

# U-Net

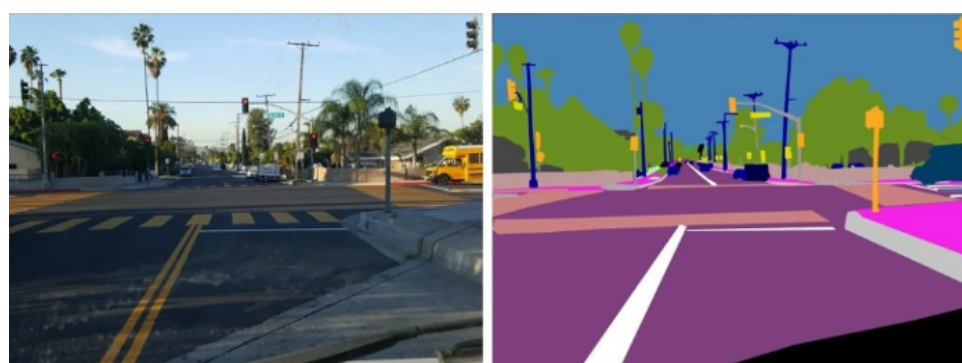
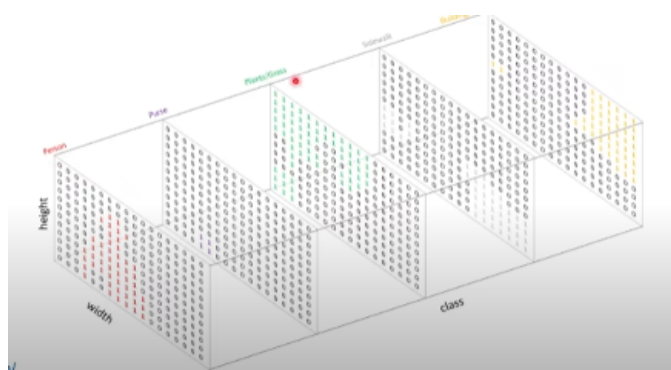
Image Segmentation의 대표적 논문

Semantic Segmentation을 하는데 이는 이미지 내에 있는 각 물체들을 의미 있는 단위로 분할하는 작업을 의미합니다. 목표는 한장의 분할 맵을 생성하는 것을 목표로 한다.

각 픽셀마다 N개의 클래스에 대한 확률을 가져야 하므로 정답은 ( 높이 X 너비 X N )형태를 갖습니다. 각 픽셀마다 원 핫 인코딩 형식을 따릅니다.

색칠을 하지는 않아도 되지만 객체의 픽셀 단위로 클래스 별로 구분을 하는 방법입니다.

일반적으로 픽셀마다 클래스 분류를 하기 위한 데이터 세트 생성 비용은 매우 비싸다. -> 픽셀마다 클래스 지정을 다 해주어야 하기 때문, 하지만 U-NET은 일반적인 CNN 분류 모델의 형식을 크게 바꾸지 않고 학습할 수 있는 장점이 있다.



## Abstract

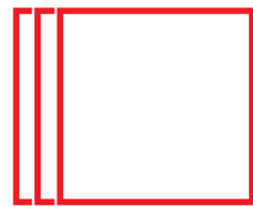
FCN 구조를 응용한 구조를 사용함.

Data augmentation을이용해 few labeled data를 가지고 적절한 학습이 가능하다.

Overlap-tile기법을 사용하여 resolution의 구애를 받지 않는다.

## Introduction

기존 classification의 방식은 이미지 속 하나의 객체가 어떤 클래스인지 구분해내는 것이었습니다. 하지만, 의학 쪽에서 한 이미지 안에 여러 개의 세포가 있는 경우가 있었고 이는 픽셀별로 클래스를 분류하는 classification방식이 필요했습니다.



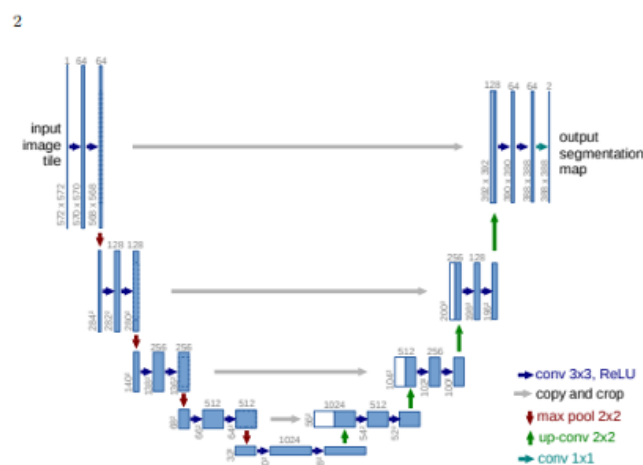
기존 방식의 sliding window



U-net의 patch 탐색 방식

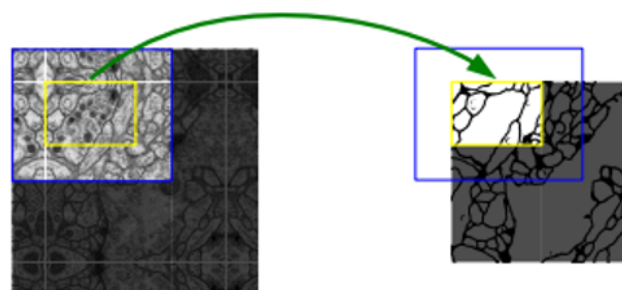
이를 위해서 sliding window 방식이 존재했지만, 이러한 방식은 중복되는 부분이 생겨 학습하는데 많은 시간이 사용되었습니다. 그리고 sliding window 방식은 픽셀 주변을 전부 이용하는 방식으로 주변의 문맥을 이용하여 예측하는데 사용하는데, 많은 문맥 정보를 담기 위해 큰 patch를 담을 시 max-pooling layer를 많이 걸치게 됩니다. 이를 해결하기 위해서 작은 patch를 사용한다면 적은 문맥 정보가 담겨 accuracy가 낮아지는 trade-off현상이 발생했습니다. 이를 해결하기 위해 U-NET에서는 patch 탐색 방식을 제안합니다.

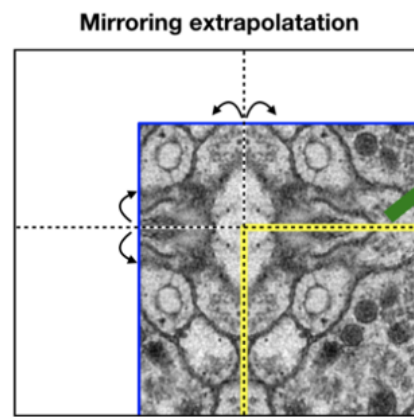
또한, U-NET 방식의 네트워크를 제안합니다.



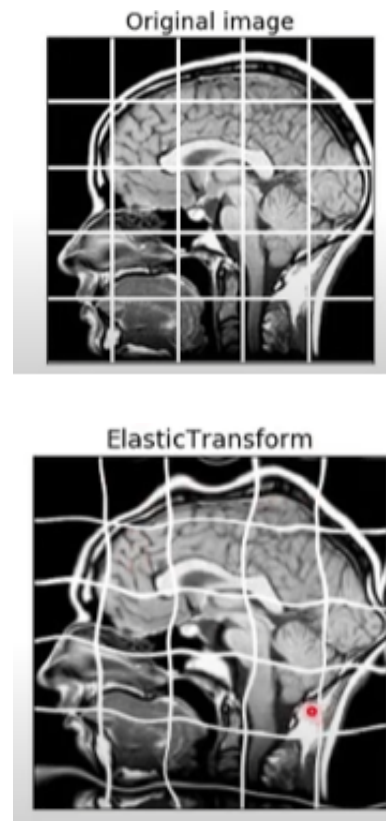
U-NET은 U자형으로 생긴 네트워크를 가져 U-NET으로 이름이 붙어졌습니다. 왼쪽 부분부터 순차적으로 진행이 되며 왼쪽은 해상도를 수축하며 채널을 확장시키는 Cotracting path, 오른쪽은 해상도를 확장시키며 채널을 줄이는 Expanding path입니다. 또한 가운데에서 수축 경로에서 확장 경로로 전환되는 Bottle Neck 구조를 가집니다.

맨 왼쪽에서 들어오는 이미지의 크기와 출력되는 이미지의 해상도 크기는 다르게 됩니다. 예시로, 아래 파란색 네모를 input에 넣으면 노란색 부분만을 결과로 도출해낼 수 있습니다. 따라서 노란색 부분의 결과를 알기 위해서는 조금 더 큰 input을 집어넣게 됩니다. 기존 CNN방식에서 정보가 담기지 않은 부분을 패딩 처리를 통해 크기를 맞춰주었지만, U-NET에서는 Mirroring extrapolation기법을 사용하여 이미지 안쪽을 거울로 반사된 것처럼 정보를 추가해주었다.



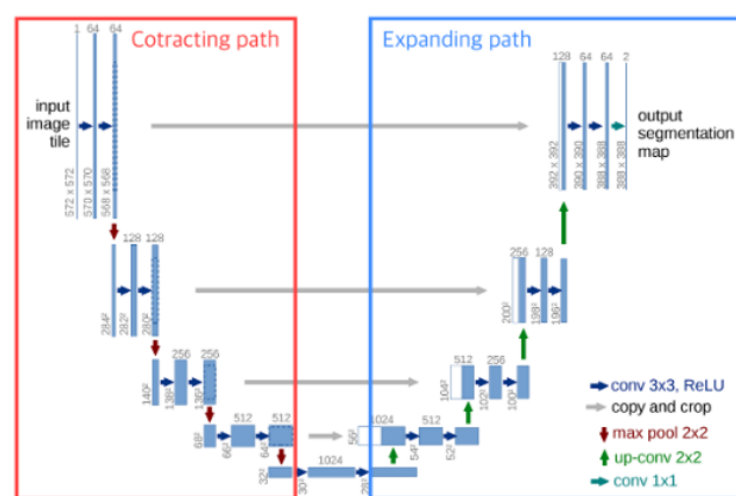


또한, 의료 분야의 경우 데이터가 많이 존재하지 않기 때문에 shift, rotation 방식을 사용하였고, 추가적으로 elastic deformation 방식을 사용하여 이미지의 비선형성을 주어 데이터를 생성하였다.

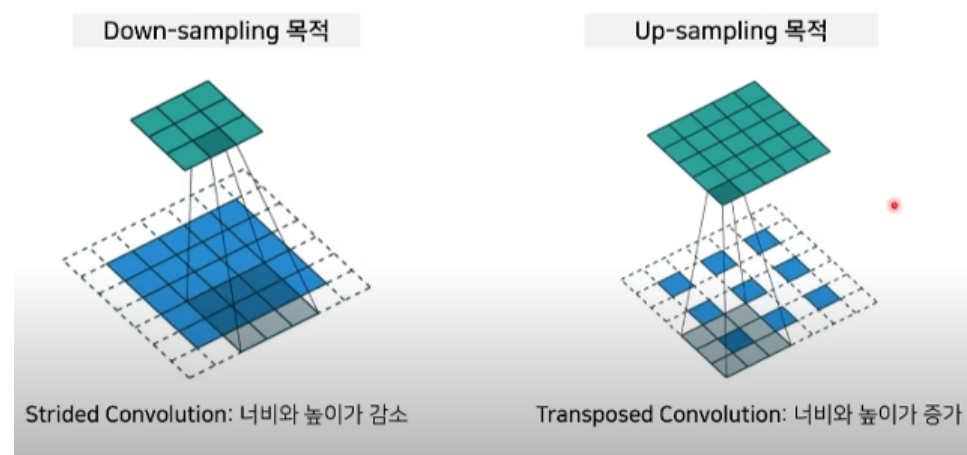


Cell segmentation에서 중요한 과제 중 하나는 동일한 class 개체 간의 분리였다. 특히, 세포의 경우 세포들끼리 인접한 경우가 많았고 이는 중요한 문제였다. U-NET은 경계를 포착하기 위해 픽셀 별로 가중치 학습을 사용합니다.

## Network Architecture



왼쪽은 해상도를 수축하며 채널을 확장시키는 Contracting path, 오른쪽은 해상도를 확장시키며 채널을 줄이는 Expanding path입니다. 또한 가운데에서 수축 경로에서 확장 경로로 전환되는 Bottle Neck 구조를 가집니다. 왼쪽에서의 진행 방식은 두 개의 3x3 convolution layer 이후 2x2 max pooling이 반복됩니다. 그리고 오른쪽의 방식은 2x2 up-sampling 방식이 사용되고 왼쪽 Contracting에서 나온 피쳐 맵과의 결합이 이루어집니다. 이 방식은 residual block과 비슷한 유형입니다. 그리고 다시 3x3 convolution layer가 이루어지며 이러한 과정이 반복되어지고 마지막에 1x1 convolution layer를 통해 output을 도출합니다. 마지막 1x1을 하는 이유는 원하는 클래스의 수에 따라 채널을 만들기 위함입니다. 기존 cnn과는 다르게 목적이 이미지의 class의 예측이 아닌 이미지 자체를 output으로 하기 때문에 fc layer를 진행하지 않습니다. 네트워크에는 총 23개의 conv layer가 존재합니다.



## Training

네트워크의 출력값은 픽셀 단위의 softmax에 의해 계산된다. 최종 feature map에 대한 픽셀의 예측값은

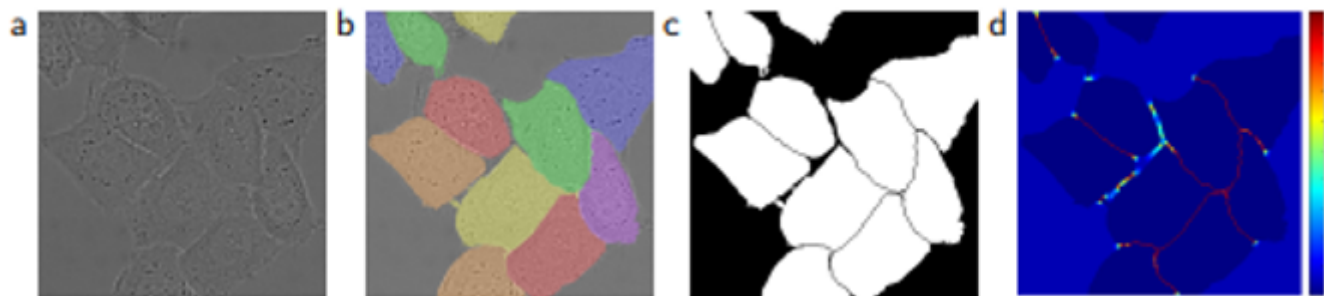
$$\hat{p}_k(\mathbf{x}) = \exp(a_k(\mathbf{x})) / \left( \sum_{k'=1}^K \exp(a_{k'}(\mathbf{x})) \right)$$

와 같이 계산된다. 각 채널별로 해당하는 클래스의 값 중 가장 큰 값으로 예측을 하게 된다.

Cell segmentation에서 중요한 과제 중 하나는 cell의 분리이다. 하지만 cell끼리는 근접해 있어 따로 분리를 하는데 신경 써야한다. U-NET은 개체 사이 작은 경계를 포착하기 위해

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

학습하는데 필요한 가중치를 설정한다. 한 개의 픽셀에서 가장 가까운 객체까지의 거리  $d_1(\mathbf{x})$ , 그 다음 객체 까지의 거리  $d_2(\mathbf{x})$ 를 이용하여 픽셀이 객체에 가까울 경우 가중치를 높여 구분을 잘해내도록 하였다.



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

위의 픽셀 예측값과 가중치 방식을 활용하여 이와 같이 로스를 정의한다.

그 밖에 Gaussian Distribution을 이용하여 모델의 파라미터를 초기화하고 학습합니다.

그리고 Optimizer에 Momentum 0.99를 부여하여 일관적이게 학습합니다.

## Experiments

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

Rank	Group name	Warping Error	Rand Error	Pixel Error
	<b>** human values **</b>	0.000005	0.0021	0.0010
1.	u-net	<b>0.000353</b>	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	<b>0.0582</b>
⋮				
10.	IDSIA-SCI	0.000653	<b>0.0189</b>	0.1027

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)	<b>0.9203</b>	<b>0.7756</b>

특히, PhC-U373데이터 세트에서는 평균 IOU 92%를 달성하였고, DIC-HeLa에서도 높은 점수와 2등 모델과의 점수 차이가 큼을 보여준다.

## Conclusion

Data augmentation을 통해 적은 양의 데이터로 학습이 가능하다.

Weighted loss를 활용해 객체를 효과적으로 분리한다.

기존, sliding window방식에서 탈피해 patch탐색 방식을 활용하여 빠른 속도를 가진다.

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