Segment Anything

Veronika Lebedyuk

Segment Anything Project

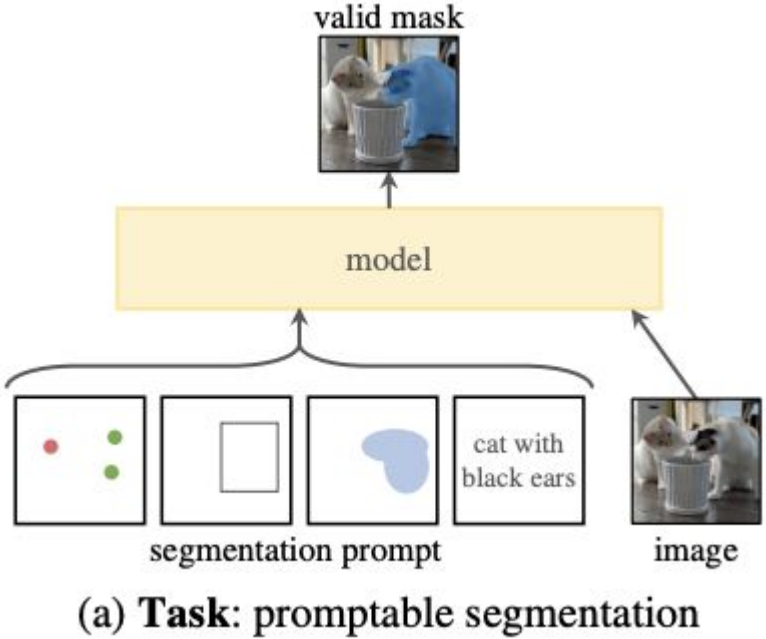
Goal: to build a foundation model for image segmentation

Introducing new:

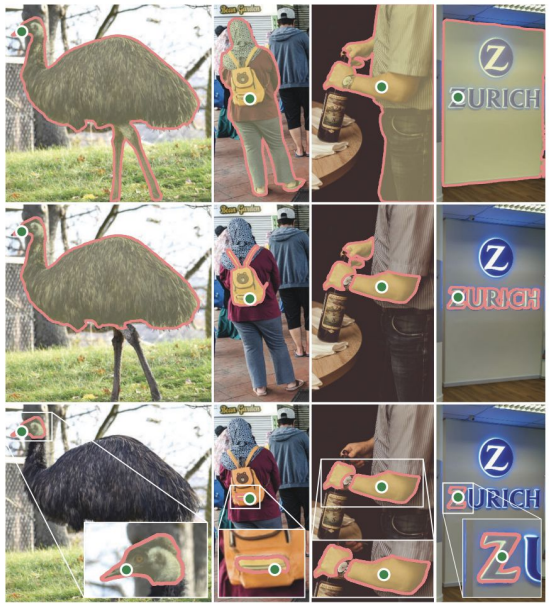
- task
- model
- dataset

Segment Anything Task

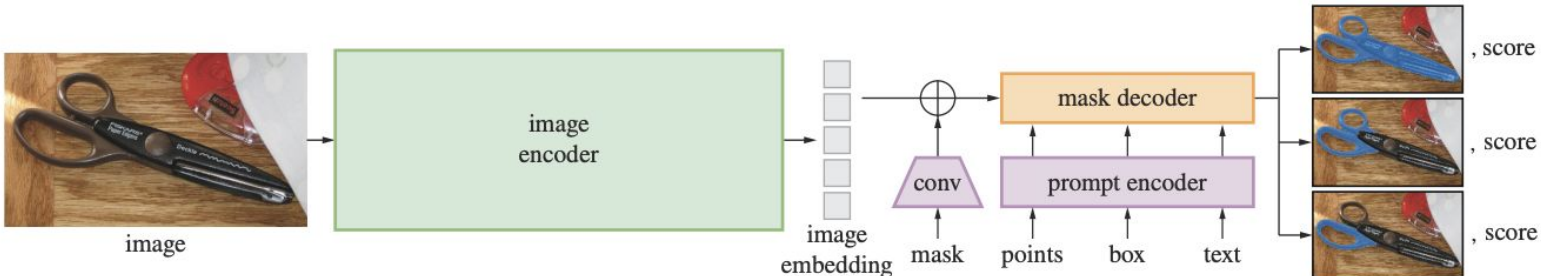
the goal is to return a valid segmentation mask given any segmentation prompt



(a) **Task**: promptable segmentation



Segment Anything Model

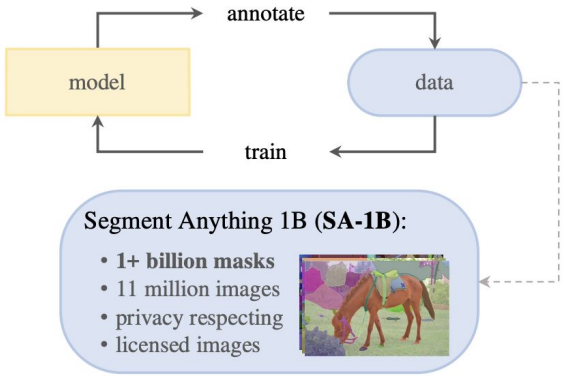


Segment Anything Dataset

Data engine: iterate between using model to assist in data collection and using the newly collected data to improve the model

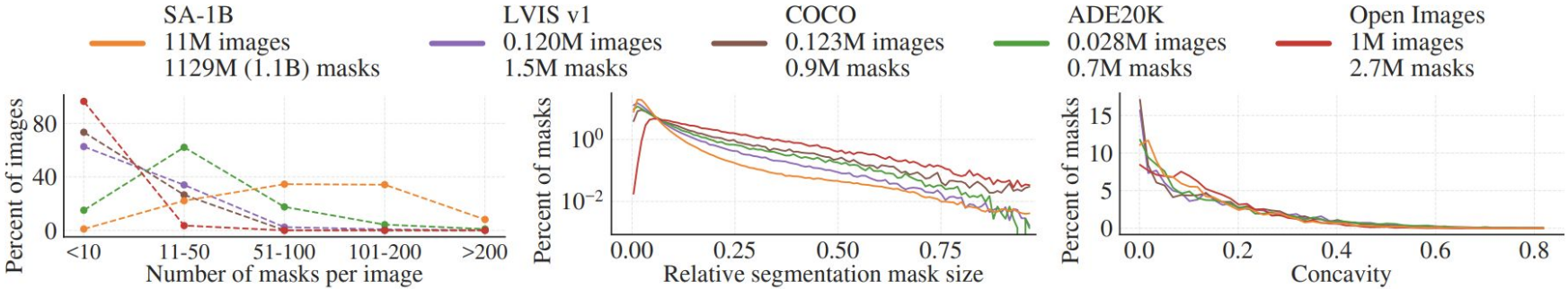
stage 0: SAM trained on public datasets

- Assisted-manual stage (interactive segmentation)
- Semi-automatic stage (increasing mask diversity)
- Fully automatic stage



Segment Anything Dataset

- SA-1B, collected fully automatically using the final stage of our data engine, has 400× more masks than any existing segmentation dataset



Segment Anything Dataset

demo

Segment Anything Dataset

demo

Zero-Shot Transfer Tasks

explore different levels of image understanding.

low-level: Edge Detection

mid-level: Object Proposals

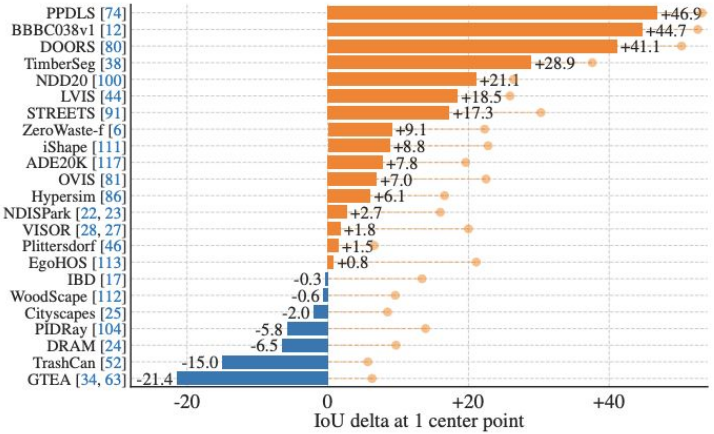
high-level: Instance Segmentation

even higher-level:Text-to-Mask

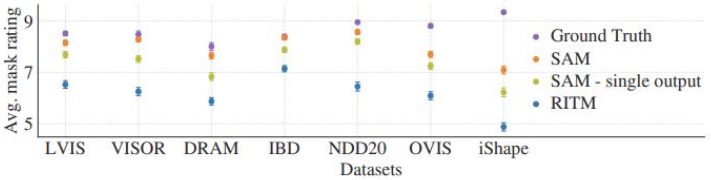
These four tasks differ significantly from the promptable segmentation task that SAM was trained on and are implemented via **prompt engineering**

Zero-Shot Single Point Valid Mask Evaluation

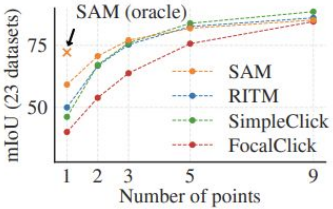
evaluate segmenting an object from a single foreground point



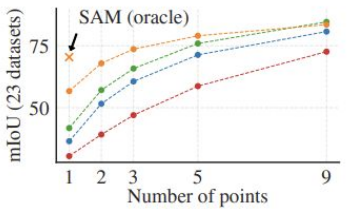
(a) SAM *vs.* RITM [92] on 23 datasets



(b) Mask quality ratings by human annotators



(c) Center points (default)



(d) Random points

Zero-Shot Edge Detection

prompt SAM with a 16×16 regular grid of foreground points

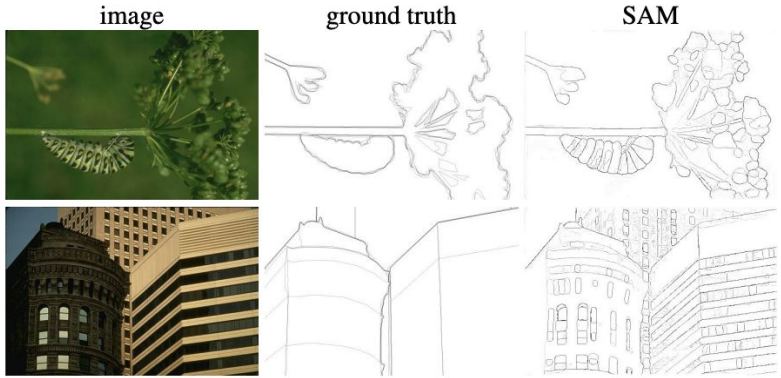


Figure 10: Zero-shot edge prediction on BSDS500. SAM was not trained to predict edge maps nor did it have access to BSDS images or annotations during training.

Zero-Shot Object Proposals and Instance Segmentation

we run a object detector (the ViTDet used before) and prompt SAM with its output boxes

	mask AR@1000						
method	all	small	med.	large	freq.	com.	rare
ViTDet-H [62]	63.0	51.7	80.8	87.0	63.1	63.3	58.3
zero-shot transfer methods:							
SAM – single out.	54.9	42.8	76.7	74.4	54.7	59.8	62.0
SAM	59.3	45.5	81.6	86.9	59.1	63.9	65.8

	COCO [66]				LVIS v1 [44]			
method	AP	AP^S	AP^M	AP^L	AP	AP^S	AP^M	AP^L
ViTDet-H [62]	51.0	32.0	54.3	68.9	46.6	35.0	58.0	66.3
zero-shot transfer methods (segmentation module only):								
SAM	46.5	30.8	51.0	61.7	44.7	32.5	57.6	65.5

Zero-Shot Text-to-Mask

prompt SAM with the extracted CLIP image embeddings as its first interaction

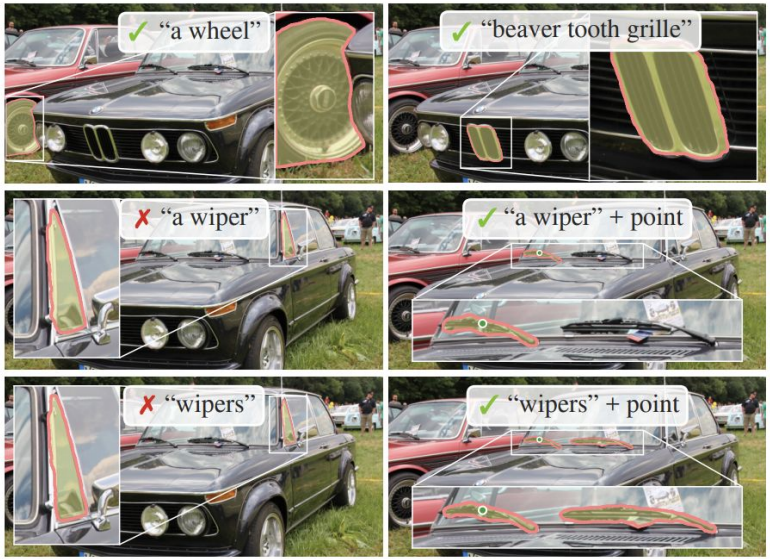


Figure 12: Zero-shot text-to-mask. SAM can work with simple and nuanced text prompts. When SAM fails to make a correct prediction, an additional point prompt can help.

Conclusion

The Segment Anything project is an attempt to lift image segmentation into the era of foundation models. Our principal contributions are a new task (promptable segmentation), model (SAM), and dataset (SA-1B) that make this leap possible.