IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Survey report on U-Net and its Variants

[1]Anitha M, [2]Chaitra P C, [3]Chethana T S, [4]Deepa N C, [5]Gurijala Pranathi

[1]Assistant Professor, [2]Assistant Professor, [3]Student, [4]Student, [5]Student

[1]Department of Computer Science and Engineering

[1]Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India

[4]Dayananda Sagar College of Engineering, Bengaluru, Karnataka, India

Abstract: Carotid artery plagues are common in many individuals. They can cause serious problems like blood clots, etc. This paper stretches an overview of the image segmentation procedures that can be accustomed to scrutinize the boundary of the carotid artery. These algorithms use 2D and 3D images. They explain the recent advancements in CNN and Unet architectures. The modifications done with respect to the original Unet architecture are shown. The changes in each step of the architecture and how it improves the model are illustrated.

***Index Terms* - Unet, encoder, decoder, CNN, Artificial neural networks**

I. INTRODUCTION

Carotid arteries supply blood to the brain, neck, and face. There are two arteries the left and the right one. They are a major part of the circulatory system. They send oxygenated blood to tissues in your head, neck, and your brain. Blood clots in carotid arteries can stop the process of pumping blood and cause serious problems.

Carotid arteries may have a build-up of plaque which can contain fat, cholesterol, calcium, and other substances. Early detection of a narrow carotid artery can reduce stroke risk. It can also help us evaluate blood flow through the artery after the removal of the plague. The data about the carotid artery can help us locate any blood clots and the placement of the stent.

Image processing methods can be used on the images of the carotid artery to obtain a clear boundary of it. It helps to evaluate whether the artery is healthy or not. Image processing methods such as image segmentation can be beneficial to analyze the artery. Image segmentation is a procedure used to trace an object and create a boundary around the object. It is a method of grouping pixels in an image into distinctive categories. Image segmentation can include traditional techniques such as thresholding, region-based segmentation, edge-based, or clustering.

Semantic segmentation of images by use of deep learning techniques like neural networks can help on taming the segmentation of images. Artificial neural networks consist of three different layers with nodes and these layers are- the input layer, hidden layers that can either be one or more than one layer, and the output layer. They contain connected nodes with weights associated with them and data passes through those nodes only when they are activated. Convolutional neural networks are specifically used to process pixel data. They are used in object detection and segmenting images.

Convolutional neural networks contain three different layers which are convolution layer, the pooling layer, and the fully connected layer. Feature extraction is deployed by convolution layer. The convolution layer deals with obtaining large features from the image. It uses a kernel to do element-wise multiplication operations also called Hadamard product. The kernel can be applied with or without padding around the image. Without padding, the image size decreases and it is called the same padding. With padding, the image size remains the same and it is called valid padding. The convolution layer may consist of more than one layer.

There are a couple of pooling techniques. Besides max pooling, average pooling is one such technique. Max pooling produces the determined value from the image portion enclosed by the image and average pooling produces the average value of the image portion enclosed by the kernel. A fully connected layer accomplishes the classification based on the feature extracted from the prior layers.

II. UNET AND ITS VERSIONS

UNet is an FCNN or a fully convolutional neural network that is U-shaped. UNet was originally developed to replace the typical sliding window technique of convolutional neural networks. UNet was found to possess increased accuracy with very few input images, this made UNet more reliable in image processing and segmentations specifically medical images which are usually ultrasound or MRI images in JPEG, PNG, or DICOM format. UNet: Convolutional Networks used for Biomedical Image Segmentation, elaborates on the typical UNet architecture [1].

The below table compares the different Unet models illustrated in this paper.

Table 2.1 Comparison of Unet architectures

Architecture	Dice coefficient	Modifications	Improvements
Unet	77.5%	No modifications	generates higher-level feature maps
Unet++	82.90%	The skip connections among the contracting and the expanding paths	full-resolution feature maps which can be used to enhance the quality of the output obtained
Unet 3+	95.52%	makes use of full-scale skip connections and allowing it to chain the low-level details with high-level semantics of the feature maps on various scales	record fine-grained details with coarse-grained semantics on full scales
Swin-Unet	90.00%	It contains transformers to produce excellent performance and generalization ability	It has excellent performance and generalization ability
RIC Unet	80.00%	Residual blocks, multi-scale, and channel attention processes are put into use.	This network is better because of cost effectiveness and better diagnosis.
Residual Unet	94.28%	before every down-sampling, the channels are doubled compared to the upper layer; at the same time, before every up-sampling, the channels of the convolution kernel is halfed, hence effectively reducing the bottleneck of the model	The model gets a higher mean value with a minor standard deviation on every segmentation.
DU Unet	72.31%	It adds 2D and 3D convolutions in encoding stage.	This model improves performance compared to the 2D model and effectively utilizes 3D information.
Fully Dense Unet	87.00%	the input image goes through multilevel decomposition in the contracting path of the FD-UNet	The FD-UNet is observed to be a better and more compact CNN.
B-U-Unet	94.54%	This model involves the addition of a specification layer in amongst each inner layer and activation layer and between different convolution layers and activation layers to improve the Unet model.	Usage of this model decreases the processing time and improves the quality of the image

R2U-net	86.16%	Normal connections are replaced by recurrent Convolutional layer (RCL), based on RCNN and RRCNN, and perform concatenation of complete feature maps instead of reducing them and then copying them from the encoding path to the decoding path.	performs in-detail feature extraction and also makes efficient use of time steps to retrieve better and strong features.

2.1 UNet

The UNet architecture contains four convolutional sub-networks in the encoding or contracting path. Similarly, there are four identical convolutional sub-networks in the decoding or expanding path, consistent with the encoding path forming the U-shaped architecture hence this architecture or network is called UNet. Each sub-network consists of three convolutional layers which in turn contain many channels. On the encoding path, the sum of channels or feature maps increases with the decrease in image size by half and this is called down-sampling. While in the decoding path, the reverse happens where the image size is doubled as the sum of feature maps or channels decreases, which is called up-sampling. Here, the blue arrows represent the activation function and 3x3 filters that are applied to the images passing through each channel or layer. Here we use the ReLU activation function. ReLU (rectified linear activation function), is a non-linear function or piecewise linear function which outcomes the input directly if it is positive, otherwise, results zero. It is generally the most commonly used activation function in convolutional neural networks and multilayer perceptron. The red arrows represent the 2x2 maximum pooling, this operation selects the maximum element in the feature maps. The Gainsboro arrows corresponds to the skip connections that perform concatenation of the feature maps from the encoding sub-networks to the decoding sub-networks. The feature map passed down by the contracting path is cropped to match the dimension for concatenation. These skip connections make the UNet architecture unique from the other convolutional neural networks. The green arrows correspond to the 2x2 up-sampling operation [1]. Finally, the light blue arrow at the final layer in the decoding path represents the 1x1 convolution which is employed in mapping each 64- component feature vector to the anticipated number of classes [1]. On whole, the network comprises of 23 convolutional layers [1]. The final layer apportions a class label to each pixel value.

The UNet architecture gradually generates higher-level feature maps as the network gets deeper. The final feature map from the series of convolutional layers contains high-level and contextual information. And hence, the output is of high resolution.

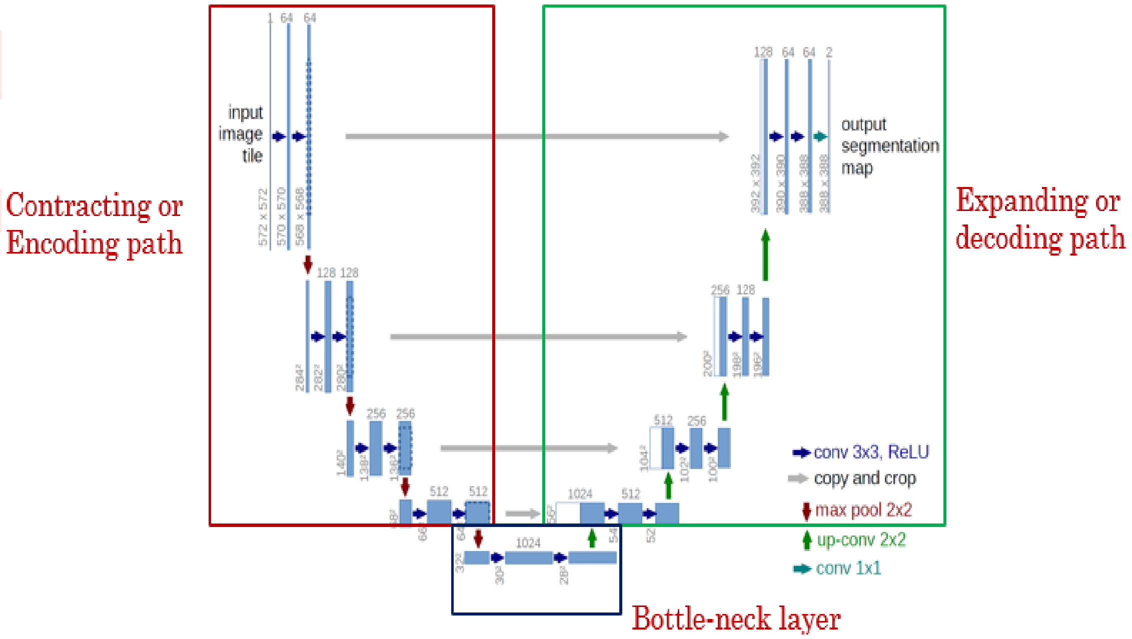

Fig. 1. *Unet*

2.2 UNet++

UNet++ is an advanced architecture of UNet. It was evident through their recently published paper that UNet++ architecture is much more advanced than normal UNet architecture. The skip connections amongst the contracting path and the expanding path mark the major difference between the two architectures. The skip connections have always been successful in retrieving highly-detailed elements and developing segmentation filters even for highly complicated images [2]. The UNet++ architecture contains a cluster of sub-networks in amongst the contracting and expanding paths instead of simple skip connections. And these clusters of sub-networks are in turn connected through many skip connections. Due to this, the input image passes through many sub-networks, channels, and filters within each sub-network in the cluster. This allows the output image to be more detailed and also This unique architecture of UNet++ makes it easier to produce output, as it gives accurate output on passing a smaller number of input images.

Just like in base UNet architecture, the UNet++ architecture contains the contracting path with an encoder sub-network and then an expanding path that consists of a decoding sub-network [2]. There is no bottle-neck layer here. The contracting path is also called the backbone of the UNet++ architecture [2]. As mentioned before, UNet++ consists of "Re-designed skip connections" [2] which makes it distinctive from
the original architecture of UNet. Because of the presence of these re-designed skip connections, the feature maps taken from the encoder sub-networks are not bluntly gotten by the decoding sub-networks. Instead, they pass through a cluster of convolutional network layers before reaching the decoding sub-network [2]. The number of sub-networks present between the two paths is determined based on the pyramid level [2]. As shown in the figure below (Fig. 2.), the number of skip connections between $X^{0,0}$ and $X^{0,4}$ is three, which include, $X^{0,1}$, $X^{0,2}$, and $X^{0,3}$.
The general mathematical representation of this is as follows:

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

Here,
i – represents the index of the contracting path sub-layer
j – represents the index of the sub-networks in the re-designed skip connections
H(·)- represents convolutional and activation function operations
U(·)- represents the up-sampling/decoding or expanding path layers.[2]

The proposed architecture makes use of deep supervision techniques which involve accurate mode and fast mode [2] while the base UNet architecture implements data aggregation techniques [1].

Overall, the output obtained is more detailed using UNet++ when compared to UNet, as the UNet++ architecture develops full-resolution feature maps [2] which can be used to enhance the quality of the output obtained.

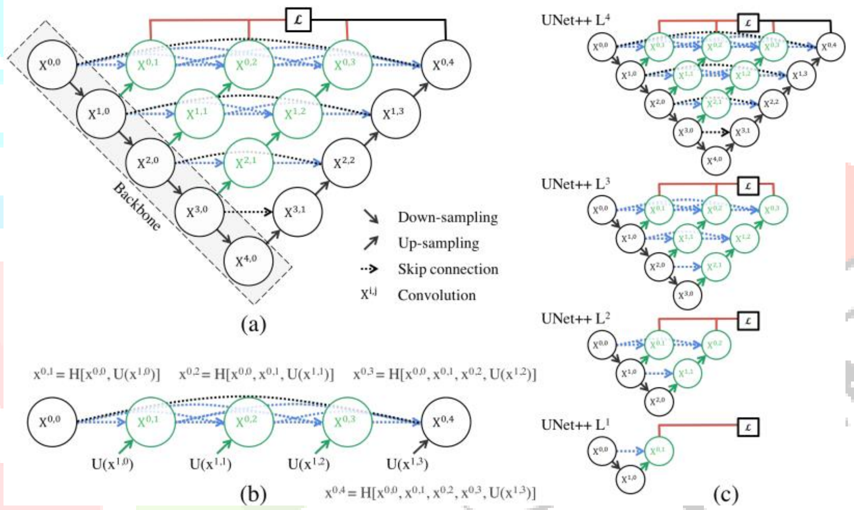

Fig. 2. *a) Contracting/encoding path or backbone and expanding or decoding path of UNet++ architecture [2], b) re-designed skip connections in-detail [2], c) UNet++ pruned at inference time using deep supervision techniques [2].*

2.3 UNet3+

UNet++ can sometimes be a complex architecture. Especially for the datasets that may require many convolutional layers with many clusters of sub-networks connected by several skip connections amongst the contracting path and the expanding paths of the UNet architecture. Even though UNet++ produces an accurate and detailed output when compared to the original UNet architecture, it does not utilize the advantages offered by "Full-Scale skip connections" [3] failing to learn the accurate position and outline of the image.

UNet3+ is a novel architecture[3]. The basic idea of this UNet3+ architecture is to simplify the UNet++ architecture and simultaneously produce more accurate output when compared with the outputs produced by UNet and UNet++. UNet3+ architecture makes use of full-scale skip connections amongst the contracting path and the expanding path allowing it to chain the low-level details and the high-level semantics of the feature maps on various scales [3].

UNet3+ also makes use of deep supervision techniques, where, the architecture picks up tiered representations from the full-scale grouped feature maps [3]. The projected method works efficiently, particularly for images with varying scales [3]. With increased accuracy, UNet3+ architecture also reduces the network parameters due to its simplified architecture and hence increases computation efficiency [3]. The paper [3] also discusses the hybrid loss function and draws up a classification-guided module which is essential in improving the image outline besides decreasing over-segmentation, which in turn produces an accurate output [3]. The figure below (***Fig. 3***.) shows the UNet3+ architecture in comparison with UNet and UNet++ architectures.

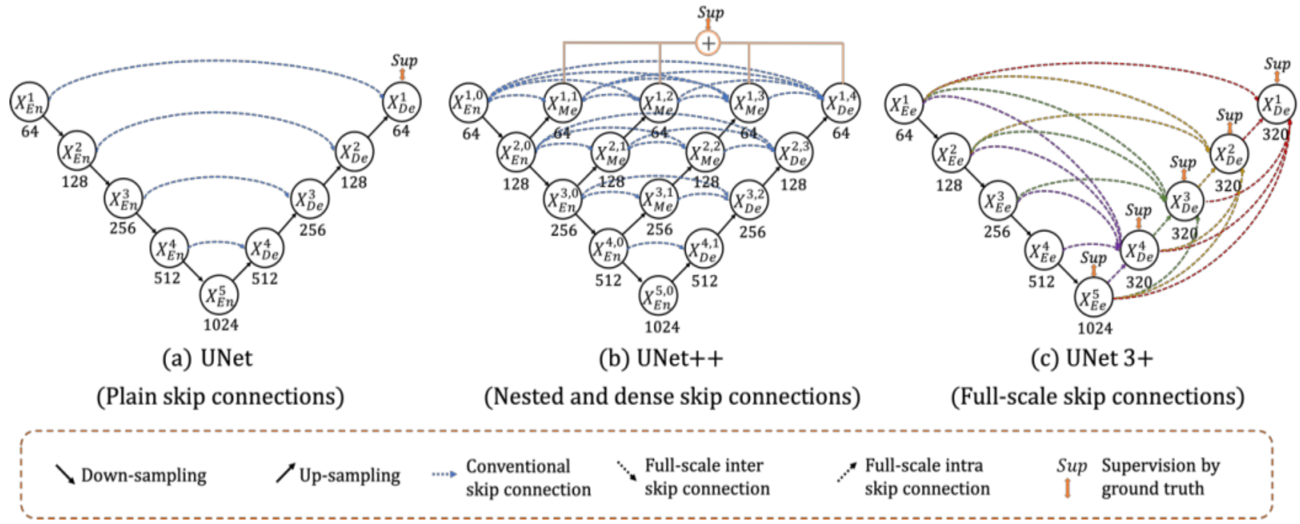

***Fig. 3**. a) UNet architecture with plain skip connections [3], b) UNet++ architecture including nested and dense skip connections [3], and c) UNet3+ architecture with full-scale skip connections [3]*

These full-scale skip connections not only connect sub-networks in the contracting path with the sub-networks in the expanding path but also connect sub-networks within the expanding or decoding path with each other [3]. This way UNet3+ combines minor and same-scale feature maps in the encoder with the decoder, it also combines the larger-scale feature maps within the decoder sub-networks, in order to record fine-grained details and coarse-grained semantics on full scales [3].

2.4 Swin-Unet

Swin-Unet is a type of Unet architecture in the medical field. It contains transformers to produce excellent performance and generalization ability. Components of Swin-Unet include encoder, bottleneck, decoder, and also skip connections. The Swin Transformer block is used to build the encoder, bottleneck, and decoder. The non-overlapping image patches are formed by input medical images. The Transformer-based encoder is fed with tokens to learn deep feature representations. Tokens are patches. The decoder up-samples the extracted context features using patch expanding layer[4]. These are combined with the multi-scale features in the encoder through skip connections to retain spatial resolution in feature maps and move to segmentation next.

The figure shows the overall architecture of Swin-Unet.

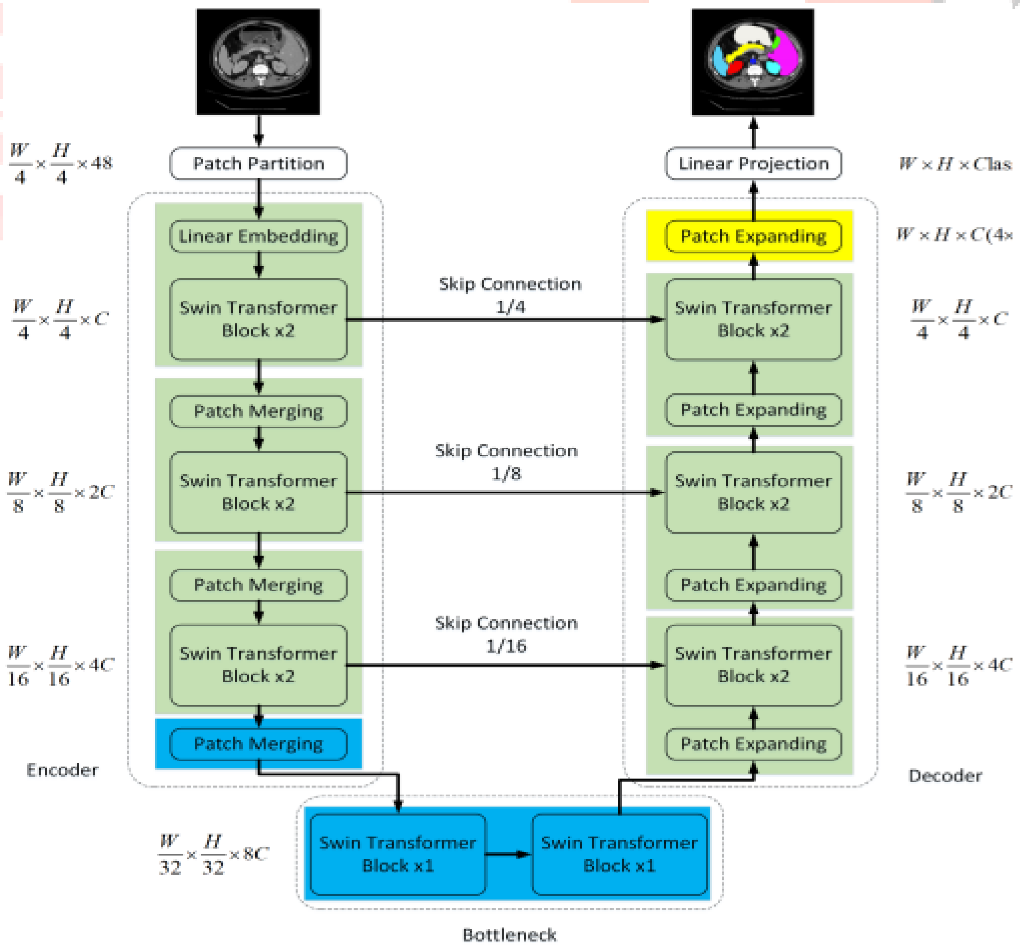

***Fig. 4**. Swin-Unet architecture where the contracting path and the expanding path are constructed on the swin transformer block [4].*

The medical imaginings are divided into non-overlapping patches with a patch size 4 x 4 to help encoder transform them into sequence embeddings. The feature dimension of every patch becomes 4 x 4 x 3 = 48 using such a partition approach[4]. Then, a linear embedding layer is added to the determined feature dimension to an arbitrary dimension. Many Swin Transformer blocks with patch merging layers are passed through transformed patch tokens to get the tiered feature representations. Especially, downsampling and incrementing dimension are managed by the patch merging layer, and feature representation learning is managed by the Swin Transformer block.

The decoder comprises a Swin Transformer block and patch expanding layer[4]. To act against the loss of spatial information with down-sampling the background features are added to multiscale features from the encoder through skip connections. Contrasting the patch merging layer, the patch expanding layer is formed to do up-sampling. Feature maps of adjacent dimensions are reshaped by the patch expanding layer into greater feature maps which have 2× up-sampling of resolution. At last, 4x up-sampling is done to regain the resolution of feature maps to input resolution by the last patch expanding, and then the pixel-level segmentation predictions are obtained by applying a linear projection layer on the up-sampled features[4]. It has excellent performance and generalization ability

2.5 RIC-Unet

RIC-Unet is an Unet and inception net-based architecture called "Residual-inception-channel attention-Unet" [5]. It is developed as an architecture that is suitable for pathological or histological images which are used to determine the morphological characteristics and population of cells and their contents.

The typical architecture of RIC-Unet is composed of three sections: The left part which is identical to the base Unet architecture and contains a ReLu layer used to perform activation function, four blocks of decoding sub-networks that perform upsampling, and a final convolutional layer that produces the final output. On the right is another ReLu layer followed by four DC blocks and another convolutional layer that gives the output. The Third section is the middle section which contains four RI blocks and a convolutional layer that connects the other two sections through DRI modules. The middle section, which contains RI blocks takes the input images while the other two sections, containing decoder blocks and DC blocks produce two final outputs. This way the outputs obtained can be compared easily.

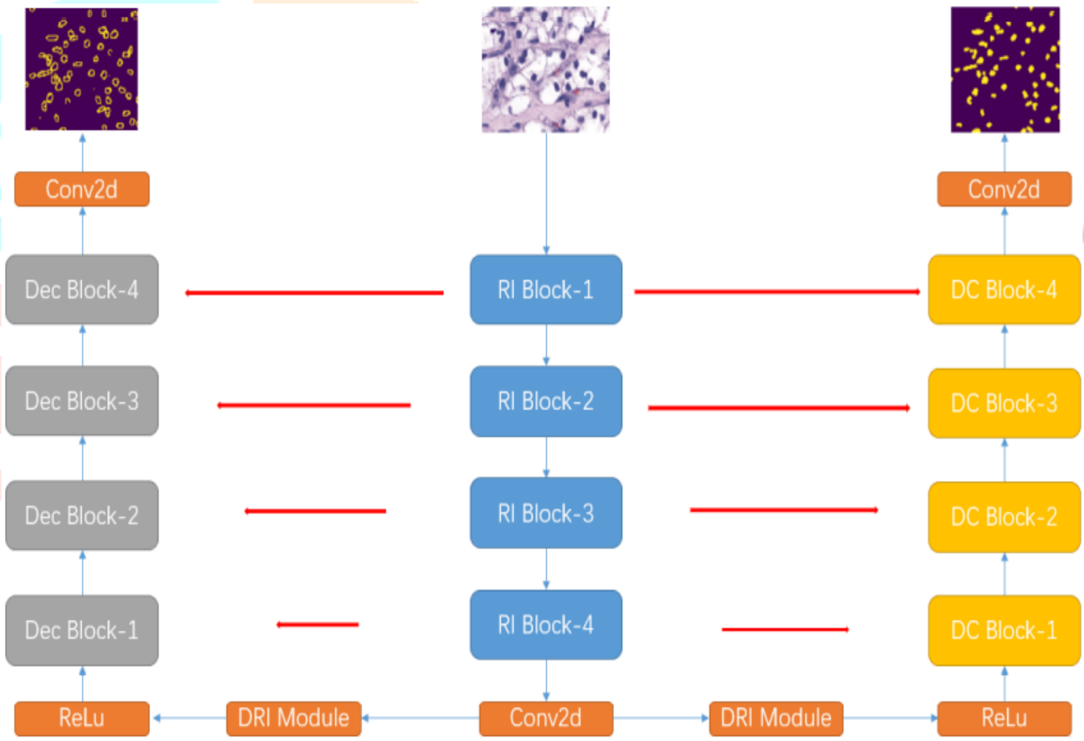

Fig. 5. *RIC-Unet architecture with three sections: the first section with decoder blocks, identical to base Unet outputs nuclei's contour mask, the second section with RI blocks, takes input images, and the third section with DC blocks, outputs original nuclei's mask.*[5]

Here, cell contours are used while distinguishing the dense cells besides decreasing faults in the object level [5]. The RI block and DC block make this RIC-Unet architecture different from the base Unet architecture. As the RI blocks take input images, These RI blocks form the encoding phase where multi-scale details are retrieved [5]. The residual module (RES module in ***Fig. 6****.)* and the inception module (DIR module in ***Fig. 6***.) from the RI block [5]. The input image is first processed by the convolutional layer with two strides for the feature map and a kernel of size 3x3. Then it is processed by the DRI module, which is an intricated inception module. This DRI module retrieves critical details with different kernel sizes inside receptive fields [5]. The DC blocks and the decoding blocks are responsible for producing output. While decoding blocks work just like base Unet, the DC blocks make use of the channel-attention-block module (CAB module in ***Fig. 6***.) in the upsampling block for picking the best features from the down-sampling part and make efficient use of these features [5]. Following the CAB module is the deconvolution layer, which aggregates features with varied resolutions [5].

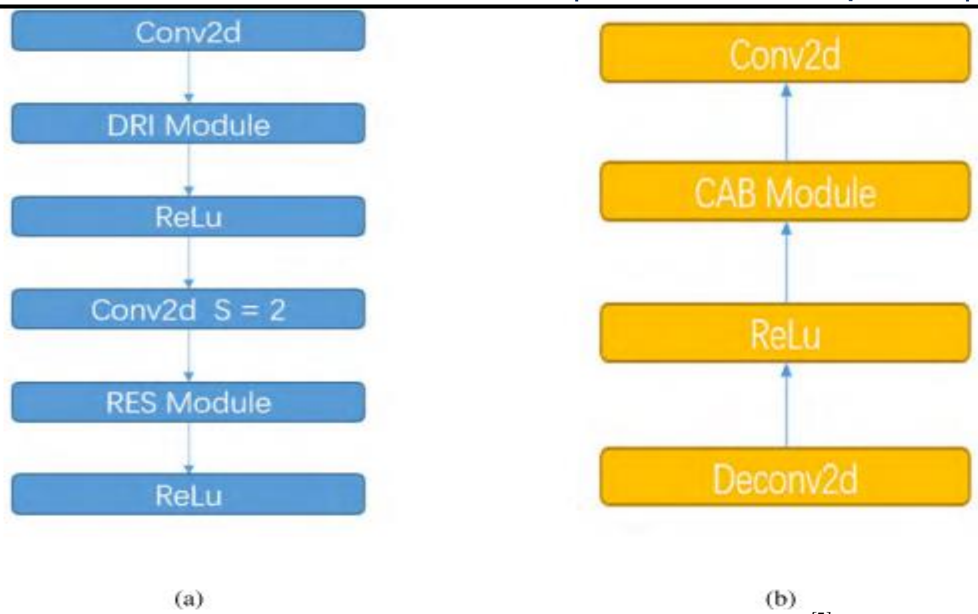

Fig. 6. *a) RI block and b) DC block of RIC-Unet architecture* [5].

2.6 Residual Unet

Residual Unet was introduced to overturn the difficulty of pixel imbalance and to increase the generalization capacity of the model. The main modifications compared to unet in this model include: First, prior to each down-sampling, the number of channels is switched to double compared to the number of the upper layer; simultaneously, before each up-sampling, the number of channels of convolution kernel is halved, hence successfully decreasing the bottleneck of the model[6]. Secondly, utilizing convolution with a step size of 2 as alternative of pooling can alleviate the information loss caused by directly discarding features and it can decrease memory consumption. Thirdly, in case of difficulty with slow model convergence, it incorporates remaining structure to advance the model, to accelerate the convergence[6]. Simultaneously, the skip connections in the network can improve the network performance by elevating information propagation and decreasing the number of parameters.

The given figure shows the residual unet architecture. The red cube represents the feature map. The blue arrow shows the batch normalization, ReLU activation function, with convolution on the input data[6]. The purple arrow shows the convolution process. The dashed blue arrow shows the forward propagation of the flow of information in the network without any operation[6].

The red arrow shows convolution with a kernel size of 2 x 2 and with step size of 2. The count of convolution kernels is double the count of channels of the input feature map, and its main reason is to down-sample the feature map[6].

The green arrow represents transposed-convolution. The number of convolution kernels is halved compared to the count of channels of the input feature map, and its main purpose is to up-sample the feature map.

The adding layer adds the correlating positions of two inputs and the concatenation layer adds the two inputs in the channel dimension.

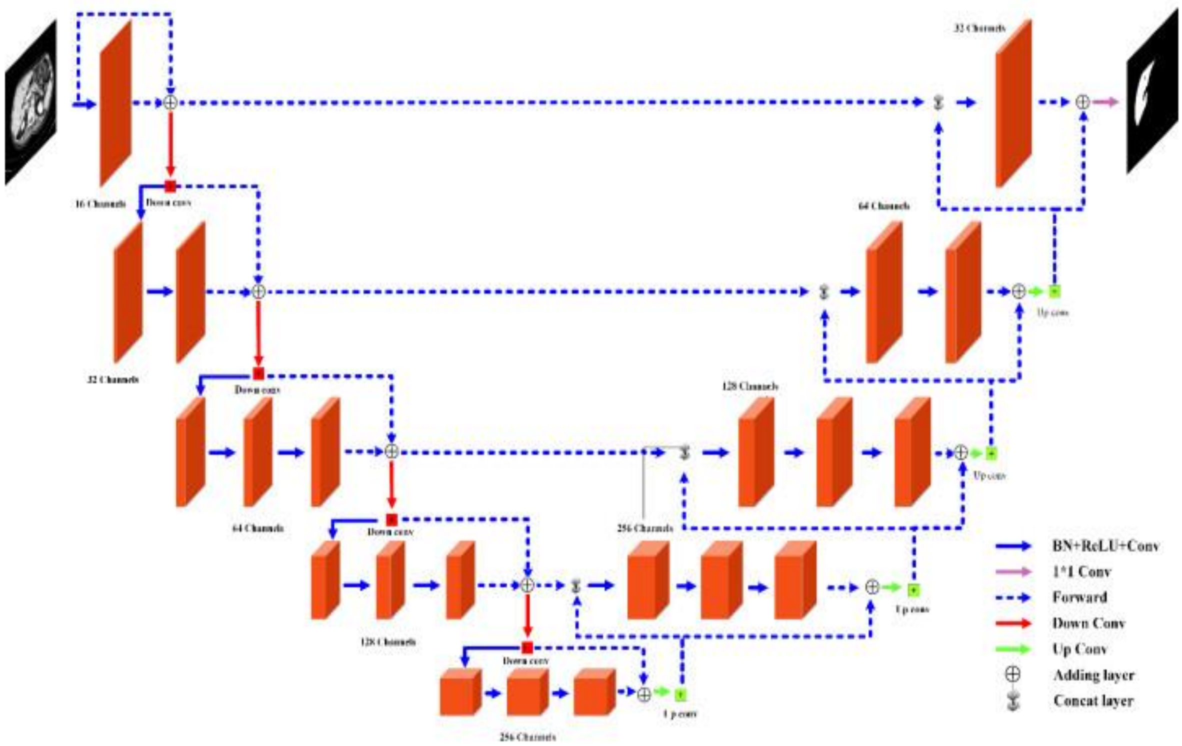

Fig. 7. *Residual Unet*

2.7 DU Unet

Dimension-fusion-UNet also called DU Unet is an improved version of Unet. It adds 2D and 3D convolution creatively in the encoding stage. The suggested architecture generates improved segmentation than 2D networks, and it needs comparatively less computation time than 3D networks.

It is similar to the unet in the case of a symmetrical structure. It had two parts encoder and a decoder which combine high-level features and low-level fine-grained features[7]. In DU Unet, the encoder has two dimensions where both 2D and 3D convolutions carry out down-sampling in their appropriate dimensions, then the results are combined. This model helps in identifying small lesion areas and refines edges as the 2D networks are integrated with 3D information. The trainable parameters are extremely increased as networks deepen and the dimension process is only at the beginning of the model. The dimensionality part contains 3D dimensionality reduction, channel excitation with dimensional fusion[7]. The Squeeze and excite part starts the connection amongst different channels by weighing the feature channels. It is used to improve the combined effect of 2D and 3D features.

In every dimension transform part, we decrease the dimensions of the 3D branch feature map in beginning and later combine them with the 2D branch following the SE weighted respectively[7]. This model improves performance compared to the 2D model and effectively utilizes 3D information.

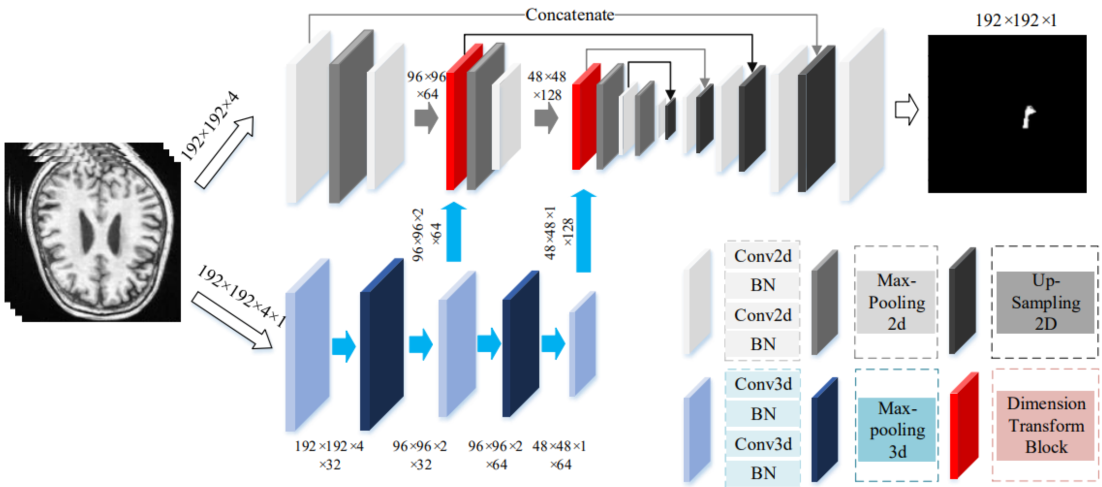

***Fig. 8**. DU Unet*

2.8 Fully Dense Unet

F D Unet encompasses dense connectivity within the contracting and expanding paths of Unet[8]. In the contracting path multilevel decomposition is done in the image. In the contracting path spatial dimensions of the feature maps are lessened by max pooling[8]. The local features together with global features are extracted which are used for artifact removal in different spatial scales. The image which is unsampled using the deconvolutional operator is combined with the original image to produce image Y. Deconvolution is basically transposed convolution. After deconvolution, the feature maps coming from the encoding path are combined with those of the decoding path. This helps as higher-resolution images from the encoding path are used.

The F D Unet uses dense blocks instead of 3x3 convolution operators[8]. The convolutional layers are attached to all subsequent layers along with channel-wise concatenation in a dense block which has unneeded features and also supports learning various set of features[8]. Heavy connections lead to an increase in depth which can cause gradient information to be lost and creates a vanishing gradient problem. They can be addressed by introducing many connections to let gradient information that may be backpropagated. It removes the vanishing gradient problem.

The results from FD Unet show that it gives a higher quality of images than Unet and works better even with a corrupted initial image.

The figure shows the expanding and contracting paths of FD Unet.

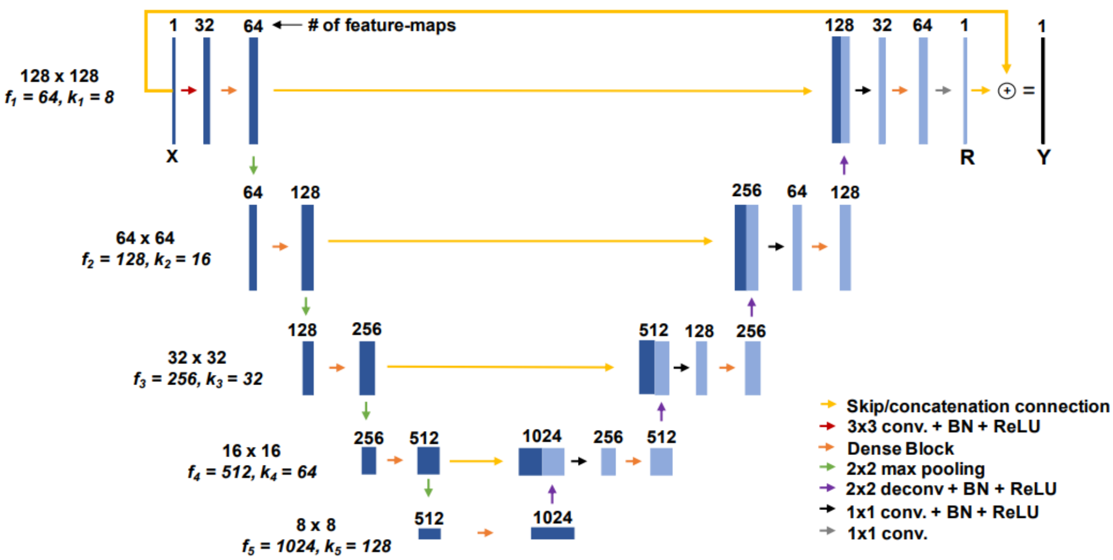

***Fig. 9**. Fully Dense Unet*

2.9 BN-U-Net

BN-U-Net is an improved model of Unet. Usage of this model decreases the processing time and improves the quality of the image. This model involves the addition of a specification layer (BN layer) in amongst each inner layer and activation layer and also amongst different convolution layers and activation layers to improve the Unet model.

In this model, data is normalized in batch at some convolution layer[9]. R^a_{xy} is the outcome of layer y which has X neurons and is trained for a^{th} data[9]. F_{xy} is average outcome of neuron x in layer y[9]. K_{xy} is the standard deviation of neuron x in layer y. The result P as soon as batch normalization applied is

$$P = \frac{R^a_{xy} - F_{xy}}{K_{xy}}. \quad (1)$$

The mean value of neuron outcome is

$$F_{xy} = \frac{1}{t}\sum_{a=1}^{t} R^a_{xy}. \quad (2)$$

Standard deviation is

$$K_{xy} = \sqrt{b + \frac{1}{t}\sum_{a=1}^{t}\left(R^a_{xy} - F_{xy}\right)^2}. \quad (3)$$

The following is the output obtained by the specification layer activation function. In this, b and t are constants and input n augments the weight through weight L[9].

$$U = f\left(BN\left(\sum_{a=1}^{t} L_x N_x\right)\right). \quad (4)$$

This model has significant improvement in accuracy, sensitivity, and specificity compared to FCNN and Unet models. The segmentation time is greatly shortened.

2.10 R2U-Net

R2U-Net is a recurrent residual U-shaped network architecture. R2U-net was built based on RRCNN, that is, the Recurrent residual convolutional neural network and basic U-net. There are several features of R2U-Net that differentiate it from the base Unet architecture. R2U-net makes use of both recurrent and residual neural networks. RCNN was successful in object recognition tasks making efficient use of various benchmarks [10]. This RCNN is modified making use of the “recurrent Convolutional layers” (RCL) [10] forming an RRCNN architecture. RCL deploys discrete time stamps based on which the operations in convolutional layers within RCNN are performed. This can be mathematically represented as follow:

$$O^l_{ijk}(t) = \left(w^f_k\right)^T * x_l^{f(i,j)}(t) + (w^r_k)^T * x_l^{r(i,j)}(t-1) + b_k$$

Where,
$O^l_{ijk}(t)$ - output of the network with time stamp t,
x_l, – input at the l^{th} layer of the RRCNN w,
(i, j) – pixel position,
w_k – weights of the k^{th} feature map, and
b_k – bias [10].

The output of the network or RCL is given as an input to the ReLU function and the output from the ReLU function is used for downsampling or upsampling operations. Here, in R2U-net, the final results from the RCNN are passed to the residual unit as shown in ***Fig. 11**. d).* The typical architecture of R2U-net is as represented in ***Fig. 12***. It has a similar U-net-like architecture the contracting and expanding paths’ sub-networks have RCL built-in.

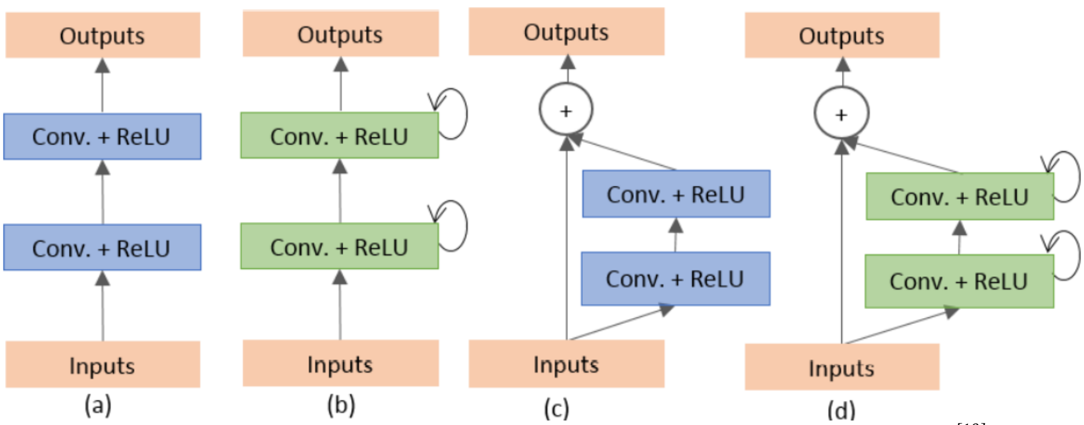

***Fig. 10**. a) FCNN units, b) RCNN units, c) Residual CNN units, d) RRCNN units [10].*

R2U-net is more efficient than U-net as it replaces regular connections with RCL, which performs in-detail feature extraction and also makes efficient use of time steps to retrieve better and strong features[10]. R2U-net also makes use of simple concatenation of a complete set of feature maps rather than cropping and copying them, this way the exactness of the outputs obtained Is increased.

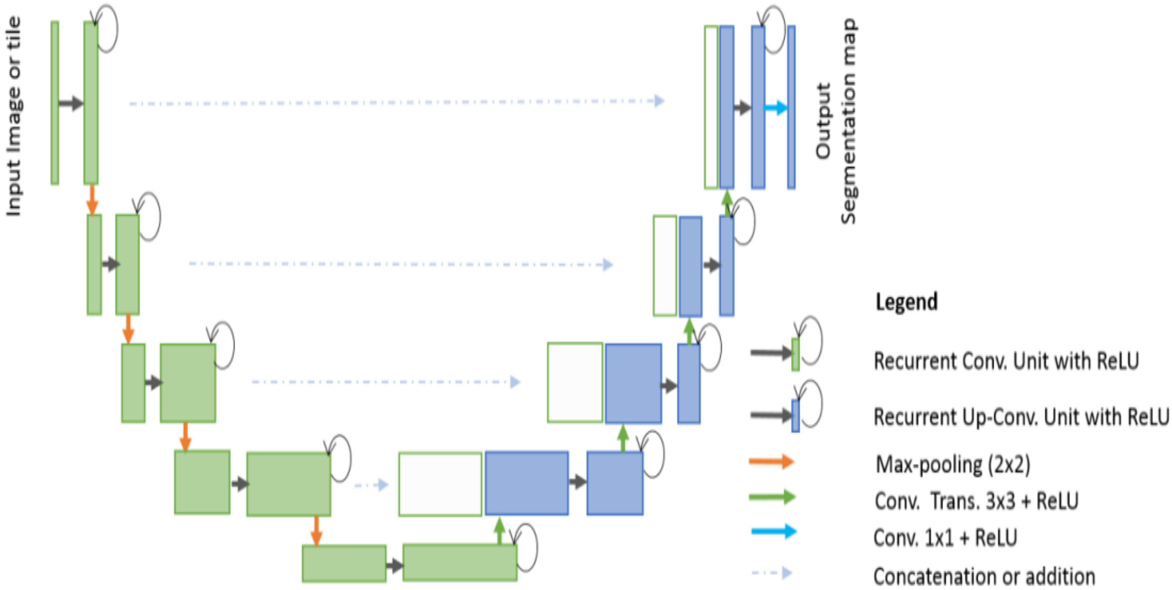

Fig. 11. *R2U-Net architecture with RCL[10].*

R2U-Net was developed for better performance and accuracy of the outputs obtained.

III. Conclusion

Medical images are obtained using specialized devices for instance, computed tomography (CT scanners), magnetic resonance imaging (MRI scanners), ultrasound scanners, etc. Medical image preprocessing is necessary to identify the condition or morphology of an organ or a cell. It makes the diagnosis process easier and most of the time can even produce accurate predictions of the patient's condition. With medical image processing gaining more importance, the techniques or the processes involved are often tedious due to the quality of the medical images. Due to these issues surrounding the concept of medical image processing, there are many techniques developed to work on the same. Specifically deep learning techniques such as CNN, FCNN, RCNN, etc., have been successful in this process involving classification, segmentation, pre-processing, etc. And one such technique is U-Net. U-Net is an FCNN-based architecture that is extensively applied in medical image segmentation. Due to its veracity in producing proper segmentation outputs, U-Net architecture was often experimented on, and as a result, there are many versions of this base U-Net architecture that is unique in their ways. This paper highlights the architectural differences in ten different U-Net-based architectures. The main idea was to produce a proper understanding of different versions of U-Net besides comparing them with the base architecture. This paper also portrays the supremacy of U-Net and its versions in producing accurate results by considering every critical detail of medical images.

References

[1] Ronneberger, Olaf & Fischer, Philipp & Brox, Thomas. (2015). UNet: Convolutional Networks for Biomedical Image Segmentation. LNCS. 9351. 234-241. 10.1007/978-3-319-24574-4_28.

[2] Zhou, Zongwei & Rahman Siddiquee, Md Mahfuzur & Tajbakhsh, Nima & Liang, Jianming. (2018). UNet++: A Nested UNet Architecture for Medical Image Segmentation: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings. 10.1007/978-3-030-00889-5_1.

[3] Huang, Huimin & Lin, Lanfen & Tong, Ruofeng & Hu, Hongjie & Qiaowei, Zhang & Iwamoto, Yutaro & Han, Xian-Hua & Chen, Yen-Wei & Wu, Jian. (2020). UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation.

[4] Cao, Hu & Wang, Yueyue & Chen, Joy & Jiang, Dongsheng & Zhang, Xiaopeng & Tian, Qi & Wang, Manning. (2021). Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation.

[5] ZITAO ZENG , WEIHAO XIE, YUNZHE ZHANG, AND YAO LU (2019).RIC-Unet: An Improved Neural Network Based on Unet for Nuclei Segmentation in Histology Images

[6] Pricing, Jinke Wang, Xiangyang Zhang, Chunlei Ji, Lubiao Zhou, Haiying Wang. An improved residual U-Net with morphological-based loss function for automatic liver segmentation in computed tomography

[7] Yongjin Zhou, Member, IEEE, Weijian Huang, Pei Dong, Yong Xia, and Shanshan Wang, Member, IEEE.D-UNet: a dimension-fusion U shape network for chronic stroke lesion segmentation

[8] Steven Guan, Amir A. Khan, Siddhartha Sikdar, and Parag V. Chitnis. Fully Dense UNet for 2D Sparse Photoacoustic Tomography Artifact Removal.

[9] Qingfeng Zhang, Yun Du, Zhiqiang Wei, Hengping Liu, Xiaoxia Yang, and Dongfang Zhao. Spine Medical Image Segmentation Based on Deep Learning

[10] Alom, Md. Zahangir & Hasan, Mahmudul & Yakopcic, Chris & Taha, Tarek & Asari, Vijayan. (2018). Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation.