

Employing GRU to combine feature maps in DeeplabV3 for a better segmentation model

Mahmood Haithami¹, Amr Ahmed¹, Iman Yi Liao¹, Hamid Jalab²

1. University of Nottingham Malaysia, 43500 Semenyih, Malaysia

2. Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia

Abstract

In this paper we aim to enhance the segmentation capabilities of DeeplabV3 by employing Gated Recurrent Neural Network (GRU). A 1-by-1 convolution in DeeplabV3 were replaced by GRU after the Atrous Spatial Pyramid Pooling (ASSP) layer to combine the input feature maps. The convolution and GRU have sharable parameters, though, the latter has gates that enable/disable the contribution of each input feature map. The experiments on unseen test sets demonstrate that employing GRU instead of convolution would produce better segmentation results. The used datasets are public datasets provided by MedAl competition.

Keywords: Segmentation; deep learning; GRU; Polyp; Instrument

Introduction

Colon polyp segmentation is considered a challenging task as a polyp could have various forms and does not have clear borders in some cases [1]. Furthermore, the lack of large and representative datasets in the endoscopy domain is a persistent challenge that still exists today. Different techniques have been proposed in the literature to address such challenges by applying image augmentations, transfer learning, and ensemble learning [2]. However, very few works have studied the application of Recurrent Neural Network (RNN) models [2, 3]. In this paper, we attempt to employ Gated Recurrent Neural Network (GRU) (i.e., a variant of RNN model) in DeeplabV3 [4] to enhance its segmentation capabilities.

Method

Figure 1 shows the difference between DeeplabV3 [4] and the proposed model. We employ the GRU instead of the 1-by-1 convolution to do the projection of the five feature maps produced by the Atrous Spatial Pyramid

Pooling layer (ASPP) [4], as illustrated in Figure 2. Mathematically, this mapping can be expressed as follows:

$$G: \mathbb{R}^{5C \times H \times W} \to \mathbb{R}^{C \times H \times W} \tag{1}$$

Where C, H, and W refers to number of channels, height, and width of the feature map. The motivation of using GRU is to utilize the weights sharing technique (as in the convolution method) meanwhile employing gates to regulate the flow of information across different feature maps. Hence, it provides a better method than the 1-by-1 convolution used in the original DeeplabV3 for combining/projecting the input feature maps into one feature map.

The experiments were conducted as follows. 1) All images were resized to 265-by-300 to reduce the computational complexity. 2) Each of the Instrument as well as Polyp datasets were shuffled and divided into training, validation, and testing subsets with a percentage of 80%, 10%, and 10%, respectively. 3) The evaluation metrics are mean Intersection-over-Union (mIOU), Dice, and Accuracy. The code for the proposed model can be found at:https://github.com/mss3331/ Proposed-model-for-MedAl21

Results

To demonstrate the performance of the proposed model, we employed two different datasets, i.e., Kvasir-Seg [5] and Kvasir-Instrument [6], for polyp segmentation and instrument segmentation, respectively [7]. Both datasets were provided in the MedAl competition as training datasets [7]. The performance of the proposed model versus the other state-of-the-art SegNet and DeeplabV3 are summarized in Table 1 and Table 2.

It is noticed that the proposed model performed better than SegNet and DeeplabV3 in all experiments except for the Instrument validation set. For the test sets,



b)Using a GRU instead of 1x1 Conv(proposed)

Figure 1: This figure contrasts the original DeeplabV3 with the proposed modification.a) refers to the original DeeplabV3, b) refers to the proposed modification. A GRU is added after the ASSP layer to capture the contextual information between the input feature maps. This figure is reworked from [4].



Figure 2: Proposed GRU model to combine/project 5 feature maps into one feature map. Meanwhile the original DeeplabV3 uses 1-by-1 convolution.

Table 1: Results on the Instrument dataset for both validation and test sets. The best results are highlighted. The reported results in this table based on the data splits explained in the Method section

Instrument database	mIOU		Dice		Accuracy	
	val	test	val	test	val	test
SegNet	91.4%	91.43 %	94%	94.62 %	98.5%	98.56%
DeepLabV3	94.9%	93.75%	97.1%	96.46%	99.2%	98.96%
Proposed	94.8%	94.28%	97%	96.85%	99.2%	99.01%

Table 2: Results on the Polyp dataset for both validation and test sets. The best results are highlighted. The reported results in this table based on the data splits explained in the Method section

Polyp database	mIOU		Dice		Accuracy	
	val	test	val	test	val	test
SegNet	78.8%	78.5%	85.2%	85%	93.8%	93%
DeepLabV3	88.5%	86%	92.9%	91%	96.7%	95.3%
Proposed	89.2%	87.2%	93.1%	91.7%	96.7%	95.9%

the proposed model achieved the best results for both Instrument and Polyp datasets which indicates that the proposed model can generalize better on unseen samples. For the test sets, the proposed model on average achieved a better Dice than DeeplabV3 (i.e., 0.545%). On the other hand, the Dice difference between the proposed model and DeeplabV3 on average is smaller on the validation set (i.e., Dice= 0.05%) than that on the test sets.

This is an indication that the added GRU model enhanced the capabilities of DeeplabV3. Even though both convolution and GRU have sharable parameters, the latter has gates which enable/disable the sequenced feature maps to contribute to the output. This capability of GRU enabled it to capture the contextual information within the feature maps.

Conclusion and future works

In this paper, we proposed an enhancement to the stateof-the-art DeeplabV3 by using GRU to combine the feature maps produced by ASSP layer instead of using 1-by-1 convolution. The motivation of using GRU is to capture the contextual information between the feature maps. The achieved results suggest that using a GRU as a projection function enhances the segmentation of DeeplabV3 on unseen test sets. Nevertheless, to further confirm this conclusion, re-sampling techniques such as cross-validation and bootstrap should be employed.

In the future, different types of RNN models would be used instead, such as LSTM [8]. Moreover, the recent Attention mechanism could be applied as well [9] to capture interrelationship between consecutive feature maps.

References

- Jia X, Mai X, Cui Y, Yuan Y, Xing X, Seo H, Xing L, and Meng MQH. Automatic polyp recognition in colonoscopy images using deep learning and two-stage pyramidal feature prediction. IEEE Transactions on Automation Science and Engineering 2020; 17:1570–84
- 2. Sánchez-Peralta LF, Bote-Curiel L, Picón A, Sánchez-Margallo FM, and Pagador JB. Deep learning to find colorectal polyps in colonoscopy: A systematic literature review. Artificial intelligence in medicine 2020 :101923
- Haithami M, Ahmed A, Liao I, and Jalab H. An embedded recurrent neural network-based model for endoscopic semantic segmentation. CEUR Workshop Proceedings 2021; 2886:59–68
- 4. Chen LC, Papandreou G, Schroff F, and Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 2017
- Jha D, Smedsrud PH, Riegler MA, Halvorsen P, Lange T de, Johansen D, and Johansen HD. Kvasir-seg: A segmented polyp dataset. 2020 :451–62
- 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. 2021 :218–29
- 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
- 8. Hochreiter S and Schmidhuber J. Long short-term memory. Neural computation 1997; 9:1735–80
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, and Polosukhin I. Attention is all you need. 2017 :5998–6008