

의료 이미지 데이터 기반 모델 성능 극대화 전략

초록

질병 판별을 위한 이미지 데이터가 부족한 상황에서 데이터를 증강하여 예측 성능을 최적화하는 전략에 대하여 연구한다.

1. 서론

최근, MRI 기반 영상 분석으로 인지장애를 판별해내는 솔루션을 개발하거나 악성 종양을 탐지해내는 등, 의료 진단에서의 인공지능 기술의 적용이 확대되고 있다.

하지만, 의료 데이터는 다른 데이터보다 수집과 전처리 과정에서 많은 시간과 비용이 소모된다. 부족한 데이터 개수는 모델 성능 저하로 이어지기 때문에 생명과 직결된 의료 분야에서 많은 문제를 발생시킬 수 있다. 특히, 정상인의 수가 환자 수보다 훨씬 많기 때문에 환자 데이터가 부족한 것이 매우 큰 문제로 작용한다.

본 연구에서는 해당 문제에 대한 돌파구를 제시하기 위해 망막 데이터를 활용한 연구를 진행한다. 당뇨망막병증은 안저검사를 활용하여 시각적으로 판별가능하기 때문에, 모델의 학습에 이미지 데이터를 활용한다. 이미지에 data augmentation, Semi-Supervised Learning, GAN 등의 기술을 결합한다. Semi-Supervised Learning에서는 transfer learning의 아이디어를 일부 벤치마킹한다. 망막 데이터가 아닌, 구형

데이터와 GAN 이미지를 라벨링 없는 데이터로 간주하여 학습을 수행한다. 이러한 과정을 통해 의료 데이터가 부족한 상황에서, 유의미한 데이터를 생성하고 의료 진단기의 성능을 최적화시킬 수 있는 방법을 모색한다.

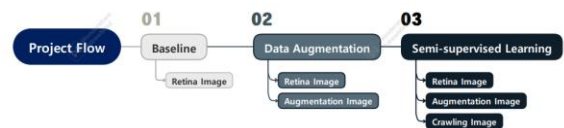
2. 본론

0) 당뇨망막병증

망막의 미세혈관이 손상되는 질병으로, 비증식성 망막병증과 증식성 망막병증으로 구분된다.[1] 비증식성 망막병증은 혈청이 새거나, 혈관이 막히는 현상이며, 증식성 망막병증은 신생혈관이 생기는 현상을 말한다. 안저 검사, 형광 안저혈관 조영, 초음파 검사 등 크게 5가지의 진단 방법이 존재한다.[2] 본 연구에서 활용되는 이미지는 당뇨망막병증 판별에 가장 중요한 안저검사를 통해 추출된 데이터이다.

1) 시스템 구성

[사진1]



본 연구는 [사진1]의 흐름으로 진행된다. Baseline 모델은 VGG11로 구성되며, 부족한 망막 데이터를 보강하기 위해 다양한 조합의 data augmentation을 적용한다. 추가로, Semi-supervised Learning을 진행하며

구형 크롤링 이미지와 GAN 이미지와 같은 가짜 망막 이미지들을 라벨이 없는 데이터로 간주하여 학습에 활용한다.

2) 데이터 전처리

Kaggle의 Diabetic Retinopathy(당뇨망막병증)[3] 데이터를 이진분류에 사용할 수 있도록 전처리를 수행한다.

[표1]

Raw	Count
Mild	2443
Moderate	5292
Proliferate_DR	708
Severe	873
No_DR	25810

[표2]

Preprocessing		Count
Train	DR	2266
	No_DR	2266
Test	DR	566
	No_DR	566

[표1]에서 보이는 것과 같이 데이터는 정상(No_DR) 25000개, 환자(No_DR 외 데이터) 10000개로 이루어져 있으며, 환자의 데이터 개수가 전체 데이터의 30% 정도로 개수가 훨씬 적음을 확인할 수 있다. 환자 데이터는 질병의 진행 정도에 따라 Mild ~ Severe의 4개 class로 구분되며, 이진 분류를 위해 질병 class인 DR class 1개로 병합하였다. 의료 데이터가 부족한 보편적인 상황을 재현하고자, 데이터의 개수가 가장 적은 Proliferate_DR를 기준으로 각각의 데이터를 708개씩 추출하여, [표2]와 같이 병합하였다. 또한, train : test의 비율을 8 : 2로 두어 split하여 훈련과 평가를 위한 데이터를 분리하였다.

3) 모델 구성

본 연구의 Baseline 모델은 VGG11[4]로 구성한다. 망막 데이터가 부족한 상황에서 VGG16[4], VGG19[4] 등 매우 높은 복잡도의 모델은 과대적합이 발생할 수 있기 때문에, 적절한 복잡도를 가진 모델을 구현하였다.

4) Data augmentation

Data augmentation은 데이터에 특정한 규칙으로 변형을 주어 적은 데이터의 개수를 보완하는 기법이다.

본 프로젝트에서는 horizontal, rotation, gaussian blur 등 다양한 augmentation 조합을 시도해 보며 가장 높은 성능을 내는 조합을 추출한다.

5) 크롤링

1) 유사한 데이터 크롤링

원하는 검색어 list를 지정한 후, selenium 크롤링을 수행한다. '구글 이미지' 사이트에서 검색된 이미지들을 사용자가 지정한 개수만큼 저장한다.

2) 다양한 알고리즘 활용한 이미지 전처리 - 코드 설계

Hough Circle Transform을 사용하여 원을 추출한 후, 반지름 기준으로 ROI 영역을 설정하여 원 주변 일정 범위를 Crop한다. 이후 쌍선형보간법을 활용하여 망막 데이터와 같은 해상도(224x224)로 사이즈를 조절한다.

6) GAN

GAN[5] 알고리즘은 생성자와 구별자의 상호경쟁으로 진짜 이미지와 비슷한 가짜 이미지를 만들어 내는 알고리즘이다. GAN, DC GAN[6]으로 생성된 이미지를 Noisy Student 알고리즘에 활용하여 성능 변화를 분석한다.

7) Semi-Supervised Learning

본 연구에서는 Self-Supervised Learning 알고리즘으로 NoisyStudent[7]을 적용하였다. unlabeled image를 활용하여 모델의 성능을 향상시키는 알고리즘인데, 이에 transfer learning의 아이디어를 융합한다. 망막 데이터가 아닌, 육안 상 망막 데이터와 유사한 크

롤링한 구형 이미지와 GAN 이미지에 augmentation을 적용한 후, 이를 라벨링 되지 않은 망막 데이터로 가정하여 실험을 진행하는 것이다.

3. 실험 결과

1) baseline

[표3]

lr	accuracy			loss		
	train	validation	test	train	validation	test
1.00E-02	0.5731	0.5856	0.59	0.676	0.6707	0.666
7.50E-05	0.7671	0.6513	0.66	0.4538	0.6652	0.597
5.00E-05	0.869	0.6716	0.7289	0.3053	0.837	0.5352
2.50E-05	0.9691	0.6642	0.8	0.08794	1.367	0.456
1.00E-05	0.9641	0.659	0.6314	0.1102	1.12	0.6721
1.00E-06	0.691	0.6081	0.6	0.5803	0.6691	0.664

우선, VGG11을 baseline으로 learning rate를 변경하면서 각 loss를 측정하였다. [표3]의 첫 번째 행을 보면, learning rate가 매우 큰 0.01에서는 학습이 제대로 되지 않고 과소적합 되었다. 5번째 행의 learning rate가 1e-5인 경우에서 train accuracy가 매우 크고, loss가 매우 작다. 하지만, validation과의 accuracy가 30% 정도의 큰 차이를 보면 과대적합 되었음을 알 수 있다. 2번째 행에서는, learning rate를 더 크게 조절하여 learning rate가 7.5e-5인 지점에서 local minimum에서 벗어나 train과 validation, test가 적절히 조화를 이루는 것을 확인할 수 있었다.

2) augmentation

[표4]

lr	augmentation	accuracy			loss		
		train	validation	test	train	validation	test
7.50E-05	None	0.9968	0.6511	0.7202	9.13E-03	2.232	0.5676
	Rs RO(RH, RV, RA), N	0.7153	0.6564	0.7731	0.5386	0.5959	0.4919
	Rs RO(RH, RV, RA, RR180), N	0.6419	0.5964	0.6759	0.62	0.657	0.5684
	Rs RO(RH, RV, RA, RR180, GB), N	0.5942	0.5618	0.6027	0.6649	0.6816	0.6654
	Rs, RO(RH, RV, RA, RB, RAF, RP), N	0.6082	0.6252	0.671	0.6534	0.6506	0.6085

Rs, RO(RH, RV, GB), N	0.7392	0.6512	0.7173	0.5136	0.648	0.5472
Rs, RO(RH, RV, RA, GB), N	0.6455	0.6316	0.6574	0.6188	0.6471	0.6167
Rs RO(RH, RV, RA), N + GAN284	0.6858	0.6427	0.6021	0.5705	0.626	0.6639

Rs : Resize RO : RandomOrder RH : RandomHorizontal RV : RandomVertical

RA : RandomAutoContrast

GB : GaussianBlur RAF : RandomAffine

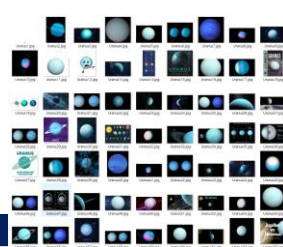
RP : RandomPerspective N : Normalization

[표4]에서 두 번째 행의 모델은 augmentation을 수행하지 않은 것에 비해 전체적으로 높은 성능을 보였다. 또한, training과 validation 사이의 격차를 크게 줄여 regularization과 같은 역할을 수행하는 것을 확인할 수 있었다. 그러나, 3가지 이상의 augmentation을 수행하거나, GaussianBlur transform을 수행한 다른 모델의 경우 성능이 저하되었다.

가장 높은 성능을 보여주었던 augmentation 기법에 GAN에서 생성된 이미지 284개를 DR 데이터셋에 추가하여 학습을 진행한 결과 모든 accuracy, loss 지표가 악화됨을 확인할 수 있었으며, 이는 augmentation을 지나치게 많이 수행하는 경우 모델의 성능에 악영향을 끼칠 수 있음을 보여준다. 해상도가 낮은 일반 GAN을 사용한 것은 Gaussian Blur를 사용한 것과 유사한 효과를 발휘하기 때문임을 추측해볼 수 있다.

3) 구형 데이터 크롤링 및 전처리

[사진2] 구형 이미지 크롤링



[사진3] 크롤링 이미지 전처리

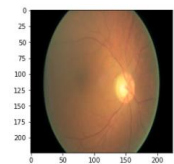


[사진2]의 불규칙한 크롤링 데이터에서 Hough Circle Transform을 사용하여 구형 데이터를 추출한 후 주위 영역 일부와 함께 저장하여 망막 데이터와 유사한 형태를 띠는 [사진3] 이미지를 추출했다. 크롤링된 데이터 4400개 중, 망막 데이터와 유사

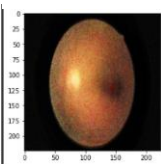
한 형태의 이미지 2000개를 선별하여 준지도 학습에 활용하였다.

4) GAN & DC GAN

[사진4] 원본

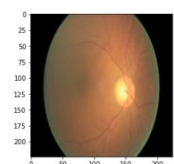


[사진5] GAN



[사진4]은 원본, [사진5]은 GAN으로 만들어낸 이미지이다. GAN 모델은 선형 layer를 사용하였기 때문에, DC GAN과 원본보다 낮은 해상도의 이미지가 출력되었다. GAN으로 만들어낸 망막 이미지를 사용했을 때 모델 성능이 저하된 것은 당뇨병성 질환의 특징에 있다. 본문에서 제시한 것과 같이, 안저검사를 통해 혈관의 상태와 황반을 확인하는데, GAN이 도출한 이미지는 혈관이 생성되지 않았으며, Blur를 한 것과 같은 형태의 이미지가 형성되어 망막의 상태를 판별하는 것이 불가능했다.

[사진6] 원본



[사진7] DC GAN



이미지의 해상도를 높이기 위해 CNN[8]을 사용하여 DC GAN을 구성한 후 실험을 진행한 결과, 조금 더 높은 해상도의 이미지를 생성했다. 망막의 혈관까지 일부 구현해내었기 때문에, 안저검사를 통해 망막 상태를 구별할 수 있을 것이라 기대하였다.

GAN으로 만들어낸 No_DR, DR 이미지 각각 2800개, DC GAN으로 만들어낸 이미지를 총 2839개를 사용하여 Semi-Supervised Learning에 활용하였다.

5) Semi-Supervised Learning

(1) 크롤링 이미지 활용

최대 test accuracy가 0.76이 출력되어 baseline과 거의 동일한 성능이 출력되었다. epoch 과정을 보았을 때, train accuracy가 0.99, valid accuracy가 0.61가 출력되어 오버피팅 되었음을 확인할 수 있다.

(2) DC GAN 이미지 활용

최대 test accuracy가 0.78이 출력되어 baseline보다 근소하게 높은 성능을 보였다. 해당 데이터 역시 train accuracy가 0.98, valid accuracy가 0.64가 출력되어 오버피팅 되었음을 확인할 수 있다.

이는 구형 크롤링 이미지, GAN의 해상도, DC GAN의 해상도가 매우 낮게 설정되었기 때문이며, 구형 크롤링 데이터 -> GAN -> DC GAN 순서대로 망막에 가까운 형상을 띄지만 가장 중요한 미세한 황반, 혈관의 모양을 생성해내는 것은 어려웠다. 혈관의 모양을 생성해낸 DC GAN 이미지마저도, Gaussian Blur transform을 수행한 것보다도 흐린 이미지를 출력해냈기 때문에, 위의 transform 과정에서 Gaussian Blur가 학습에 방해가 되었던 것처럼 모델의 성능을 하락시키는 원인이 되었다는 것을 추측해볼 수 있다.

4. 결론

적절한 data augmentation 기법은 예측 정확도 측면에서 의료 데이터를 기반으로 한 모델의 성능을 소폭 상승시킬 수 있었으며, regularization 역할도 수행하였다. 지나치게 많은 augmentation 사용, 일반 GAN과 같은 해상도가 낮은 데이터 활용, 데이터의 이미지를 blur시키는 행위 등은 모델의 성능을 오히려 저하시켰다.

크롤링 데이터와 DC GAN으로 수집한 데이터에 NoisyStudent를 적용하여 라벨을 붙인 후, 학습한 모델은 오버피팅되었다. 이는, 환자 구분에 중요한 미세

한 정보들을 제대로 학습하지 못했기 때문이며, 데이터의 낮은 해상도의 이미지를 학습에 사용하여 도출된 결론이라고 할 수 있다.

본 연구에서 도출된 결론은 하나이다. 모델의 성능 향상에는 증강된 데이터의 '개수'보다는 '유사성'이 더 중요하다는 것이다. baseline부터 semi-supervised learning에 이르기까지 Gaussian Blur와 같이 데이터의 중요한 정보를 잃게 만드는 모든 기법들은 모델 성능의 저하로 이끌었으며, Style GAN과 같이 실제 데이터와 매우 유사한 이미지를 생성하는 모델이나 해당 도메인에서 모델의 중요한 정보를 소실시키지 않는 적절한 transform을 사용하는 경우에서만 모델 성능의 향상을 이끌 수 있다는 것을 확인할 수 있었다.

참고 문헌

- [1] 손준홍, 「실명의 위험요인, 당뇨병성 망막병증」, 사단법인 한국당뇨협회, 2004. pp. 20~23.
- [2] "서울아산병원", [Internet]. Available : <https://www.amc.seoul.kr/asan/healthinfo/disease/diseaseDetail.do?contentId=31182>
- [3] SOVIT RANJAN RATH, "Diabetic Retinopathy 2015 Data Colored Resized", Kaggle, 2019. <<https://www.kaggle.com/datasets/sovitath/diabetic-retinopathy-2015-data-colored-resized>>
- [4] Karen Simonyan, Andrew Zisserman, "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION", ICLR, 2015. pp.1~8.
- [5] Ian J. Goodfellow , Jean Pouget-Abadie† , Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair† , Aaron Courville, Yoshua Bengio§, "Generative Adversarial Nets", University of Montreal, 2014.
- [6] Alec Radford, Luke Metz, Soumith Chintala , "Unsupervised Representation Learning with Deep

Convolutional Generative Adversarial Networks", University of Cornell, 2016. pp.2~11.

[7] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, Quoc V. Le, "Self-training with Noisy Student improves ImageNet classification", University of Cornell, 2020. pp.2~9.

[8] 박진선, 홍충선, "GAN 기반 의료 영상 학습을 위한 데이터 증강 기법", 한국컴퓨터종합학술대회논문, 2022.pp.2~5

[9]Helen Hong, Jaeeun Yoo, Yoogin Jang, "assessment and Analysis of Fidelity Diversity for GAN-based Medical Image Generative model", '한국컴퓨터그래픽스학회',