

SEGMENT ANYTHING

Kirillov et al.

Meta AI Research

Presented by William Guimont-Martin

2024-02-21

Segment Anything

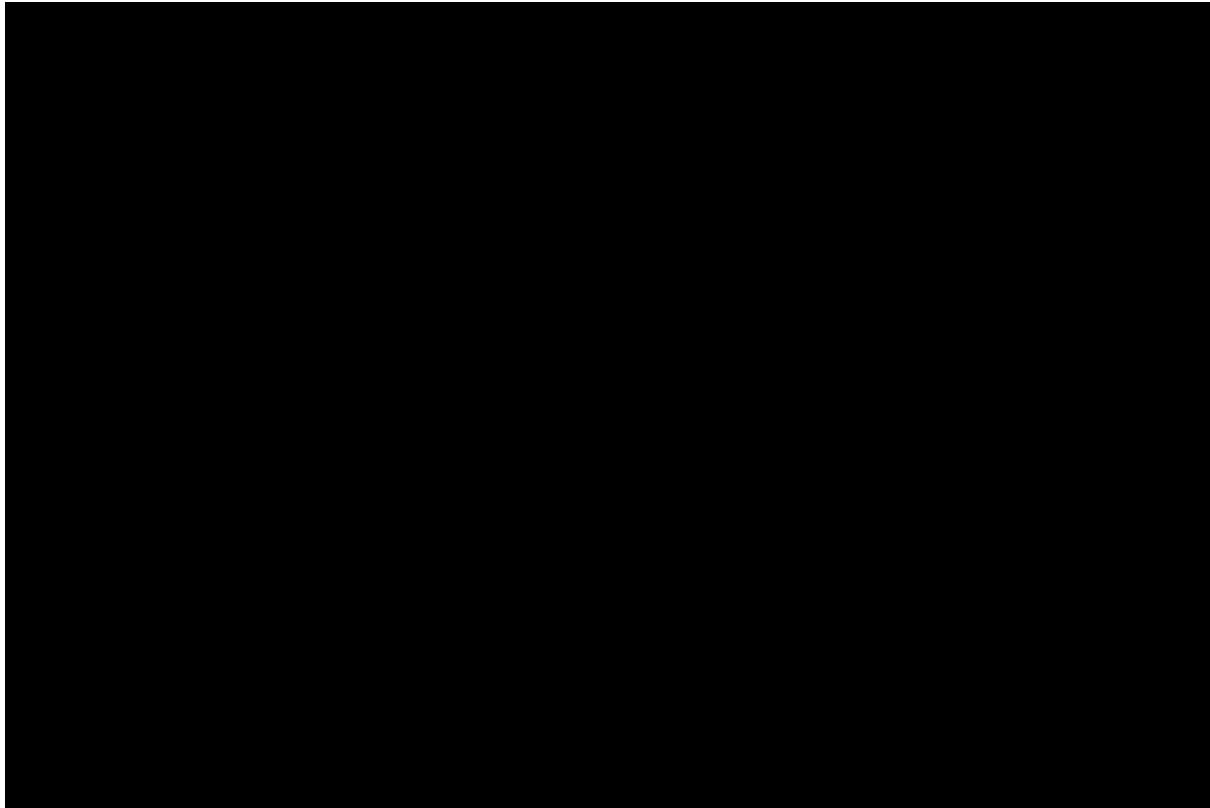
Motivation

- LLMs show strong zero- and few-shot generalization
 - New tasks using **prompt engineering**
- Segment Anything as a **vision foundation model** for segmentation
 - Composable in a larger system to solve various tasks



Segment Anything

A small example

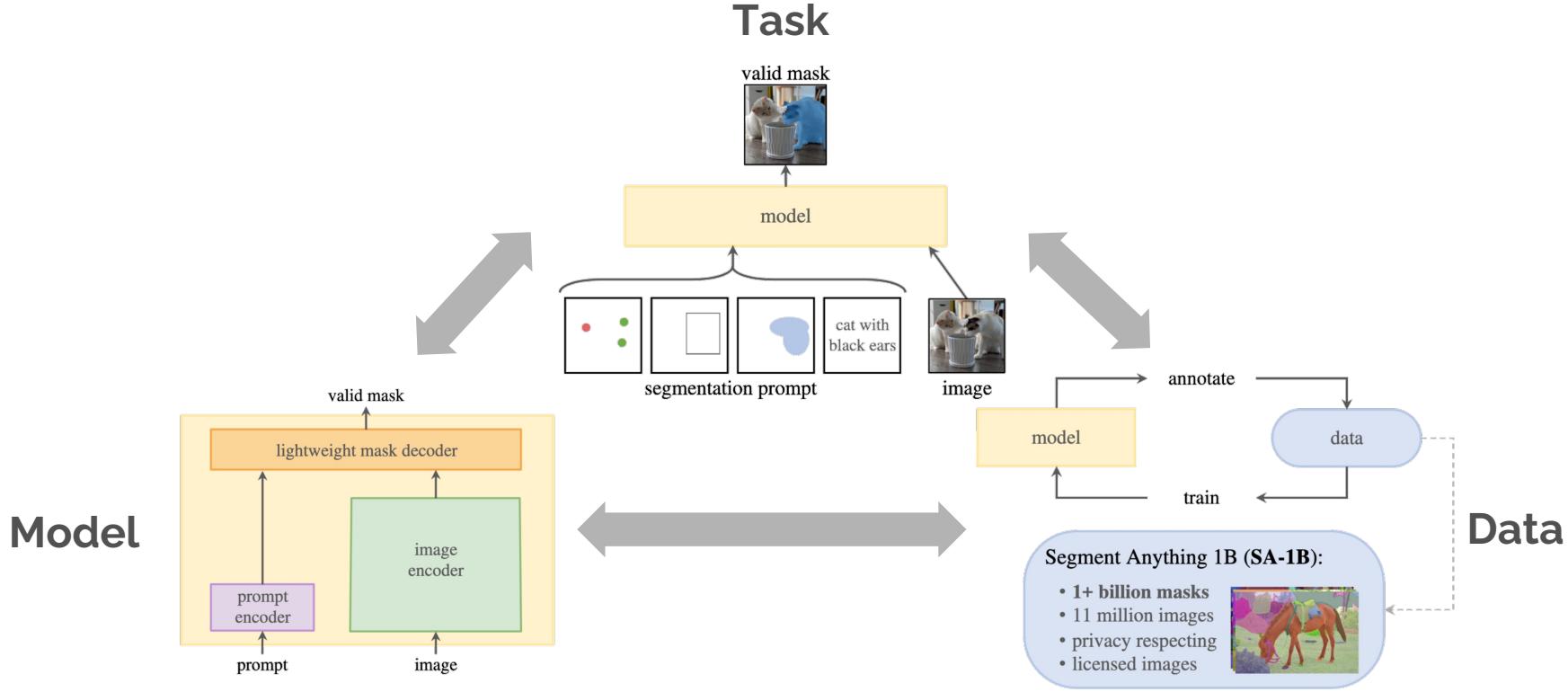


Segment Anything

Foundation model for image segmentation

- What **tasks** will enable zero-shot generalization?
- What is the corresponding **model** architecture?
- What **data** can power this task and model?

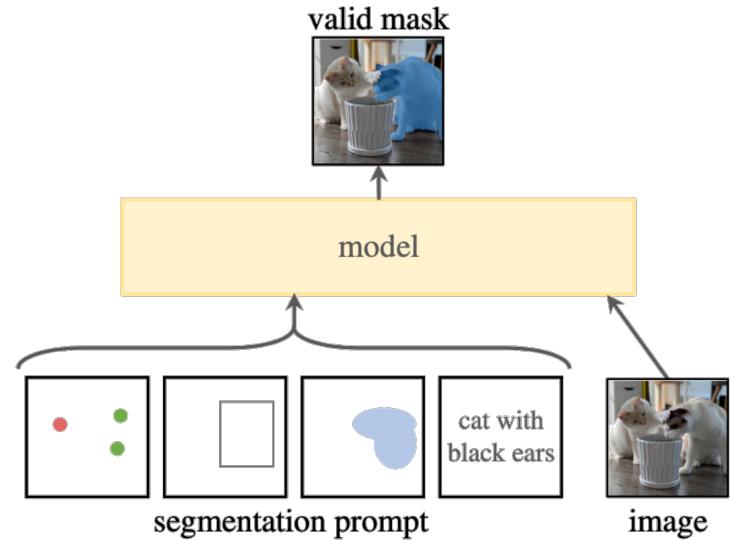
Segment Anything



Task

A promptable segmentation task

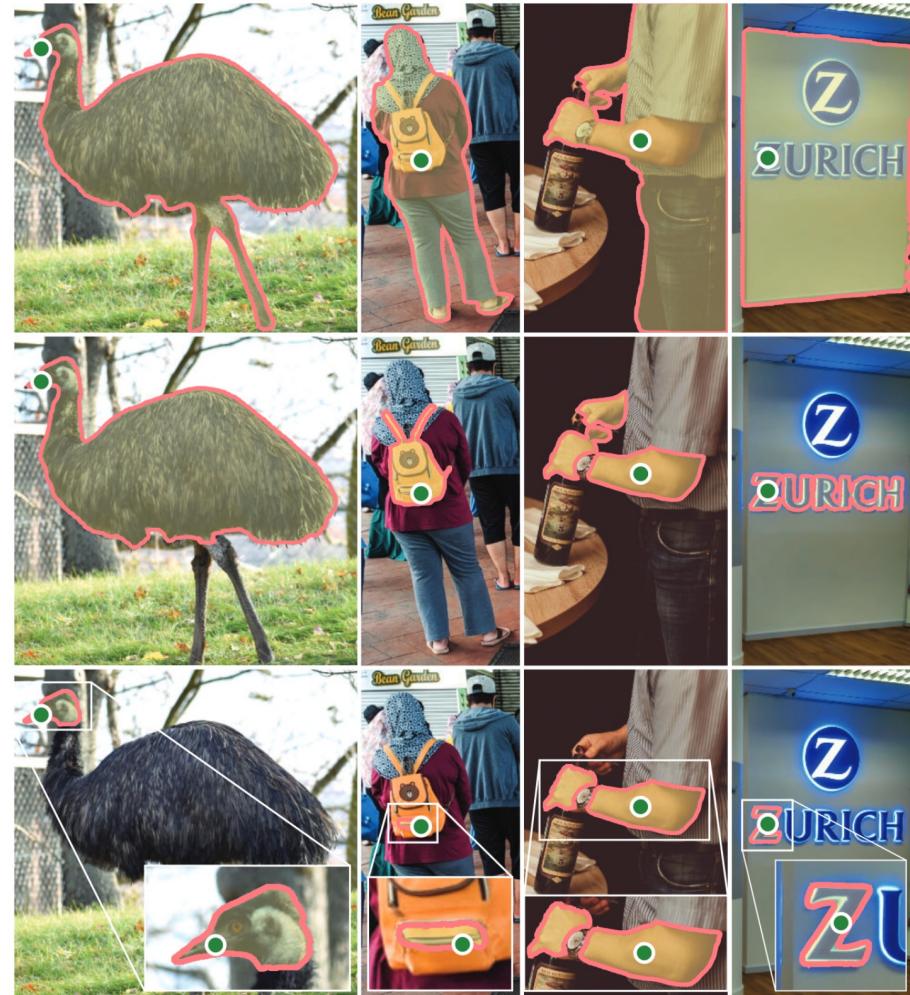
- Prompts in multiple formats
- Zero-shot generalization to new tasks
 - As a component of a larger system
 - Edge detection
 - Instance segmentation
 - Text-to-masks
- Natural pre-training task
 - Generate prompts from annotation



Task

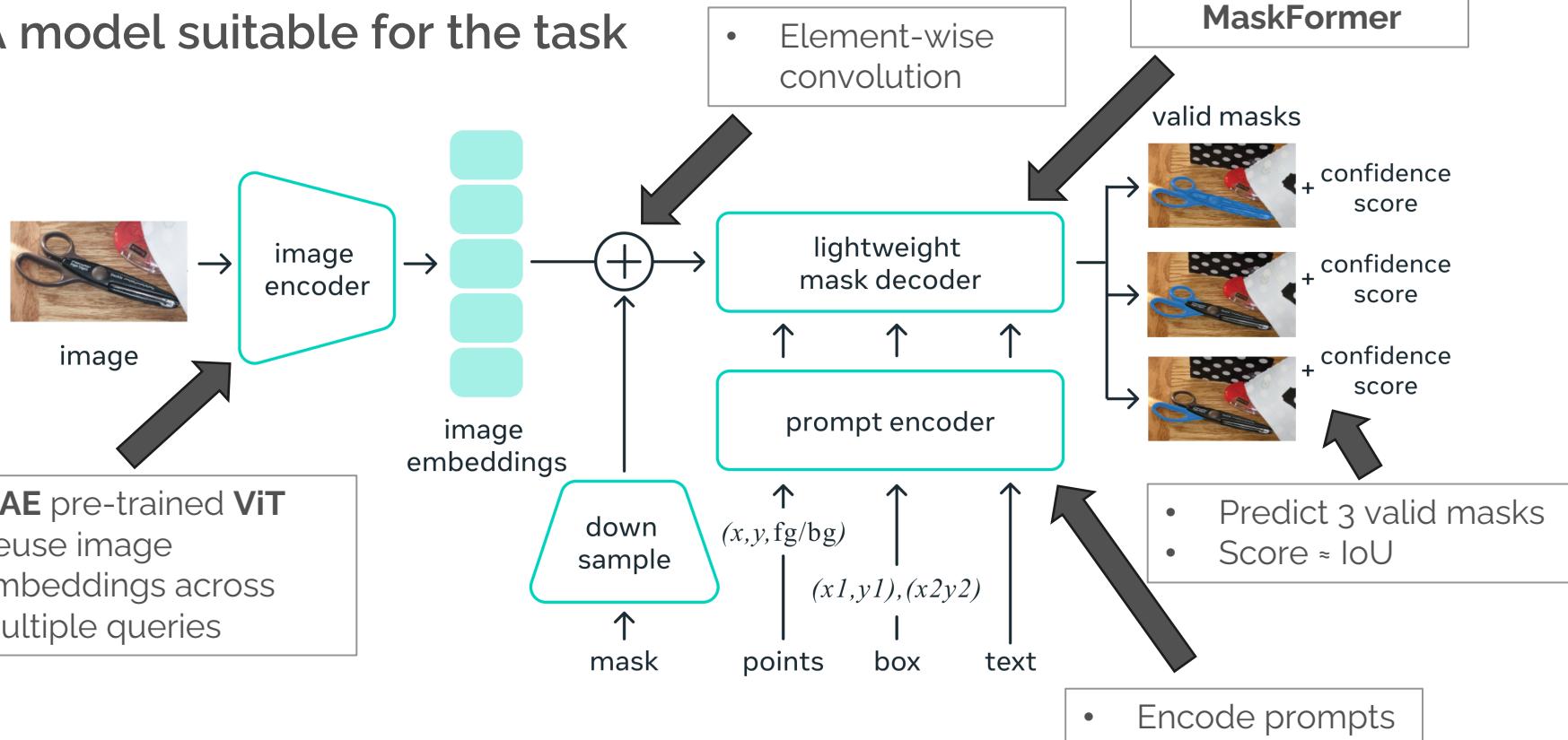
Prompt ambiguity

- Prompts can be ambiguous
- Part vs Whole
- Predict valid masks



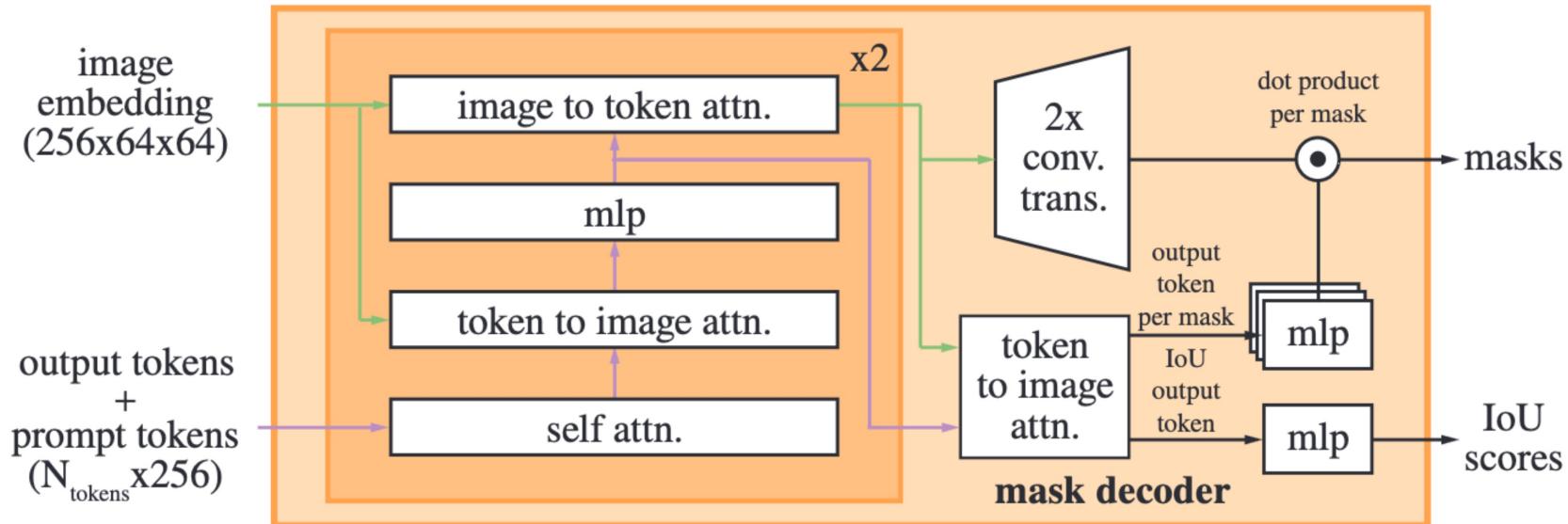
Model - SAM

A model suitable for the task



Model - SAM

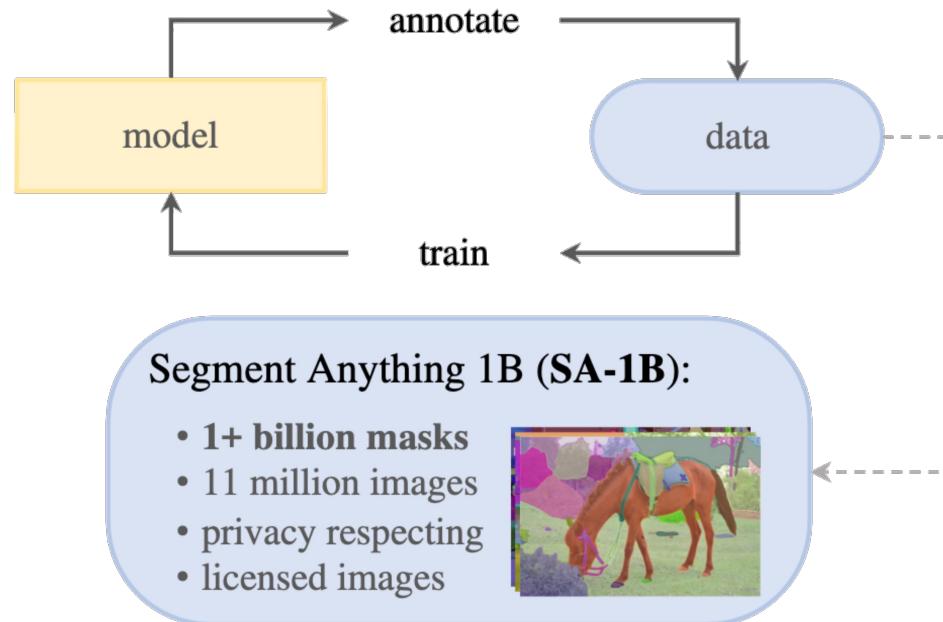
Mask decoder



Data

Data engine to power SAM

- There is not a lot of mask data
- Assisted-manual stage
 - Train on public mask dataset
 - Annotation tool powered by SAM
 - 6 retraining
 - **4.3M** masks from **120k** images
- Semi-automatic stage
 - Increase diversity
 - Manual annotation of unannotated objects
 - 5 retraining
 - **10.2M** masks from **180k** images
- Fully automatic stage
 - Enough masks
 - Ambiguity-aware model + confidence
 - Stable masks + NMS + Zoom-in
 - **1.1B** masks from **11M** images



SA-1B

A large masks dataset

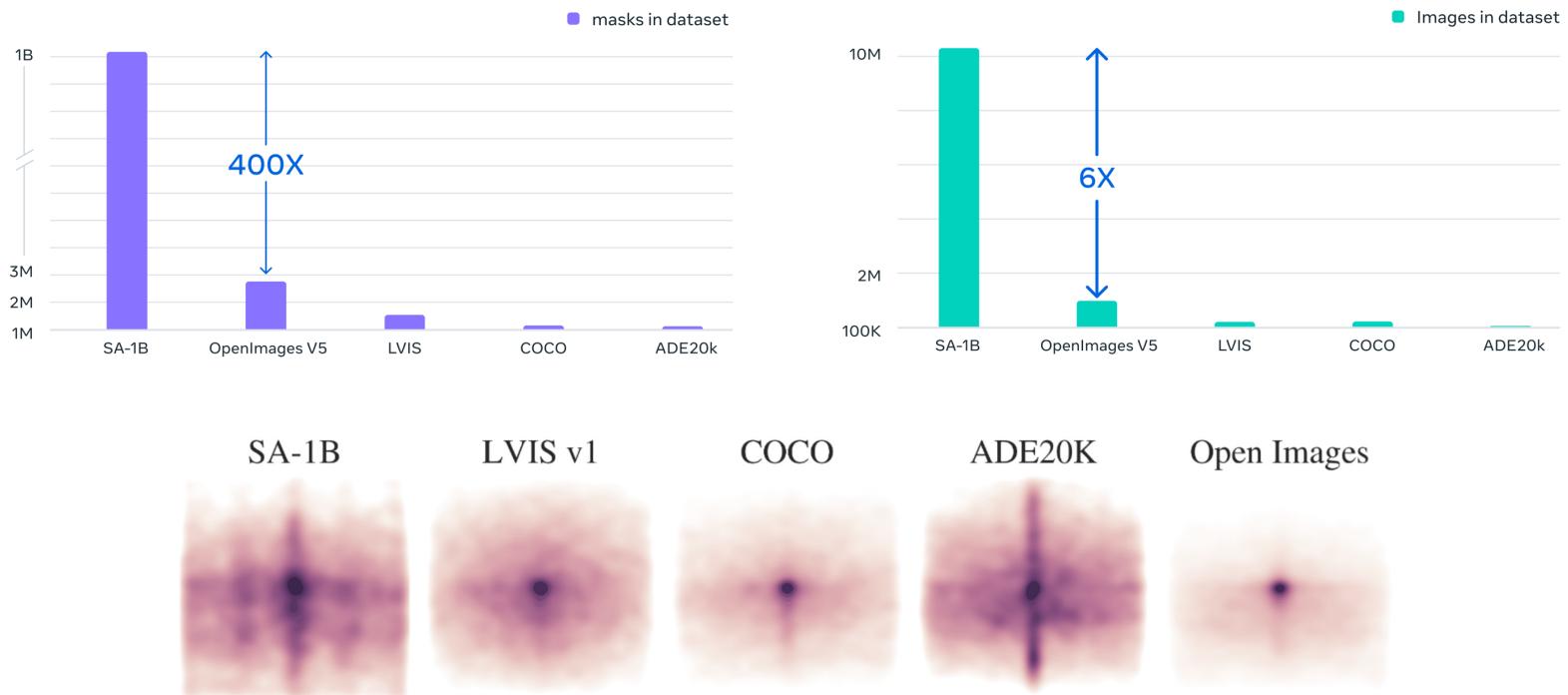


Figure 5: Image-size normalized mask center distributions.

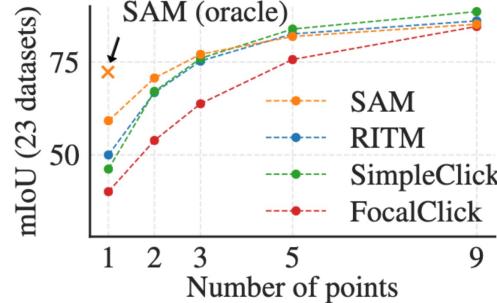
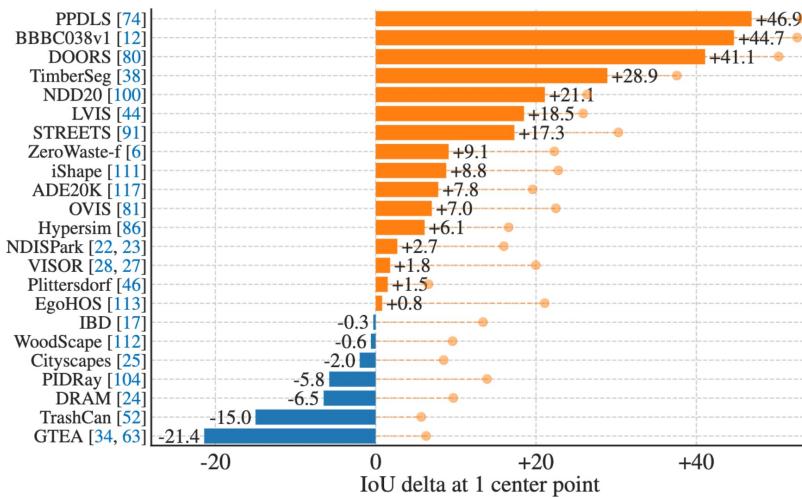
Zero-Shot Transfer Experiments

Test on 23 diverse datasets

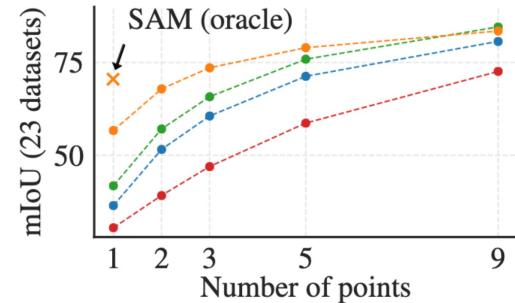


Zero-Shot Transfer Experiments

Results



(c) Center points (default)



(d) Random points

Zero-Shot on Other Tasks

Edge detection

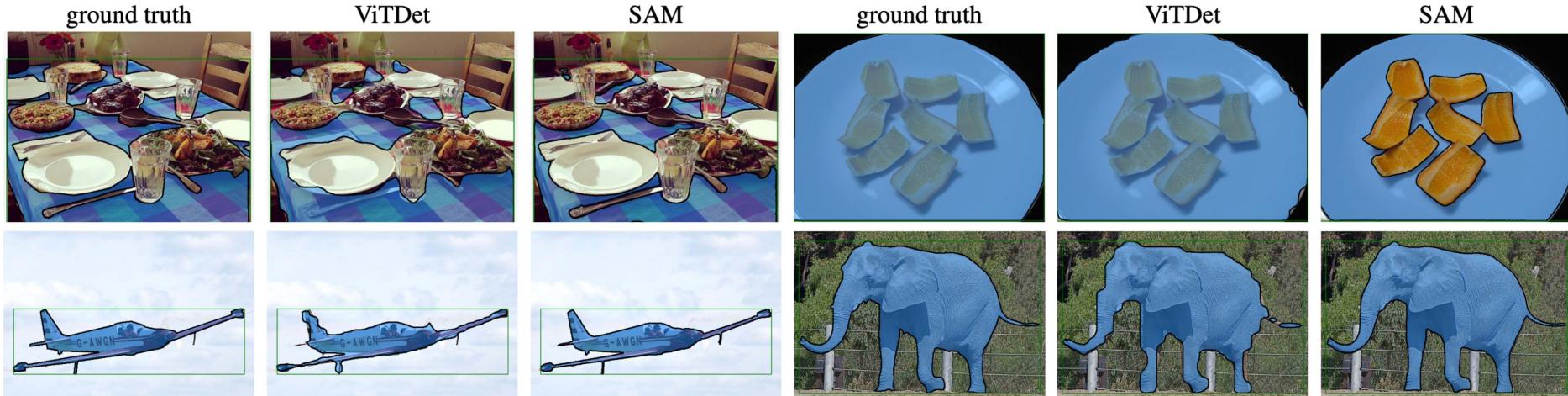
- Masks generation
- NMS
- Sobel filter on unthresholded mask



Zero-Shot on Other Tasks

Instance segmentation

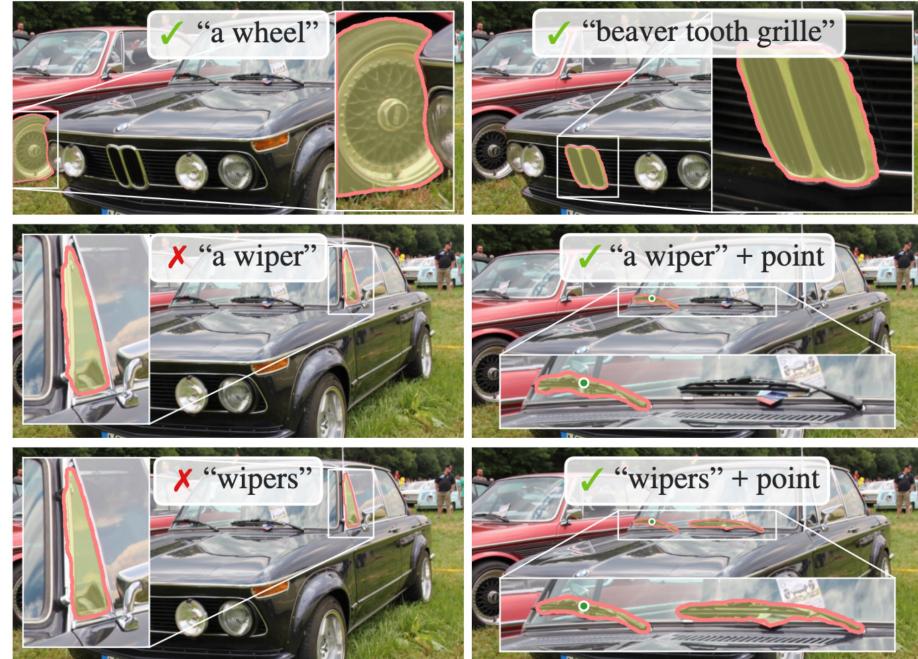
- Object detector bounding boxes → SAM



Zero-Shot on Other Tasks

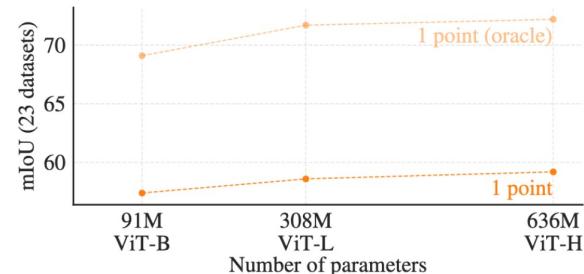
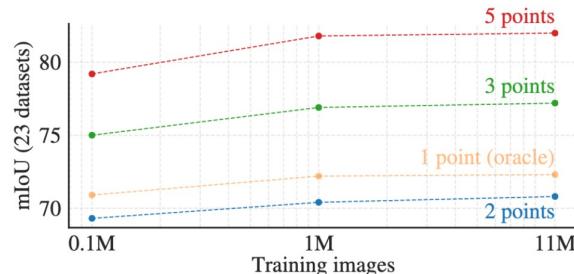
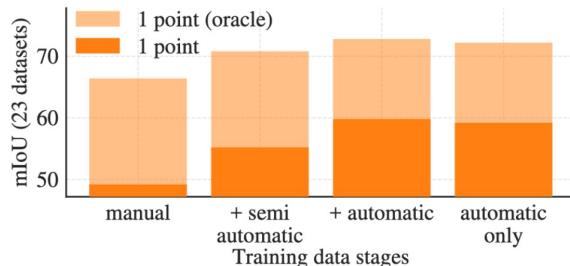
Text-to-masks

- CLIP image embeddings during training
- CLIP text embeddings at inference
- Point help solve ambiguity



Ablations

- Oversampling by 10x manual and semi-automatic data
- Stages help to better handle ambiguity
- 10% of SA-1B comparable to 100%
- Saturated scaling with bigger architectures



Try it on your own pictures

<https://segment-anything.com/demo>



?

Questions?

?

