Efficient Mask Correction for Click-Based Interactive Image Segmentation

Fei Du, Jianlong Yuan, Zhibin Wang, Fan Wang
Alibaba Group

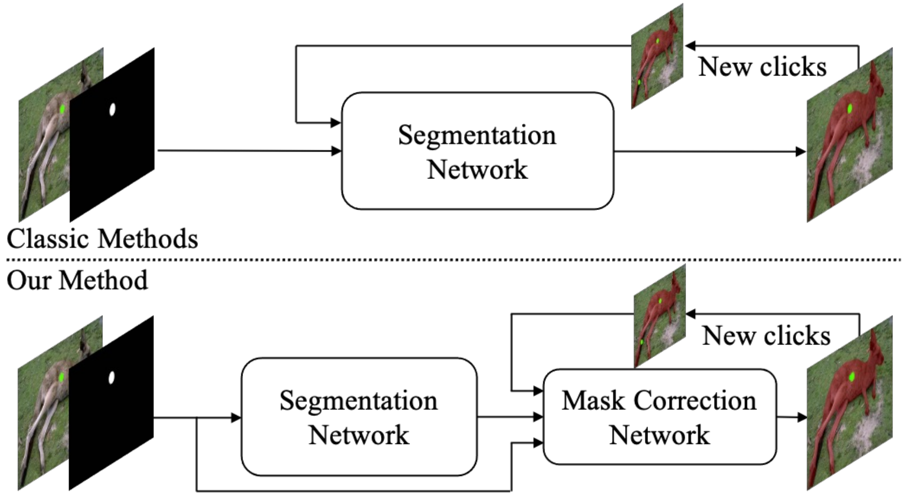
THU-PM-207

Preview

- **Interactive segmentation:**
  - Extract target masks with the input of positive/negative clicks.
  - An attractive way to simplify the mask annotation process.

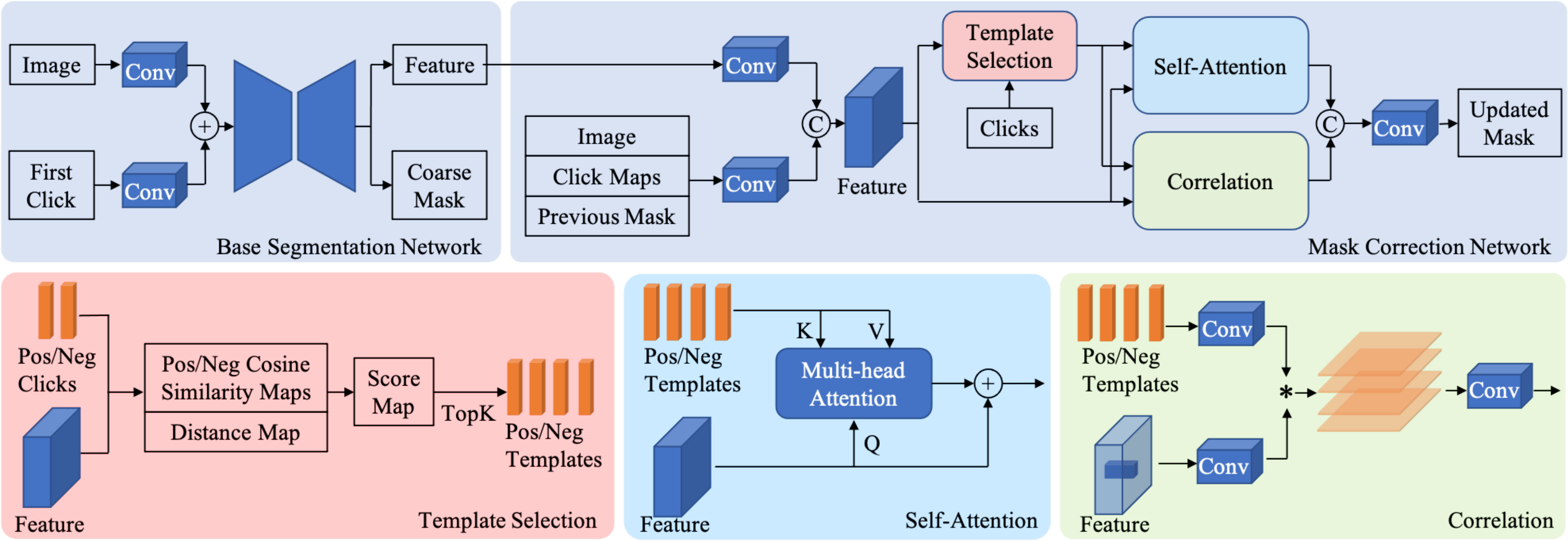
- **Classic Methods**:
  - Run a segmentation network to update the masks Iteratively.
  - Time-consuming especially when a strong segmentation network is used.

- **Proposed Method**:
  - Correct the masks via a lightweight mask correction network.
  - Improve the performance via two click-based feature augmentation modules.



Proposed method

- **Framework:**
  - A base segmentation network is used to extract target-aware features and predict the first coarse mask.
  - The mask correction network corrects the coarse mask and iteratively updates the mask with the input of user clicks.
  - Two click-guided feature augmentation modules are proposed to improve the network.



Experiments

- Speed on CPUs per click

Method	Params/MB	FLOPs/G	Speed/ms
CDNet-ResNet34$_{384}$ [7]	23.5	56.7	3339
f-BRS-ResNet50$_{400}$ [39]	31.4	84.6	2373
RITM-hrnet18s$_{400}$ [40]	4.22	8.96	634
RITM-hrnet18$_{400}$ [40]	10.0	15.4	1103
RITM-hrnet32$_{400}$ [40]	31.0	40.4	1635
FocalClick-hrnet18s$_{256}$ [8]	4.23	3.82	358
FocalClick-hrnet32$_{256}$ [8]	31.0	17.1	728
FocalClick-SegB0$_{256}$ [8]	3.74	1.94	207
FocalClick-SegB3$_{256}$ [8]	45.6	12.9	805
Ours-hrnet18s-FirstClick$_{384}$	4.33	9.35	752
Ours-hrnet18-FirstClick$_{384}$	10.3	15.3	1237
Ours-hrnet32-FirstClick$_{384}$	31.2	40.5	1745
Ours-SegB0-FirstClick$_{384}$	3.84	5.38	528
Ours-SegB3-FirstClick$_{384}$	45.9	32.3	2006
Ours-MaskCorrection-C64$_{384}$	0.11	1.09	183
Ours-MaskCorrection-C96$_{384}$	0.22	2.25	280

- Speed on GPUs per click (ms)

	HRNet18s	HRNet32	SegB0	SegB3
RITM	30 / 22	59 / 40	-	-
FocalClick	35 / 26	61 / 45	21 / 16	44 / 34
Ours-FirstClick	38 / 30	70 / 47	23 / 17	54 / 36
Ours-MaskCorrection	9 / 7	10 / 7	9 / 7	10 / 7

- Total evaluation time on SBD dataset

	NoC@90	Time@90	NoC@95	Time@95
FocalClick-hrnet18s	6.79	34min	12.78	62min
FocalClick-hrnet32	6.51	49min	12.50	85min
FocalClick-SegB0	6.86	23min	12.73	39min
FocalClick-SegB3	5.59	34min	11.55	63min
Ours-hrnet18s	6.16	19min	12.47	30min
Ours-hrnet32	5.65	21min	11.90	32min
Ours-SegB0	6.21	16min	12.45	27min
Ours-SegB3	5.57	20min	11.65	29min

Experiments

- Competitive performance on five benchmarks

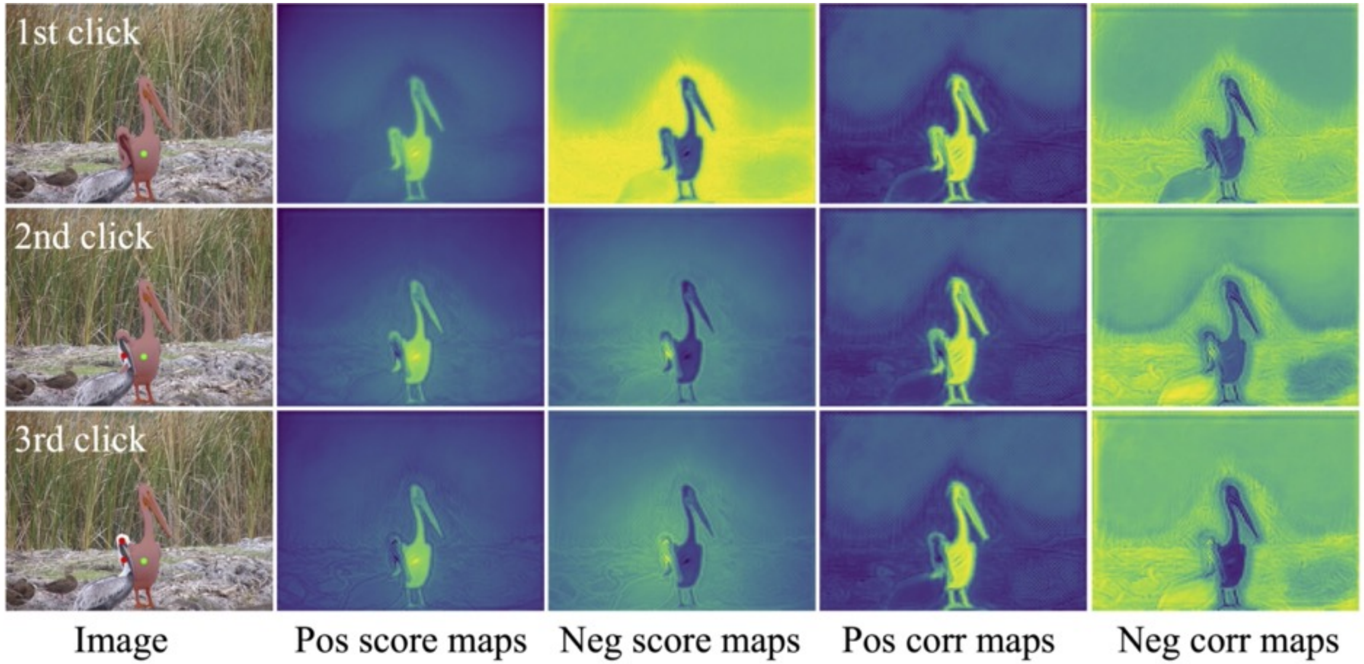
Methods	Train Data	GrabCut NoC@85	GrabCut NoC@90	Berkeley NoC@90	SBD NoC@85	SBD NoC@90	DAVIS NoC@85	DAVIS NoC@90	Pascal NoC@85
f-BRS-B-ResNet50 [39]	SBD	2.50	2.98	4.34	5.06	8.08	5.39	7.81	
CDNet-ResNet50 [7]	SBD	2.22	2.64	3.69	4.37	7.87	5.17	6.66	-
FocusCut-ResNet50 [26]	SBD	1.60	1.78	3.44	3.62	5.66	5.00	6.38	-
FocalClick-hrnet18s [8]	SBD	1.86	2.06	3.14	4.30	6.52	4.92	6.48	-
RITM-hrnet18 [40]	SBD	1.76	2.04	3.22	3.39	5.43	4.94	6.71	-
Ours-hrnet18s	SBD	1.82	1.92	3.26	3.58	5.79	5.23	6.88	2.47
Ours-hrnet18	SBD	1.74	1.84	3.03	3.38	5.51	5.05	6.71	2.37
TransClick-segformerB4 [11]	C+L	1.52	1.60	1.60	3.44	5.63	3.68	5.06	2.08
FocalClick-segformerB0 [8]	C+L	**1.40**	1.66	**2.27**	4.56	6.86	**4.04**	**5.49**	2.97
Ours-segformerB0	C+L	1.56	**1.64**	2.40	**3.95**	**6.21**	4.48	5.53	**2.65**
FocalClick-segformerB3 [8]	C+L	1.44	1.50	**1.92**	3.53	5.59	**3.61**	**4.90**	2.46
Ours-segformerB3	C+L	**1.42**	**1.48**	2.35	**3.44**	**5.57**	4.49	5.69	**2.23**
RITM-hrnet18s [40]	C+L	1.54	1.68	**2.60**	4.04	6.48	4.70	5.98	2.57
FocalClick-hrnet18s [8]	C+L	1.48	1.62	2.66	4.43	6.79	**3.90**	**5.25**	2.93
Ours-hrnet18s	C+L	**1.40**	**1.52**	2.68	**3.86**	**6.16**	4.42	5.66	**2.37**
RITM-hrnet18 [40]	C+L	1.42	1.54	**2.26**	3.80	6.06	4.36	5.74	**2.28**
EdgeFlow-hrnet18 [15]	C+L	1.60	1.72	2.40	-	-	4.54	5.77	2.50
Ours-hrnet18	C+L	**1.38**	**1.50**	2.30	**3.69**	**5.93**	**4.34**	**5.59**	2.37
RITM-hrnet32 [40]	C+L	1.46	1.56	2.10	3.59	5.71	4.11	5.34	2.57
FocalClick-hrnet32 [8]	C+L	1.64	1.80	2.36	4.24	6.51	**4.01**	5.39	2.80
PseudoClick-hrnet32 [29]	C+L	-	1.50	**2.08**	-	5.68	4.09	**5.27**	**1.94**
Ours-hrnet32	C+L	**1.30**	**1.42**	2.35	**3.55**	**5.65**	4.29	5.33	2.22

Experiments

- Ablation study
  - The first click helps extract target-aware features.
  - The correlation module helps learn target outlines.
  - The self-attention module propagates the click information.
  - The template selection module enriches click features.

1st Click	Corr	Self-Att	TS	DAVIS NoC@90	SBD NoC@90	Pascal NoC@90
				7.11	8.32	4.38
✓				6.98	6.83	3.19
✓	✓			6.32	6.30	3.09
✓		✓		6.02	6.33	2.98
✓	✓	✓		5.86	6.24	2.89
✓	✓	✓	✓	5.66	6.16	2.84

Table 4. Ablation study on three challenging benchmarks. HR-Net18s is used as the base segmentation network. *1st Click* denotes that we input the first click to the base segmentation network, *Corr* denotes the correlation module, *Self-Att* denotes the self-attention module, and *TS* denotes the template selection module.



	Negative Templates	DAVIS NoC@90	SBD NoC@90	Pascal NoC@90
Correlation Module		6.10	6.39	3.01
	✓	6.09	6.26	2.96
Self-Attention Module		6.04	6.55	3.22
	✓	5.92	6.29	2.99

Table 5. The performance of the correlation and self-attention modules with and without the adoption of the negative templates.

Conclusion

- Propose an efficient click-based interactive segmentation method.
- The method saves much inference time from the second click.
- Propose click-guided correlation and self-attention modules to exploit the click information to boost performance.
- Experimental results on five datasets show the effectiveness and efficiency of our method.
- Limitation: cannot save inference time of the first click.
- Code will be released at https://github.com/feiaxyt/EMC-Click

Thanks!