HAI Assignment 2 Report

MultiResUNet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation

Anonymous CVPR submission

Paper ID ****

Abstract

Medical image segmentation has seen a boost in recent years because of Deep Learning. In this aspect, U-Net is the most well-known deep network but in Jan 2020, MultiResUNet [1] has been proposed as a successor of traditional UNet, by Samsung, and Bangladesh University of Engineering and Technology. MultiResUNet outperforms U-Net on majority of datasets. This report contains re-implementation of MultiResUNet and comparison of outputs generated by MultiResUnet, UNet++ (another model deemed to be successor of UNet) and UNet on four publically available datasets. The performance of MultiResUNet model is motivating in majority of cases.

1. UNet vs MultiResUNet

1.1. UNet

The U-net network (Figure 1) can be divided into two part: The first part is Encoder, it employs a standard CNN architecture which includes two successive 3x3 convolution layer followed by a ReLU activation unit and a max-pooling layer in each block of the contracting path. This pattern is repeated multiple times. Then second is Decoder, it uses 2×2 Up-convolution and two times of 3×3 convolution is done to recover the size of segmentation map. While doing this model can get advanced features, but also loss the localization information. Thus, in the end, Using Skip connections, each block of the decoder is connected to the corresponding block of the encoder. This helps to give the localization information from contraction path to expansion path. At the final stage, an additional 1x1 convolution is applied to reduce the feature map to the required number of channels and produce the segmented image.

1.2. Problems identified in UNet

The authors of paper [1] identifies two major issues with existing state of the art UNet.

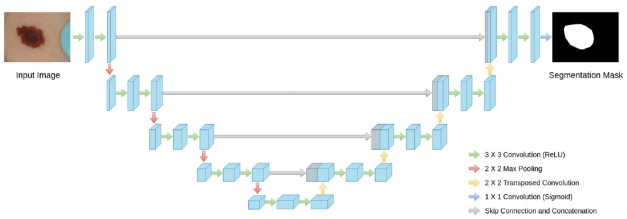

Figure 1. UNet Architecture

1. **Variation of Scale in Medical Images:** In most of the cases, the images of cell necluei, organ, tumor, etc. which contain various modalities are of irregular and different scales. Therefore, a network must be capable of analysing objects of various sizes.

2. **Probable Semantic Gap between the Corresponding Levels of Encoder-Decoder:** In skip connections, preservation of spatial features that gets lost during the pooling operation is done. But, the features in up-sampling layers (decoder) are complex or of high level when compared to features in early layers of network (encoder) since they are computed at the very deep layers of the network. As a result, they undergo additional processing, and the concatenation of these two seemingly irreconcilable sets of features may result in some discrepancy throughout the learning process.

1.3. MultiResUNet

For the sequence of two convolutional layers at each level in the original U-Net, they are replaced by the proposed MultiRes block (Figure 2). First, start with a simple Inception-like block by using 3×3, 5×5 and 7×7 convolutional filters in parallel, to reconcile spatial features from different context size. Then, large filter is factorized into a succession of 3 × 3 filters. Finally, MultiRes block is established, by increasing the number of filters in the successive three layers gradually and adding a residual connection, along with 1×1 filters for

conserving dimensions. This is similar to the DenseBlock in DenseNet with the residual path, originated in ResNet. For the ResPath (skip connections), there are 3×3 and 1×1 filters. Number of 3×3 and 1×1 filters depends on the level inside the network (Figure 3). These additional non-linear operations are expected to reduce the semantic gap between encoder and decoder features. In order for fair comparison with U-Net, similar number of parameters should be maintained between two models: W = α x U, where U and W are the number of filters in one convolutional layer in U-Net and MultiResUNet respectively. α = 1.67 is used. Thus, the filter numbers as shown below are multipled by α already. The architecture of MultiResUnet is shown in Figure 4.

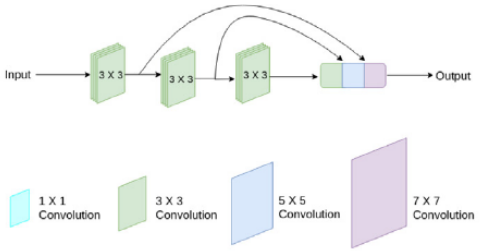

Figure 2. Developing the proposed MultiRes block

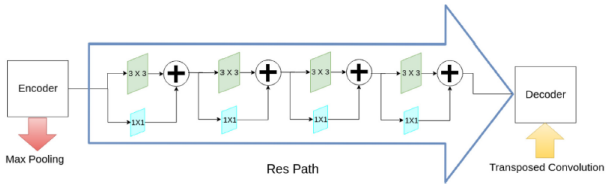

Figure 3. Proposed ResPath

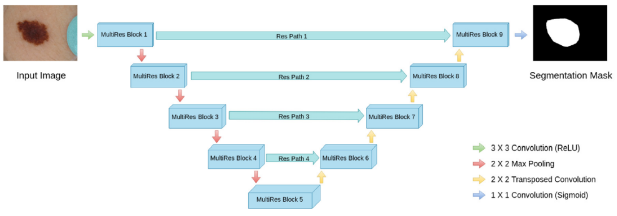

Figure 4. MultiResUNet Architecture

2. Datasets

Medical imaging datasets are more difficult to curate than typical computer vision datasets. So, the proposed model was tested on four unique two-dimensional datasets. Description of the dataset is given in Table 1:

Modality	Dataset	No. of images	Input Resolution
Microscopy (Cell nuclei)	Data Science Bowl 2018	670	256 x 256
Electron Microscopy	ISBI-2012	30	256 × 256
Endoscopy	CVC-ClinicDB	612	256 × 192
Dermoscopy	ISIC-2018	2594	256 × 192

Table 1. Datasets Overview

3. Results

The points of interest in semantic segmentation usually comprise of a small portion of the whole image. Therefore , metrics like precision and recall are inadequate, leading to a deceptive sense of superiority. Hence , the Jaccard Index is frequently used to assess and compare image segmentation. The models were trained with the help of Keras for 150 epochs using Adam optimizer. Authors of MultiResUNet claims that results do not show any further improvements after 150 epochs but while reimplementation of same it was found that ISBI-2012 dataset require atleast 200 epochs to get better results.

Dataset	Re-implementation Results	Author Results
Data Science Bowl 2018	84.65	———— (Not used)
ISBI-2012	88.10	87.95
CVC-ClinicDB	84.55	82.06
ISIC-2018	80.70	80.30

Table 2. MultiResUnet Results : Jaccard Index (in %) (IoU)

4. Conclusion

In conclusion, Re-implementation results almost resembles the proposed results in MultiResUnet paper. Furthermore, on the immensely challenging photos, U-Net tended to over-segment, under-segment, make incorrect predictions, and even totally overlook the objects. Although the proposed MultiResUNet's segmentations were not flawless, but it outperformed the traditional U-Net in the vast majority of circumstances.As a result, it can be said that MultiResUNet architecture has the potential to succeed the traditional U-Net architecture.

References

[1] N. Ibtehaz and M. S. Rahman. Multiresunet : Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, 121:74–87, 2020.