UNet++

Redesigning Skip Connections to Exploit

0. Abstract

(1) Abstract

Unet++의 개요

FCN 및 U-Net의 경우 2가지의 한계점이 존재합니다.

- 데이터셋에 맞은 모델의 최적 깊이를 알 수가 없습니다. 그래서 비용을 들여서 이를 찾아내거나 여러 깊이의 모델들을 앙상블하는 비효율적인 작업이 필요합니다.
- Skip Connection이 동일한 깊이를 가지는 인코더와 디코더만 연결되는 제한적인 구조를 가집니다.

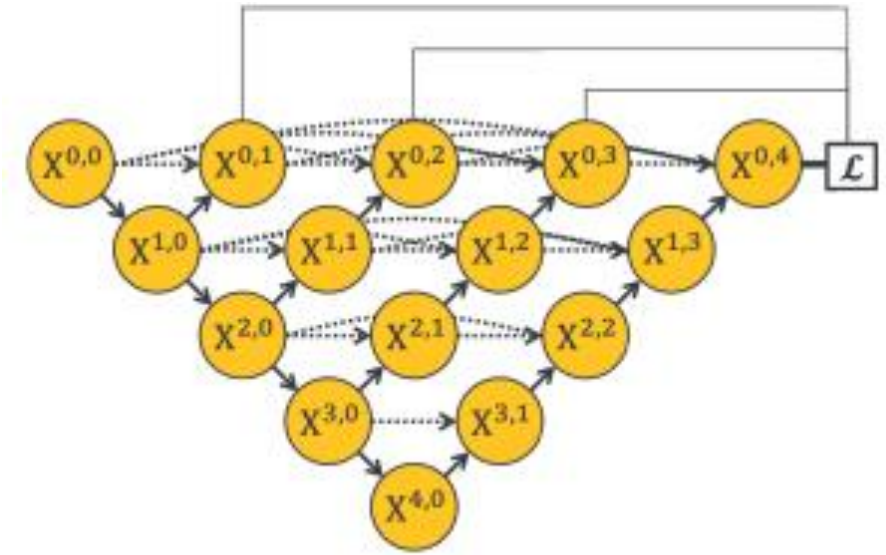
출처 : UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation

0. Abstract

(1) Abstract

Unet++의 개요

이러한 2가지의 한계점을 극복하기 위해서 UNet++에서는 새로운 형태의 아키텍처를 제시합니다.



(g) UNet++

- 인코더를 공유하는 다양한 깊이의 U-Net을 만들어서 deep supervision을 이용해서 함께 학습하고 앙상블하는 형태를 제안합니다.
- skip connection을 동일한 깊이에서의 특징맵들이 모두 결합하도록해서 유연한 특징맵을 만들어줍니다.
- Pruning을 통해서 추론 속도를 올리는 방법을 제안합니다.

출처 : UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation

0. Abstract

(1) Abstract

Unet++의 개요

위의 과정을 통해서 만든 UNet++를 6개의 다른 이미지 데이터셋에 적용하였고 다음의 결과를 얻었습니다.

- UNet++는 6개의 데이터셋에 대해서 일관성있는 높은 성능을 보입니다.
- UNet++는 다양한 크기의 객체에 대해서 높은 성능의 세그멘테이션 품질을 보입니다.
- Mask RCNN에 새로운 Skip Connection을 적용한 Mask RCNN++의 경우 Instance Segmentation에서 높은 성능을 보입니다.
- Pruning을 적용한 UNet++는 높은 성능을 유지하면서 빠른 추론속도를 보입니다.

출처 : UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation

1. Introduction

- (0) Overview

Encoder-Decoder Network의 한계

Abstact에서 언급한 것처럼 기존의 encoder-decoder 형식의 모델들의 경우 2가지의 한계점이 존재합니다.

- 첫째, 데이터 셋마다 최적의 깊이가 다릅니다. 그렇기에 이를 찾아주거나 다양한 깊이의 모델들을 각각 학습한 후에 결합하는 형식의 방법들이 제안되었습니다. 하지만, 이러한 접근은 인코더를 공유하지 않고 각각 돌아간다는 점에서 비효율적입니다. 특히, 이렇게 독립적으로 학습하게되면 multi-task learning의 장점이 없기도 합니다.

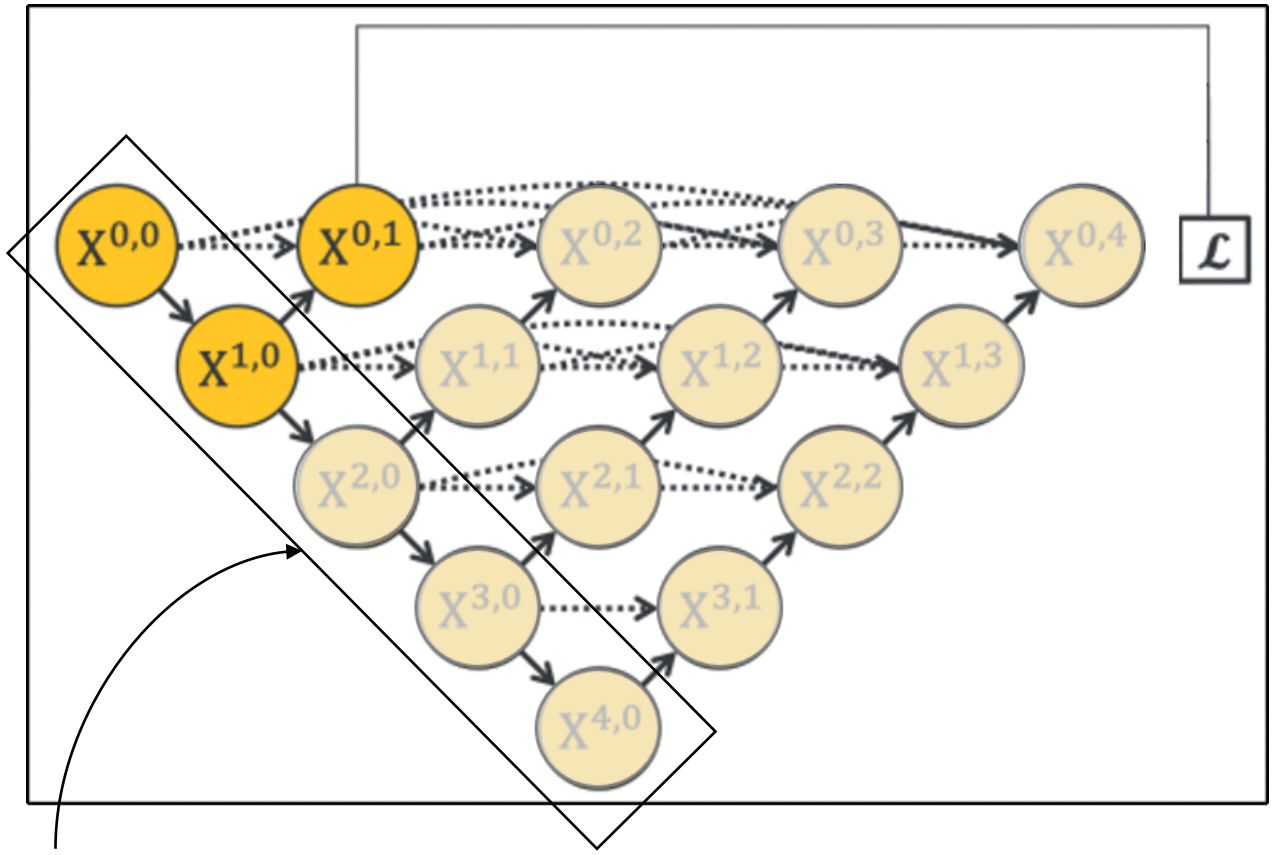
- 둘째, Skip Connections의 디자인이 불필요하게 제한적입니다. 같은 크기의 특징맵을 가지는 경우의 인코더와 디코더가 결합하는 구조는 너무 약합니다.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

1. Introduction

- (0) Overview

Encoder-Decoder Network의 한계를 극복



인코더를 여러 개의 UNet들이 공유하는 형태

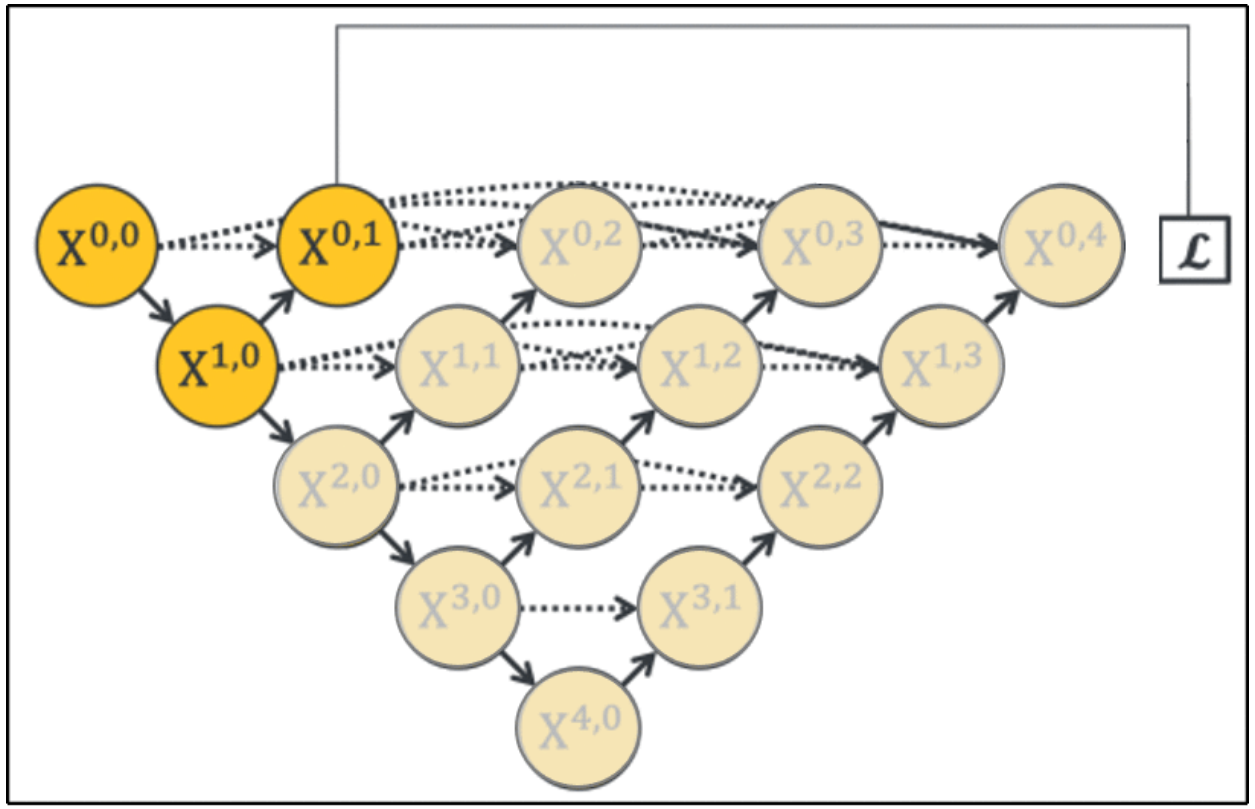
- UNet++는 인코더를 공유하는 형태로 다양한 깊이에서의 학습을 공유합니다. Deep Supervision을 통해서 이미지의 표현(representation)을 공유하면서 같이 학습하고 깊이를 선택할 필요가 없습니다.
- 성능의 향상뿐만아니라 Pruning을 통해서 추론 속도를 올릴 수 있다는 장점도 있습니다.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

1. Introduction

- (0) Overview

Encoder-Decoder Network의 한계를 극복



- UNet++는 인코더를 공유하는 형태로 다양한 깊이에서의 학습을 공유합니다. Deep Supervision을 통해서 이미지의 표현(representation)을 공유하면서 같이 학습하고 깊이를 선택할 필요가 없습니다.
- 성능의 향상뿐만아니라 Pruning을 통해서 추론 속도를 올릴 수 있다는 장점도 있습니다.
- UNet++는 인코더와 디코더의 동일한 크기의 특징 맵만 결합되는 제한적인 연결을 사용하지 않습니다. 이를 Dense하게 Connection 함으로써 다양한 특징 맵의 특성들을 디코어의 특징 맵과 결합시켜줍니다.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

1. Introduction

(0) Overview

UNet++의 주요 공헌

그 결과 6개의 데이터셋에 대해서 높은 성능을 달성했고 모델의 주요 공헌을 정리하면 아래와 같습니다.

1. UNet++에서 다양한 깊이의 U-Nets의 내부적인 앙상블을 도입하여 다양한 크기의 객체에 대해 성능 향상을 보였습니다.
2. 우리는 UNet++에서 Skip Connection을 재설계하여 디코더의 유연한 특징 결합을 가능하게 합니다. 이는 동일한 깊이의 인코더와 디코더를 결합하는 U-Net에 비해 성능적으로 많은 향상을 가져옵니다.
3. Pruning을 통해 성능은 유지하면서 추론 속도만 향상시키는 UNet++ 방법을 제안합니다.
4. UNet++에 내장된 다양한 깊이의 U-Nets의 훈련은 U-Net 간의 협업 학습을 통해 개별 U-Net들을 학습하는 것 보다 나은 성능을 보입니다.
5. UNet++의 경우 다양한 인코더 백본을 가져와서 학습이 가능하고 다양한 의료 영상 데이터에서 높은 성능을 보임으로서 확장성 및 활용 가능성을 증명합니다.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

2. Proposed Network Architecture : UNet++

(1) Motivation behind the new architecture

Motivation

TABLE I: Ablation study on U-Nets of varying depths alongside with the new variants of U-Nets proposed in this work. U-Net L^d refers to a U-Net with a depth of d (Fig. 1(a-d)). U-Net^e, UNet+, and UNet++ are the new variants of U-Net, which are depicted in Fig. 1(e-g). "DS" denotes deeply supervised training followed by average voting. Intersection over union (IoU) is used as the metric for comparison (mean±s.d. %).

Architecture	DS	Params	EM	Cell	Brain Tumor
U-Net L^1	✗	0.1M	86.83±0.43	88.58±1.68	86.90±2.25
U-Net L^2	✗	0.5M	87.59±0.34	89.39±1.64	88.71±1.45
U-Net L^3	✗	1.9M	88.16±0.29	90.14±1.57	89.62±1.41
U-Net (L^4)	✗	7.8M	88.30±0.24	88.73±1.64	89.21±1.55
U-Net^e	✓	8.7M	88.33±0.23	90.72±1.51	90.19±0.83
UNet+	✗	8.7M	88.39±0.15	90.71±1.25	90.70±0.91
UNet+	✓	8.7M	88.89±0.12	91.18±1.13	91.15±0.65
UNet++	✗	9.0M	88.92±0.14	91.03±1.34	90.86±0.81
UNet++	✓	9.0M	**89.33**±0.10	**91.21**±0.98	**91.21**±0.68

EM은 깊이가 깊으면 모델의 성능이 향상

Cell, Brain Tumor는 깊이가 깊은 경우 점수 하락

1. U-Net의 깊이를 깊게 한다고 성능이 반드시 좋은 것은 아닙니다.
2. 모델의 최적 깊이는 데이터셋마다 다릅니다. (EM : L4, Cell : L3, Brain : L3)

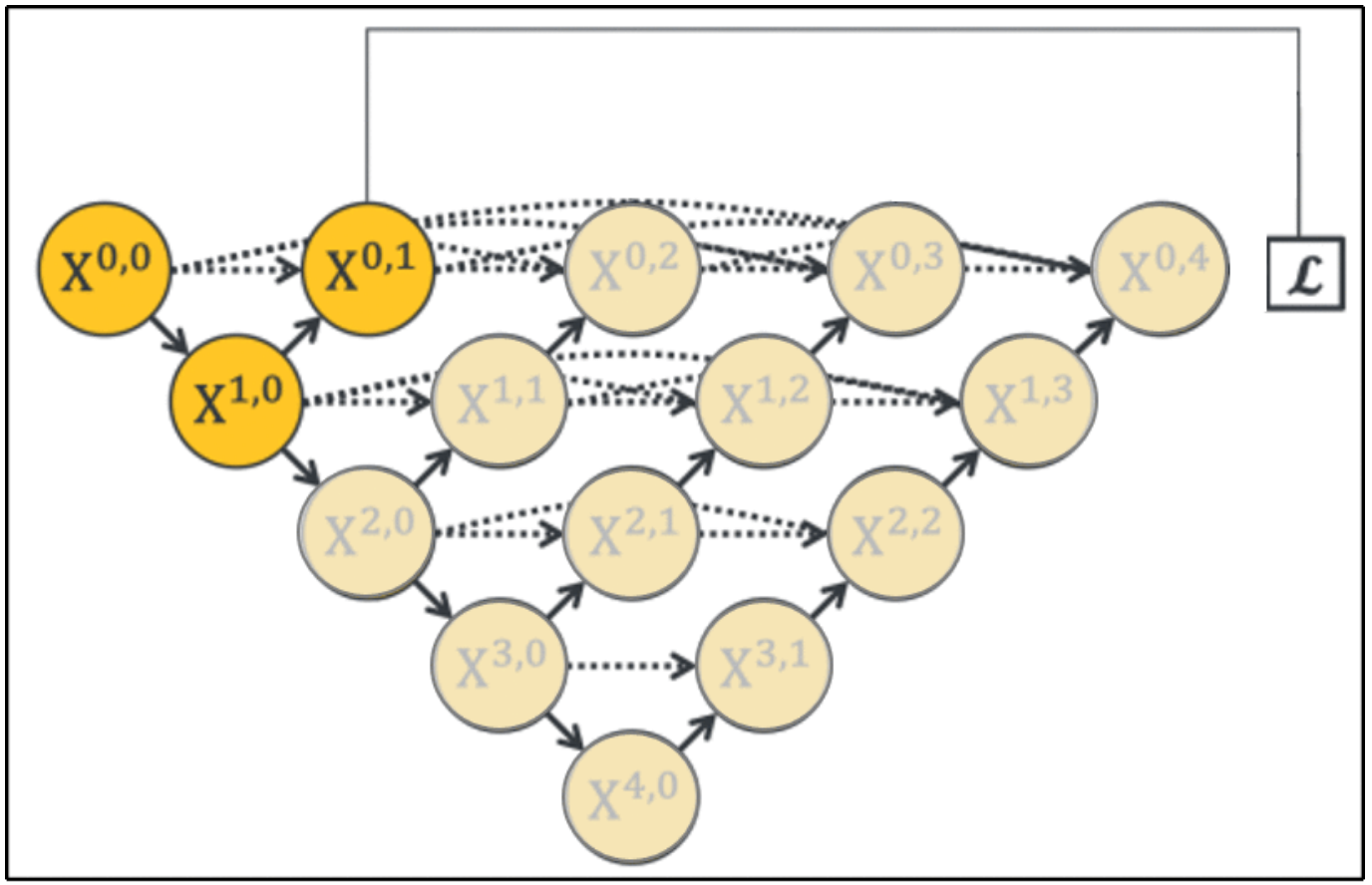
출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

2. Proposed Network Architecture : UNet++

(1) Motivation behind the new architecture

Solution

모든 깊이의 UNet을 결합시키자!!



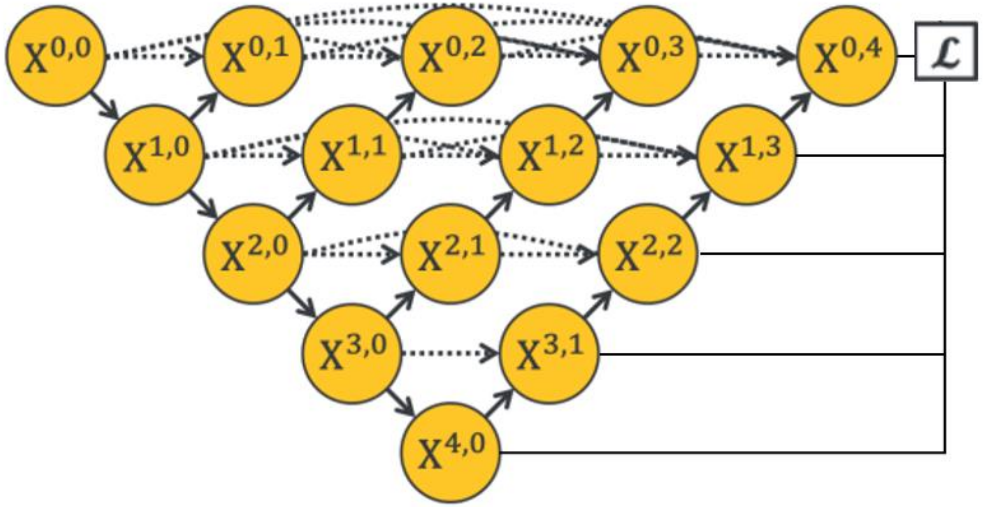
UNet++의 Deep Supervision

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

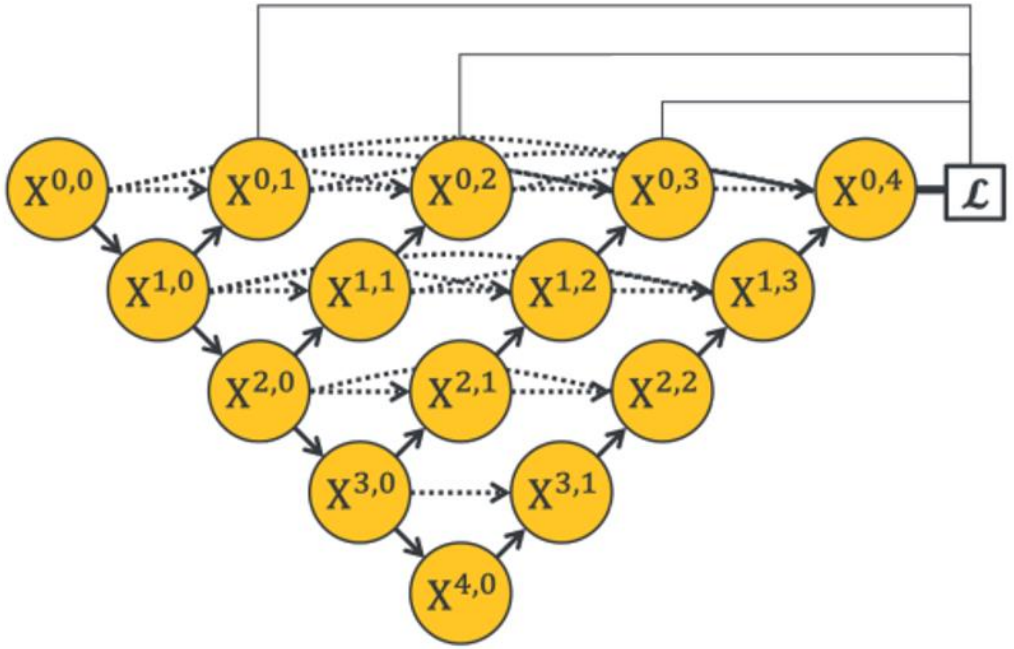
2. Proposed Network Architecture : UNet++

(1) Motivation behind the new architecture

Solution



일반적인 방식의 Deep Supervision



UNet++의 Deep Supervision

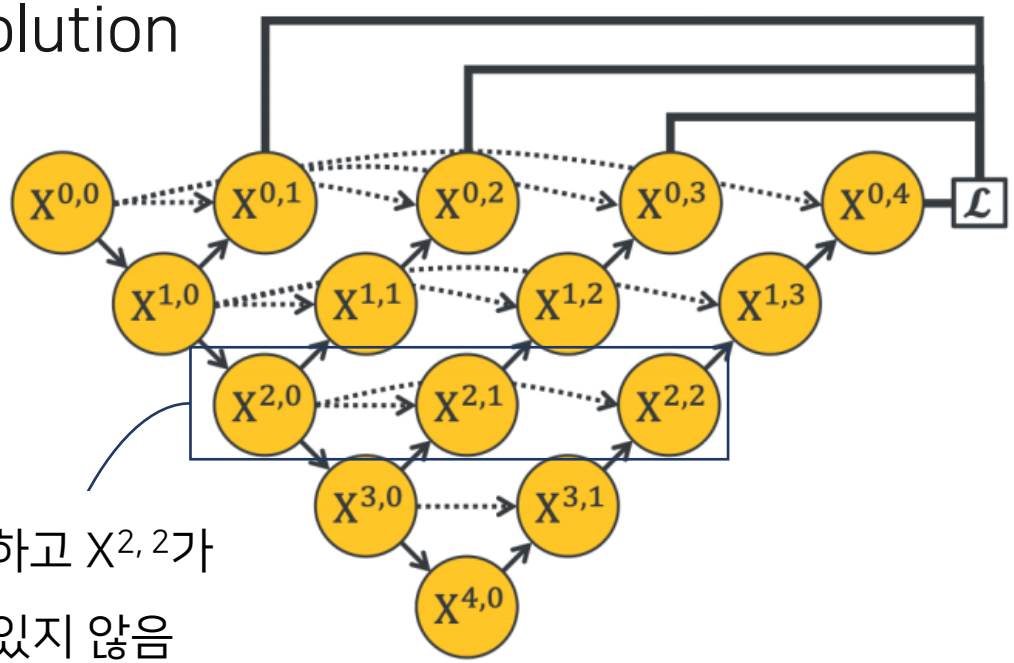
- Deep Supervision의 경우 기존의 논문들과는 다르게 $X^{4-j,\,j}$가 아닌 $X^{0,\,j}$에 둠으로써 U-Net의 구조들을 앙상블한 형식입니다.
- 즉, 4개의 서로 다른 깊이를 가지는 U-Net이 결합된 형태입니다.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

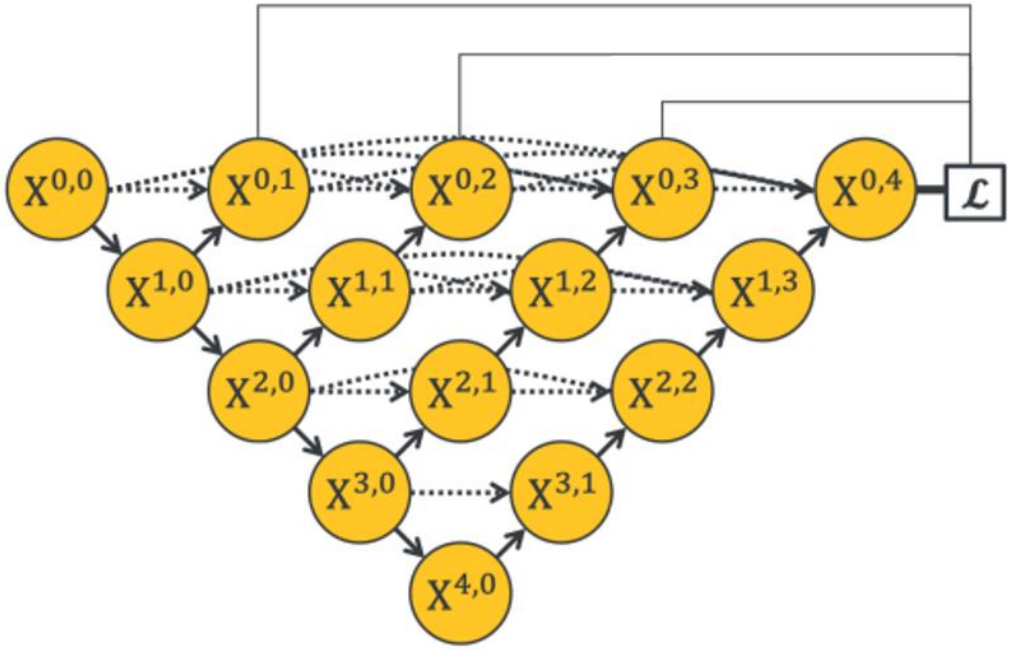
2. Proposed Network Architecture : UNet++

(1) Motivation behind the new architecture

Solution



U-Net^e



UNet++의 Deep Supervision

노드 $X^{2,1}$하고 $X^{2,2}$가 연결 되어있지 않음

- 디코더 $X^{4-j,j}$가 분리되어 더 깊은 U-Nets는 앙상블에서 더 낮은 U-Net의 디코더에 신호를 제공하지 않습니다.
  - 노드 $X^{2,1}$의 경우 3번째 U-Net의 디코더인데 4번째 디코더를 구성하는 $X^{2,2}$하고 연결 되어있지 않아서 정보의 손실이 발생합니다.
  - U-Net^{e}의 디코더들의 경우 불필요하게 같은 크기의 특징 맵만 결합하기에 객체의 크기에 유연하지 못한 단점이 있습니다.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

2. Proposed Network Architecture : UNet++

(2) Technical Details

Nested Convolution

$$x^{i,j} = \begin{cases} \mathcal{H}\left(\mathcal{D}(x^{i-1,j})\right), & j = 0 \\ \mathcal{H}\left(\left[\left[x^{i,k}\right]_{k=0}^{j-1}, \mathcal{U}(x^{i+1,j-1})\right]\right), & j > 0 \end{cases}$$

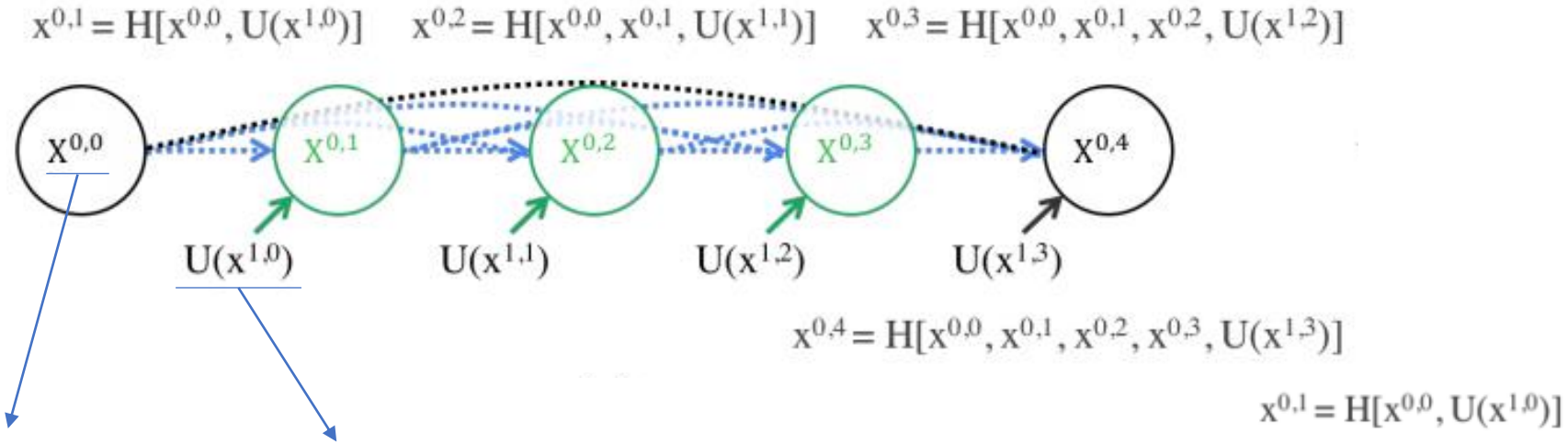
- H : Convolution Operation
- D : Down Sampling
- U : Up Sampling
- [] : Concatenation

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

2. Proposed Network Architecture : UNet++

(2) Technical Details

Nested Convolution



- $x^{0,1}$의 경우 $x^{0,0}$과 $x^{1,0}$을 Upsampling한 결과를 결합한 형태($[x^{0,0}, U(x^{1,0})]$)에 Convolution을 적용한 형태입니다.

$x^{0,1} = H[x^{0,0}, U(x^{1,0})]$ → $x^{0,4} = H[x^{0,0}, x^{0,1}, x^{0,2}, x^{0,3}, U(x^{1,3})]$

- 디코더 쪽으로 갈 수록 내부의 입력 값들을 늘어나고 마지막 디코어 $x^{0,4}$의 경우 모든 입력을 받습니다.

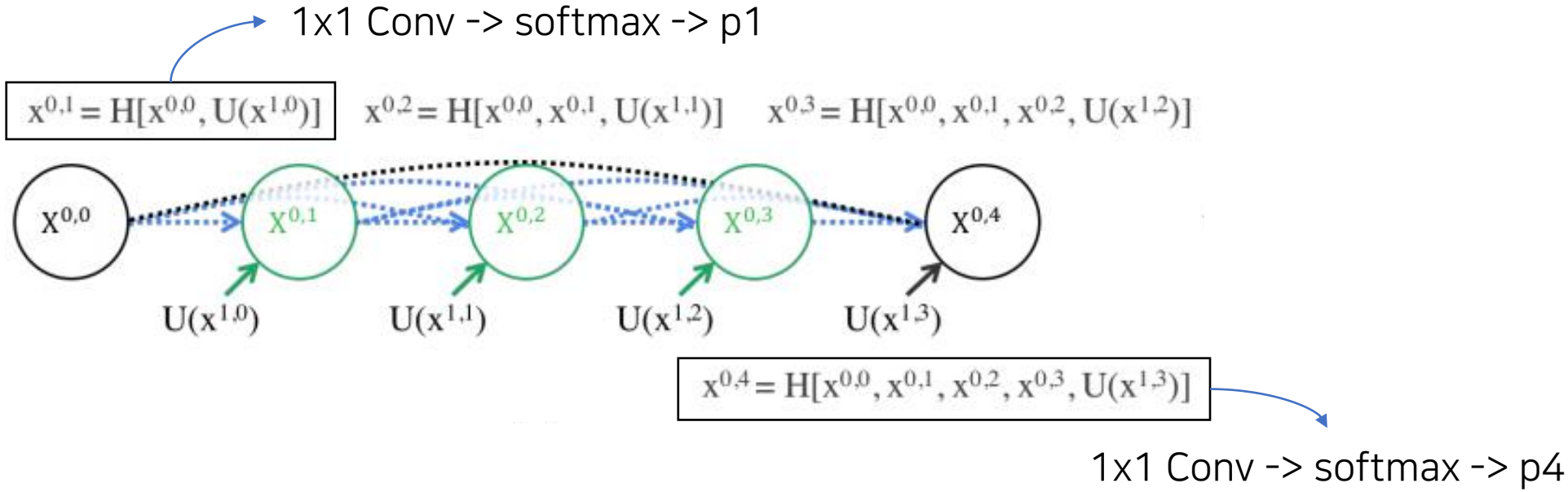
- j = 0인 경우 바로 윗 노드가 다운 샘플된 값에 Convolution을 수행한 결과가 됩니다.
  - 예를 들어 $x^{1,0}$의 경우에는 $H(D(x^{0,0}))$으로 인코더의 바로 윗 노드가 다운 샘플된 값에 Convolution을 수행한 결과가 됩니다.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

2. Proposed Network Architecture : UNet++

(2) Technical Details

Deep Supervision



- 총 4개의 Probability가 생깁니다. 그리고 이 4개에 대해서 각각 Loss를 계산해주고 Average를 취하면 Deep Supervision의 값이 계산됩니다.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

2. Proposed Network Architecture : UNet++

(2) Technical Details

Deep Supervision

$$\mathcal{L}(Y, P) = -\frac{1}{N}\sum_{c=1}^{C}\sum_{n=1}^{N}\left(y_{n,c}\log p_{n,c} + \frac{2y_{n,c}p_{n,c}}{y_{n,c}^2 + p_{n,c}^2}\right)$$

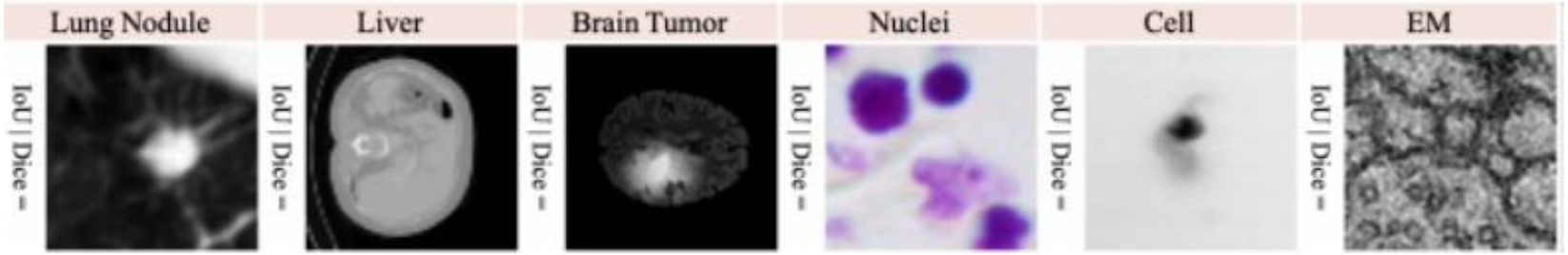
- N : 하나의 배치내의 픽셀 수
- n^{th} : 배치내의 픽셀 번호
- C : 클래스의 개수
- $y_{n,c}$: 타겟 레이블
- $p_{n,c}$: 예측 레이블

- Pixel Wise Cross Entropy + Soft Dice Coefficient의 하이브리드 로스

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

3. Experiments
(1) Datasets

6개의 데이터셋



1. Electron Microscopy (EM)
   - 30 images (512 x 512)
   - 2 classes
   - 96 x 96 패치를 적용하고 sliding window를 사용해서 패치의 절반이 겹치게 한 후 겹치는 부분은 aggregate를 적용
2. Cell
   - training 212 / validation 70 / test 72 이미지
   - 2 classes
3. Nuclei
   - training 335 / validation 134 / test 201 이미지
   - 96 x 96 패치를 적용하고 sliding window를 사용해서 32 pixel stride 적용
4. Brain Tumor
   - 256 x 256 의 30명의 환자들에 대한 데이터셋
5. Liver
   - training 100 / validation 15 / test 15 환자 데이터셋
6. Lung Nodule
   - training 510 / validation 100 / test 408 이미지
   - 64 x 64 x 64 crop 적용

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

3. Experiments

(2) Baseline and implementation

실험 결과

EM	Sensitivity	Specificity	F1 score	F2 score
U-Net	91.21±2.18	83.55±1.62	87.21±1.88	89.56±2.06
UNet++	92.87±2.08	84.94±1.55	88.73±1.79	91.17±1.96
p-value	0.018	0.008	0.013	0.016
Cell	Sensitivity	Specificity	F1 score	F2 score
U-Net	94.04±2.36	96.10±0.75	81.25±2.62	88.47±2.49
UNet++	95.88±2.59	96.76±0.65	84.34±2.52	90.90±2.57
p-value	0.025	0.005	5.00*e*-4	0.004
Nuclei	Sensitivity	Specificity	F1 score	F2 score
U-Net	93.57±4.30	93.94±0.87	83.64±2.97	89.33±3.71
UNet++	97.28±4.85	96.30±0.94	90.14±3.82	94.29±4.41
p-value	0.015	5.35*e*-10	6.75*e*-7	4.47*e*-4
Brain Tumor	Sensitivity	Specificity	F1 score	F2 score
U-Net	94.00±1.15	97.52±0.78	88.42±2.61	91.68±1.77
UNet++	95.81±1.25	98.01±0.67	90.83±2.46	93.75±1.77
p-value	2.90*e*-5	0.042	0.005	7.03*e*-3
Liver	Sensitivity	Specificity	F1 score	F2 score
U-Net	91.22±2.02	98.48±0.43	86.19±2.84	89.14±2.37
UNet++	93.15±1.88	98.74±0.36	88.54±2.57	91.25±2.18
p-value	0.003	0.046	0.010	0.006
Lung Nodule	Sensitivity	Specificity	F1 score	F2 score
U-Net	94.95±1.31	97.27±0.47	83.98±1.94	90.24±1.60
UNet++	95.83±0.86	97.81±0.40	86.78±1.66	91.99±1.22
p-value	0.018	3.25*e*-3	1.92*e*-5	4.27*e*-3

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

4. Results

(1) Semantic Segmentation Results

TABLE IV: Semantic segmentation results measured by IoU (mean±s.d. %) for U-Net, wide U-Net, UNet+ (our intermediate proposal), and UNet++ (our final proposal). Both UNet+ and UNet++ are evaluated with and without deep supervision (DS). We have performed independent two sample t-test between U-Net [5] vs. others for 20 independent trials and highlighted boxes in red when the differences are statistically significant ($p < 0.05$).

Architecture	DS	Params	2D Application					Architecture	DS	Params	3D Application
			EM	Cell	Nuclei	Brain Tumor†	Liver				Lung Nodule
U-Net [5]	✗	7.8M	88.30±0.24	88.73±1.64	90.57±1.26	89.21±1.55	79.90±1.38	V-Net [28]	✗	22.6M	71.17±4.53
wide U-Net	✗	9.1M	88.37±0.13	88.91±1.43	90.47±1.15	89.35±1.49	80.25±1.31	wide V-Net	✗	27.0M	73.12±3.99
UNet+	✗	8.7M	88.39±0.15	90.71±1.25	91.73±1.09	90.70±0.91	79.62±1.20	VNet+	✗	25.3M	75.93±2.93
UNet+	✓	8.7M	88.89±0.12	91.18±1.13	92.04±0.89	91.15±0.65	**82.83±0.92**	VNet+	✓	25.3M	76.72±2.48
UNet++	✗	9.0M	88.92±0.14	91.03±1.34	**92.44±1.20**	90.86±0.81	82.51±1.29	VNet++	✗	26.2M	76.24±3.11
UNet++	✓	9.0M	**89.33±0.10**	**91.21±0.98**	92.37±0.98	**91.21±0.68**	82.60±1.11	VNet++	✓	26.2M	**77.05±2.42**

† The winner in BraTS 2013 holds a "complete" Dice of 92% vs. 90.83%±2.46% (our UNet++ with deep supervision).

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

4. Results

(1) Semantic Segmentation Results

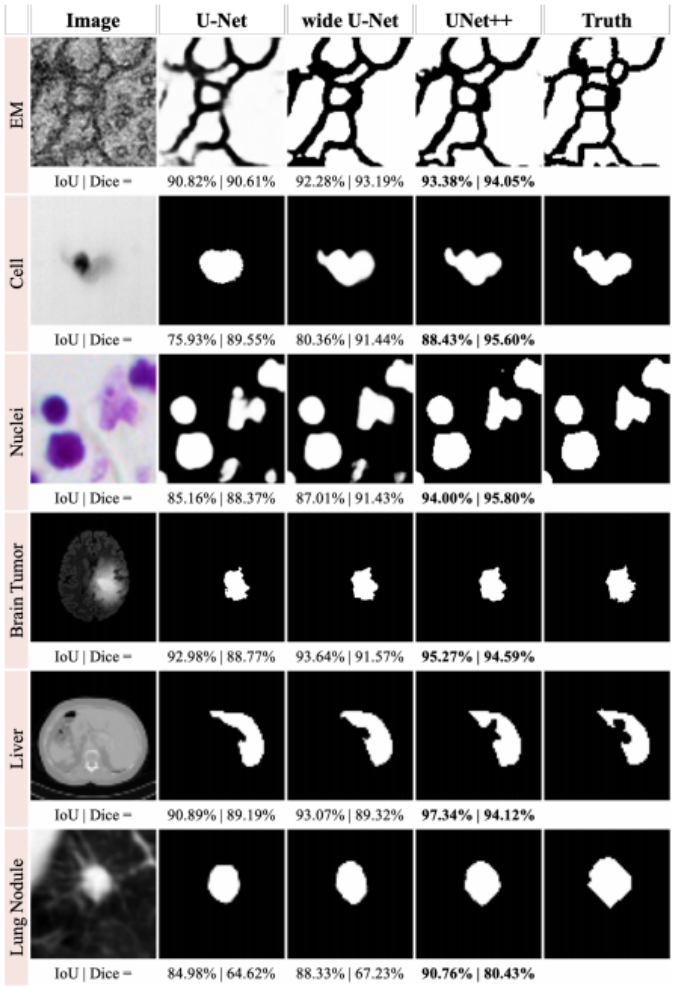
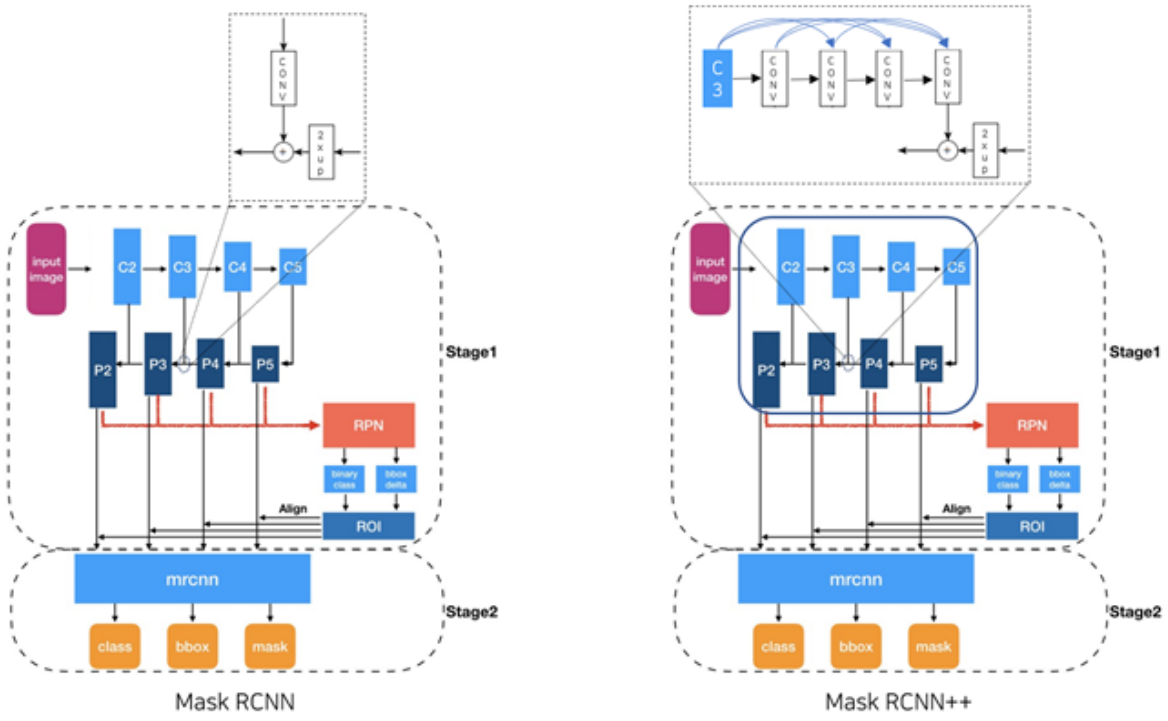


Fig. 3: Qualitative comparison among U-Net, wide U-Net, and UNet++; showing segmentation results for our six distinct biomedical image segmentation applications. They include various 2D and 3D modalities. The corresponding quantitative scores are provided at the bottom of each prediction (IoU | Dice).

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

4. Results

(2) Instance Segmentation Results



출처 : https://alittlepain833.medium.com/simple-understanding-of-mask-rcnn-134b5b330e95

TABLE V: Redesigned skip connections improve both semantic and instance segmentation for the task of nuclei segmentation. We use Mask R-CNN for instance segmentation and U-Net for semantic segmentation in this comparison.

Architecture	Backbone	IoU	Dice	Score
U-Net	resnet101	91.03	75.73	0.244
UNet++	resnet101	**92.55**	**89.74**	**0.327**
Mask R-CNN [12]	resnet101	93.28	87.91	0.401
Mask RCNN++†	resnet101	**95.10**	**91.36**	**0.414**

†Mask R-CNN with UNet++ design in its feature pyramid.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

5. Discussion
(1) Overview

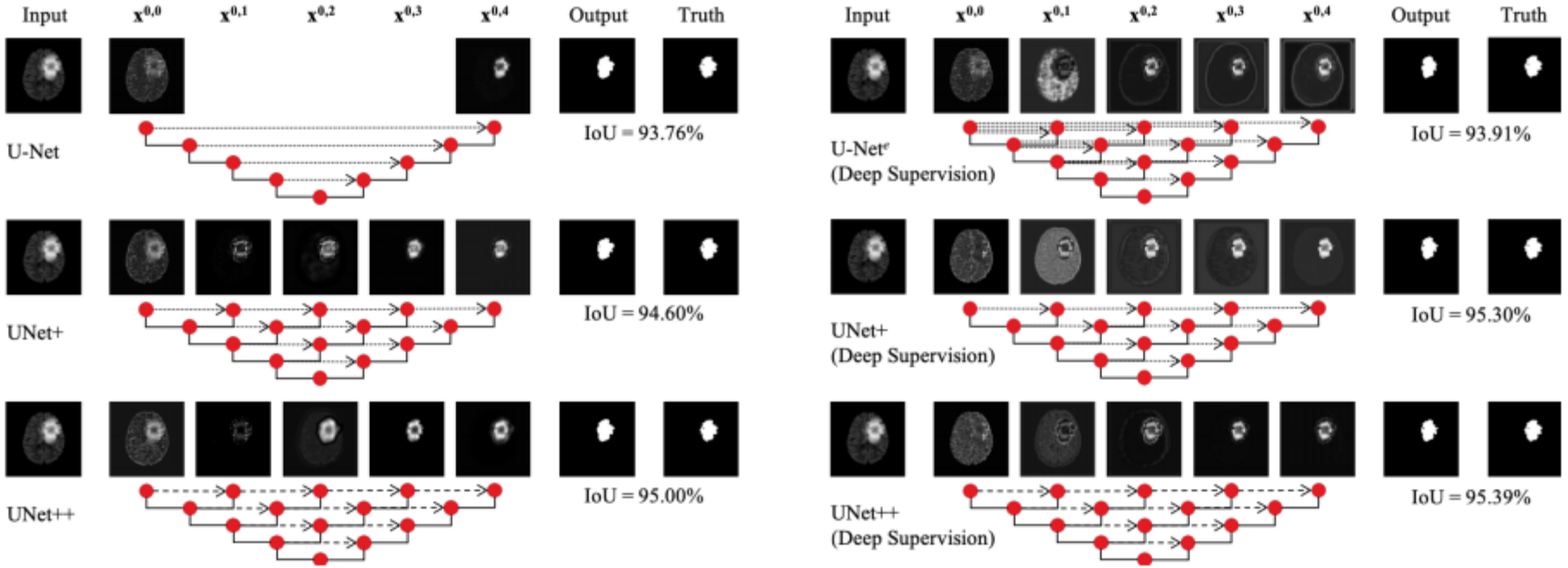
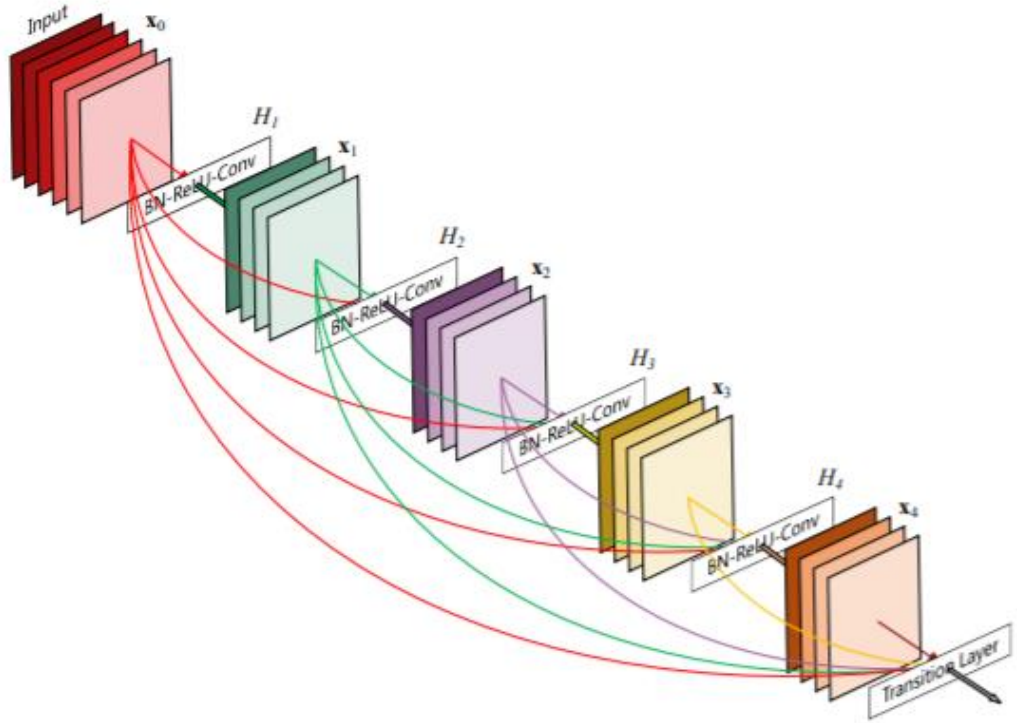


Fig. 8: Visualization and comparison of feature maps from early, intermediate, and late layers along the top most skip connection for brain tumor images. Here, the dot arrows denote plain skip connection in U-Net and UNet+, while the dash arrows denote dense connections introduced in UNet++.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

7. Conclusion

• (1) Advantages

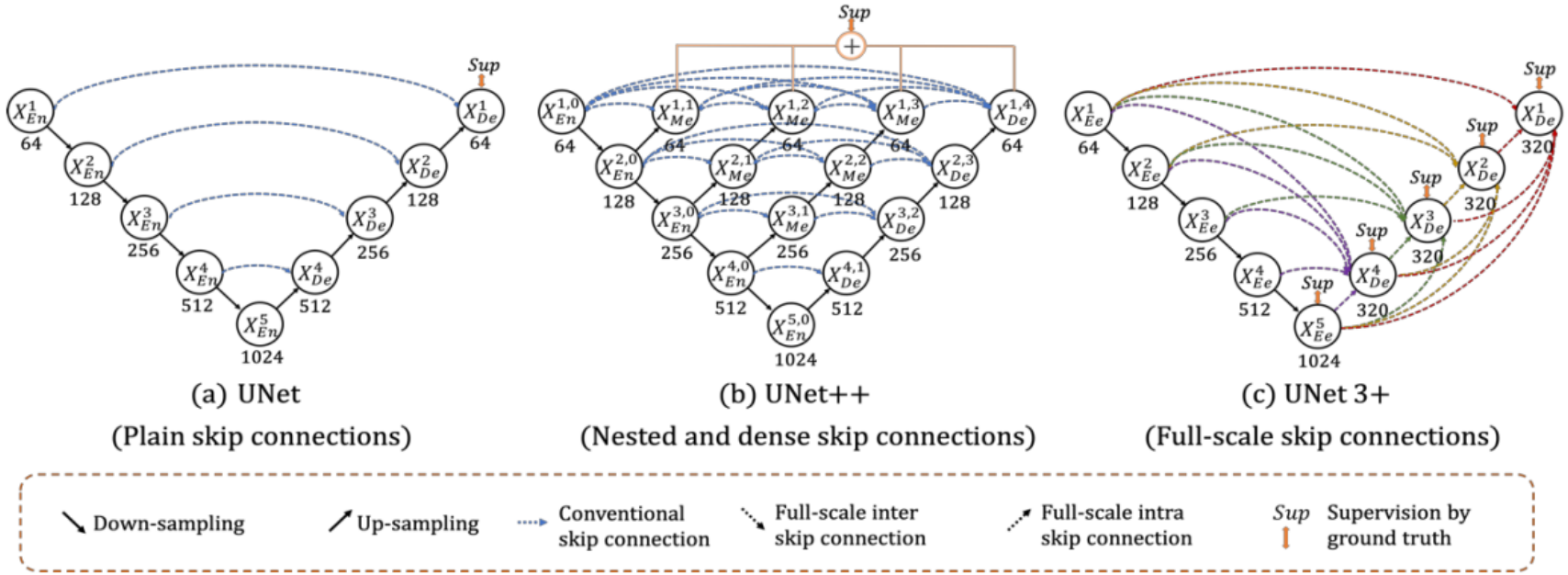


- Dense하게 연결하는 방식은 Densenet에서도 이미 나왔던 개념인데 이를 잘 접목해서 활용한 것 같습니다.
- wide U-Net을 도입해서 UNet++의 점수 상승 요인이 파라미터가 많아져서가 아닌 것을 보여주는게 인상깊습니다. 이 부분이 있기에 논리의 빈틈이 줄어든 것 같습니다.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

7. Conclusion

• (2) Disadvantages



- 처음에 읽었을 때는 Same Scale의 특징 맵의 한계를 극복 했다는게 다른 크기의 특징 맵도 결합해서 작은 객체와 큰 객체도 탐지했다는 의미인 줄 알았습니다. 하지만, 계속 읽어보니 인코더의 같은 계층의 노드만 사용한게 아니라 인코더와 디코더만 연결했다는 것을 같은 깊이의 노드를 전부 연결하고 업샘플링을 통해서 더 깊은 노드들의 정보도 활용한다는 의미 같습니다. 실제로 위와 같은 방법을 한계를 극복하기 위해서 UNet 3+라는 논문에서는 작은 특징맵과 큰 특징맵도 받아서 해결하려는 시도가 있었습니다.
- TABLE IV의 파라미터의 수를 보면 7.8M에서 9.0M으로 크게 증가하지는 않았지만 메모리 관점에서는 정보를 계속 저장해야하는 문제가 있습니다.

출처 : O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015

감사합니다