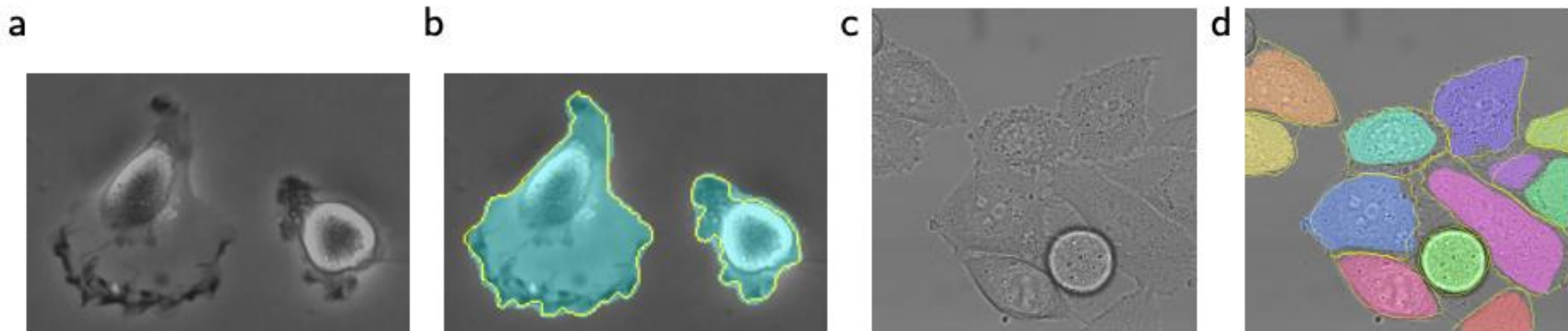


시각지능 A – Semantic Segmentation

U-Net

U-Net 개요

- ISBI cell tracking challenge 2015 우승
 - 현미경으로 촬영한 세포 이미지 Segmentation
- 적은 데이터셋을 Augmentation으로 극복함
- Contracting path와 Expanding path 구조로 정밀한 추론
- 512 * 512 Mask를 1초도 안 되는 빠른 시간 내에 출력

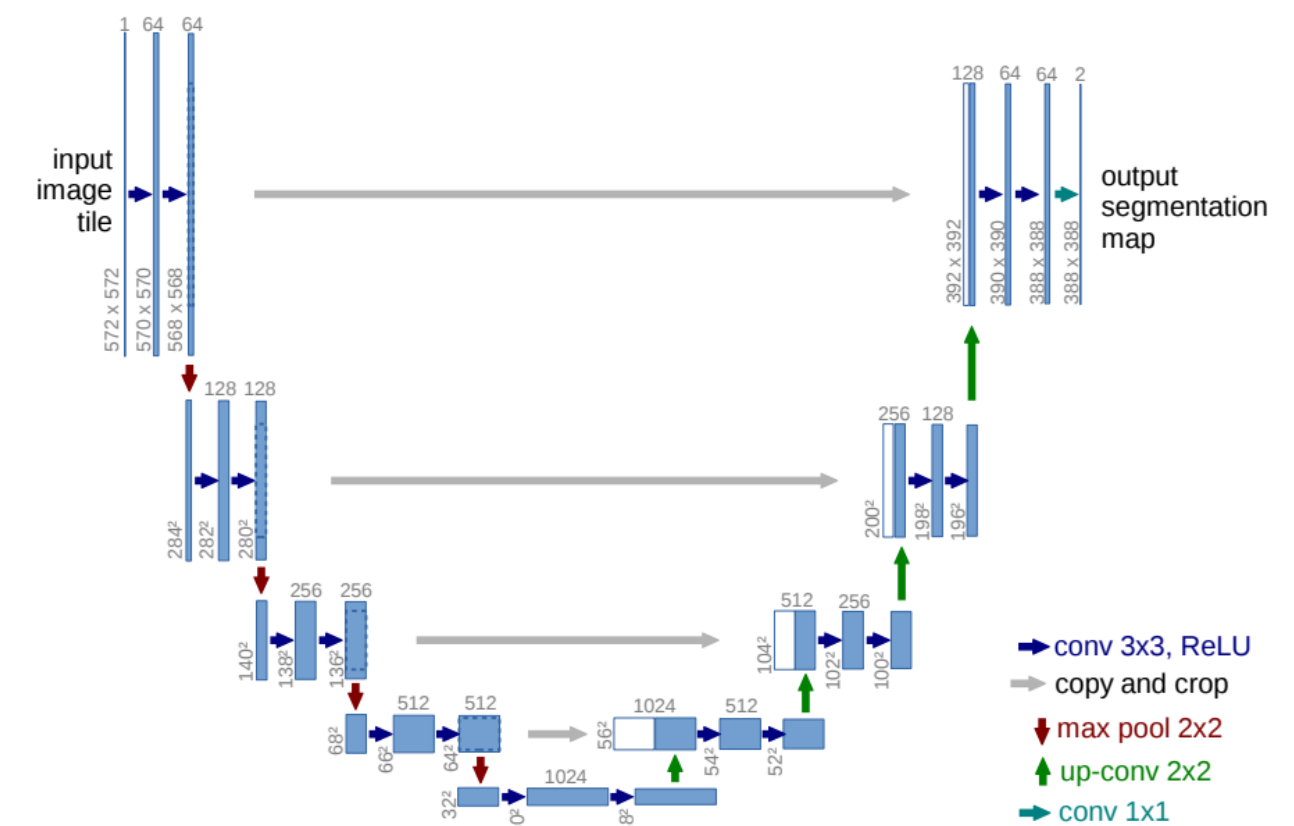
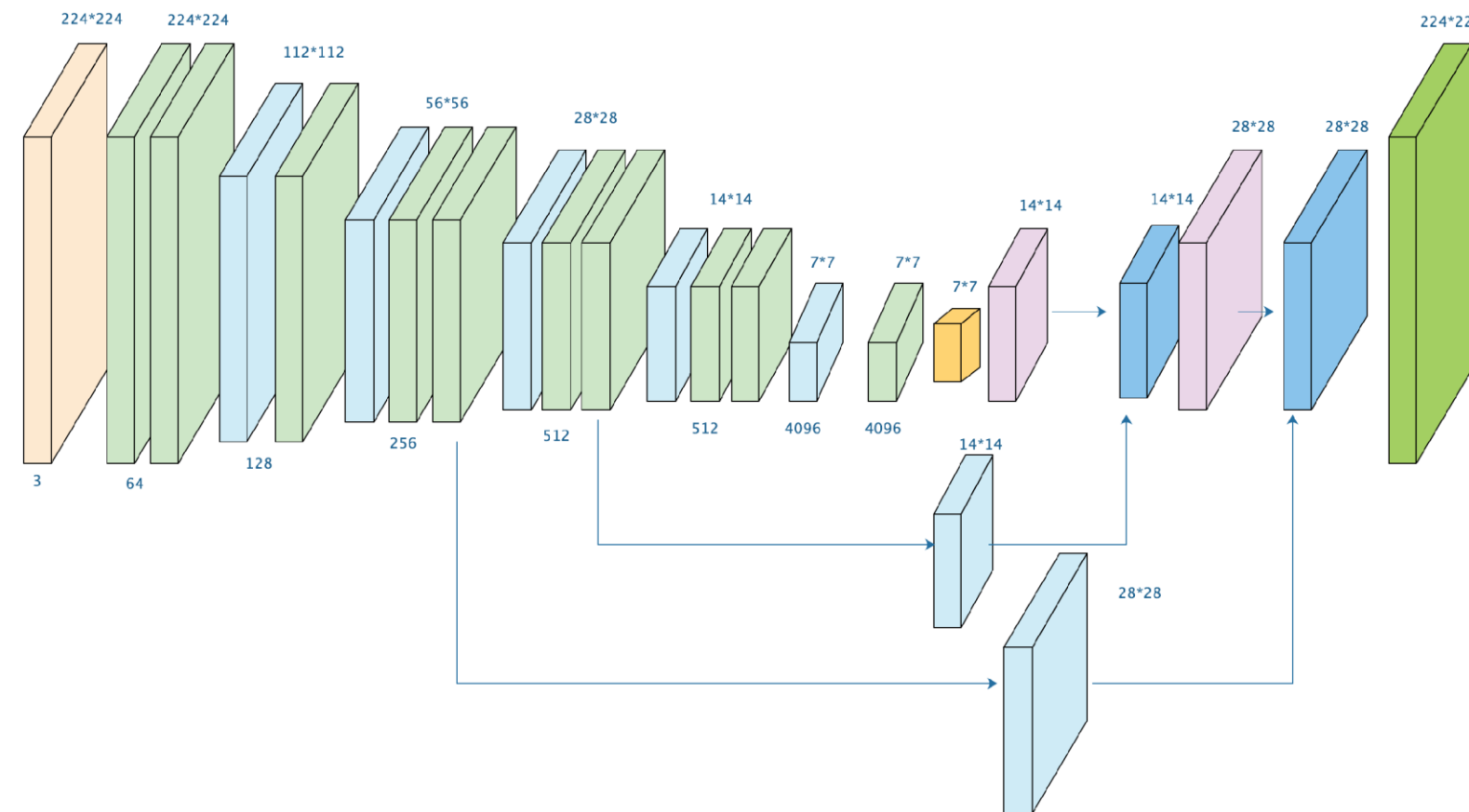


FCN to U-Net

- U-Net은 FCN의 구조를 성공적으로 개량한 모델
 - Convolutionalization: 모든 층을 Convolution layer 기반으로 설계
 - Deconvolution: Feature map을 Upsampling 하기 위해 Decoder 구조 사용
 - Skip architecture: 지역(Appearance) 정보와 추상적(Semantic) 정보를 모두 고려하여, 정밀한 Segmentation mask 생성

U-Net - FCN과의 차이점

- Upsampling 과정에서, Feature map의 Channel 수를 확장
 - FCN: 21(Class 수)
 - U-Net: 1024→512 →256 →128 →64 →2
- 입력 이미지와 출력 마스크의 크기가 다름
 - FCN
 - In: (224, 224, 3)
 - Out: (224, 224, 21)
 - U-Net
 - In: (572, 572, 1)
 - Out: (388, 388, 2)

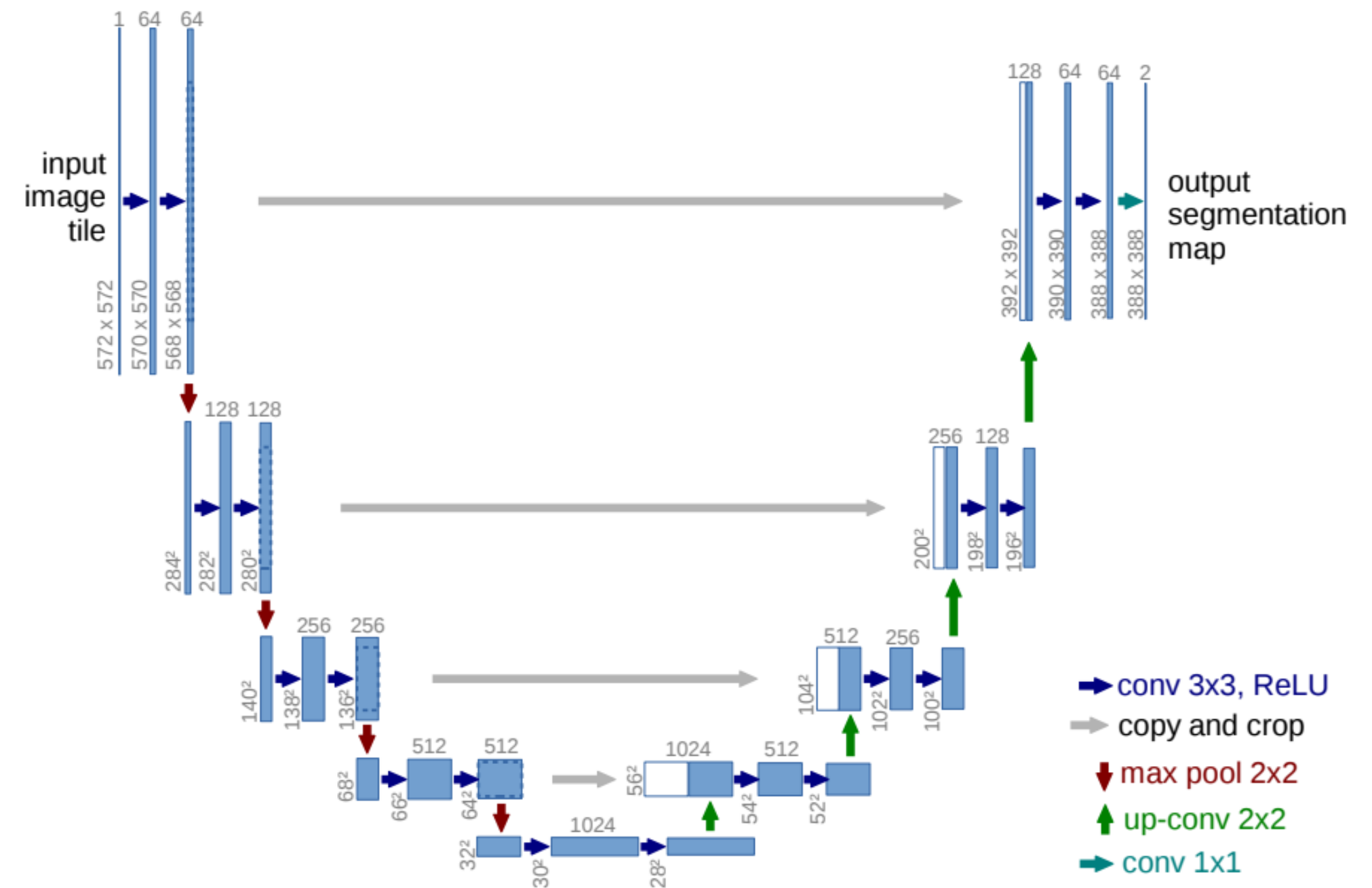


U-Net – 이전 알고리즘과의 차이점

- Sliding-window approach
 - 이미지에서 각 픽셀의 주변 패치를 네트워크에 입력으로 주어 해당 픽셀의 클래스를 예측하는 방식
 - 각 패치마다 네트워크를 개별적으로 실행해야 하기 때문에, 한 장의 이미지를 처리하는 데 시간이 매우 오래 걸림
 - 네트워크가 한 번에 처리하는 패치의 크기가 제한적이므로, 넓은 문맥 정보를 활용하기 어려움
- Fully Convolutional Networks(FCN):
 - Upsampling을 통한 고해상도 복원이 완벽하지 않았고, 이로 인해 정확한 경계 예측이 어려움

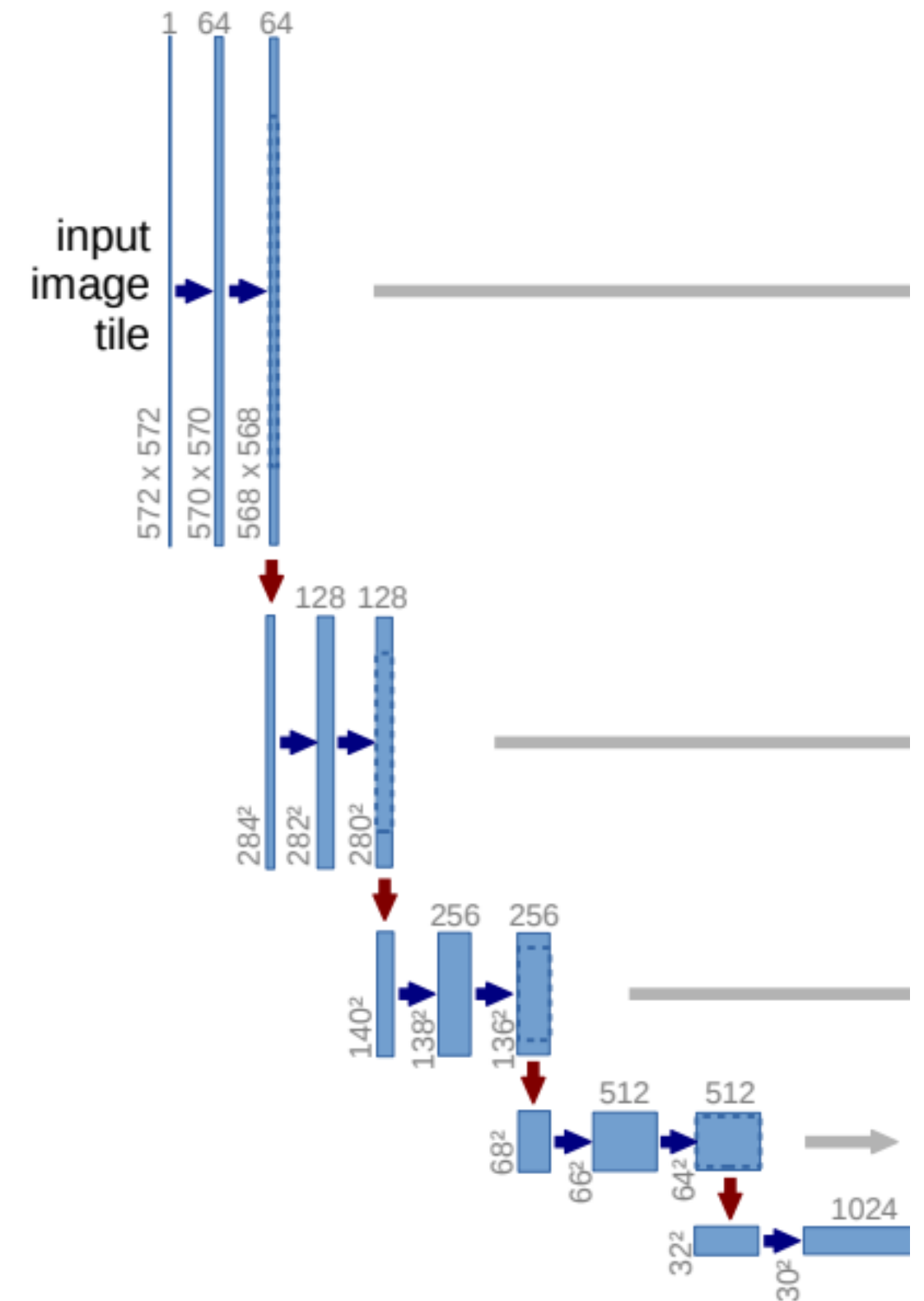
U-Net의 모델 구조

- 대칭점을 기준으로 기능 단위가 나뉨
 - 좌측: Contracting path
 - Backbone(Encoder)
 - 이미지 Feature extraction
 - 우측: Expansive path
 - Head(Decoder)
 - 이미지 Feature map을 Upsampling
 - 두 부분을 이어주는 3 개의 Skip architecture



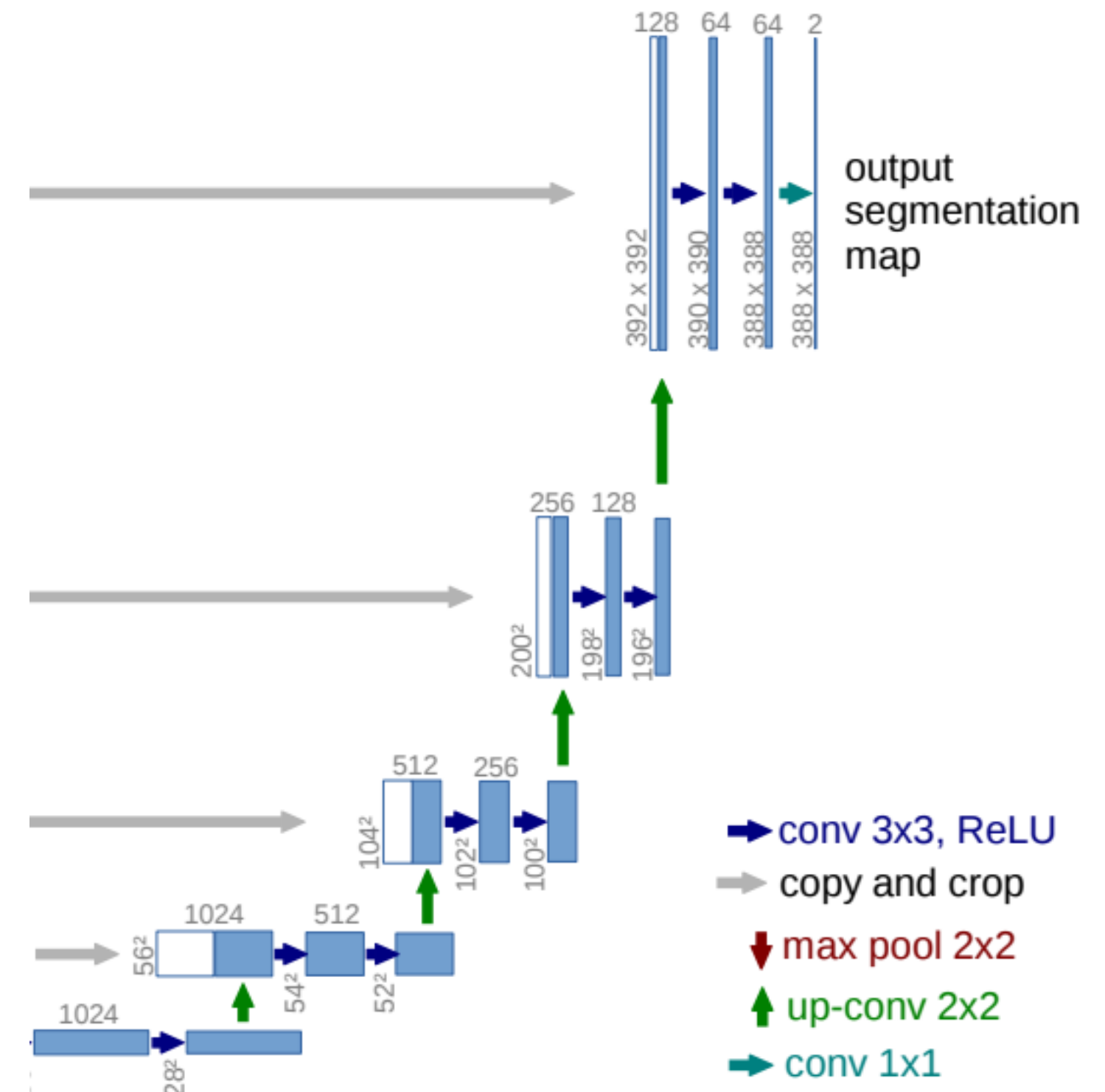
U-Net의 모델 구조

- Contracting path
 - 일반적인 CNN 구조
 - 아래 구조가 반복
 - 입력 이미지/Feature map에 3x3 Conv 두 번 적용(Padding 없음)
 - ReLU 활성화 함수
 - 2x2 Max pooling으로 이미지 크기 반감
 - 매 구조가 반복될 때마다, Feature map을 Expansive path로 전달



U-Net의 모델 구조

- Expansive path
 - 아래 구조가 반복
 - Upsampling을 통해 Feature map의 크기(w, h)를 두 배 증폭
 - Contracting path의 중간 Feature map과 결합(Skip connection)
 - 3x3 Conv 두 번 적용(Padding 없음)
 - 마지막 Feature map에 대하여 1x1 Conv 적용(Bottleneck)
 - Channel depth를 조정하기 위함

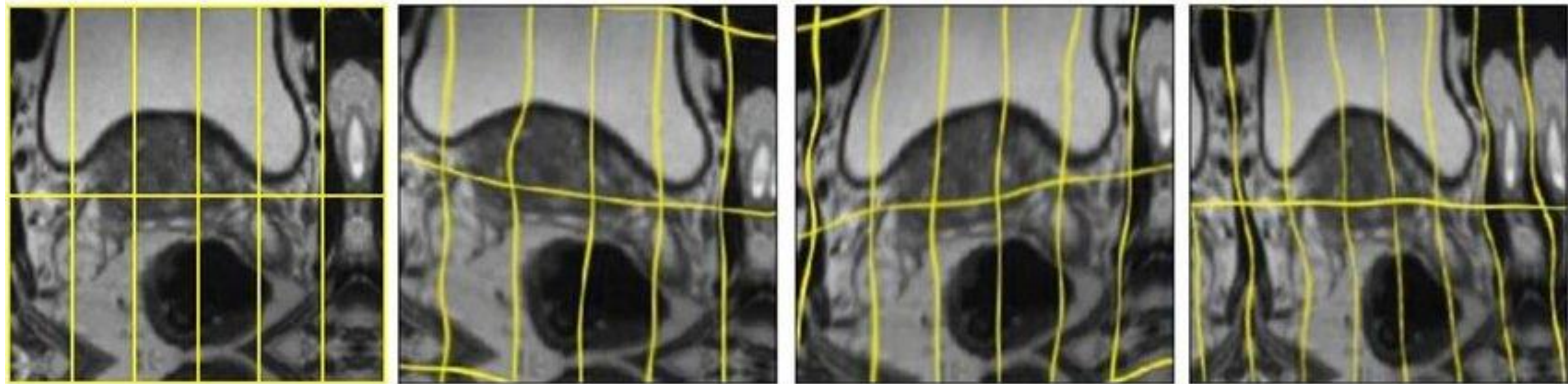


Data Augmentation

- 해당 대회에서 제공된 이미지의 해상도는 높지만, 이미지의 양이 부족
 - 의료데이터 특성상 데이터를 모으기가 어렵고
 - Annotation을 위해 전문적인 지식이 필요하기 때문
- 이미지가 부족할 경우, 과적합이 쉽게 발생함
 - 데이터가 적어 노이즈를 학습하기 쉬움
 - 현실에 있는 데이터의 다양한 측면을 학습하지 못하고, 데이터셋에 포함된 표현에만 국한됨

Data Augmentation

- 데이터 증강(Data augmentation)을 통해 학습에 유의미한 데이터를 늘릴 수 있음
- U-Net에서는 아래와 같은 Augmentation을 사용
 - 회전(Rotation)
 - 이동(Shift)
 - Elastic deformation: 실제 체내 조직에서도 자주 발견되는 현상이기에, 효과적인 증강법

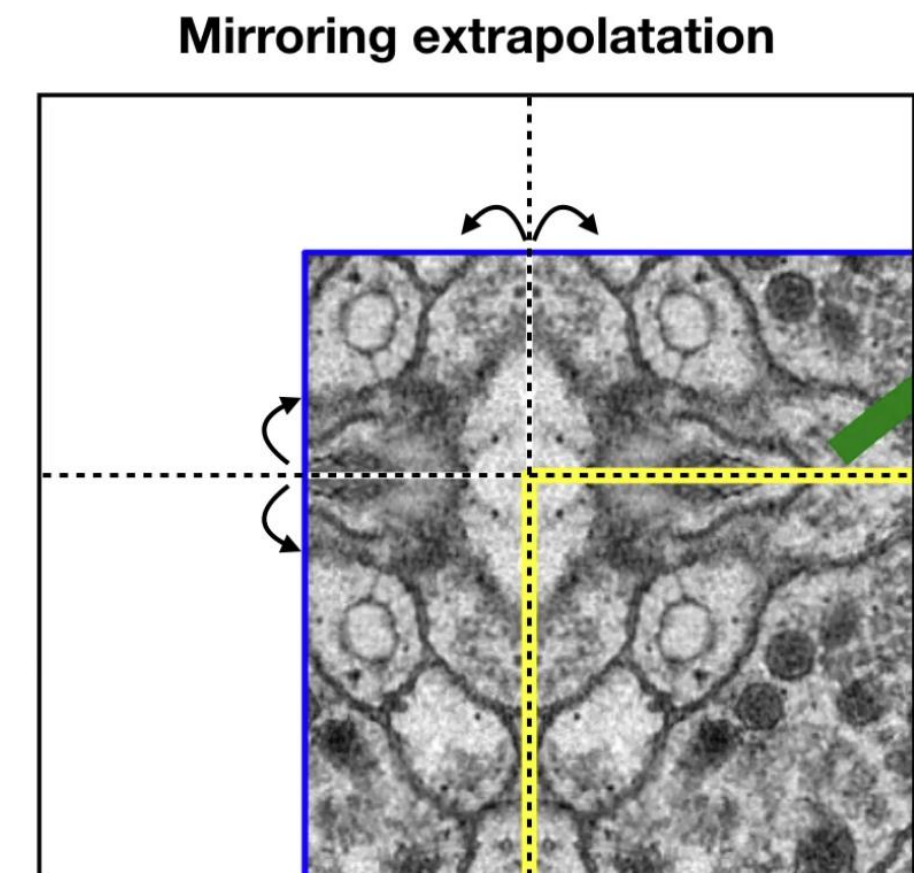
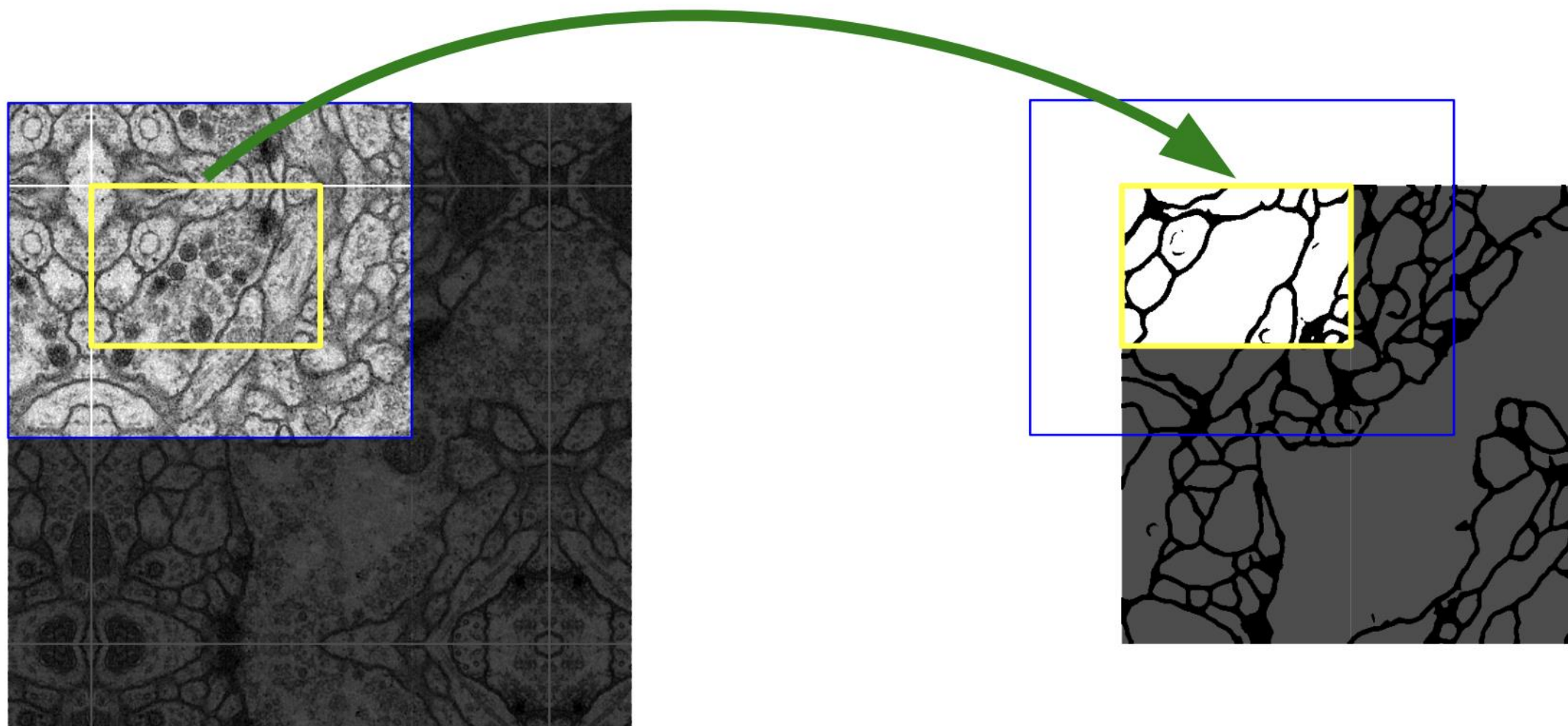


Overlap-Tile Strategy

- 의료 이미지는 주로 고해상도로 촬영됨
 - 한 번에 이미지 전체를 학습하기엔 GPU, 메모리 등 리소스가 많이 소요됨
 - 이미지를 분할하여 학습시키는 것이 효율적
- 이미지를 분할할 경우 정밀한 학습이 어려워짐
 - 주변 정보를 적게 활용하기에 표현 능력이 떨어짐
 - 특히 이미지의 가장자리에서 학습할 정보가 감소함

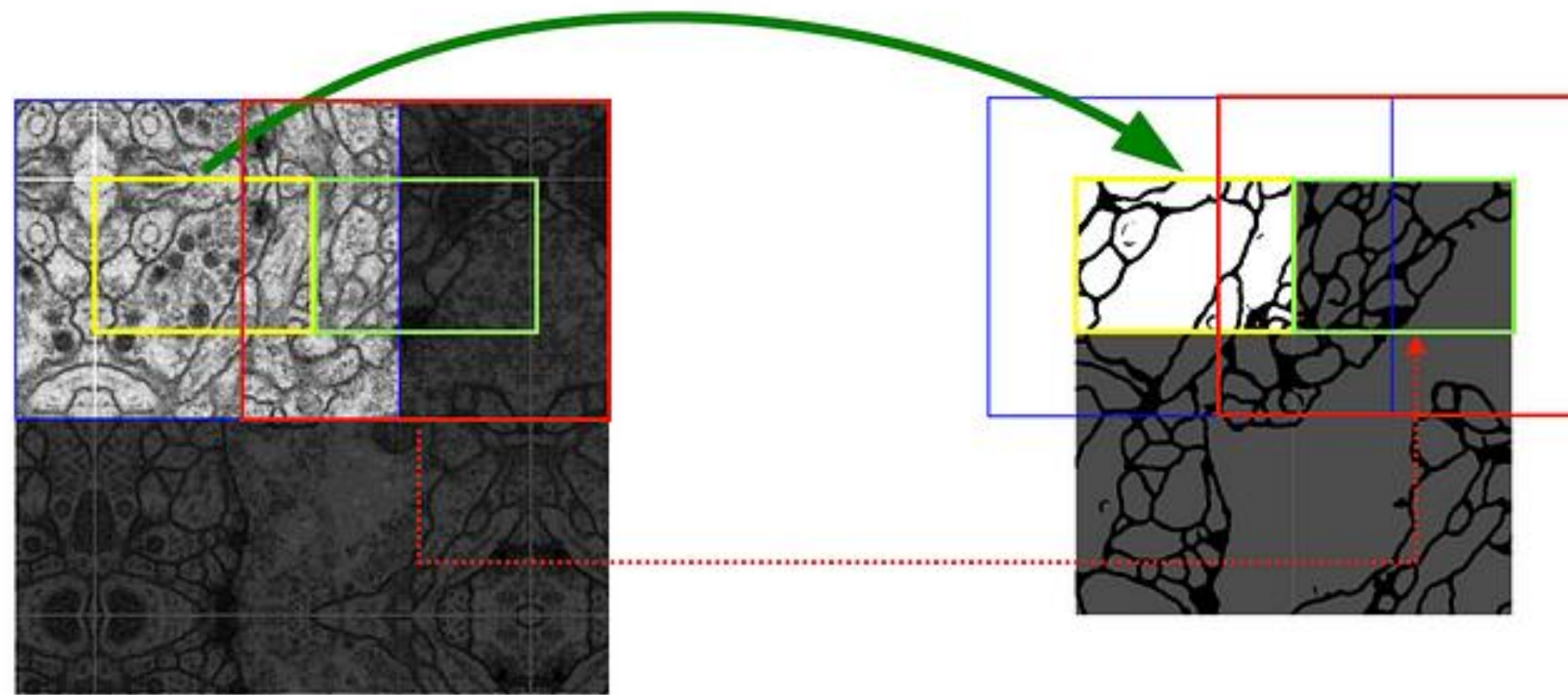
Overlap-Tile Strategy

- Overlap-tile strategy: 큰 이미지를 입력, 중심의 제한된 부분만을 학습하는 전략
 - 푸른 영역을 학습하고, 노란 영역을 추론
 - 노란색 외곽 중, 여백은 이미지를 대칭시켜 채움(Mirroring extrapolation)



Overlap-Tile Strategy

- 다음 이미지(녹색)를 추론하기 위해, 바로 전에 학습한 입력 이미지의 일부분이 포함됨
- 이렇게 겹치는 부분이 존재하므로, "Overlap-tile" strategy라 명명함
- 적은 고해상도 이미지를 분할하여 활용하며, Zero padding보다 유의미한 정보를 학습



U-Net 모델의 성능

- IoU(Intersection over Union)으로 모델 성능 평가
 - PhC-U373 데이터셋(a, b)에 대해 0.9203 (차순위 0.83)
 - 해당 데이터셋은 35장의 학습용 이미지로 구성
 - DIC-HELA 데이터셋(c, d)에 대해 0.7756 (차순위 0.4607)
 - 해당 데이터셋은 20장의 학습용 이미지로 구성

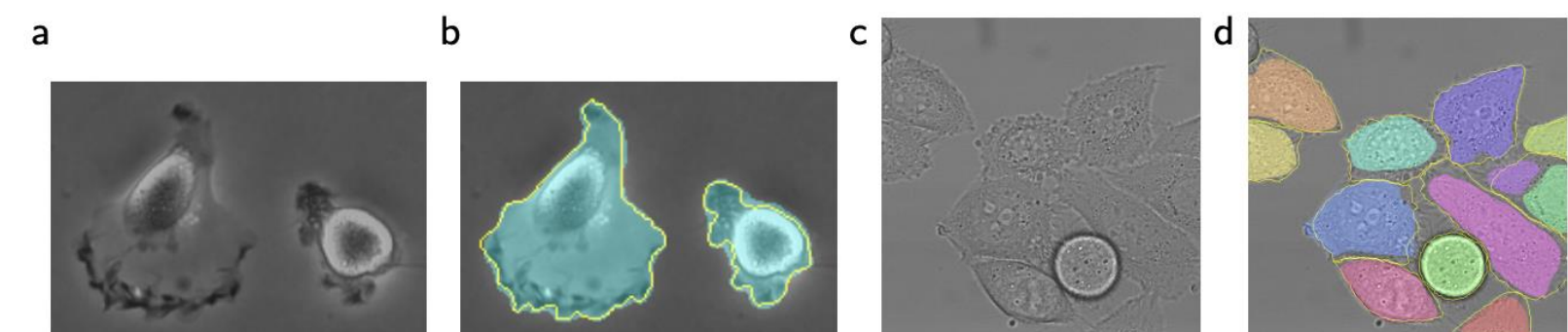


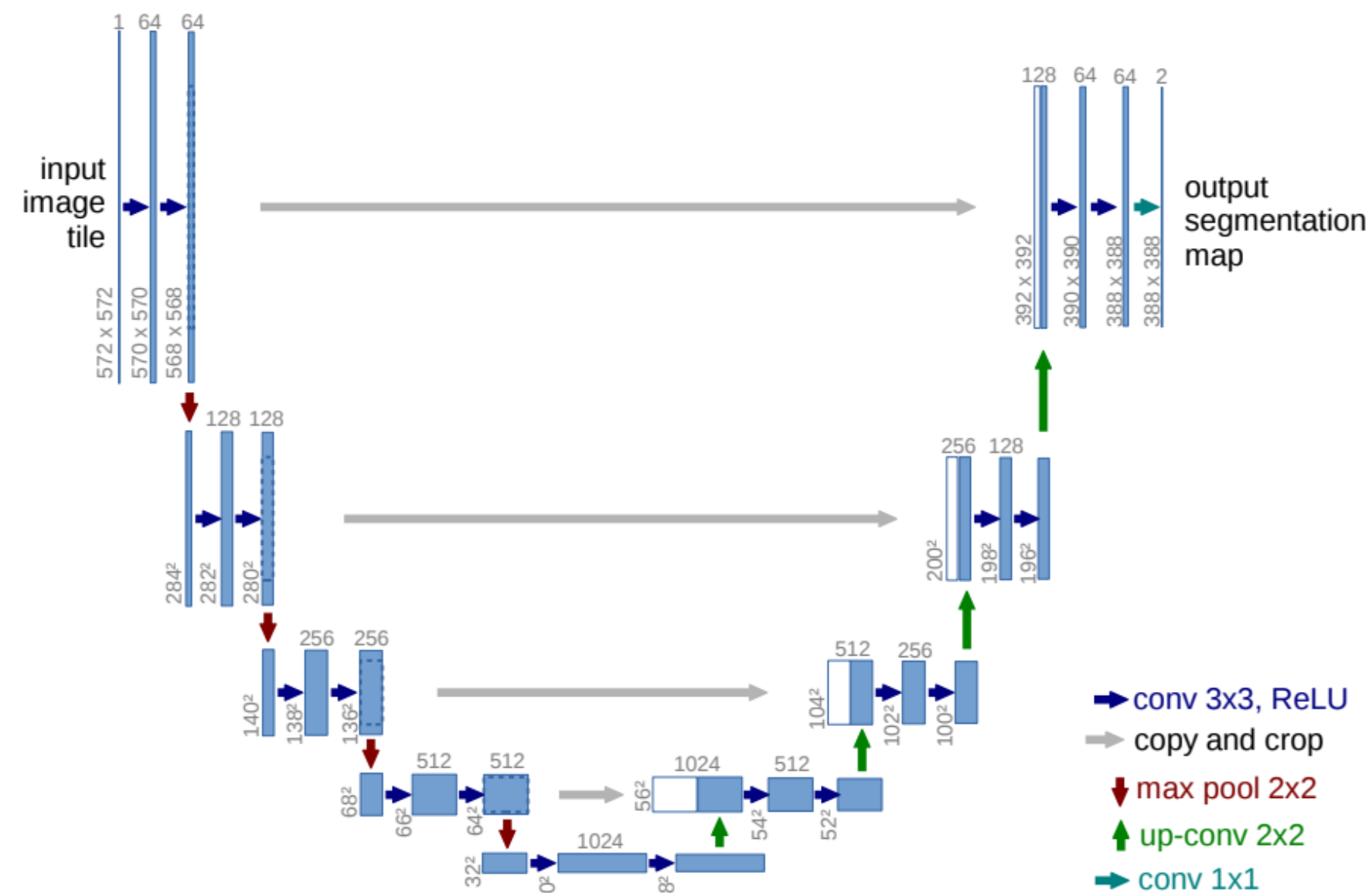
Fig. 4. Result on the ISBI cell tracking challenge. (a) part of an input image of the “PhC-U373” data set. (b) Segmentation result (cyan mask) with manual ground truth (yellow border) (c) input image of the “DIC-HeLa” data set. (d) Segmentation result (random colored masks) with manual ground truth (yellow 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)	0.9203	0.7756

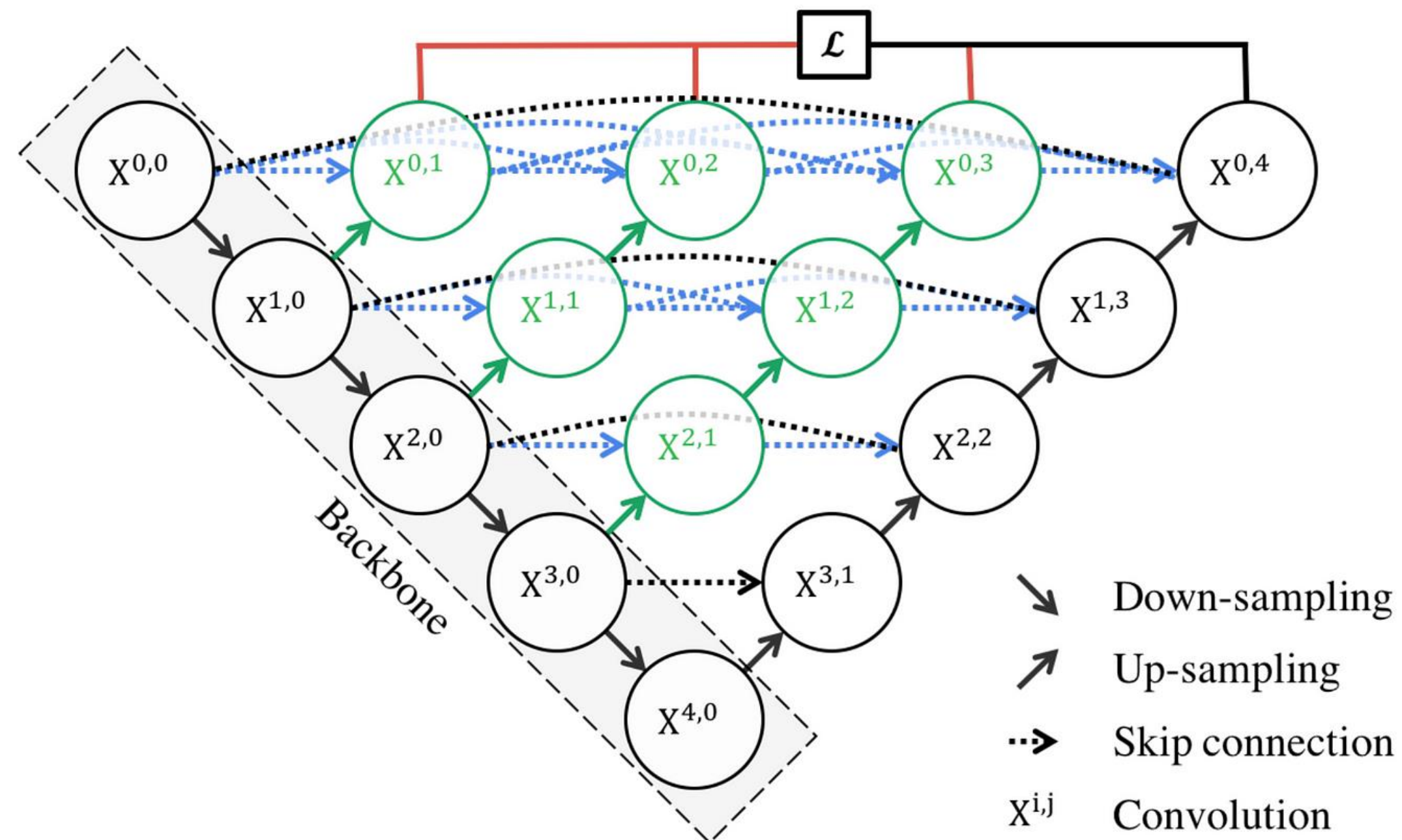
U-Net++의 등장 배경

- U-Net의 한계점을 제시
 - Contracting path와 Expanding path가 동일한 크기(w, h)의 Feature map에 대하여 연결이 되어있지만, 의미상 간극(Semantic gap)을 줄일 여지가 있음
 - 두 Path를 거친 Feature map의 정보가 의미상 유사하다면, 모델 최적화가 더 쉬워질 것



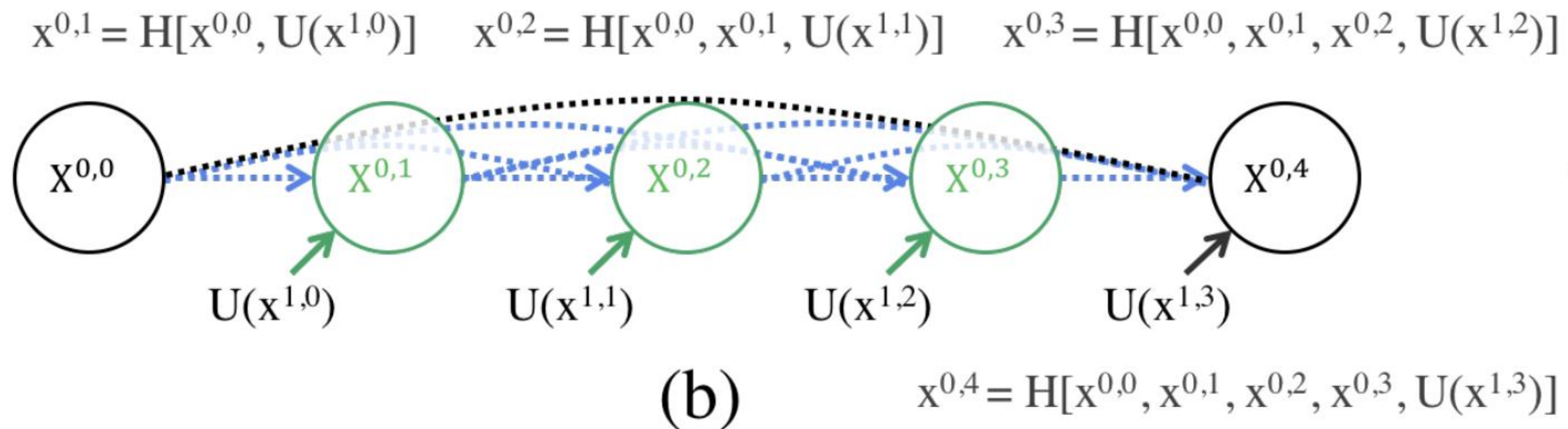
U-Net++의 구조

- (Nested)Dense skip connection: Encoder와 Decoder 사이에 더 많은 연결선 추가
- U-Net에 비해 추가된 구조는 파란 선과 초록색 노드로 표현



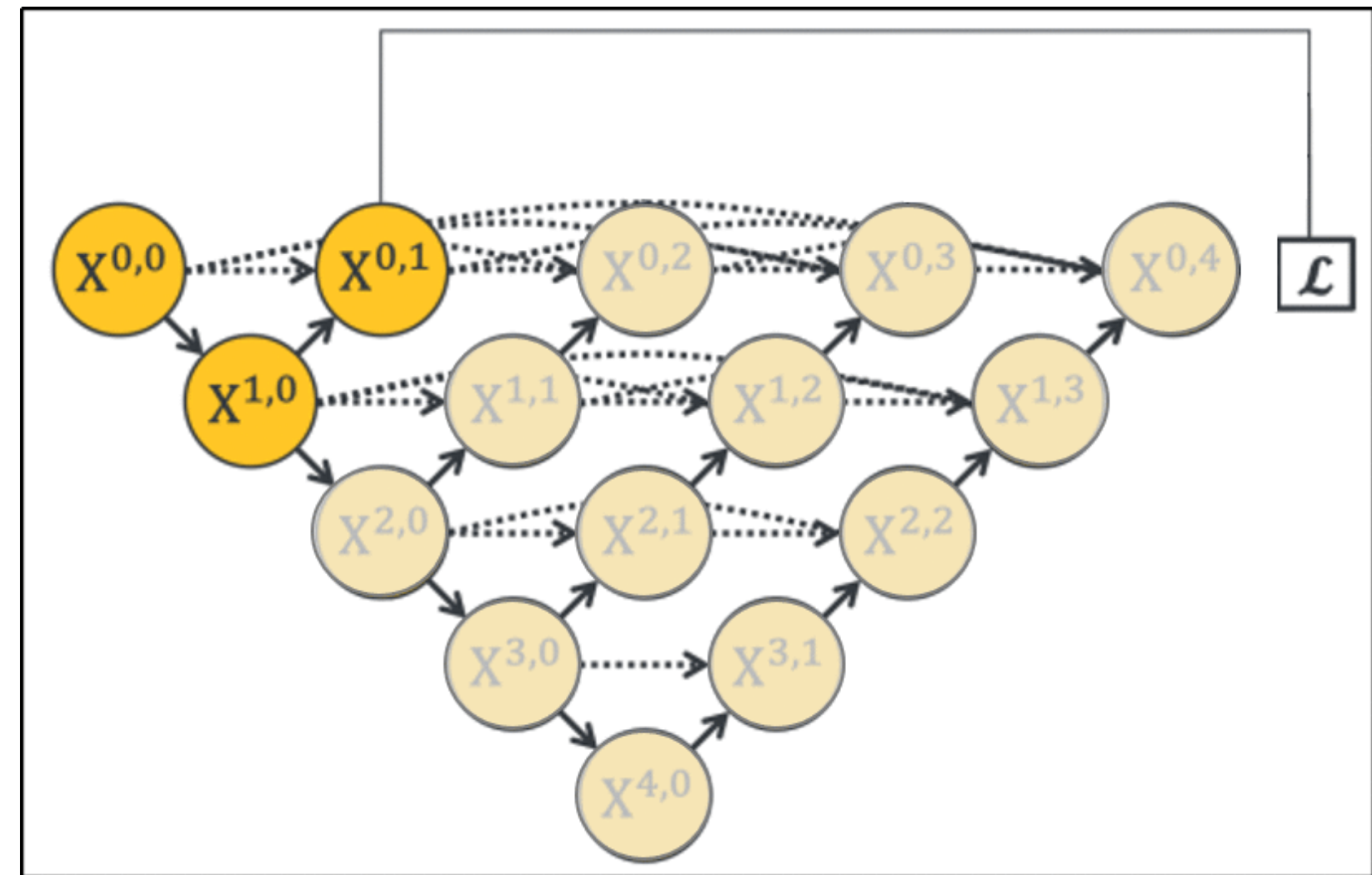
U-Net++의 구조 – Dense Skip Connection

- Encoder-Decoder 사이에 추가적인 Skip connection을 배치하면, Mask가 다양한 깊이의 정보를 활용할 수 있도록 촉진시킬 수 있음
- 기존 U-Net은 동일한 깊이에 위치한 Encoder와 Decoder만을 연결함
- 이는 Decoding시 정보를 다각적으로 고려하지 못하는 단점의 원인임
- Deep Supervision method로 모델을 학습할 수 있음



Deep Supervision

- Deep supervision은 모델 내부 중간 계층에서 추가적인 손실을 계산하여 계층별 학습을 유도하는 방법
 - 추가 손실은 $x^{0,n}$ 에서 계산됨
 - 우측 그림과 같이, 모델은 4개의 작은 U-Net이 조합된 구조로 해석할 수 있음
 - 이는 다양한 깊이의 모델들을 동시에 학습함을 의미하며
 - 데이터에 따라 특정 모델이 억제되거나 성능이 부각됨
 - 즉, 데이터와 이미지 속 객체의 크기에 Robust함



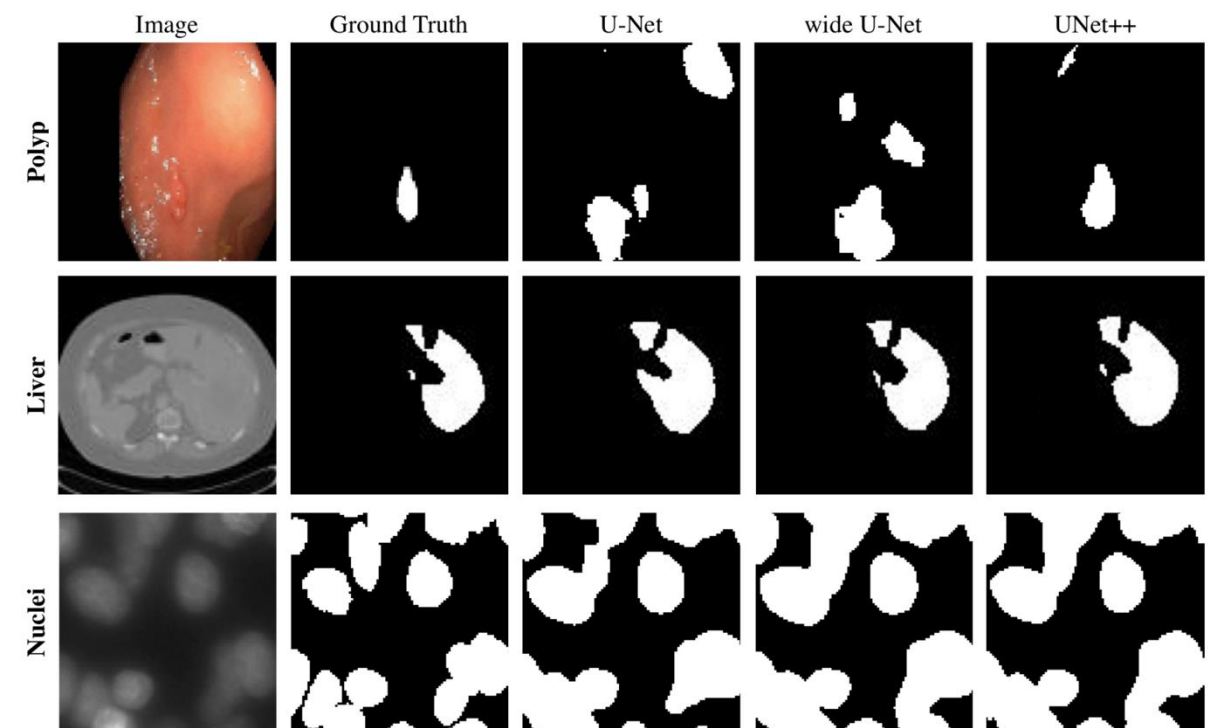
UNet++의 Deep Supervision

U-Net++의 성능

- 다양한 의료 영상에 대하여, U-Net보다 전부 우월한 성능을 보임
- Deep supervision을 활용할 경우, 추가적인 성능 향상도 관찰할 수 있음

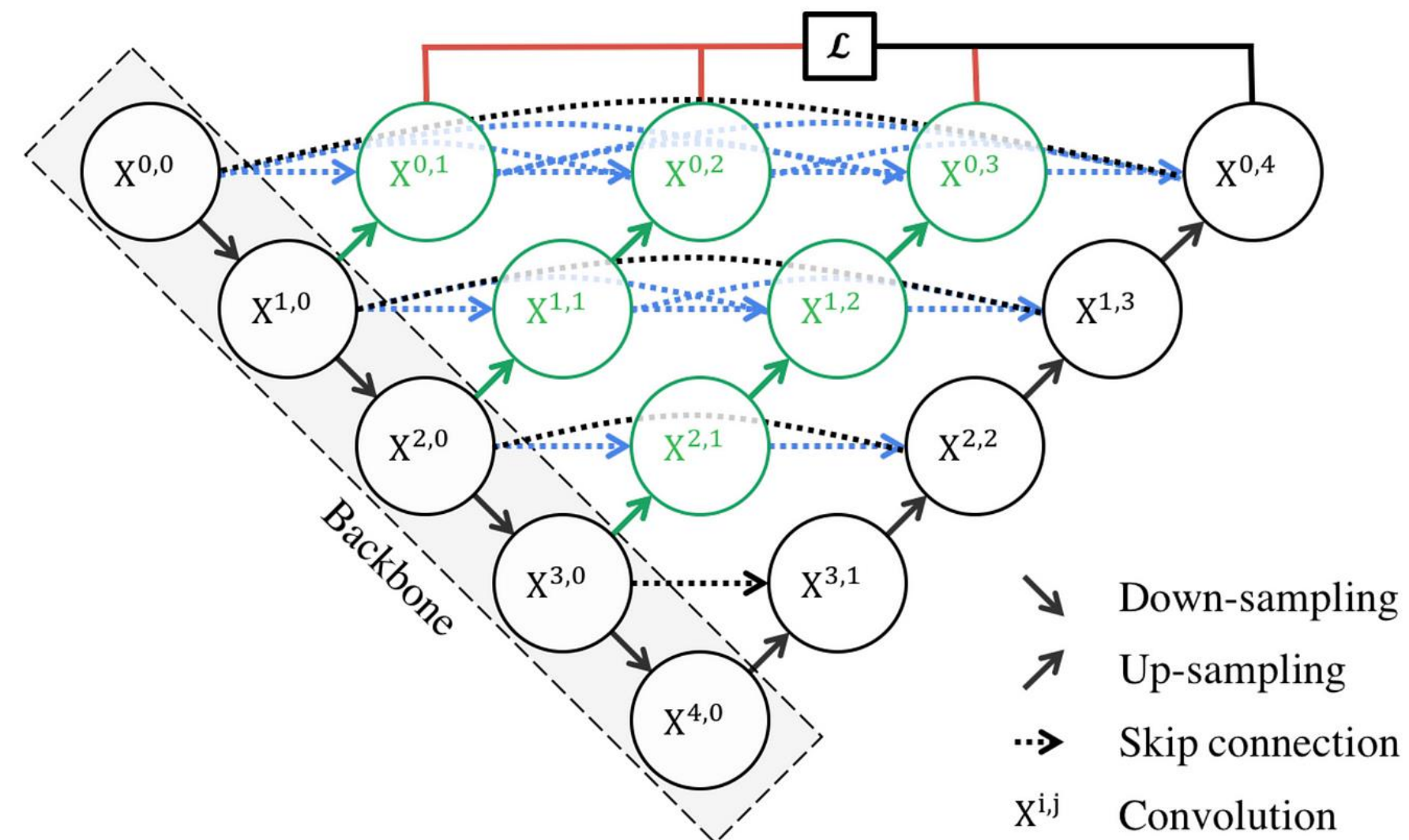
Table 3: Segmentation results (IoU: %) for U-Net, wide U-Net and our suggested architecture UNet++ with and without deep supervision (DS).

Architecture	Params	Dataset			
		cell nuclei	colon polyp	liver	lung nodule
U-Net [9]	7.76M	90.77	30.08	76.62	71.47
Wide U-Net	9.13M	90.92	30.14	76.58	73.38
UNet++ w/o DS	9.04M	92.63	33.45	79.70	76.44
UNet++ w/ DS	9.04M	92.52	32.12	82.90	77.21



U-Net3+의 등장배경

- U-Net++의 Dense skip connection 한계 지적
 - Encoder와 Decoder의 의미상 간극은 줄일 수 있었음
 - 그러나 Skip connection이 불필요하게 많아 Full-scale에 대한 정보가 부족함
 - Full scale: 모든 크기의 Feature map 정보를 의미



U-Net3+의 구조

- Full-scale skip Connection
 - Decoder에서 Feature map을 Upscaling할 때, 모든 Scale의 feature를 직접적으로 활용

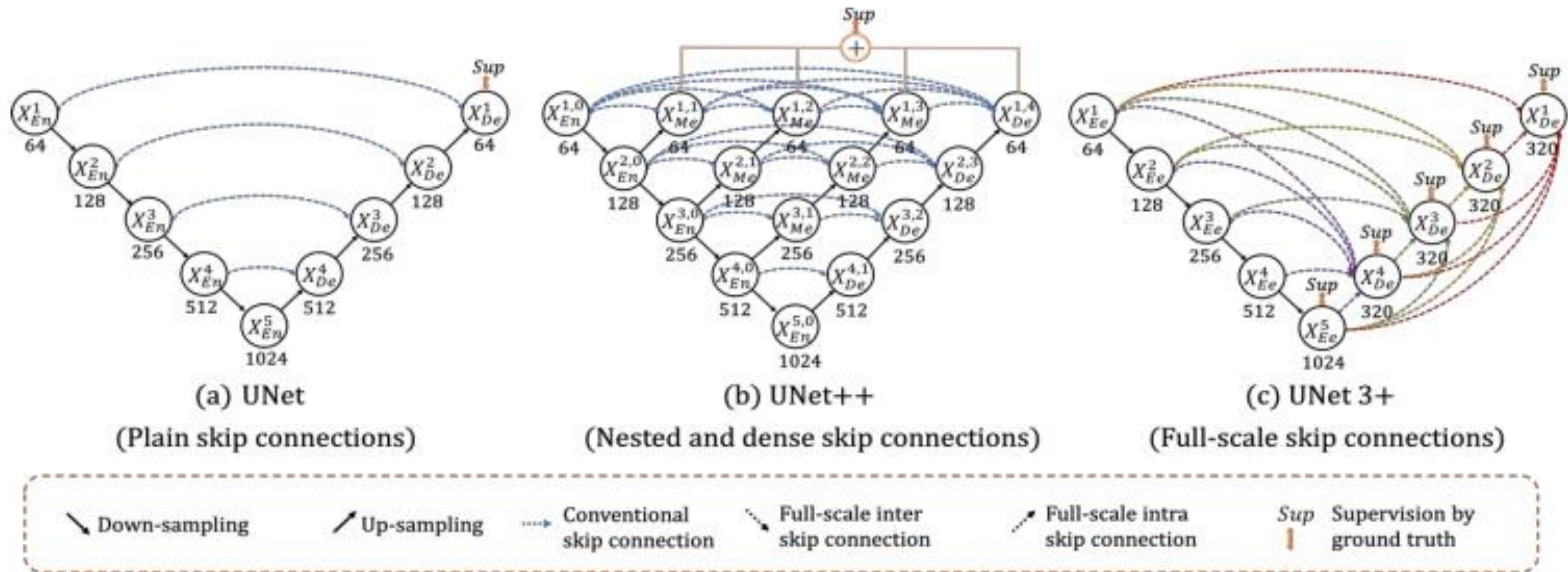
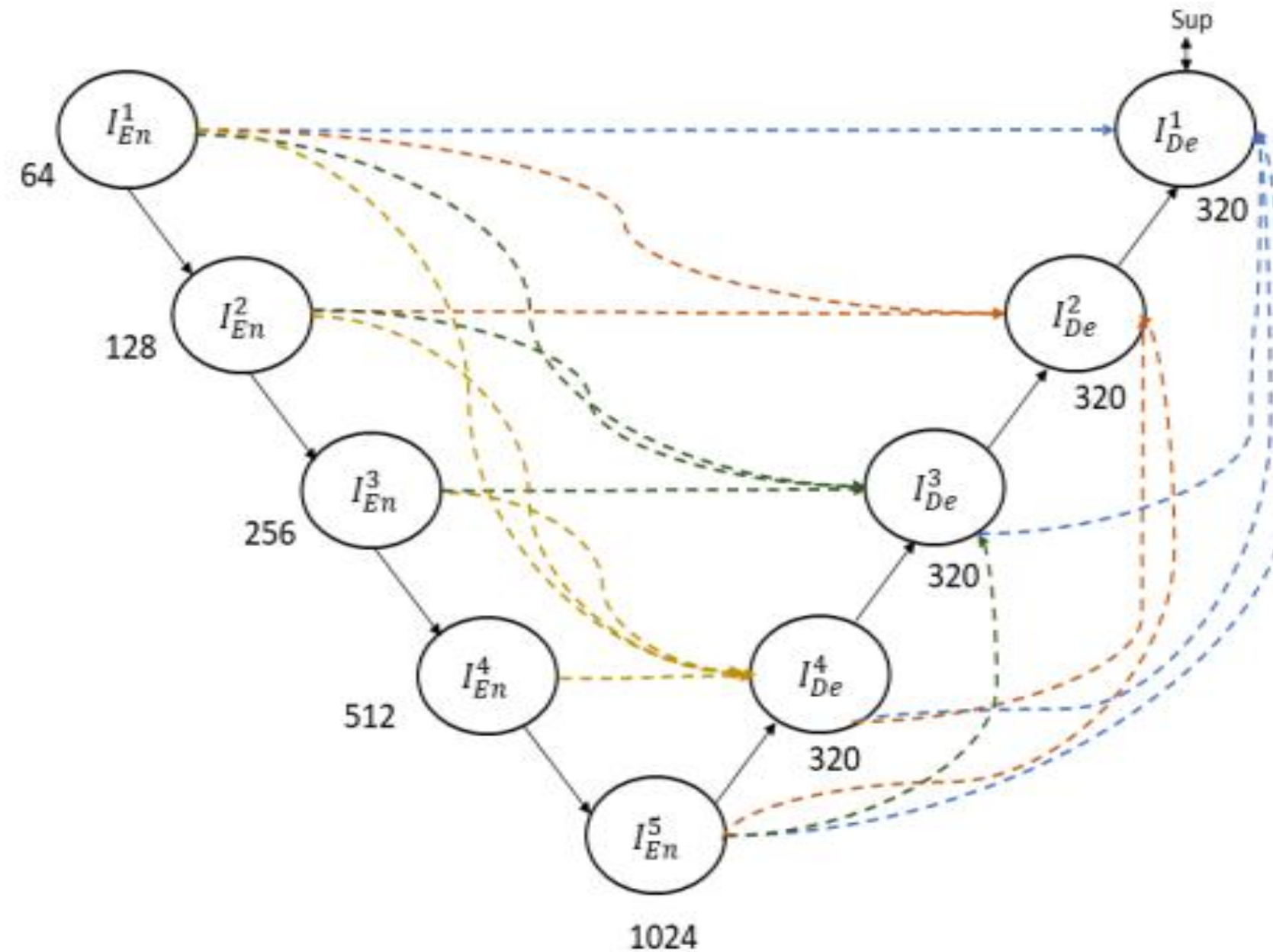


Fig .1: Comparison of U-Net (a), U-Net++(b) and proposed U-Net 3+ (c). The depth of each node is presented below the circle.

U-Net3+ - Full-Scale Skip Connection

- Semantic information과 Appearance information의 연관성을 세밀하게 학습
- 즉, 의료 영상 이미지의 경계를 더욱 명확하게 분할할 수 있음
- 또한, U-Net/U-Net++에 비하여 적은 파라미터를 사용



U-Net3+의 성능

- Liver dataset, Spleen dataset에 대하여 성능 비교 평가
 - 평가 지표로 Dice metrics (Dice coefficient)을 사용
 - 두 샘플의 유사성을 측정
 - 모델 간의 파라미터 사용량도 평가
- 적은 파라미터로, 우수한 성능을 달성

Architecture	Vgg-16			ResNet-101			$Dice_{average}$
	Params	$Dice_{liver}$	$Dice_{spleen}$	Params	$Dice_{liver}$	$Dice_{spleen}$	
UNet	39.39M	0.9206	0.9023	55.90M	0.9387	0.9332	0.9237
UNet++	47.18M	0.9278	0.9230	63.76M	0.9475	0.9423	0.9352
UNet 3+ w/o DS	26.97M	0.9489	0.9437	43.55M	0.9580	0.9539	0.9511
UNet 3+	26.97M	0.9550	0.9496	43.55M	0.9601	0.9560	0.9552