NORDIC MACHINE INTELLIGENCE

Nordic Machine Intelligence, MedAI 2021
https://doi.org/10.5617/nmi.9142

Explainable U-Net model for Medical Image Segmentation

Sahadev Poudel [1], Sang-Woong Lee[2] E-mail any correspondence to: slee@gachon.ac.kr

1. *Department of IT Convergence Engineering, Gachon University, Seongnam, South Korea*
2. *Department of Software, Gachon University, Seongnam, South Korea*

Abstract

In a nutshell, we propose a simple, efficient, and explainable deep learning-based U-Net algorithm to tackle the MedAI challenge 2021, focusing on precise segmentation of polyps and instruments in addition to a focus on transparency of the proposed algorithms. We develop a straightforward encoder-decoder-based algorithm for the task above. In addition we also make an effort to make the network as simple as possible. Especially, we focus on input resolution and width of the model to find the optimal settings for the network. We perform ablation studies to determine these settings.

Keywords: artificial intelligence; deep learning; medical imaging

Introduction

The aim of *MedAI* is the automatic segmentation of polyps and instruments with a focus on transparency of the applied machine learning-based algorithms [1]. For the polyp segmentation task, 1,000 polyp images with their corresponding segmentation masks labeled by medical experts are provided. Similarly, a set of 590 instrument images is provided for the instrument segmentation task. The third task includes a focus on the transparency of the algorithms by making for example the model explainable. Explainable and interpretable algorithms would be better suited to be deployed in clinical practice to provide the medical experts a better understanding and ultimately aiding in quantitative analysis, treatment organization, and follow-ups. Convolutional neural networks (CNNs), particularly fully convolutional networks (FCN) encoder-decoder with skip-connection techniques are widely popular and have achieved tremendous success in a myriad of medical applications such as polyp detection [2], skin disease classification [3] and Covid-19 segmentation on CT and X-ray images [4].

Recently, several attention-based methods have been proposed to increase the feature representation capabilities. Primarily, spatial and channel attention are widely popular and are used in computer vision tasks. However, these methods have poor explainability capabilities and often faced transparency issue. For this challenge, we inspect the potential of a simple U-Net model, where we only increase the width and the input resolution, to show that simpler architectures can also achieve acceptable performance. A simple encoder-decoder-based convolutional neural network (CNN) is introduced to perform efficient polyp and instrument segmentation using the provided data.

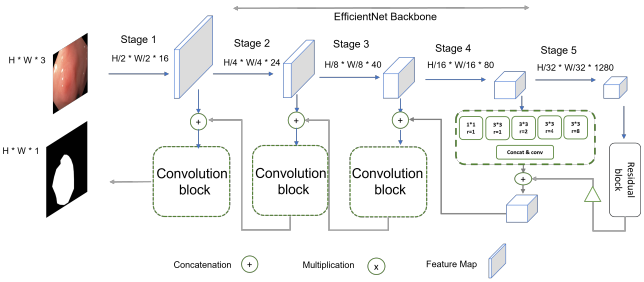

Figure 1: Overview of our proposed architecture.

Materials and methods

This section presents Ef-UNet, a simple, powerful, and efficient segmentation method for medical image segmentation. As shown in Figure 1, Ef-UNet consists of two parts: (1) a UNet encoder which uses EfficientNet [5] as a backbone that generates different semantic details in multiple stages; and (2) a decoder that integrates spatial information at different-stage and generates a final precise segmentation mask.

Table 1: The metrics calculated on the testing set results. Note that the first row indicates the polyp dataset results, and the second row indicates the results for the instrument segmentation.

Method	Jc	DSC	Rec	Pre	Accuracy	F1-Score
Method 1	0.8116	0.8669	0.8852	0.8922	0.9715	0.8669
Method 1	0.9178	0.9528	0.9687	0.9441	0.9901	0.9528

Given an image size of $H \times W \times 3$, we first feed the image into an encoder block (EfficientNet as an backbone) and obtain multi level features at $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$ and $\frac{1}{32}$ of the original input. The features generated at the fifth stage of the encoder side are passed through a residual block and later are upsampled for the skip-connection process. By leveraging the idea from Deeplab to exploit global features using different dilation rates, the output of stage 4 at the encoder side is passed through a convolution layer with different dilation rates and concatenated and convolved. Note that 256, 128, 64, 32, and 16 channels are set for stage 5, stage 4, stage 3, stage 2, and stage 1 of the decoder, respectively. Similarly, a similar setting is applied after increasing the channel widths.

Materials and Methods

For the evaluation, we have used the dataset provided by the challenge: Kvasir-SEG [6], and Kvasir Instrument segmentation [7]. We implemented all methods in the Pytorch framework with version 1.8.0. We utilized Adam optimizer while training with a learning rate of 10^{-4}, batch size of 16, and iteration of 200 epochs. The training was conducted on a V100 with two GPUs. We used the Dice loss function for the training of each method. Further, we used Dice Coefficient Score (DSC), Jaccard index (Jc), precision (Pre), and recall (Rec) for the quantitative evaluation.

Ablation Studies

We performed ablation study on 224×224 and 448×448 input resolution settings with different network widths as explained below:

- Method 1 applies the proposed method with 448×448 input resolution and width of [256, 128, 64, 32, 16].
- Method 2 extends method 1 with width of [512, 256, 128, 64, 32]
- Method 3 applies the proposed method with 224×224 input resolution and width of [256, 128, 64, 32, 16].
- Method 4 extends method 3 with width of [512, 256, 128, 64, 32]

Table 2: Result for the instrument validation subset.

Method	DSC	Jc	Rec	Pre
Method 1	**0.9579**	**0.9197**	**0.9553**	**0.9612**
Method 2	0.9443	0.8953	0.9338	0.9562
Method 3	0.9491	0.9039	0.9366	0.9629
Method 4	0.9428	0.8931	0.9383	0.9494

Table 3: Result for the Kvasir-SEG validation subset.

Method	DSC	Jc	Rec	Pre
Method 1	0.9220	0.8567	0.9107	0.9362
Method 2	0.9306	0.872	0.9259	0.9375
Method 3	0.9203	0.8535	0.9149	0.9283
Method 4	**0.9318**	**0.8739**	**0.9337**	**0.9315**

Discussion

The visual comparison is illustrated in Figure 2. By visualizing the attention maps at different layers, we can better analyze how the network behaves when increasing the input resolution and width. From Table 1, it can be seen that Method 1 achieved the best accuracy on the testing subset, indicating that increasing width and the input resolution can significantly improve the performance. It also shows the effectiveness and explainability of the network, unlike the existing attention-based network with poor explainability. From Table 2, it can be observed that Method 1 achieved outstanding accuracy on the validation set of the instrument segmentation dataset with 0.9579 of dice coefficient score and 0.9197 of Jaccard index. It shows that the accuracy is increased with increasing network input resolution. However, the same is not the case when increasing network width, and the reason could be the overfitting issues due to increasing network complexity. Similarly, in Kvasir-SEG, increment on accuracy with increasing width, and similar increased accuracy can be observed when increasing input resolution (see Table 3).

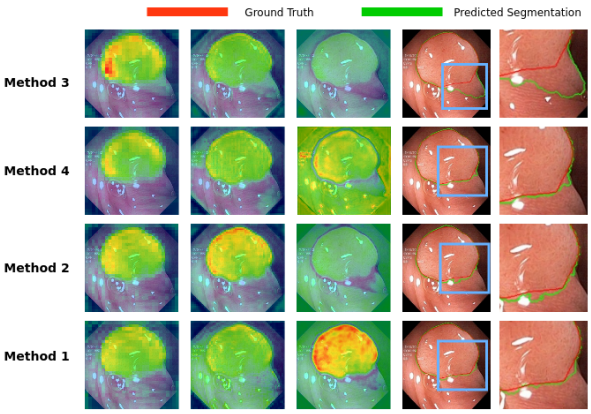

Figure 2: Visual comparison between different methods

Conclusion

This paper presented a simple and efficient UNet model for accurate segmentation of polyps and instruments. We focused on increasing input resolution and channel width to increase the performance. Despite the increasing complexities, we show that it has better explainability and can gain trust for deployment in a clinical environment. We plan to continue researching efficient and lighter models and further improve the results.

Acknowledgments

This work was partly supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2020-0-01907, Development of Smart Signage Technology for Automatic Classification of Non-face-to-face Examination and Patient Status Based on AI, 50%) and by the GRRC program of Gyeonggi province [GRRC-Gachon2020 (B02), AI-based Medical Information Analysis, 50%].

References

1. Hicks S, Jha D, Thambawita V, Riegler M, Halvorsen P, Singstad B, Gaur S, Pettersen K, Goodwin M, Parasa S, and Lange T de. MedAI: Transparency in Medical Image Segmentation. Nordic Machine Intelligence 2021. DOI: 10.5617/nmi.9140
2. Poudel S and Lee SW. Deep multi-scale attentional features for medical image segmentation. Applied Soft Computing 2021; 109:107445
3. Yuan Y, Chao M, and Lo YC. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. IEEE transactions on medical imaging 2017; 36:1876–86
4. Amyar A, Modzelewski R, Li H, and Ruan S. Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation. Computers in Biology and Medicine 2020; 126:104037
5. Tan M and Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*. PMLR. 2019 :6105–14
6. Jha D, Smedsrud PH, Riegler MA, Halvorsen P, Lange T de, Johansen D, and Johansen HD. Kvasir-seg: A segmented polyp dataset. *International Conference on Multimedia Modeling*. Springer. 2020 :451–62
7. Jha D, Ali S, Emanuelsen K, Hicks SA, Thambawita V, Garcia-Ceja E, Riegler MA, Lange T de, Schmidt PT, Johansen HD, et al. Kvasir-instrument: Diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy. *International Conference on Multimedia Modeling*. Springer. 2021 :218–29
8. Celebi ME, Codella N, and Halpern A. Dermoscopy image analysis: overview and future directions. IEEE journal of biomedical and health informatics 2019; 23:474–8
9. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, and Sánchez CI. A survey on deep learning in medical image analysis. Medical image analysis 2017; 42:60–88
10. Long J, Shelhamer E, and Darrell T. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015 :3431–40
11. Ronneberger O, Fischer P, and Brox T. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015 :234–41
12. Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, Han X, Chen YW, and Wu J. Unet 3+: A full-scale connected unet for medical image segmentation. *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020 :1055–9