

CLL788 Process Data Analytics

Term Paper



ResUNet++: An Advanced Architecture for Medical Image Segmentation

Advisor: Prof. Hariprasad Kodamana

Contributor:

Rishav Kumar Rajak (2018CH70302)

Abstract

During colonoscopy examinations, accurate computer-aided polyp detection and segmentation can benefit endoscopists in resecting problematic tissue and reducing the likelihood of polyps developing into cancer. The author proposes ResUNet++ [1], an upgraded ResUNet architecture for colonoscopic image segmentation, as a first step toward constructing a fully automated model for pixel-wise polyp segmentation. Experimental tests indicate that the proposed architecture yields satisfactory segmentation results on publicly available datasets. Furthermore, ResUNet++ beats U-Net and ResUNet, two of the most popular state-of-the-art deep learning architectures, by getting high evaluation scores for the Kvasir-SEG dataset [2], with a dice coefficient of 78.09 percent and a mean Intersection over Union (mIoU) of 67.71 percent.

Introduction

Colorectal cancer (CRC) is one of the most common cancer-related causes of death globally. Polyps are precursors to this form of cancer; thus, physicians should look for them early during colonoscopy tests. It is common practice to resect neoplastic tumours to lower CRC risk (for example, adenomatous polyps). Many adenomatous polyps are unfortunately overlooked during endoscopic investigations. A CAD (Computer-Aided Detection) system that can highlight the positions of polyps in the video stream from the endoscope in real-time can operate as a second observer, perhaps attracting the endoscopist's attention to the polyps displayed on the monitor. It's less likely that any polyps may go unnoticed due to this. The authors also want our CAD system to have pixel-wise segmentation capability to identify the specific regions of interest within each abnormal image. This is an essential improvement over pure anomaly detection approaches, which only identify whether there is something abnormal in an image.

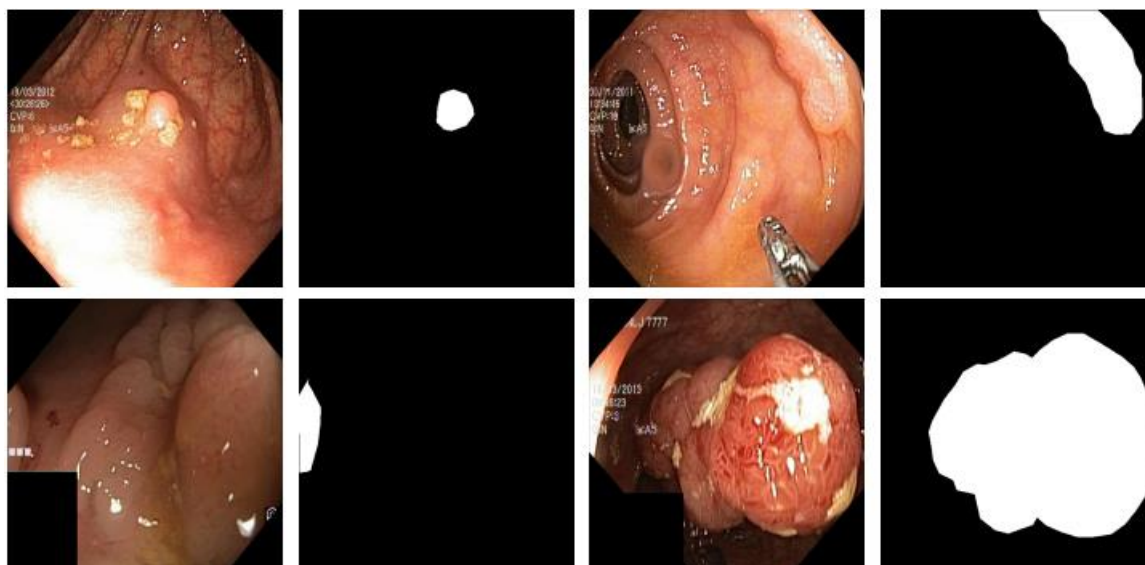


Fig 1. From the Kvasir-SEG dataset, examples of polyp pictures and their related masks. The original photos are represented by the first and third columns, while the ground truth is represented by the second and fourth columns.

The significant expenses of obtaining and identifying appropriate medical information for training and testing are a major issue in building an accurate CAD system for polyps. As seen in Figure 1, polyps arise in a broad range of forms, sizes, colours, and appearances. There is a lot of interclass similarities and intraclass variance among the four basic types of polyps: adenoma serrated, hyperplastic, and mixed (rare). Backdrop object similarity can also be high, for example, when sections of a polyp are covered

with excrement or merge into the background mucosa. Despite the fact that these considerations make authors work difficult, they believe there is still a lot of room for building a system with clinically acceptable performance.

The author examines how semantic segmentation-based approaches for medical image analysis might be utilised to improve the performance of automatic polyp segmentation and identification. This is motivated by the recent success of semantic segmentation-based approaches for medical image analysis. U-Net is a well-known deep learning architecture for semantic segmentation in biomedical applications. ResUNet is another outstanding architecture for image segmentation.

The ResUNet++ architecture for medical picture segmentation is proposed in this research. They have evaluated our model on both datasets. In comparison to the popular U-Net and ResUNet designs, our experimental results show that the enhanced model is efficient and achieves a performance gain.

In summary, the paper's contributions are as follows:

- 1) The authors present ResUNet++, a semantic segmentation neural network that uses residual blocks, squeezing and excitation blocks, Atrous Spatial Pyramidal Pooling (ASPP), and attention blocks. When compared to other state-of-the-art approaches, ResUNet++ greatly improved the colorectal polyp segmentation findings. With a lesser number of photos, the recommended architecture works effectively.
- 2) The authors created the new Kvasir-SEG dataset by annotating the polyp class from the Kvasir dataset [\[2\]](#) with the aid of an expert gastroenterologist. To encourage the development of novel approaches and repeatable research, they are making this polyp segmentation dataset available to the scientific community.

Architecture

The ResUNet++ architecture is based on the Deep Residual U-Net (ResUNet), which is a deep residual learning and U-Net-based architecture. The ResUNet++ design makes use of residual blocks, squeeze and excitation blocks, ASPP, and the attention block.

The residual block propagates information across layers, allowing a deeper neural network to be built that can tackle the deterioration problem in each encoder. This enhances the interdependencies across channels while lowering the computational cost. One stem block is followed by three encoder blocks, ASPP, and three decoder blocks in the proposed ResUNet++ architecture. Figure 2 depicts the proposed ResUNet++ architecture in block diagram form. The residual unit is a mix of batch normalisation, Rectified Linear Unit (ReLU) activation, and convolutional layers, as shown in the block diagram.

Each encoder block is made up of two 3x3 convolutional blocks that are connected by an identity mapping. A batch normalisation layer, a ReLU activation layer, and a convolutional layer are all included in each convolution block. The identity mapping connects the encoder block's input and output. At the first convolutional layer of the encoder block, a strided convolution layer is used to lower the spatial dimension of the feature maps by half. The squeeze-and-excitation block receives the encoder block's output. The ASPP functions as a bridge, allowing the filters' field of vision to be expanded to cover a larger environment. Similarly, residual units are present in the decoding route. The attention block, which comes before each unit, boosts the efficiency of feature maps. Following that, feature maps from the

lower level are nearest-neighbour up-sampled, and feature maps from their associated encoding route are concatenated.

The decoder block's output is routed via ASPP, and then the segmentation map is generated using a 1x1 convolution with sigmoid activation. The squeeze-and-excitation blocks are light blue, the ASPP block is dark red, and the attention block is light green in the ResUNet++ extension. The subsections that follow provide a brief description of each part.

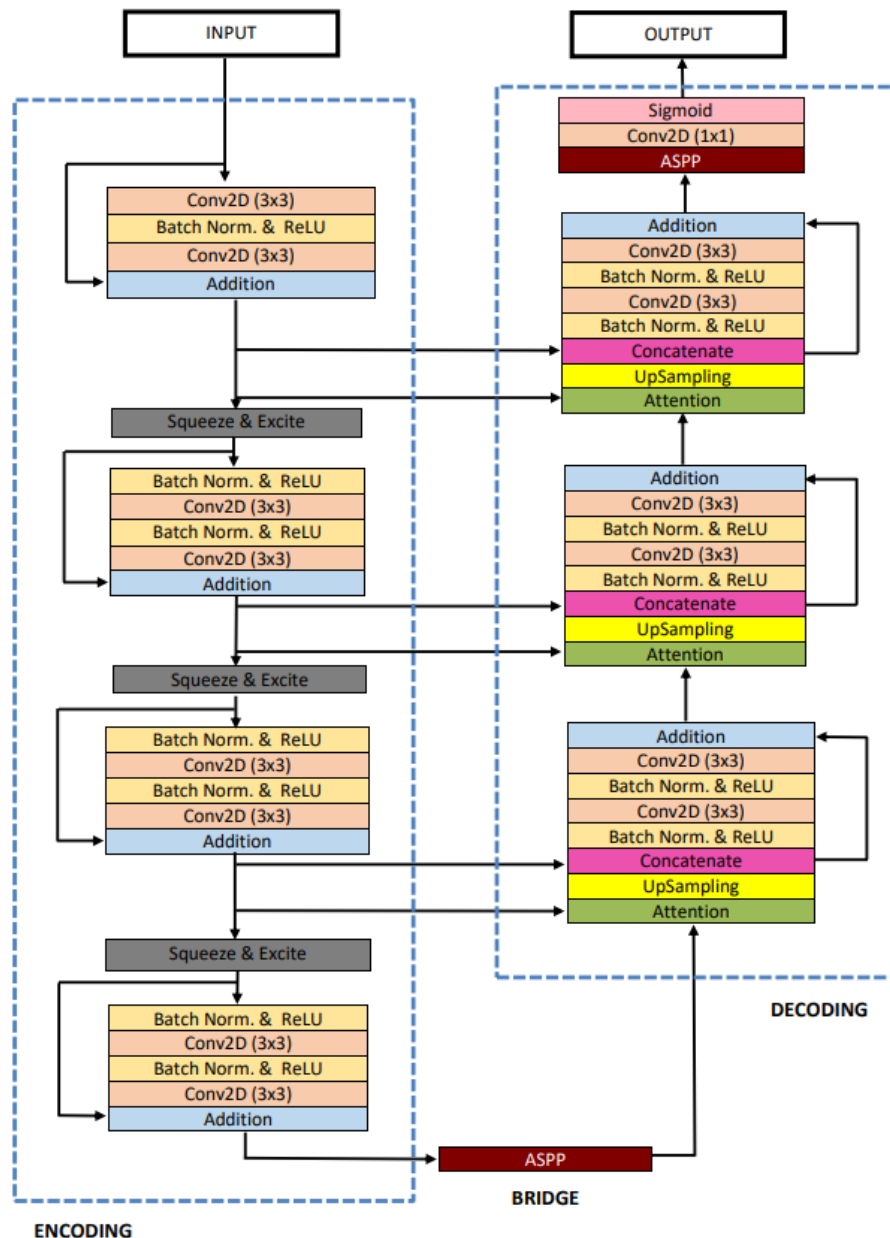


Fig 2. ResUNet++ Architecture

A. Residual Unit

Deeper neural networks are more difficult to train than shallower neural networks. The accuracy of a deep neural network may be improved by increasing the network depth. ResUNet employs entire residual units prior to activation. The deep residual unit makes the deep network simple to train, and the skip connection within the networks aids in information propagation without degradation, improving the design of the neural network by lowering parameters while maintaining or improving performance on

the semantic segmentation task. The authors chose ResUNet as the backbone architecture because of these benefits.

B. Squeeze and Excitation Unit

By re-calibrating the features responses using accurate modelling inter-dependencies across the channels, the squeeze-and-excitation network improves the network's representational power. The squeeze and excite block's purpose is to guarantee that the network's sensitivity to key features increases while irrelevant features are suppressed. This objective is accomplished in two phases. Squeeze (global information embedding) is the initial phase, in which each channel is compressed using global average pooling to get channel-wise statistics. Excitation (active calibration) is the second stage, which seeks to capture the channel-wise relationships fully. The squeeze and excitation blocks are stacked with the residual block in the proposed architecture to promote effective generalisation across diverse datasets and improve the network's performance.

C. Atrous Spatial Pyramidal Pooling

ASPP is based on spatial pyramidal pooling, which effectively resamples features at many scales. The contextual information is gathered at several scales in ASPP, and the input feature map is fused using numerous parallel atrous convolutions with varied speeds. Controlling the field-of-view via atrous convolution enables for accurate capture of multi-scale information. As illustrated in Figure 2, ASPP works as a bridge between the encoder and the decoder in our proposed design. The ASPP model has demonstrated promising results on numerous segmentation tasks by offering multi-scale information. As a result, they employ ASPP to acquire the necessary multi-scale data for the semantic segmentation job.

D. Attention Unit

In Natural Language Processing, the attention mechanism is most used. It concentrates on a subset of its input. It has also been used in semantic segmentation tasks such as pixel-wise prediction. The neural network's attention mechanism decides which sections of the network demand more attention. The attention mechanism also lowers the expense of encoding the information in each polyp picture into a fixed-dimensional vector. The key benefit of the attention mechanism is that it is simple, can be applied to any input size, and improves the quality of features, which improves the outcomes.

The encoder feature maps, and the decoder feature maps are directly concatenated in the previous two systems, U-Net and ResUNet. The authors implemented the attention block in the decoder component of our design to be able to focus on the critical parts of the feature maps, inspired by the success of attention mechanisms in both NLP and computer vision applications.

Experiments

The authors used two publicly accessible datasets to train, verify, and test models to assess the ResUNet++ architecture. According to the authors, our ResUNet++ models outperformed those trained using U-Net and ResUNet, according to the authors.

A. Dataset

Each pixel in the training photos must be identified as belonging to either the polyp class or the non-polyp class for the job of polyp image segmentation. The Kvasir-SEG dataset[2], which comprises of 1,000 polyp pictures and their related ground truth masks annotated by experienced endoscopists from Oslo University Hospital, is used to evaluate ResUNet++ (Norway). Figure 1 shows examples of pictures and their related masks from the Kvasir-SEG dataset.

B. Dataset Preprocessing

Within the same dataset, the picture size varies. Both the datasets utilised in the study contain photos of various resolutions. I cropped the images with a crop margin of 320x320 to enhance the training dataset for optimum GPU use and to shorten the training time. Before feeding the photos to the model, the images are shrunk to 256x256 pixels. To enhance the number of training examples, I employed data augmentation techniques such as Elastic Transformation, horizontal flip, vertical flip and grid distortion.

C. Implementation Details

All architectures were implemented using the Keras framework with TensorFlow as backend. I experimented on my personal laptop with dual Intel(R) i5 8th Gen CPU@2.70GHz, 8GB of DDR4-2667MHz DRAM. I start the training with a batch size of 2, and the proposed architecture is optimized by Adam optimizer. The learning rate of the algorithm is set to $1e-4$. The dataset is split as a ratio of 8:1:1, train, valid and test set respectively. I trained the model for 30 epochs.

Results

To optimize ResUNet++ architecture, I explored several sets of hyperparameters (e.g., learning rate, number of epochs, optimizer, batch size, and filter size). Hyperparameter tweaking may be done by manually training the models with multiple sets of hyperparameters and analysing their outcomes.

For the Kvasir-SEG dataset, Table I compares the results of authors and mine.

Method	Dice Coefficient	mIoU	Precision	Recall
ResUNet++ (Author's)	0.8133	0.7927	0.8774	0.7064
ResUNet++ (Mine)	0.7809	0.6771	0.8482	0.7093

Table 1. Performance Metrics

Due to computational disadvantage, I cannot train the model for larger epochs but at less epoch, my results are pretty decent. Additionally, I have plotted the dice coefficient vs epochs graph for training and validation set, Figure 3, the mean IoU vs epochs graph for training and validation set, Figure 4, the dice loss vs epochs graph for training and validation set, Figure 5.

Figure 6 shows the results of the predicted mask on test set.

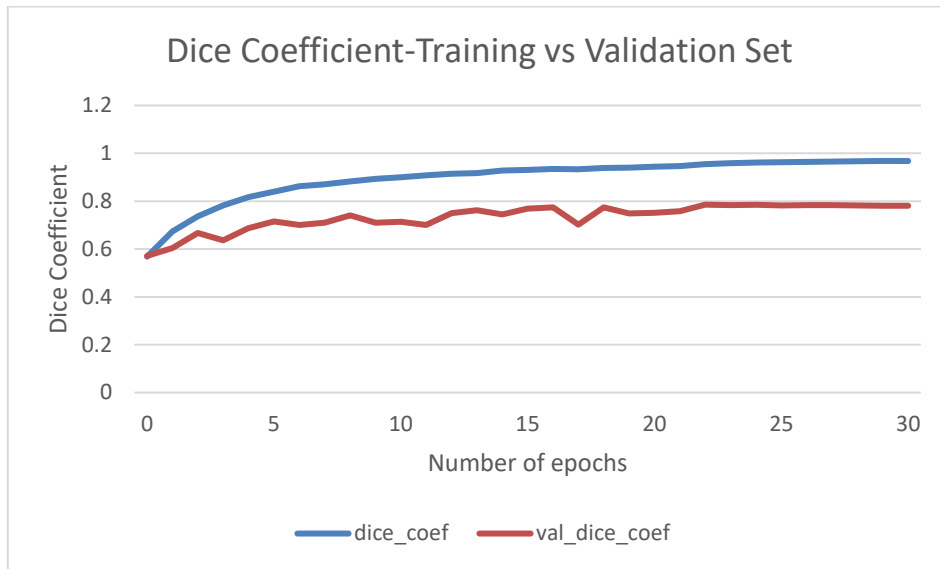


Fig 3. Dice Coefficient

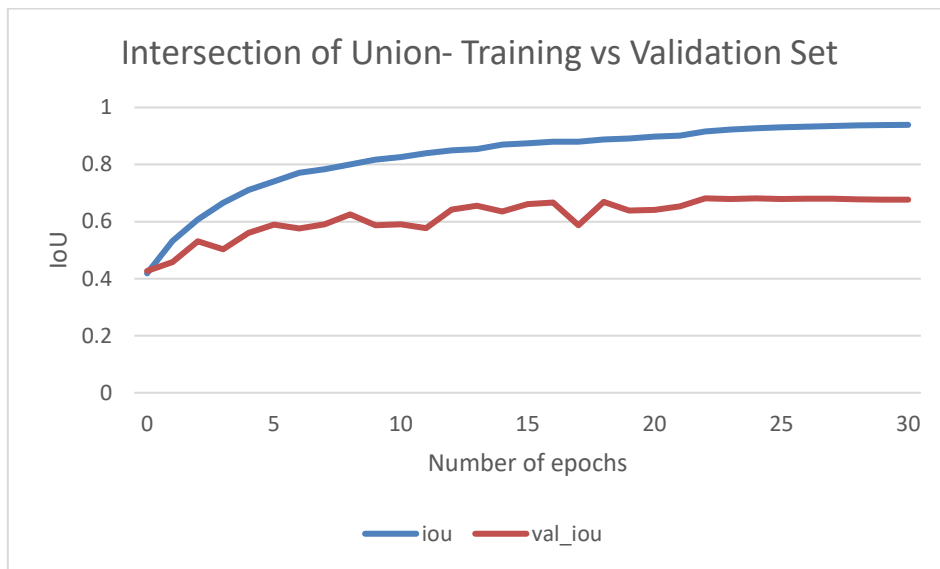


Fig 4. Intersection of Union

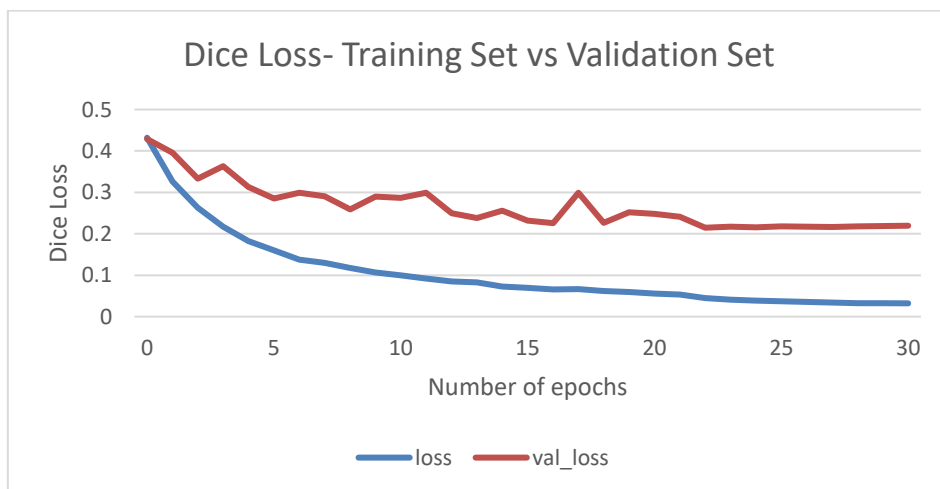


Fig 5. Dice Loss

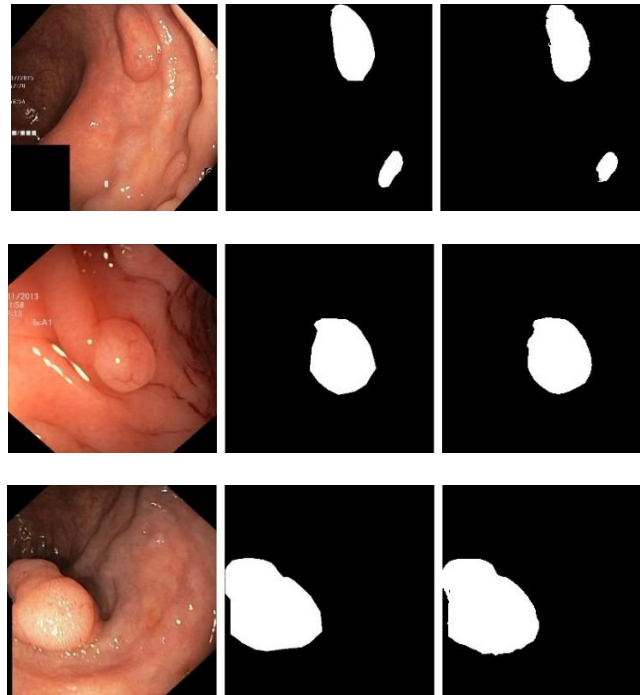


Fig 6. Image (Leftmost), Ground Truth (Middle), Prediction (Rightmost)

Discussion

I believe that by increasing the dataset size, using additional augmentation approaches, and performing certain postprocessing steps, the model's performance may be further enhanced. I trained the model to obtain improved performance despite the suggested architecture's increased number of parameters. To sum up, ResUNet++'s use should not be confined to biomedical image segmentation; it might also be extended to natural image segmentation and other pixel-wise classification tasks that need more comprehensive validations.

Conclusion

ResUNet++ is an architecture described in this work to meet the requirement for more precise segmentation of colorectal polyps identified during colonoscopy examinations. Residual units, squeeze and excitation units, ASPP, and attention units are all used in the proposed design. The proposed ResUNet++ design outperforms the state-of-the-art U-Net and ResUNet architectures in terms of delivering semantically correct predictions, according to a comprehensive assessment utilising various dataset. The suggested architecture may be a solid basis for additional exploration to build a therapeutically relevant strategy to achieve the generalizability objective. Postprocessing methods might be used on the model to improve segmentation results even more.

References

1. Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Dag Johansen, Thomas de Lange, Pal Halvorsen, Havard D. Johansen et.al ResUNet++: An Advanced Architecture for Medical Image Segmentation
2. D. Jha, P. H. Smedsrud, M. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. Johansen, "Kvasir-seg: A segmented polyp dataset," in International Conference on Multimedia Modeling. Springer, 2020. [Online]. Available: <https://datasets.simula.no/kvasir-seg/>