



9주차 논문 리뷰: U-Net: Convolutional Networks for Biomedical Image Segmentation

0. Abstract

- 데이터 증강(Data Augmentation)을 적극적으로 활용하여 적은 수의 annotated training samples을 효율적으로 사용할 수 있는 **네트워크 구조와 훈련 전략**을 제시
→ 문맥 정보를 포착하는 **contracting path** + 정확한 로컬라이제이션을 가능하게 하는 **symmetric expanding path**
- 소수의 이미지로도 끝까지 학습될 수 있으며, 이전 방식(슬라이딩 윈도우 기반의 컨볼루션 네트워크)보다 높은 성능

1. Introduction

- Deep Convolutional Networks는 여러 시각 인식 과제에서 기존의 기술보다 높은 성능을 보임
- 컨볼루션 신경망의 일반적인 용도: 이미지 분류로, 이미지 하나당 하나의 클래스 레이블을 예측하는 작업
- 생의학 이미지 처리와 같은 특정 분야에서는 **Localization** 작업이 중요
→ **Localization**: 특정 이미지 내에서 각 픽셀이나 작은 영역에 고유한 레이블을 부여하여 위치를 파악하는 기술

→ 의료 영상에서 수천 개의 학습 이미지를 얻는 것은 쉽지 않기 때문에 효율적인 학습이 필수적

1) U-Net 구조

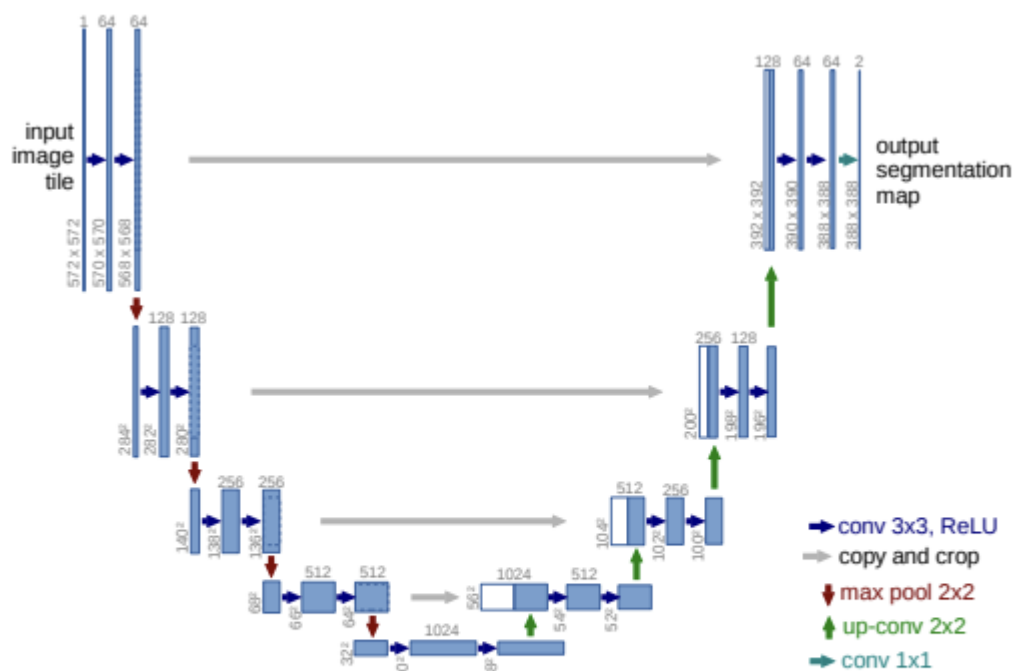


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

- 32×32 픽셀 이미지 타일을 예시로 사용하여 픽셀 수준의 분할(Segmentation)을 수행
- 각 픽셀에 대한 클래스 레이블을 예측하기 위해 **슬라이딩 윈도우(sliding window)**를 통해 지역 정보를 입력으로 받아 활용
- 일반적인 컨볼루션 네트워크의 contracting path와 반대로 **expansive path**를 결합
 - 이미지의 해상도를 높이면서 로컬라이제이션을 수행할 수 있도록
 - 신경망이 넓은 문맥 정보와 세밀한 로컬라이제이션 정보를 동시에 다룰 수 있도록



contracting path: 이미지의 특징을 추출하면서 점차 해상도를 낮춰 이미지의 요약된 정보 만들기

expansive path: 다시 해상도를 높이며 원본 이미지의 크기로 복구하고, 각 픽셀에 대한 분류를 수행

2) Overlap-Tile Strategy

- U-Net은 큰 이미지를 처리할 때 경계선 문제를 해결하기 위해 overlap-tile 전략 사용
- 입력 이미지의 일부가 부족할 경우 그 부분을 **대칭적으로 채워 넣어** 이미지 경계에서 발생할 수 있는 문제를 방지
- 생의학 분야에서 큰 이미지를 분할할 때 유용
- GPU 메모리의 제한으로 인해 대용량 이미지 전체를 처리하기 어려울 때 적합

3) 데이터 증강과 훈련 방법

- 생의학 이미지에서는 훈련 데이터가 적은 경우가 많기 때문에, U-Net은 데이터 증강 (**Data Augmentation**) 기법을 사용해 데이터를 변형하여 신경망이 다양한 변형에 대해 학습할 수 있도록 함
- **Elastic Deformation**을 적용하여 실제 조직이 변형되는 모습을 시뮬레이션
→ 모델이 다양한 형태에 대해 학습하고 일반화할 수 있게 하여, 적은 데이터로도 효과적인 학습을 가능하게 함

2. Network Architecture

- [contracting path + expansive path]로 구성됨
- contracting path: 왼쪽 & expansive path: 오른쪽
- 전형적인 컨볼루션 신경망의 구조를 따름

- **contracting path**

- 3×3 크기의 컨볼루션 연산 반복
- 각 컨볼루션 후에는 ReLU 활성화 함수와 2x2 크기의 맥스 풀링(max pooling) 연산 적용
- 맥스 풀링은 다운샘플링(downsampling)을 위해 적용됨

- **expansive path**

- 특성 맵을 upsampling하는 단계로 시작
- 2×2 크기의 up-convolution을 통해 특성 채널 수를 절반으로 줄임
- contracting path에서 가져온 해당 특성 맵을 expansive path와 연결
- 3×3 컨볼루션을 두 번 수행하며 ReLU 활성화 함수를 적용

⇒ 경계 픽셀 손실로 인해 잘린 특성 맵의 크기가 줄어들기 때문에 **cropping** 필요

- 마지막 레이어: 1x1 convolution이 적용되어, 각 64차원의 feature 벡터를 원하는 클래스 수에 맞게 변환
- 전체 네트워크는 총 23개의 convolution레이어로 구성
- 출력 세그멘테이션 맵을 원활하게 타일링하기 위해서는, 입력 이미지의 크기를 적절하게 설정해야 함
 - ex) 네트워크가 2x2 크기의 맥스 풀링을 할 때 이미지 크기가 짝수일 경우에는 정확히 나누어 떨어지지만, 홀수 크기일 경우에는 경계에서 문제가 발생해 출력이 매끄럽지 않을 수 있음
 - 모든 2x2 맥스 풀링 연산이 짝수 크기의 x, y 축을 가진 레이어에 적용되도록 설정



ReLU: 음수를 0으로 변환하고 양수는 그대로 유지하는 활성화 함수

Max Pooling: 입력 특성 맵에서 특정 영역의 최대값을 선택해 해상도를 줄이는 연산으로, 다운샘플링과 중요한 정보 추출에 사용

Up-Convolution: 업샘플링을 통해 이미지의 해상도를 높이는 과정으로, 픽셀 레벨의 세밀한 정보를 재구성할 때 사용

Cropping: 경계에서 발생하는 정보 손실을 보완하기 위해 feature 맵의 일부를 잘라내는 과정

3. Training

- 입력 이미지와 그에 대응하는 segmentation maps을 사용하여 네트워크를 훈련
 - **Stochastic Gradient Descent(확률적 경사 하강법)**을 적용
 - 네트워크 구현에는 **Caffe 라이브러리** 사용
- padding 없는 컨볼루션을 사용하기 때문에 출력 이미지의 크기가 입력보다 작아짐
- GPU 메모리 사용량을 줄이기 위해 큰 크기의 타일 입력 또는 큰 배치 크기를 사용하여 단일 이미지로 배치 크기를 줄임

⇒ 이전 단계의 많은 학습 샘플을 활용해 현재 최적화 단계의 업데이트에 반영 가능

- **Energy Function:** 최종 특성 맵에 대해 각 픽셀마다 소프트맥스(soft-max)와 교차 엔트로피(cross-entropy) 손실 함수를 사용하여 계산
 - 소프트맥스 함수: 픽셀 위치 x 에서 클래스 k 의 예측 확률 $p_k(x)$ 를 제공
 - $a_k(x)$: 픽셀 위치에서 k 번째 클래스의 활성화 값
 - 교차 엔트로피 손실: 각 위치에서 예측 값이 실제 레이블에서 벗어난 정도를 페널티로 부여

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

→ E: 손실값

→ $w(x)$: 특정 픽셀의 중요도를 반영하는 가중치 맵

→ $p_l(x)$: 소프트맥스를 통해 얻은 픽셀 x 에서의 클래스 l 확률

- **weight map**: 특정 클래스 픽셀의 빈도를 보정하거나 세포 간 경계와 같은 중요한 영역을 학습하도록 도움

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

→ $w_c(x)$: 클래스 불균형을 보정하는 가중치

→ w_0 : 경계 가중치를 조절하는 상수 (예: 실험에서 10으로 설정)

→ $d_1(x)$, $d_2(x)$: 가장 가까운 세포 및 두 번째로 가까운 세포의 경계까지의 거리

→ σ : 거리 가중치의 표준편차, 경계 영역을 조정하는 파라미터

• 초기화 & 가중치 설정

- 심층 신경망에서 여러 컨볼루션 층과 서로 다른 경로를 사용할 때 가중치 초기화 매우 중요
- 네트워크 일부가 비활성화되거나 과도하게 활성화되는 현상을 방지하기 위해, 평균이 0이고 분산이 $1/\text{root}(N)$ 인 정규 분포에서 가중치를 초기화

• Data Augmentation

- 적은 학습 샘플로 네트워크가 불변성(invariance)과 robustness를 학습하게 하는 데 필수적
 - 현미경 이미지의 경우, shift invariance과 rotation invariance이 필요 & 변형과 회색 값 변화에 대한 robustness 중요
- ⇒ 학습 샘플이 많지 않을 때, 네트워크가 다양한 변형에 대해 잘 대응하도록 도와줌
- random elastic deformations을 학습 샘플에 적용: 소수의 주석이 달린 이미지로도 segmentation 네트워크를 훈련하는 데 중요한 개념으로 작용
 - 3×3 그리드에서 랜덤 변위 벡터를 사용하여 매끄러운 변형을 생성
 - 변위는 10픽셀의 표준편차를 가진 가우시안 분포에서 샘플링됨

- 각 픽셀의 변위는 bicubic interpolation 통해 계산됨
- contracting path 끝에 드롭아웃 레이어를 추가하여 추가적인 암시적 데이터 증강을 수행
 - ⇒ 네트워크가 과적합을 방지하고 더 다양한 데이터 패턴에 대해 강인해지도록 도움



- **Shift Invariance:** 이미지가 이동해도 네트워크가 동일한 출력을 내는 성질
- **Rotation Invariance:** 이미지가 회전해도 네트워크가 동일하게 인식하는 성질
- **Elastic Deformation:** 이미지를 탄성 변형시켜 다양한 형태의 데이터로 변환하는 방법으로, 주로 생의학 이미지 증강에 사용
- **Bicubic Interpolation:** 변형된 이미지를 보다 매끄럽게 보정하는 방법으로, 주어진 점 주위의 16개 픽셀을 사용하여 값을 보간
- **Dropout Layer:** 학습 시 일부 뉴런을 무작위로 비활성화하여 과적합을 방지하고 모델의 일반화 능력을 향상시키는 기법

4. Experiments

- U-Net을 세 가지 다른 Segmentation 과제에 적용
1. 전자 현미경 기록에서 신경 구조를 Segmentation
 - 데이터셋: ISBI 2012 EM Segmentation 챌린지에서 제공한 데이터
 - 512×512 픽셀 크기의 이미지 30장
 - 평가 지표: "워핑 에러(warping error)", "랜덤 에러(Rand error)", "픽셀 에러(pixel error)"
 - 입력 데이터의 7가지 회전 버전을 평균하여 추가 전처리나 후처리 없이 **워핑 에러 0.0003529, 랜덤 에러 0.0382**
 - 랜덤 에러 측면에서는 더 나은 성능

2. ISBI 셀 추적 챌린지 2015에 참여하여 "PhC-U373" 데이터셋과 "DIC-HeLa" 데이터셋을 이용한 세포 Segmentation 작업

- PhC-U373 데이터셋에는 Glioblastoma-astrocytoma 세포가 포함됨
→ 위상 대비 현미경 이미지를 사용해 기록
- U-Net은 평균 IOU(Intersection over Union) 92%를 기록했으며, 이는 기존 최고 알고리즘보다 9% 높은 성과
- DIC-HeLa 데이터셋에는 **DIC 현미경**을 통해 기록된 HeLa 세포가 포함됨
→ 77.56% IOU를 달성 = 기존 최고 알고리즘보다 46% 이상 높은 성과



- **EM Segmentation Challenge:** 전자 현미경(EM)을 사용한 이미지 Segmentation 과제를 다루는 챌린지로, 생의학 이미징에서 중요한 역할
- **Warping Error:** 예상되는 위치와 실제 위치의 차이를 측정하는 오류로, Segmentation 경계가 얼마나 정확하게 잡혔는지 평가하는 지표
- **Rand Error:** Segmentation에서 클러스터 간 관계의 일치도를 측정하는 오류로, 세포나 구조를 정확하게 분리하는지를 평가
- **IOU (Intersection over Union):** 예측된 Segmentation 영역과 실제 영역이 얼마나 겹치는지 측정하는 지표로, 높을수록 성능이 좋음을 의미
- **DIC (Differential Interference Contrast) Microscopy:** 차등 간섭 대비 현미경으로, 투명한 세포 구조를 관찰할 때 사용

5. Conclusion

- U-Net 아키텍처는 다양한 생의학 Segmentation 작업에서 매우 우수한 성능을 보여줌
 - **elastic deformation을 사용한 데이터 증강** 덕분에 가능했으며, 이를 통해 **네트워크 크기가 적은 수의 주석이 달린 이미지로도 효과적으로 학습 가능**
 - NVIDIA Titan GPU(6GB 메모리)를 사용했을 때 학습 시간은 약 10시간 ⇒ 매우 합리적인 수준의 학습 시간

- 논문에서는 Caffe를 기반으로 한 전체 구현 및 학습된 네트워크를 제공



- **lastic Deformation:** 데이터 증강 기법 중 하나로, 이미지를 탄성적으로 변형시켜 다양한 형태의 데이터를 생성하여 네트워크가 다양한 변형에 강인성을 가지도록 학습
- **Caffe:** 딥러닝 프레임워크 중 하나로, 특히 이미지 처리 작업에서 자주 사용