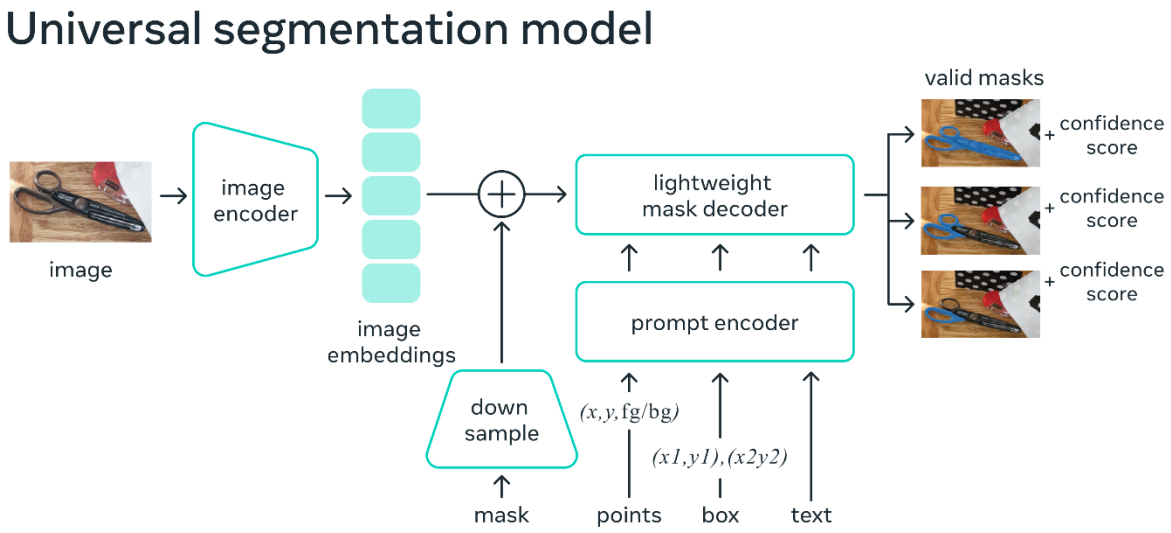
Paper Review

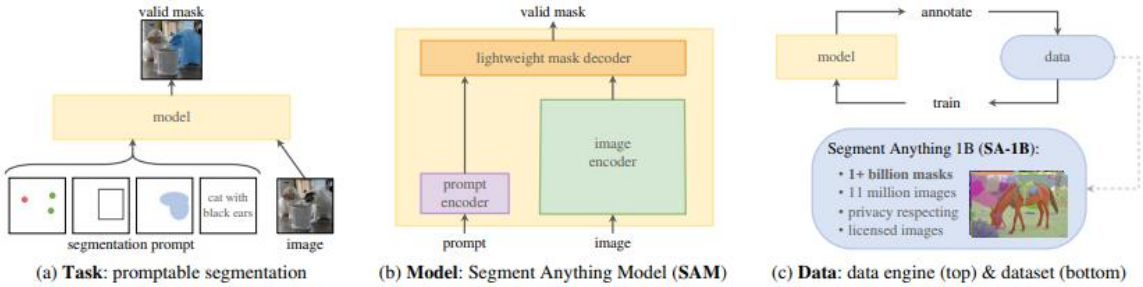
SAM - Segment Any Model

- Meta's Segment Anything Model (SAM) is the latest breakthrough in computer vision. The team created a foundation model for image segmentation that can adapt to various downstream tasks using *prompt engineering*.

SAM's Network Architecture and Design



1. The promptable segmentation task to enable zero-shot generalization.
2. The model architecture.
3. The dataset that powers the task and model.



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

Promptable Task

SAM was trained on millions of images and over a billion masks to return a valid segmentation mask for any prompt. The prompt, in this case, is the segmentation task and can be foreground/background points, a rough box or mask, clicks, text, or, in general, any information indicating what to segment in an image. The task is also used as the pre-training objective for the model.

Model

SAM's architecture comprises three components that work together to return a valid segmentation mask:

- An image encoder to generate one-time image embeddings.
- A prompt encoder that embeds the prompts.
- A lightweights mask decoder that combines the embeddings from the prompt and image encoders.

Data Engine and Dataset

A data engine is needed to power the tasks and improve the dataset and model. The data engine has three stages:

- Assisted-manual, where SAM assists annotators in annotating masks, similar to a classic interactive segmentation setup.
- Semi-automatic, where SAM can automatically generate masks for a subset of objects by prompting it with likely object locations, and annotators focus on annotating the remaining objects, helping increase mask diversity.
- Fully automatic, where human annotators prompt SAM with a regular grid of foreground points, yielding on average 100 high-quality masks per image.

The data engine builds the large segment anything 1-billion mask dataset Meta AI released.

Why the paper and model is interesting to me?

- SAM can segment objects by simply clicking or interactively selecting points to include or exclude from the object. We can also create segmentations by drawing bounding boxes or segmenting regions with a polygon tool and it will snap to the object.

- SAM has the ability to identify and generate masks for all objects present in an image automatically.

References

https://github.com/facebookresearch/segment-anything

https://segment-anything.com/demo

https://ai.facebook.com/blog/segment-anything-foundation-model-image-segmentation/

https://arxiv.org/abs/2304.02643