

U-Net: Convolutional Networks for Biomedical Image Segmentation

0. Abstract

Data Augmentation을 활용하여, 제한된 주석 데이터를 더 효율적으로 사용
contracting path로 문맥을 포착하고, expanding path로 정확한 위치를 파악.
속도가 빨라, 최신 GPU에서 512x512 이미지를 1초 미만에 세분화 가능

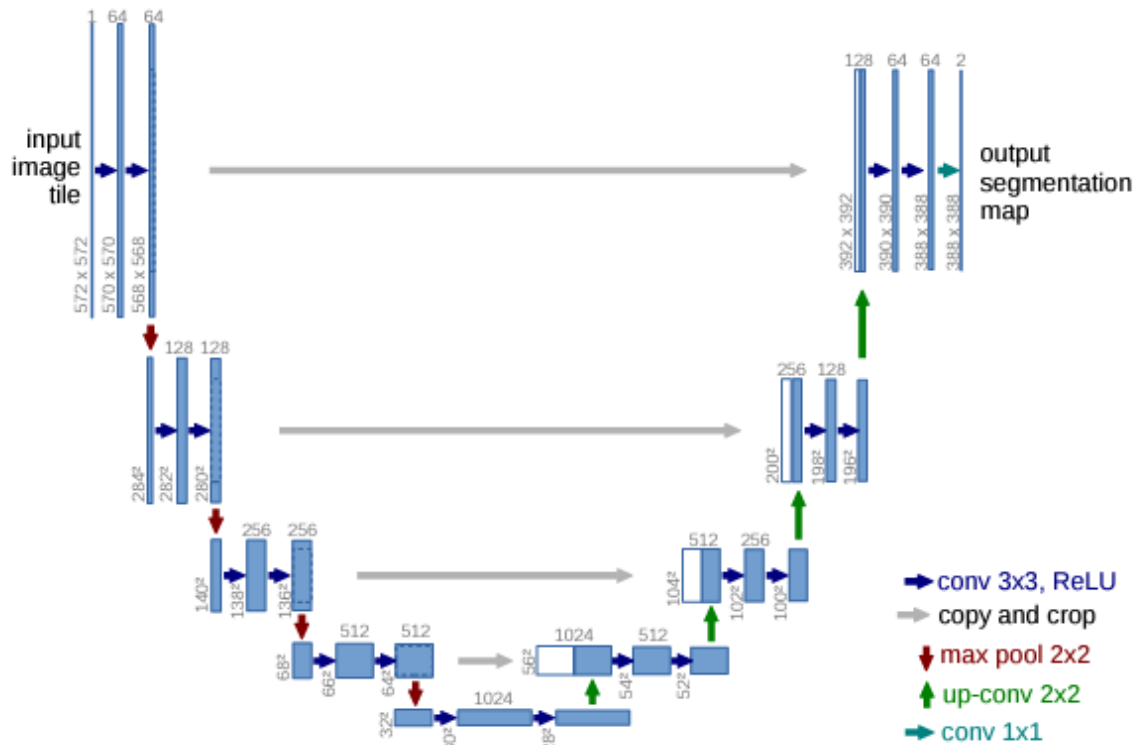
1. Introduction

deep convolutional 네트워크가 이미지 인식 태스크에서 성능이 뛰어났음. 학습 세트 크기와 네트워크 사이즈 제한으로 성능 제한이 있음. 8겹과 수백만개의 파라미터로 이루어진 ImageNet 데이터셋을 지도 학습해 극복. 더 크고 깊은 네트워크 훈련 가능

분류할 때는 이미지를 단일 클래스 라벨로 했었는데 클래스 라벨이 각 픽셀에 배당되도록 바뀜. sliding window 기법 사용하면 localize 가능하고 학습 데이터의 패치가 커져 학습하기 편해짐.

단점은 각 패치를 따로 진행해야해서 느리고, 겹쳐진 패치로 효율성이 떨어짐. localization 정확도와 문맥 사이의 트레이드 오프가 존재함. 패치가 커질수록 max pooling layer가 많이 필요하고 localization 정확도가 떨어짐. 반대로 작은 패치는 문맥을 조금만 살핌.

이 논문에서는 더 정교한 아키텍처로 fully convolutional network를 다룸. 약간의 학습 이미지로 더 정확한 세그멘테이션 가능. 연속적인 레이어 사용. 풀링은 업샘플링으로 교체.



업샘플링이 중요한 요소임. 피쳐 채널이 많아 고해상도 층에서 문맥 정보를 전파할 수 있음. 그 결과 expansive path 가 contracting path와 거의 대칭되어 U자 형태를 가지게 됨.

학습 데이터가 매우 적어 과도한 데이터 어그멘테이션도 사용함. elastic deformation을 적용해 학습 데이터에 적용함.

cell segmentation 태스크에서 또 다른 문제로는 붙어있는 세포가 문제인데 weight loss 로 백그라운드 라벨을 분리해 붙어있는 셀들을 분리

2. Network Architecture

왼쪽의 contracting path와 오른쪽은 expansive path로 구성됨.

contracting path: 전형적인 convolutional 네트워크 구조를 따름 2개의 3x3 convolution과 각각 ReLu 활성화 함수를 지나고 2x2 max pooling, stride 2 해서 다운 샘플링함. 각 다운 샘플링 단계에서 피쳐 채널 수를 두배함.

Expansive path: 2x2 up-convolution으로 너비와 높이 2배로 늘리고 커널 개수는 반으로 줄어듦. contracting path에서 잘린 피쳐 맵을 합쳐 2개의 3x3 convolution을 하고 최종 레이어에 1x1 convolution 진행

최종적으로 네트워크는 23개의 convolutional layer를 가지고 있음. seamless tiling을 하기 위해 인풋 타일 사이즈를 맥스 풀링할때 x,y에 모두 적용해야 함

3. Training

unpadded convolution 때문에 출력 이미지는 인풋 이미지보다 작아짐. 오버헤드를 최소화하기 GPU 메모리 사용을 최대화하기 위해 큰 인풋 타일 사용.

에너지 함수는 마지막 얻은 피쳐 맵에 픽셀 단위로 soft max 사용. cross entropy loss function 적용.

$$E = \sum_{\mathbf{x} \in \Omega} w(\mathbf{x}) \log(p_{\ell(\mathbf{x})}(\mathbf{x}))$$

cross entropy loss로 각 위치에서 실제 값과 벗어난 정도를 페널티 줌.

$w(\mathbf{x})$ 는 클래스 빈도의 균형을 맞추기 위한 가중치 맵

$$w(\mathbf{x}) = w_c(\mathbf{x}) + w_0 \cdot \exp\left(-\frac{(d_1(\mathbf{x}) + d_2(\mathbf{x}))^2}{2\sigma^2}\right)$$

실험에서는 w_0 을 10으로 설정하고 시그마를 5로 설정.

deep network에서 많은 컨볼루션 레이어와 다른 경로를 지날 때 가중치의 초기화가 중요함. 각 네트워크가 과도한 활성화를 줄 수 있고 전혀 기여하지 않을 때도 있음. 초기 가중치를 이상적으로 설정하기 위해 각 피쳐맵을 unit variance와 비슷하게 설정. 주로 초기값을 가우시안 분포에서 도출함.

3.1 Data Augmentation

네트워크가 불변성과 robust한 특성을 가지게 하기 위해 data augmentation이 필요.

작은 이미지에서 우리는 이동과 회전 불변성이 필요함. 특히 랜덤 변형을 통해 학습 데이터를 증가시킬 필요가 있음.

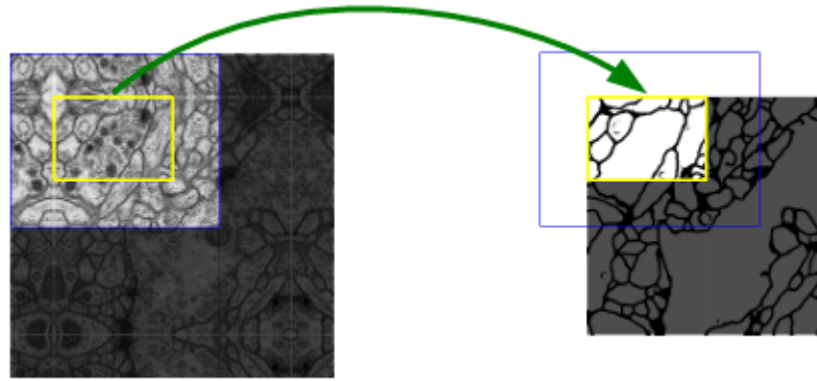
smooth 변형을 위해 3x3 그리드의 변위 벡터를 이용함.

픽셀 단위 변형, drop out 레이어 사용

4. Experiments

U-net을 3가지 다른 세그멘테이션 태스크에 적용

1. 전자 현미경에서 신경 구조 세그멘테이션



EM 세그멘테이션 챌린지에서 시작됨. 학습 데이터는 512x512 픽셀로 이루어진 이미지 30개

각 이미지는 하얀 세포와 검정색의 세포막으로 구성된 ground truth segmentation map을 만듦.

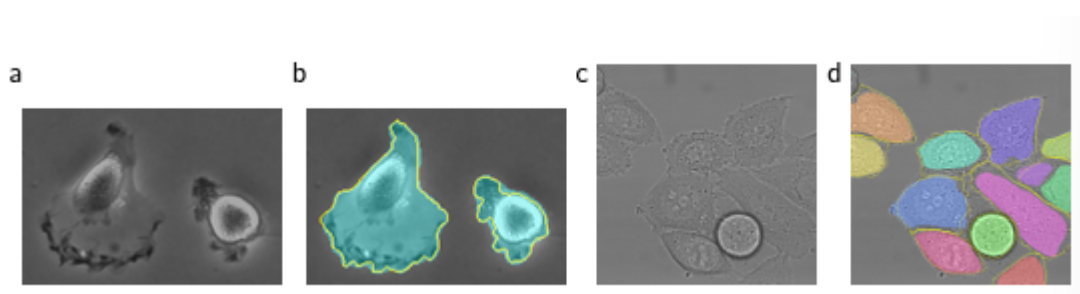
평가지표: warping error, rand error, pixel error

Rank	Group name	Warping Error	Rand Error	Pixel Error
	** human values **	0.000005	0.0021	0.0010
1.	u-net	0.000353	0.0382	0.0611
2.	DIVE-SCI	0.000355	0.0305	0.0584
3.	IDSIA [1]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	0.0582
	⋮			
10.	IDSIA-SCI	0.000653	0.0189	0.1027

warping error에서 U-Net이 가장 성능이 좋음

2. ISBI cell tracking 챌린지의 데이터로 실험 적용

35개의 부분적으로 표시된 이미지 포함. 위상 대비 현미경 이미지



평균 Intersection over Union 92% 달성. 세포라고 구분한 영역과 실제 세포의 영역이 겹치는 정도로 성능 측정함.

5. Conclusion

U-Net은 다양한 biomedical 세그멘테이션 적용에서 매우 좋은 성능을 보임.

elastic deformation이 포함된 data augmentation 덕분에 적은 데이터셋만으로 효율적인 시간 동안 학습 가능.

U-Net의 구조가 다양한 태스크에 쉽게 응용될 수 있을 것임.