NORDIC MACHINE INTELLIGENCE

Nordic Machine Intelligence, MedAI 2021
https://doi.org/10.5617/nmi.9116

Dual Parallel Reverse Attention Edge Network : DPRA-EdgeNet

Debayan Bhattacharya[1,2], Christian Betz[1], Dennis Eggert[1], and Alexander Schlaefer[2]

[1]Klinik und Poliklinik für Hals-, Nasen- und Ohrenheilkunde, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany
[2]Institute of Medical Technology and Intelligent Systems, Hamburg University of Technology, Hamburg, Germany

Abstract
In this paper, we propose the Dual Parallel Reverse Attention Edge Network (DPRA-EdgeNet), an architecture that jointly learns to segment an object and its edges. We compare our model against three popular segmentation models and demonstrate that our model improves the segmentation accuracy on the Kvasir-SEG dataset and the Kvasir-Instrument dataset.

Keywords: artificial intelligence; polyp segmentation; deep learning;

Introduction

MedAI: Transparency in Medical Image Segmentation [1] is the first challenge that will be held at *Nordic AI meet*. It consists of three tasks: (i) segmentation of polyps, (ii) segmentation of surgical instruments and, (iii) a transparency task which necessitates the requirement for transparent research in Medical Artificial Intelligence.

To this end, we propose a novel architecture called Dual Parallel Reverse Attention Edge Net (DPRA-EdgeNet). Our architecture is built upon the state-of-the-art PraNet [2]. To that end, our contribution is three fold: (i) Instead of only learning the object (polyp/instrument) mask, we learn the edge and the object mask simultaneously, (ii) we borrow the Reverse Attention (RA) modules from PraNet and extend it to provide attention to edge features and, (iii) we build novel decoder blocks that use spatial and channel attention to combine edge and object features and refine them using RA estimates.

Methods

Figure 1 shows the schematic diagram of our DPRA-EdgeNet which utilizes HarDNet [3] as backbone, Receptive Field Blocks (RFB) [4], Cascaded Partial Decoder (CPD) [5], two cascades of RA blocks [2] and novel edge and object detection blocks.

Encoder Architecture

The HarDNet architecture improved on the original Densenet [6] by systematically reducing the number of short connections. We have used this architecture as our backbone because it has less inference time. Encoder features from four levels are passed onto the RFBs. We use RFBs as this module reduces the channel dimension while strengthening the deeper features.

The output of the four RFBs are simultaneously passed to two CPDs after going through the Residual Block (RB) and the Squeeze and Excitation Block (SE) [7]. The RBs translate the common RFB features into object and edge features. The two concurrent CPDs combine four levels of features and outputs an initial estimate of the object and edge map, respectively. The initial estimates of the CPDs are passed to two cascades of RA blocks. The RA blocks are responsible for pruning the object and the edge features, respectively. The two cascades of RA blocks progressively mine for object or edge features outside the previously estimated map with the help of the high level encoder features. Fan *et al.* [2] have a mathematical formulation of the same.

Decoder Architecture

The decoder architecture decodes the object and edge segmentation maps. The Edge Decoder Block uses the RA edge estimates as attention gates to refine the high-level edge features and preceding decoder block features. This is followed by a concatenation of two levels of features followed by a sequence of RB and SE. Formally,

$$X_i^{e*} = F_T(X_{i+1}^e * \sigma(\eta_{i+1}) + X_{i+1}^e) \odot (\phi_i * \sigma(\eta_i) + \phi_i) \quad (1)$$

$$X_i^e = F_{SE}(F_{RB}(F_{SE}(F_{RB}(X_i^{e*})))) \quad (2)$$

where X_i^e is the output edge features of the i-th Edge Decoder Block. X_i^{e*} denotes the intermediate edge features. η_i is features from i-th Edge Reverse Attention (ERA) Block. ϕ_i is the high level edge features as shown in Figure 1. the $\sigma(.)$ represents the sigmoid function while

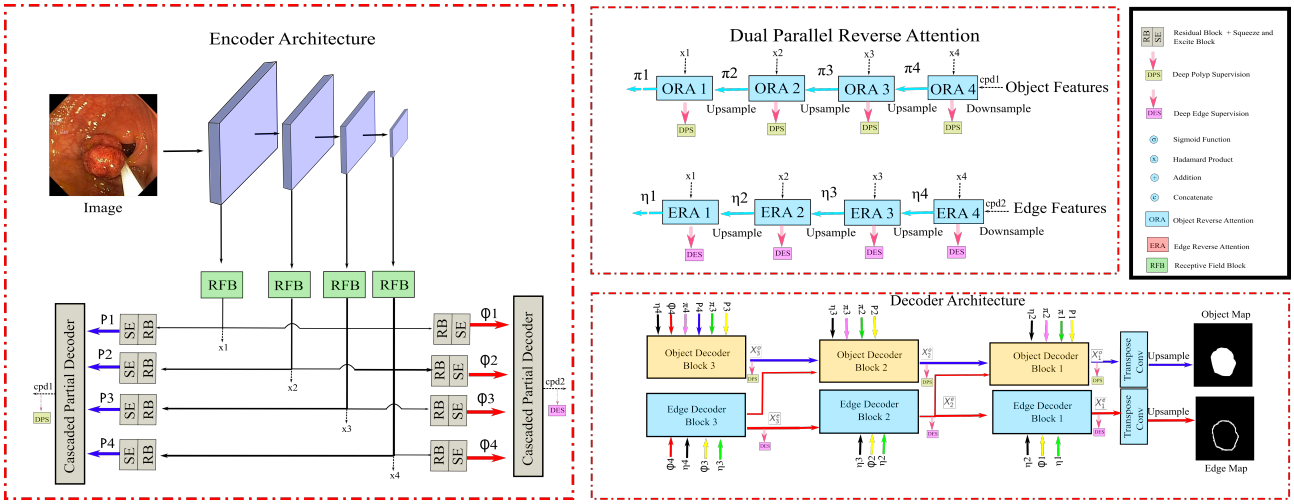

Figure 1: Schematic Diagram of DPRA-EdgeNet.

$\odot$ signifies the concatenation operation. $F_T(.)$, $F_{RB}(.)$ and $F_{SE}(.)$ denotes transpose convolution, RB and SE, respectively. For the sake of brevity, we do not show the parameters for the convolution and the SE. Finally, in the case of $i = 3$, ϕ_{i+1} is used instead of X^e_{i+1} in equation (1).

The i-th Object Decoder Block uses edge features from the $i-1$-th Edge Decoder Block and object features from $i-1$-th Object Decoder Block. Firstly, object and edge features are enhanced using corresponding map estimates from the Object Reverse Attention (ORA) and ERA blocks. This is followed by concatenation of the lower resolution object and edge features. The concatenated features are passed through a RB which provides spatial attention. The spatial attention is followed by SE which provides channel attention. Finally, the transformed lower resolution object features are passed through a transpose convolution block to match the resolution of the higher resolution object decoder features. The high and low resolution object decoder features are concatenated after which they are passed through series of RBs and SEs. Formally,

$$X_i^{o*} = F_{RB}((X^e_{i+1} * \sigma(\eta_{i+1}) + X^e_{i+1}) \odot (X^o_{i+1} * \sigma(\pi_{i+1}) + X^o_{i+1})) \quad (3)$$

$$X_i^{o**} = \sigma(X_i^{o*}) * X^o_{i+1} \quad (4)$$

$$X_i^{o***} = (P_i * \sigma(\pi_i) + P_i) \odot F_T(F_{SE}(X_i^{o**}) + X_i^{o*} + X^o_{i+1}) \quad (5)$$

$$X_i^o = F_{SE}(F_{RB}(F_{SE}(F_{RB}(X_i^{o***})))) \quad (6)$$

where X_i^{o*}, X_i^{o**} and X_i^{o***} are intermediate features for the i-th Object Decoder Block. X_i^o is the output of the i-th Object Decoder Block. P_i is the high level object feature as shown in Figure 1. We have used Deep Supervision[8] to learn intermediate edge and object concepts as it introduces regularization and promotes smooth flow of gradients[9].

Results and Discussion

We perform a five-fold cross validation experiments with the Kvasir-SEG [10] and Kvasir-Instruments [11] dataset. The dataset was split into 80:10:10 for training, validation and test split. We compare our DPRA-EdgeNet against U-Net [12], SegNet [13] and PraNet. The results of our experiments are shown in Table 1. Our experiments demonstrate that joint learning of edge and object segmentation in conjunction with dual reverse attention improves all the segmentation metrics. DPRA-EdgeNet focuses on boundary cues to segment objects more precisely using a decoder block which fuses segmentation and edge features from previous decoder blocks. Simultaneously, the decoder block takes advantage of the reverse attention blocks to prune the segmentation and the edge features.

Table 1: Results of five-fold cross validation experiments

Dataset	Model	DSC	mIoU	Precision	Recall
Kvasir-SEG	UNet	0.85 ± 0.01	0.77 ± 0.02	0.88 ± 0.01	0.86 ± 0.02
Kvasir-SEG	SegNet	0.88 ± 0.01	0.81 ± 0.01	0.90 ± 0.005	0.91 ± 0.04
Kvasir-SEG	PraNet	0.90 ± 0.005	0.85 ± 0.007	0.91 ± 0.008	0.92 ± 0.002
Kvasir-SEG	DPRA-EdgeNet	$\mathbf{0.92 \pm 0.003}$	$\mathbf{0.86 \pm 0.004}$	$\mathbf{0.93 \pm 0.007}$	$\mathbf{0.93 \pm 0.007}$
Kvasir-Instruments	U-Net	0.93 ± 0.002	0.89 ± 0.003	0.94 ± 0.02	0.93 ± 0.02
Kvasir-Instruments	SegNet	0.94 ± 0.003	0.89 ± 0.02	0.94 ± 0.002	0.94 ± 0.006
Kvasir-Instruments	PraNet	0.95 ± 0.002	0.92 ± 0.002	0.95 ± 0.001	0.96 ± 0.003
Kvasir-Instruments	DPRA-EdgeNet	$\mathbf{0.96 \pm 0.002}$	$\mathbf{0.93 \pm 0.003}$	$\mathbf{0.96 \pm 0.003}$	$\mathbf{0.96 \pm 0.003}$

Table 2: Results on the two test sets used in the competition

Dataset	Number of images	DSC	mIoU	Precision	Recall
Intruments	300	0.96	0.92	0.95	0.97
Polyps	300	0.89	0.83	0.91	0.90

In Table 2, we show the results of the DPRA-EdgeNet with the highest Dice Score on the two test sets used in the competition. We observe that the segmentation metrics on the polyp test set is lower that the instrument test set. We believe this to be the case for two reasons: (1) The heterogeneous appearance of the polyps makes the task of segmenting more challenging in comparison to segmenting instruments which have definitive appearance. (2) The polyp test set contains images of polyps from different datasets other than the Kvasir-SEG. As such, the test set distribution differs from the training set distribution. A future work would be to train models that learn domain invariant features.

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
2. Fan DP, Ji GP, Zhou T, Chen G, Fu H, Shen J, and Shao L. PraNet: Parallel Reverse Attention Network for Polyp Segmentation. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Ed. by Martel AL, Abolmaesumi P, Stoyanov D, Mateus D, Zuluaga MA, Zhou SK, Racoceanu D, and Joskowicz L. Cham: Springer International Publishing, 2020 :263–73
3. Chao P, Kao CY, Ruan YS, Huang CH, and Lin YL. HarDNet: A Low Memory Traffic Network. Available from: https://arxiv.org/pdf/1909.00948
4. Liu S, Di Huang, and Wang Y. Receptive Field Block Net for Accurate and Fast Object Detection. Available from: https://arxiv.org/pdf/1711.07767
5. Wu Z, Su L, and Huang Q. Cascaded Partial Decoder for Fast and Accurate Salient Object Detection. Available from: https://arxiv.org/pdf/1904.08739
6. Huang G, Liu Z, van der Maaten L, and Weinberger KQ. Densely Connected Convolutional Networks. Available from: https://arxiv.org/pdf/1608.06993
7. Hu J, Shen L, Albanie S, Sun G, and Wu E. Squeeze-and-Excitation Networks. Available from: https://arxiv.org/pdf/1709.01507
8. Wang L, Lee CY, Tu Z, and Lazebnik S. Training Deeper Convolutional Networks with Deep Supervision. Available from: https://arxiv.org/pdf/1505.02496
9. Lee CY, Xie S, Gallagher P, Zhang Z, and Tu Z. Deeply-Supervised Nets. Available from: https://arxiv.org/pdf/1409.5185
10. Jha D, Smedsrud PH, Riegler MA, Halvorsen P, Lange T de, Johansen D, and Johansen HD. Kvasir-SEG: A Segmented Polyp Dataset. Available from: https://arxiv.org/pdf/1911.07069
11. Jha D, Ali S, Emanuelsen K, Hicks SA, Thambawita V, Garcia-Ceja E, Riegler MA, Lange T de, Schmidt PT, Johansen HD, Johansen D, and Halvorsen P. Kvasir-Instrument: Diagnostic and Therapeutic Tool Segmentation Dataset in Gastrointestinal Endoscopy. *MultiMedia Modeling*. Ed. by Lokoč J, Skopal T, Schoeffmann K, Mezaris V, Li X, Vrochidis S, and Patras I. Cham: Springer International Publishing, 2021 :218–29
12. Ronneberger O, Fischer P, and Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Available from: https://arxiv.org/pdf/1505.04597
13. Badrinarayanan V, Kendall A, and Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. Available from: https://arxiv.org/pdf/1511.00561