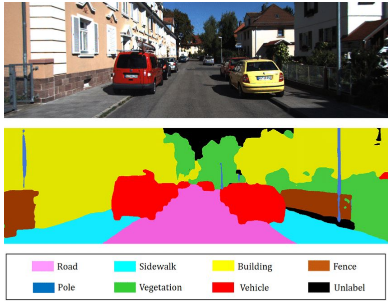
UNET 3+
A FULL-SCALE CONNECTED UNET
FOR
MEDICAL IMAGE SEGMENTATION

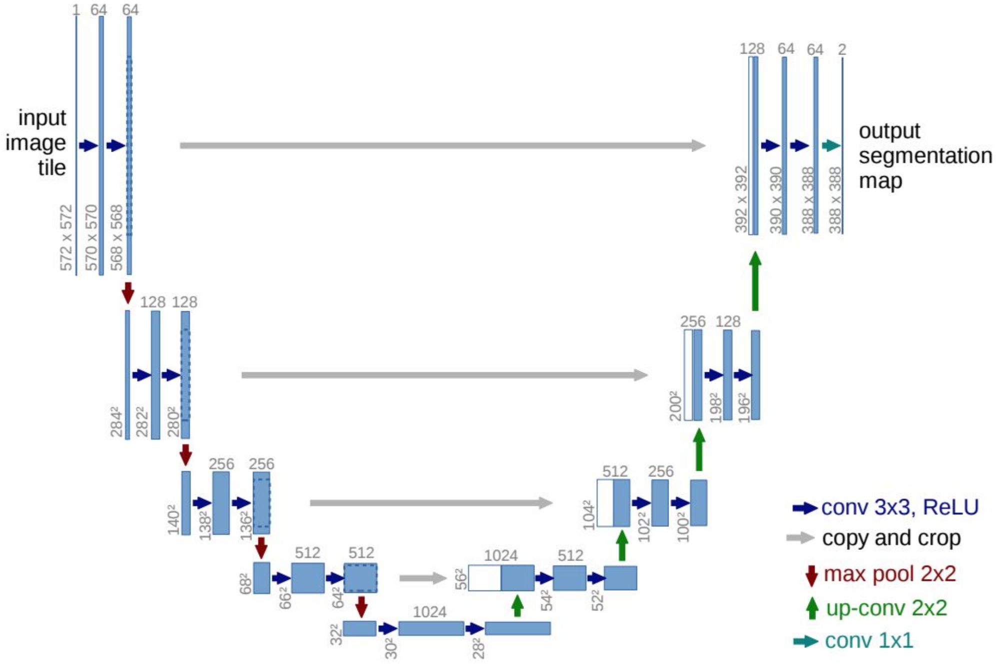
U-NET: CONVOLUTIONAL NETWORKS FOR BIOMEDICAL IMAGE SEGMENTATION-2015



- Partition the image into different segments which each of them represents a different entity.

- Convolutional Neural Networks gave decent results in easier image segmentation problems but it hasn't made any good progress on complex ones. That’s where UNet comes in the picture. **UNet was first designed especially for medical image segmentation.**
- CNN works well in classification problems as the image is converted into a vector which used further for classification. But in image segmentation, we not only need to convert feature map into a vector but also reconstruct an image from this vector.

U-NET: CONVOLUTIONAL NETWORKS FOR BIOMEDICAL IMAGE SEGMENTATION-2015



The u-net comprises of three parts an encoder/contraction path(left side), the bottleneck and a decoder/expansion path(right side).

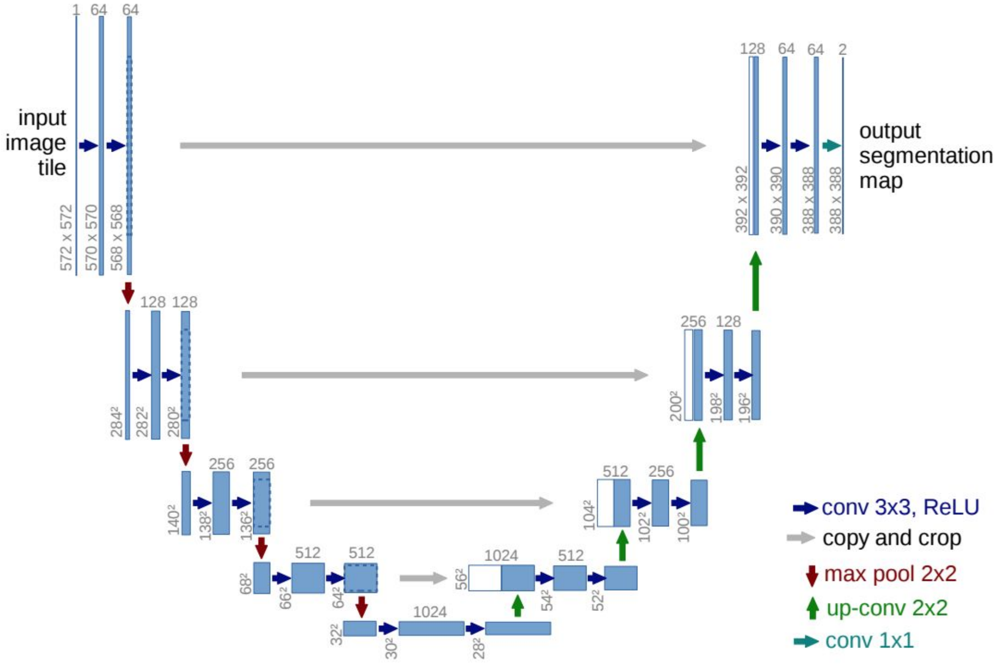
Contraction part

Each block takes an input applies two 3X3 convolution layers followed by a 2X2 max pooling. The number of kernels or feature maps after each block doubles so that architecture can learn the complex structures effectively.

Bottleneck

It mediates between the contraction layer and the expansion layer. It uses two 3X3 CNN layers followed by 2X2 upsampling layer.

U-NET: CONVOLUTIONAL NETWORKS FOR BIOMEDICAL IMAGE SEGMENTATION-2015



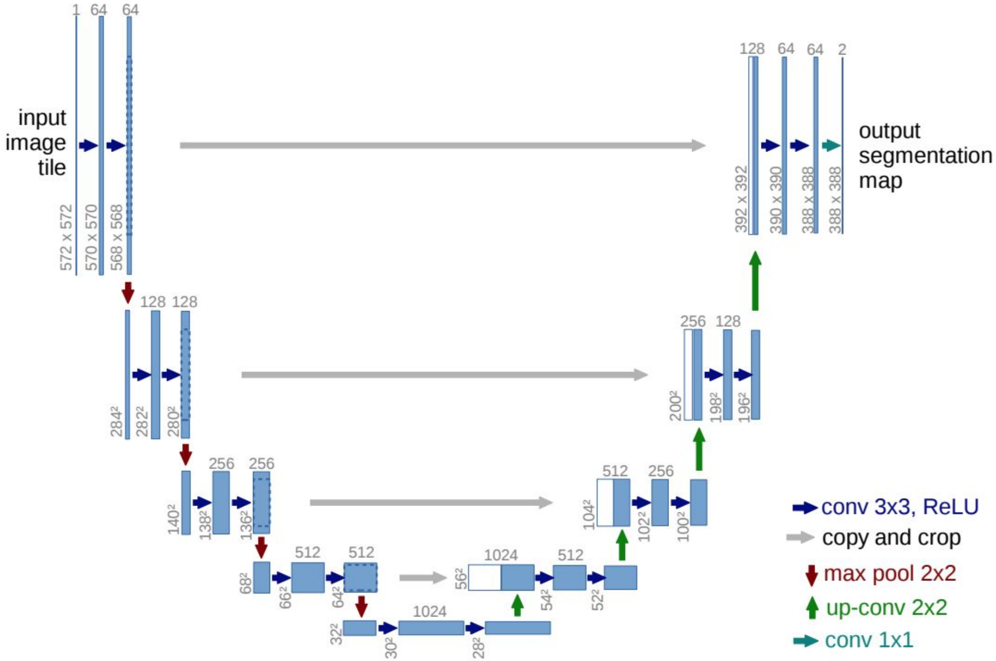
Expansion part

Each block passes the input to two 3X3 CNN layers followed by a 2X2 upsampling layer. Also after each block number of feature maps used by convolutional layer get half to maintain symmetry. However, every time the input is also get appended by feature maps of the corresponding contraction layer.

However, the above process **reduces the “where”** though it **increases the “what”**. That means, we can get advanced features, but we also loss the localization information.

Thus, after each up-conv, we also have **concatenation of feature maps (gray arrows) that are with the same level**. This helps to **give the localization information from contraction path to expansion path**.

U-NET: CONVOLUTIONAL NETWORKS FOR BIOMEDICAL IMAGE SEGMENTATION-LOSS FUNCTION

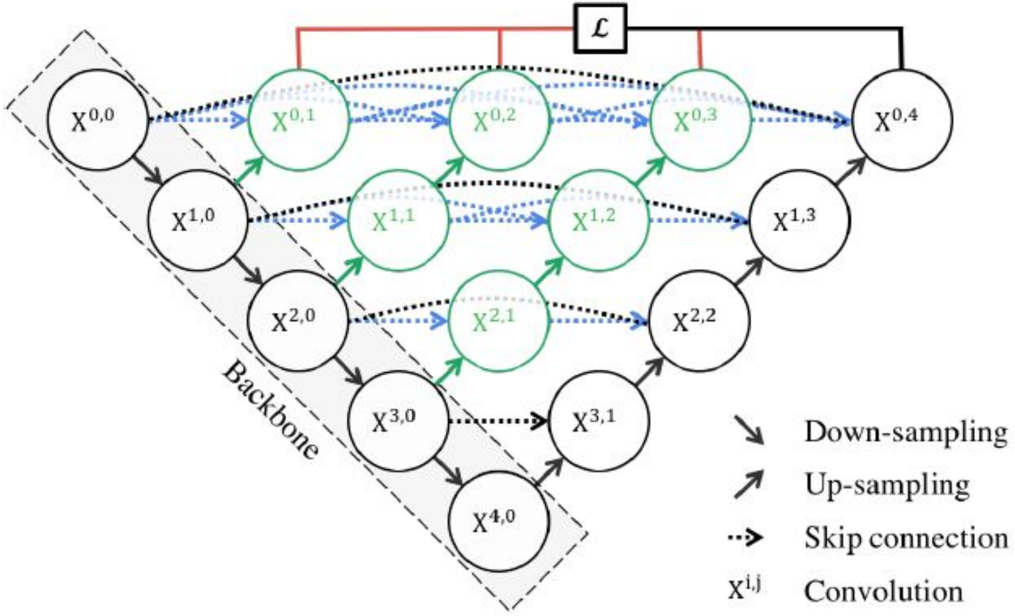


First of all pixel-wise softmax applied on the resultant image which is followed by cross-entropy loss function.

U-NET: CONVOLUTIONAL NETWORKS FOR BIOMEDICAL IMAGE SEGMENTATION-KEY FEATURES

1. **U-Net learns segmentation in an end-to-end setting.**
   You input a raw image and get a segmentation map as the output.
2. **U-Net is able to precisely localize and distinguish borders.**
   Performs classification on every pixel so that the input and output share the same size.**UNET is able to localise and distinguish borders is by doing classification on every pixel, so the input and output share the same size.**
3. **U-Net uses very few annotated images.**
   Data augmentation with elastic deformations reduces the number of annotated images required for training.
   - Along with the usual shift, rotation, and color adjustments, they added elastic deformations. This was done with a coarse (3x3) grid of random displacements.
   - This allows the network to learn invariance to such deformations, without the need to see these transformations in the annotated image corpus. This is important in biomedical segmentation since deformation is the most common variation in tissue and realistic deformations can be simulated efficiently.

U-NET++: A Nested U-Net Architecture



UNet++ Architecture

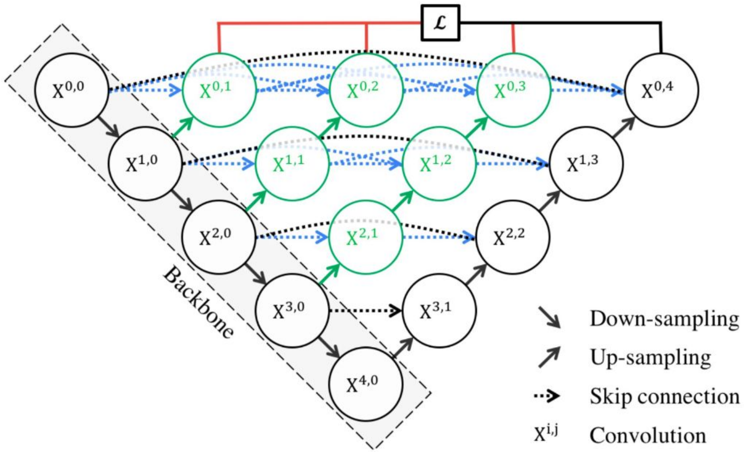
UNet++ have 3 additions to the original U-Net:

1. redesigned skip pathways (shown in green)
2. dense skip connections (shown in blue)
3. deep supervision (shown in red)

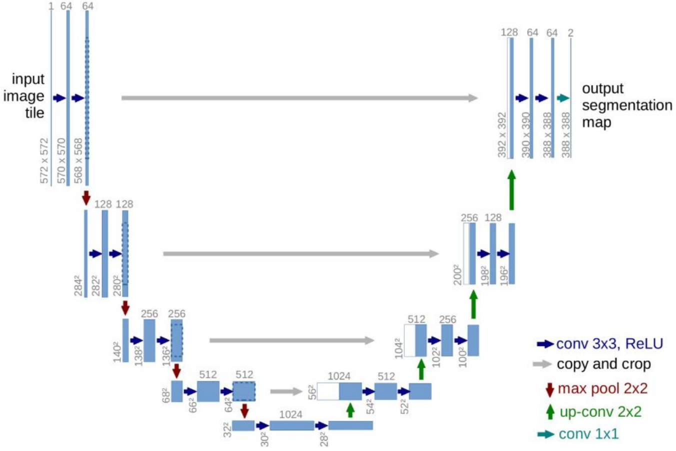
UNet++ architecture gives image segmentation accuracy more than the UNet.

U-NET++: Redesigned Skip Pathways (In Green)

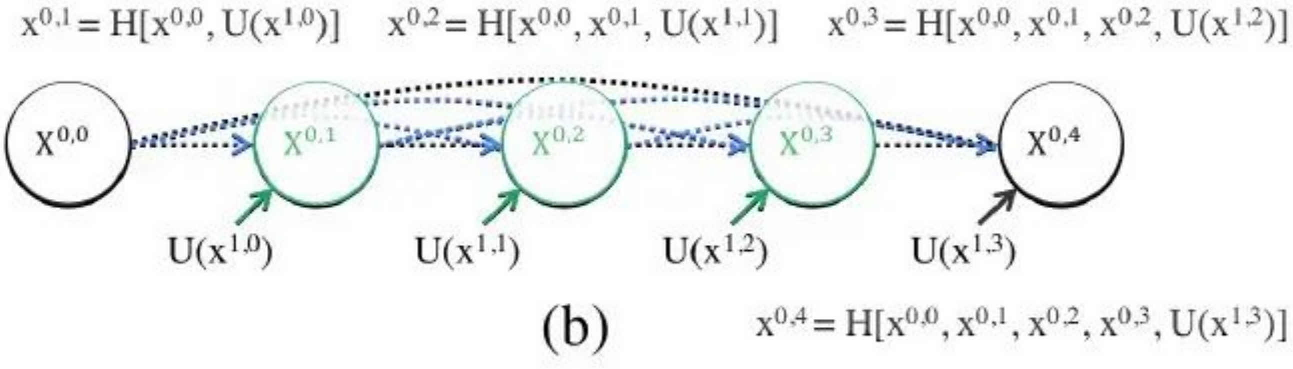
UNet++



U-Net



U-NET++: Redesigned Skip Pathways (In Green)


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

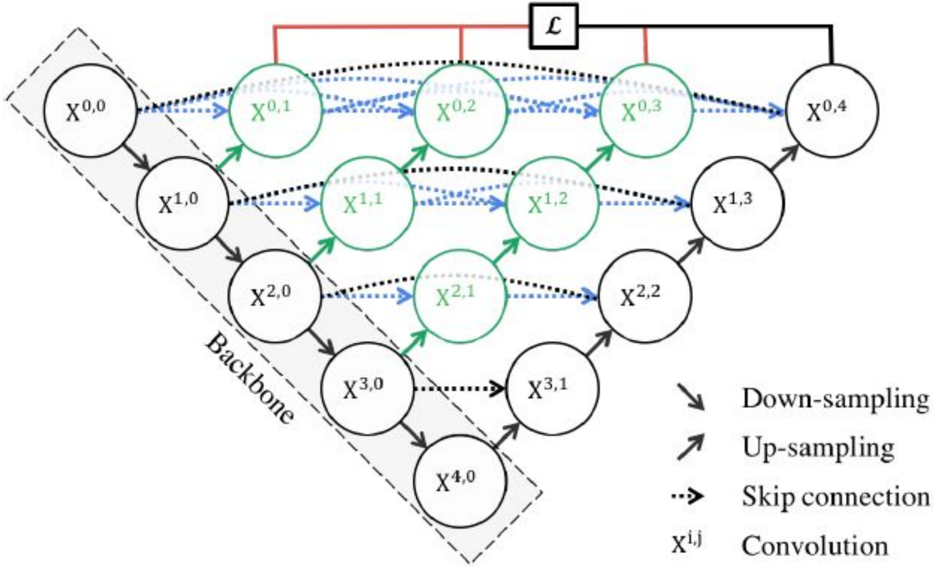
(b)

Each convolution layer is preceded by a concatenation layer that fuses the output from the previous convolution layer of the same dense block with the corresponding up-sampled output of the lower dense block.

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

where *H*() is a convolution operation followed by an activation function, *U*() denotes an up-sampling layer, and [ ] denotes the concatenation layer.

U-NET++: Dense Skip Connections (In Blue)

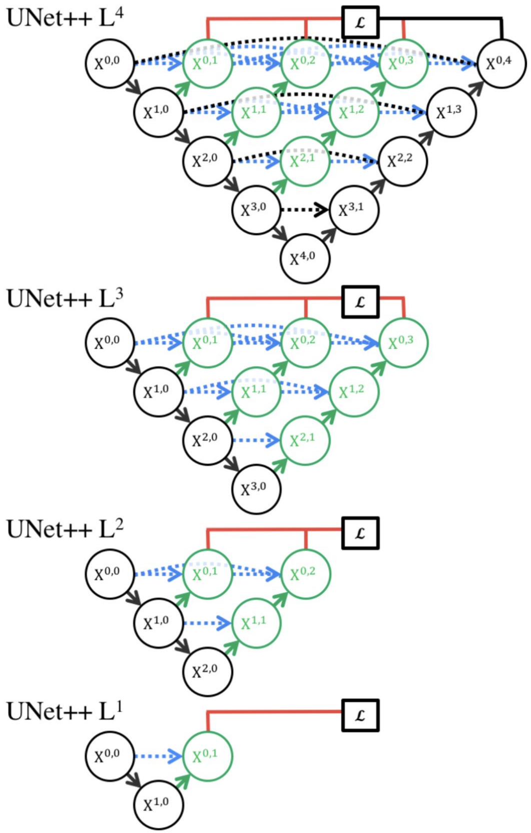


UNet++ Architecture

1. Dense blocks are inspired by DenseNet with the purpose to improve segmentation accuracy and improves gradient flow.

2. Dense skip connections ensure that all prior feature maps are accumulated and arrive at the current node because of the dense convolution block along each skip pathway. **The main idea behind is to bridge the semantic gap between the feature maps of the encoder and decoder prior to fusion.**

U-NET++: Deep Supervision (In Red)



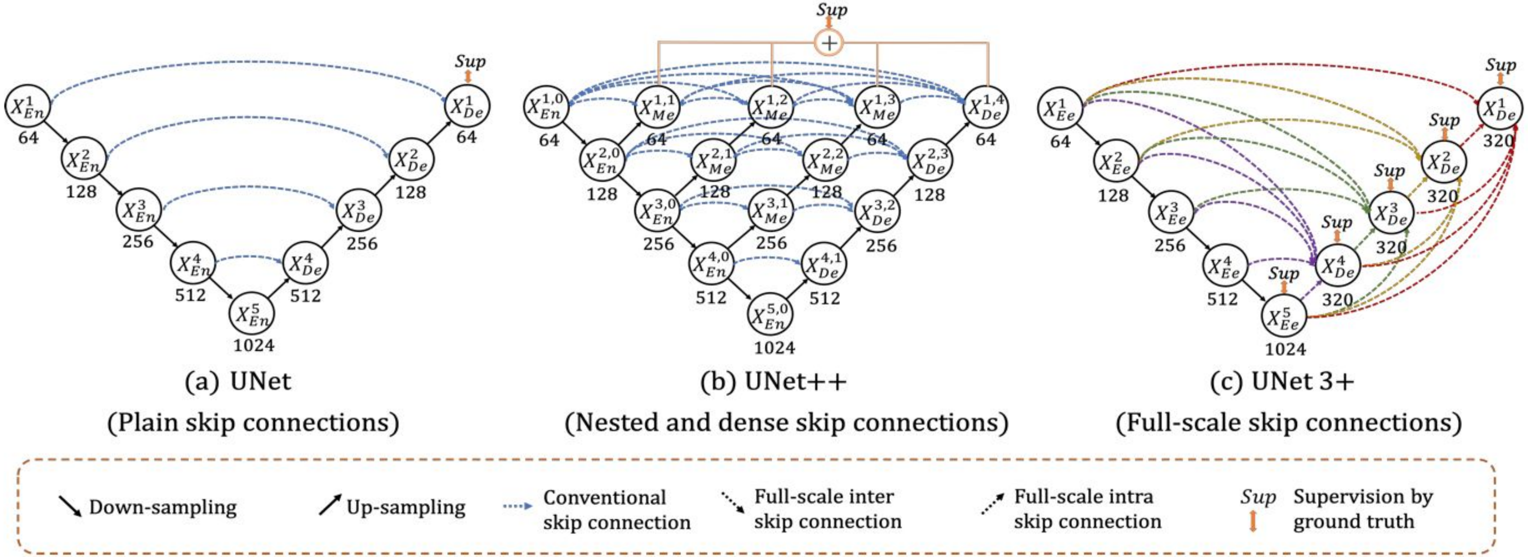
Deep supervision was introduced to the model so that it could decide the level of model pruning and speed gain.

This enabled the model to operate in two modes:

1) **accurate mode** wherein the outputs from all segmentation branches are averaged;

2) **fast mode** wherein the final segmentation map is selected from only one of the segmentation branches.

U-NET 3+:A Full-Scale Connected UNET For Medical Image Segmentation-Full-Scale Skip Connections-2020



(a) UNet
(Plain skip connections)

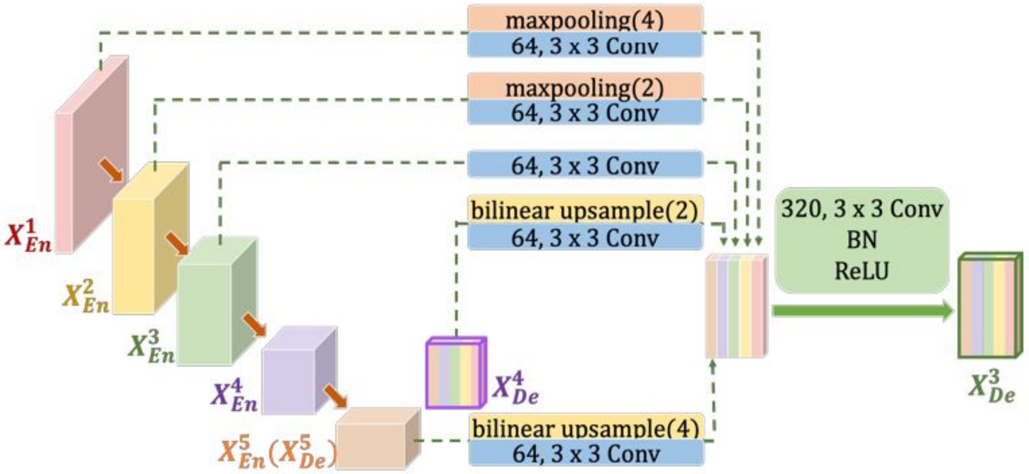
(b) UNet++
(Nested and dense skip connections)

(c) UNet 3+
(Full-scale skip connections)

Left: UNet, Middle UNet++, Right: UNet 3+

Full-scale skip connections
Full-scale Deep Supervision
Classification-guided Module (CGM)

U-NET 3+:A Full-Scale Connected UNET For Medical Image Segmentation--1.Full-Scale Skip Connections



Redesigned Skip Connection of U-Net++
in U-NET3+ as Full-Scale Skip Connection
shows how the 3rd level decoder get its input

Similar to the UNet, the feature map from the same-scale encoder layer 3 are directly received in the decoder. In contrast to the UNet, a set of inter encoder-decode skip connections delivers the low-level detailed information from the smaller-scale encoder layer 1 and 2(by applying non-overlapping max pooling operation; while a chain of intra decoder skip connections transmits the high-level semantic information from larger-scale decoder layer 4 and 5 , by utilizing bilinear interpolation

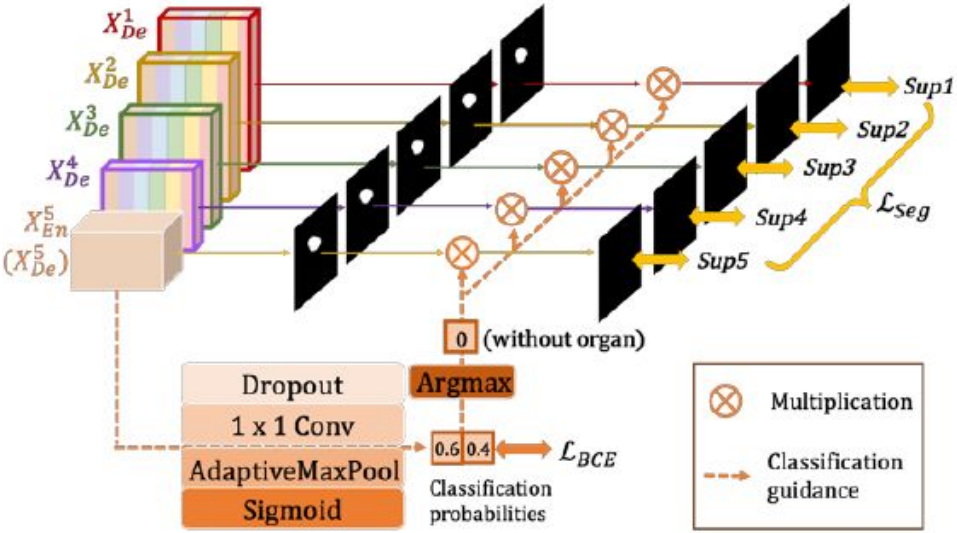
feature aggregation mechanism on the concatenated feature map from five scales, which consists of 320 filters of size 3 × 3, a batch normalization and a ReLU activation function.

U-NET 3+:A Full-Scale Connected UNET For Medical Image Segmentation--2.Full-Scale Deep Supervision

2. Full-Scale Deep Supervision

- In order **to learn hierarchical representations from the fullscale aggregated feature maps**, the **full-scale deep supervision** is further adopted in the UNet 3+.
- Compared with the deep supervision performed on generated full-resolution feature map in UNet++, the proposed **UNet 3+ yields a side output from each decoder stage, which is supervised by the ground truth.**
- The proposed UNet 3+ yields a side output from each decoder stage, which is supervised by the ground truth. **To realize deep supervision, the last layer of each decoder stage is fed into a plain 3 × 3 convolution layer followed by a bilinear up-sampling and a sigmoid function**

U-NET 3+:A Full-Scale Connected UNET For Medical Image Segmentation--3.Classification-Guided Module(CGM)



Full-Scale Deep Supervision with Classification-Guided Module (CGM) will try to detect **whether the object is present first**, before trying to segment the organ.

- As shown in figure, after passing a series of operations including dropout, convolution, max pooling and sigmoid, **a 2-dimensional tensor is produced from the deepest-level *X5En*, each of which represents the probability of with/without organs.**
- With the help of the **argmax function**, **2-dimensional tensor is transferred into a single output of {0,1}**, which denotes with/without organs.
- Subsequently, **the single classification output is multiplied with the side segmentation output.**
- **Binary cross entropy loss** function is used to train the CGM.

U-NET 3+:A Full-Scale Connected UNET For Medical Image Segmentation-2020

2.2 Loss Function

- The paper proposed a new composite loss function in order to further exploit the full-scale information.

$$\ell_{seg} = \ell_{fl} + \ell_{ms-ssim} + \ell_{iou}$$

- The new loss function is defined as the sum of **focal loss (fl)**, **multi-scale structural similarity index loss (ms-ssim)**, and **intersection over union (IoU) loss** (ms-ssim loss will panelize the fuzzy organ boundary predictions heavier, and therefore enhance the organ boundary segmentation.)

$$\ell_{ms-ssim} = 1 - \prod_{m=1}^{M} \left(\frac{2\mu_p\mu_g + C_1}{\mu_p^2 + \mu_g^2 + C_1} \right)^{\beta_m} \left(\frac{2\sigma_{pg} + C_2}{\sigma_p^2 + \sigma_g^2 + C_2} \right)^{\gamma_m}$$

U-NET 3+:A Full-Scale Connected UNET For Medical Image Segmentation-2020

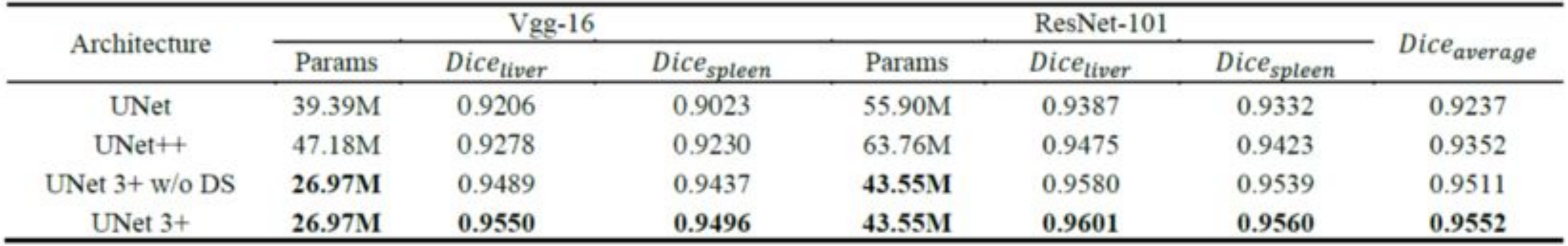
Experimental Results

Datasets

- The method was validated on two organs: the liver and spleen.
- The dataset for **liver segmentation** is obtained from the **ISBI LiTS 2017 Challenge**. It contains 131 contrast-enhanced 3D abdominal CT scans, of which 103 and 28 volumes are used for training and testing, respectively.
- The **spleen dataset** from the hospital passed the ethic approvals, containing 40 and 9 CT volumes for training and testing.
- Images are cropped to 320×320.

U-NET 3+:A Full-Scale Connected UNET For Medical Image Segmentation-2020

Comparison of UNet, UNet++, the proposed UNet 3+ without deep supervision (DS) and UNet 3+ on liver and spleen datasets in terms of Dice metrics. The best results are highlighted in bold. The loss function used in each method is focal loss.



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**

- Based on the backbone of Vgg-16 and ResNet-101, Table 1 compares UNet, UNet++ and the proposed UNet 3+ architecture in terms of the number of parameters and segmentation accuracy on both liver and spleen datasets. As seen, UNet 3+ without deep supervision achieves a surpassing performance over UNet and UNet++, UNet 3+ combined with full-scale deep supervision further improved.

U-NET 3+:A Full-Scale Connected UNET For Medical Image Segmentation-2020

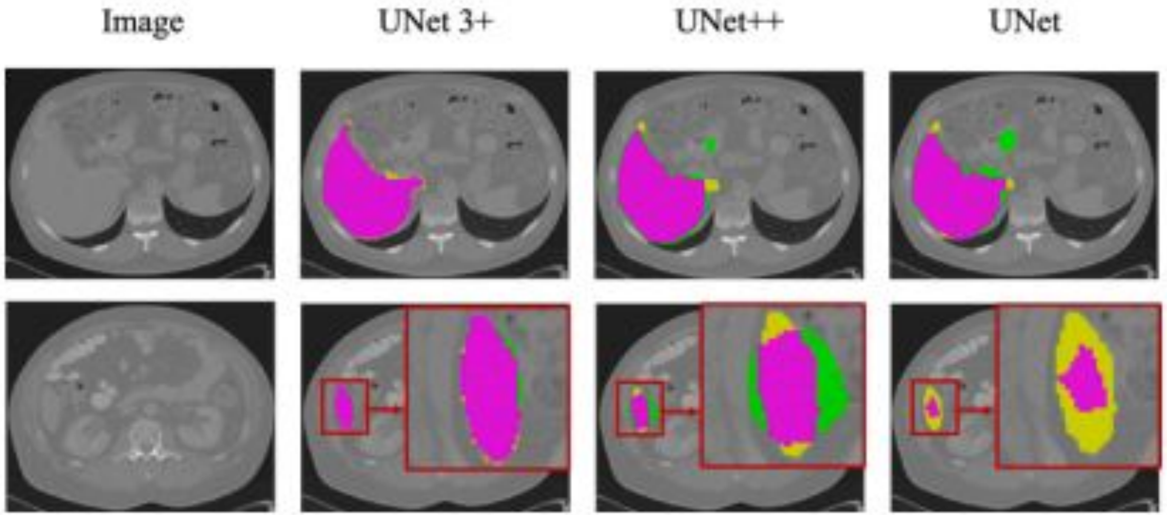
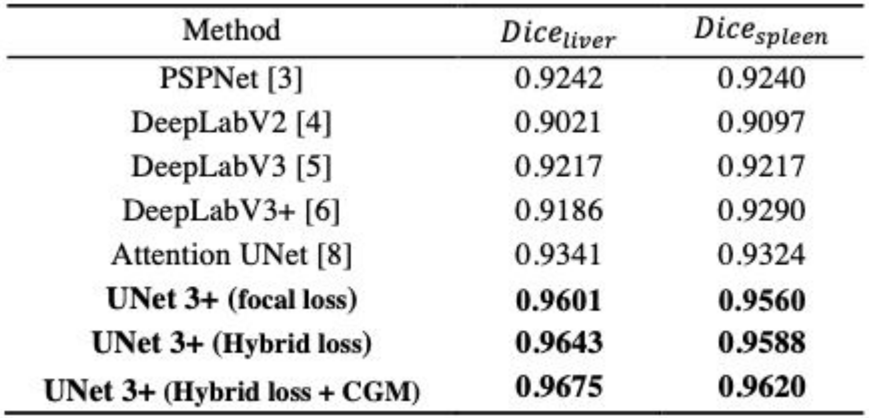


Fig. 4: Qualitative comparisons of ResNet-101-based UNet, UNet++, and proposed UNet 3+ on liver dataset. Purple areas: true positive (TP); Yellow areas: false negative (FN); Green areas: the false positive (FP).

The segmentation results of ResNet-101-based UNet, UNet++ and UNet 3+ with full-scale deep supervision on liver datasets. It can be observed that our proposed method not only **accurately localizes organs but also produces coherent boundaries**, even in small object circumstances

U-NET 3+:A Full-Scale Connected UNET For Medical Image Segmentation-2020



Method	$Dice_{liver}$	$Dice_{spleen}$
PSPNet [3]	0.9242	0.9240
DeepLabV2 [4]	0.9021	0.9097
DeepLabV3 [5]	0.9217	0.9217
DeepLabV3+ [6]	0.9186	0.9290
Attention UNet [8]	0.9341	0.9324
UNet 3+ (focal loss)	**0.9601**	**0.9560**
UNet 3+ (Hybrid loss)	**0.9643**	**0.9588**
UNet 3+ (Hybrid loss + CGM)	**0.9675**	**0.9620**

Comparison with State of the Art

Comparison of UNet 3+ and other 5 state-of-the-art methods. The best results are highlighted in bold

The proposed **hybrid loss function greatly improves the performance** by taking pixel-, patch-, map-level optimization into consideration.

Moreover, taking advantages of the **classification-guidance module** (CGM), UNet 3+ skillfully **avoids the over-segmentation in complex background.**

Finally, **UNet 3+ outperforms Attention UNet, PSPNet, DeepLabV2, DeepLabV3 and DeepLabv3+.**

THANKS FOR LISTENING, THOSE OF YOU WHO LISTENED.