

GAN 기반 의료 데이터 증강을 통한 분류 성능 향상

박길한⁰¹ 이재환¹ 이익규¹ 김지수² 허재필²

¹성균관대학교 컴퓨터교육과

²성균관대학교 소프트웨어학과

{ a01152a, jwlee508, dlrb0521, jisu7935, jaepilheo }@skku.edu

Improving Classification results on Medical Dataset

Augmentation using GAN

Gilhan Park⁰¹ Jaehwan Lee¹ Ikgyu Lee¹ Jisoo Kim² Jae-Pil Heo²

¹Department of Computer Education, Sungkyunkwan University

²Department of Computing and Informatics, Sungkyunkwan University

요 약

의료 분야의 특성상 데이터 개방성이 낮고 절대적인 데이터의 수가 부족하다. 또한 의료 데이터는 특유의 민감성 때문에 반드시 연구에 필요한 최소 수준의 데이터 요청 및 폐쇄된 환경에서만 활용이 가능해 데이터 접근이 매우 제한적이다. 한정된 의료 데이터의 특성을 해결하기 위한 연구가 다양하게 진행되고 있다. 본 논문에서는 피부암 데이터를 분류하기 위한 모델의 학습에 필요한 의료 데이터의 수를 늘리기 위한 방법으로 GAN을 통한 Data Augmentation을 위해서 소량의 데이터를 학습하고 이를 통해 데이터의 수를 증강한다. 또한 GAN의 모델 별로 실제 이미지와 가짜 이미지의 비율이 어느정도 수준에서 기준이 되는 분류 모델의 정확도에 가까워지는지 실험해본다.

1. 서 론

인공지능 기술의 발전과 의료영상의 발전에 따라 의료 분야에서 인공지능 기술이 적용되고 있을 뿐 아니라 다양한 연구가 제안되고 있다. 하지만 정확한 진단과 치료를 우선하는 의료 분야에서 인공지능 기술 적용의 확장을 위해서는 뛰어난 성능이 우선 시 되어야 한다. 그러므로 충분한 고품질의 의료데이터를 활용하는 것이 중요해졌다. 하지만 의료 데이터는 특유의 민감성 때문에 반드시 연구에 필요한 최소 수준의 데이터 요청 및 폐쇄된 환경에서만 활용이 가능해 데이터 접근이 매우 제한적이다. 양질의 데이터를 확보하는 데에도 한계가 있다. 질병의 특성 때문에 데이터 자체가 희귀하거나 데이터가 특정 그룹에 편향되어 있어 데이터 불균형 문제가 발생한다. 질환을 앓고 있는 환자가 정상인에 비해 소수이기 때문에, 질병으로 분류(Classification)되는 데이터는 전체 데이터의 극히 일부만을 차지한다. 의료 데이터의 정제와 가공 또한 많은 비용과 전문 인력 투입이 필요하므로 어려움을 겪는다. 의료 데이터는 전문적인 영역으로, Labeling 하기 위해서는 의료 전문가의 협력이 필수적이다. 하지만 전문적인 인력을 Labeling과 같은 단순 작업에 이용하는 것은 산업 구조에서 비효율성을 초래한다.

이러한 의료 데이터의 문제를 해결하기 위해 의료 데이터의 수를 늘리기에 적합한 이미지 생성 모델인 Generative Adversarial Network (GAN)[1, 2, 3]을

탐구하고, 합성 의료 데이터를 생성하여 Data Augmentation을 진행한다.

합성 데이터란 직접 확보하지 못하는 특정 환경에 적용할 수 있는 모든 데이터를 뜻한다. 합성 의료 데이터 생성 기술을 통해 실제와 유사한 가짜 의료 데이터를 생성함으로써 의료 데이터 부족 문제를 해결할 수 있다. 이는 소량의 원본(original) 데이터로 빠르고 저렴하게 합성 의료 데이터를 생성할 수 있다는 장점이 있다. 더불어 데이터 라벨링 작업을 위한 시간과 비용을 절감할 수 있으며 민감 정보를 포함하고 있는 실제 의료 데이터를 대체할 수 있다.

본 연구는 절대적으로 부족한 의료 데이터에 여러 GAN을 적용하여 Data Augmentation을 진행한다. 이를 통해 의료 데이터 분류(Classification) 모델의 성능을 향상시키고자 한다. 실제 이미지와 가짜 이미지의 비율을 변화시켜 어느 수준에서 실제 이미지만을 사용하여 학습한 분류 모델의 정확도에 근접할 수 있는지 실험한다. 더 나아가 의료 데이터와 같은 특성을 가진 유사 분야에 본 연구 방식을 적용하여 문제를 해결할 수 있다.

2. 제안된 방법

2.1 Generative Adversarial Network[1]

Generative Adversarial Network(GAN)[1]은 이미지 생성 모델로, 이미지를 생성하거나, 변환, 합성 및 고해상도에 사용된다. GAN은 이미지를 생성해주는 Generator와 생성한 이미지와 실제 이미지를 판별해주는 Discriminator의 적대적인 상호작용에 의해서 학습이 된다. Generator는 학습을 통해

본 논문은 과학기술정보통신부·경찰청/KIPOT의 폴리 스랩 2.0 사업(No. 210121M06)과 인공지능대학원치 원사업(No. 2019-0-00421)의 지원을 받아 수행된 연구임.

Discriminator가 이미지를 실제인지 생성한 이미지인지 판별할 수 없을 정도로 실제와 유사한 이미지를 생성하도록 학습을 하게 되고, Discriminator는 이미지가 진짜인지 아닌지를 판별할 수 있도록 학습된다. 위와 같은 원리로 GAN은 현실에서 존재할 법한 이미지를 생성해낼 수 있다.

2.2 PGGAN[2]

PGGAN(Progressive Growing of GAN)[2]은 안정적인 고해상도의 이미지를 생성하기 위해 개발된 모델이다. 일반적인 GAN의 경우 Generator가 생성한 고해상도의 이미지가 실제 이미지인지 가짜 이미지인지에 대해서 Discriminator가 구분하기 쉬워지기 때문에 고해상도의 이미지를 만들기 어렵다. 또한 메모리의 한계 때문에 GAN의 훈련 역시 어렵다. PGGAN[2]은 훈련 초기 Generator와 Discriminator를 저해상도 이미지 수준에서 학습을 하고, 점진적으로 Layer를 동시에 추가하였으며 또한 새롭게 추가된 Layer를 smooth하게, fade in하게 넣어주어 해상도를 높여가는 알고리즘을 가진 모델이다.

2.3 Image to Image Translation

Image to Image Translation[3, 4]이란 특정 도메인을 가진 이미지 데이터를 Input으로 주요 특징은 유지한 상태로 원하는 속성을 가지도록 스타일을 변화시키는 생성 모델의 한 분야이다. 이 분야는 GAN의 등장 이후 활발하게 연구가 진행되었으며, 흑백 이미지에 컬러를 입힌다든지, 낮을 밤으로 바꾸거나, 특정 동물을 다른 동물로 바꾸는 등 여러 범위에서 응용되었다. 본 논문에서는 Image to Image Translation에 대표적인 변형 알고리즘인 CycleGAN[3]을 사용한다. CycleGAN은 짝을 이루는 데이터를 요구하던 이전 모델[4]과 다르게 짝을 이루지 않는 데이터를 사용하고, 생성한 데이터를 다시 원복해야 하므로 Generator와 Discriminator가 하나씩 더 추가된다. 본 연구에서 각각 양성종양과 악성종양 데이터의 특징을 학습하여 양성종양 데이터를 악성종양의 특징으로 변환시켜 Data의 불균형을 해결하려고 한다.

3. 실험

3.1 데이터셋 정의

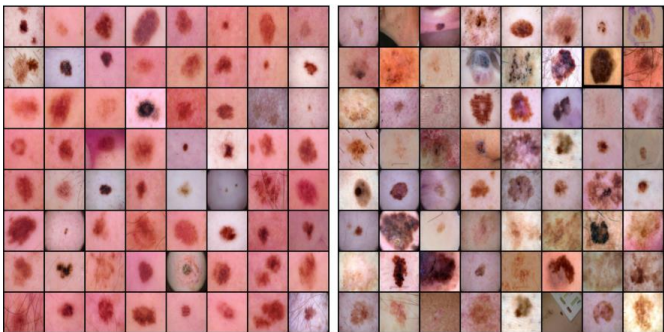


그림 1 양성종양 및 악성종양 데이터

본 연구에서 사용한 데이터셋은 Kaggle에 업로드 되어있는 Skin Cancer Image 데이터로 benign(양성)과 malignant(악성) 종양 class의 이진 분류를 목적으로 사용되는 데이터셋이다. 모두 224 x 224 크기의 image이다. 기준이 되는 분류 모델을 학습하는 benign과 malignant의 image는 각각 1440, 1197개이며, test하는 benign과 malignant의 image는 각각 360, 300개이다.

3.2 시나리오

본 연구에서는 실질적인 활용 방안을 고려하여 소량의 악성종양 데이터셋과 다량의 양성종양 데이터셋을 가지고 있는 상황을 가정한다. 그리고 여러가지 GAN 모델을 통하여 악성종양 데이터셋을 생성하여 Data의 불균형을 해결하며 분류 성능 측정의 기준이 되는 간단한 CNN 모델을 통해 성능을 확인해본다.

총 3가지 종류의 GAN 모델인 DCGAN, PGGAN, CycleGAN을 실험 모델로 선택하였으며, 분류 모델은 기본적인 2-Block의 Pytorch CNN모델을 사용하였다. 그리고 실제 악성종양 이미지의 개수의 수준(50, 100, 200, 500)에 따라 기준이 되는 분류 모델에서 학습되어지는 양성종양 데이터셋의 개수(1197개)에 맞춰 Data Augmentation을 통해 실제 개수인 1197개의 데이터를 맞추어 기존의 성능과 비교한다.

이를 통해 GAN의 모델 종류 별로 Real Image와 Fake Image의 비율이 어느 정도 수준에서 기준이 되는 분류 모델의 정확도에 가까워지는지 실험해 본다.

3.3 실험 결과

실험은 DCGAN[5], PGGAN[2], CycleGAN[3] 세가지의 모델로 기본적인 Data Augmentation을 진행하였고, [그림 2]는 그를 통해 생성된 Fake image 샘플이다. 위에서부터 DCGAN, PGGAN, CycleGAN의 모델을 사용하였으며, 학습에는 악성종양 100장, 200장, 500장을 학습시킨 모델의 Fake image 생성 샘플이다. 분류 모델은 Pytorch 기반의 간단한 CNN 분류 모델로 구성하였으며, 아래 [표 1]은 Data imbalance인 경우(예. 양성종양 1440장, 악성종양 100장) 분류 성능을 나타낸 것이며 [표 2]는 Data imbalance를 해결하기 위해 각각의 GAN 모델로 Fake image를 생성하여 악성종양 이미지를 1197개로 Augmentation하여 분류하였을 때의 성능이다. 빈 칸의 값은 각각 'Accuracy'/'AUC' 값을 나타낸다.

DCGAN[5]의 경우 Original GAN에서 Layer만 CNN 신경망으로 교체된 비교적 오래된 초창기 GAN 모델이고 최근 GAN 모델에 비해 성능이 많이 떨어져 소량의 이미지 데이터로는 데이터의 특징을 잘 찾지 못하며 다소 흐릿한 이미지가 생성된다.

PGGAN은 충분한 이미지 데이터를 가지고 있을 때 고품질의 이미지를 생성한다. 하지만 50장 같이

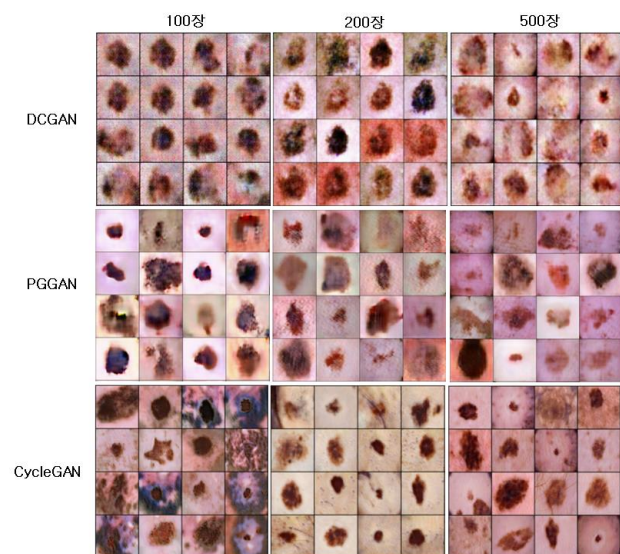


그림 2 제한된 악성종양 데이터 개수로 학습된 GAN 모델의 Fake Image 샘플

소량의 데이터를 학습시키는 경우 사용할 수 있는 이미지를 생성하는 데에는 한계가 존재한다.

CycleGAN은 두 GAN과는 다르게 이미지의 속성을 변화시키므로 종양 부위가 검거나, 다양한 분화상태를 띄는 등 다른 GAN에 비해 악성종양의 구별되는 특징을 잘 잡아내어 생성한다.

표 1 Data Imbalance 상태에서의 분류 성능

모델	악성 종양 이미지의 개수 (양성 종양 1440장)			
	50 Images	100 Images	200 Images	500 Images
Baseline	0.58/0.54	0.58/0.54	0.62/0.59	0.81/0.81

표 2 GAN을 통해 Data Imbalance를 해결한 상태에서의 성능

모델	GAN 학습에 사용된 악성 종양 이미지 개수 (Real 악성 종양 이미지 1197개 분류 성능 0.83/0.83)			
	50 Images	100 Images	200 Images	500 Images
DCGAN	0.67/0.64	0.57/0.53	0.67/0.64	0.82/0.81
PGGAN	0.58/0.54	0.61/0.57	0.73/0.71	0.81/0.80
CycleGAN	0.57/0.53	0.62/0.58	0.68/0.65	0.82/0.82

악성 종양 클래스의 모든 이미지를 사용하여 분류한 모델의 성능이 '0.83/0.83'이라면 [표 2]의 결과를 통해 알 수 있는 것은 50장의 경우 분류 모델이 어느정도 안정적인 성능을 보이기 때문에, 200장의 이미지를 사용하여 GAN을 학습하고 이를 통해 Fake Image를 생성한 실험 환경에서 성능의 향상이 제일 높았다. 특히 PGGAN 같은 경우 기존 Baseline의 Accuracy인 0.58보다 0.15의 정확도가 오른 0.73의 Accuracy를 보인다. 전체적으로 보았을 때 GAN을 사용하여 Data Imbalance를 해결하였을 때 더 좋은 성능을 보임을 알 수 있다.

4. 결론

GAN을 사용하여 Fake Image를 생성하고 이를 통해 데이터 불균형을 해결하려 한 본 연구의 실험 결과를 보았을 때, 학습 데이터가 불균형한 상태에서 분류 모델을 학습한 경우보다는 대부분의 모든 GAN을 통한 실험에서 눈에 띄게 성능이 올라갔다. 하지만 전체 Real 학습 데이터 셋을 사용한 기준 모델에는 미치지 못하는 것을 알 수 있다. 학습에 이용되는 Real Image 데이터를 최소한 500장은 사용해야 기준 모델에 버금가는 성능이 나온다. 성능의 향상을 위해 추후에 Image Augmentation 기법을 적용하였을 때의 결과와 비교해보고, 다른 발전된 최신 GAN을 통한 추가 실험을 진행하여 성능의 변화를 보고자 한다. 또한 단순 이진 분류의 데이터가 아닌 다중 분류의 데이터에서 성능 비교도 추후 진행해보려 한다.

5. 참고문헌

- [1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Advances in Neural Information Processing Systems (NIPS), pages 2672–2680, 2014.
- [2] Karras, Tero, et al. Progressive Growing of GANs for Improved Quality, Stability, and Variation. arXiv preprint arXiv:1710.10196, 2017.
- [3] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.
- [4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [5] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015.