Polyp Segmentation from Colonoscopy Images

Hui Che

2020.08.27

香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

Paper List

- Selective Feature Aggregation Network with Area-Boundary Constraints for Polyp Segmentation (MICCAI 2019)

- PraNet: Parallel Reverse Attention Network for Polyp Segmentation (MICCAI 2020)

Colorectal Polyps

Significance

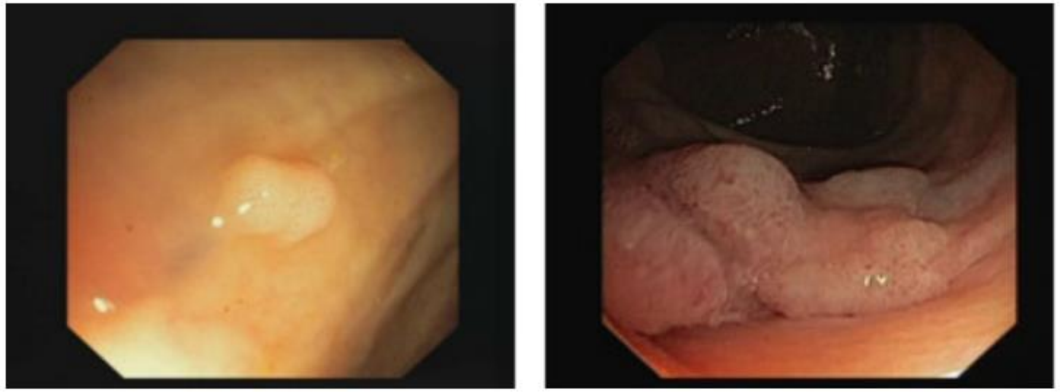
- Colorectal cancer is the third leading cause of cancer-related deaths.
- In clinical practice, segmenting polyps from colonoscopy images is of great importance since it provides **valuable information for diagnosis and surgery**.

Colorectal Polyps

Challenges

(i) the same type of polyps has a diversity of size, color and texture.

(ii) the boundary between a polyp and its surrounding mucosa is not sharp



Zhang Ruifei, et al. “Adaptive Context Selection for Polyp Segmentation." (2020).

Feature Aggregation & Area-Boundary Constraints

Selective Feature Aggregation Network with Area-Boundary Constraints for Polyp Segmentation

Yuqi Fang[1], Cheng Chen[1], Yixuan Yuan[2(✉)], and Kai-yu Tong[1(✉)]

[1] Department of Biomedical Engineering, The Chinese University of Hong Kong, Sha Tin, Hong Kong
kytong@cuhk.edu.hk

[2] Department of Electrical Engineering, City University of Hong Kong, Kowloon Tong, Hong Kong
yxyuan.ee@cityu.edu.hk

Feature Aggregation & Area-Boundary Constraints

Contributions

- Challenge 1 =>
  - Optimize the ***skip concatenations***.
  - Enrich the ***diversity of receptive fields*** at each layer.
- Challenge 2 =>
  - Utilize ***area and boundary information simultaneously*** in polyp segmentation.
  - ***Boundary-sensitive loss function***.

Feature Aggregation & Area-Boundary Constraints

Network Architecture

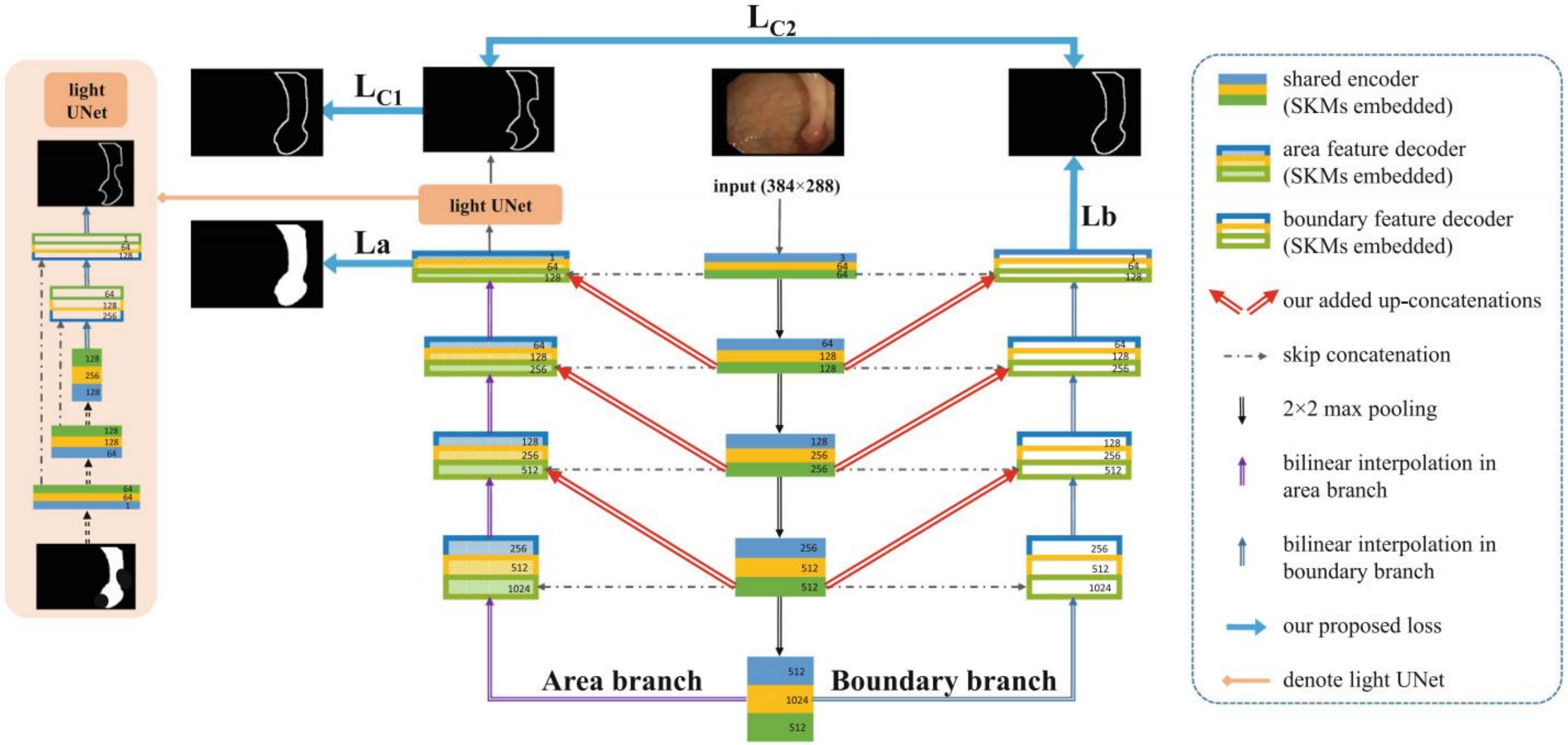
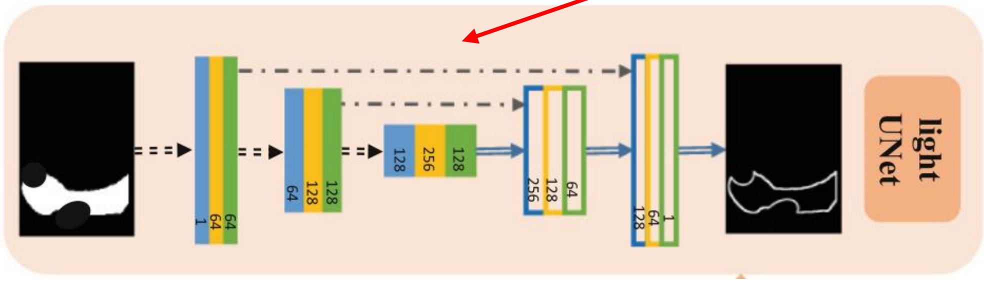


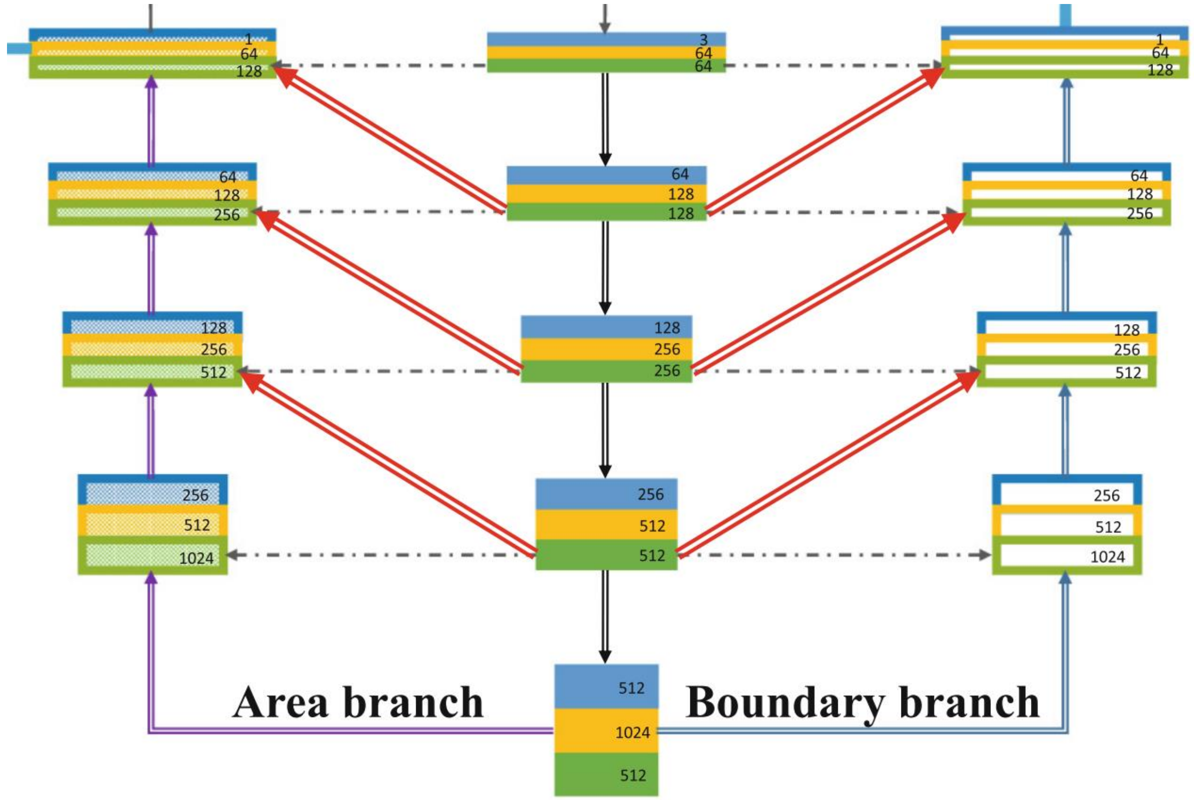
Fig. 1. Illustration of the selective feature aggregation network with area-boundary constraints. Numbers in each block represents the number of feature channels. (Color figure online)

Feature Aggregation & Area-Boundary Constraints

Selective Feature Aggregation

Skip concatenations in a parallel manner





Up-Concatenations

Selective Feature Aggregation

Selective Kernel Module (SKM)

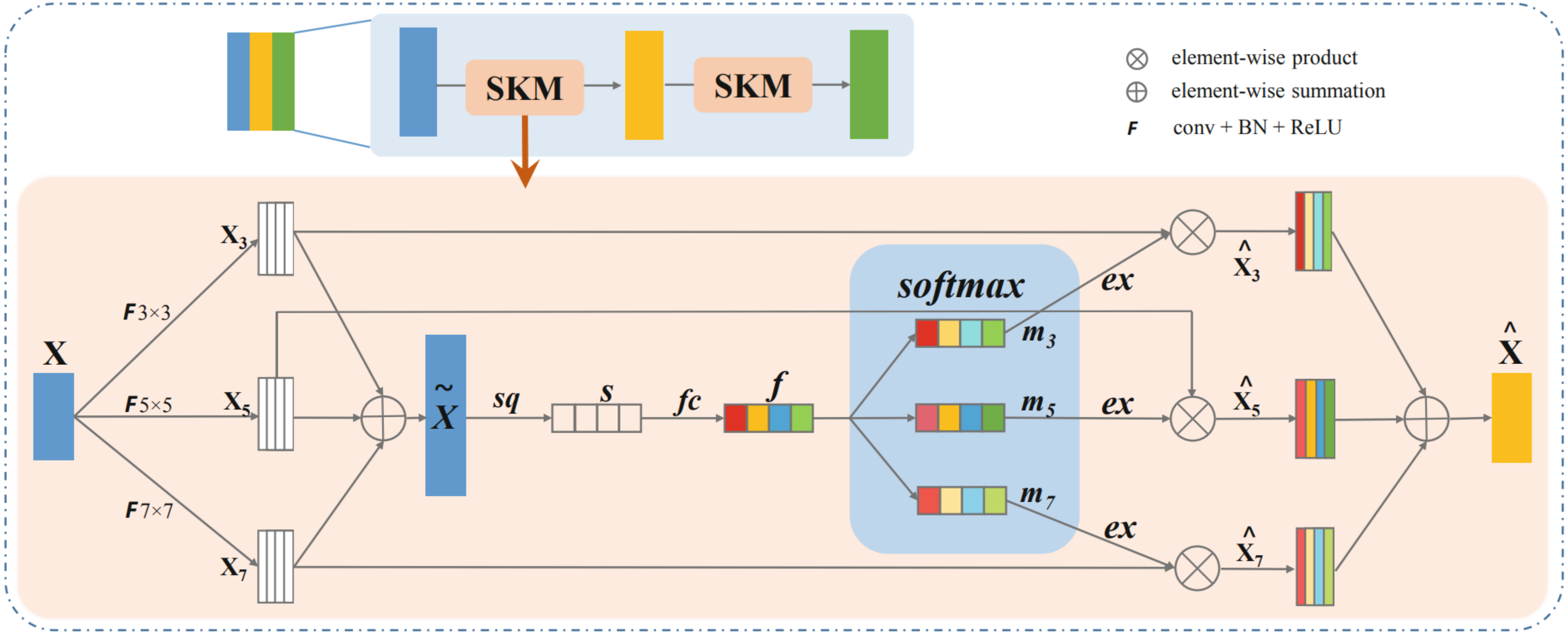


Fig. 2. Selective Kernel Module (SKM). sq: squeeze; fc: fully connected; ex: excitation.

Boundary-Sensitive Loss

Area Loss:
(area ground truth label)

$$L_a = -\sum_i z_i log(m_i) + (1 - \frac{2\sum_i m_i z_i + \varepsilon}{\sum_i m_i + \sum_i z_i + \varepsilon})$$

binary cross-entropy loss + *dice loss*

Boundary Loss:
(boundary ground truth label)

$$L_b = -\sum_i y_i log(p_i)$$

Area-Boundary Constraint Loss:

- aims to model the dependency between areas and boundaries.

Feature Aggregation & Area-Boundary Constraints

Boundary-Sensitive Loss

Area-Boundary Constraint Loss:

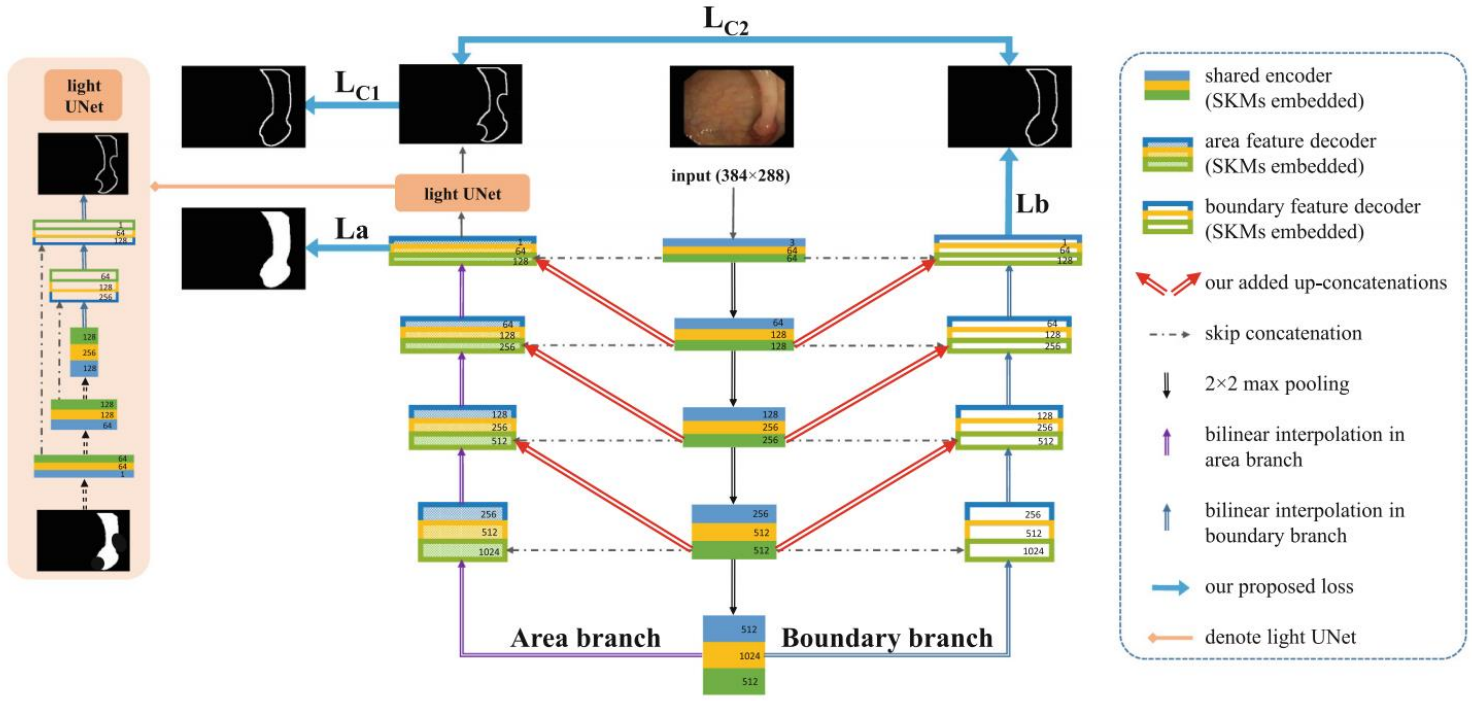


Fig. 1. Illustration of the selective feature aggregation network with area-boundary constraints. Numbers in each block represents the number of feature channels. (Color figure online)

$$L_{C1} = -\sum_i y_i log(q_i)$$

To minimize the difference between edge detector results and boundary ground truth.

Feature Aggregation & Area-Boundary Constraints

Boundary-Sensitive Loss

Area-Boundary Constraint Loss:

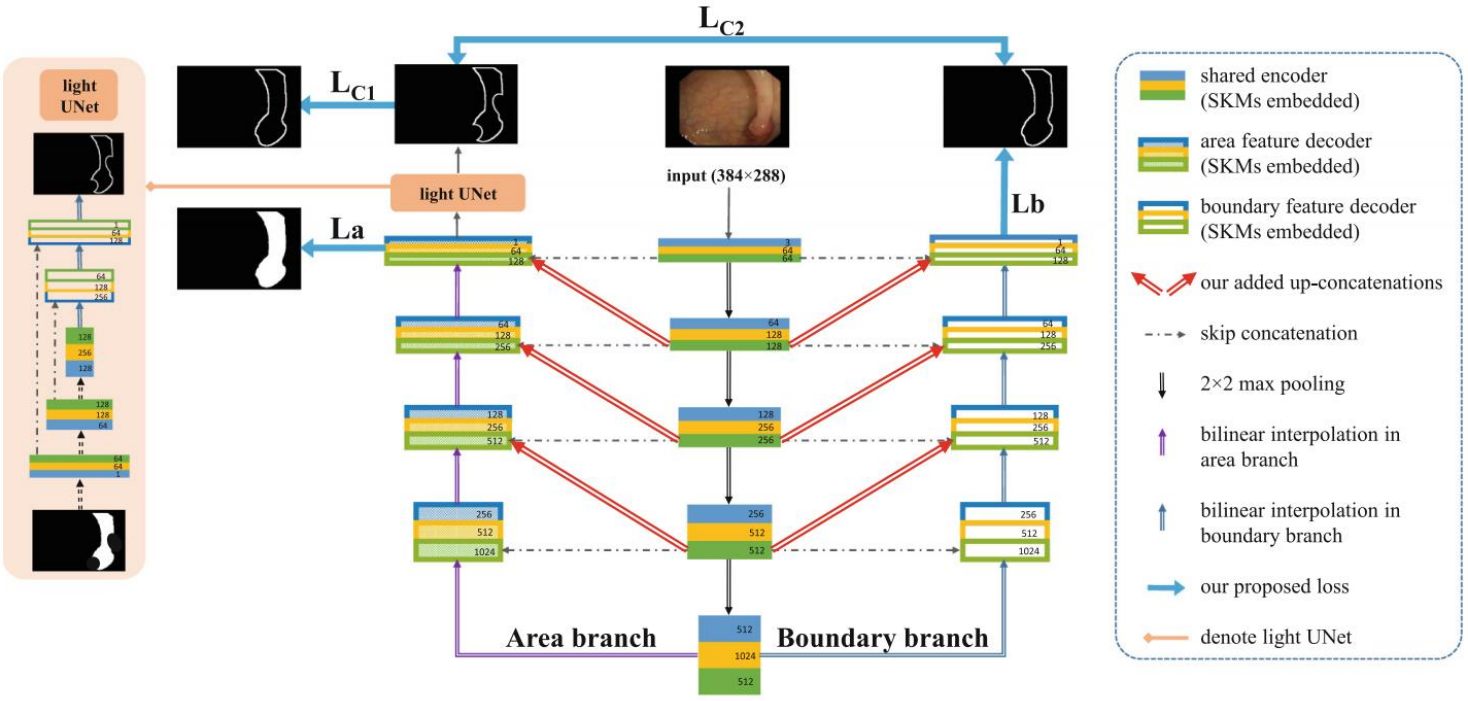


Fig. 1. Illustration of the selective feature aggregation network with area-boundary constraints. Numbers in each block represents the number of feature channels. (Color figure online)

$$L_{C2} = D_{KL}(P||Q) + D_{KL}(Q||P) = -\sum_i p_i log(\frac{q_i}{p_i}) - \sum_i q_i log(\frac{p_i}{q_i})$$

Kullback-leibler divergence

To minimize the difference between edge detector results and outputs of boundary branch.

Feature Aggregation & Area-Boundary Constraints

Boundary-Sensitive Loss

The Final Loss Function:

$$L_{total} = w_a L_a + w_b L_b + w_{C1} L_{C1} + w_{C2} L_{C2}$$

where w_a, w_b, and w_{C1} are set to 1, w_{C2} is set to 0.5 by empirical studies.

Feature Aggregation & Area-Boundary Constraints

Comparative Experiments

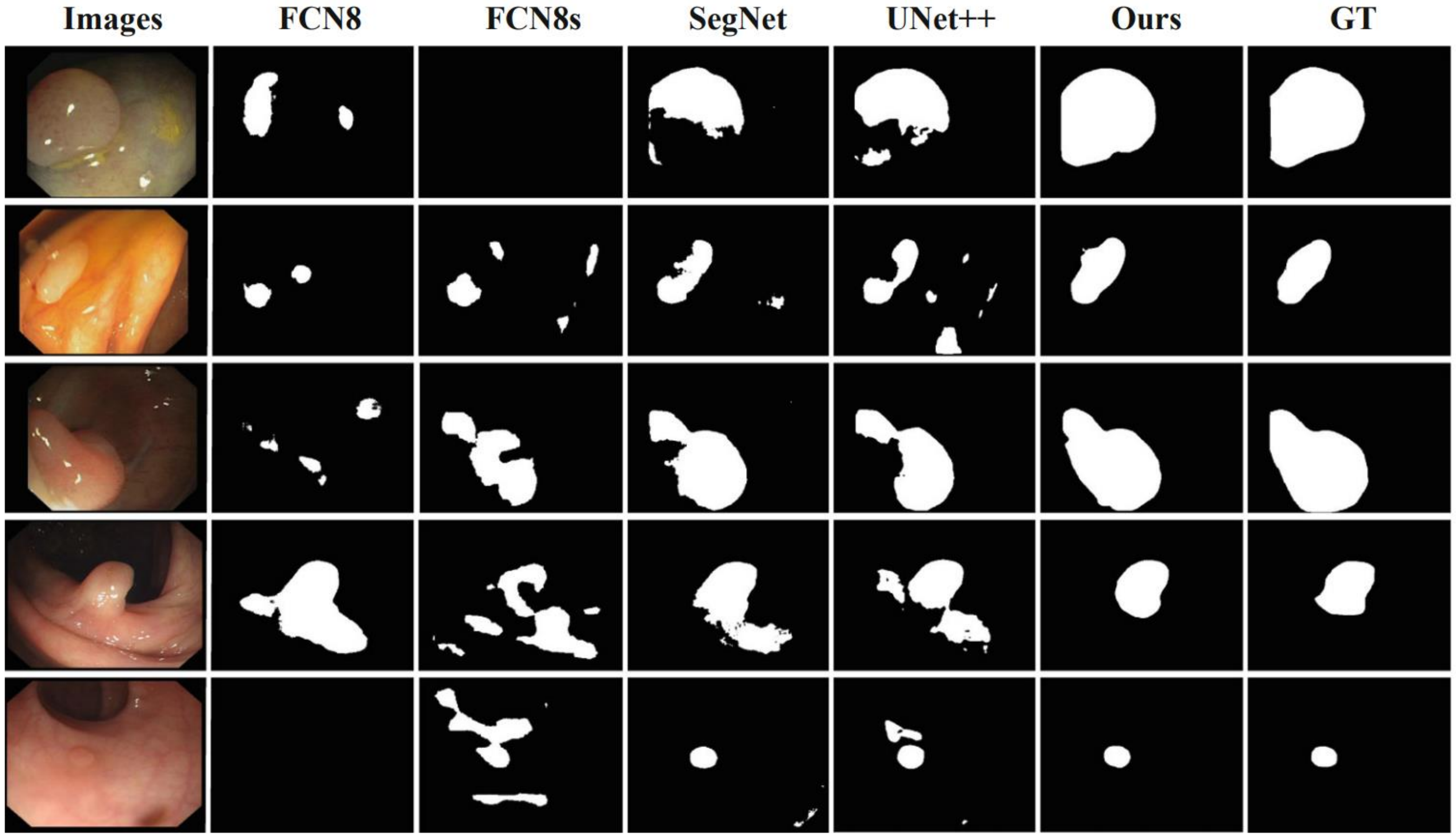


Fig. 3. Polyp segmentation results of different methods.

Comparative Experiments

Table 1. Comparison with different baselines and other state-of-the-art methods. 'UNet': the typical UNet with area branch; 'Up': up-concatenations; 'SKM': selective kernel module; 'bd': two-branch model with L_a and L_b; 'L_{C1}': the first constraint loss; 'Ours': the two-branch model with total loss, i.e. L_a, L_b, L_{C1} and L_{C2}.

Methods	*Rec*	*Spec*	*Prec*	*Dice*	IoU_P	IoU_B	*mIoU*	*Acc*
FCN8 [3]	53.38	98.83	78.11	55.70	45.48	93.69	69.59	93.97
FCN8s [2]	62.81	98.08	72.60	58.52	45.99	93.22	69.61	93.56
SegNet [12]	81.70	99.03	85.12	79.29	70.33	95.71	83.02	95.99
UNet++ [13]	80.68	99.24	85.31	78.55	69.83	95.71	82.77	95.97
Ours	**83.84**	99.43	90.19	**83.08**	**76.23**	**96.44**	**86.33**	**96.68**
UNet	82.29	99.08	86.13	80.45	72.10	95.84	83.97	96.12
UNet+Up	82.53	99.18	87.13	80.85	72.74	95.88	84.31	96.16
UNet+Up+SKM	82.27	99.37	88.71	81.51	74.41	96.18	85.30	96.41
UNet+Up+SKM+bd	83.30	99.44	90.13	82.61	75.84	96.39	86.12	96.62
UNet+Up+SKM+bd+L_{C1}	83.51	**99.44**	**90.20**	82.97	76.16	96.44	86.30	96.68

PraNet

PraNet: Parallel Reverse Attention Network for Polyp Segmentation

Deng-Ping Fan[1], Ge-Peng Ji[2], Tao Zhou[1], Geng Chen[1],
Huazhu Fu[1] ✉, Jianbing Shen[1] ✉, and Ling Shao[3,1]

[1] Inception Institute of Artificial Intelligence, Abu Dhabi, UAE.
[2] School of Computer Science, Wuhan University, Hubei, China.
[3] Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE.
{huazhu.fu, jianbing.shen}@inceptioniai.org
https://github.com/DengPingFan/PraNet

PraNet

Motivation

- **The fact**: during polyp annotation, clinicians ***first roughly locate a polyp*** and then ***accurately extract its silhouette mask*** according to the ***local features***.

Strategy

- The ***area and boundary*** are two key characteristics that distinguish normal tissues and polyps.
- **First** predict ***coarse areas*** and **then** implicitly ***model the boundaries*** by means of reverse attention.

PraNet

Contributions

- Challenge 1 =>
  - Aggregate the features in high-level layers using a ***parallel partial decoder*** (PPD).
  - Generate a global map as the initial ***guidance area*** for the following components.
- Challenge 2 =>
  - Mine the ***boundary cues*** using the reverse attention (RA) module.
- Extra =>
  - ***Real-time*** segmentation efficiency (~**50fps**).

PraNet

Network Architecture

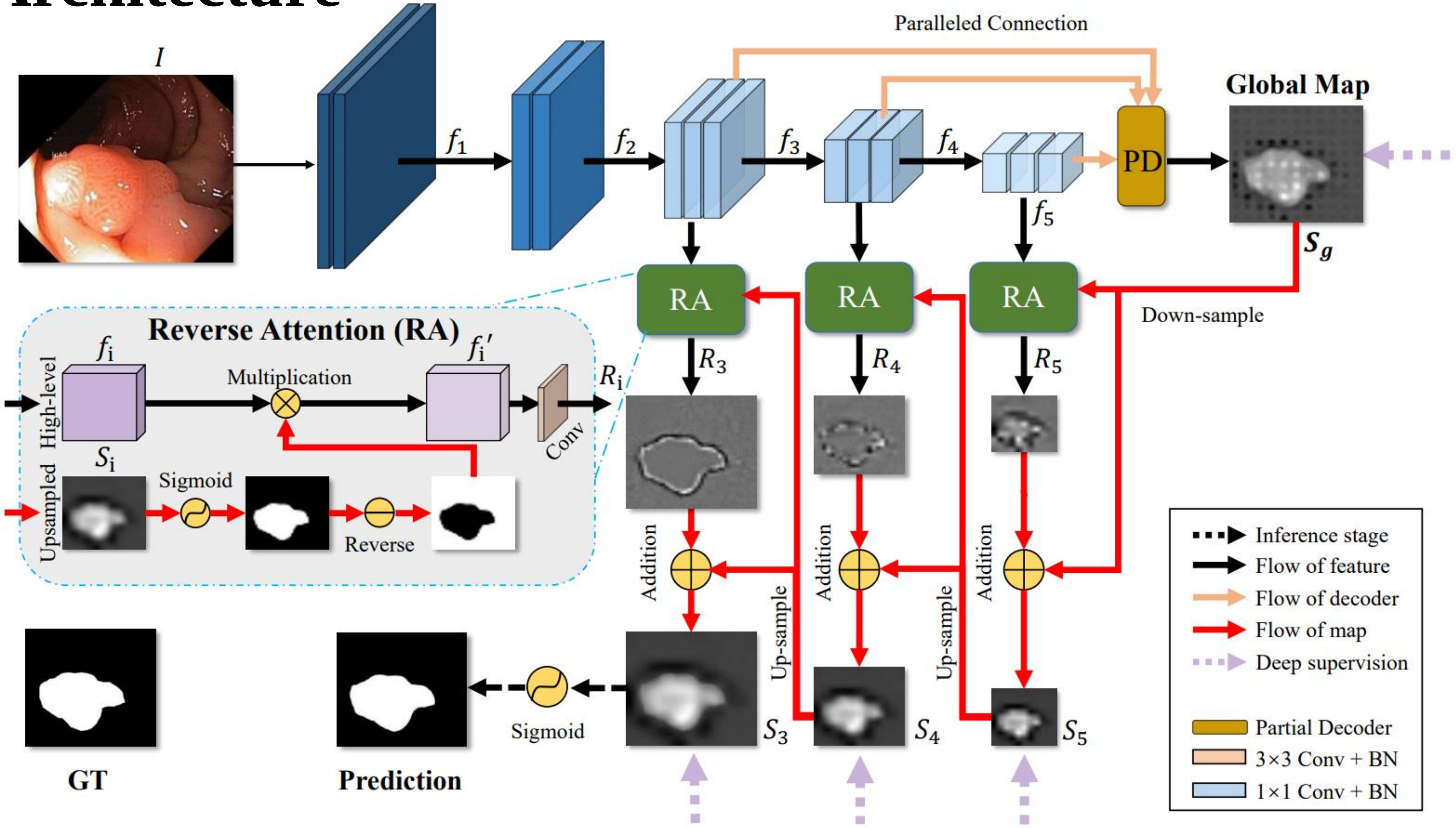
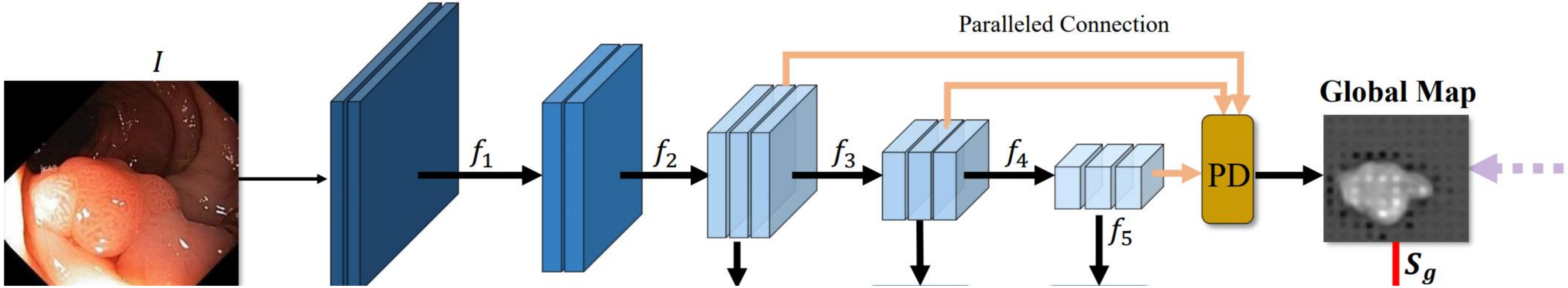


Fig. 1: Overview of the proposed *PraNet*, which consists of three reverse attention modules with a parallel partial decoder connection. See § 2 for details.

PraNet

Feature Aggregating via Parallel Partial Decoder



- ***Res2Net-based backbone network*** to extract five levels of features;
- ***Partial decoder*** $p_d(\cdot)$ to aggregate the high-level features with a paralleled connection.

  \- The partial decoder feature is computed by

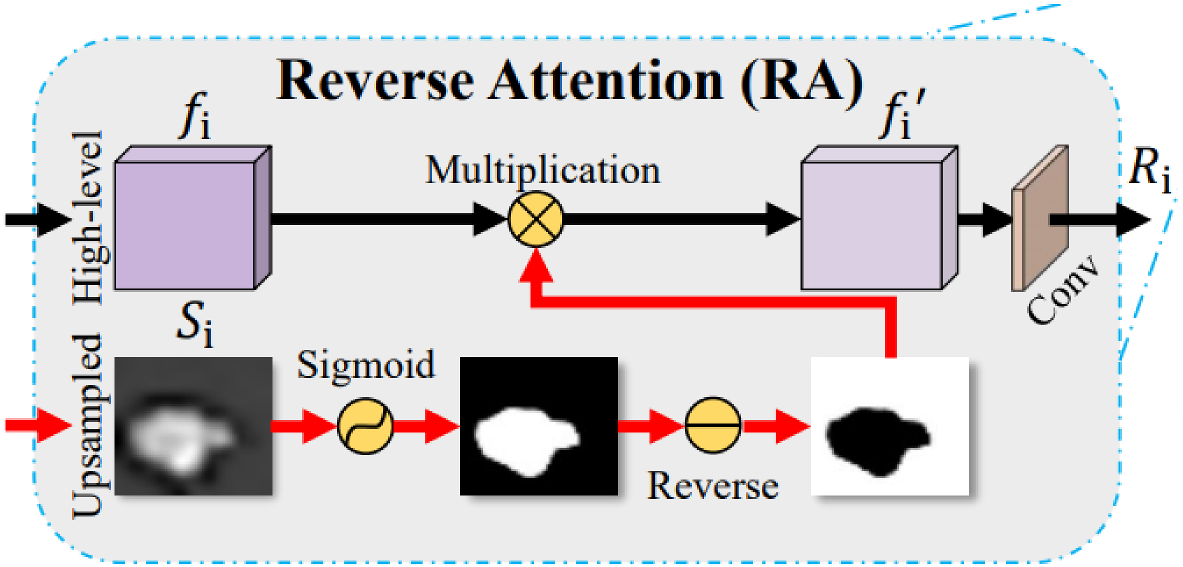
$$\boldsymbol{PD} = p_d(f_3, f_4, f_5)$$

  to obtain a global map S_g.

PraNet

Reverse Attention Module

An erasing foreground object manner



PraNet

Reverse Attention Module

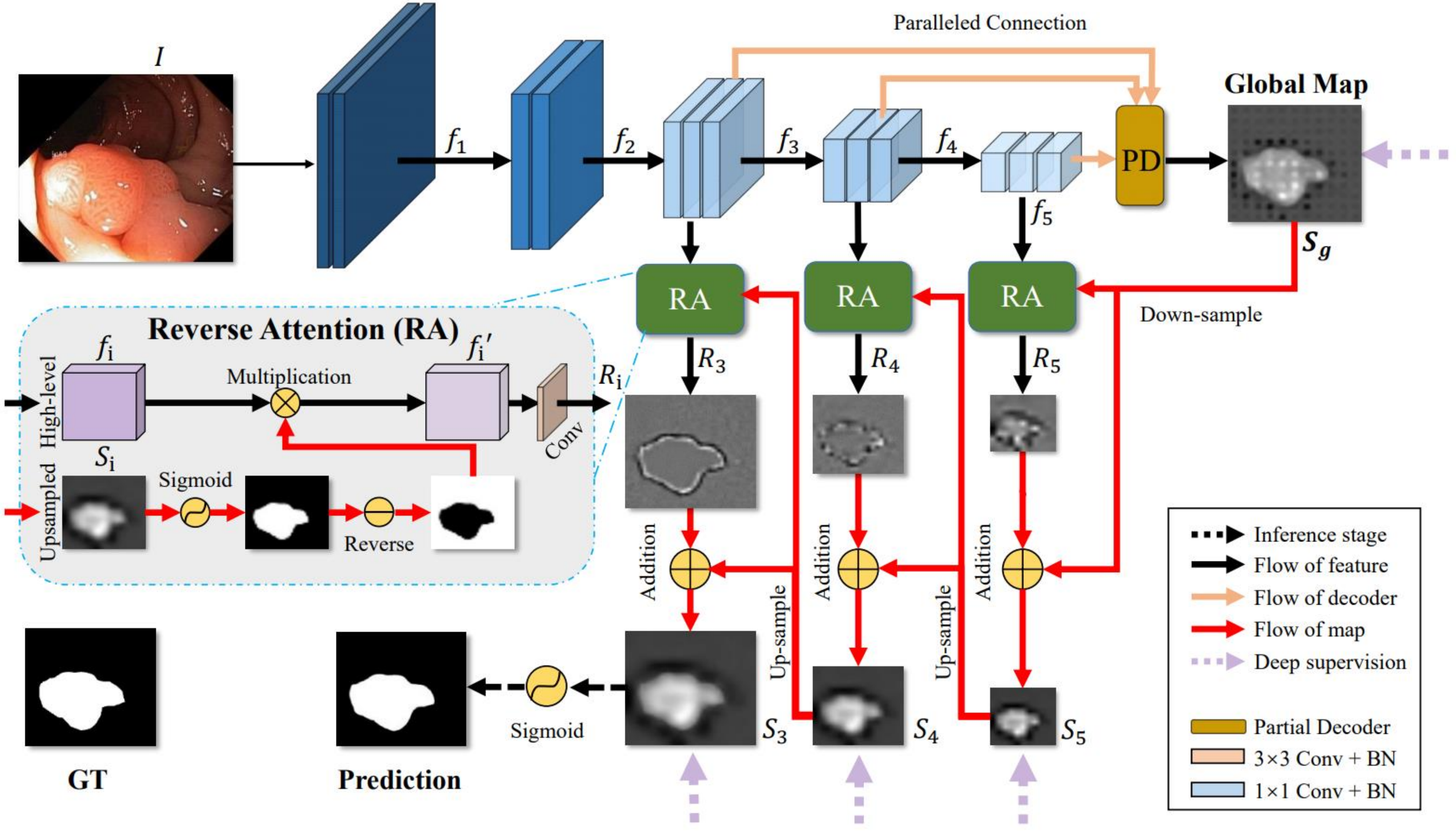


Fig. 1: Overview of the proposed *PraNet*, which consists of three reverse attention modules with a parallel partial decoder connection. See § 2 for details.

PraNet

- Loss Function: $$\mathcal{L} = \mathcal{L}_{IoU}^{w} + \mathcal{L}_{BCE}^{w}$$

the weighted IoU loss + the binary cross entropy (BCE) loss

- **Deep supervision** for the three side-outputs ($S_3, S_4\ and S_5$) and the global map S_g .

The Final Loss Function:

$$\mathcal{L}_{total} = \mathcal{L}(G, S_g^{up}) + \sum_{i=3}^{i=5} \mathcal{L}(G, S_i^{up})$$

PraNet

Experiments – Learning Ability

Table 1: Quantitative results on Kvasir [15] and CVC-612 [2] datasets. 'n/a' denotes that the results are not available. '†' represents evaluation scores from [16].

	Methods	mean Dice	mean IoU	F_{β}^{w}	S_{α}	E_{ϕ}^{max}	MAE
Kvasir	U-Net (MICCAI'15) [22]	0.818	0.746	0.794	0.858	0.893	0.055
	U-Net++ (TMI'19) [39]	0.821	0.743	0.808	0.862	0.910	0.048
	ResUNet-mod† [35]	0.791	n/a	n/a	n/a	n/a	n/a
	ResUNet++† [16]	0.813	0.793	n/a	n/a	n/a	n/a
	SFA (MICCAI'19) [10]	0.723	0.611	0.670	0.782	0.849	0.075
	***PraNet* (Ours)**	**0.898**	**0.840**	**0.885**	**0.915**	**0.948**	**0.030**
CVC-612	U-Net (MICCAI'15) [22]	0.823	0.755	0.811	0.889	0.954	0.019
	U-Net++ (TMI'19) [39]	0.794	0.729	0.785	0.873	0.931	0.022
	ResUNet-mod† [35]	0.779	n/a	n/a	n/a	n/a	n/a
	ResUNet++† [16]	0.796	0.796	n/a	n/a	n/a	n/a
	SFA (MICCAI'19) [10]	0.700	0.607	0.647	0.793	0.885	0.042
	***PraNet* (Ours)**	**0.899**	**0.849**	**0.896**	**0.936**	**0.979**	**0.009**

PraNet

Experiments – Generalization Capability

Table 2: Quantitative results on CVC-ColonDB [24], ETIS [23], and `test set` (CVC-T) of EndoScene [25] datasets. SFA [10] results are generated using the released code.

	Methods	mean Dice	mean IoU	F_{β}^{w}	S_{α}	E_{ϕ}^{max}	MAE
ColonDB	U-Net(MICCAI'15) [22]	0.512	0.444	0.498	0.712	0.776	0.061
	U-Net++(TMI'19) [39]	0.483	0.410	0.467	0.691	0.760	0.064
	SFA (MICCAI'19) [10]	0.469	0.347	0.379	0.634	0.765	0.094
	***PraNet* (Ours)**	**0.709**	**0.640**	**0.696**	**0.819**	**0.869**	**0.045**
ETIS	U-Net (MICCAI'15) [22]	0.398	0.335	0.366	0.684	0.740	0.036
	U-Net++ (TMI'19) [39]	0.401	0.344	0.390	0.683	0.776	0.035
	SFA (MICCAI'19) [10]	0.297	0.217	0.231	0.557	0.633	0.109
	***PraNet* (Ours)**	**0.628**	**0.567**	**0.600**	**0.794**	**0.841**	**0.031**
CVC-T	U-Net (MICCAI'15) [22]	0.710	0.627	0.684	0.843	0.876	0.022
	U-Net++ (TMI'19) [39]	0.707	0.624	0.687	0.839	0.898	0.018
	SFA (MICCAI'19) [10]	0.467	0.329	0.341	0.640	0.817	0.065
	***PraNet* (Ours)**	**0.871**	**0.797**	**0.843**	**0.925**	**0.972**	**0.010**

Experiments – Training and Inference Analysis

Table 3: Training and inference analysis (same platform) on CVC-ClinicDB [2] dataset. We record the #epochs when the model converges. Lr = learning rate.

	Methods	Epoch	Lr	Training	Inference	mean Dice
CVC-612	U-Net (MICCAI'15) [22]	30	3e-4	∼40 minutes	∼8fps	0.823
	U-Net++ (TMI'19) [39]	30	3e-4	∼45 minutes	∼7fps	0.794
	SFA (MICCAI'19) [10]	500	1e-2	>20 hours	∼40fps	0.700
	***PraNet* (Ours)**	**20**	**1e-4**	**∼30 minutes**	**∼50fps**	**0.899**

PraNet

Experiments – Qualitative Results

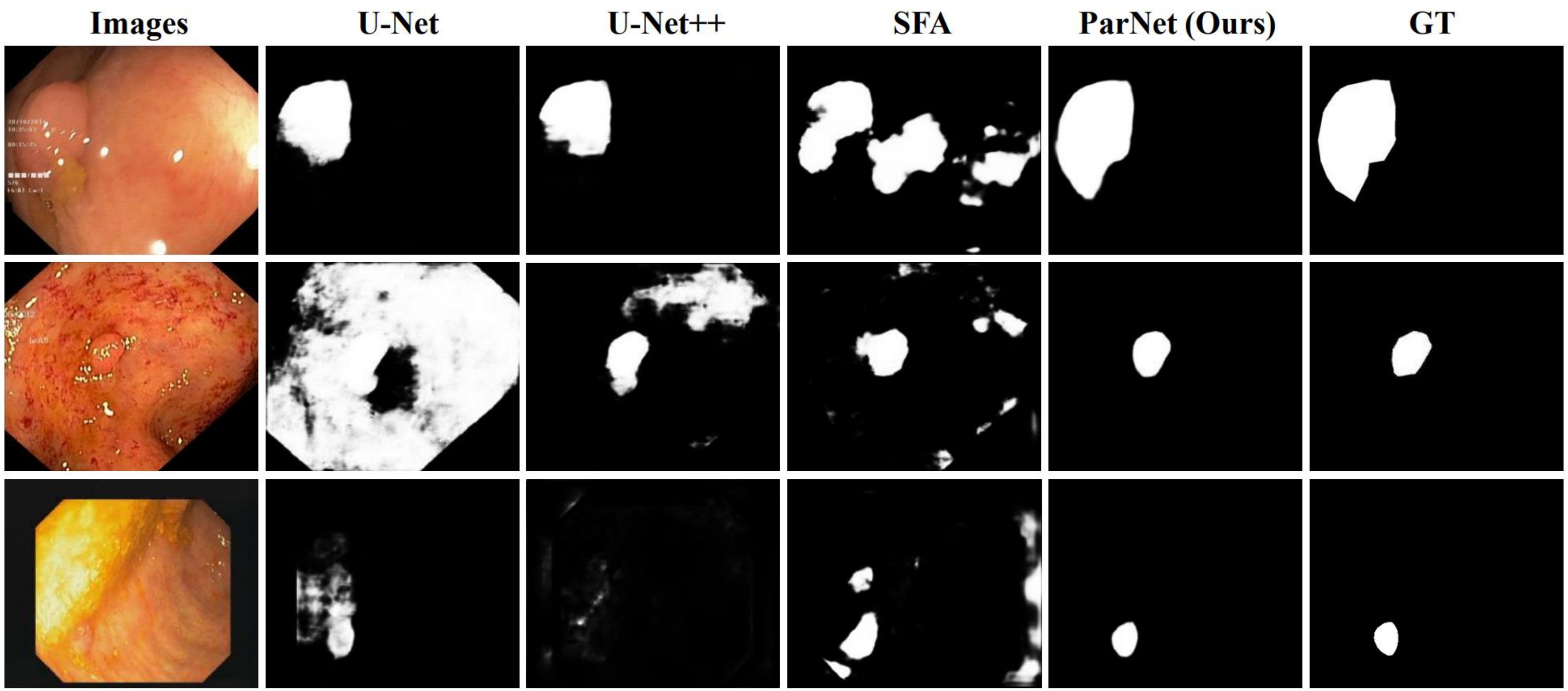


Fig. 2: Qualitative results of different methods.

PraNet

Ablation Study

Table 4: Ablation study for *PraNet* on the CVC-612 and CVC300 datasets.

Settings	CVC-612 (*seen*)			CVC300 (*unseen*)		
	mean Dice	mean IoU	S_α	mean Dice	mean IoU	S_α
Backbone (No.1)	0.747	0.668	0.735	0.726	0.631	0.670
PPD + Backbone (No.2)	0.865	0.798	0.902	0.824	0.734	0.893
RA + Backbone (No.3)	0.888	0.845	0.912	0.871	**0.800**	0.888
PPD + RA + Backbone (No.4)	**0.899**	**0.849**	**0.936**	**0.871**	0.797	**0.925**

- ***Effectiveness of PPD***
- ***Effectiveness of RA***
- ***Effectiveness of PPD & RA***

Summary

- Multi-scale information;
- Global and local information;
- Feature selection;
- Boundary and area constraint.

香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

Thanks for watching!

Hui Che

2020.08.27