International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Modified U-Net Models for Coarse Tumor Segmentation

Rakshith Rai[1], Pavan Kamath[2], Prasanna[3], Rakesh Kotian[4], Vivek Sharma[5]

[1-4]Student, Dept. of CSE, Alvas Institute of Engineering and Technology, Moodubidri, India.
[5]Assistant Professor, Dept. of CSE, Alvas Institute of Engineering and Technology, Moodubidri, India.

---***---

Abstract - *U-net is a state of architecture which is used for biomedical image segmentation. It consists of contracting path and a expanding path. As the u-net architecture goes deeper vanishing gradient problem occurs which affects the context and the precise localization resulting in poor segmentation results. In this paper we propose two architectures named U-net+Resnet and U-net+Densenet. The proposed models involves the fusion of residual blocks and dense blocks to u-net architecture. By using resnet blocks in the contracting and expanding part of u-net the network can be essentially deeper and by using dense blocks in the contracting part of u-net the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers. The modified architectures have several advantages as compared to u-net as network goes deeper they lighten the vanishing-gradient problems, encourage feature reuse which results in good segmentation accuracy. We evaluate the modified u-net models on coarse breast tumor data sets. The modified u-net models obtain good improvements than the u-net model.*

***Key Words*: Gradient, Localization, Segmentation, Relu, Pooling**

1. INTRODUCTION

Convolutional neural networks[12] a class of deep neural networks cause expectional results in natural and biomedical classification and segmentation tasks they have become the dominant machine learning approach for visual object recognition. However the abnormal use of convolutional networks is on classification tasks, where the output to an image is a single class label .However to process biomedical image processing, the desired output should include localization which is a class label that is supposed to be assigned to each pixel. To solve this problem an architecture U-net[1] convolutional neural network was introduced which consisted of contracting path and expanding path which provided good localization of pixel. This network won the EM segmentation challenge at ISBI 2012. As cnn networks goes deeper a new problem arises as information about the input or gradient passes through many layers, it can vanish till it reaches the beginning or end of the network. More layers is better but the network goes deep the model weights cannot be updated through back propagation of the error gradient. When the network depth increases, accuracy gets saturated and then degrades rapidly and adding more layers to a suitably deep model leads to higher training error. This can be accomplished by using Resnet's[2] shortcut connection which can be be substantially deeper, more accurate, and efficient to train. Another deep network DenseNets[5] have several advantages they lighten the vanishing-gradient problem, strengthen feature propagation and encourage feature reuse. This network enables maximum information flow between layers because each layer is connected to another in feed-forward fashion.

In this paper, we build upon a more elegant architectures, so called U-net+Resnet and U-net+Densenet. We modify and extend this architectures such that it works with very few training images and yields more precise segmentations. The main idea in U-net+Resnet is to use shortcut connection which is used to skip one or more layers and use two residual block at each contracting path and expanding path of the u-net architecture. In U-net+Resnet we combine features through summation before passing into the next layer. In U-net+Densenet we use denseblock at the contracting path of the u-net unlike in resnet the features maps are summed here the layers are concatenated. Each layer reads the state from its preceding layer and writes to the subsequent layer. This help for better improved flow of information and gradients throughout the network, which yields good segmentation accuracy then u-net.

2. LITERATURE SURVEY

Jonathan Long, Evan Shelhamer, Trevor Darrell and UC Berkeley[3] they have defined a novel architecture that combines semantic information from a deep layers and fine layer to produce accurate and detailed segmentations. Vijay Badrinarayanan, Alex Kendall and Roberto Cipolla[4] proposed a novel and practical deep fully convolutional neural network architecture for semantic pixel-wise segmentation termed SegNet samples more efficiently. The u-net architecture consists two parts a contracting path which captures context and a expanding path which is symmetric that enables precise localization. In their work they show that u-net network can be trained end-to-end from very few images.

Simon J´egou, Michal Drozdzal, David Vazquez1, Adriana Romero1 and Yoshua Bengio[6] to deal with segmentation problem, extended Densenet network. Many improvements were done on the u-net architecture for improving accuracy like in[7][10]. Sachin Mehta, EzgiMercan, Jamen Bartlett, Donald Weaver, Joann G. Elmore and Linda Shapiro[11] proposed y-net architecture for joint classification and segmentation their method efficiently segments different types of tissues in breast biopsy images while simultaneously predicts a discriminative map for identifying important areas in an image.

To train a Deep neural network is very hard. KaimingHe, Xiangyu Zhang, ShaoqingRen and Jian Sun [3] in their work they present a residual learning framework to ease the training of networks. They explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. As the networks goes deeper vanishing gradient problem occurs which causes poor accuracy and degradation of the model. There solution presented in their work is by construction to the deeper model the added layers are identity mapping, and the other layers are copied from the learned shallower model. Shortcut connections are introduced which add a linear layer connected from the network input to the output.

Simon Gao Huang, Liu Tsinghua, Laurens van Kilian and Weinberger Cornell[5] in their work has shown that convolutional networks can be essentially deeper and much efficient to train if they contain shorter connections between layers close to the input and those close to the output. In the their work they introduce dense convolutional Network which connects each layer to every other layer in a feed-forward fashion. In each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers which is done by concatenation operation. There are many advantages of this network some includes solving the vanishing-gradient problem, encourage feature reuse. It has also been shown that their proposed DenseNet network obtain significant improvements over many state-of-the art methods.

3. NETWORK SPECIFICATIONS

3.1. U-net+Densenet

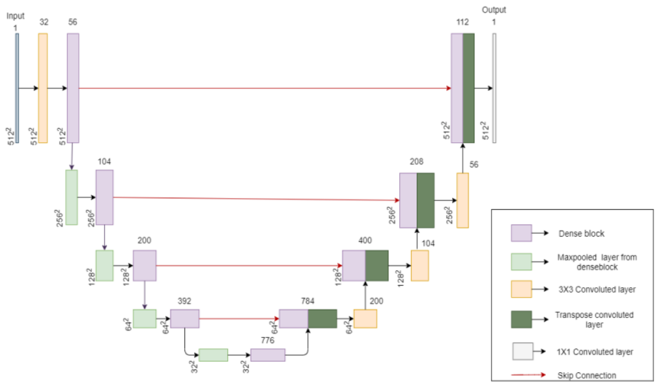

Fig-1. U-net+Densenet architecture

The proposed architecture of U-net+Densenet is presented in fig 1 for clear illustration we use blocks with different colors to indicate different layers. For simplification purpose each convolution operation in the above architecture is followed by batch normalization, relu and dropout which is discarded in above figure. The left part of architecture represents the contracting part. The encoder starts with an input image of channel 1 then convoluted with filter size 3x3 which is given as input to the dense-block. Max-pooling is only the pooling layer used in the down-sampling part of the architecture. After the operation inside the dense-block the concatenated feature output from the dense-block is max-pooled with stride of 2 and layer is given as input to the dense-block. The operation inside dense-block is depicted in fig 2.

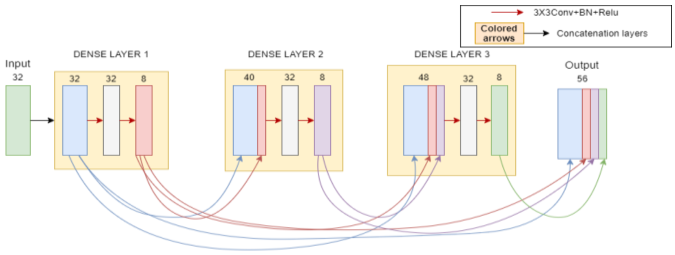

Fig-2. Dense-Block

Fig 3.2 represents a Dense-block consisting of three layers. Growth rate K is used to learn feature maps at each dense-block which is initially set to 8 in our architecture. In down-sampling the outputs from max-pooled operation is taken as input to the first layer of dense-block. In each layer there is 3*3 convolution with filters same as the max-pooled layer output which is then convoluted for filter size 3*3 with filters which is equal to growth rate. In dense-layer1 block the first layer and the last layer is concatenated which is given as input to the dense-layer 2 and same operation follows in dense-layer 2 and dense-layer 3 which finally produces a concatenated output feature dimension. In each contracting path the growth rate is multiplied by a factor of 2.

The expanding part adopts the U-net architecture. After the down-sampling operations the feature maps are up-sampled by transposed convolution with filters same as of previous contracting path layer and stride factor 2 which is then concatenated channel wise with feature maps with similar in size with the corresponding contracting path and then convoluted for filter size 3*3 with previous contracting path layer filters. The final output segmented image of 1x1 convolutions is obtained. This helps for deeper network since the proposed architecture has more numbers of convolution layers than U-net and the vanishing gradient problem for deeper network is overcome. The contracting path has concatenated feature outputs at the ends where reused features helps in favorable reconstruction of output segmented image.

3.2. U-net+Resnet

The second proposed architecture U-net+Resnet is presented in fig 3.

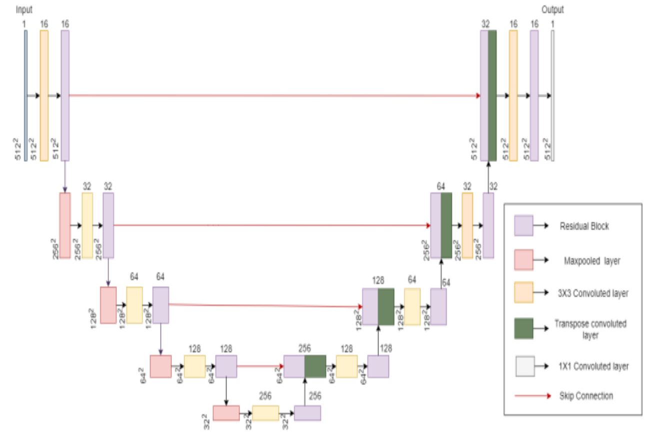

Fig -3. U-net+Resnet Architecture

Each convolution operation is followed by batch normalization, relu and dropout. The left part of architecture shows the contracting path with residual block fusion .The encoder starts with input image of channel1 which is convoluted with filter size 3*3 and given as input to the residual block. The down-sampling path consists of output from previous Residual block which is being max-pooled with stride factor 2 and the max-pooled layer is then convoluted for filter size 3*3 and given as input to residual block successively in the contracting path. The operation inside the Residual block is depicted in Fig 4.

In Each Residual block there are two residual block layers which is shown in fig 4. In each layer there is two 3*3 Convolution layers and each convolution operation is followed by batch normalization and relu operations. Skip connections adds feature maps of input layer with the output of the residual block layer.

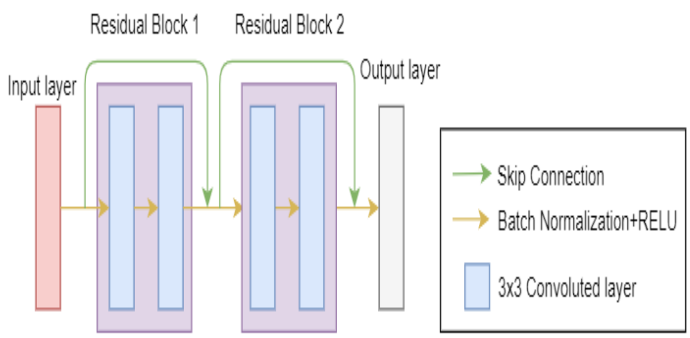

Fig 4. Residual Block

The Up-sampling part also consists residual blocks. The feature maps obtained are up-sampled by transposed convolution with filters same as of corresponding contracting path features which is then concatenated channel wise with feature maps on the corresponding contracting path and is convoluted for 3*3 convolution and this given as input to the residual block. The output from residual block is then transposed convoluted. The final output segmented image of 1x1 convolution is obtained. This proposed architecture helps for deeper network by solving vanishing gradient problem due to skip connections and more number of convolutions as compared to general u-net.

4. EXPERIMENT

In this section we demonstrate the training strategies and evaluation of two proposed modified u-net models. The modified u-net models is compared with the existing u-net model. For evaluation we used coarse breast tumor data sets.

4.1 Datasets

The data sets is composed of whole slide images. There are 3 classes of data sets used in our experiment. The class namely are benign, in-situ carcinoma and invasive carcinoma. The data set contains 216 images distributed as benign: 56, in-situ carcinoma: 60, invasive carcinoma:100 which are colored images with size of 512*512.Each image comes with a ground truth where cancerous area is labeled as white and non-cancerous area is labeled as black. Training is done on each class. From each class data set we pick 10% validation images and two testing images which are chosen randomly and the rest for training. We divide the pixel values by 255 so they are in the range 0 to 1.

4.2. Training

In this sub-section the training of modified u-net models is being conferred. The networks have almost identical hyper-parameters with Adam as optimization algorithm. The first proposed architecture U-net+Resnet which is fusion of residual blocks to u-net is trained with batch size of 8 for 30 epochs. The second proposed architecture U-net+Densenet which is fusion of dense blocks to u-net is trained with batch size of 4 for 23 epochs .The U-net architecture is trained for 40 epochs with batch size of 4.The learning rate is fixed to 1e-4 and the dropout is set to 0.035 which is same to all the networks. We use Keras deep learning framework for implementation and training of the architectures. Loss function used in our experiment is binary cross entropy. The network is trained on a Google Colab'sK80 Tesla, and the training is stopped when the validation loss no longer decreases.

4.3. Evaluation

The proposed architectures and the U-net architecture is evaluated on the three class validation data sets. The evaluation criteria's used for measuring the performance of network are validation accuracy, loss and mean intersection of union. The intersection over union is calculated for a validation data by the formula

Intersection over union= (Area of Intersection)/(Area of Union)

The mean Intersection over union is calculated over each class on validation set.

Classes	Mean IOU		
	Unet	Unet+ Resnet	Unet+ Densenet
Benign	0.60	0.81	0.79
Insitu carcinoma	0.54	0.77	0.78
Invasive Carcinoma	0.58	0.76	0.78

Table-1: Mean-Iou Calculation

Table 1 shows the comparison result of mean-Iou of modified u-net models and the u-net model. The modified models outperforms the u-net models in terms of mean intersection over union values.

Table 2 shows the validation loss and accuracy of proposed architecture and u-net architecture.The losses of proposed models are much lower than compared to existing u-net and accuracy has been around 11% more than the existing u-net architecture.

Fig 5 represents the comparison between the predicted outputs of U-net,U-net+Resnet, U-net+Densenet. The predicted outputs of the proposed methods are better than the existing u-net architecture hence yielding better intersection over union value.

Classes	Validation					
	Unet		Unet+Resnet		Unet+Densenet	
	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy
Benign	0.44	0.77	0.22	0.91	0.27	0.89
Insitu Carcinoma	0.45	0.78	0.22	0.90	0.25	0.90
Invasive Carcinoma	0.39	0.78	0.28	0.89	0.21	0.89

Table-2: Loss and accuracy

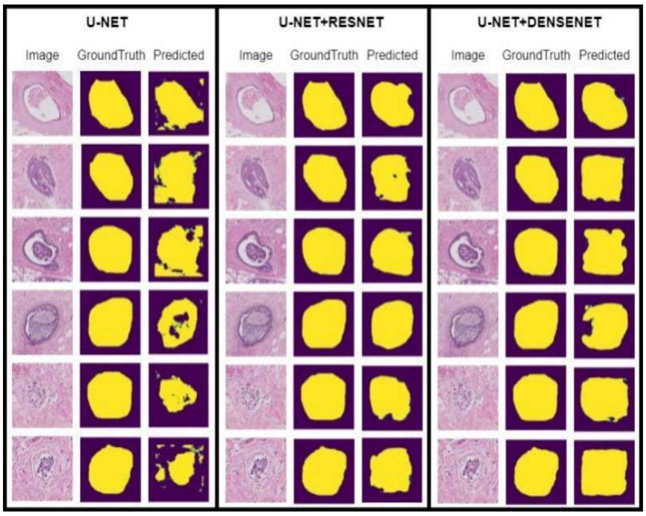

Fig-5: Segmentation output

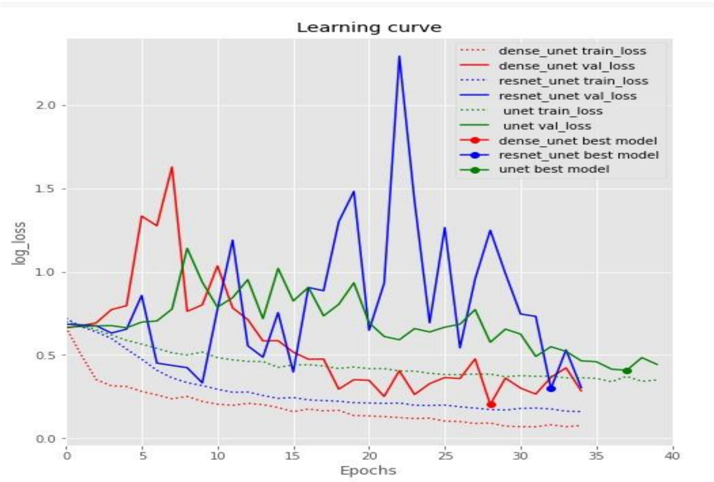

Fig -6: Graph of U-net+Densenet, U-net+Resnet and U-net on the validation data set of Benign class.

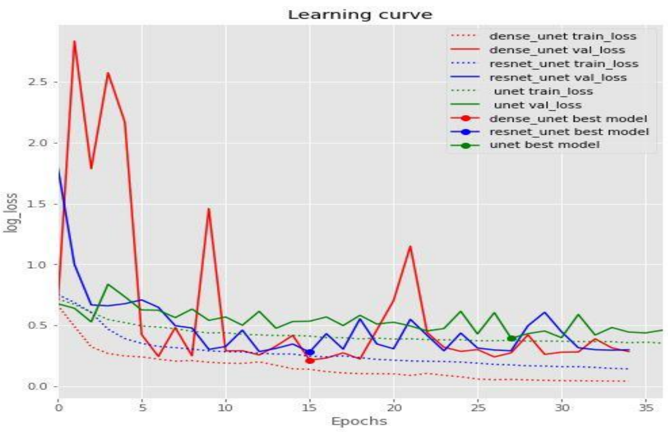

Fig -7: Graph of U-net+Densenet, U-net+Resnet and U-net on the validation data set of invasive carcinoma class.

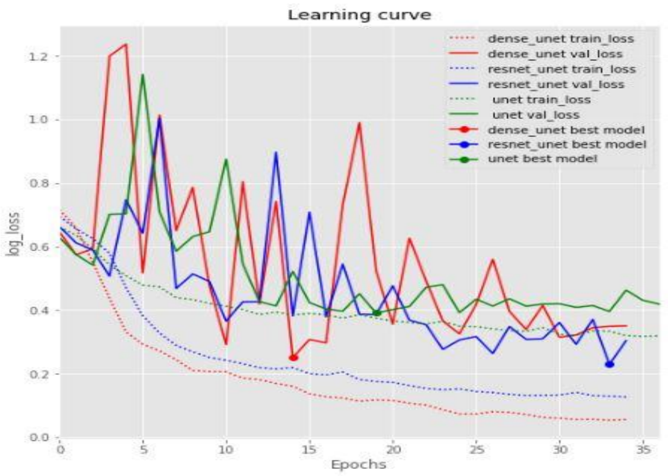

Fig -8: Graph of U-net+Densenet, U-net+Resnet and U-net on the validation data set of in-situ carcinoma class

5. CONCLUSION

In this work, we propose the modified u-net models named U-net+Resnet and U-net+Densenet for coarse tumor segmentation. The U-net+Resnet combines the residual blocks in encoder and decoder part of U-net and the U-net+Densenet combines dense-blocks in the encoder part of u-net. The comparative experiment shows that the proposed modified u-net models outperformed segmentation results of u-net on coarse tumor data sets. This work can also be extended for Y-net architecture and other biomedical image tasks.

REFERENCES

[1] Olaf Ronneberger, Philipp Fischer,Thomas Brox: U-Net-Convolutional Networks for Biomedical Image Segmentation,arXiv:1505.04597v1 2015.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Deep Residual Learning for Image Recognition,Cvpr 2016.

[3] Jonathan Long,Evan Shelhamer ,Trevor Darrell,UC Berkeley: Fully Convolutional Networks for Semantic Segmentation, arXiv:1411.4038v2, 2015.

[4] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, arXiv:1511.00561v3, 2016.

[5] Gao Huang, Liu Tsinghua, Laurens van Kilian ,Q. Weinberger Cornell: Densely Connected Convolutional Networks, arXiv:1608.06993v5 2018.

[6] Simon J´egou, Michal Drozdzal, David Vazquez1,Adriana Romero1,Yoshua Bengio1:The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation, arXiv:1611.09326v2,2016.

[7] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu Pheng-Ann Heng: H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volume, arXiv:1709.07330v3 2018.

[8] Jindong Jiang, Lunan Zheng, Fei Luo, and Zhijun Zhang RedNet: Residual Encoder-Decoder Network for indoor RGB-D Semantic Segmentation, arXiv:1806.01054v2 ,2018.

[9] Li Xu,Jimmy SJ. Ren ,Ce Liu ,Jiaya : Deep Convolutional Neural Network for Image Deconvolution 2017.

[10] Thorsten Falk1, Dominic Mai1, Robert Bensch1, Özgün Çiçek1, Ahmed AbdulkadirYassine Marrakchi1, Anton Böhm1, Jan Deubner, Zoe Jäckel, Katharina Seiwald, Alexander Dovzhenko, Olaf Tietz, Cristina Dal Bosco, Sean Walsh, Deniz Saltukoglu, Tuan Leng Tay, Marco Prinz, Klaus Palme, Matias Simons2, Ilka Diester, Thomas Brox1, and Olaf Ronneberger1: U-Net – Deep Learning for Cell Counting, Detection, and Morphometry, DOI: http://dx.doi.org/10.1038/s41592- 018-0261-2,2019

[11] Sachin Mehta, Ezgi Mercan, Jamen Bartlett, Donald Weaver, Joann G. Elmore, Linda Shapiro1: Y-Net: Joint Segmentation and Classification for Diagnosis of Breast Biopsy Images: arXiv:1806.01313v1, 2018.

[12] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classifification with deep convolutional neural networks. In: NIPS. pp. 1106–1114 (2012).