

SEGMENT ANYTHING

“The first foundation model for image segmentation”.

<https://arxiv.org/pdf/2002.05709>

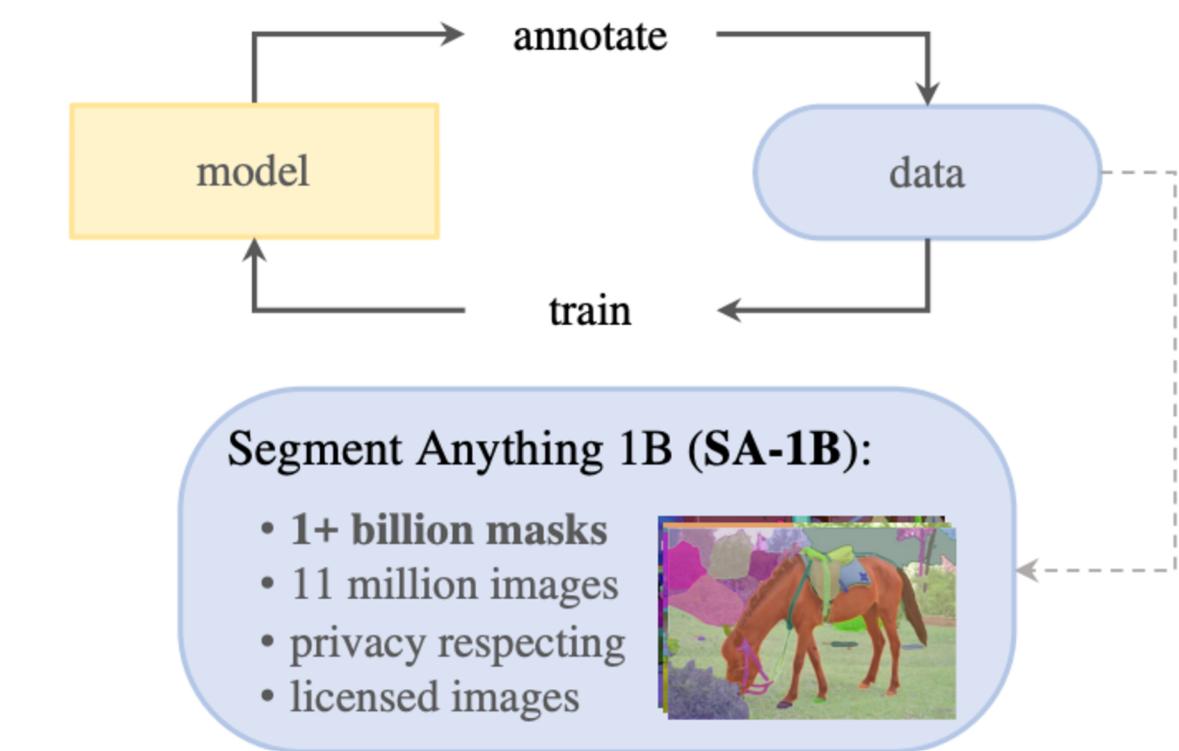
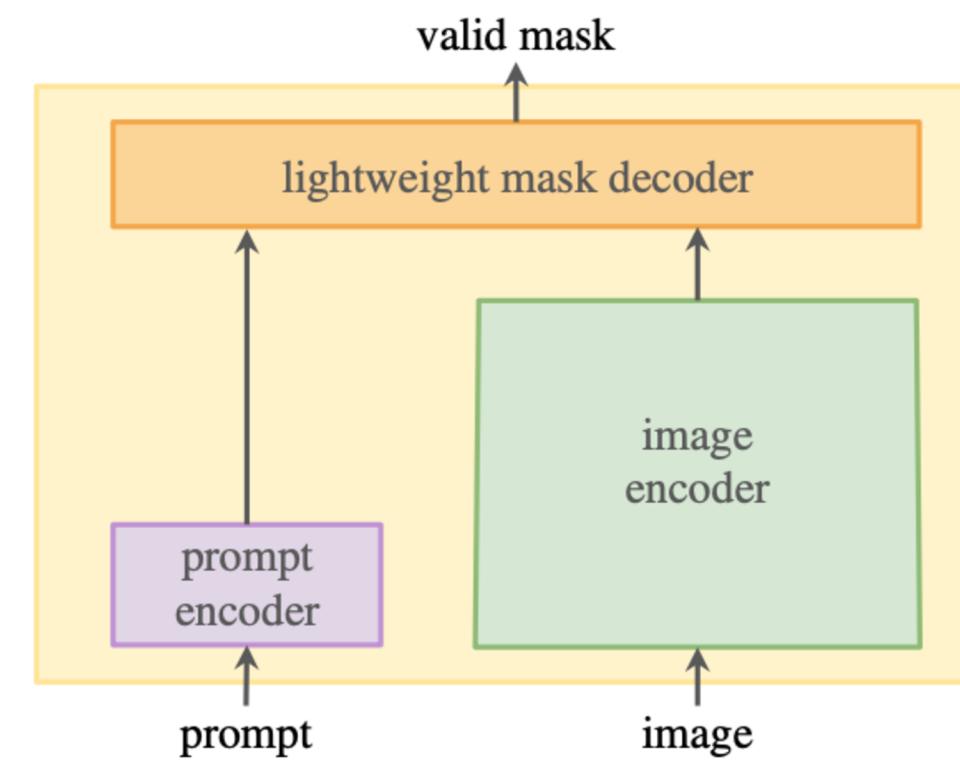
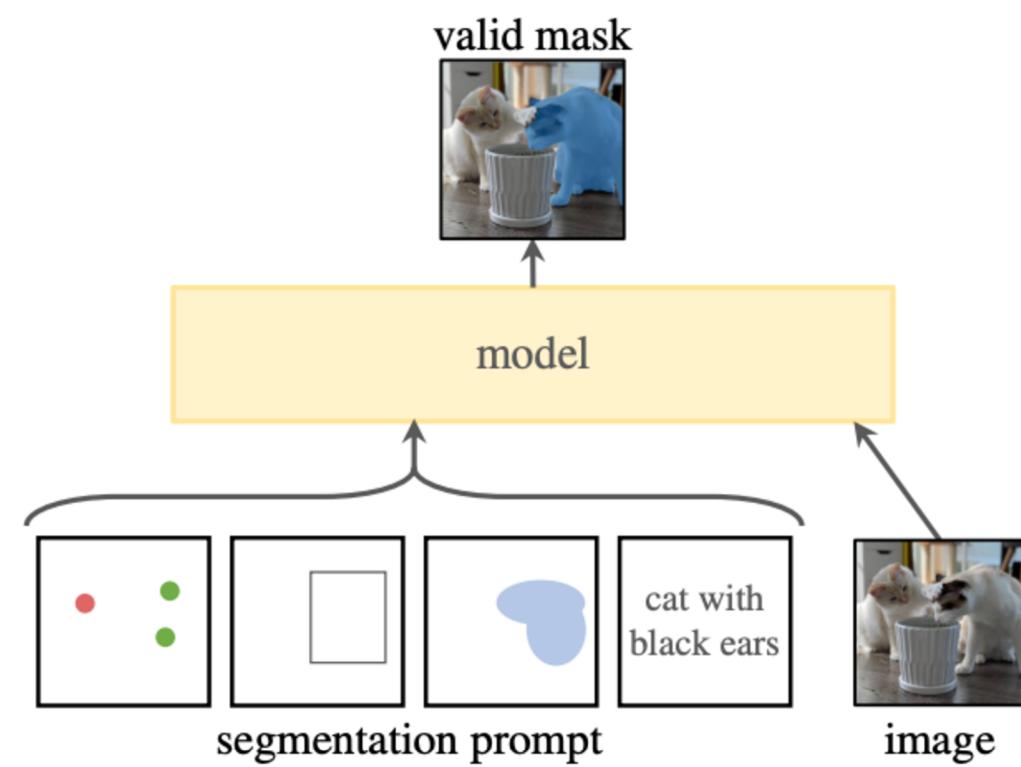
META AI RESEARCH - FAIR 2023

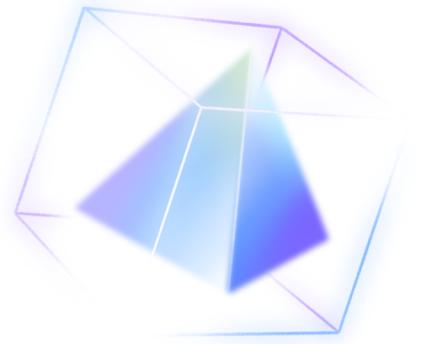
Presented By
Nyein Chan Soe

WHAT is SAM?

Segment Anything (SAM) is a full segmentation project consisting of a

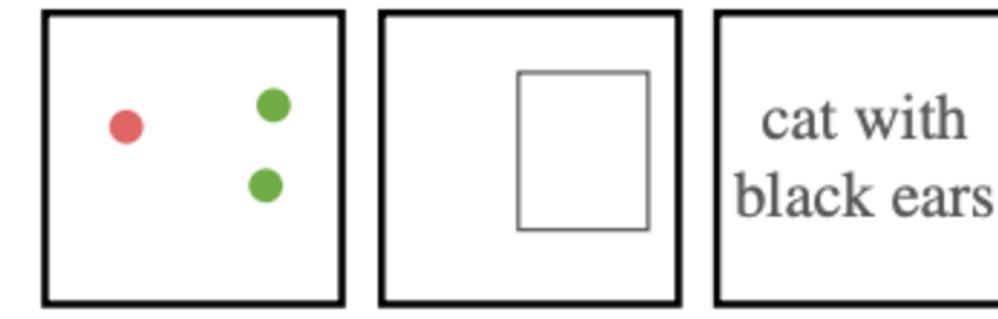
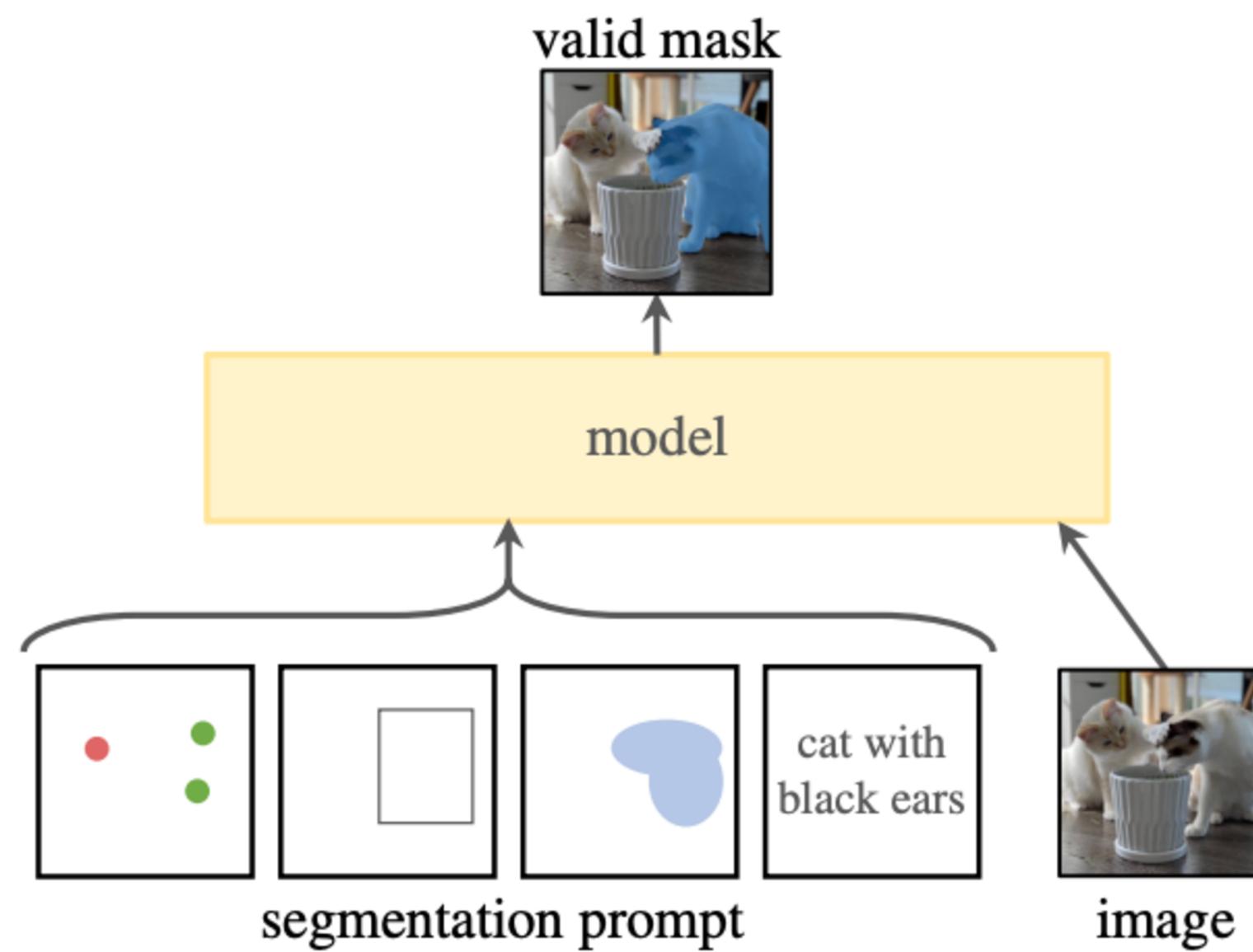
- Promptable Segmentation Task
- The SAM model
- Large-scale Data Engine





Promptable Task

User Inputs (prompts) are used to output segmentation masks

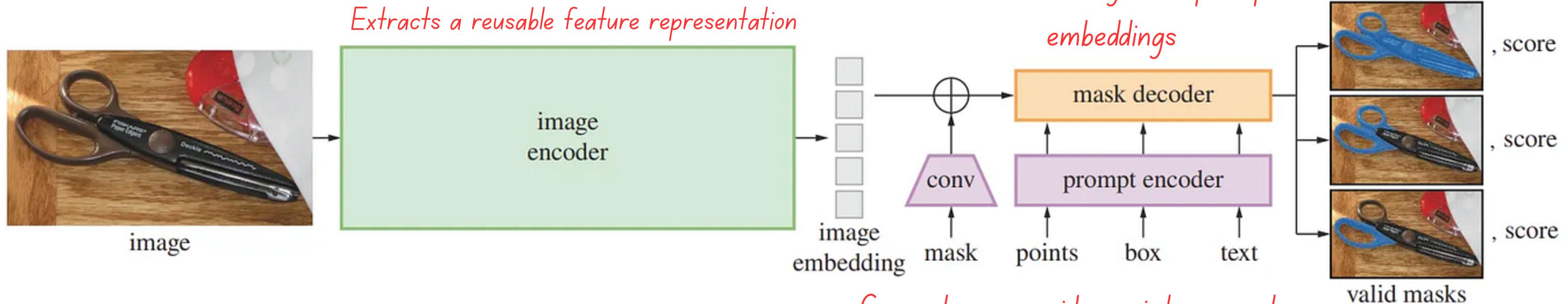


Sparse Inputs



Dense Inputs

SAM Model Arch

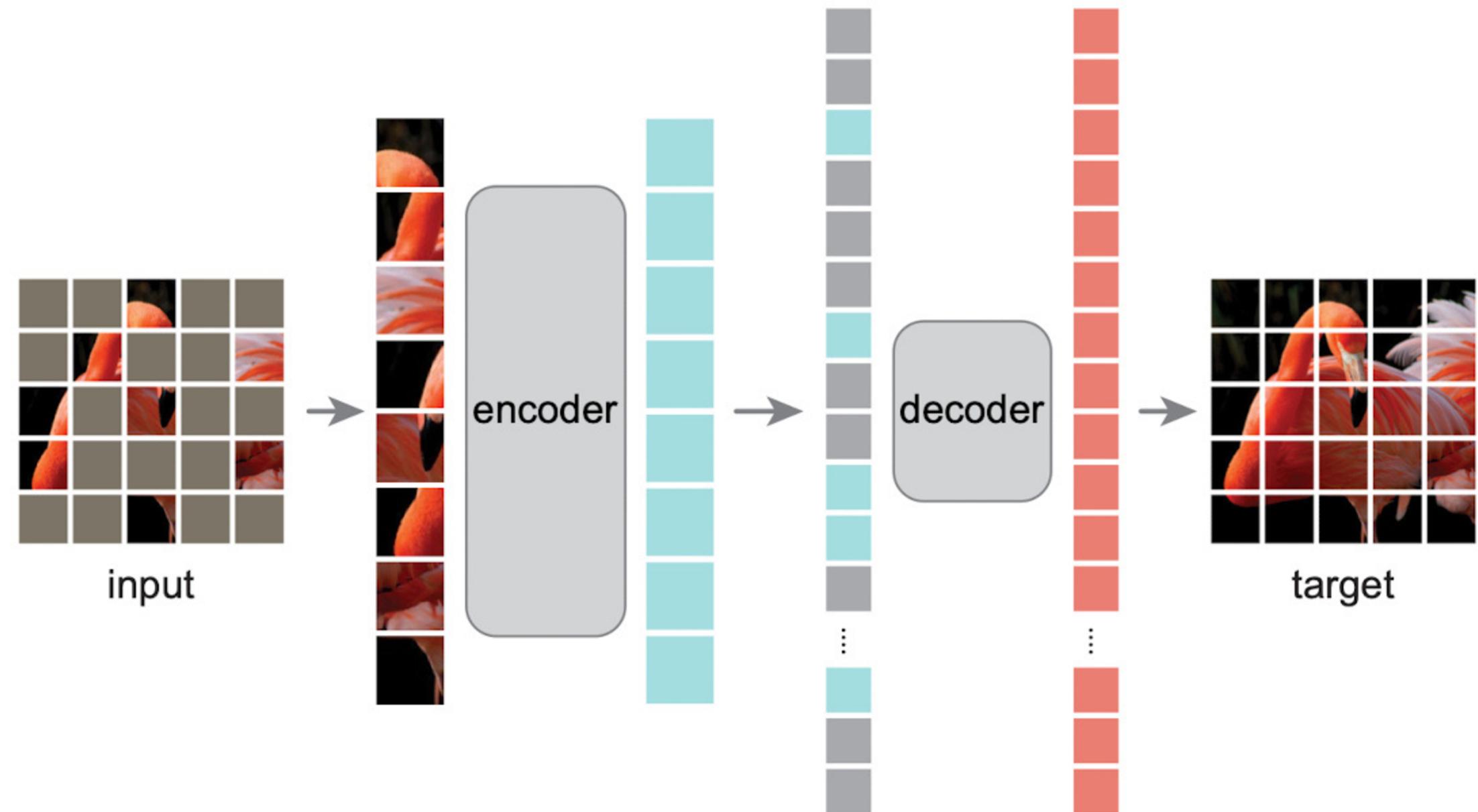


1. Image Encoder (ViT-based)
2. Prompt Encoder
3. Masked Decoder

Encoder pretraining idea

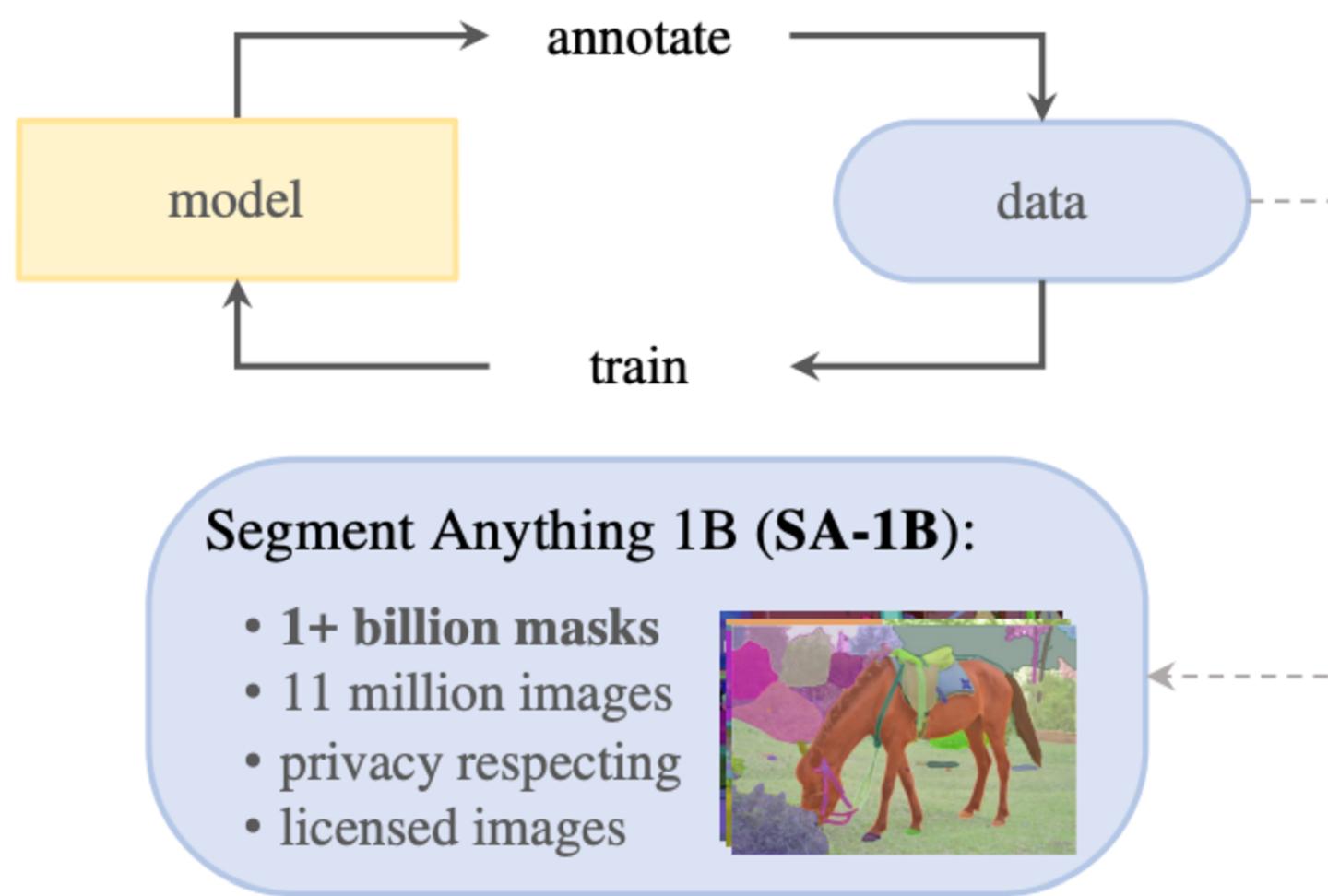
MAE to pretrain ViT
based Encoder

* Output From backbone is 16x
scaled down embeddings with
256 channels



Data Engine

If **SAM** is a factory that produce masks
Data Engine makes the factory improves itself



Why need it?

- Small segmentations datasets
- Labels are time consuming and expensive
- Doesn't scale much as classification labels

SA-1B

- Massive 11 million images with 1 billion masks
- Currently largest image segmentation dataset

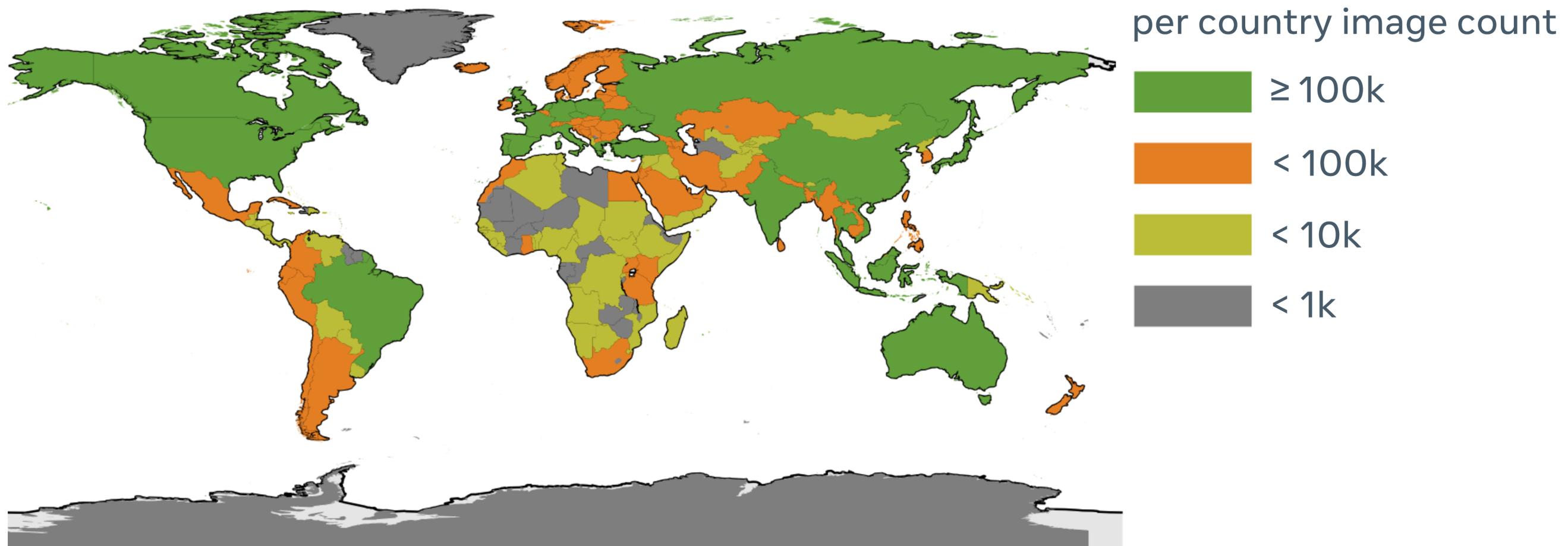
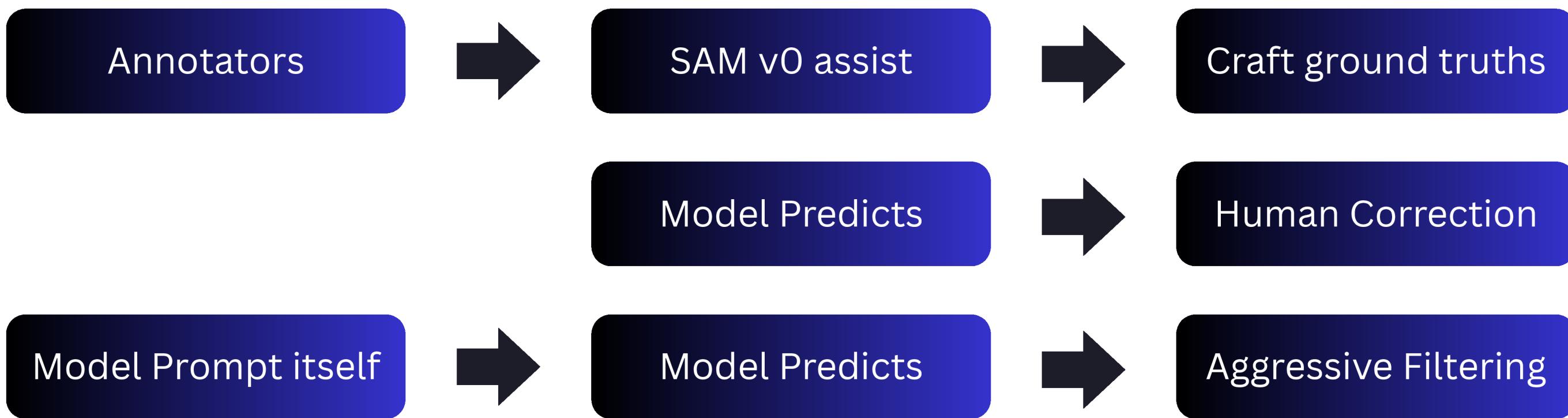


Figure 7: Estimated geographic distribution of SA-1B images. Most of the world's countries have more than 1000 images in SA-1B, and the three countries with the most images are from different parts of the world.

How SA-1B is created

- Model Assisted Manual Annotation
- Semi Automatic
- Fully Automatic



Quality Validations



Validate Quality of the Stage3 outputs by comparing
auto created masks vs professionally corrected masks

- 94% of auto-masks had > 90% IoU
- 97% of auto-masks had > 75% IoU

SAM's masks match expert masks about as well as humans match each other.

Evaluation Setup

Tested On

Zero-shot interactive segmentation across 23 diverse segmentation datasets

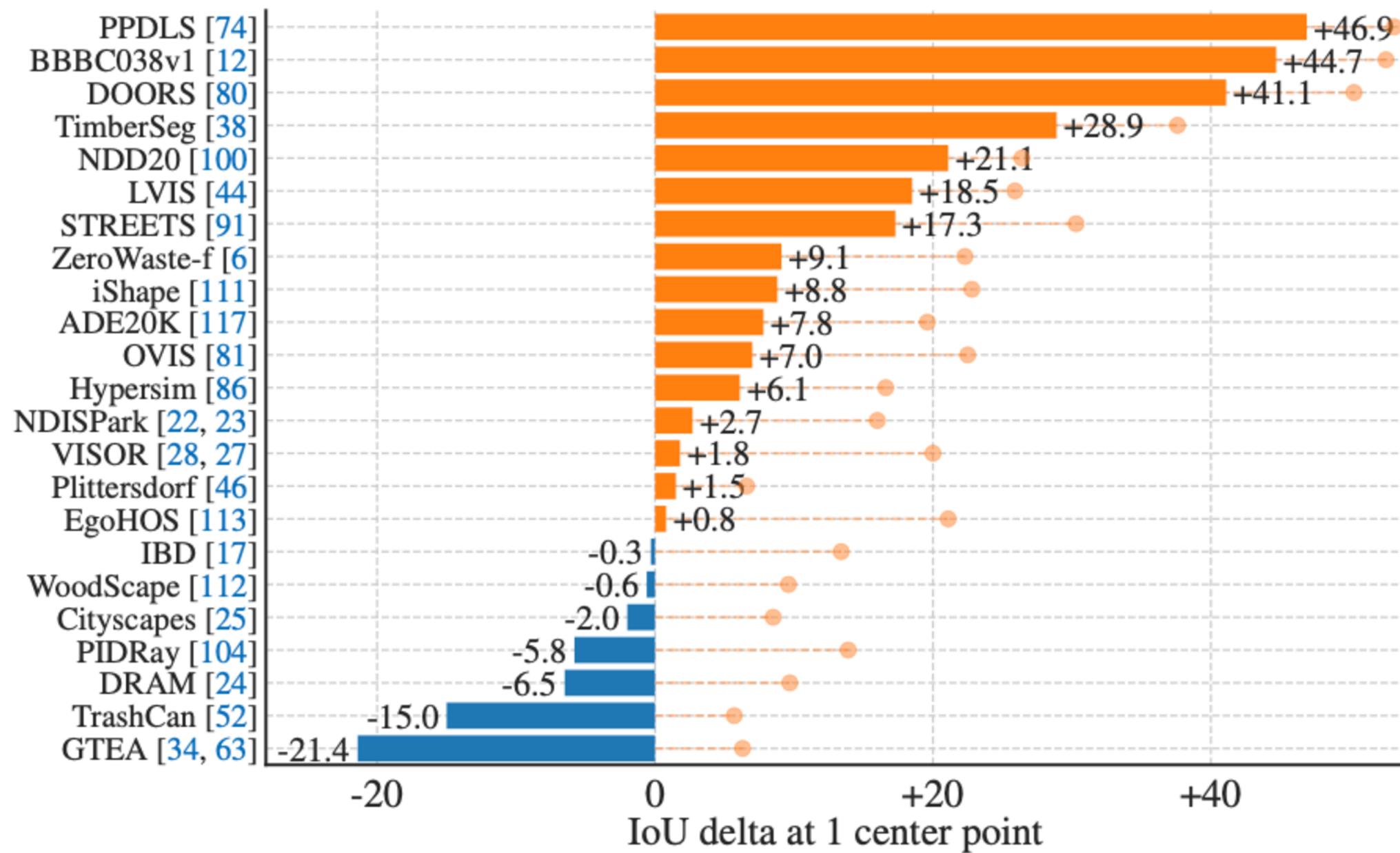
Compared Against

Strong interactive methods: **RITM**, **SimpleClick**, **FocalClick**

Metrics

mIoU and Human evaluations

Evaluation Results



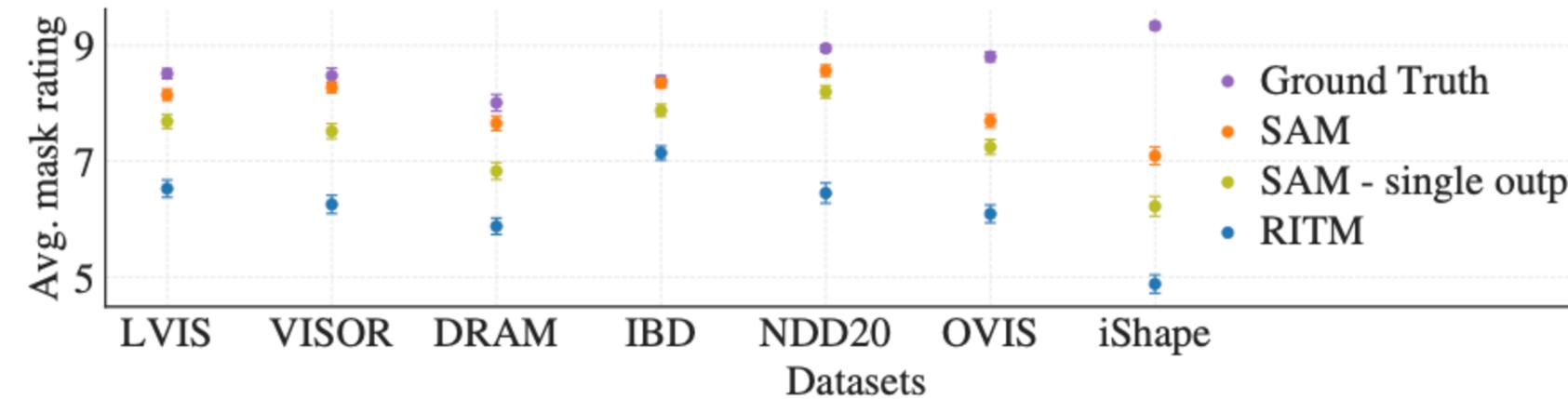
(a) SAM vs. RITM [92] on 23 datasets

Mean IoU of SAM and
the strongest single
point segmenter

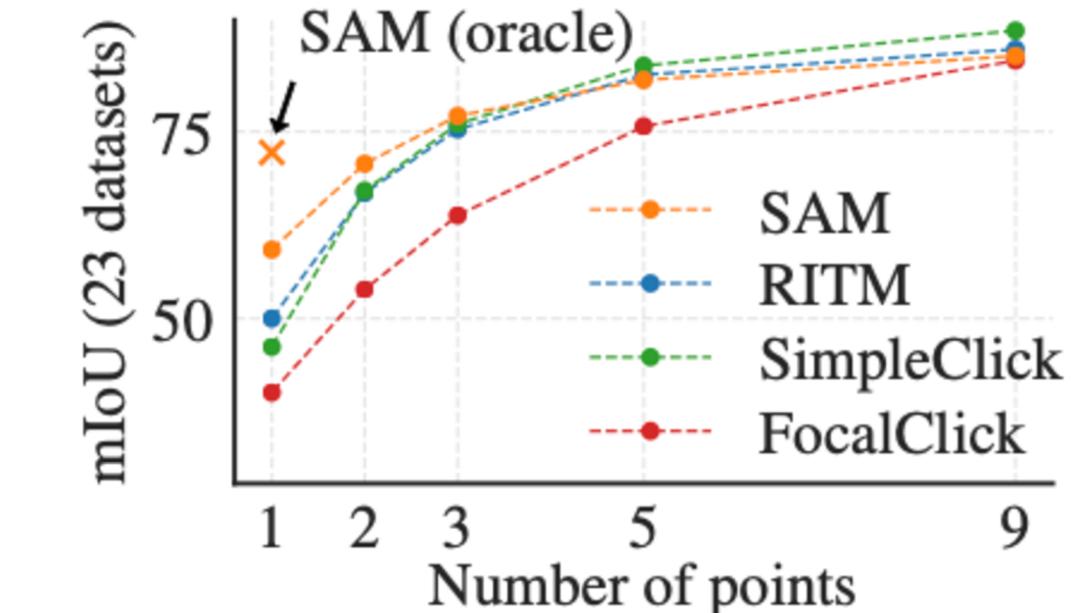
Evaluation Results

Center point: the click prompt is placed at the center

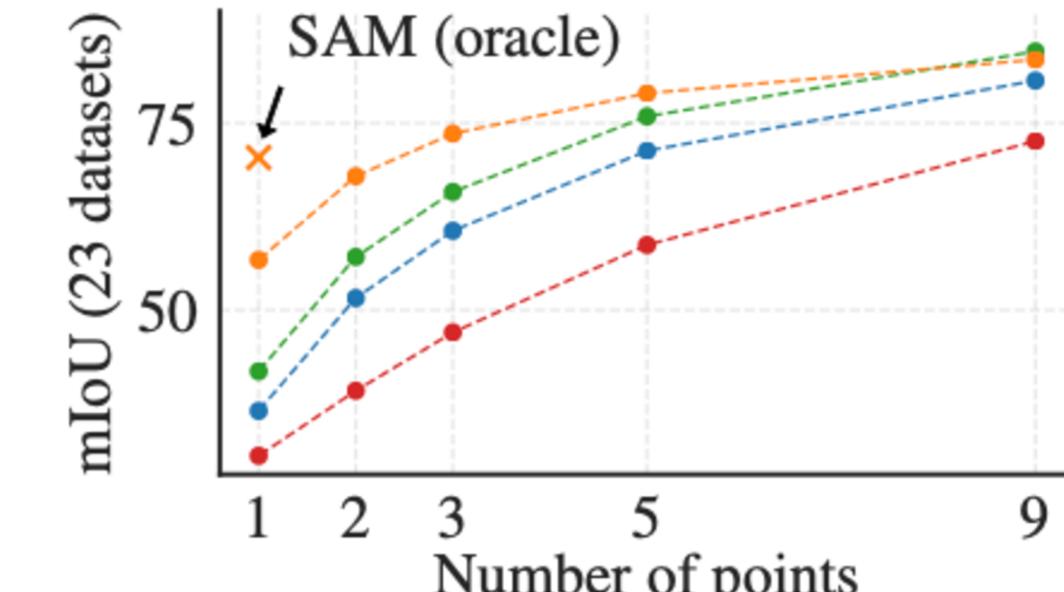
Random point: the click prompt is placed at a random location



(b) Mask quality ratings by human annotators



(c) Center points (default)



(d) Random points

Limitations of SAM1

- Can miss fine structures
- Sometimes hallucinates small disconnected regions
- Boundaries may be less crisp than “zoom-in” methods
- With many prompt points, specialized interactive methods can outperform SAM
- Text-to-mask is still exploratory / not fully robust
- Domain-specific tools may beat SAM in their niche domains

Thank You!

For your time