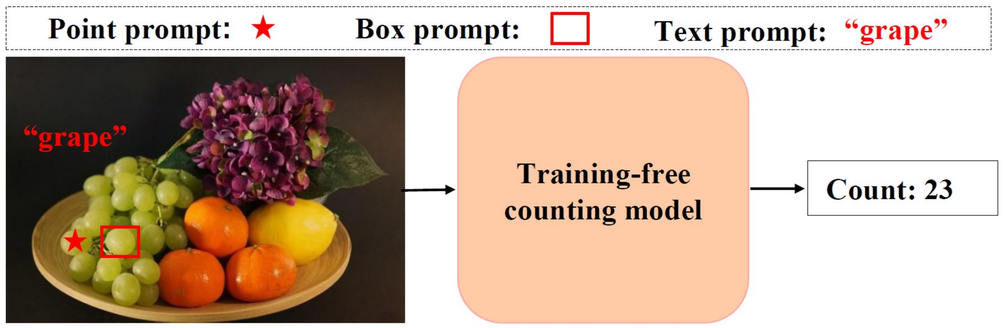
Training-free Object Counting with Prompts

Zenglin Shi[1], Ying Sun[1, 2], Mengmi Zhang[1, 2, 3]

[1]I2R, Agency for Science, Technology and Research, Singapore
[2]CFAR, Agency for Science, Technology and Research, Singapore
[3]Nanyang Technological University (NTU), Singapore

Goal of this work



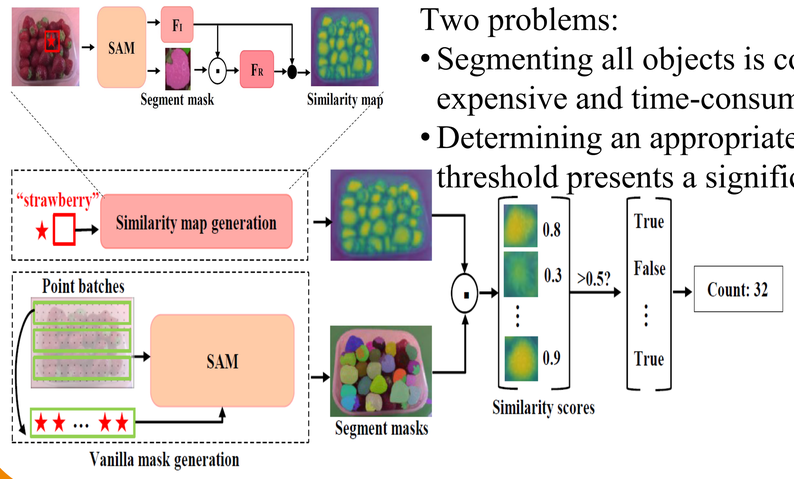
Building a **training-free counting** model, where we can specify what to count with **prompts**

Contributions

- The class-agnostic counting task is reformulated as a prompt-based segmentation problem to eliminate the need for extensive data collection and model training.
- A new prior-guided mask generation method is proposed to improve the segmentation efficiency and accuracy by incorporating three types of priors into SAM.
- A new two-stage approach for counting objects specified through text is proposed by integrating reference object selection with the prior-guided mask generation method.

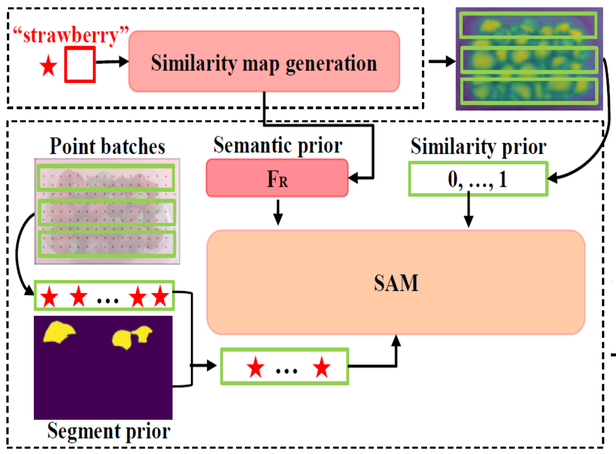
Methods

Counting by segmentation with vanilla SAM

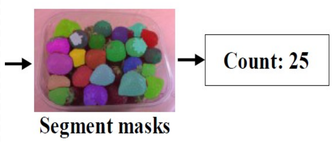


Two problems:
- Segmenting all objects is computationally expensive and time-consuming.
- Determining an appropriate similarity score threshold presents a significant challenge.

Counting by segmentation with prior-guided SAM



Three priors:
- Similarity prior: Create positive and negative points as SAM prompts by utilizing a similarity map.
- Segment prior: Remove redundant positive points by leveraging an overall segment map.
- Semantic prior: Use reference object's semantic feature enables SAM to identify target objects better.



Two advantages:
- Only target objects are segmented
- Improved efficiency and accuracy.

Experiments

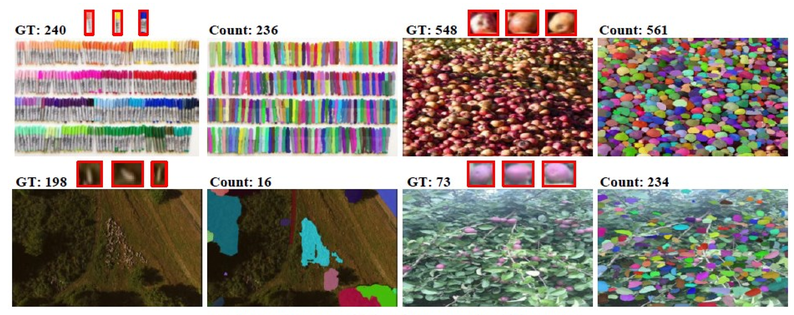
Metrics: Mean Absolute Error (MAE); Root Mean Square Error (RMSE); Normalized Relative Error (NAE); Squared Relative Error (SRE)

Counting with point and box prompts

	Training	Prompt	FSC-147				CARPK			
			MAE ↓	RMSE ↓	NAE ↓	SRE ↓	MAE ↓	RMSE ↓	NAE ↓	SRE ↓
GMN [13]	Yes	box	26.52	124.57	-	-7.48	9.90	-	-	
FamNet+ [17]	Yes	box	22.08	99.54	0.44	6.45	18.19	33.66	-	-
CFOCNet+ [27]	Yes	box	22.10	112.71	-	-	-	-	-	-
BMNet+ [19]	Yes	box	14.62	91.83	0.25	2.74	5.76	7.83	-	-
SAM	No	N.A.	42.48	137.50	1.14	8.13	16.97	20.57	0.70	5.30
Ours (vanilla)	No	box	26.29	137.89	0.38	4.38	15.67	19.44	0.67	5.06
Ours	No	box	**19.95**	**132.16**	**0.29**	**3.80**	**10.97**	**14.24**	**0.48**	**3.70**
Ours (vanilla)	No	point	25.18	137.62	0.37	4.34	15.67	19.44	0.67	5.06
Ours	No	point	20.10	132.83	0.30	3.87	11.01	14.34	0.51	3.89

Ablation study

Three types of priors			FSC-147			
Similarity	Segment	Semantic	MAE ↓	RMSE ↓	NAE ↓	SRE ↓
			42.48	137.59	1.14	8.13
✓			21.36	134.07	**0.27**	4.29
	✓		26.14	134.98	0.51	4.84
		✓	37.17	134.86	1.12	8.19
✓	✓		20.38	134.32	0.31	3.89
✓		✓	20.83	133.16	0.38	5.29
✓	✓	✓	**19.95**	**132.16**	0.29	**3.80**

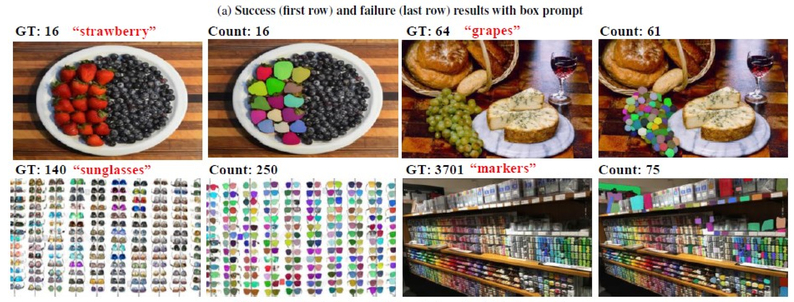


(a) Success (first row) and failure (last row) results with box prompt

Counting with text prompt

	Training	FSC-147			
		MAE ↓	RMSE ↓	NAE ↓	SRE ↓
Xu *et al.* [24]	Yes	22.09	115.17	0.34	3.74
SAM	No	42.48	137.50	1.14	8.13
Ours (vanilla)	No	32.86	142.89	0.44	5.12
Ours	No	**24.79**	**137.15**	**0.37**	**4.52**

- Perform much better than the vanilla method and original SAM
- Competitive results compared to learning-based approaches
- Succeed in the scenes where objects are individually visible
- Fail in extreme scenes where individual objects are hard to distinguish



(b) Success (first row) and failure (last row) with text prompt

DEEP NEUROCOGNITION LAB

Code