



Hand Bone Image Segmentation 프로젝트 Wrap up Report(수정본)

1. Introduction

1-1. Problem Statement

뼈는 우리 몸의 구조와 기능에 중요한 영향을 미치기 때문에, 의료 진단 및 치료 계획의 개발을 위해서는 정확한 뼈 분할이 필요하다. 이를 위해 X-ray 사진으로 **뼈의 종류를 예측하고 단위 구조로 분할하는 모델**이 필요하다. 특히 가려진 뼈 또는 모양이 비슷한 뼈를 정확하게 분할해야 의료 계획에 신뢰성을 높일 수 있기 때문에 우수한 성능의 모델이 필수적이다.

1-2. Competition Specifics

- Dataset : 손의 X-ray 사진, 사람 별로 두 장의 이미지가 존재, 총 1100장(550명)
- Evaluation metric : Dice coefficient
- Experiment environment : GPU V100 server

2. Methods

2-1. Model

	Description
unet 3+	기본 unet3+ 모델. 3*512*512 image를 입력받아 29*512*512의 mask를 출력
unet3+ deepsupervision(ds)	기본 unet3+에서 최종 출력층 뿐만 아니라 각 layer 별로 segmentation task를 수행함.
unet3+ ds 2048	unet 3+ ds에 최종 출력층에 upsampling, 3*3 depthwise convolution을 통해 29*2048*2048이 mask를 출력
unet3+ ds 2048 eff	unet3+ ds 1024에 layer별 channel을 1/2이상 줄이고, layer별

downsampling rate를 4배로 늘림. 또한 deepsupervision 수행시 layer 별 mask를 4배만 upsampling하여 parameter 수를 기존 unet3+ ds 1024 대비 1/4로 줄여 3*1024*1024 image를 입력받아 29*2048*2048 로 출력이 가능하게 한 모델

3. Experiment

3-1. Data

Dataset	Descriptions
train	기본 제공 train 데이터 800장

3-2. Models

Architecture	Backbone	Descriptions	Public score
FCN	ResNet101	200 epochs, BCEWithLogitsLoss	0.9487
U-Net	EfficientNet-B0	200 epochs, BCEWithLogitsLoss	0.9462
swin	swin transformer	26000 iters, BCEDiceLoss	0.9507
DeepLabV3Plus	ResNet101	20000 iters, BCEDiceLoss	0.9528
unet 3+	X	iou + bce loss,	0.9493
unet3+ deepsupervision(ds)	X	bce + dice loss, 50 epochs	0.9541
unet3+ ds 2048	X	bce + dice loss, 31epochs	0.9674
unet3+ ds 2048	X	bce + dice loss, with augmentation	0.9703
unet3+ ds 2048 eff	X	bce + dice loss, 13epochs, with augmentation	0.9709

SOTA 모델을 포함하여 다양한 모델을 loss, epoch을 변경하며 실험하였지만 public score 가 0.96 이상 나오지 않았다. 이미지의 해상도가 중요함을 파악하고 input 및 output의 해상도를 높이는 방향으로 모델링을 하였다.

3-3. Augmentations

Augmentation	Descriptions	Validation Score
--------------	--------------	------------------

Augmentation	Descriptions	Validation Score
None	augmentation 미적용	0.9441
데이터 증강		
HorizontalFlip	p=0.5	0.9471
GridDropout	p=0.5	0.9447
Cutout	p=0.5	0.9419
Cutout/Copy and paste	p=0.5	0.942
Rotate 1	10 degrees, p=0.5	0.9456
Rotate 2	30 degrees, p=0.5	0.945
전처리		
sharpness	sharpness_factor=2	0.9564
gamma	gamma=2	0.9561
saturation	saturation_factor=2	0.956
contrast	contrast_factor=2	0.9551
brightness	bright_factor=2	0.9521

이번 대회 데이터셋은 손 뼈의 x-ray 사진으로, 패턴이 단조롭기 때문에 모델의 일반화를 향상시키거나 ground truth의 클래스별 특징을 극대화하는 augmentation이 중요할 것이라 추측하고 비교 실험을 진행하였다. 실험 모델은 Unet 3+ ds 2048이며 최대 에폭은 8이다.

데이터 증강 기법은 training set에만 적용하고 validation set에는 적용하지 않았다. X-ray 데이터셋은 좌우의 손이 한 세트로 이루어져 있으므로 **HorizontalFlip**으로 한 사람당 총 4개의 손을 학습에 활용하였다. 또 segmentation의 특성상 픽셀 별 의존도를 줄이기 위해 **GridDropout**을 적용하였다. 모델의 정확도가 떨어지는 몇가지 클래스에 대해 **Copy and Paste**를 적용하였으나 크게 성능 향상이 있진 않았다. 몇 가지 데이터에서 손 방향이 꺾여있는 것을 관찰했기 때문에 **Rotate** 또한 적용하였다. 10도와 30도로 진행하였는데 10도가 더 성능이 좋았다.

전처리 기법은 training set과 validation set 모두에 적용하였다. sharpness, gamma, saturation, contrast, brightness 총 다섯 가지를 적용했으며 그중 **sharpness**가 성능이 가장 높았다. 여러 전처리 augmentation을 섞어서 조합을 만들어 학습을 진행하였으나 단일로 sharpness를 사용했을 때 validation score가 가장 높아 최종적으로 이를 채택하였다.


위의 표에서 확인할 수 있듯이, 아무것도 적용하지 않은 비교군(0.9441)에 비해 sharpness(0.9564), horizontalflip(0.9471), griddropout(0.9447)을 적용한 실험들은 validation 성능이 향상된 것을 확인할 수 있다.


3-4. Ensemble

Model	Group K-Fold	epochs	input	output
unet3+ ds 2048	0	33(그 이상)	512	2048
unet3+ ds 2048	1	22	512	2048
unet3+ ds 2048	2	69	512	2048
unet3+ ds 2048	3	31	512	2048
unet3+ ds 2048 eff	0	13	1024	2048

Group k-Fold로 폴드를 5개로 나누어 soft_voting 앙상블을 진행하였다. 50회 이상 학습한 모델에서는 horizontal flip에 과적합되어 검지와 약지를 동시에 예측하는 것을 EDA로 확인하였다. 그래서 개별 성능이 가장 낮은 4번 폴드는 제외하고 학습 에폭이 50회 이상인 폴드와 50회 이하인 폴드를 섞어서 활용하였다. 또 입출력 해상도가 512인 다른 모델들과 달리 1024인 모델을 포함하였다. 이렇게 **에폭수와 데이터셋 폴드, 입출력 해상도가 다른 환경에서 학습을 진행함으로써 앙상블에서 일반화를 극대화**시킬 수 있었다.

4. Conclusion

6 (-)	CV_06조		0.9730	56	3d
----------	--------	---	--------	----	----

7 (1 ↓)	CV_06조		0.9733	56	3d
------------	--------	---	--------	----	----

	Rank	Dice coefficient
Public Scores	6	0.9730
Private Scores	7	0.9733

5. 자체 평가 의견

잘했던 점

- 앙상블 과정이 전반적으로 순조로웠다. K-Fold Ensemble을 위해 데이터를 나누고 모델을 통일하는 과정이 신속했다.
- Augmentation을 다양하게 진행하였고 실제로 큰 효과를 얻을 수 있었다.

- 기존 UNet을 개조하여 해상도를 높이는 것이 성공적이었다.

시도 했으나 잘 되지 않았던 것들

- SOTA 모델 실험(MMSegmentation 버전 이슈)
- TTA

아쉬웠던 점들

- EDA를 하지 않았다.

프로젝트를 통해 배운점 또는 시사점

- Augmentation을 통한 성능 향상이 크다.
- K-Fold Ensemble을 위해 미리 데이터를 나누는 것이 효과적이다.

6. 개인회고

오서영

학습 목표를 달성하기 위해 시도한 점

- segmentation 프레임워크 이해

segmentation 분야에서 활용되는 프레임워크에는 크게 pytorch와 mmsegmentation이 있습니다. 이번 대회를 통해 두 프레임워크 환경에서 fcn 모델을 학습시켰으며 소스코드를 통해 작동방식을 이해해 보았습니다.

- augmentation

학습 데이터의 단조로운 패턴을 극대화하기 위해 augmentation 비교 실험을 진행하였습니다. deep 3+ ds 2048 기준 test score이 augmentation을 적용하지 않았을 때(0.9674)에 비해 적용했을 때(0.9703) 성능이 크게 향상되었습니다.

마주한 한계/아쉬웠던 점

- mmsegmentation 라이브러리 버전 충돌

pytorch에 비해 mmsegmentation은 라이브러리 버전에 민감한 프레임워크였습니다. onepeace와 같은 SOTA 모델을 활용해보고자 하였는데 라이브러리 버전이 계속해서 충돌하여 시간적인 제약으로 실제 학습시켜보지 못한 점이 아쉽습니다.

- 이미지 해상도 변형

이번 대회 of 핵심은 서버 용량이 수용 가능한 선에서 이미지 해상도를 원본과 최대한 흡사하게 학습을 진행하는 것이었습니다. 이 점을 일찍 파악했다면 제가 실험했던 swin이나 fcn resnet101과 같은 모델에 적용하여 성능을 높일 수 있었을 텐데, 늦게 확인하여 실험해보지 못한 점이 아쉽습니다.

이현구

학습 목표를 달성하기 위해 시도한 점

- unet, unet 3+의 구조를 완벽하게 파악하였다. 또한 unet 3+의 경우 논문 저자의 코드를 완벽히 파악하고 parameter 수는 줄이면서 고해상도의 mask output을 얻기 위해 모델을 수정하였다.
해당 시도를 바탕으로 unet3+ ds 2048, unet3+ ds 2048 eff를 제작하였으며, unet3+ ds 2048는 unet 3+ ds 대비 dice score가 0.0133점 상승하였고, unet3+ ds 2048 eff는 unet3+ ds 2048 대비 0.0006점 상승하였다. unet3+ ds 2048 eff가 시간 관계상 unet3+ ds 2048의 절반 정도의 epoch 밖에 학습을 하지 못 한 점을 감안하면 더 높은 성능 향상도 기대할 수 있다.
- weighted softvoting으로 기존 softvoting 대비 dice score가 0.0004점 상승하였다.

마주한 한계/아쉬웠던 점

- 학습 시간이 오래 걸려 다양한 실험을 하지 못해 아쉬웠다.
- 최종적으로 제작한 모델이 가장 높은 점수를 보여주었지만 학습 시간으로 인해 약 절반 정도 학습 후 결과를 제출한 점이 아쉬웠다.
- 다양하게 출력층을 수정하고 학습하는 과정에서 기학습된 모델들의 일부 weight들을 가져와서 학습을 진행하였으면 더 빠르게 실험을 진행할 수 있었을 것 같다.

정현진

학습 목표를 달성하기 위해 시도한 점 및 새롭게 시도한 변화 및 효과

이번 프로젝트의 개인적인 학습 목표는 SOTA 모델을 학습시켜 보는 것이었다.

DeepLabV3는 mmSegmentation의 config에 포함되어 있기는 했지만 SOTA 모델 중 하나임으로 사용해 보려고 노력하였다.

지난 object detection 대회에서는 mmDetection 라이브러리가 어떻게 작동하는지 파악하지 못해 미리 작성되어 있는 코드를 그대로 사용하기만 했다. 그러나 이번 대회에서는 mmSegmentation이 작동하는 원리를 오피스 아워를 통해 알 수 있었고, 이를 바탕으로 정

의되어 있는 모델의 구조, loss 함수 등을 변경하는 등 mmSegmentation을 자유롭게 현재 task에 맞추어 사용할 수 있었다.

마주한 한계/아쉬웠던 점

UNet과 DeepLabv3 등 리서치한 모델이 좋은 성능을 내지 못해 최종적으로 사용하지 못한 점이 아쉽다.

이번 프로젝트에서는 모델 실험과 간단한 augmentation 실험만 했다. 데이터에 맞춰 조금 더 다양한 기법을 시도해 보지 못한 것이 이번 프로젝트의 한계점이라고 생각한다.

다음 프로젝트에서 시도해보고 싶은 점

이번 프로젝트는 빨리 학습을 시켜야 한다는 생각 때문에 EDA를 소홀히 했는데, 다음 프로젝트에서는 EDA를 꼼꼼히 하여 미리 프로젝트의 방향을 정하고 다양한 실험을 진행하고 싶다.

김가영

학습 목표를 달성하기 위해 시도한 점

모델의 성능을 끌어올리기 위해 다양한 실험을 진행하였다. 특히 대체적으로 겹쳐있는 클래스가 성능에 많은 영향을 주었는데 해당 내용을 파악하기 위해 로그를 확인하는 등 노력을 기울였다.

다양한 버전의 모델을 올려보며 성능을 비교해보고 시드 실험 등 실험의 작은 부분도 확인하고자 노력했다. 해당 내용을 Optuna로 구현하여 자동화 하였으나 학습 시간이 오래 걸려서 지표를 다양하게 써보기엔 시간적인 한계가 존재했다.

마주한 한계/아쉬웠던 점

실험의 특성 상 한번에 한 가지의 변수만 바꾸어 진행해야 유의미한 결과를 얻을 수 있는데 학습 시간이 오래 걸려 실험이 매우 길어진 것이 아쉽다.

다양한 라이브러리를 통해 실험을 진행하고 싶었으나 내부 환경 설정 문제로 시간을 많이 빼앗긴 점이 아쉽다. 이번 대회에선 데이터의 부족으로 Transformer 기반의 모델의 성능이 좋지 않았는데, 개인적인 실험을 진행할 때 하나의 모델을 잡고 꾸준히 디벨롭 했다면 더 좋았을 것 같다.

임지윤

학습 목표를 달성하기 위해 시도한 점

FCN의 백본 교체 실험을 진행하였는데, 해당 백본에 대한 FCN의 기학습 가중치가 없었다. 따라서 기존 PyTorch에서 제공하는 ResNet101 버전보다 성능이 떨어질 수밖에 없었다.

MMSegmentation 또한 도전하였는데, 아직 multi label을 지원하지 않아서 SOTA 모델을 시도하는데 어려움이 많았다. 여기에 시간을 많이 허비해서 아쉬움이 많다.

데이터 증강 실험을 전 대회보다 많이 진행해서 좋았다. 데이터 증강이 이러한 단조로운 데이터에서 많은 성능 향상을 주는 것을 잘 느낄 수 있었다.

전반적으로 개인적으로는 많은 실력 향상이 있었다고 느낀다. 성공적이진 않았지만 SOTA 모델을 도전하는 과정에서 MMSegmentation 라이브러리를 하나하나 뜯어보았고 이해하는데 많은 도움이 되었다. 그 전에는 PyTorch에서 백본을 바꾸기 위해 클래스를 다 뜯어보았는데 이 과정이 그 다음 과정을 수행하는데 도움을 줬다.

마주한 한계/아쉬웠던 점

그외에도 TTA 또한 진행하였는데, 기존에 생각했던 방식으로 진행하려다 팀원의 권유로 바꿨었다. 그러다가 시간이 촉박하여 서두르다가 잘 안 된 것 같다. 그래도 완성본을 제출하였는데, AI Stages 서버가 몇분동안 터져서 제출을 못 했다. 다음부터는 시간 관리에 신경써야 할 것 같다.

다음 프로젝트에서 시도해보고 싶은 점

이번 프로젝트에선 EDA를 아예 하지 않은 것 같다. 또한 팀원들도 다들 지친 것이 느껴졌다. 나 또한 지속적으로 늦게 자다 보니 수면 패턴이 망가졌고, 신경이 날카로워진 것 같다. 다음 프로젝트에서는 자기 관리에 힘 써서 긍정적 에너지로 임하고 싶다.