays Generated Using Progressively Growing GANs." SN Computer Science 2.4, pp. 1-17, 2021.
- [6] Kim, Heon-Gyu. "미국의 국립보건연구원." The Science & Technology 2.3, pp. 65-70, 1969.
- [7] Karras, Tero, et al. "Progressive growing of gans for improved quality, stability, and variation." arXiv preprint arXiv:1710.10196 , 2017.