**BIOMEDICAL IMAGE SEGMENTATION USING**

**U-NET++ ARCHITECTURE WITH VGG19 ENCODER**

***A Thesis/Project Submitted in partial fulfillment of the***

***Requirement for the award of the Degree***

***of***

**MASTER OF TECHNOLOGY**

**IN**

### COMPUTER SCIENCE AND ENGINEERING

### BY

### SHIV HARI TEWARI

### (MT/CS/45001/19)

### 

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

### BIRLA INSTITUTE OFTECHNOLOGY

### MESRA - 835215, RANCHI

**(Noida Campus)**

**DECLARATION CERTIFICATE**

This is to certify that the work presented in the thesis entitled **“Biomedical image segmentation using U-Net++ with VGG19 encoder”** in partial fulfillment of the requirement for the award of Degree of **Master of Technology in Computer Science and Engineering** of Birla Institute of Technology Mesra, Ranchi is an authentic work carried out under my supervision and guidance. To the best of my knowledge, the content of this thesis does not form a basis for the award of any previous Degree to anyone else.

Date: Guided by: Dr. Charu Wahi

Dept. of Computer Science Sign:

Birla Institute of Technology, Mesra,

Ranchi.

(Noida Campus)

**CERTIFICATE OF APPROVAL**

The foregoing thesis entitled **“Biomedical Image segmentation using U-Net++ architecture with VGG19 encoder”,** is hereby approved as a creditable study of research topic and has been presented in satisfactory manner to warrant its acceptance as prerequisite to the degree for which it has been submitted.

It is understood that by this approval, the undersigned do not necessarily endorse any conclusion drawn or opinion expressed therein, but approve the thesis for the purpose for which it is submitted.

Internal Examiner: External Examiner:

Director:

**Acknowledgement**

First and foremost we wish to thank my Supervisor Asst. Prof. Dr. Charu Wahi (Department of Computer Science and Engineering, Birla Institute of Technology, Mesra) for her kind blessings to me. She allowed me the freedom to explore, while at the same time provided with invaluable sight without which this Project would not have been possible. We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the Department for their kind assistance and cooperation during the development of our project.

SHIV HARI TEWARI

(MT/CS/45001/19)

**ABSTRACT**

**Biomedical Image segmentation using U-Net++ architecture with VGG19 encoder**

The image segmentation is widely used for the various computer vision tasks such as object localization and recognition, boundary detection, and medical imaging. This thesis proposes a deep learning architecture for the biomedical image segmentation task using Nested U-Net or U-Net ++ architecture added with the VGG19 encoder. The U-Net++ architecture is a densely connected network based on the famous U-Net architecture

.

The basic U-Net ++ architecture is series of dense skip connections which use the VGG16 encoder and backbone of densely connected U-Net architectures. We propose the novel architecture which is based on the Nested U-Net architecture (U-Net++) added with the VGG19 encoder.

This thesis includes the application of our model with its improved score, methodologies, the architecture and the detailed description of the work done including the snapshots of output, the IoU score and dice loss and detailed improvements than the U-Net++ with VGG16.

**Table of Contents**

1. Introduction
   1. Motivation
   2. Purpose
   3. Approach
2. Problem Statement
3. Literature Review

3.1. U-Net and U-Net based Models

3.2. U-Net++ architecture

1. Experimental Works

4.1. Implementations of image segmentation architectures

4.2. Selection of appropriate U-Net based architecture

4.3. Why U-Net++ or Nested U-Net

4.4. Selection of appropriate encoder

4.5. Why VGG19 encoder

4.6. Implementation of U-Net++ with VGG19

1. Results and Discussion

5.1. Experimental Results and comparison

5.2. Future Discussion

1. Conclusions
2. Future scope of work
3. References

**List of Figures**

Figure 1.1: The U-Net Architecture

Figure 1.2: The Nested U-Net (U-Net++) architecture

Figure 3.1: Basic U-Net Architecture

Figure 3.2: U-Net architecture for the 3D 128x128x128 voxels segmentation

Figure 3.3: Double U-Net Architecture

Figure 3.4: Nested U-Net architecture (U-Net++)

Figure 3.5: Bi-directional Convolution LSTM U-Net

Figure 3.6: SCAU-Net Architecture

Figure 3.7: Structure of U2-Net

Figure 3.8: The U-Net++ architecture

Figure 4.1: Fast Fully Connected Network

Figure 4.2: Gated SCNN architecture

Figure 4.3: Mask R-CNN

Figure 4.4: Segmentation output of Basic U-Net architecture

Figure 4.5: Nested U-Net architecture segmentation output

Figure 4.6: Segmentation output of SCAU-Net

Figure 4.7: The Residual block Network

Figure 4.8: The structure of DenseNet

Figure 4.9: The structure of Attention-Net

Figure 4.10: The structure of VGG encoder

Figure 4.11: Training set images with the masks

Figure 4.12: images after augmentation

Figure 4.13: Training of VGG19 Encoder

Figure 4.14: Training of VGG19 Encoder

Figure 4.15: IoU and Dice loss plot

Figure 4.16: Predictions on validation images set

Figure 5.1: Segmentation results of U-Net++ architecture with VGG16

Figure 5.2: Segmentation results of U-Net++ architecture with VGG19

**CHAPTER 1**

**INTRODUCTION**

The image segmentation is a prominent concept of computer vision. It is the process of partitioning the pixels of an image into “segments” associated with different classes. The main goal of segmentation is often to simplify the representation of an image such that it is easier to interpret, analyze, and understand. Image segmentation has been used in a variety of computer vision tasks, such as object localization, boundary detection, medical imaging, and recognition. In essence, these tasks are performed by assigning each pixel in an image to a certain label based on similar attributes, such as texture, color, intensity, or distance metrics. The result of image segmentation is a set of segments that collectively cover the entirety of an image.

The thesis focuses on the architectures based on U-Net for the image segmentation process. The U-Net architecture is proposed in 2015 by Olaf Ronneberger *et. al*. The basic U-Net is network based on encoder- decoder network. It has two paths, the path which does the encoding is known expansion path and other path is called the contraction path. The basic U-Net architecture is the network of plain skip connections. The layers are connected directly to each others in this network. The skip connections are extra connections between nodes in different layers of neural networks that skip one more layers of non-linear processing.

The figure 1.1 shows the network of the basic U-Net architecture.

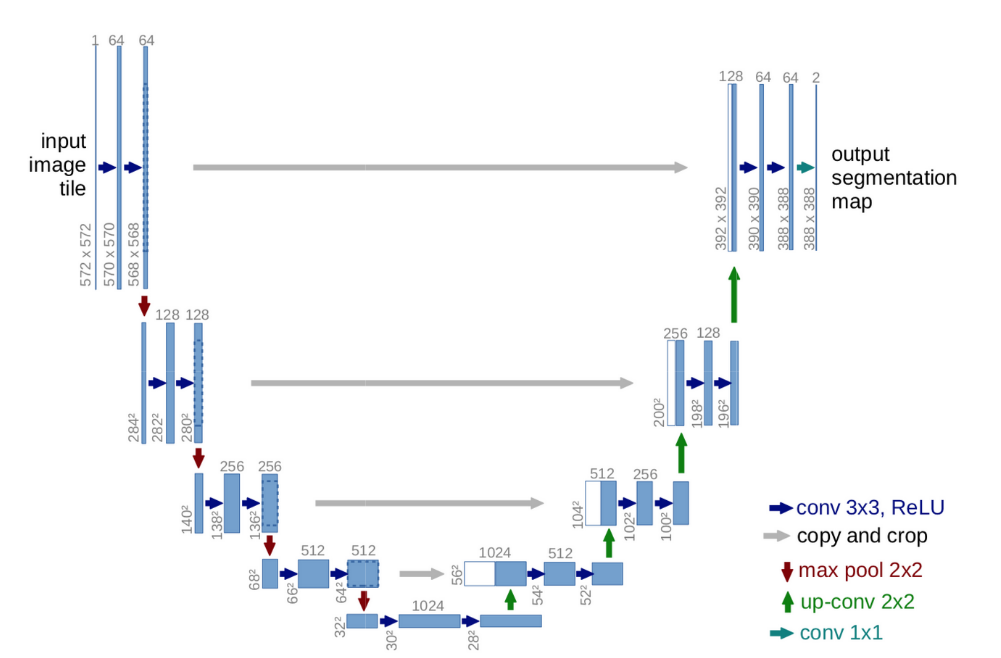


Figure 1.1: The U-Net Architecture

In our project we have worked on the architecture which is a network of densely connected U-Net architecture also known as U-Net++ or Nested U-Net architecture. The nested U-Net architecture is densely connected convolutional U-Net network. It is a network of dense skip connections, the dense skip connection or fully connected layered connection, are a type of layer in a deep neural network that use a linear operation where every input is connected to every output by a weight.

Figure 1.2 shows the architecture of the Nested U-Net (U-Net++).

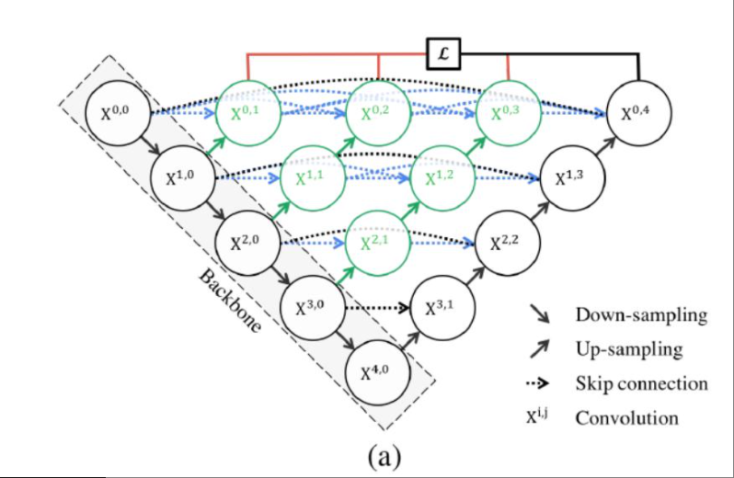


Figure 1.2: The Nested U-Net (U-Net++) architecture

The basic U-Net++ architecture used to run with densely connected U-Net as backbone and VGG16 encoder. In our project we have worked on the U-Net++ architecture by combining it with the latest VGG19 encoder. There are several types of encoders used in the U-Net based architectures. The basic U-Net architecture uses the VGG11 encoder, in the other U-Net based architectures like V-Net, U-Net ++ use VGG16 encoder. The latest proposed U-Net based architectures which are Double U-Net and U-Net 3+ are using VGG19 encoders.

In our project we have added the latest VGG19 encoder with the U-Net++ network and calculated the loss functions using the same dataset which is used to train the original U-Net++ with the VGG16 encoder, these datasets are EM-segmentation challenge datasets, these datasets contains the images of lungs and cell nuclei.

* 1. **Motivation**

Many researchers have proposed the different types of image segmentation architectures that are based on U-Net. The architectures based on the U-Net are U-Net++, V-Net, double U-Net etc.. These architectures are mostly use the pre-trained encoders like VGG16, VGG11 which are trained by imagenet, some other architectures based on U-Net use encoders like attention-net for example Spatial Attention U-Net and Spatial Channel Attention U-Net (SCAU-Net) these both architectures use attention-net encoders. Thus, our topic of the project is to design the model which is based on the densely connected U-Net with the latest VGG19 encoder.

As per the current situations in the medical field we need some robust image segmentation architecture which can which can produce the output efficiently and can clearly indicate the abnormalities in the images taken as input. To make our model more efficient we added our model with latest and famous VGG19 encoder.

**1.2 Purpose**

The main purpose of this project is twofold:

* Build an efficient image segmentation architecture using U-Net++ architecture.
* Learning about the different types of encoders which are used to train the architecture.
* Combining the densely connected nested U-Net with the latest VGG19 encoder.

**1.3 Approach**

In preparation of designing our model, we decided to analyze the different approaches in the literature with the intention of learning and understanding the advantages and shortcomings of each approach. After our analysis we decided to follow that approach that combines both a scalable model and latest encoding techniques. Such an approach will enable to overcome the limits of U-Net with obsolete encoder and use the advantages of both. Nested U-Net architectures are quite successful, but under certain domain and data characteristics different encoders may achieve unlike results. For our intended model we have decided to try out a way of introducing different encoders into our model by combining them with U-Net++ architecture. To be more specific encoding features are utilized to train the network and process its input image. In order to find the best way to integrate our intended with the encoder, we studied various other models and encoding strategies that exist for the U-Net++ architecture. After our analysis we decided to go for VGG19 encoder for our model. This encoder is pre-trained network which is trained by the imagenet algorithm and adding this encoder made the improvement in the results and gave better loss function calculations.

In the following sections of this document, we are going to discuss more in details each of our approaches. But before diving into the discussion of the approaches, we are going to define briefly what U-Net++ or Nested U-Net architecture is. We are also going to study briefly about the U-Net and U-Net based architectures in order to know what the different types of U-Net based architectures in the literature are and what are their strengths and weaknesses. Furthermore, we are going to study the U-Net++ in the context of the image segmentation and elicit the requirements which will enable us to make our decisions at design time. After discussing in details the problems we want to solve as well as the solutions we propose, we are going to draw conclusions regarding the advantages of our approach and also the difficulties met during the implementation. Finally we are going to coin in the enhancements that can be made on the implemented system in the future.

**CHAPTER 2**

**Problem Statement**

Image segmentation plays an important role in the medical imaging. In current situation an efficient architecture which can produce accurate output in less training time. In the medical imaging we need to identify the abnormality much clearly. So, our aim is to create a model which can perform image segmentation efficiently and takes less time to train. To make such model we studied U-Net based architectures and choose the U-Net++ for our model. Reason behind choosing U-Net++ architecture is that it has U-Net as a backbone network.

The U-Net architecture and U-Net based models have proposed by various researchers. These architectures are using different types of encoders to get trained. We can classify the U-Net based architectures on the basis of encoders. The basic U-Net, U-Net++ and double U-Net uses VGG encoders. The SCAU-Net and SAU-Net uses attention-net encoder. Other models like U2 – Net uses residual-net and U-Net3+ uses densenet encoder.

The encoders play a major role in training the model for image segmentation. The encoder takes an image as input and generates high dimensional feature vectors. It also aggregate features at multiple levels. In the image segmentation architectures the image is entered as input and the output is produced in the form of segmentation mask.

Our aim is to create a model which is based on the U-Net architecture and has the latest pre-trained encoder. The VGG19 encoder is the latest pre-trained encoder which is trained using imagenet algorithm. The VGG19 encoder is pre-trained so we don’t need to train it from scratch.

**CHAPTER 3**

**Literature Review**

**3.1 U-Net and U-Net based Models**

Several researchers have proposed the models based on the basic U-Net architecture. These models can be classified on the basis of various parameters such as skip connections, types of encoders etc.

In 2015, Olaf Ronneberger *et al*. developed the first U-Net architecture using CNN with plain skip connections for image segmentation. It connects the encoding and decoding layers without any interference. The training of encoding sequence is done by VGG-11 encoder. The 3X3 convolution layers are added in the segmentation tiles. For the decoding 2X2 up-sampling layers are added in the architecture.

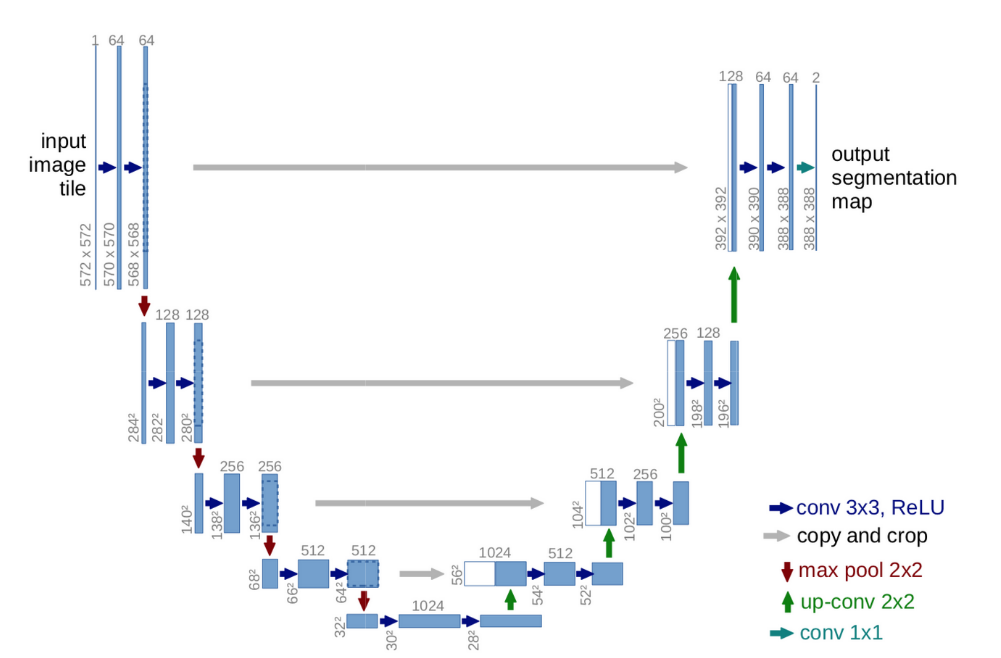


Figure 3.1: Basic U-Net Architecture

The size of U-Net must be comparable with the features included otherwise result could be ambiguous. The U-Net architecture takes significant amount of time to train. It requires relatively high GPU memory unit for larger images.

In 2017 Fabian Isensee *et al.* proposed a robust segmentation algorithm in the form of CNN inspired from U-Net architecture. This network is designed to process the large 3D input blocks of 128x128X128 voxels. The network comprises the context aggregation pathways. It encodes increasingly abstract representations of inputs as we go deeper in the network. The activations in the context pathway are computed by the context modules. The figure 3.2 shows the architecture which can be used for the segmentation of the larger 128X128X128 3D voxels.

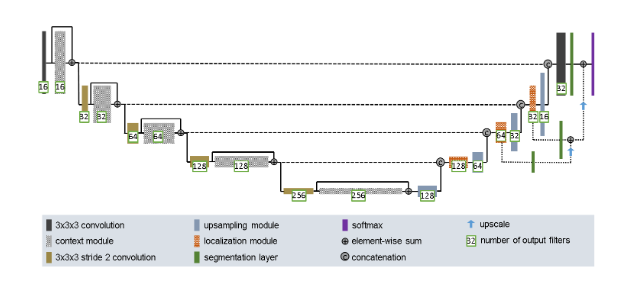


Figure 3.2: U-Net architecture for the 3D 128x128x128 voxels segmentation

In 2018, Debesh Jha *et al.* proposed the novel architecture named Double U-Net. It is a combination of two U-Net stacked on the top of each other. In the first U-Net network of the architecture the network is trained using VGG19 encoder. The decoder performs 2X2 bi-linear up-sampling on input features. It doubles the dimensions of input feature maps. The output is generated by concatenation process followed by two 3X3 convolution operations. Figure 3.3 shows the architecture of Double U-Net.

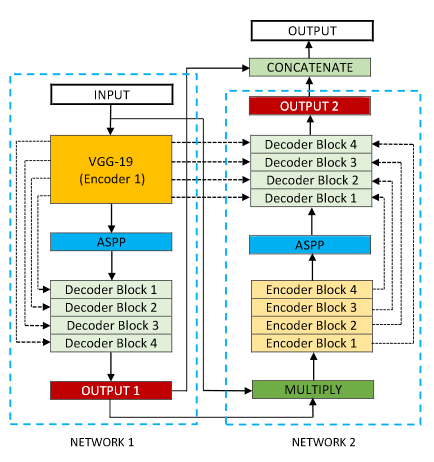


Figure 3.3: Double U-Net Architecture

It is possible to integrate more CNN blocks in Double U-Net. The network may require a powerful GPU to process the input. It uses more parameters than U-Net which leads to an increase in training time.

In the same year Zongwei Zhou *et al.* have proposed a nested U-Net architecture, termed as U-Net++. The U-Net++ has encoder and decoder connected through the series of nested dense convolution blocks.

U-Net++ can be deployed as the backbone of Basic U-Net architecture by replacing the plain skip connections with nested dense skip pathways. The U-Net++ is trained and evaluated on the lung segmentation dataset. Figure 3.4 shows the detailed architecture of U-Net++.

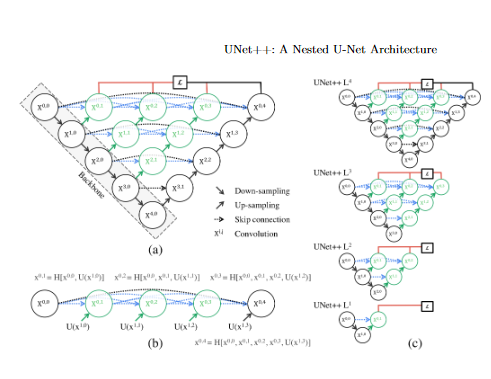


Figure 3.4: Nested U-Net architecture (U-Net++)

The dense skip connections in U-Net++ takes significant amount of time to get trained. It uses pre-trained VGG16 encoder with the nested dense skip connections. It also requires an efficient GPU memory like basic U-Net and Double U-Net.

In 2019 Reza Azad *et al.* has proposed a Bi-directional Convolutional model termed as BCDU-Net. This model has used the application Bi-directional Convolutional LSTM combined with the basic U-Net. Figure 3.5 shows the architecture of BCDU-Net.

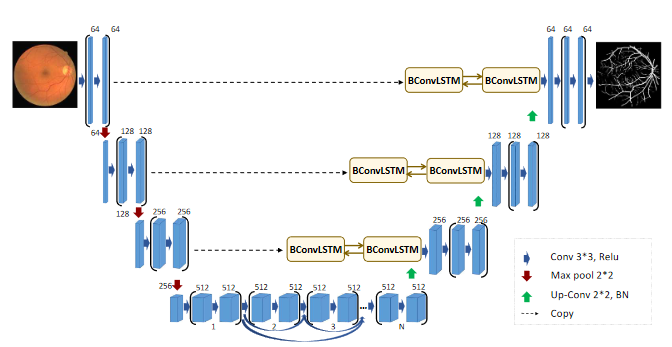


Figure 3.5: Bi-directional Convolution LSTM U-Net

The network is trained on the ISIC lung segmentation dataset, this model outperforms all the baseline models and its predecessors. It has used Keras and Tensorflow for implementation. Bi-directional connected convolutional LSTM which makes it work more efficiently than the other baseline models and gives better accuracy. After each up-sampling it takes too much time in training and performing the batch normalization process.

In 2020, Peng *et al.* has proposed the novel architecture. The architecture has added the attention network encoder, thus it is named as Spatial Channel Attention U-Net (SCAU-Net). The figure 3.6 shows the architecture of SCAU-Net.

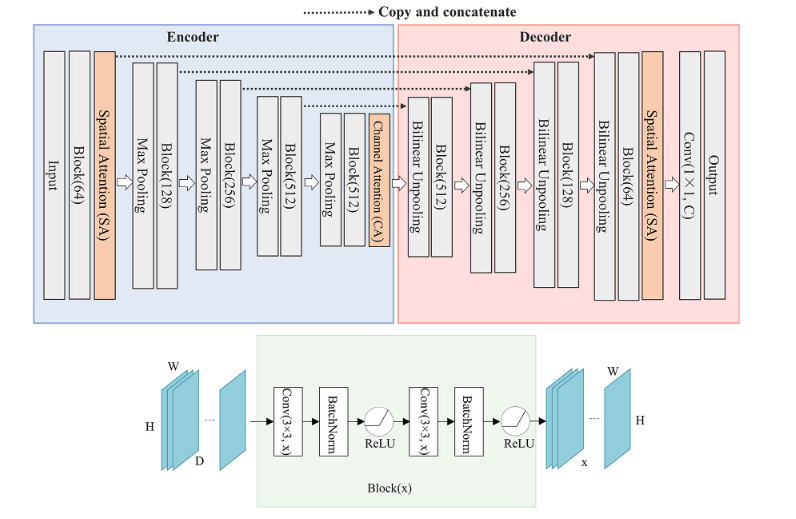


Figure 3.6: SCAU-Net Architecture

It can be considered as one of the most powerful image segmentation models, since it uses channel attention mechanism for segmentation. The use of spatial attention helps to retrieve detailed information easily.

It has multi-scale skip connections which tend to use unnecessary information and computational sources. Contextual information of low level encoder is insufficient and leads to the poor pixel wise recognition.

In 2020, Xuebin Qin *et al.* 2020 has proposed a model known U2-Net. This model is trained using residual-net encoder added with U-Net. They took advantage of residual U-blocks with U-Net named Residual U-Net or Res U-Net. It is nested level U-structured architecture. This architecture enables us to train a deep network from scratch without using backbones from image classification tasks. Figure 3.7 shows the structure of U2-Net.

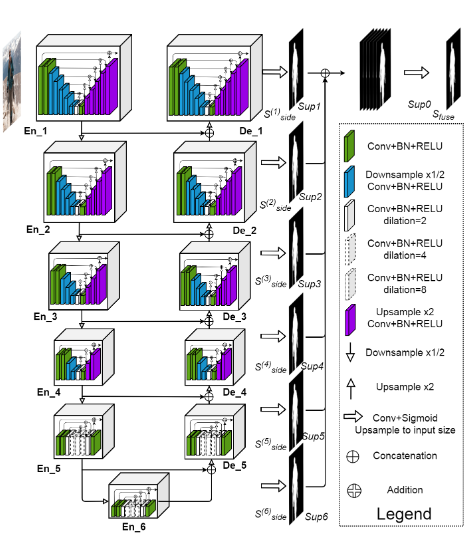


Figure 3.7: Structure of U2-Net

**3.2 U-Net++ architecture**

Proposed by Zongwei Zhou *et al.* in 2018 the U-Net++ is a new segmentation architecture based on nested and dense skip connections. This designed skip connection reduces the gap between the feature maps of the encoding and decoding sub-networks. This segmentation architecture is based on this assumption: **Before fusing the high-resolution feature map from the encoder network with the corresponding semantically rich feature map from the decoder network, the model is gradually enriched, and the model can effectively capture the fine-grained details of the foreground object.** When the feature maps from the decoder and encoder networks are semantically similar, the network will handle easier learning tasks.

**Difference from U-Net architecture:**

The main idea behind Nested U-Net or U-Net++ was to bridge the semantic gap between encoding and decoding feature maps before fusion.

* There is a convolutional layer on the jumping path to bridge the semantic gap between encoding and decoding feature maps
* There are tight skip connections on the jump path, which improves the gradient flow
* In-depth supervision is added, model can be pruned and improved, or in the worst case, performance equivalent to using only one loss layer can be achieved.

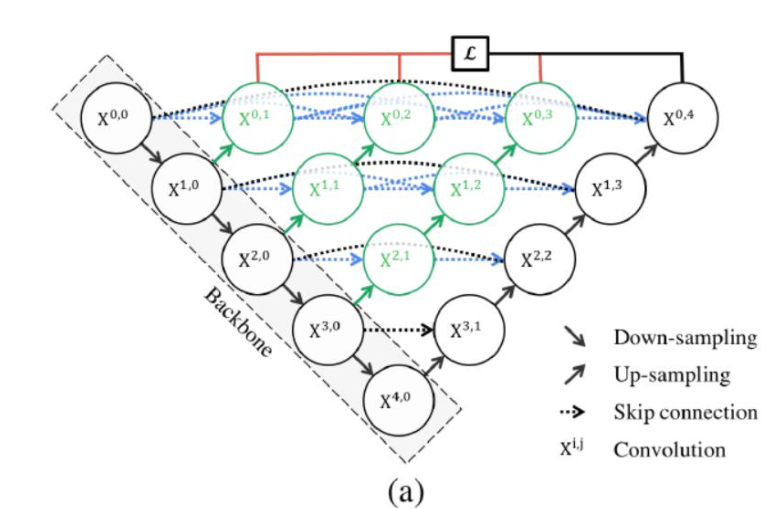


Figure 3.8: The U-Net++ architecture

**In-Depth Supervision:**

Deep supervision can make the model run in two modes:

1. Accurate mode, in which the output of all subdivision branches is averaged.
2. Fast mode, in which only the final segmentation map is selected from one of the segmentation branches, and its selection determines the degree of model pruning and speed gain.

**CHAPTER 4**

**Experimental Works**

**4.1 Implementations of Image segmentation architectures**

There are various architectures for the image segmentation process. In the image segmentation architectures the encoder extracts features from the image through filters. The decoder is responsible for generating the final output which is usually a segmentation mask containing the outline of the object. Some of the image segmentation architectures that we have studied are:

### FastFCN —Fast Fully-connected network

In this architecture, a Joint Pyramid Upsampling(JPU) module is used to replace [dilated convolutions since they consume a lot of memory and time](https://arxiv.org/pdf/1808.08931.pdf). It uses a fully-connected network at its core while applying JPU for upsampling. JPU upsamples the low-resolution feature maps to high-resolution feature maps. Figure 4.1 shows the Fast FCN architecture.

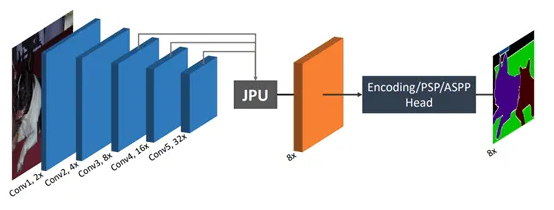
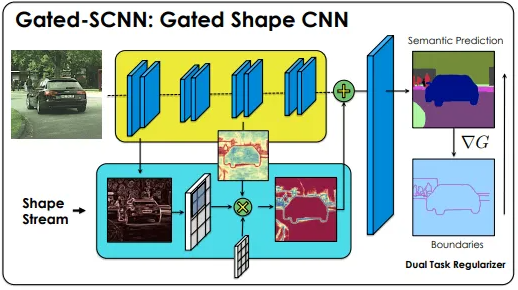


Figure 4.1: Fast Fully Connected Network

1. **Gated shape CNN—Gated SCNN**

This architecture consists of a two-stream CNN architecture. In this model, a separate branch is used to process image shape information. The shape stream is used to process boundary information. Figure 4.2 shows the Gated SCNN architecture.

****  
Figure 4.2: Gated SCNN architecture

1. **Mask R-CNN**

In this [architecture](https://github.com/facebookresearch/Detectron), objects are classified and localized using a bounding box and semantic segmentation that classifies each pixel into a set of categories. Every region of interest gets a segmentation mask. A class label and a bounding box are produced as the final output. The architecture is an extension of the Faster R-CNN. The Faster R-CNN is made up of a deep convolutional network that proposes the regions and a detector that utilizes the regions. Figure 4.3 shows the Mask R-CNN architecture.

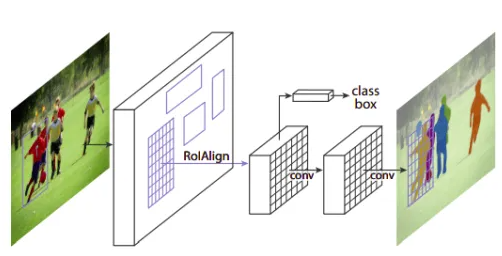


Figure 4.3: Mask R-CNN

1. **U-Net Architecture**

U-Net is a convolutional neural network originally developed for segmenting biomedical images. When visualized its architecture looks like the letter U and hence the name U-Net. Its architecture is made up of two parts, the left part — the contraction path and the right part — the expansion path. The purpose of the contracting path is to capture context while the role of the expansion path is to aid in precise localization.

U-Net is made up of an expansion path on the right and a contracting path on the left. The contracting path is made up of two three-by-three convolutions. The convolutions are followed by a rectified linear unit and a two-by-two max-pooling computation for downsampling.

**4.2 Selection of appropriate U-Net Based architectures**

For the implementation of our project we selected the U-Net as the basic architecture. To selected the suitable image segmentation architecture we studied and implemented many architectures. These architectures were, The Basic U-Net, U-Net++, Double U-Net, 3D Conv U-Net, Ladder-Net (UUU-Net) architecture, SCAU-Net, Dense U-Net (U-Net with Dense Net) and BCDU-Net.

1. **Implementation of Basic U-Net:**

The Basic U-Net architecture proposed Olaf Ronneberger *et al.* was trained and implemented on the ISBI cell tracking dataset. So, we also took the same dataset to train the U-Net. Figure 4.4 shows the segmentation output of Basic U-Net.

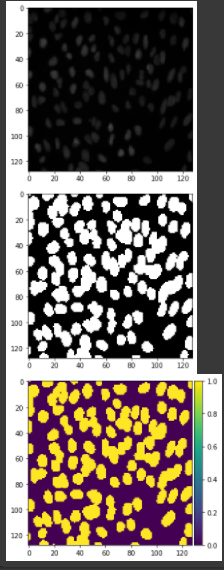


Figure 4.4: Segmentation output of Basic U-Net architecture

1. **Implementation of the U-Net++ architecture:**

At the initial level while selecting the suitable models we implemented the U-Net++ or Nested U-Net architecture on the Nuclei segmentation dataset. After training the U-Net++ architecture gave the results shown in figure 4.5.

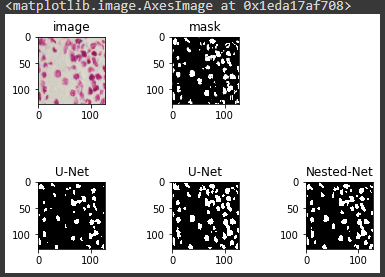


Figure 4.5: Nested U-Net architecture segmentation output

1. **Implementation of SCAU-Net:**

The SCAU-Net uses the attention-net encoder combined with U-Net architecture. It has two modes the spatial and channel attention. The spatial attention mode is added in the encoding network and both spatial and channel attention modes are added in decoder. Figure 4.6 shows the segmentation output of SCAU-Net.

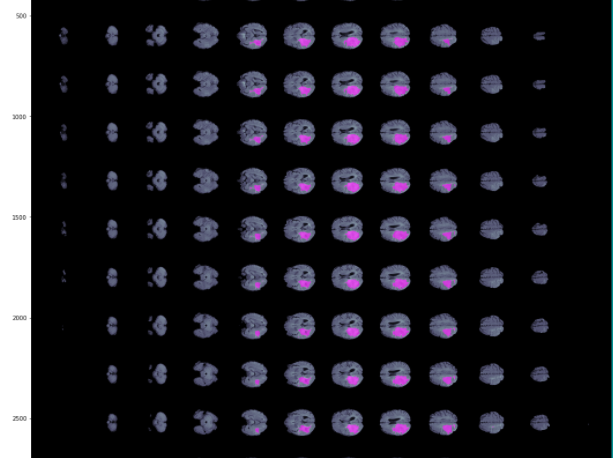


Figure 4.6: Segmentation output of SCAU-Net

**4.3 Why U-Net++ or Nested U-Net Architecture:**

Since there are many recently proposed architecture, one question may rise here that why U-Net++ architecture. There many qualities of U-Net++ which isn’t available in other novel architectures. The reasons for selecting U-Net++ for our working model are:

1. **It has convolution layers on skip pathways**, which bridges the semantic gap between encoder and decoder feature maps.
2. **It has dense skip connections on skip pathways**, which improves gradient flow.
3. **It has deep supervision, which enables model pruning**and improves or in the worst case achieves comparable performance to using only one loss layer.
4. The U-Net architecture is added as the backbone network of the whole model.
5. The densely connected convolution makes it more robust than the other baseline models and newly proposed models.
6. It has nested U-Net models connected through the series of U-Net architectures added with densenet.

Hence these are the reasons for selecting the U-Net++ architecture over the other U-Net based models.

**4.4 Selection of appropriate encoder network:**

In the semantic segmentation or in the image segmentation process the encoder plays major role in training the network. The U-Net and U-Net based architecture use various types of encoders to generate the feature vectors. To select the encoder for our model we have studied the architectures of encoding networks in the image segmentation. The encoders we have studied are, Residual-Net encoder, DenseNet, Attention-Net, SegNet and VGG encoders ( VGG11, VGG16 and VGG19). We have selected the VGG19 encoder for our modified U-Net++ model. These encoders are deep learning networks used to train the segmentation models. The various encoders we have studied are briefly described below.

1. **Residual-Net Encoder:**

ResNet, short for Residual Network is a specific type of neural network that was introduced in 2015 by Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun in their paper “Deep Residual Learning for Image Recognition”. When introduced it easily replaced the VGG16 encoder by observing relative improvements of 28%. The ResNet is made of Residual blocks. Figure 4.7 shows the architecture of residual block.

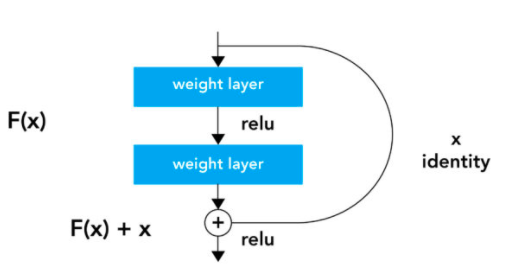


Figure 4.7: The Residual block Network

1. **DenseNet Encoder:**

DenseNet is quite similar to ResNet with some fundamental differences. ResNet uses an additive method (+) that merges the previous layer (identity) with the future layer, whereas DenseNet concatenates (.) the output of the previous layer with the future layer. The figure 4.8 shows the architecture of the DenseNet.

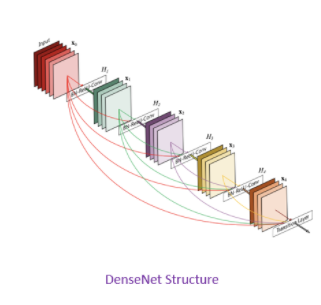
****

Figure 4.8: The structure of DenseNet

1. **Attention-Net Encoder:**

The attention mechanism emerged as an improvement over the encoder decoder-based neural machine translation system in [natural language processing (NLP)](https://courses.analyticsvidhya.com/courses/natural-language-processing-nlp?utm_source=blog&utm_medium=comprehensive-guide-attention-mechanism-deep-learning). Later, this mechanism, or its variants, was used in other applications, including [computer vision](https://courses.analyticsvidhya.com/courses/computer-vision-using-deep-learning-version2?utm_source=blog&utm_medium=comprehensive-guide-attention-mechanism-deep-learning), speech processing, etc. The figure 4.9 shows the architecture of Attention-Net mechanism.

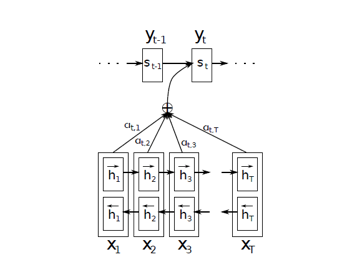


Figure 4.9: The structure of Attention-Net

1. **VGG Encoders:**

The VGG Network was introduced by the researchers at Visual Graphics Group at Oxford (hence the name VGG). This network is specially characterized by its pyramidal shape, where the bottom layers which are closer to the image are wide, whereas the top layers are deep. Figure 4.10 shows the architecture of VGG encoder network.

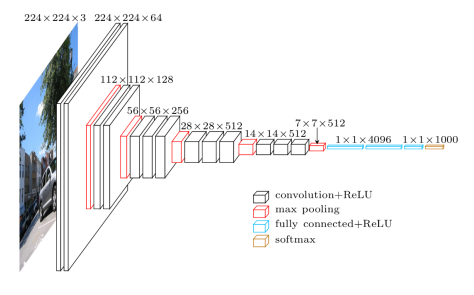


Figure 4.10: The structure of VGG encoder

VGG contains subsequent convolutional layers followed by pooling layers. The pooling layers are responsible for making the layers narrower. It is pre-trained network and that’s why we selected the VGG19 encoder for our modified U-Net++ model.

**4.5 Why VGG19 Encoder for the model:**

The VGG19 encoder is latest pre-trained network which is trained by ImageNet. Unlke other encoders the VGG19 encoder focuses on the window sizes and strides in the first convolutional layer. VGG addresses another very important aspect of CNNs which is depth. Since it works on the depth it is easy to work on deep convolution neural networks like U-Net++ architecture. The other advantages of VGG19 encoders are, VGG takes in a 224x224 pixel RGB image. For the ImageNet competition, the images cropped out the center 224x224 patches in each image to keep the input image size consistent. The convolutional layers in VGG use a very small receptive field (3x3, the smallest possible size that still captures left/right and up/down). There are also 1x1 convolution filters which act as a linear transformation of the input, which is followed by a ReLU unit. The convolution stride is fixed to 1 pixel so that the spatial resolution is preserved after convolution. VGG19 has three fully-connected layers: the first two have 4096 channels each and the third has 1000 channels, 1 for each class. VGG incorporates 1x1 convolutional layers to make the decision function more non-linear without changing the receptive fields. Hence these are the advantages because which we choose VGG19 encoder for our model.

**4.6. Implementation of U-Net++ architecture with VGG19:**

To perform the image segmentation process we took the same U-Net++ architecture which was proposed by Zongwei Zhou *et al.* They trained their model on the lung segmentation dataset given in EM segmentation challenge, so we took the same dataset to train the architecture with VGG19 encoder. To run the model we require the GPU(Graphical Processing Unit) so we choose the Kaggle notebook to run the model. The kaggle notebook provides virtual GPU and TPU (Tensor Processing Unit) for 38 hours. The U-Net++ architecture is evaluated with the help of Dice loss functions and IoU( intersection over union) score. We calculated the same loss functions for U-Net++ architecture added with the latest VGG19 encoder. We did the comparative studies to see the improvements in the model and the latest encoder improved the loss functions. The figure 4.11 shows the training set images with generated masks.

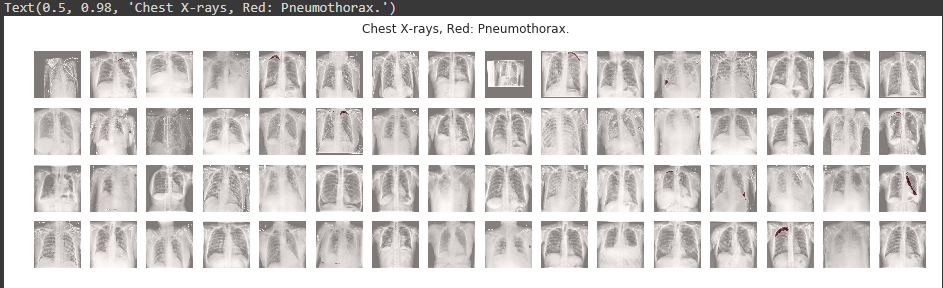


Figure 4.11: Training set images with the masks

After the training and mask generation the next step is the augmentation of the dataset. The augmentation process expands the dataset to visualize image clearly. Figure 4.12 shows the augmented images dataset.

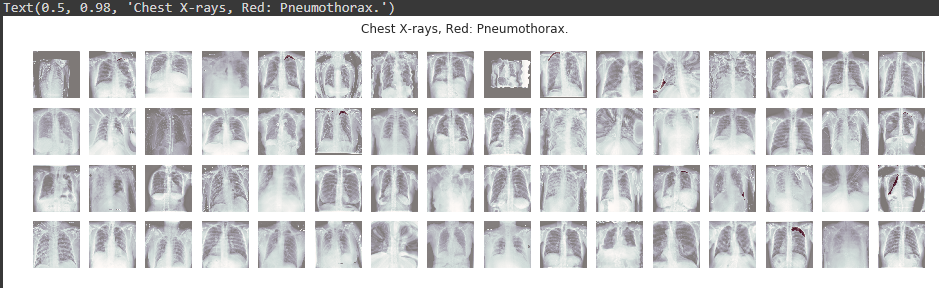


Figure 4.12: images after augmentation

In the above picture we can see the red areas which the infections in lungs which cause pneumonia. After the images augmentation we trained our encoder VGG19. The encoder takes input as an image and generates a high dimensional vector. The figure 4.13 shows the training of VGG19 encoder.

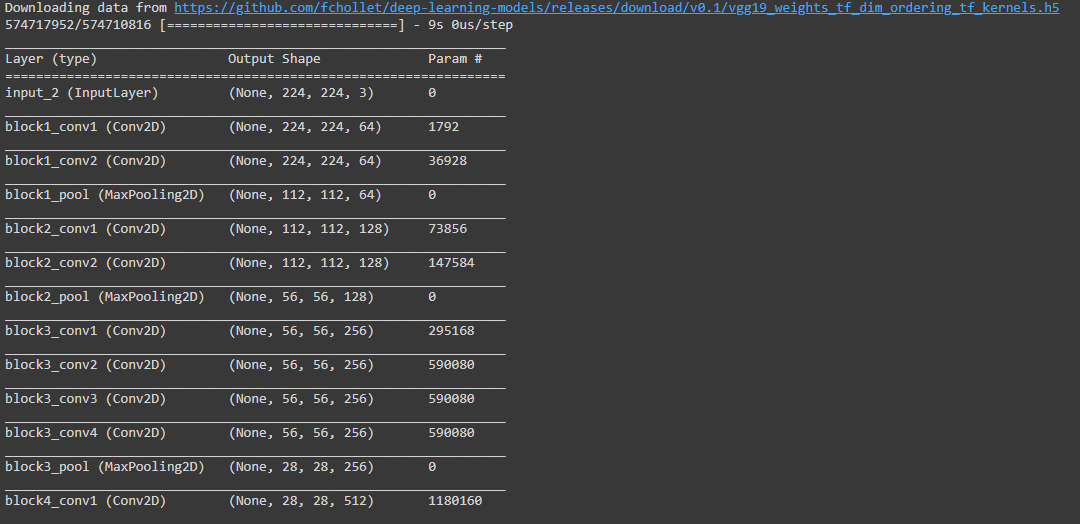


Figure 4.13: Training of VGG19 Encoder

After the training of VGG19 encoder we trained the loss function calculation i.e. IoU and Dice loss function calculations and trained the model to calculate the loss function. We plotted the IoU and Dice loss. Figure 4.14 shows the plots of IoU and Dice loss.

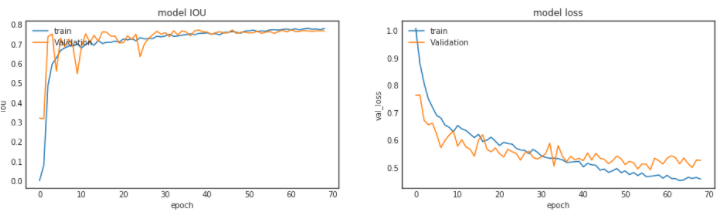


Figure 4.14: IoU and Dice loss plot

After the plotting and training of loss functions we trained our model to plot the predictions of images validations dataset. The predicted plot showed the areas of abnormality clearly in red color which signifies the infection while the green colored area shows the healthy lungs. Figure 4.15 shows the plotted predictions for validations set.

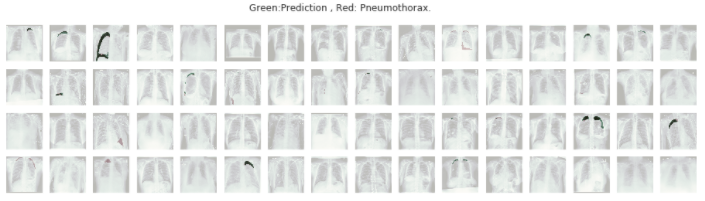


Figure 4.15: Predictions on validation images set

After the plotting of the validation set we calculated the threshold value and compared it with IoU score using plot. The figure 4.16 shows the comparison between threshold values and IoU score.

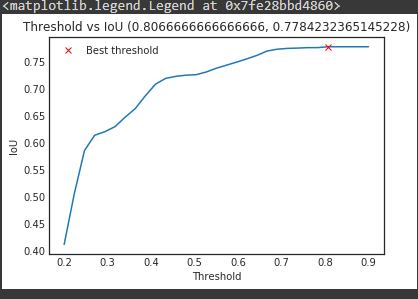


Figure 4.16: Threshold Vs IoU

After the completion of the loss function training and plotting, we finally plot the prediction of our prepared test set. The figure 4.17 shows the prediction of test set.

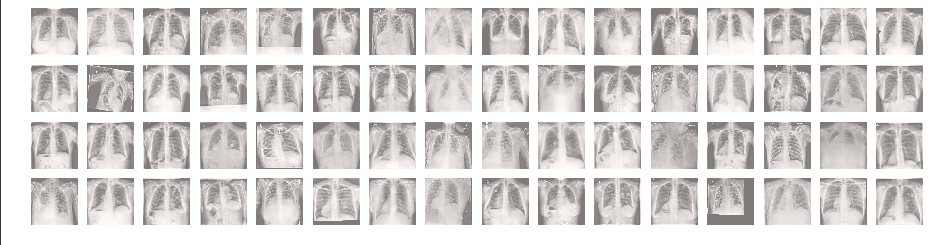


Figure 4.17: The test set prediction

In the above figure the green color shows the prediction in the images test set and the gray color shows the healthy lungs images. Thus this is the process of producing the segmentation output from given images dataset.

**CHAPTER 5**

**Results and Discussions**

* 1. **Results and comparisons**

To calculate the result of our model we studied various loss functions. We implemented some of the loss functions which are easy to calculate. The U-Net++ proposed by Zongwei Zhou *et al.* in 2018 they have calculated the dice loss and IoU score of their model. Their model was U-Net++ architecture and they have used the VGG16 encoder added with DenseNet. In our modified model we added U-Net++ architecture with the VGG19 encoder and used efficient-net with the DenseNet. In our model we have also calculated the same loss functions i.e. Dice loss and IoU score, so that we can compare our result with the original model.

1. **Dice loss function:**

This loss is obtained by calculating smooth [dice coefficient](https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%93Dice_coefficient)function. This loss is the most commonly used loss is segmentation problems. The formula of dice coefficient calculation is given below.



1. **Intersection over union(IoU) loss function:**

The IoU-balanced classification loss aims at increasing the gradient of samples with high IoU and decreasing the gradient of samples with low IoU. In this way, the localization accuracy of machine learning models is increased.



The table 5.1 shows the comparison of loss functions calculations between our model and original model.

|  |  |  |
| --- | --- | --- |
| Loss Functions | Original Model (U-Net++ with VGG16) | Comparison (U-Net++ with VGG19) |
| Dice Loss | 0.8043 | 0.9230 |
| Intersection over Union (IoU) | 0.9076 | 0.8066 |

Table 5.1: Loss Functions comparison between Original model and Modified U-Net++ with VGG19

Thus from the above table we can conclude that U-Net++ architecture works better with the latest VGG19 encoder and can give better output than the original model.

We also compared the segmentation results produced by U-Net++ architecture with VGG16 and DenseNet and U-Net++ architecture with VGG19, DenseNet and Efficient-Net architecture. Figure 5.1 and 5.2 shows the segmentation results of both the architectures.

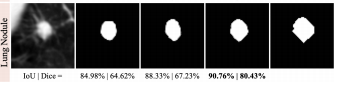


Figure 5.1: Segmentation results of U-Net++ architecture with VGG16

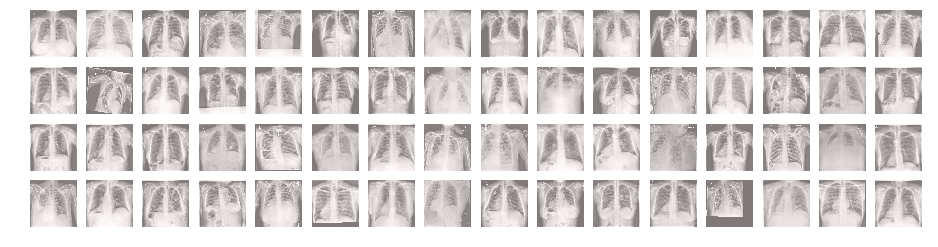


Figure 5.2: Segmentation results of U-Net++ architecture with VGG19

From the above images we can clearly see that U-Net++ architecture works better with the latest pre-trained VGG19 encoder and can produce better results.

* 1. **Future Discussions:**

The U-Net++ architecture is proposed in 2018 and after that many researchers proposed various architectures based on U-Net which were using Nested U-Net like U-Net++ architecture. Nested architectures like U2 – Net architecture which is proposed in 2020 are given to replace the U-Net++ architecture as obsolete image segmentation technique but since they don’t have deep supervision and dense convolution, they couldn’t replace the Nested U-Net or U-Net++ architecture. The U2 – Net architecture was using ResNet as encoder while U-Net++ use the VGG16 which performs better than the Residual Net or ResNet. Thus, U-Net++architecture can be used further because of its nested architecture and its robustness. The Nested Skip pathways in the U-Net++ allows it to generate the full resolution feature maps. Thus we can say that for now U-Net++ architecture which is also known Nested U-Net architecture can’t be replaced by any other architecture. It can work efficiently and can produce better results with any type of encoders. So, we can conclude that in future the U-Net based architectures can replace their backbone network from U-Net to U-Net++ architecture.

**CHAPTER 6**

**CONCLUSIONS**

We have presented a novel architecture, named UNet++ architecture with VGG19 encoder, for more accurate image segmentation. The improved performance by our model is attributed to its nested structure and redesigned skip connections, which aim to address two key challenges of the U-Net++ with VGG16:

1) Unknown depth of the optimal architecture.

2) The unnecessarily restrictive design of skip connections.

3) The limits for encoder restrict its efficiency it can work efficiently with any encoder and produces the results relatively.

We have evaluated UNet++ with VGG19 encoder using biomedical imaging dataset and demonstrated consistent performance improvement over various baseline architectures for semantic segmentation and meta framework for instance segmentation.

**CHAPTER 7**

**Future Scope of our work**

We have trained the Nested U-Net architecture with the VGG19 encoder. In our model we have also added the Efficient-Net network and Xception-Net library which improved the segmentation results of original Nested U-Net network which was trained using VGG16 encoder and DenseNet network. The deep supervision of this architecture makes it easy to add with the new encoder. In future we can use the U-Net++ architecture for Brain tumor segmentation which can create an evolutionary change in the field of medical imaging. The medical image segmentation always stood as a challenge for the image segmentation architecture. It is always difficult of indentify the area of abnormality in the medical images. We created a modified architecture which does the segmentation process much efficiently than its predecessor model and thus we concluded that U-Net++ architecture can be added with the many architectures for image segmentation too. Thus in future this architecture can become a baseline model for the other models based on the Nested U-Net architectures and uses dense skip connections. In the future this model can be trained and perform better than our model if added with some efficient encoder, like VGG or DenseNet or SegNet etc. Thus this model can become the backbone of upcoming models and endorse widely the usefulness deep supervisions and dense connections.

**References**

1. Medical Image Segmentation Using Deep Learning: A Survey, Tao Lei, Senior Member, IEEE, Risheng Wang, Yong Wan, Bingtao Zhang, Hongying Meng, Senior Member, IEEE and Asoke K. Nandi, Fellow, IEEE.2020
2. Spatio-Temporal FCN for Semantic Video Segmentation: Mohsen Fayyaz, Mohammad Hajizadeh Saffar, Mohammad Sabokrou, Mahmood Fathy, Reinhard Klette, Fay Huang.2016
3. CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances: YuzhuJi,HaijunZhang, ZhaoZhang,MingLiu
4. Gated-SCNN: Gated Shape CNNs for Semantic Segmentation: Towaki Takikawa, David Acuna, Varun Jampani, Sanja Fidler.2019
5. U-Net: Convolutional Networks for BiomedicalImage Segmentation : Olaf Ronneberger, Philipp Fischer, and Thomas Brox, 2015
6. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning: Hoo-Chang Shin, Member, IEEE, Holger R. Roth, Mingchen Gao, Le Lu, Senior Member, IEEE, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, Ronald M. Summers , 2016.
7. Deep Learning using Rectified Linear Units (ReLU): Abien Fred M. Agarap 2019.
8. SKIP CONNECTIONS ELIMINATE SINGULARITIES: A. Emin Orhan, Xaq Pitkow, 2018
9. Image Super-Resolution Using Dense Skip Connections: Tong Tong, Gen Li, Xiejie Liu, Qinquan Gao
10. The Importance of Skip Connections in Biomedical Image Segmentation: Michal Drozdzal, Eugene Vorontsov, Gabriel Chartrand, Samuel Kadoury and Chris Pal
11. Brain Tumor Segmentation and RadiomicsSurvival Prediction: Contribution to the BRATS 2017 Challenge: Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus,and Klaus H. Maier-Hein.
12. DoubleU-Net: A Deep Convolutional NeuralNetwork for Medical Image Segmentation : Debesh Jha, Michael A. Riegler, Dag Johansen, 2018
13. Deep Volumetric Feature Encoding for Biomedical Images: Brian Avants,Elliot Greenblatt, Jacob Hesterman, Nicholas Tustison, 2019
14. UNet++: A Nested U-Net Architecturefor Medical Image Segmentation: Zongwei Zhou, MdMahfuzur Rahman Siddiquee,NimaTajbakhsh, and Jianming Liang, 2018
15. Building Detection on Aerial Images Using U-NET Neural Networks: Leonid Ivanovsky, Vladimir Khryashchev, Vladimir Pavlov , Anna Ostrovskaya.[ISSN 2305-7254], 2018
16. A Probabilistic U-Net for Segmentation of Ambiguous Images: Simon A. A. Kohli, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R. Ledsam, Klaus H. Maier-Hein,S. M. Ali Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. [NeurIPS 2018]
17. Bi-Directional ConvLSTM U-Net with Densley Connected Convolutions: Reza Azad, Maryam Asadi-Aghbolaghi, Mahmood Fathy , 2019
18. nnU-Net for Brain Tumor Segmentation :Fabian Isensee, Paul F. Jager, Peter M. Full, Philipp, and KlausH.Maier-Hein.
19. SCAU-Net: Spatial-Channel Attention U-Net for Gland Segmentation: [Peng Zhao](https://www.frontiersin.org/people/u/898011), [Jindi Zhang](https://www.frontiersin.org/people/u/934835), [Weijia Fang](https://www.frontiersin.org/people/u/825617) and [Shuiguang Deng](https://www.frontiersin.org/people/u/910579) [ORIGINAL RESEARCH published: 03 July 2020doi: 10.3389/fbioe.2020.00670]
20. U^2-Net: Going Deeper with Nested U-Structure for Salient Object Detection:[Xuebin Qin](https://arxiv.org/search/cs?searchtype=author&query=Qin%2C+X), [Zichen Zhang](https://arxiv.org/search/cs?searchtype=author&query=Zhang%2C+Z), [Chenyang Huang](https://arxiv.org/search/cs?searchtype=author&query=Huang%2C+C), [Masood Dehghan](https://arxiv.org/search/cs?searchtype=author&query=Dehghan%2C+M), [Osmar R. Zaiane](https://arxiv.org/search/cs?searchtype=author&query=Zaiane%2C+O+R), [Martin Jagersand](https://arxiv.org/search/cs?searchtype=author&query=Jagersand%2C+M)[2020]
21. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation: [Huimin Huang](https://arxiv.org/search/eess?searchtype=author&query=Huang%2C+H), [Lanfen Lin](https://arxiv.org/search/eess?searchtype=author&query=Lin%2C+L), [Ruofeng Tong](https://arxiv.org/search/eess?searchtype=author&query=Tong%2C+R), [Hongjie Hu](https://arxiv.org/search/eess?searchtype=author&query=Hu%2C+H), [Qiaowei Zhang](https://arxiv.org/search/eess?searchtype=author&query=Zhang%2C+Q), [Yutaro Iwamoto](https://arxiv.org/search/eess?searchtype=author&query=Iwamoto%2C+Y), [Xianhua Han](https://arxiv.org/search/eess?searchtype=author&query=Han%2C+X), [Yen-Wei Chen](https://arxiv.org/search/eess?searchtype=author&query=Chen%2C+Y), [Jian Wu](https://arxiv.org/search/eess?searchtype=author&query=Wu%2C+J) [2020]
22. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghe-mawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S.,Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu,Y., Zheng, X., Brain, G.: TensorFlow: A system for large-scale machine learning.In: OSDI. pp. 265–284 (2016)
23. Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net : Learningdense volumetric segmentation from sparse annotation. In: MICCAI. pp. 424–432(2016)
24. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann,J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for thepre-operative scans of the TCGA-LGG collection. The Cancer Imaging Archive(2017)
25. A. Is ̧ın, C. Direko ̆glu, and M. S ̧ ah, “Review of MRI-based brain tumorimage segmentation using deep learning methods,”Procedia Comput.Sci., vol. 102, pp. 317–324, 2016.
26. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani,J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiestet al., “Themultimodal brain tumor image segmentation benchmark (BRATS),”IEEE Trans. Med. Imag., vol. 34, no. 10, pp. 1993–2024, 2015
27. S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, “A survey of MRI-based medical image analysis for brain tumor studies,”Phys. Med. Biol.,vol. 58, no. 13, pp. R97–R129, 2013.
28. S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain tumor segmentation using convolutional neural networks in MRI images,”IEEE Trans.Med. Imag., vol. 35, no. 5, pp. 1240–1251, 2016
29. M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio,C. Pal, P.-M. Jodoin, and H. Larochelle, “Brain tumor segmentation withdeep neural networks,”Med. Imag. Anal., vol. 35, pp. 18–31, 2017.
30. Image Segmentation Algorithms Overview: Song Yuheng,Yan Hao