

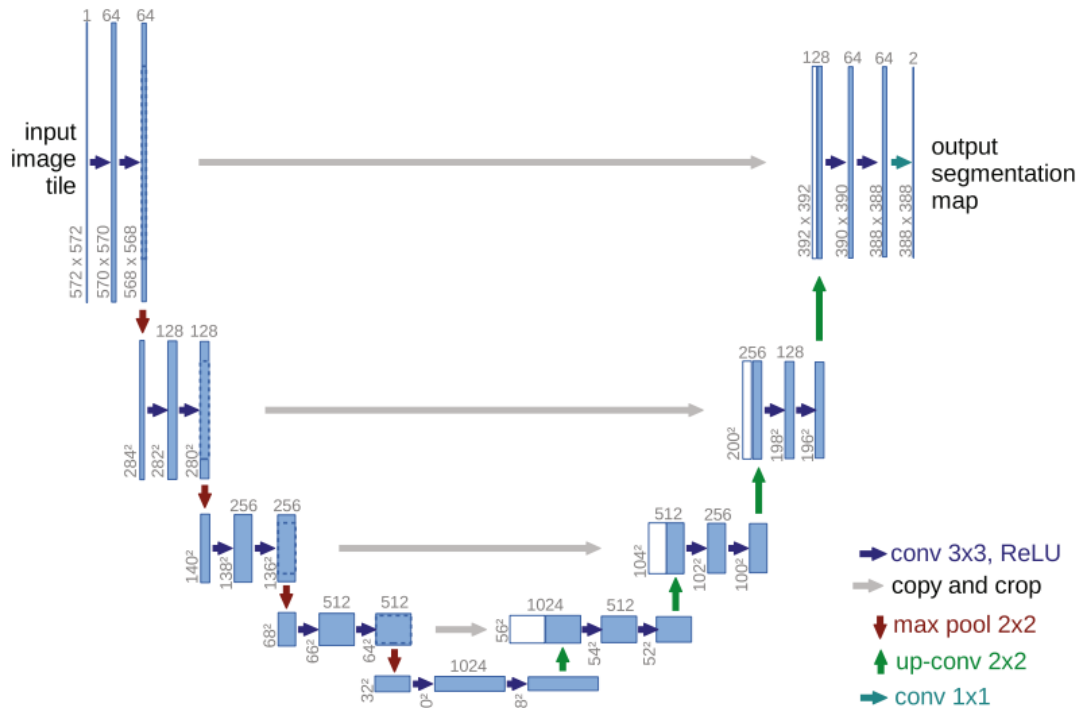


# [16주차] U-Net: Convolutional Networks for Biomedical Image Segmentation

## 1. Introduction

- 딥러닝의 발전
  - 최근 2년간 CNN은 여러 CV 작업에서 최고 성능을 기록함
  - CNN은 오래전부터 존재했지만, 훈련 데이터와 네트워크 크기의 한계로 인해 성공이 제한적이었음
  - Krizhevsky 등의 연구는 800만 개의 매개변수를 가진 8층 네트워크를 100만 개의 이미지를 가진 ImageNet 데이터셋으로 훈련하여 성공함
- 전형적인 CNN
  - 주로 분류 작업에 사용되어 이미지에 단일 클래스 레이블을 출력
  - biomedical image 처리와 같은 작업에서는 각 픽셀에 클래스 레이블을 할당해야 하는 경우가 많음
  - 대규모 훈련 이미지를 얻기 어려운 biomedical 작업에서는 Ciresan 등이 슬라이딩 윈도우 설정에서 네트워크를 훈련하여 각 픽셀의 클래스 레이블을 예측했으며, 이 네트워크는 ISBI 2012 EM 분할 챌린지에서 큰 성과를 거두었음
  - 그러나 슬라이딩 윈도우 방식은 느리고 중복이 많아 비효율적이며, 또한 큰 패치는 맥스 풀링 레이어를 많이 사용해 정확도가 떨어지고, 작은 패치는 맥락을 충분히 볼 수 없음
- 새로운 접근 방식
  - 최근의 방법들은 **여러 층의 특징을 활용**해 좋은 위치 지정과 맥락 사용이 가능하게 함
  - 본 논문에서는 "완전 컨볼루션 네트워크(fully convolutional network)"를 기반으로 한 U-Net 구조를 제안함

- U-Net은 적은 훈련 이미지로도 작동하며 더 정밀한 분할을 제공함
- U-Net 구조



- 수축 경로 + 대칭적인 확장 경로 → **U자형 아키텍처**
- 각 픽셀에 대한 전체 맥락을 이용해 **분할 지도**를 생성함

## ★ 분할 지도

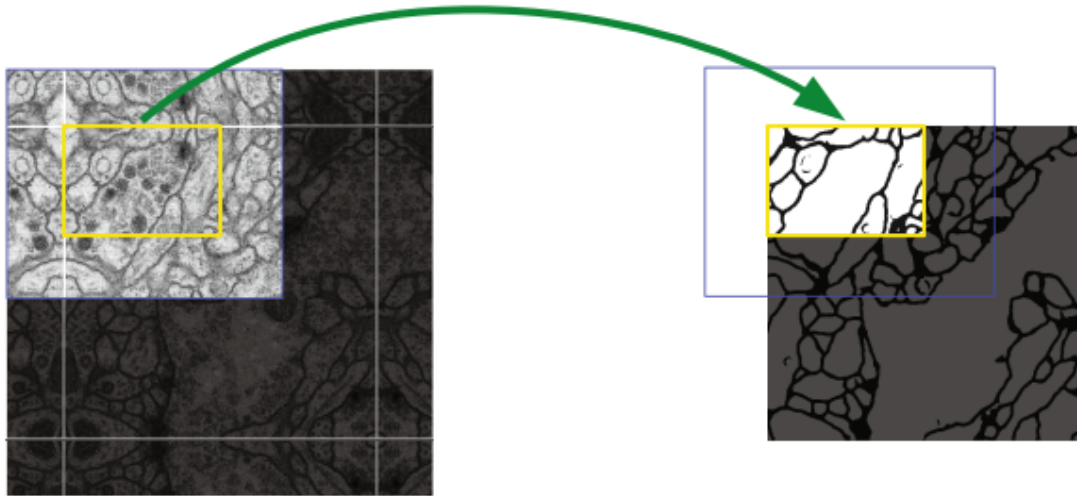
- 이미지의

**각각의 픽셀이 어떤 클래스에 속하는지**를 나타내는 지도

- ex) 의료 영상에서 종양을 분할할 때, 입력 이미지의 각 픽셀이 종양에 속하는지 아닌지를 판단하여 그 결과를 시각적으로 표현한 것

- 분할 지도는 보통

**원래 이미지와 같은 크기**를 가지며, 각 픽셀의 값은 해당 픽셀이 어떤 클래스에 속하는지를 나타냄



- 데이터 증강
  - 적은 훈련 데이터를 보완하기 위해 **탄성 변형(elastic deformation)**을 사용



### 탄성 변형

- 이미지에 랜덤한 변형을 적용하여 원본 이미지의 구조는 유지하되 **다양한 형태의 변형 이미지를 생성**하는 데이터 증강 기법

<탄성 변형의 과정>

1

**무작위 변형 필드 생성:** 먼저 이미지의 각 픽셀에 대한 이동 벡터를 생성함. 이 벡터는 무작위로 선택되며, 이미지의 각 부분을 얼마나 | 동시킬지 결정함

2

**이미지 변형:** 이 이동 벡터를 사용하여 이미지의 픽셀을 이동시켜 변형된 이미지를 생성함. 이 과정에서 이미지는 탄성이 있는 것처럼 변형됨

- ex) 얼굴 이미지를 탄성 변형하면 눈, 코, 입 등의 위치가 약간씩 이동하여 다양한 얼굴 이미지를 만들 수 있음. 하지만 원래의 중요한 특징(예: 얼굴의 구조)은 유지됨

- 네트워크가 변형에 대해 불변성을 학습할 수 있게 도와줌
  - 네트워크가 다양한 형태의 이미지를 보더라도 그 이미지가 원래 어떤 클래스에 속하는지를 정확하게 판단할 수 있도록 학습함

- ex) 다양한 변형된 이미지를 학습함으로써 실제 새로운 환자의 영상을 분석할 때도 정확한 결과를 도출할 수 있음
  - 접촉하는 객체의 분리
    - 같은 클래스의 접촉하는 객체를 분리하기 위해, 접촉하는 셀 사이의 배경 레이블에 큰 가중치를 부여하는 **가중 손실(weighted loss)**을 사용
      - 이미지에서 같은 클래스의 객체들이 서로 접촉하고 있을 때, 이를 개별 객체로 분리해야 함
      - 가중 손실: 손실 함수의 일부 영역에 더 큰 가중치를 부여하여 학습을 진행, 접촉하는 객체를 분리하는데 사용됨
  - 1 이미지에서 객체들을 레이블링
  - 2 서로 접촉하는 객체들 사이의 영역을 식별(일반적으로 객체 경계나 접촉 지점을 포함)
  - 3 접촉하는 객체 사이의 배경 레이블(경계 영역)에 높은 가중치를 부여
  - 4 네트워크가 이 경계 영역을 더 중요하게 학습
  - 5 결과적으로 접촉하는 객체들을 개별적으로 분리할 수 있게 됨
- ➡ 결과적으로 U-Net은 다양한 생의학적 분할 문제에 적용 가능하며, 기존 방법들의 한계를 극복하는 해결책을 제시함

## 2. Network Architecture

### 1 수축 경로

- 전형적인 합성곱 신경망(CNN) 구조
  - $3 \times 3$  합성곱(패딩 X)  $\rightarrow$  ReLU 활성화 함수  $\rightarrow$   $2 \times 2$  최대 풀링(stride 2, 다운샘플링)
  - 각 다운샘플링 단계에서 특징 채널(feature channels)의 수를 두 배로 증가시킴

### 2 확장 경로

- 수축 경로와 대칭을 이룸
  - 업샘플링  $\rightarrow$   $2 \times 2$  합성곱(up convolution, 특징 채널 수를 절반으로 줄임)  $\rightarrow$  수축 경로에서 잘라낸(cropped) 특징 맵과 연력(concatenation)  $\rightarrow$   $3 \times 3$  합성곱  $\rightarrow$  ReLU
  - 경계 픽셀의 손실로 인해 크롭핑이 필요함

- 1x1 합성곱: 각 64-component feature vector를 원하는 클래스 수로 매핑
- 총 합성곱 레이어 수: 23개

### 3 최종 레이어

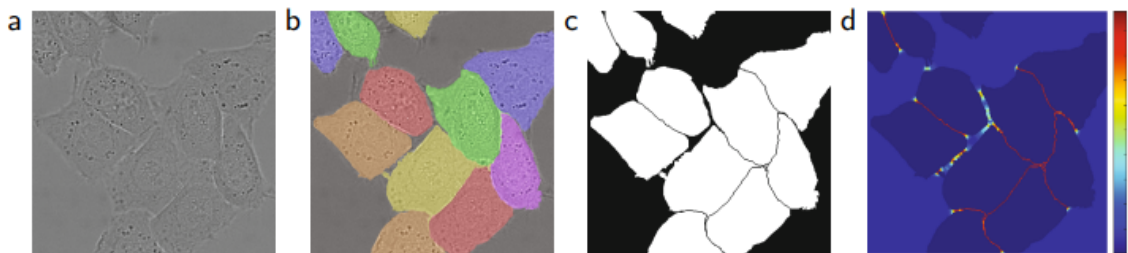
- 1x1 합성곱: 각 64-컴포넌트 특징 벡터를 원하는 클래스 수로 매핑
- 총 합성곱 레이어 수: 23개



#### Seamless tiling

- 출력 분할 맵을 원활하게 타일링하기 위해,  
**모든 2x2 최대 풀링이 짝수 크기의 층에 적용되도록** 입력 타일 크기를 선택하는 것이 중요
- 최대 풀링은 2x2 크기로 적용되기 때문에, 입력 이미지의 크기가 짝수여야 경계에서 문제가 발생하지 않음

## 3. Training



**Fig. 3.** HeLa cells on glass recorded with DIC (differential interference contrast) microscopy. (a) raw image. (b) overlay with ground truth segmentation. Different colors indicate different instances of the HeLa cells. (c) generated segmentation mask (white: foreground, black: background). (d) map with a pixel-wise loss weight to force the network to learn the border pixels.

- 훈련 데이터
  - 입력 이미지와 해당 분할 지도를 사용하여 네트워크를 훈련
  - 훈련은 Caffe 프레임워크의 확률적 경사 하강법으로 구현됨

- 출력 이미지 크기
  - 패딩이 없는 합성곱 연산 사용, 출력 이미지가 입력 이미지보다 작아짐
  - GPU 메모리를 최대한 활용하기 위해 큰 입력 타일을 사용하고 배치 크기를 줄여 한 번에 하나의 이미지를 처리함
  - → 배치 크기를 줄이는 대신 높은 모멘텀(0.99)을 사용하여 이전에 본 많은 훈련 샘플이 현재 최적화 단계에서 업데이트를 결정하도록 함
- 에너지 함수
  - 최종 특징 맵(feature map)에 대해 픽셀 단위 소프트맥스(soft-max)를 적용한 후 교차 엔트로피 손실 함수(cross entropy loss function)를 결합하여 에너지 함수를 계산함
  - **소프트맥스 함수**

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

- $a_k(x)$ : 픽셀 위치  $x$ 에서 특징 채널  $k$ 의 활성화 값
- $K$ : 클래스의 수
- $p_k(x)$ : 특정 픽셀 위치  $x$ 에서 클래스  $k$ 에 속할 확률
- **교차 엔트로피**

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

- $\Omega$ : 모든 픽셀 위치의 집합
- $c(x)$ : 픽셀  $x$ 의 실제 레이블
- $p_{\{c(x)\}}(x)$ : 픽셀  $x$ 에서 실제 레이블  $c(x)$ 에 대한 예측 확률
- $w(x)$ : 훈련에서 특정 픽셀에 더 많은 중요성을 부여하기 위해 도입된 가중치 맵
- $E$ 가 작을수록 모델의 예측이 실제 레이블과 더 일치하다는 의미, 이 값을 최소화하는 방향으로 학습
- **가중치 맵**

- 각 픽셀에 가중치를 부여하여 특정 영역이나 경계를 더 중요하게 학습하도록 만드는 맵
- 특정 클래스의 픽셀 수가 다른 클래스보다 훨씬 많거나, 서로 붙어있는 객체들 사이의 경계를 정확히 학습하는 것이 어려울 때 가중치 맵을 사용

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

- $w_c(\mathbf{x})$ : 특정 클래스의 픽셀 빈도를 균형 맞추기 위한 가중치
- $w_0 \cdot \exp\sim$ : 경계 가중치,  $w_0$ 은 경계 가중치의 크기를 조절하는 상수이며  $\sigma$ 는 경계 가중치가 얼마나 멀리 퍼질지를 결정하는 파라미터
  - $\sigma$ 값이 작으면 경계에서 멀어질수록 가중치가 급격히 줄어듦, **경계 근처의 픽셀에만 높은 가중치를 부여함**
  - $\sigma$ 값이 크면 경계에서 멀어져도 가중치가 천천히 줄어들어 **경계에서 더 멀리 떨어진 픽셀까지도 어느 정도 높은 가중치를 부여하게 됨**
- $d_1(\mathbf{x}), d_2(\mathbf{x})$ : 픽셀  $\mathbf{x}$ 에서 가장 가까운/두 번째로 가까운 셀의 경계까지의 거리
- 경계에 가까운 픽셀일수록  $d_1x$ 과  $d_2x$ 의 값이 작아지고  $\exp\sim$ 의 값이 커짐 → **경계에 가까운 픽셀에 더 높은 가중치를 부여하여 네트워크가 경계를 더 잘 학습하도록 만들**
- 초기 가중치 설정
  - 네트워크의 각 특징 맵이 대략적으로 단위 분산(분산이 1)을 가지도록 초기 가중치를 설정 → 역전파 과정에서 기울기 소실 혹은 폭발 문제가 줄어듦
  - 가우시안 분포에서 표준 편차  $\sqrt{2/N}$ 로 초기 가중치를 추출함,  $N$ 은 하나의 뉴런에 들어오는 노드 수

## 3.1 Data Augmentation

- 데이터 증강
  - 네트워크가 다양한 변형에 대해 잘 작동하도록 학습시키는 데 필수적
  - 특히 훈련 데이터가 적을 때 매우 중요
- 미세 현미경 이미지
  - 이동과 회전에 대해 불변성을 가져야 함

- 변형과 회색 값 변화에 대해 강건해야 함
- 랜덤 탄성 변형
  - 훈련 샘플을 랜덤하게 변형시키는 것이 효과적
  - 3×3 그리드에서 변위 벡터를 생성하고, 이를 통해 매끄러운 변형을 만들
    - 변위 벡터: 이미지의 각 지점이 얼마나 이동했는지를 나타내는 벡터
  - 변위는 가우시안 분포(표준 편차 10 픽셀)에서 샘플링됨
  - 각 픽셀의 변위는 비큐빅 보간법으로 계산됨
    - 이미지 변형 과정에서 각 픽셀의 새로운 위치를 결정할 때 비큐빅 보간법을 사용하여 매끄럽게 변형을 적용함
- 드롭아웃 층
  - 네트워크의 수축 경로 끝에 있는 드롭아웃 층은 추가적인 데이터 증강 역할을 함
    - 드롭아웃을 적용하면 매 학습 단계마다 네트워크의 구조가 달라짐 → 마치 여러 개의 작은 신경망(서브네트워크)을 학습하는 것과 같음
    - 특정 노드나 경로에 대한 의존성을 줄임 → 모델이 입력 데이터의 다양한 부분을 활용하여 학습하도록 만들
    - 데이터 증강 기법이 입력 데이터에 다양한 변형을 적용하여 모델이 더 일반화된 패턴을 학습하게 만드는 것과 유사함

## 4. Experiments

### 1 전자 현미경 이미지에서 신경 구조 세분화

- 목표: 전자현미경으로 촬영한 이미지에서 신경 구조를 세분화
- 데이터셋: Drosophila(초파리) 유충의 복부 신경 코드 이미지를 사용
  - 훈련 데이터: 30개의 이미지(각각 512x512 픽셀 크기)와 그에 대응하는 완전하게 주석된 세분화 지도(세포는 흰색, 막은 검정색)
  - 테스트 데이터: 공개되어 있지만 세분화 지도는 비공개, 예측된 막 확률 지도를 조직자에게 제출하여 평가
- 평가 방법: 10개의 다른 임계값에서 "warping error", "Rand error", "pixel error"를 계산
  - **warping error**: 세분화된 객체의 형태와 위치가 얼마나 잘 맞는지를 평가하는 지표, 객체의 경계가 얼마나 잘 일치하는지에 중점을 둠



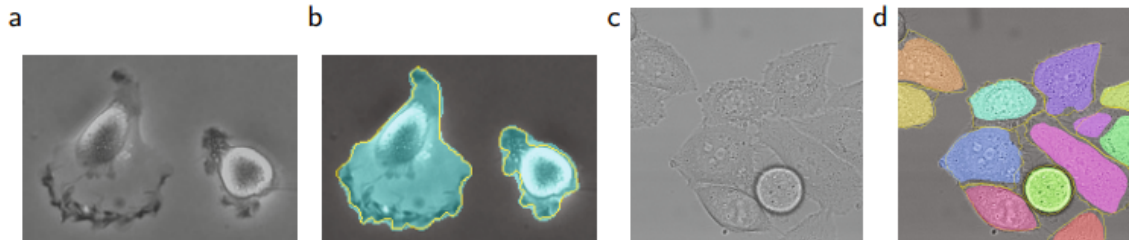
- **Rand error:** 세분화된 픽셀 쌍이 실제로 같은 객체에 속하는지 여부를 평가하는 지표, 두 픽셀이 같은 객체에 속하는지(positive pair) 또는 다른 객체에 속하는지(negative pair)에 대한 분류가 얼마나 정확한지를 측정
- **Pixel error:** 개별 픽셀 수준에서 세분화된 이미지와 실제 주석 이미지 간의 차이를 평가하는 지표, 이미지의 각 픽셀이 올바르게 분류되었는지에 중점을 둠
- 결과
  - U-Net은 입력 데이터를 7개의 회전된 버전으로 평균화하여 처리
  - Warping error: 0.0003529 (새로운 최고 기록)
  - Rand error: 0.0382
  - 이는 이전 최고 기록인 Ciresan et al.의 결과(0.000420의 warping error와 0.0504의 rand error)보다 우수함

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 [2]	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

## 2 광학 현미경 이미지에서 세포 세분화

- 목표: 광학현미경 이미지에서 세포를 세분화
- 데이터셋: ISBI Cell Tracking Challenge 2014 및 2015의 데이터
  - 첫 번째 데이터셋: PhC-U373
    - 내용: Phase contrast microscopy로 촬영한 Glioblastoma-astrocytoma U373 세포
    - 훈련 데이터: 부분적으로 주석된 35개의 이미지
    - 결과: 평균 IOU(Intersection over Union) 92%로, 두 번째로 좋은 알고리즘의 83%보다 우수함
  - 두 번째 데이터셋: DIC-HeLa

- 내용: Differential interference contrast (DIC) microscopy로 촬영한 HeLa 세포
- 훈련 데이터: 부분적으로 주석된 20개의 이미지
- 결과: 평균 IOU 77.5%로, 두 번째로 좋은 알고리즘의 46%보다 우수함



**Fig. 4.** Result on the ISBI cell tracking challenge. (a) part of an input image of the “PhC-U373” data set. (b) Segmentation result (cyan mask) with manual ground truth (yellow border) (c) input image of the “DIC-HeLa” data set. (d) Segmentation result (random colored masks) with manual ground truth (yellow border).

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

## 5. Conclusion

- **U-Net 아키텍처:** 다양한 생의학적 세분화 작업에서 뛰어난 성능을 발휘
- 탄성 변형을 이용한 데이터 증강 덕분에 이미지 데이터가 매우 적어도 괜찮으며, NVidia Titan GPU(6 GB)에서 10시간 정도의 훈련 시간만 필요함(매우 합리적)
- 유넷 아키텍처는 의생명공학 뿐만이 아니라 더 많은 작업에 쉽게 적용될 수 있을 것

## 논문에 대한 의견 및 의문점(꼭지)

➡ 유넷이 특정 생의학적 이미지 데이터셋에 대해 좋은 성능을 보이는 것은 본 논문을 통해 잘 알려져 있지만, 다른 종류의 데이터셋에서도 동일한 성능을 발휘할 수 있는지에 대한 논의가 필요하다고 생각함. 예를 들어, 다른 의료 분야나 비의료 분야에서의 적용 가능성과 성

능에 대한 연구가 중요할 수 있음. 세포나 신경구조에 대해서는 성능이 뛰어난 것으로 나타나더라도, 다른 분야의 이미지에서는 객체 분할이 어려워지는 가능성을 생각해야 할 것 같음