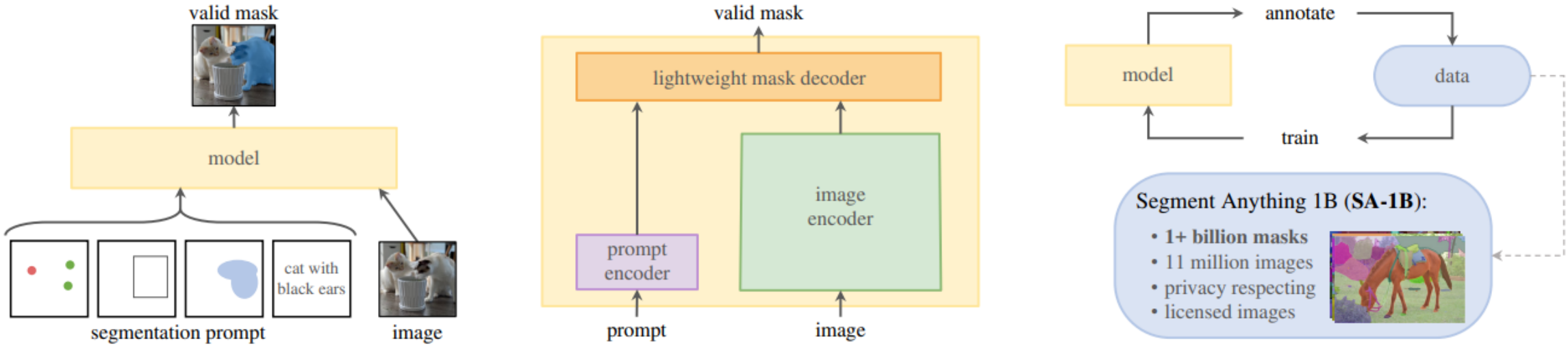
Segment Anything

Alexander Kirillov[1,2,4] Eric Mintun[2] Nikhila Ravi[1,2] Hanzi Mao[2] Chloe Rolland[3] Laura Gustafson[3]
Tete Xiao[3] Spencer Whitehead Alexander C. Berg Wan-Yen Lo Piotr Dollár[4] Ross Girshick[4]
[1]project lead [2]joint first author [3]equal contribution [4]directional lead

Meta AI Research, FAIR

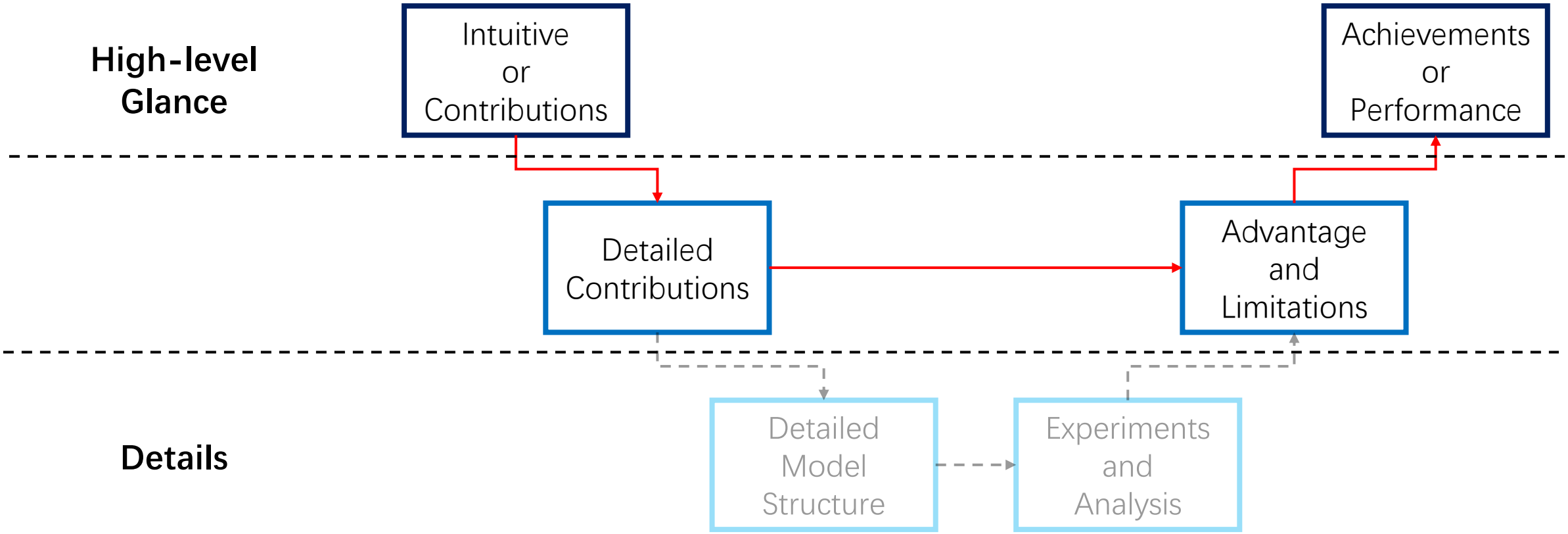


(a) **Task**: promptable segmentation (b) **Model**: Segment Anything Model (**SAM**) (c) **Data**: data engine (top) & dataset (bottom)

Figure 1: We aim to build a foundation model for segmentation by introducing three interconnected components: a prompt-able segmentation *task*, a segmentation *model* (SAM) that powers data annotation and enables zero-shot transfer to a range of tasks via prompt engineering, and a *data* engine for collecting SA-1B, our dataset of over 1 billion masks.

Segment Anything | Meta AI (segment-anything.com)

Contents



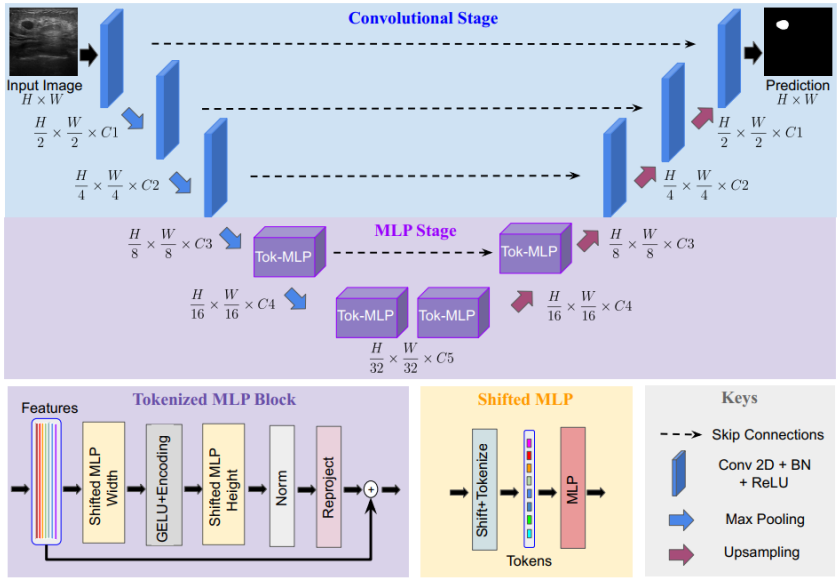
Segment Anything | Meta AI (segment-anything.com)

Contributions

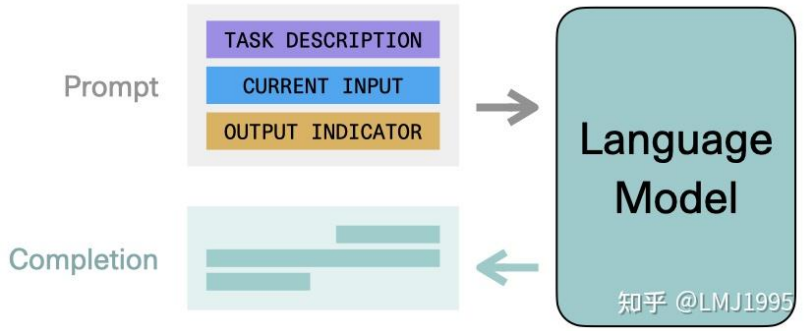
1. A promptable segmentation **TASK**
2. A **MODEL** that supports flexible prompting and can output segmentation masks in real-time when prompted to allow for interactive use
3. A diverse, large-scale source of **DATA**

Segment Anything | Meta AI (segment-anything.com)

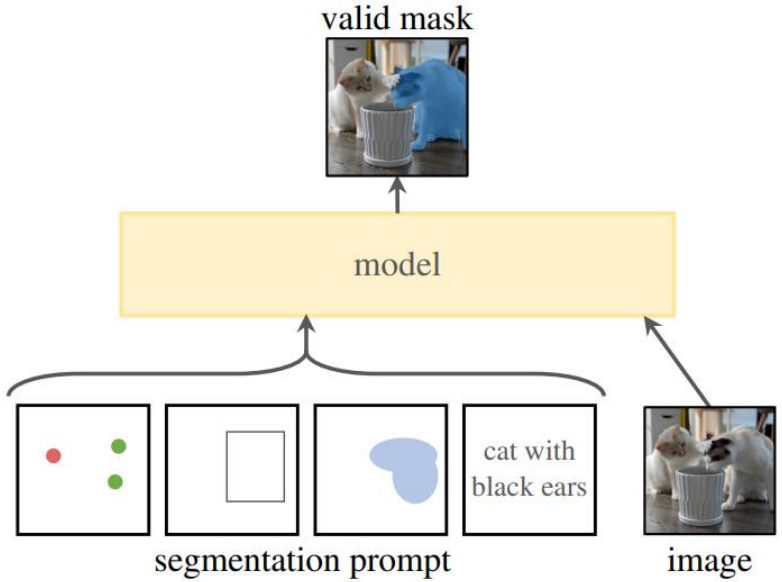
Original vs Promptable Segmentation Task



UNeXt Segmentation



Prompt Engineering

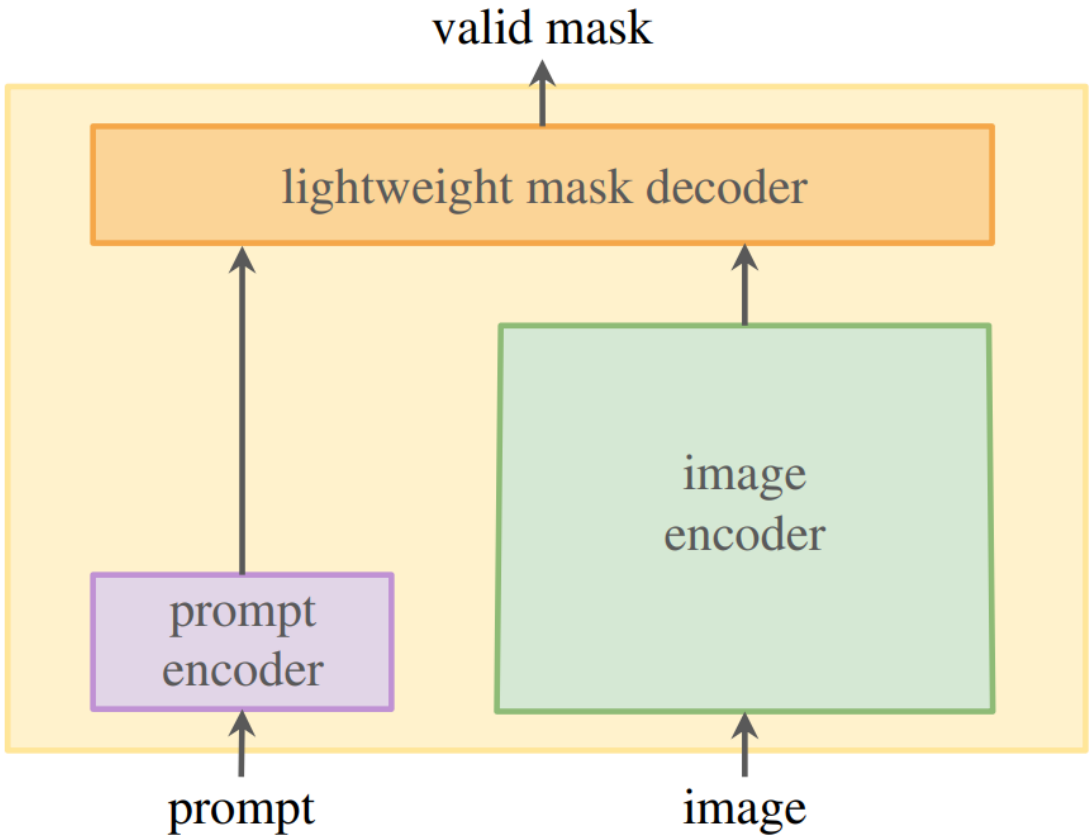


Prompt Segmentation

- Inspired by large language models pretrained on web-scale datasets, which can generalize to unseen data (Zero-shot / Few-shot)
- This capability is implemented with prompt engineering

Segment Anything | Meta AI (segment-anything.com)

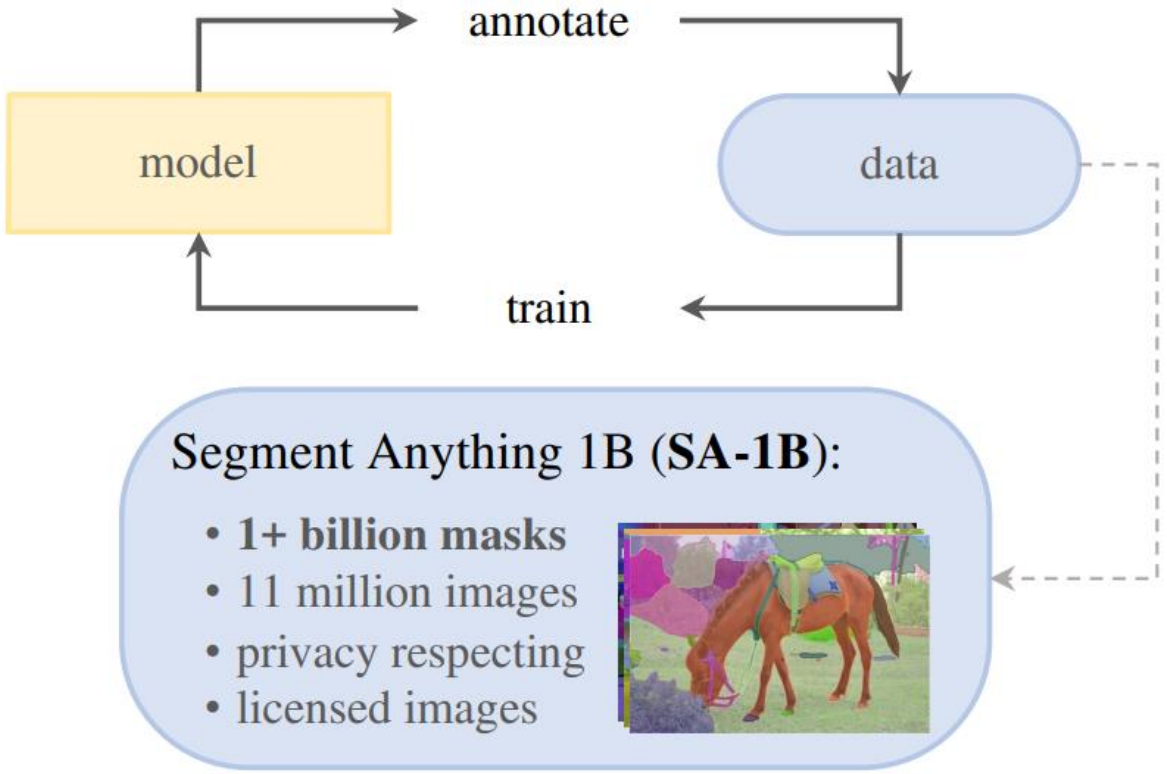
New Model for Prompt Segmentation



(b) **Model**: Segment Anything Model (**SAM**)

Segment Anything | Meta AI (segment-anything.com)

Data Engine + Dataset (SA-1B)



(c) **Data**: data engine (top) & dataset (bottom)

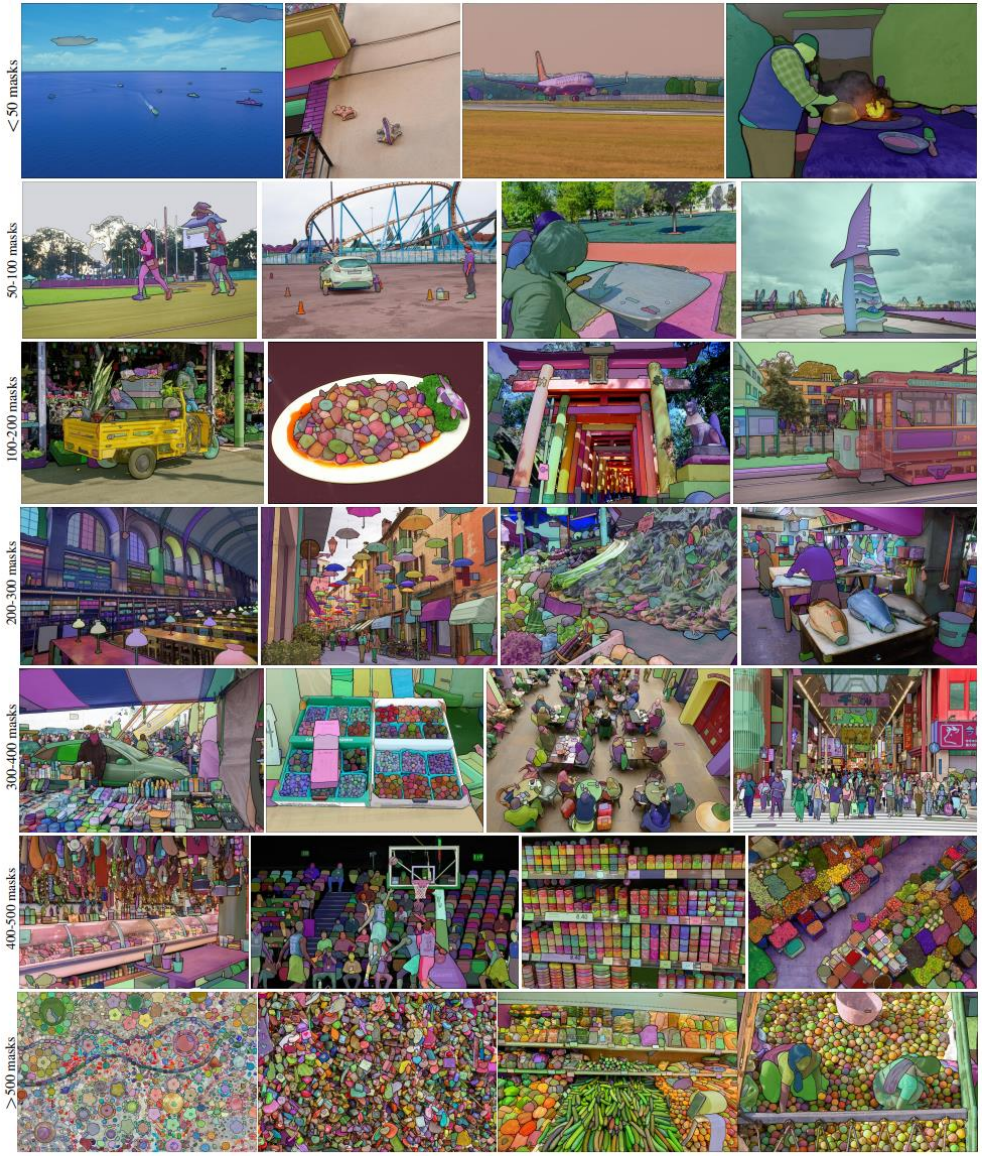


Figure 2: Example images with overlaid masks from our newly introduced dataset, **SA-1B**. SA-1B contains 11M diverse, high-resolution, licensed, and privacy protecting images and 1.1B high-quality segmentation masks. These masks were annotated *fully automatically* by SAM, and as we verify by human ratings and numerous experiments, are of high quality and diversity. We group images by number of masks per image for visualization (there are ~100 masks per image on average).

Segment Anything | Meta AI (segment-anything.com)

Demo

[Segment Anything | Meta AI (segment-anything.com)]

Promptable Segmentation Task

- Prompt:
  1. Foreground / Background Points
  2. Rough box or masks
  3. Free-form Text
- Advantage:
  1. Zero-shot transfer
  2. Generalization

Segment Anything | Meta AI (segment-anything.com)

Segment Anything Model

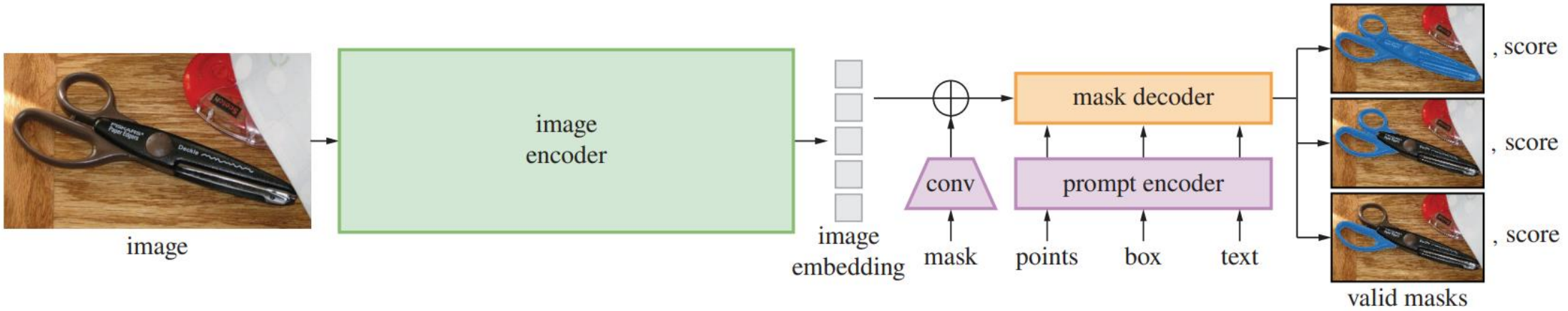


Figure 4: Segment Anything Model (SAM) overview. A heavyweight image encoder outputs an image embedding that can then be efficiently queried by a variety of input prompts to produce object masks at amortized real-time speed. For ambiguous prompts corresponding to more than one object, SAM can output multiple valid masks and associated confidence scores.

- **Model Structure**:

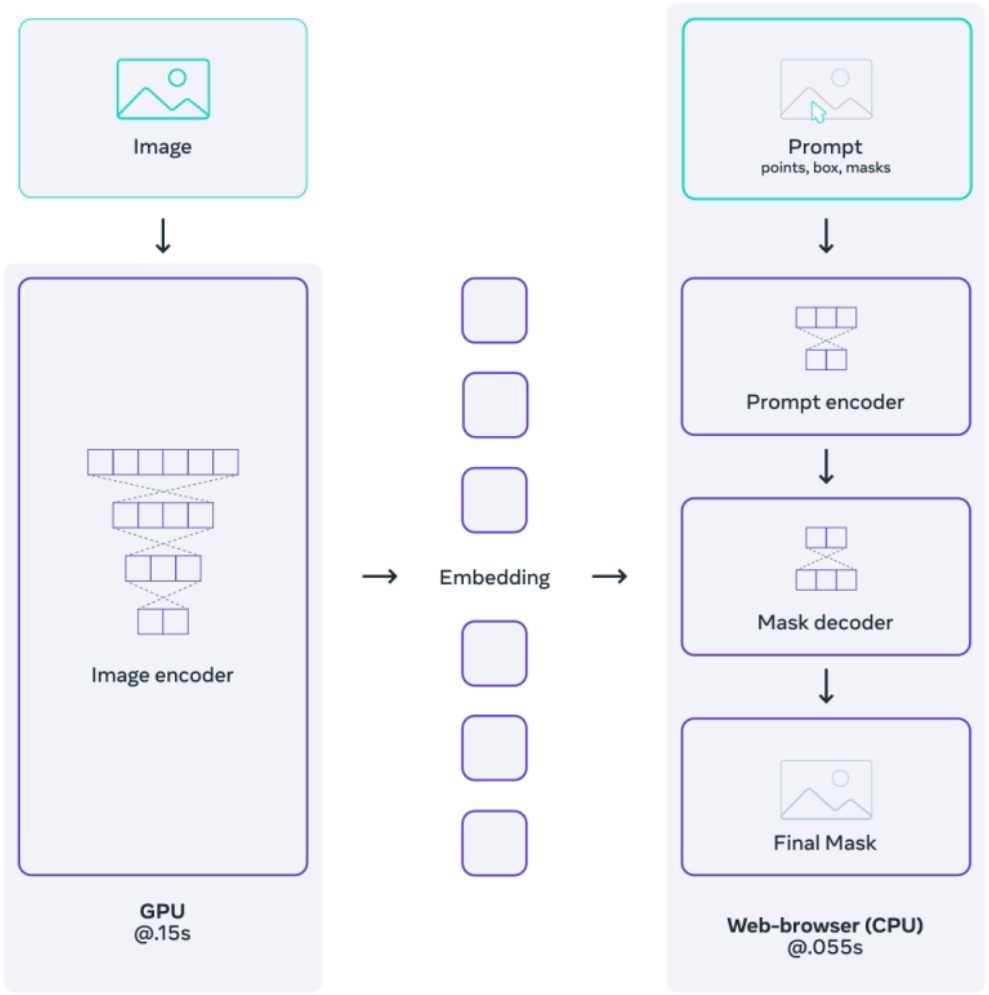
1. **Image Encoder**: MAE pre-trained ViT
2. **Prompt Encoder**
3. **Mask Decoder**: Transformer Decoder block
   followed by a dynamic mask prediction head

- **Prompt Encoder**:

1. Point and boxes: Positional Encoding
2. Text: Text Encoder in CLIP
3. Mask: Convolutions

Segment Anything | Meta AI (segment-anything.com)

Segment Anything Model



- **Ambiguity:**

Predict 3 mask outputs for nested mask
(whole, part, and subpart)

- **Efficiency:**

Precompute image embedding
Pe+md run on CPU in **~50ms**

- **Loss:**

Supervise Focal Loss + Dice Loss

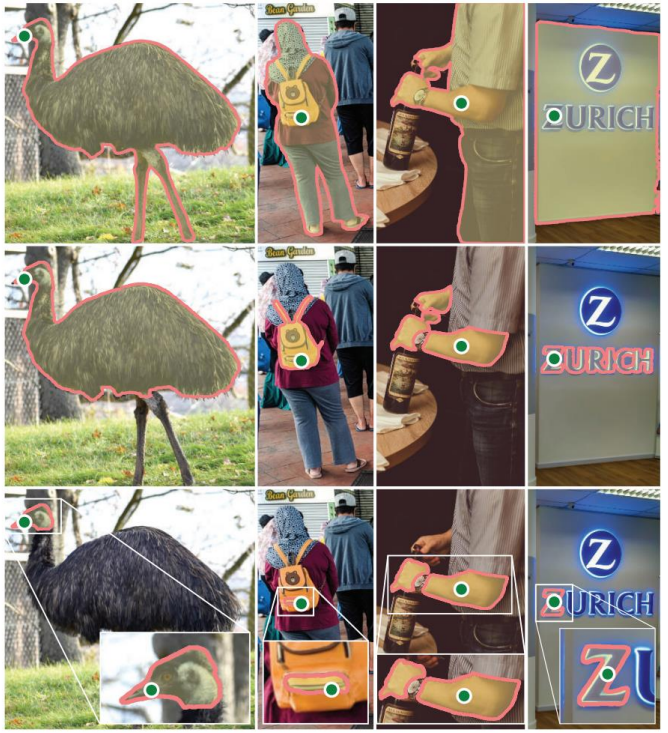
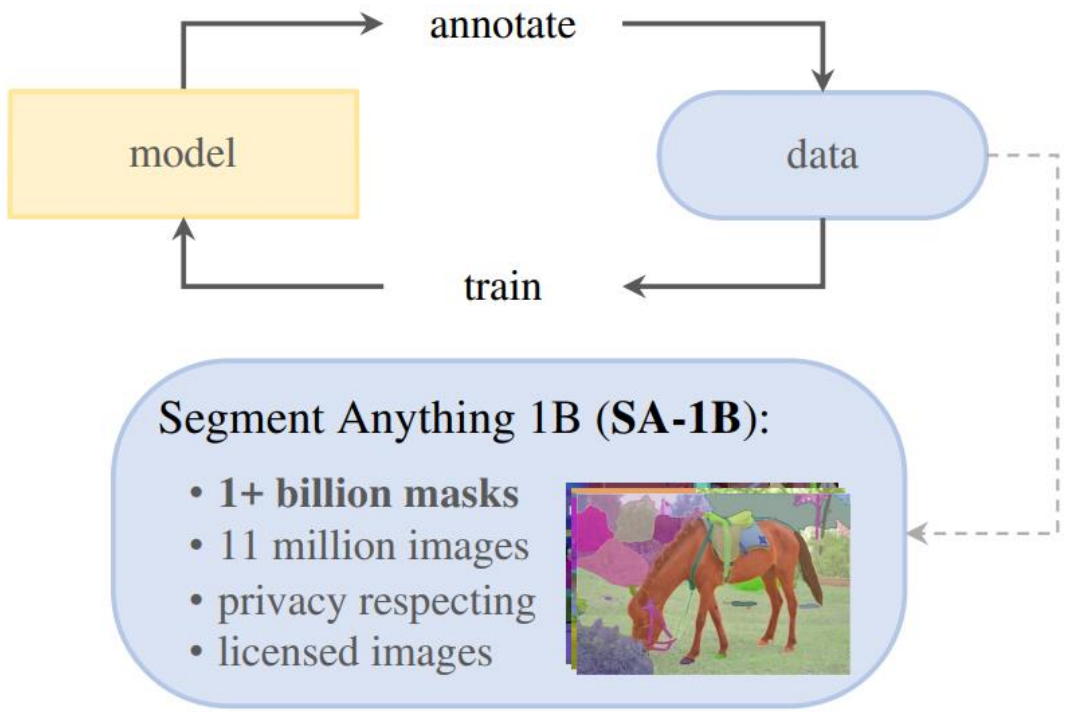


Figure 3: Each column shows 3 valid masks generated by SAM from a single ambiguous point prompt (green circle).

[Segment Anything | Meta AI (segment-anything.com)]

Data Engine



(c) **Data**: data engine (top) & dataset (bottom)

- **Assisted-manual stage:**

Training -> give new annotations -> retraining -> give new annotations -> retraining -> ⋯ (6 times)

- **Semi-automatic stage:**

Detect confident masks -> Add mask for unannotated objects -> retraining -> ⋯ (5 times)

- **Fully automatic stage:**

Cropping + Filtering + Postprocessing
Generating best mask for **SA-1B**
11M images with 1.1B masks

Segment Anything | Meta AI (segment-anything.com)

SA-1B Dataset

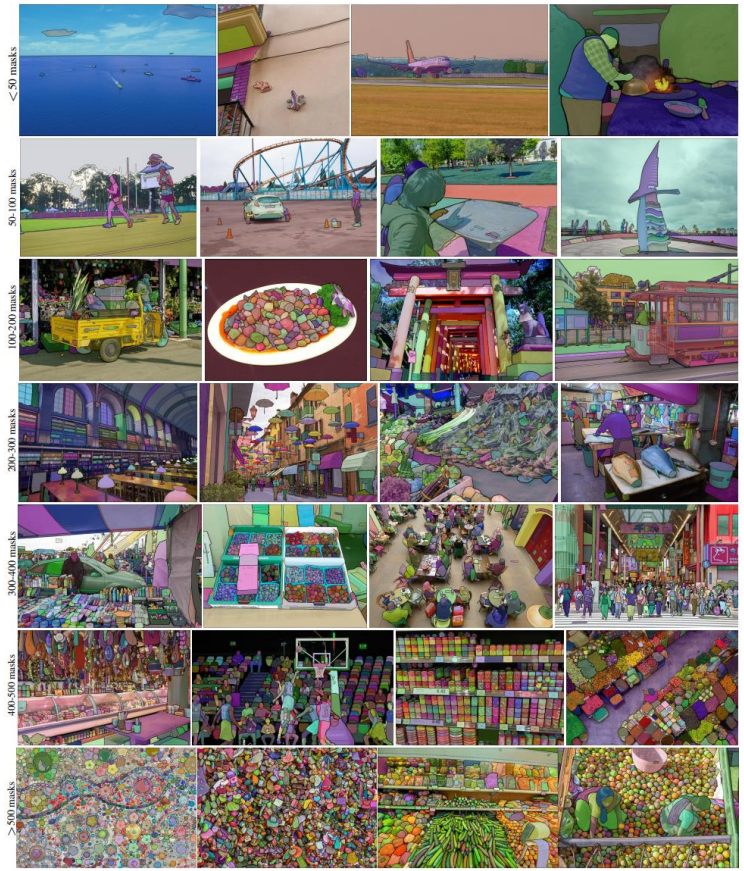


Figure 2: Example images with overlaid masks from our newly introduced dataset, **SA-1B**. SA-1B contains 11M diverse, high-resolution, licensed, and privacy protecting images and 1.1B high-quality segmentation masks. These masks were annotated *fully automatically* by SAM, and as we verify by human ratings and numerous experiments, are of high quality and diversity. We group images by number of masks per image for visualization (there are ~100 masks per image on average).

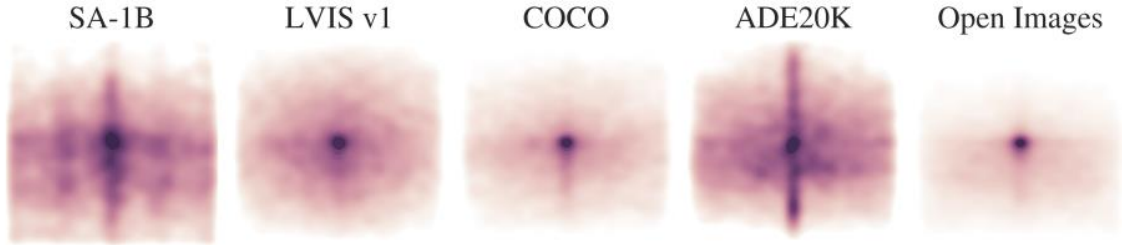


Figure 5: Image-size normalized mask center distributions.

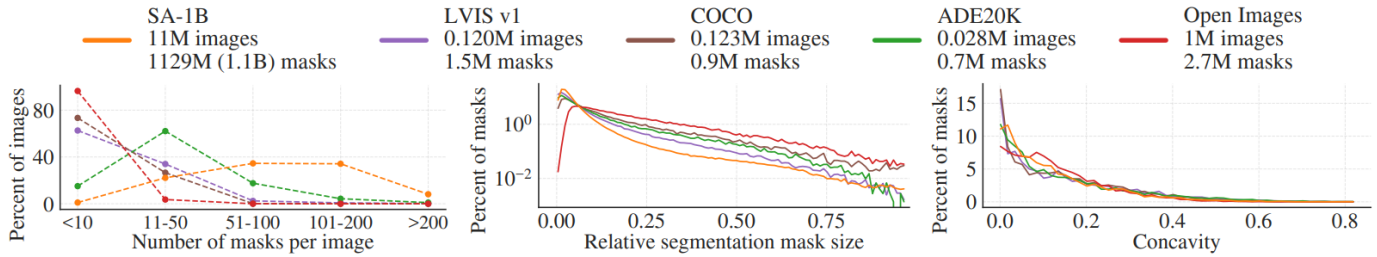


Figure 6: Dataset mask properties. The legend references the number of images and masks in each dataset. Note, that SA-1B has 11× more images and 400× more masks than the largest existing segmentation dataset Open Images [60].

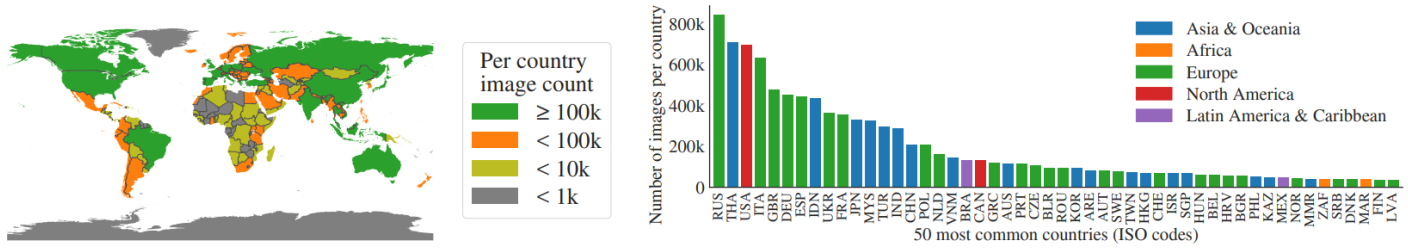


Figure 7: Estimated geographic distribution of SA-1B images. Most of the world's countries have more than 1000 images in SA-1B, and the three countries with the most images are from different parts of the world.

Segment Anything | Meta AI (segment-anything.com)

Discussion

- **Limitations**

1. Can miss fine structures, small components
2. Interactive seg can outperform when many points are provided
3. Text-to-mask is exploratory and not entirely robust
4. Unclear to define prompt in semantic and panoptic segmentation （语义和全景分割）

- **Conclusions**

1. Attempt to lift image segmentation into to the era of foundation models.
2. Contributions: **a new task (promptable segmentation), model (SAM), and dataset (SA-1B)** that make this leap possible.
3. Whether SAM achieves the status of a foundation model remains to be seen by how it is used in the community, but regardless we expect the perspective of this work, the release of over 1B masks, and our promptable segmentation model will help pave the path ahead

Segment Anything | Meta AI (segment-anything.com)