

인체조직세포 Segmentation 적용 및 최적화

CUA이 5기 CV 3팀

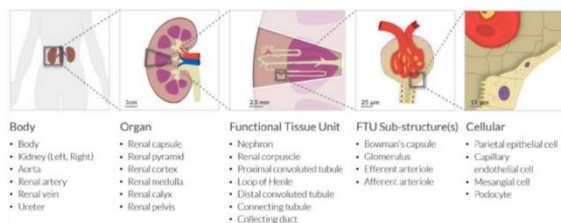
정승혁(에너지시스템공학부), 김동영(소프트웨어학부), 차수현(기계공학부), 함정훈(기계공학부)

[요약] 다양한 기관을 촬영한 사진에서 각각 FTU(Functional Tissue Unit)을 찾아보는 최종 목표 속에서 다양한 데이터를 분석해보고, 데이터에 적합한 segmentation 기법을 찾아보았다. Unet을 적용시키는 과정에서 생기는 문제점들을 분석하고, Unet보다 더 뛰어난 성능을 보일 것으로 예측되는 Swin Transformer를 Unet대신, 혹은 함께 적용시켜보고, 의료데이터에 적합한지 비교 분석해본다.

1. 서 론

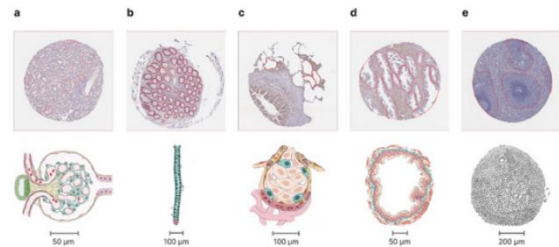
Computer Vision 분야에서 새로운 연구가 지속되고 있는 상황속에서, 분자 단위의 생체 지도를 만드는 과정 또한 시도되고 있다. Human Protein Atlas(줄여서 HPA)사는 일차적으로 Kaggle competition을 통해 신장의 주요 기능을 수행하는 세포 및 조직 단위를 segmentation 해보았고, 더 나아가 다양한 조직 및 다른 형식을 갖는 Dataset에도 적용해보는 시도를 Human BioMolecular Atlas Program(줄여서 HuBMAP)사와 함께 도전해보고 있다.

따라서 두 기관이 개최한 Kaggle competition의 최종 목표는 다섯개의 장기를 촬영한 사진에서 각각 FTU의 위치를 찾아내는 것이므로 이 대회를 진행하며 세포와 조직 구성간의 관계를 이해하고자 했다. FTU는 기능을 갖는 조직을 구성하는 가장 작은 단위로 인체구성단계를 다음과 같이 설명할 수 있으며,



[그림 1] 인체구성단계

사진 순서대로 신장, 대장, 폐, 전립선, 비장에서 FTU는 다음과 같이 나타나고 있다.



[그림 2] 신장, 대장, 폐, 전립선, 비장에서의 FTU

Segmentation은 픽셀 단위로 Classification을 수행하는 것으로, 크게 동일한 객체끼리 구분하는 Semantic Segmentation과 개별 물체에 대해 각각 masking을 진행하는 Instance Segmentation으로 구분된다. FTU가 담긴 이 데이터의 경우 Semantic Segmentation을 적용해볼 수 있다.

우선 훈련데이터와 검증데이터는 다음과 같았다.

Train Dataset : HPA
Open Test Dataset: HPA + HuBMAP
Closed Test Dataset: HuBMAP

또한 데이터셋의 label은 다음과 같음을 확인하였다.

Id/organ/img_height/img_width/pixel_size/
tissue_thickness/rle/age/sex

이후 각 기관별 FTU를 다음과 같이 분석해보았다.

[표 1] 각 기관별 FTU 분석

organ	average_area	average_area_μm	average_height	average_width	average_shape_μm
0 kidney	109827.135135	17572.341622	379.0	493.5	(151.6, 197.4)
1 largeintestine	30644.440000	4903.110400	199.5	162.0	(79.8, 64.8)
2 lung	96282.181818	15405.149091	281.5	246.5	(112.6, 98.6)
3 prostate	145201.100000	23232.176000	211.0	260.0	(84.4, 104.0)
4 spleen	332090.645161	53134.503226	1101.0	701.0	(440.4, 280.4)

2. 본 론

1) Image Augmentation

Image Augmentation은 크게 아래와 같이 분류해 볼 수 있는데, 우리가 다룰 FTU Dataset은 Data 개수가 충분치 않다고 판단하여 Basic Imagemanipulation(fliping, Cropping, Rotation, Transformation, Color space,, etc)이외에 추가로 방법을 적용시켜보기로 하였다.



[그림 3] Image Augmentation의 분류

이미지 내에는 같은 기관만 존재하고, 이 기관 내 FTU를 판별한다는 특징을 고려하여 추가 Augmentation 기법으로는 RICAP(Random Image Cropping And Patching)과 Gan Augmentation을 선정하였다.

① RICAP

무작위 사진 4개를 각각 랜덤하게 잘라서 하나의 사진으로 만들고, soft labeling을 통해 학습을 진행하게 된다. → 여기서는 같은 기관 내 이미지속에서 무작위하게 선정하였다.

② Gan Augmentation

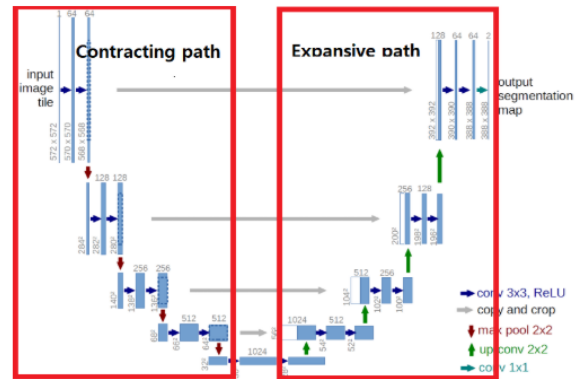
이미지 데이터 속 pertinent variance와 non-pertinent variance를 구분하고, non-pertinent variance를 최소화하는 과정을 거친다. → Gan으로 현재 SOTA 모델로 인정받는 StyleGAN을 사용하고, noise filter로 Gaussian Filter 혹은 Patch Shuffle Filter를 사용한다.

2)U-Net

U-Net은 Fully Convolution Network(FCN)을 기반으로 한 모델인데, 적은 데이터에서 정확한 Segmentation을 내기 위해 FCN의 구조를 수정했다고 할 수 있다. 우리가 augmentation하기 이전, 주어진 데이터는 수백개로 많지 않은만큼 U-Net을 우

선적으로 고려하게 되었다.

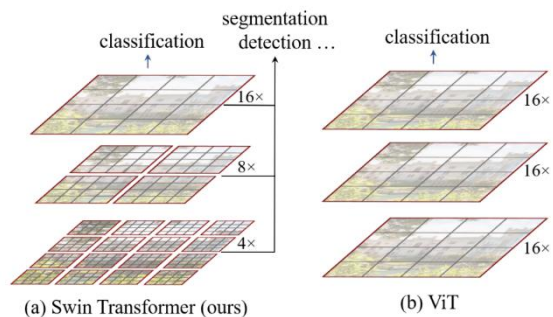
아래를 보게 되면 U-Net의 주요 특징으로 이미지의 context를 포착하는Contracting Path, feature map을 upsampling하고 포착한 context와 결합하는 expansive path가 있음을 확인할 수 있다.



[그림 4] U-Net 구조

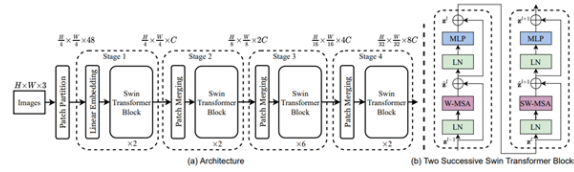
3)Swin Transformer

데이터를 고려해 U-Net을 우선적으로 사용해보았지만, 성능이 그다지 좋지 않다는 것을 확인하였다. 따라서 segmentation에서 SOTA를 달리고 있는 swin transformer를 적용해보고자 하였다.



[그림 5] swin transformer와 vision transformer

위 그림은 swin transformer와 vision transformer의 차이를 보여준다. vit는 전체 input을 patch로 나누고, 각 patch는 다른 모든 patch들에 대해 self-attention을 수행한다. swin transformer는 patch 하나를 더 작은 patch로 나누고, 각 패치마다, 그 작은 패치들끼리만 self-attention을 수행한다. 이는 computation cost를 줄여주는 역할을 한다.



[그림 6] 모델 구조의 도식화

위 그림은 모델 구조를 도식화한 것이다. stage 1에 들어가기 전에, patch partition layer를 거친다. patch partition layer는 이미지를 patch로 나누는 역할을 한다. 논문에서는, 입력 이미지를 (4, 4, 3)으로, 즉 16개의 픽셀로 나눈다. 따라서, 첫 패치 하나의 크기는 48이고, (H, W)의 이미지를 (4, 4)로 조깅기 때문에 $H/4 * W/4$ 개의 패치가 생성된다. 따라서, 출력의 형상이 (H/4, W/4, 48)이 된다.

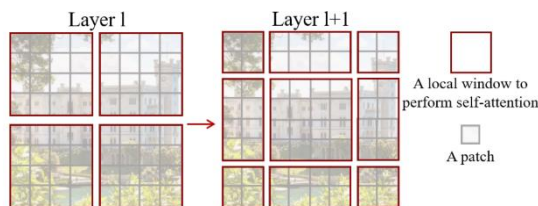
이렇게 생성된 patch들은 linear embedding layer에 들어가게 되고, 각 patch들은 C 차원으로 사영된다. (출력 형상: (H/4, W/4, C))이 patch들은 swin transformer에 들어가 계산되고, 다음 stage에서는 patch merging layer에서, patch를 인접한 4개씩 합치고(concat), 채널은 2배로 키운 뒤 똑같은 과정을 반복한다. patch는 4개씩 합치는데 채널은 2배로 늘어나는 것에 의문을 가질 수도 있는데, 논문에 자세한 설명은 적혀있지 않지만, 이렇게 제한하지 않으면 patch의 차원의 너무 커지기 때문이라 생각할 수 있을 것이다.

하나의 swin transformer block은 2개의 encoder로 구성된다. ViT에서 사용한 MSA와는 다른 W-MSA를 사용한다. W-MSA를 사용하면 각 window 안에서만 patch들끼리 self-attention을 수행하기 때문에, 연산량이 줄어든다.

$$\Omega(\text{MSA}) = 4hwC^2 + 2(hw)^2C,$$

$$\Omega(\text{W-MSA}) = 4hwC^2 + 2M^2hwC,$$

연산량은 위와 같다. 일반적으로 M은 hw보다 훨씬 작고, 상수이기 때문에 연산량이 줄어든다.



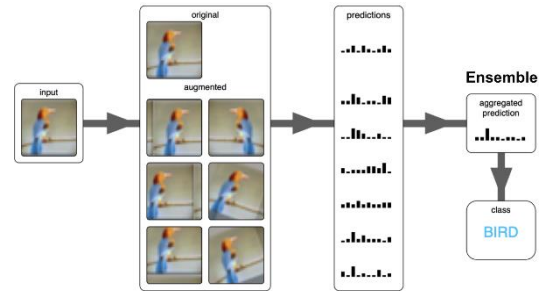
[그림 7] W-MSA와 SW-MSA

위 사진은 W-MSA와 SW-MSA를 나타낸 그림이다.

W-MSA는 window가 고정되어 있기 때문에, window를 조금 움직인 다음 self-attention을 또 수행하는데, 이것이 SW-MSA이다.

4) TTA in segmentation

모델을 학습시킨 이후에는 TTA 기법을 고려해보았다. Test 과정에서 Augmentation을 적용한 뒤 예측 확률을 평균을 통해 도출하는 기법인데, 이 역시 현재 저지 적은 데이터에서 좋은 성능을 이끌어내고자 할 때 주로 사용된다.



[그림 8] TTA 기법

5) Pseudo Labeling

마지막으로 Pseudo Labeling을 시도해보았는데, 기존에 정답이 있는 데이터에 정답이 없는 데이터를 학습시킨 결과를 접목시켜 기존에 가지고 있던 정답 Label 데이터에 기반하여 확실적인 정답 라벨을 부여하는 기법이다.

이 과정은 정답라벨이 있는 데이터로 성능이 일정 이이 확보되었을 때 유의미한 기법이므로, 그렇지 않을 경우 오히려 성능이 저해될 우려가 있으므로 앞선 과정들이 모두 성공적으로 진행되었을 때 최종적으로 적용하기로 계획하였다.

3. 결 론

다섯 가지의 인체 장기 이미지 속 FTU를 찾는 과정에서, 적용시켜본 기법들은 다음과 같다.

- ① RICAP augmentation
- ② GAN augmentation
- ③ U-Net with ASPP
- ④ Swin Transformer
- ⑤ TTA in segmentation
- ⑥ Pseudo Labeling

Semantic Segmentation에서 위 기법들이 얼마나 효율

적인지, 혹은 예상보다 못한 성능을 가지고 있는지 확인할 수 있었고 Computer Vision에서 의료데이터의 고질적인 문제, 데이터가 부족한 문제를 해결하기 위한 여러 방안이 시도되었기에 비슷한 조건을 갖춘 다른 데이터셋에서도 위 기법들을 적용해볼 수 있음을 확인할 수 있었다.

참고 문헌

Segmentation 개념, 용어, 종류 :

[https://velog.io/@cha-](https://velog.io/@cha-suyeon/%EB%94%A5%EB%9F%AC%EB%8B%9D-Segmentation1-%EA%B0%9C%EB%85%90-%EC%9A%A9%EC%96%B4-%EC%A2%85%EB%A5%98Sematic-Instance-segmentation)

[suyeon/%EB%94%A5%EB%9F%AC%EB%8B%9D-Segmentation1-%EA%B0%9C%EB%85%90-%EC%9A%A9%EC%96%B4-%EC%A2%85%EB%A5%98Sematic-Instance-segmentation](https://velog.io/@cha-suyeon/%EB%94%A5%EB%9F%AC%EB%8B%9D-Segmentation1-%EA%B0%9C%EB%85%90-%EC%9A%A9%EC%96%B4-%EC%A2%85%EB%A5%98Sematic-Instance-segmentation)

“A survey on Image Data Augmentation for Deep Learning”, Connor Shorten, 2019

[논문 리뷰] RICAP :

<https://velog.io/@cindy4741/RICAP-%EB%85%BC%EB%AC%B8-%EC%A0%95%EB%A6%AC>

“Gan Augmentation : Augmenting Training Data using Generative Adversarial Networks”, Christopher Bowles et al, 2018

“A Style-Based Generator Architecture for Generative Adversal Networks”, Tero Karras, Samuli Laine, Timo Aila, 2019

“U-Net: Convolutional Networks for Biomedical Image Segmentation”, Olaf Ronneberger, Philipp Fischer, Thomas Brox, 2015

Z Liu, "Hierarchical Vision Transformer using Shifted Windows", ICCV, 2021

https://stepup.ai/test_time_data_augmentation/