

U-net 논문 리뷰

- 일반적으로 cnn은 이미지 전체를 하나의 class label로 분류하는데 사용됨
- 의료 영상 처리에서는 픽셀 단위의 정밀한 분류가 필요(한 이미지에 대해 각 픽셀마다 클래스 레이블을 예측해야 함)
- 기존 슬라이딩 윈도우 방식을 개선 하고자 u-net 같은 네트워크가 등장함
- ISBI cell tracking challenge 2015에서 압도적으로 우승

개요 (Abstract)

딥러닝 기반의 신경망을 성공적으로 훈련하려면 수천 개의 주석된 데이터가 필요

데이터 증강(Data Augmentation) 기법으로 적은 데이터로도 효과적인 학습을 수행

U-Net은 수축 경로(Contracting Path)와 확장 경로(Expanding Path)로 구성된 대칭적 구조를 갖는 신경망.

ISBI 2015 챌린지에서 최고 성능을 기록하며 전자현미경(EM) 및 광학 현미경 이미지(Phase Contrast, DIC)에서 우수한 결과를 달성.

연구 배경 (Introduction)

- 기존 CNN은 대부분 이미지 분류(Classification)에 초점, 즉 이미지 전체를 하나의 레이블로 분류.
- 의료 영상 처리에서는 픽셀 단위의 세분화(Segmentation)가 필요함.
- Ciresan et al. (2012)은 슬라이딩 윈도우(Sliding Window) 방식을 사용해 각 픽셀을 분류하는 방법을 제안했지만, 이 방법은 속도가 매우 느리고, 위치 정보(Localization)와 문맥(Context) 정보 활용 사이의 trade-off 문제 발생

trade off

작은 패치를 사용하면 세부 정보를 더 정확하게 분류할 수 있지만, 전체 문맥 정보를 적게 활용함.

큰 패치를 사용하면 더 많은 문맥 정보를 활용할 수 있지만, 위치 정보(Localization) 정확도가 떨어짐.

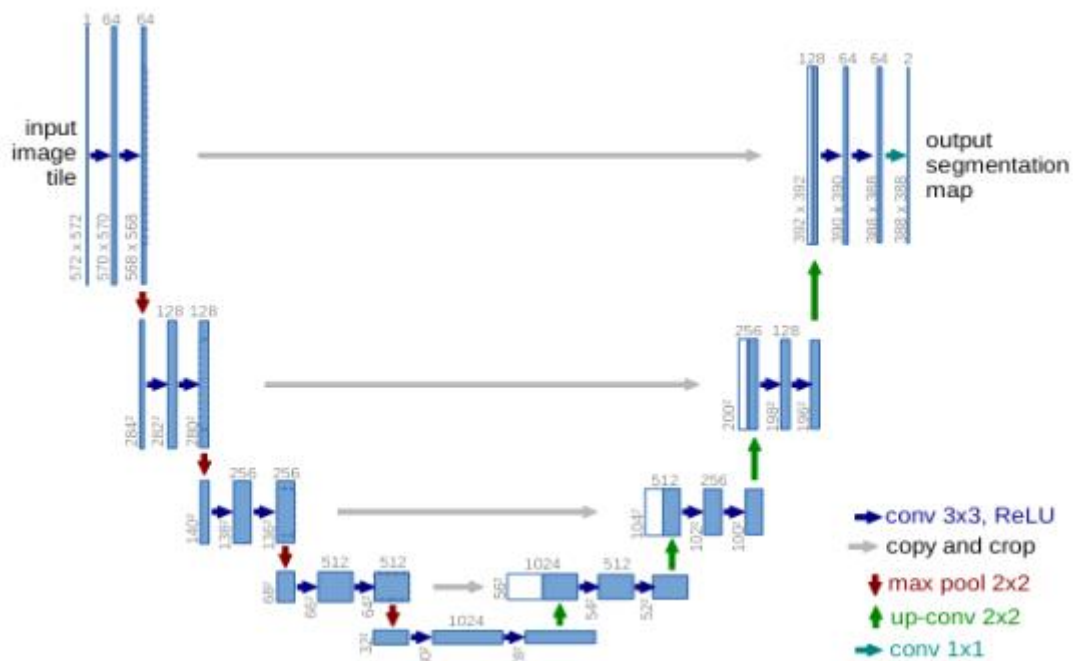
따라서 정확한 위치 정보 vs. 문맥 정보 활용 사이의 trade-off가 존재함.

해결책: U-Net

기존 Fully Convolutional Network(FCN)를 확장하여, 적은 데이터로도 효과적인 세분화(Segmentation) 가능하도록 개선.

U-Net 네트워크 구조 (Network Architecture)

- U-Net은 U자 형태의 구조로 구성
- 수축 경로, 확장 경로



위 사진 u자 왼쪽이 수축 오른쪽이 확장 경로이다

1. 수축 경로(Contracting Path)

!파란색 화살표

- 각 contracting step마다 3x3 conv 두 차례 반복
- 패딩이 없으므로 feature map 조금씩 줄어듦 -> 512->570->568
- conv에는 relu 연산 포함

!빨간색 화살표

- 2x2 maxpooling(stride:2) 연산
- feature map의 크기가 절반으로 줄어듦
- down-sampling 할 때마다 채널 수 2배로 늘어남

2. 확장 경로(Expanding Path)

!초록색 화살표

- 2x2 up-conv을 수행
- feature map의 크기가 두배로 늘어남

!파란색 화살표

- 각 expanding step마다 3x3 conv을 두 차례 반복
- 패딩이 없으므로 featuremap 조금씩 줄어듦(200->198->196)

- conv에는 rel연산 포함
- up-sampling 마다 채널의 수 절반으로 줄어듦

3. skip-architecture

!회색 화살표

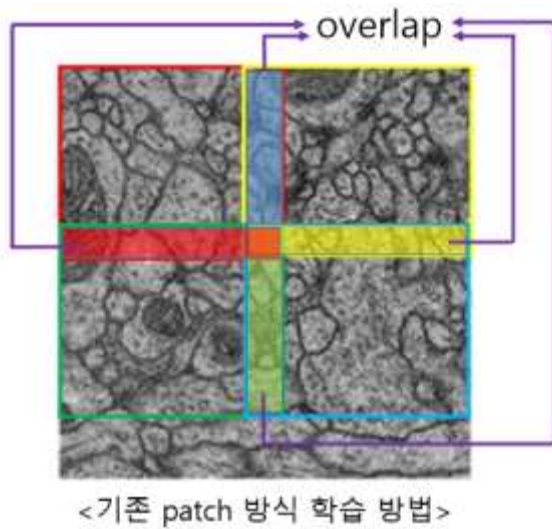
- 각 expanding step 마다 contracting path의 feature map의 테두리 부분을 자른 후 크기를 동일하게 맞추어 두 feature map을 합침
- 이 과정을 통해 더 정확한 위치 정보를 복원할 수 있다.

4. 최종 출력 (Final Layer)

1×1 합성곱(1×1 convolution)을 사용하여, 각 64차원(feature vector) 특징 벡터를 원하는 클래스 개수(class number)로 매핑
전체 네트워크는 총 23개의 합성곱 계층(convolutional layers)으로 구성

입력 이미지에 적용한 기법

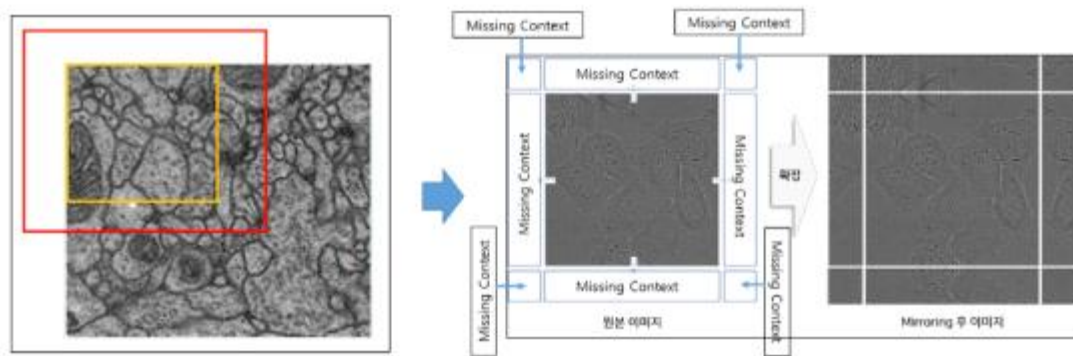
1. overlap-tile strategy



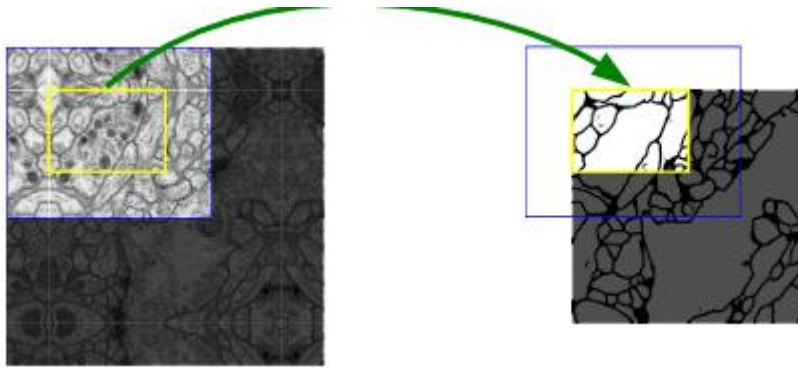
기존 patch 방식으로 학습하면 겹치는 부분이 많이 생김

위 문제 해결하기 위해 overlap-tile 기법 사용

u-net은 input사이즈(572,572)와 output(388,388) 사이즈가 다름



빨간색 bounding box와 실제 segmentaion에 해당하는 주황색 부분
주황색 부분을 모방하여 input image 크기를 늘려준다



위 그림을 보면 노란색 박스와 그 주위 이미지가 똑같음을 알 수 있음
 이렇게 bounding box를 넓혀줘서 실제 이미지 segmentation 할 때는 겹치는 것을 막을 수 있음

2. 데이터 증강(Data Augmentation)

훈련 데이터의 양을 증가시키기 위해 기존 데이터를 변형하는 기법

이 논문에서 적용된 기법은

- Shift : 이미지를 수평, 수직 이동
- Rotation : 다양한 각도로 회전
- Gray value : 픽셀의 밝기 조절
- Elastic Deformation : Pixel이 랜덤하게 다른 방향으로 뒤틀리도록 변형하는 방식

훈련 방식 (Training)

optimizer: stochastic gradient descent(SGD)

momentum:0.99

deep learning framework: caffe

batch: a single image

-> 원문 “입력 이미지와 해당되는 분할 맵을 사용하여 Caffe [6]의 확률적 경사 하강법 구현으로 네트워크를 훈련시킵니다. 언패딩된 합성곱 때문에 출력 이미지는 입력 이미지보다 일정한 테두리 폭만큼 작아집니다. 오버헤드를 최소화하고 GPU 메모리를 최대한 활용하기 위해, 우리는 큰 배치 크기보다 큰 입력 타일을 선호하여 배치를 단일 이미지로 줄입니다. 따라서 많은 수의 이전에 본 학습 샘플들이 현재 최적화 단계에서의 업데이트를 결정하도록 높은 모멘텀 (0.99)을 사용합니다.”

loss function

픽셀 단위 소프트맥스(pixel-wise soft-max) + 크로스 엔트로피(Cross Entropy) 손실 함수를 사용

softmax

$$p_k(\mathbf{x}) = \frac{\exp(a_k(\mathbf{x}))}{\sum_{k'=1}^K \exp(a_{k'}(\mathbf{x}))}$$

- $a(x)$ 는 특정 픽셀 x 의 k 번째 채널의 활성화 값(activation)입니다.
- K : 클래스 개수
- $p(x)$: 해당 픽셀이 k 번째 클래스일 확률을 나타냅니다.
- k 가 가장 높은 활성화 값을 가지는 경우 $p(x) \approx 1$, 다른 클래스의 확률은 거의 0입니다.

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

- $w(x)$: 특정 픽셀의 중요도를 조정하는 가중치
- $\ell(x)$: 픽셀 x 의 실제 클래스
- 신경망이 픽셀을 올바르게 분류할 확률이 높을수록 손실이 줄어듦

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

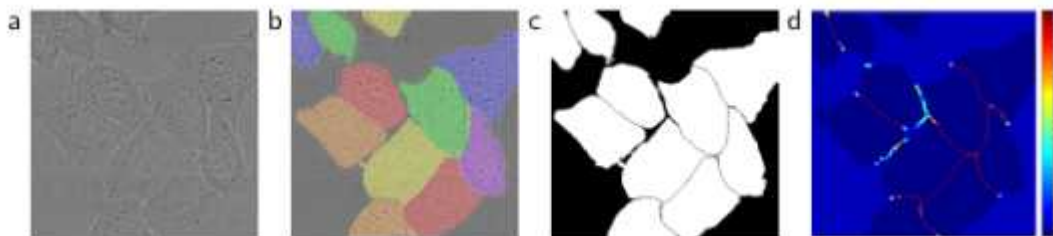
가중치 맵

- 일반적인 경우 손실 함수는 모든 픽셀을 동일하게 학습하지만 이미지 segmentation에서는 정확한 세분화를 위해 경계 픽셀이 중요
- 경계 픽셀은 중요, 배경 픽셀 상대적으로 덜 중요
- 중요한 픽셀에 더 높은 가중치 부여
- $w_c(x)$

$$frequency = \frac{PixelCount}{Total\ Number\ Of\ Pixels}$$

$$w_c(x) = inverse\ frequency = 1/frequency$$

- $d_1(x)$ 는 픽셀 x 에서 가장 가까운 세포 경계까지의 거리.
- $d_2(x)$ 는 픽셀 x 에서 두 번째로 가까운 세포 경계까지의 거리.
- 두 세포 사이에 간격이 좁을수록 큰 가중치, 간격이 높을수록 상대적으로 작은 가중치
- 실험에서는 $w_0=10$, $\sigma=5$ 픽셀로 설정
- 가중치 맵(Weight Map) 적용 → 서로 접촉하는 세포 경계에 가중치를 높여 학습 성능 향상.



(a) 원본 이미지 (Raw Image)

DIC (Differential Interference Contrast, 차등 간섭 대비) 현미경으로 촬영한 HeLa 세포의 원본 이미지.

이 이미지는 세포의 경계를 명확하게 표시하지 않음 → 따라서 **세분화(Segmentation)**가 필요함.

(b) 정답(Ground Truth) 오버레이

원본 이미지 위에 정답(Ground Truth) 세분화 정보를 덧씌운 것.

각 세포는 다른 색상으로 표시됨.

Ground Truth는 훈련 데이터로 사용되며, U-Net 모델이 학습해야 할 정답 역할을 함.

(c) 생성된 세분화 마스크 (Generated Segmentation Mask)

U-Net 모델이 예측한 세분화(Segmentation) 결과.

흰색(White) → 세포(Foreground)

검은색(Black) → 배경(Background)

이 마스크를 보면, 모델이 HeLa 세포를 어떻게 구분했는지 알 수 있음.

(d) 픽셀 단위 손실 가중치 맵 (Pixel-wise Loss Weight Map)

세포의 경계(Border) 부분에 높은 가중치(Weight)를 적용한 지도(Map).

푸른색(낮은 가중치) → 세포 내부.

빨간색(높은 가중치) → 세포 경계.

경계가 가장 중요한 영역이므로 경계를 더 정확하게 학습하도록 모델을 유도하는 역할을 함.

실험 결과 (Experiments)

U-Net은 전자현미경 및 광학 현미경 이미지 분할에서 뛰어난 성능을 보임

Ciresan et al. 의 슬라이딩 윈도우 방식(warping error 0.000420 rand error 0.0504)보다 훨씬 높은 정확도와 낮은 오류율 기록

데이터 증강(augmentation) 및 U-Net의 효율적인 구조 덕분에 적은 학습 샘플만으로도 강력한 성능 달성

ISBI 2015 챌린지에서 압도적인 성능 차이로 1위 기록

Table 1. Ranking on the EM segmentation challenge [14] (march 6th, 2015), sorted by warping 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

Table 2. Segmentation results (IOU) on the ISBI cell tracking challenge 2015.

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	0.9203	0.7756

우평균 IOU(Intersection over Union) 92% 달성했으며, 이는 두 번째로 성능이 좋은 알고

리즘(83%)보다 훨씬 높은 성능

결론

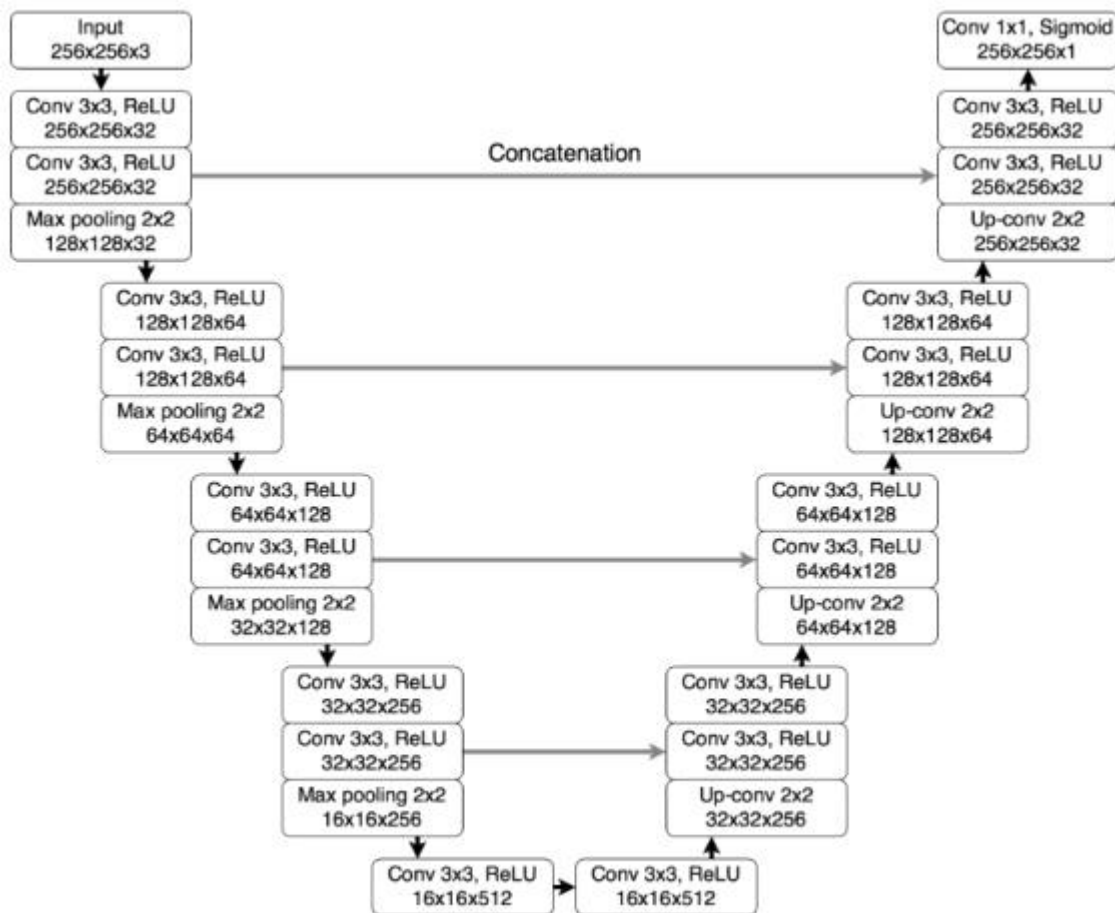
다양한 생물의학 영상 처리(Biomedical Image Processing) 문제에서 높은 성능을 보임.

데이터 증강(Data Augmentation) 기법, 특히 **탄성 변형(Elastic Deformation)**을 활용하여 적은 수의 주석(annotated) 이미지만으로도 효과적으로 학습 가능.

학습 시간(Training Time)이 매우 효율적:

****NVIDIA Titan GPU(6GB VRAM)****에서 약 10시간만에 학습 완료.

이는 기존 방법보다 빠르고 효율적임을 의미



unet 네트워크 구조

참고자료: 한요섭 블로그, unet 논문,