SoSegFormer: A Cross-scale Feature Correlated Network for Small Medical Object Segmentation

Wei Dai[1,2], Zixuan Wu[1,2], Rui Liu[1,2], Junxian Zhou[1,2], Min Wang[1,2], Tianyi Wu[1,2], Jun Liu[1,2]

[1]Centre for Robotics and Automation, City University of Hong Kong
[2]Department of Mechanical Engineering, City University of Hong Kong

CityU 香港城市大學 City University of Hong Kong   ID: 257

Introduction

In the early stages of diseases, the affected regions, such as tumour or skin lesions, are relatively small. Early detection can aid in the discovery of potential diseases, thereby increasing patient survival rates. Cell-level imaging analysis is a cutting-edge topic with various clinical applications, such as human reproduction.

Moreover, a considerable number of medical images contain numerous lesions that occupy less than 10% of the total image area. A small medical object refers to an object that occupies a relatively small area in the image, presenting a significant challenge to deep learning methods. Convolution and pooling operations used in deep learning algorithms generate lower resolution of image features, leading to the loss of the morphology characteristics of medical objects. In medical applications, researchers have leveraged attention mechanisms through convolution or pooling to extract the global feature of polyp images [1, 2]. However, these methods employ relatively small feature maps (ranging from 11×11 to 44×44) to bridge area and boundary cues, which may not adequately capture the structural details of minuscule objects.

Most recently, the vision transformer (ViT) has been introduced to process sequences of image patches to learn the inter-patch representations, which has shown immense potential in aggregating and preserving the features of small objects [3].

Methodology

The small-object segmentation with transformer (SoSegFormer) network comprises two main components: cross-scale feature map instruction and the convolution with vision transformer. The cross-scale feature map instruction module is designed to enhance the performance of extracting the features of small medical objects (i.e., orange arrow in Fig. 1). To improve the identification of medical object features and their correlations, the correlated features are further processed by convolutions integrated with transformers (i.e., green arrow in Fig. 1).
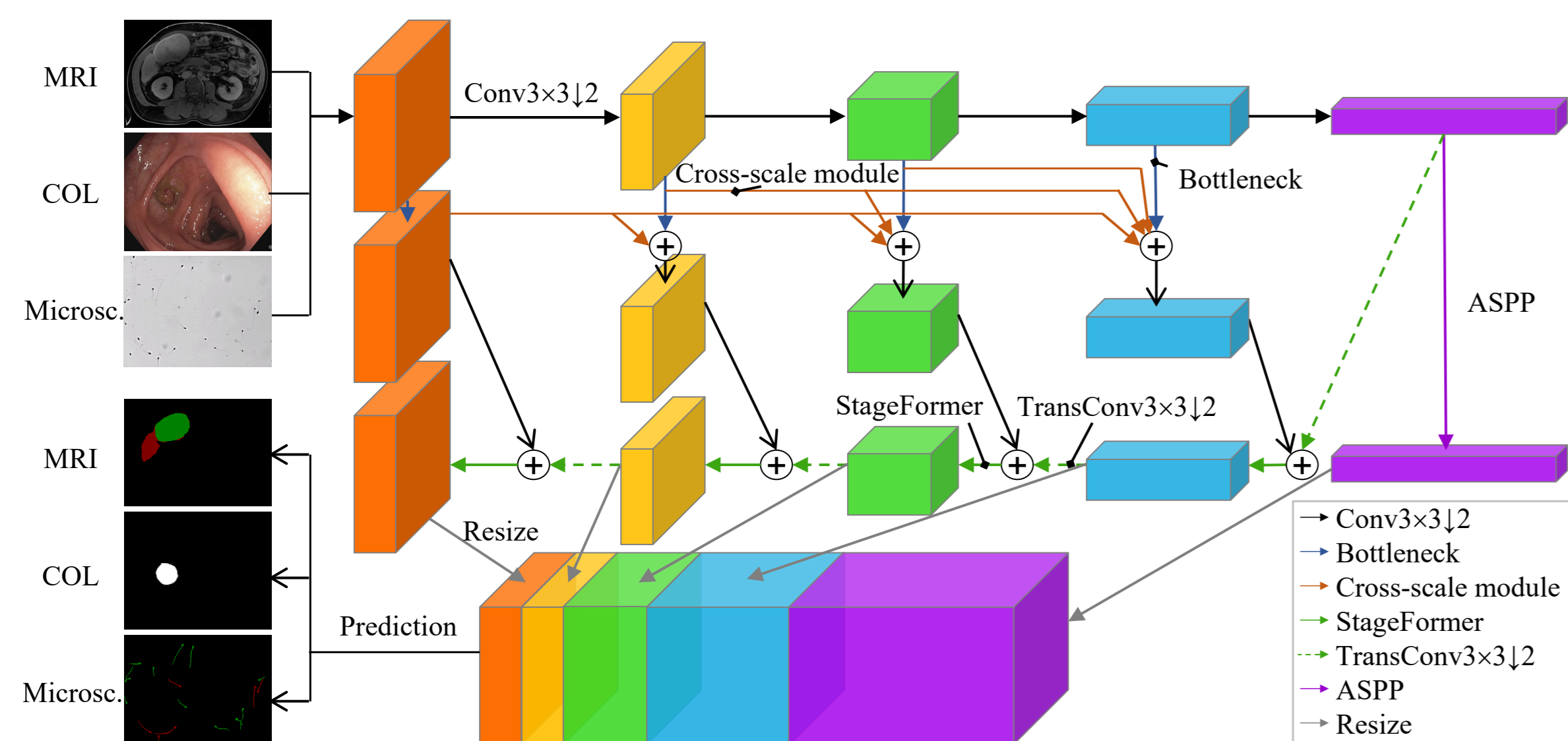

Figure 1: Architecture of small-object segmentation with transformer (SoSegFormer) network. MRI is magnetic resonance imaging; COL is colonoscopy; Miscosc. is microscopy imaging.

The proposed stage transformer (StageFormer) is based on the convolution-transformer hybrid structure, capable of learning the local and global representations of an input medical image simultaneously.
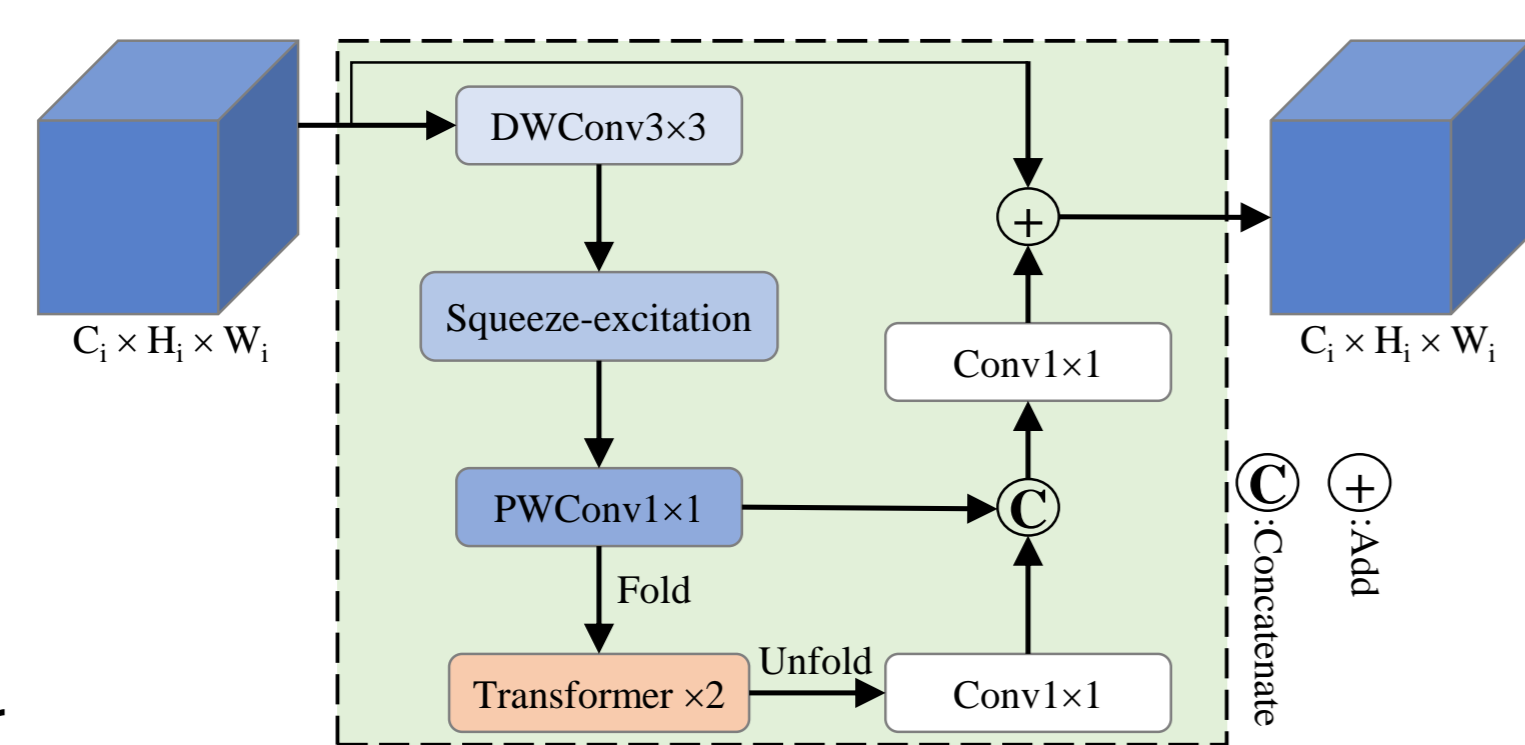

Figure 2: An illustration of stage transformer (StageFormer). StageFormer correlates local and global features using convolution with transformer operations.

Experimental Results

SoSegFormer outperforms other SOTA methods in all used metrics covering ultra-small, small, and all medical object segmentation in both datasets.

For instance, SoSegFormer achieved the highest segmention mDice with 87.02% & 80.91%, 87.73% & 90.98%, and 89.15% & 93.06% in ultra-small, small, and all medical object segmentation in ATLAS and PolypGen datasets.

Table 1: Quantitative results in the ATLAS and PolypGen datasets. Results vary across area ratios of medical objects below 1% (ultra-small), 10% (small), and 100% (all). The best results are highlighted.

	Methods	mIoU ultra small	mIoU small	mIoU all	mDice ultra small	mDice small	mDice all	MAE ($\times10^{-4}$) ultra small	MAE small	MAE all	Precision ultra small	Precision small	Precision all	F2 Score ultra small	F2 Score small	F2 Score all
ATLAS	U-Net (MICCAI'15) [4]	79.36	76.84	77.72	81.44	84.18	85.39	0.60	0.83	0.95	80.33	87.21	87.90	82.68	83.20	84.34
	U-Net++ (MICCAI'18) [5]	80.31	77.97	77.23	82.80	84.65	84.46	0.39	0.65	0.77	82.79	87.52	88.14	82.80	83.79	83.34
	PraNet (MICCAI'20) [1]	82.12	80.57	80.79	85.32	87.12	87.62	0.30	0.58	0.73	86.64	89.35	89.77	84.69	86.08	86.62
	HRNet (TPAMI'20) [6]	80.14	77.43	80.02	82.56	84.07	87.13	0.41	0.62	0.72	81.96	89.03	91.35	83.01	82.62	85.57
	CaraNet (JMI'23) [2]	82.23	78.00	79.93	85.45	85.31	87.29	0.27	0.74	0.84	87.82	87.43	88.88	84.48	84.31	86.60
	CFANet (PR'23) [7]	81.94	79.46	80.13	85.05	86.50	87.42	0.24	0.78	0.93	88.19	87.55	87.49	83.92	85.92	87.38
	SoSegFormer (Ours)	83.51	81.20	82.52	87.02	87.73	89.15	0.19	0.57	0.70	93.88	91.42	92.54	85.10	86.25	87.69
PolypGen	U-Net (MICCAI'15) [4]	70.92	74.94	76.60	73.22	83.50	85.45	2.51	1.20	1.59	71.96	84.99	90.96	74.97	82.67	82.98
	U-Net++ (MICCAI'18) [5]	72.42	77.85	80.21	75.71	85.93	88.15	2.53	1.05	1.35	73.51	86.78	91.86	78.96	85.44	86.31
	PraNet (MICCAI'20) [1]	75.50	83.23	86.36	79.96	90.00	92.30	1.45	0.72	0.90	77.21	92.44	94.94	82.83	88.68	90.91
	HRNet (TPAMI'20) [6]	68.17	76.32	81.59	69.82	84.73	89.15	7.37	1.25	1.31	69.94	82.67	90.32	70.80	86.12	88.49
	CaraNet (JMI'23) [2]	75.33	83.55	86.83	79.75	90.22	92.60	1.57	0.72	0.89	76.81	91.95	94.20	83.06	89.26	91.72
	CFANet (PR'23) [7]	74.15	83.87	86.82	78.21	90.45	92.60	2.06	0.69	0.88	75.28	92.48	94.77	82.14	89.33	91.43
	SoSegFormer (Ours)	76.22	84.63	87.56	80.91	90.98	93.06	1.37	0.67	0.83	77.92	91.79	94.80	84.04	90.50	92.11

Tab. 2 shows that the SoSegFormer network secured the first place in sperm segmentation on the SemSperm dataset, attaining 54.45% in mIoU, 65.17% in mDice, 4.25×10^{-4} in MAE, 68.25% in Precision, and 63.70% F2 Score.

Interestingly, the performance of SoSegFormer became notably superior in ultra-small medical object segmentation, surpassing other SOTA models by at least 4.29% in mIoU, 5.45% in mDice, 4.47% in Precision, and 5.12% in F2 Score. Conversely, those values are less than 2% in ATLAS and PolypGen datasets.

Table 2: Quantitative results in the SemSperm dataset. All sperms in SemSperm datasets occupy below 1% area in images. The best results are highlighted.

	Methods	mIoU	mDice	MAE*	Precision	F2 Score
SemSperm	U-Net (MICCAI'15) [4]	45.85	53.26	4.89	63.52	51.87
	U-Net++ (MICCAI'18) [5]	47.89	56.05	5.11	64.74	55.50
	PraNet (MICCAI'20) [1]	45.56	53.87	4.78	62.27	51.02
	HRNet (TPAMI'20) [6]	50.16	59.72	4.84	62.90	58.58
	CaraNet (JMI'23) [2]	46.27	54.94	4.50	59.69	53.07
	CFANet (PR'23) [7]	42.65	59.62	5.69	52.23	48.47
	SoSegFormer (Ours)	54.45	65.17	4.25	68.26	63.70

*MAE unit: $\times10^{-4}$

Negative Case Analysis

The proposed SoSegFormer network can not only distinguish all tumours, livers, polyps, and sperms positions but preferably recover the morphologies of these medical objects in the image.

In contrast, other SOTA methods either mistakenly categorise tumour regions as liver regions [see pink dashed regions in Fig. 3 (c)(e)] or struggle to differentiate ultra-small polyps and tumours from the background [see Fig. 3 (b-g)]. Besides, other SOTA methods either missed the broken-line shape or misclassified the abnormal sperm head as the normal sperm head.
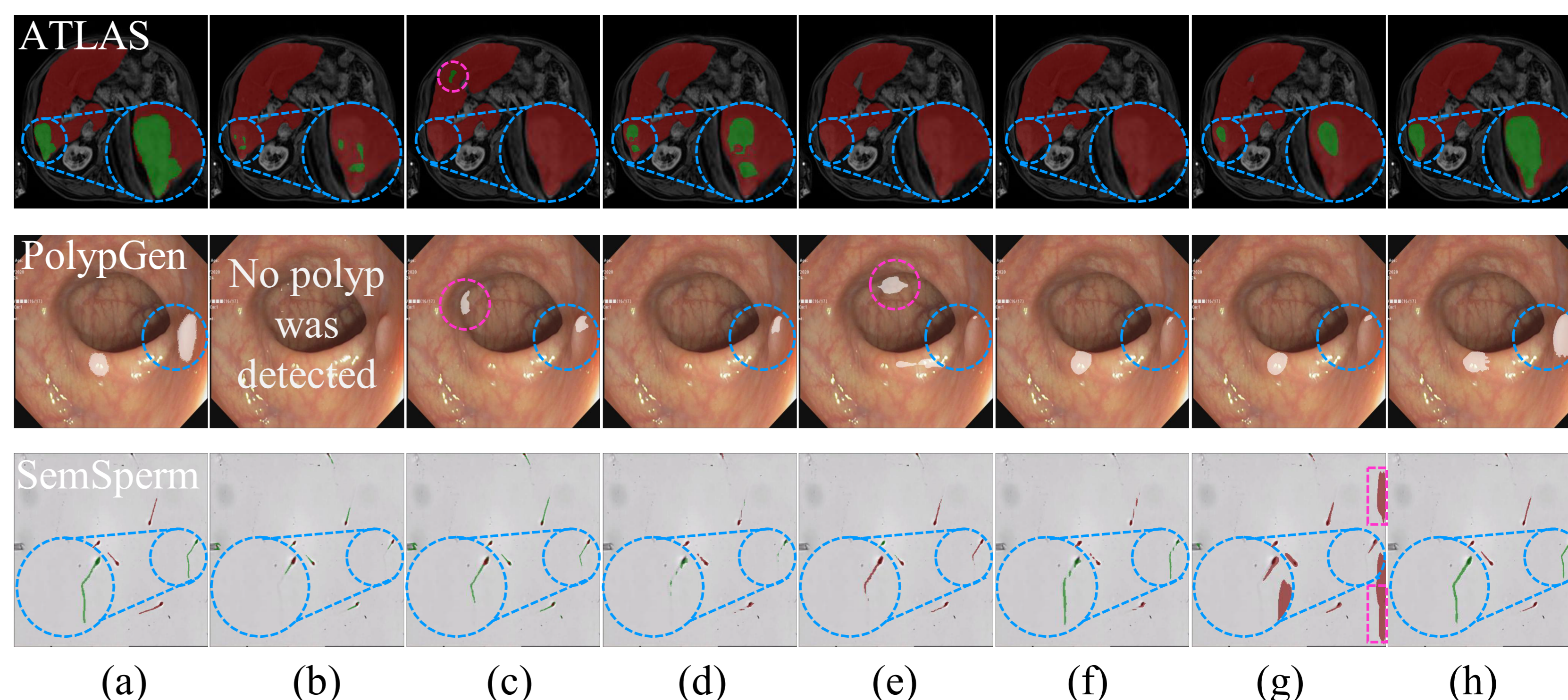

Figure 3: Visualisation of segmentation (a) ground truth and results using (b) U-Net, (c) U-Net++, (d) PraNet, (e) HRNet, (f) CaraNet, (g) CFANet, and (h) SoSegFormer (ours).

Resources

Reference Details.pdf

Download Paper.pdf