Segment Anything Model

Meta AI Research @ ICCV’23

Presented by

Chenhui Zhao | chuizhao@umich.edu

Outline

1. Recap image segmentation task
2. Introduce Segment Anything Model
3. Experiment results
4. SAM in medical image analysis

Image segmentation tasks

Semantic Segmentation[1]

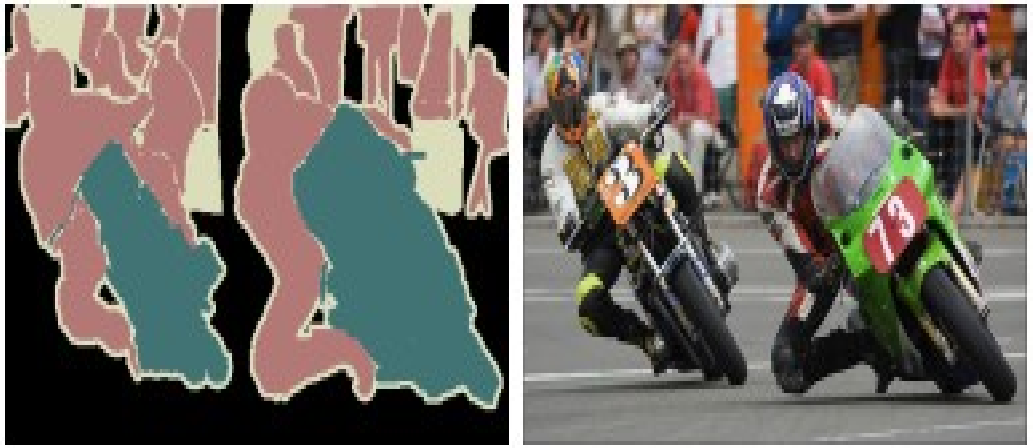
Instance

Panoptic Segmentation[3]

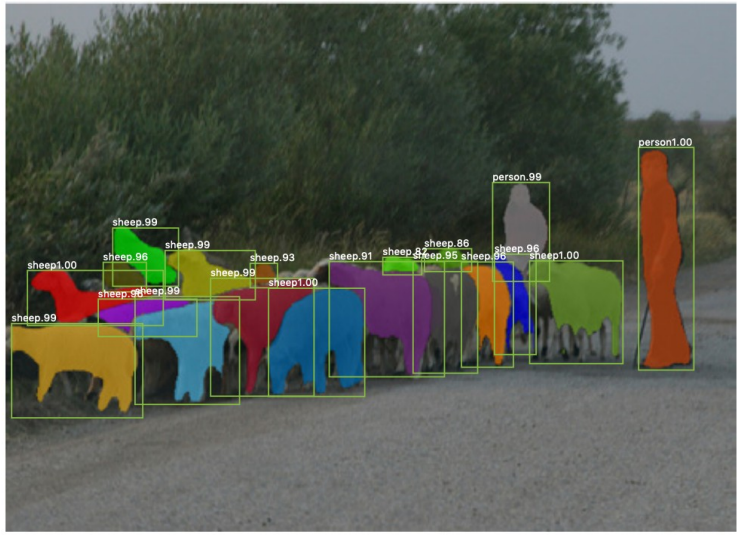
1. *Fully Convolutional Networks for Semantic Segmentation (CVPR'2015)*
2. *Mask-RCNN (CVPR' 2017)*
3. *Per-Pixel Classification is Not All You Need for Semantic Segmentation (NeurIPS' 2021)*

Semantic Segmentation and Instance Segmentation

Fully Convolution Network[1]



Mask-RCNN[2]



[1] *Fully Convolutional Networks for Semantic Segmentation (CVPR'2015)*

[2] *Mask-RCNN (CVPR' 2017)*

Panoptic Segmentation

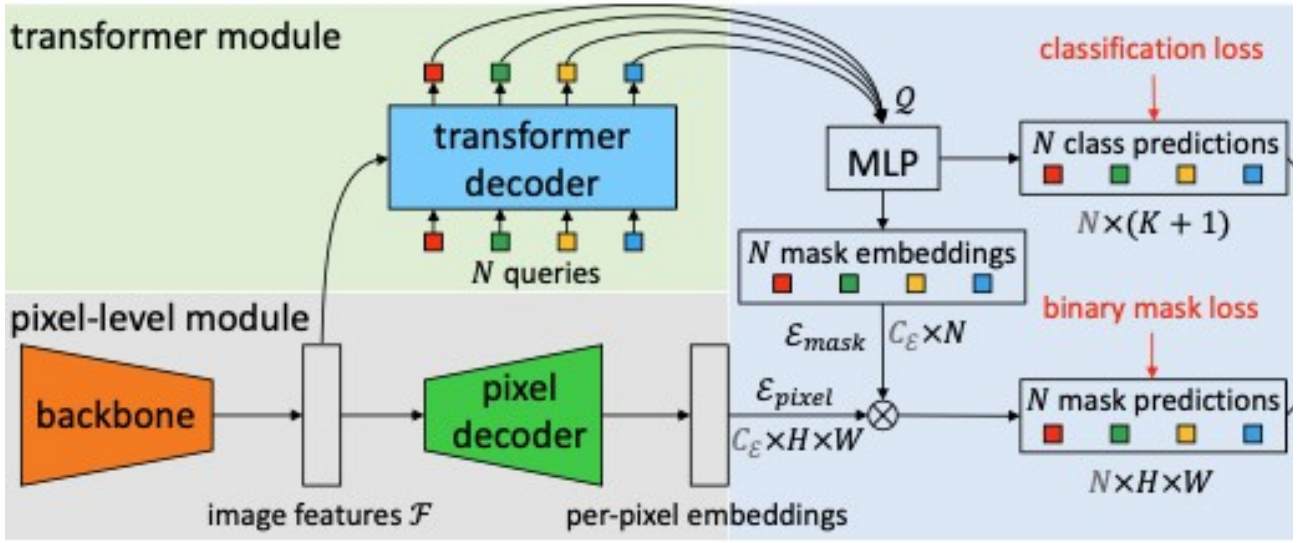
MaskFormer: Per-Pixel Classification is Not All You Need for Semantic Segmentation

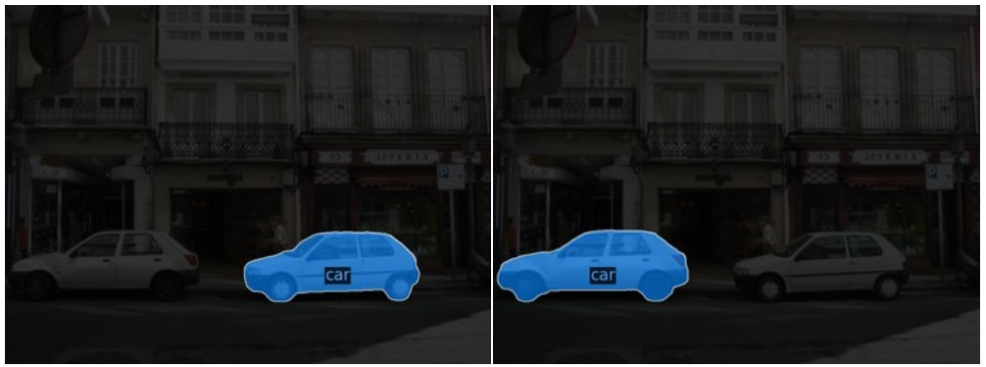
Meta AI Research @ NeurIPS’ 2021

To combine these inputs, we take inspiration from Transformer segmentation models [14, 20]...

Panoptic Segmentation

MaskFormer[3]





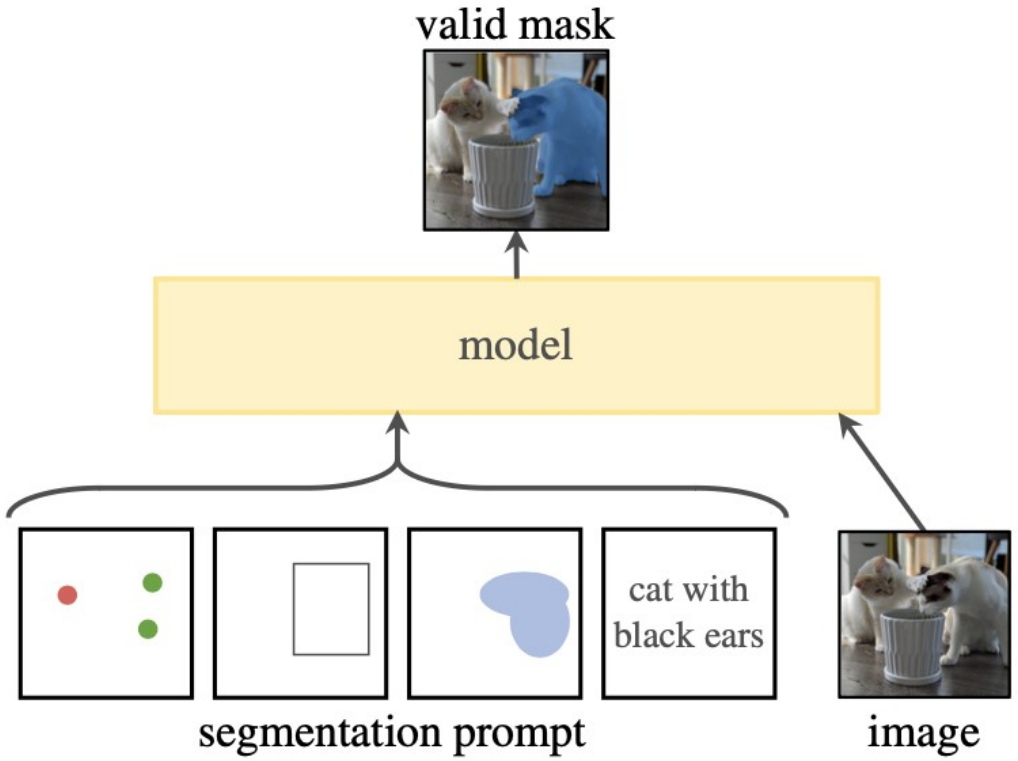
[3] *Per-Pixel Classification is Not All You Need for Semantic Segmentation (NeurIPS' 2021)*

Segment Anything Model

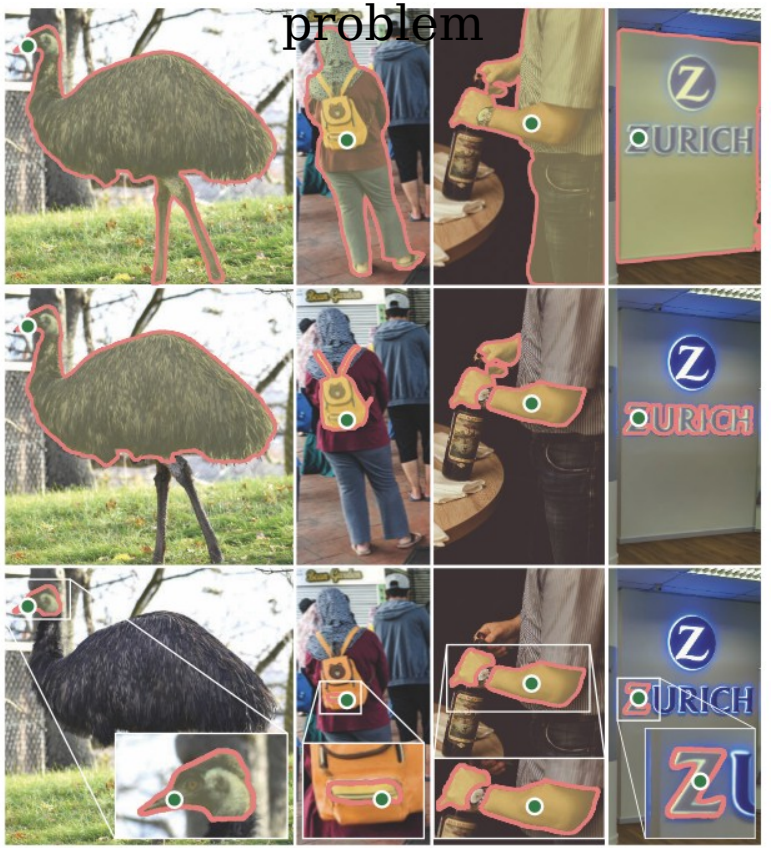
https://segment-anything.com/

What **task** will enable zero-shot generalization?

Promptable segmentation task



Ambiguous
problem



What is the corresponding **model architecture**?

Model architecture overview

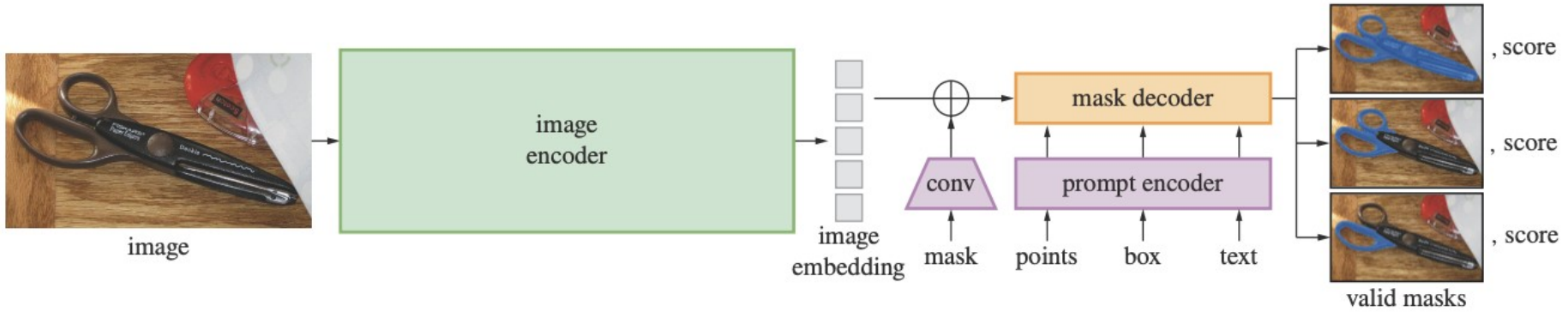


Image encoder and prompt encoder

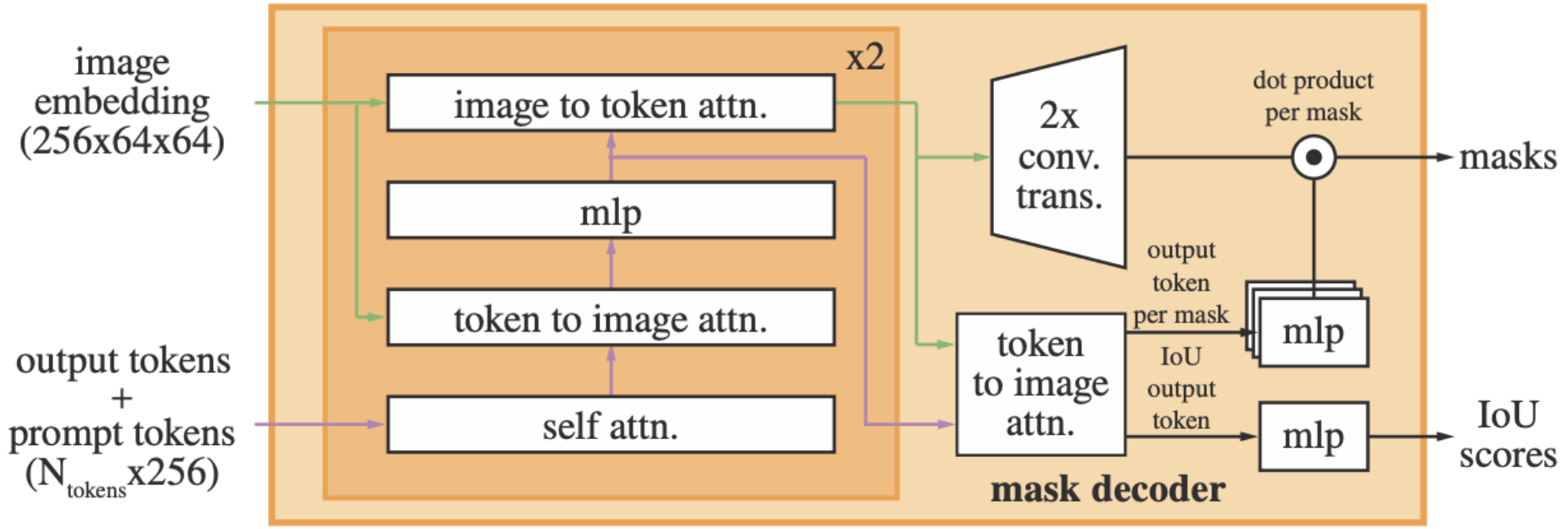
Image encoder

MAE pre-trained Vision Transformer with minimal adaptations[4] to process high resolution inputs.

Prompt encoder

[4] Exploring Plain Vision Transformer Backbones for Object Detection (ECCV’ 2022)

Mask decoder



MaskFormer: Use object query to get a binary mask for each object

SAM: Use output token to get a binary mask for each prompt.

What **data** can power this task and model?

Data engine

Assisted-manual stage

They collected 4.3M masks from 120k images in this stage.

Semi-automatic stage

During this stage they collected an additional 5.9M masks in 180k images (for a total of 10.2M masks).

Fully automatic stage

We applied fully automatic mask generation to all 11M images in our dataset, producing a total of 1.1B high-quality masks.

Fully automatic stage

1. How to automatically generate mask?
2. How to train the model?

Automatic segmentation



@step 1: Prompt the model with a 32 × 32 regular grid of foreground points. Each point will predict a set of masks that may correspond to valid objects.

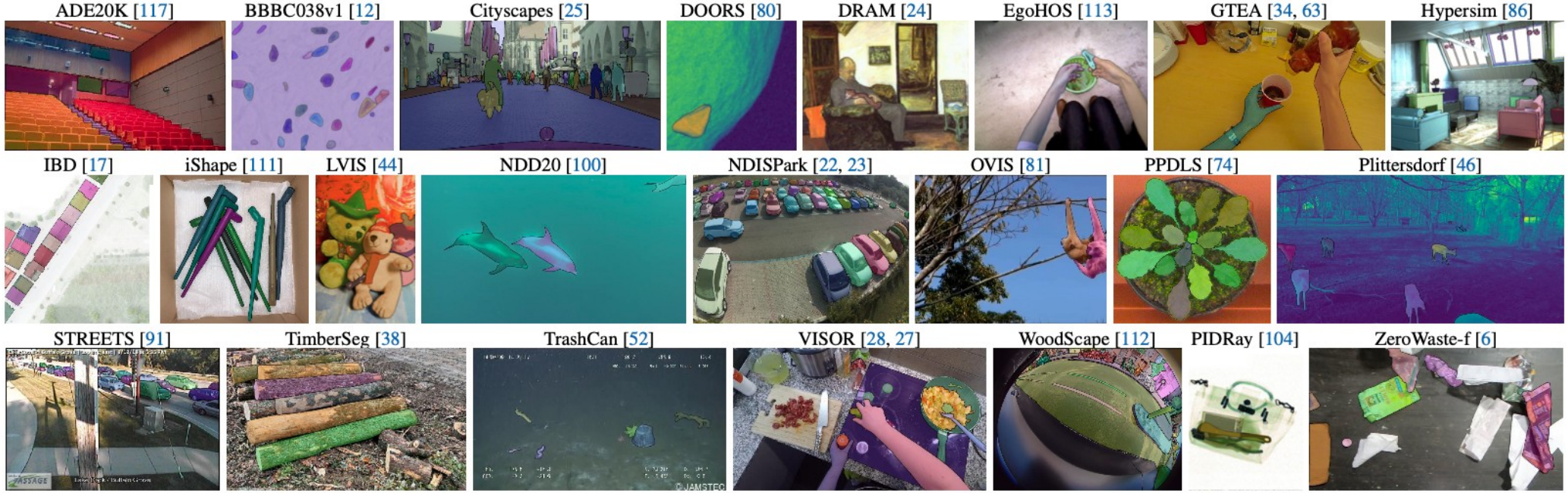
@step 2: Select stable masks (They consider a mask stable if thresholding the probability map at $0.5 - \delta$ and $0.5 + \delta$ results in similar masks).

@step 3: Apply non-maximal suppression to filter duplicates.

Interactive training strategy

```
# Main training loop
for (img, gt) in loader: # Load a minibatch 'img' with corresponding ground truth 'gt’
   embed = image_encoder(img) # Obtain image embedding from the image encoder
   prompts = prompt_encoder(points or box) # Initialize prompts either as a point from 'gt' or a noisy box derived
     from 'gt'
   iou, logits = mask_decoder(embed, prompts, iteration=0) # Generate IOU prediction and logits from mask decoder

   L = criterion(iou, logits, gt) # Compute loss
   L.backward() # Backpropagate errors
   update(image_encoder, prompt_encoder, mask_decoder) # Perform parameters update using AdamW optimizer

   points = select(logits, gt) # Select new prompts for next iteration from falsely predicted areas

   # Iterative refinement loop
   for i in iteration:
      prompts = prompt_encoder(points, logits) # Set prompts for the current iteration, using logits from the last
        iteration
      iou, logits = mask_decoder(embed.detach(), prompts, iteration=i) # Gradients won‘t propagate back to the image
        encoder

      L = criterion(iou, logits, gt) # Compute loss
      L.backward() # Backpropagate errors
      update(prompt_encoder, mask_decoder) # Perform parameters update using AdamW optimizer

      points = select(logits, gt) # Select new prompts for next iteration from falsely predicted areas
```

Experiments



Zero-shot edge detection; Zero-shot object proposals

Zero-shot Instance segmentation; Zero-shot single point valid mask evaluation

Zero-shot edge detection

method	year	ODS	OIS	AP	R50
HED [108]	2015	.788	.808	.840	.923
EDETR [79]	2022	.840	.858	.896	.930
zero-shot transfer methods:					
Sobel filter	1968	.539	-	-	-
Canny [13]	1986	.600	.640	.580	-
Felz-Hutt [35]	2004	.610	.640	.560	-
SAM	2023	.768	.786	.794	.928

@step 1: Prompt the model with a 16 times 16 regular grid of foreground points. Each point will predict a set of masks that may correspond to valid objects.

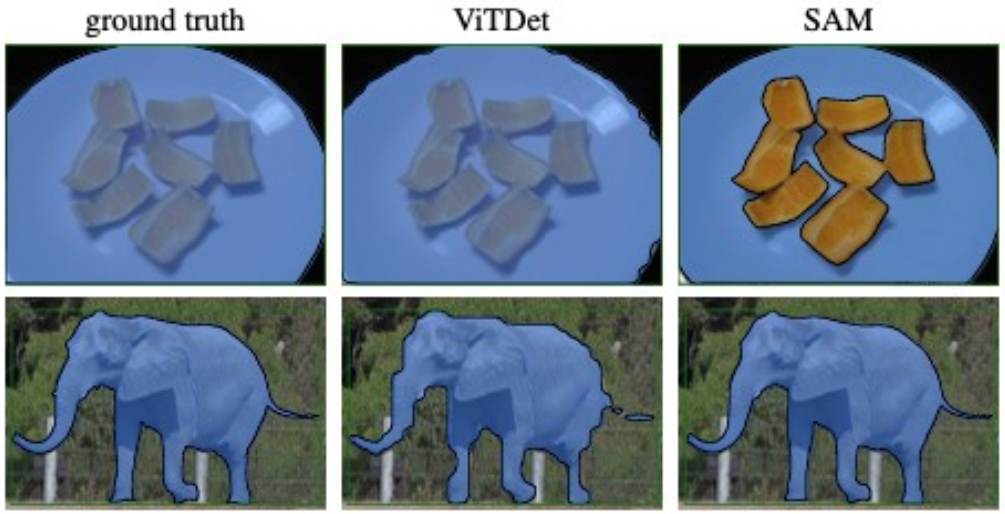
@step 2: Apply non-maximal suppression to filter duplicates.

@step 3: Then they apply a Sobel filter to the remaining masks' probability maps

Zero-shot Instance segmentation

	COCO [66]				LVIS v1 [44]			
method	AP	AP^S	AP^M	AP^L	AP	AP^S	AP^M	AP^L
ViTDet-H [62]	51.0	32.0	54.3	68.9	46.6	35.0	58.0	66.3
zero-shot transfer methods (segmentation module only):								
SAM	46.5	30.8	51.0	61.7	44.7	32.5	57.6	65.5

They run an object detector (the ViTDet[4] used before) and prompt SAM with its output boxes.



[4] Exploring Plain Vision Transformer Backbones for Object Detection (ECCV' 2022)

Quiz

Automatic segmentation

In SAM, which of the following options will represent one of these points?

a. A learnable parameter represents foreground point.

b. A learnable parameter represents background point.

c. A learnable parameter represents left-up corner.

d. A learnable parameter represents right-down corner.

Segment Anything

Why do we set all (32 × 32) points as foreground points but not include background points?

The model predicts 32 × 32 (× 3) binary masks, where each point is referring to the foreground for its corresponding mask, regardless of whether the mask represents a so-call background element (such as sand or grass) or a foreground subject (like people).

AutoSAM: Adapting SAM to Medical Images by Overloading the Prompt Encoder

Input Augmentation with SAM

Medical SAM Adapter

Segment Anything in Medical Images

Polyp-SAM: Transfer SAM for Polyp Segmentation

SAM for Medical Imaging: Experimental Study

Segment Anything Model for Medical Images?

SAM-Med2D

SAM on Medical Images: A Comprehensive Study on Three Prompt Modes

When SAM Meets Medical Images

SAM Meets Robotic Surgery

Medical Image Demo

https://segment-anything.com/