Poster: THU-AM-286

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Highlight

Interactive Segmentation as Gaussian Process Classification

Minghao Zhou [1,2], Hong Wang [2] (✉), Qian Zhao[1], Yuexiang Li [2], Yawen Huang [2], Deyu Meng[1,3,4], Yefeng Zheng [2]

[1]Xi’an Jiaotong University [2]Tencent Jarvis Lab [3]Pengcheng Laboratory

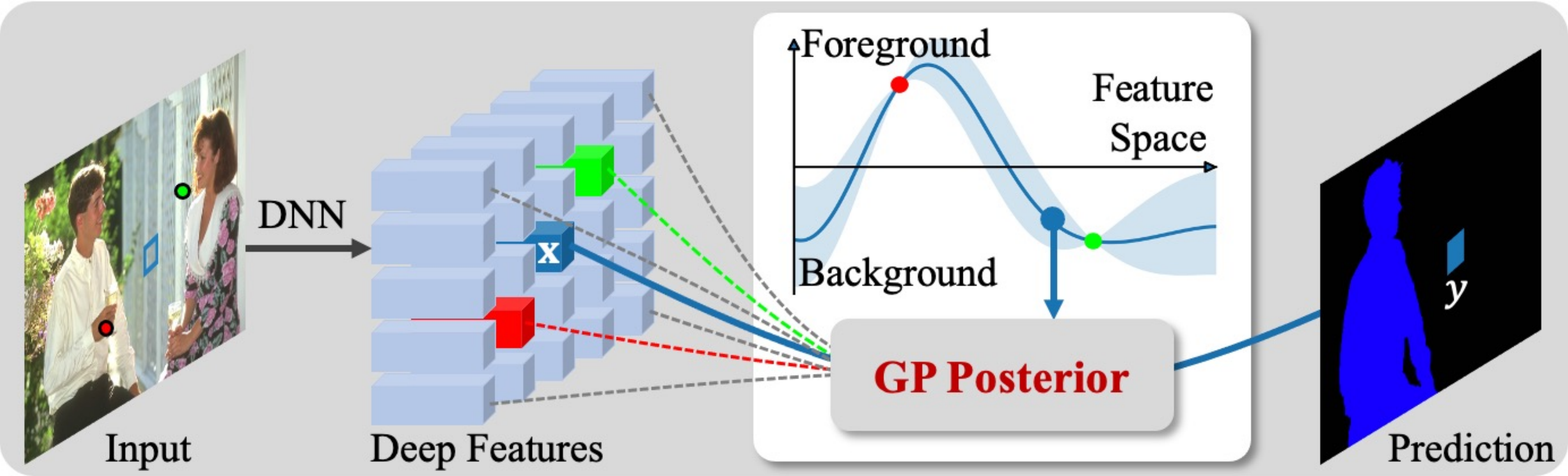
[4]Macau University of Science and Technology

Overview

Overview

Brand-new Perspective for Interactive Segmentation

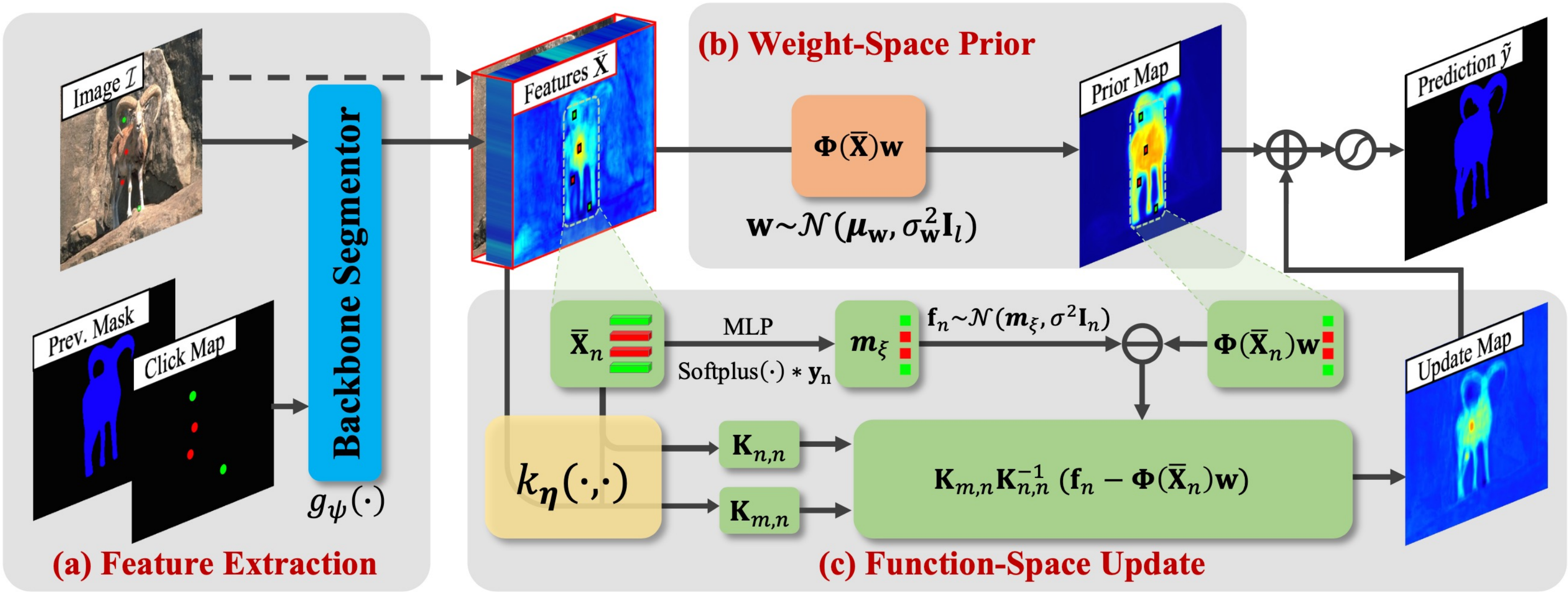
For each image, we model interactive segmentation as a Gaussian process classification task, with clicked/unclicked pixels as training/testing data.



Overview

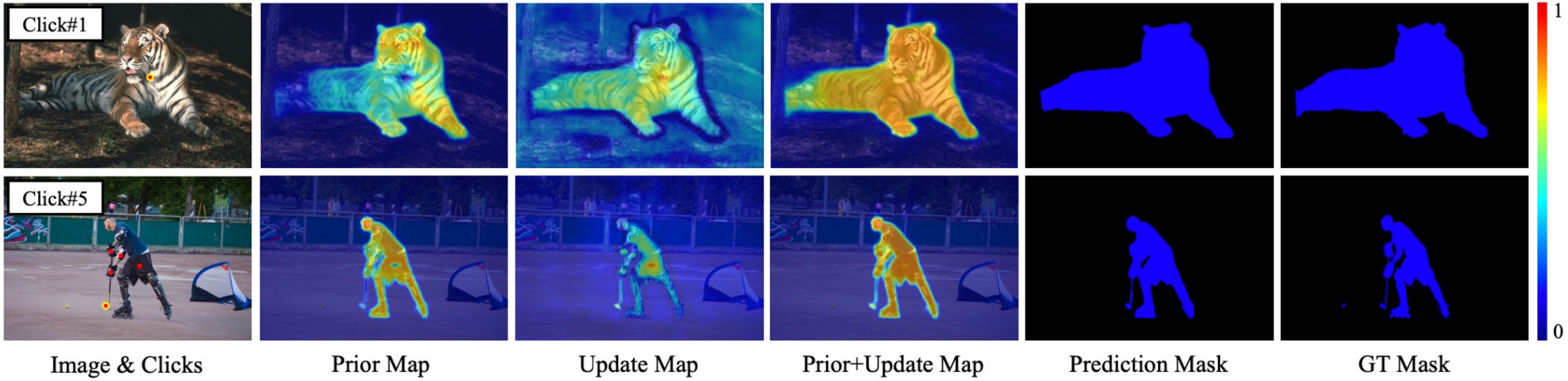
Gaussian Process Classification-based Interactive Segmentation (GPCIS) Model

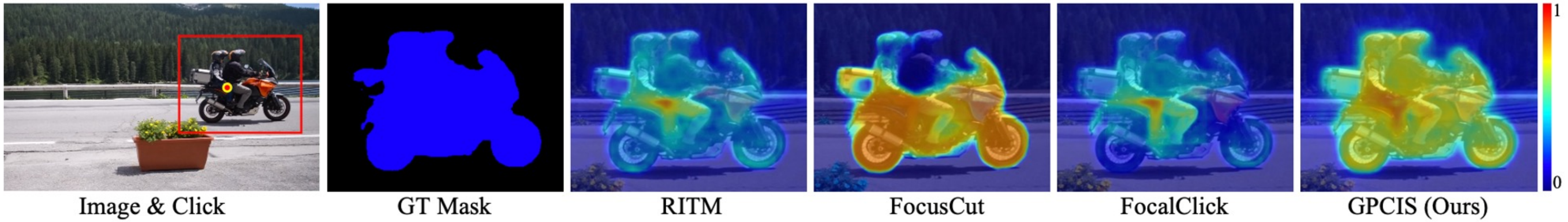
- Explicitly guide the information propagation procedure;
- Provide theoretical support for accurate predictions at clicks;
- Concise framework and clear working mechanism.



Overview

Experiments for Model Verification and Performance Evaluation

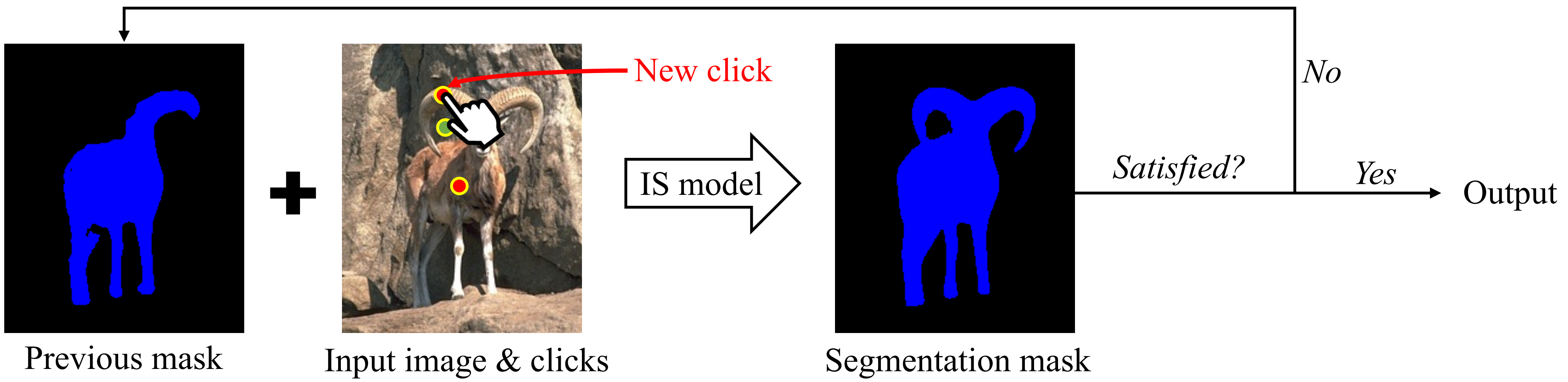




Introduction

Introduction

Interactive Segmentation (IS) Task

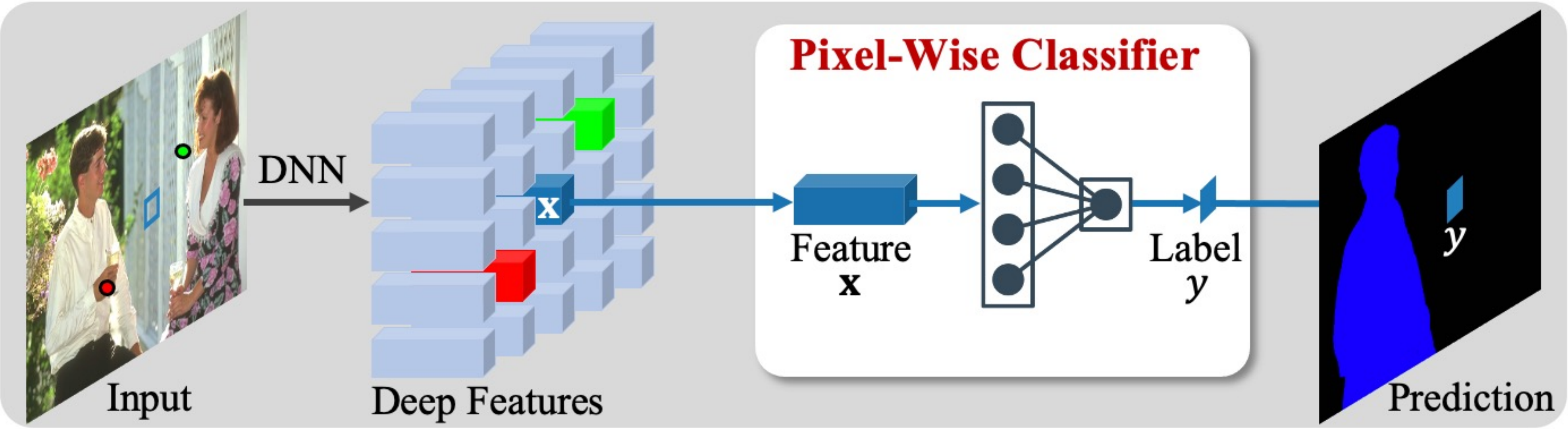


Introduction

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Current Deep Learning-based Interactive Segmentation Models

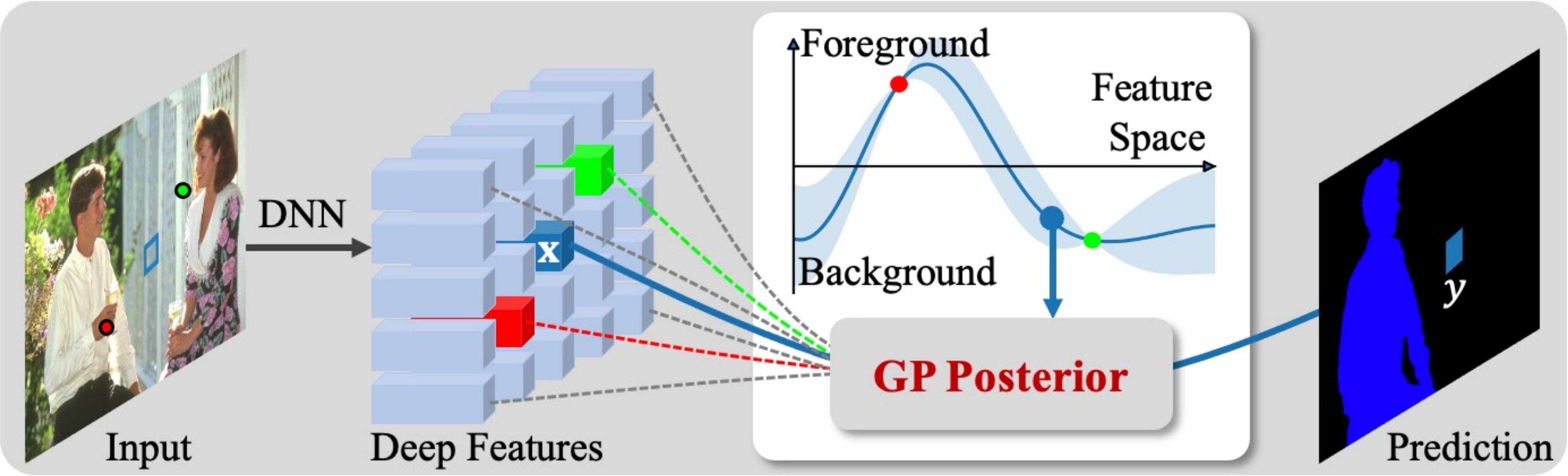
- Generally perform pixel-wise classification without specific designs;
- No explicit theoretical support that the clicked regions can be properly activated and correctly classified.



Introduction

Gaussian Process(GP)-based Framework for Interactive Segmentation

- Explicitly measure the relations between data points with a kernel function;
- Promote accurate predictions for training data;
- Can be flexibly integrated with deep networks via deep kernel learning.



Methodology

Methodology

Model Formulation

$\mathbf{X}_n/\mathbf{X}_*$: features of clicked/unclicked pixels;

$\boldsymbol{y}_n/\boldsymbol{y}_*$: labels of clicked/unclicked pixels;

$\mathbf{f}_*$: latent classification scores of unclicked pixels.

Goal: posterior probability of labels $\boldsymbol{y}_*$ at unclicked pixels

$$p(\mathbf{y}_*|\mathbf{X}_*,\mathbf{X}_n,\mathbf{y}_n) = \int \boxed{p(\mathbf{y}_*|\mathbf{f}_*)}\,\boxed{p(\mathbf{f}_*|\mathbf{X}_*,\mathbf{X}_n,\mathbf{y}_n)}\,\mathrm{d}\mathbf{f}_*$$

Monte Carlo Approximate

$p(\mathbf{y}_*|\mathbf{X}_*,\mathbf{X}_n,\mathbf{y}_n) \approx sig(\tilde{\mathbf{f}}_*)$

$$p(\mathbf{y}_*|\mathbf{f}_*) = \Pi_{c=1}^{*} sig(y_c f_c)$$

GP posterior

key step

Sample $\tilde{\mathbf{f}}_*$

Methodology

Overcome the Intractability of the GP Posterior

GP Posterior: $p(\mathbf{f}_*|\mathbf{X}_*,\mathbf{X}_n,\mathbf{y}_n)=\int \underbrace{\boxed{p(\mathbf{f}_*|\mathbf{X}_*,\mathbf{X}_n,\mathbf{f}_n)}}_{\text{Gaussian}}\boxed{p(\mathbf{f}_n|\mathbf{X}_n,\mathbf{y}_n)}\mathrm{d}\mathbf{f}_n$ intractable

$$p(\mathbf{f}_n|\mathbf{X}_n,\mathbf{y}_n) \propto \underbrace{\boxed{p(\mathbf{y}_n|\mathbf{X}_n,\mathbf{f}_n)}}_{\Pi_{c=1}^{n} s(y_c f_c)\ \text{non-Gaussian}}\underbrace{\boxed{p(\mathbf{f}_n|\mathbf{X}_n)}}_{\text{Gaussian}}$$

non-Gaussian

Approximate with $q(\mathbf{f}_n|\mathbf{X}_n,\mathbf{y}_n)=\mathcal{N}(\mathbf{m}_\xi(\mathbf{X}_n,\mathbf{y}_n),\sigma^2\mathbf{I}_n)$

where $\mathbf{m}_\xi(\mathbf{X}_n,\mathbf{y}_n)=\mathrm{Softplus}(\mathrm{MLP}_\xi(\mathbf{X}_n))*\mathbf{y}_n$

$$\min_q D_{KL}(q(\mathbf{f}_n|\mathbf{X}_n,\mathbf{y}_n)||p(\mathbf{f}_n|\mathbf{X}_n,\mathbf{y}_n))$$

Methodology

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Achieve Efficient Sampling of the GP Posterior

Approximated tractable GP posterior: $p(\mathbf{f}_*|\mathbf{X}_*,\mathbf{X}_n,\mathbf{y}_n) \sim \mathcal{N}(\boldsymbol{\mu}_{*|n}, \mathbf{K}_{*,*|n})$

Standard approach to draw samples

$$\tilde{\mathbf{f}}_* = \mu_{*|n} + \underbrace{\boxed{\mathbf{K}_{*,*|n}^{1/2}}}_{O(*^3)}\zeta \text{ with } \zeta \sim \mathcal{N}(0, \mathbf{I}_n)$$

We adopt a decoupled sampling framework[1] with the linear complexity $O(*)$ as:

$$\tilde{\mathbf{f}}_* = \underbrace{\mathbf{\Phi}(\mathbf{X}_*)\mathbf{w}}_{\text{weight-space prior}} + \underbrace{\mathbf{K}_{*,n}\mathbf{K}_{n,n}^{-1}(\mathbf{f}_n - \mathbf{\Phi}(\mathbf{X}_n)\mathbf{w})}_{\text{function-space update}},$$

[1] Wilson, et al. Efficiently sampling functions from Gaussian process posteriors. ICML. 2020

Methodology

Remark: Theoretical Support for Accurate Predictions at Clicks

$$\tilde{\mathbf{f}}_* = \underbrace{\mathbf{\Phi}(\mathbf{X}_*)\mathbf{w}}_{\text{weight-space prior}} + \underbrace{\mathbf{K}_{*,n}\mathbf{K}_{n,n}^{-1}(\mathbf{f}_n - \mathbf{\Phi}(\mathbf{X}_n)\mathbf{w})}_{\text{function-space update}},$$

Replacing $*$ with n

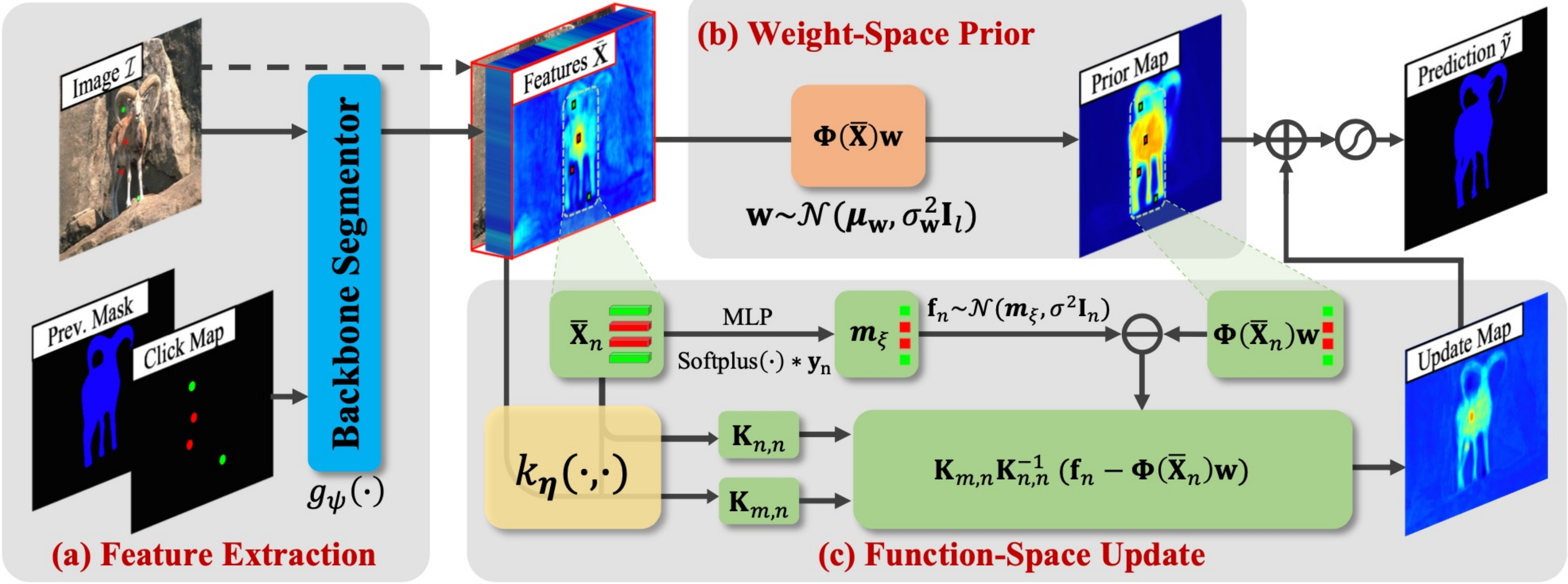
$$\tilde{\mathbf{f}}_n \approx \mathbf{f}_n \approx \mathbf{m}_\xi = \underbrace{\boxed{\mathrm{Softplus}(\mathrm{MLP}_\xi(\mathbf{X}_n))}}_{>0} * \mathbf{y}_n$$

Methodology

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

GPCIS Framework

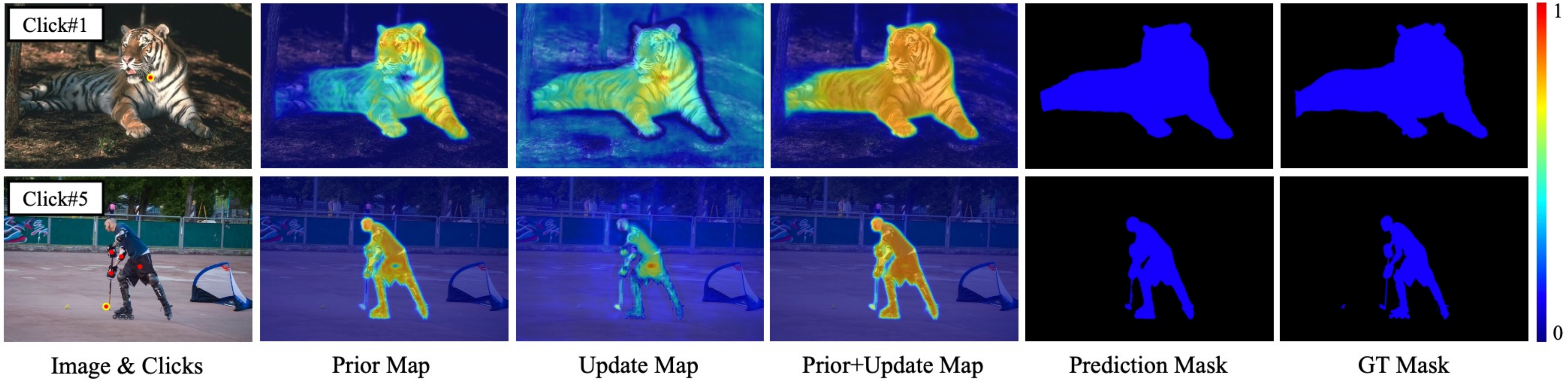
$$\tilde{\mathbf{f}}_* = \underbrace{\mathbf{\Phi}(\mathbf{X}_*)\mathbf{w}}_{\text{weight-space prior}} + \underbrace{\mathbf{K}_{*,n}\mathbf{K}_{n,n}^{-1}(\mathbf{f}_n - \mathbf{\Phi}(\mathbf{X}_n)\mathbf{w})}_{\text{function-space update}},$$



Experiments

Experiments

Visualization of the Decoupled GP Posterior



Experiments

Accuracy at Clicked Pixels

Table 1. The effect of ϵ^2 on the NoIC of our proposed GPCIS with the backbone segmentor ResNet50 on the DAVIS dataset [36].

ϵ^2	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
NoIC	36	30	21	15	15	8	2

NoIC : the Number of Incorrectly classified Clicks over a testing dataset

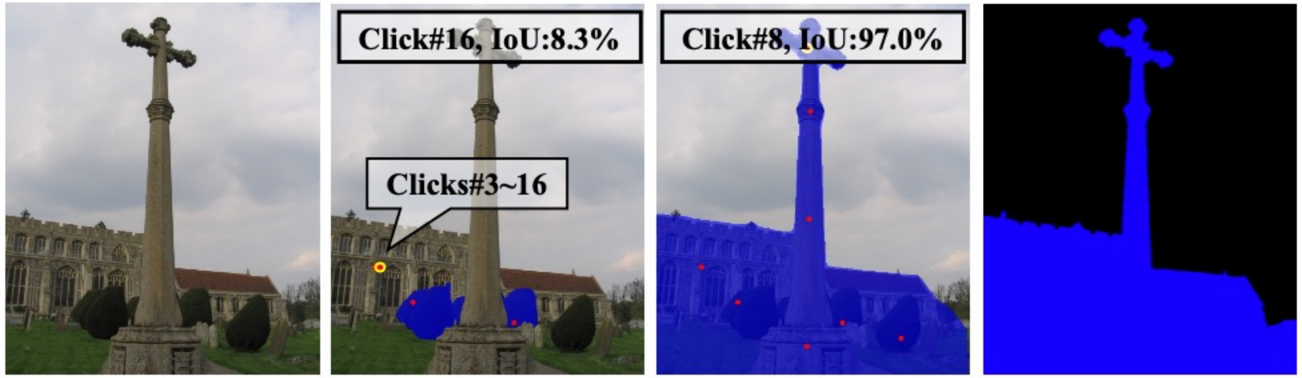


Image RITM GPCIS (Ours) GT Mask

Experiments

NoC Performance over Four Evaluation Datasets

Table 2. NoC@85 and NoC@90 of different competing methods on four datasets, *i.e.*, GrabCut, Berkeley, SBD, and DAVIS. '*' denotes the models trained on the Augmented PASCAL VOC dataset [9, 13]. Bold and underlined results indicate the top 1st and 2nd rank, respectively.

Backbone	Method	GrabCut [38]		Berkeley [32]		SBD [13]		DAVIS [36]		Average	
		NoC@85	NoC@90	NoC@85	NoC@90	NoC@85	NoC@90	NoC@85	NoC@90	NoC@85	NoC@90
DeepLab-LargeFOV [4]	* RIS-Net [24] *('17)*	-	5.00	-	6.03	-	-	-	-	-	-
CAN [53]	LD [23] *('18)*	3.20	4.79	-	-	7.41	10.78	5.95	9.57	-	-
FCN [29]	*DOS [51] *('16)*	5.08	6.08	-	-	9.22	12.80	9.03	12.58	-	-
	*CMG [31] *('19)*	-	3.58	-	5.60	-	-	-	-	-	-
DenseNet [17]	BRS [18] *('19)*	2.60	3.60	-	5.08	6.59	9.78	5.58	8.24	-	6.68
Xception-65 [8]	*CA [22] *('20)*	-	3.07	-	4.94	-	-	5.16	-	-	-
SegFormerB0-S2 [7, 50]	RITM [41] *('21)*	1.62	1.82	**1.84**	2.92	4.26	6.38	4.65	6.13	3.09	4.31
	FocalClick [7] *('22)*	1.66	1.90	-	3.14	4.34	6.51	5.02	7.06	-	4.65
	GPCIS *(Ours)*	**1.60**	**1.76**	**1.84**	**2.70**	**4.16**	**6.28**	**4.45**	**6.04**	**3.01**	**4.20**
HRNet18s-S2 [7, 43]	RITM [41] *('21)*	2.00	2.24	2.13	3.19	4.29	6.36	4.89	6.54	3.33	4.58
	FocalClick [7] *('22)*	1.86	2.06	-	3.14	4.30	6.52	4.92	6.48	-	4.55
	GPCIS *(Ours)*	**1.74**	**1.94**	**1.83**	**2.65**	**4.28**	**6.25**	**4.62**	**6.16**	**3.12**	**4.25**
ResNet50 [15]	*FCANet [26] *('20)*	2.18	2.62	-	4.66	-	-	5.54	8.83	-	-
	f-BRS-B [40] *('20)*	2.20	2.64	2.17	4.22	4.55	7.45	5.44	7.81	3.59	5.53
	CDNet [6] *('21)*	2.22	2.64	-	3.69	4.37	7.87	5.17	6.66	-	5.22
	RITM [41] *('21)*	2.16	2.30	1.90	2.95	3.97	5.92	4.56	6.05	3.15	4.31
	FocusCut [25] *('22)*	**1.60**	**1.78**	1.86	3.44	**3.62**	**5.66**	5.00	6.38	3.02	4.32
	FocalClick [7] *('22)*	1.92	2.14	1.87	2.86	3.84	5.82	4.61	6.01	3.06	4.21
	GPCIS *(Ours)*	1.64	1.82	**1.60**	**2.60**	3.80	5.71	**4.37**	**5.89**	**2.85**	**4.00**

Experiments

Performance on Different Metrics

Table 3. Quantitative evaluation on different metrics, and comparisons on parameters and inference time. Here the backbone segmentor is ResNet50, and Second Per Click (SPC) is averagely computed over DAVIS with the testing image size of 384×384 on an NVIDIA V100 GPU. Lower NoC_{100}@90, NoF_{100}@90, NoIC, #Params, SPC and higher IoU&1, IoU&5 indicate better performance.

Method	Berkeley [32]					DAVIS [36]					#Params (MB)	SPC (ms)
	NoC_{100}@90	NoF_{100}@90	IoU&1	IoU&5	NoIC	NoC_{100}@90	NoF_{100}@90	IoU&1	IoU&5	NoIC		
f-BRS-B [40]	6.21	2	77.06%	85.00%	1	22.62	57	70.97%	83.87%	**0**	39.44	116.53
CDNet [6]	-	-	-	-	-	18.59	48	-	-	-	39.90	57.76
RITM [41]	3.75	**1**	76.88%	94.66%	2	18.09	51	72.89%	89.14%	74	39.48	**34.24**
FocusCut [25]	4.63	**1**	78.89%	92.89%	1	19.00	45	72.71%	87.58%	6	40.36	950.68
FocalClick [7]	4.46	2	75.59%	94.90%	**0**	17.74	49	70.76%	88.90%	42	39.50	41.80
GPCIS *(Ours)*	**3.36**	**1**	**79.43%**	**95.11%**	**0**	**17.03**	**44**	**75.67%**	**89.60%**	2	**39.39**	38.82

Experiments

Quality Performance

1 Click

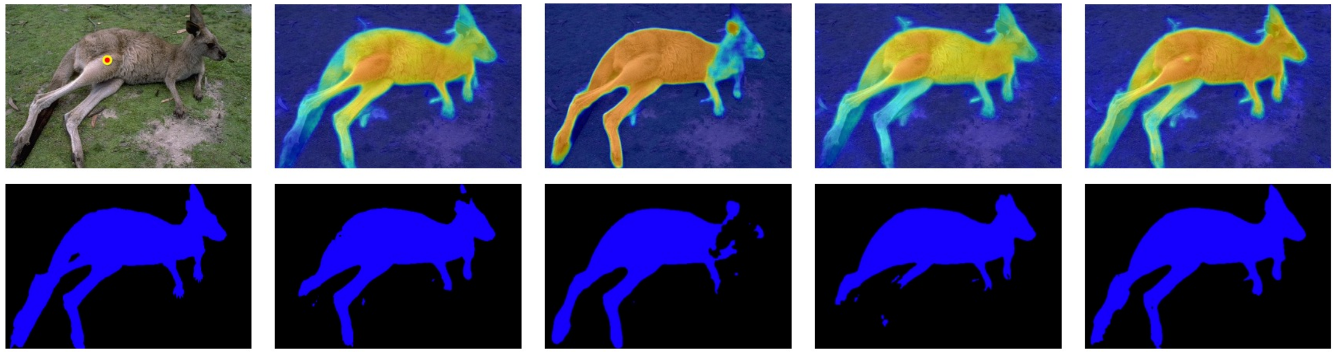


Image & GT Mask | RITM (IoU 84.81%) | FocusCut (IoU 78.57%) | FocalClick (IoU 79.36%) | GPCIS (IoU **94.62%**)

2 Clicks

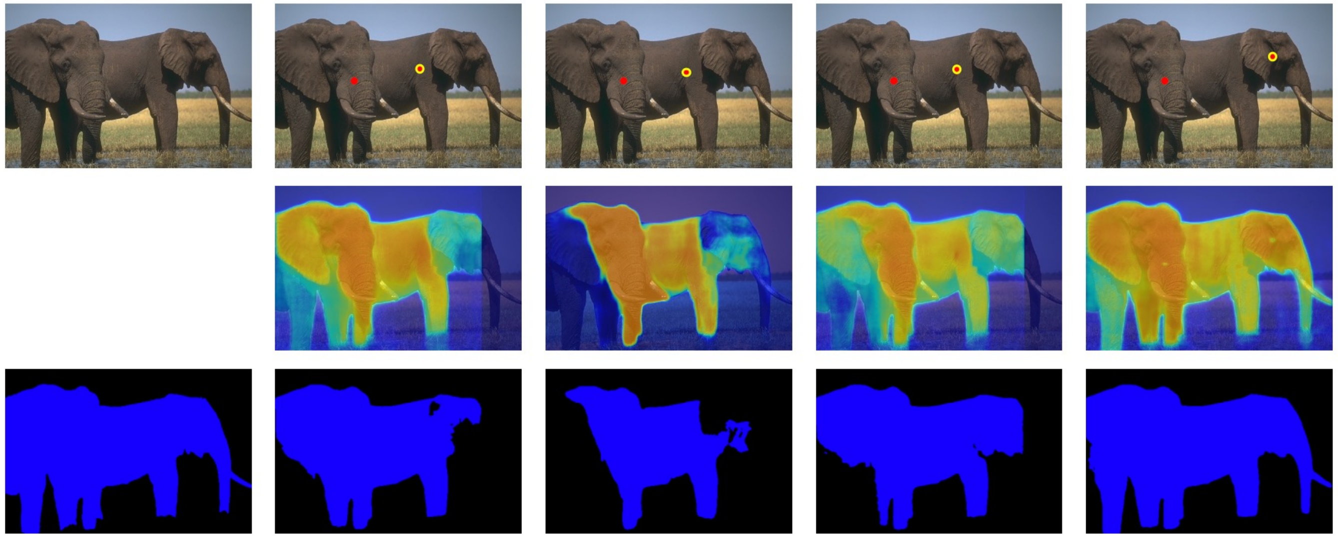


Image & GT Mask | RITM (IoU 78.15%) | FocusCut (IoU 54.44%) | FocalClick (IoU 78.37%) | GPCIS (IoU **96.60%**)

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Thank you!

Github Page

Virtual Site

woshizhouminghao@stu.xjtu.edu.cn