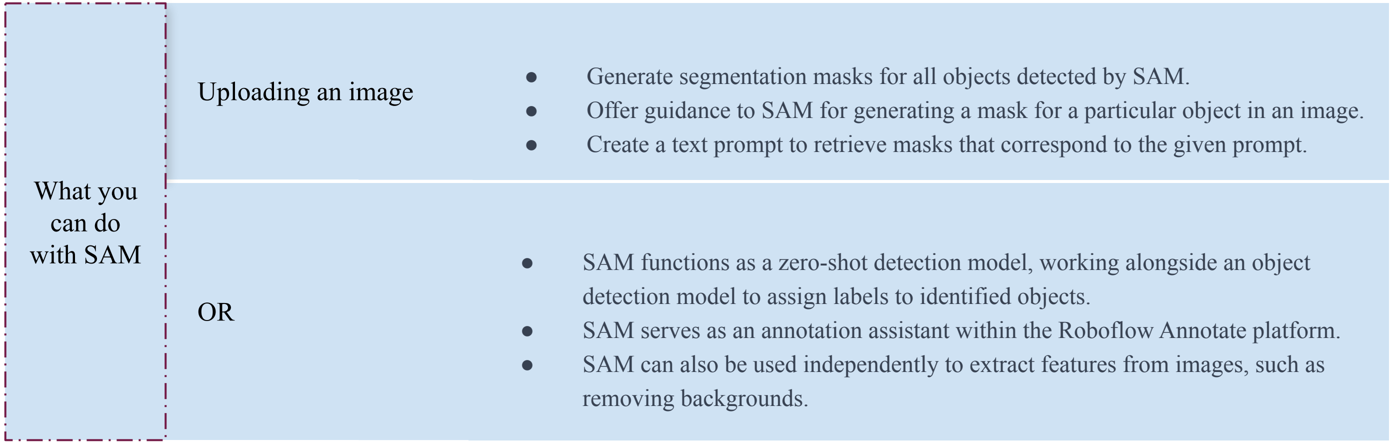
Segment Anything

By Meta Research

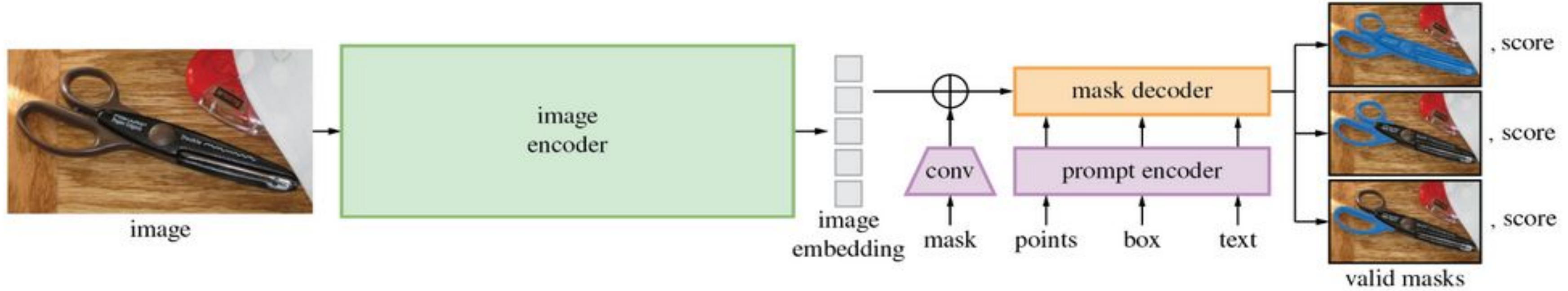
What SAM is about ?

- The Segment Anything Model (SAM) is an instance segmentation model developed by Meta Research and released in April, 2023 [1].



What you can do with SAM	Uploading an image	● Generate segmentation masks for all objects detected by SAM. ● Offer guidance to SAM for generating a mask for a particular object in an image. ● Create a text prompt to retrieve masks that correspond to the given prompt.
	OR	● SAM functions as a zero-shot detection model, working alongside an object detection model to assign labels to identified objects. ● SAM serves as an annotation assistant within the Roboflow Annotate platform. ● SAM can also be used independently to extract features from images, such as removing backgrounds.

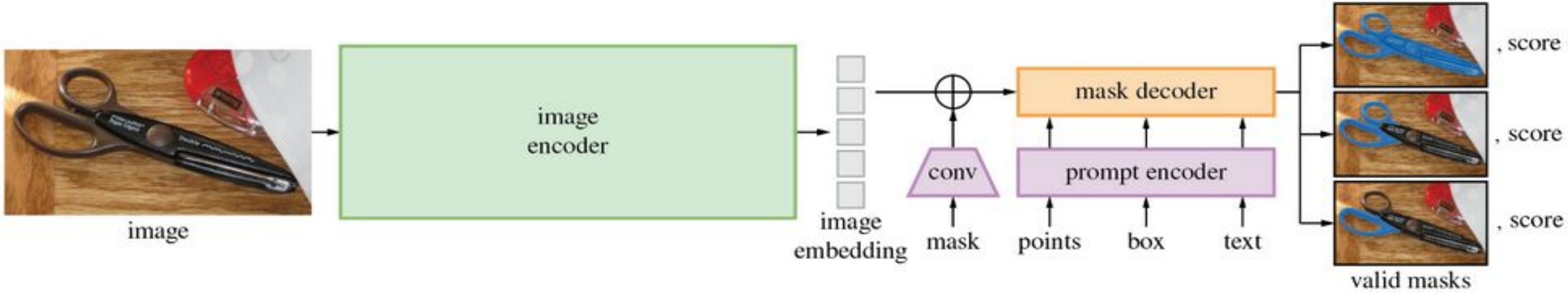
Architecture of SAM



[2] Rath, S. (2023) *Segment anything – a foundation model for image segmentation*, *LearnOpenCV*. Available at: https://learnopencv.com/segment-anything/ (Accessed: 27 May 2023).

- The image undergoes encoding by an image encoder, generating a single embedding representing the entire image. Additionally, there exists a prompt decoder for points, boxes, or text as prompts.
- For point prompts, the x and y coordinates, along with foreground and background information, are fed into the encoder. For box prompts, the bounding box coordinates serve as input to the encoder. As for text prompts (not available at the time of writing this), the tokens are used as input.
- If a mask is provided as input, it directly undergoes downsampling via 2D convolutional layers. Subsequently, the model concatenates this downscaled mask with the image embedding to obtain the final vector representation.
- Any vector derived from the prompt vector plus the image embedding then passes through a lightweight decoder, responsible for generating the ultimate segmentation mask. The output includes possible valid masks along with confidence scores.

Architecture of SAM



[2] Rath, S. (2023) *Segment anything – a foundation model for image segmentation*, *LearnOpenCV*. Available at: https://learnopencv.com/segment-anything/ (Accessed: 27 May 2023).

- The image encoder in SAM is based on a pre-trained Vision Transformer model, specifically the MAE architecture. This encoder plays a crucial role in extracting relevant features from the input image.
- Regarding the prompt encoder, different types of prompts are handled differently. Points and bounding boxes are treated as sparse inputs and are encoded using positional encodings, which are then combined with learned embeddings. For text prompts, SAM utilizes the text encoder from CLIP (Contrastive Language-Image Pretraining). In the case of masks as prompts, downsampling occurs through convolutional layers, and the resulting embedding is element-wise summed with the input image embedding to incorporate the mask information.

Secret in the “Recipe”

- The dataset on which a deep learning model is trained forms the foundation of any groundbreaking model, and the Segmentation Anything Model (SAM) is no exception.
- Segment Anything was trained on 11 million images and 1.1 billion segmentation masks. Its final dataset is called SA-1B dataset. Such a dataset is surely needed to train a model of Segment Anything capability. But we also know that such datasets do not exist, and manually annotating so many images is impossible.

The authors of the Segment Anything dataset product their dataset through three stages:

1. **Assisted Manual**: Annotators annotate alongside SAM to pick all masks in an image.
2. **Semi-Automatic**: Annotators are asked to only annotate masks for which SAM cannot render a confident prediction.
3. **Full-Auto**: SAM is allowed to fully predict masks given its ability to sort out ambiguous masks via a full sweep.

In the first stage, the annotators used a pretrained SAM model to interactively segment objects in images in the browser. The image embeddings were precomputed to make the annotation process seamless and real-time. After the first stage, the dataset consisted of 4.3 million masks from 120k images. The Segment Anything Model was retrained on this dataset.

In the final ‘fully automatic stage’, the annotation was entirely done by SAM, By this stage, it had already been trained on more than 10M masks which made it possible. Automatic mask generation was applied on 11M images resulting in 1.1B masks.

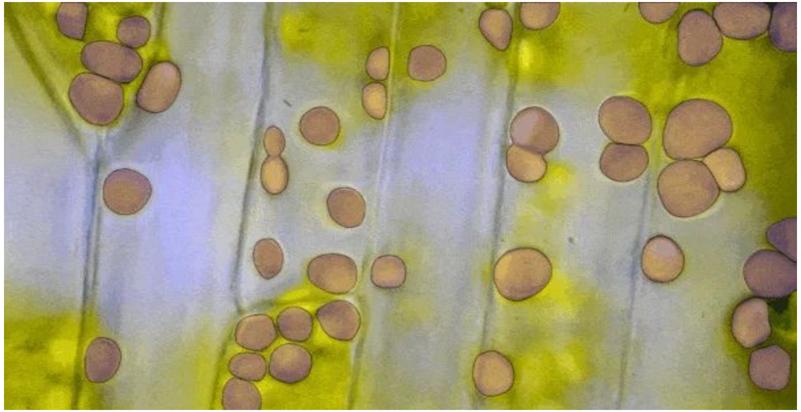
Promises of SAM

- **Prompt-Based Segmentation Task**: SAM is specifically designed to excel in prompt-based segmentation tasks, enabling it to generate accurate segmentation masks in response to various prompts, including spatial or text-based cues that identify specific objects.
- **Advanced Architecture**: SAM leverages a sophisticated architecture comprising an efficient image encoder, a prompt encoder, and a lightweight mask decoder. This architectural design empowers SAM with the ability to handle diverse prompts, compute masks in real-time, and exhibit awareness of ambiguities in segmentation.
- **SA-1B Dataset**: The Segment Anything project introduces the remarkable SA-1B dataset, consisting of over 11 million images and more than 1 billion masks. This dataset stands as the largest and most comprehensive segmentation dataset available, providing SAM with an extensive and diverse training data source.
- **Zero-Shot Performance**: SAM showcases exceptional zero-shot performance across a wide range of segmentation tasks. This allows SAM to be readily employed in various applications without the need for extensive customization, thanks to its prompt engineering capabilities.

Use Cases of SAM

- **Augmented Reality:** SAM's ability to identify common objects stands as a significant milestone in the area of augmented reality.
- **Bio-Medical Image Segmentation:** SAM excels in the challenging task of segmenting medical images and cell microscopy. It can effortlessly assist in segmenting cell-microscopy images without requiring any retraining.
- **Integration with Diffusion Models:** SAM can be seamlessly integrated with diffusion-based image generation models, automating various tasks, including mask creation during inpainting processes. This integration streamlines the workflow and enhances efficiency.

Carla's Icing Tips
3 Steps for Better Icing
Icing Made Simple
Not actual product footage. Footage is strictly for illustrative purposes only.



[2] Rath, S. (2023) *Segment anything – a foundation model for image segmentation*, *LearnOpenCV*. Available at: https://learnopencv.com/segment-anything/ (Accessed: 27 May 2023).

References

[1] Rath, S. (2023) *Segment anything – a foundation model for image segmentation*, *LearnOpenCV*. Available at:

https://learnopencv.com/segment-anything/ (Accessed: 27 May 2023).

[2] Rath, S. (2023) *Segment anything – a foundation model for image segmentation*, *LearnOpenCV*. Available at:

https://learnopencv.com/segment-anything/ (Accessed: 27 May 2023).

[3]Solawetz, J. (2023) *What is segment anything (Sam)? the ultimate guide*, *Roboflow Blog*. Available at:

https://blog.roboflow.com/segment-anything-breakdown/ (Accessed: 27 May 2023).

[4] *Ai's New breakthrough: Segment anything model (Sam)* (2023) *AIFT*. Available at:

https://hkaift.com/ais-new-breakthrough-segment-anything-model-sam/ (Accessed: 27 May 2023).