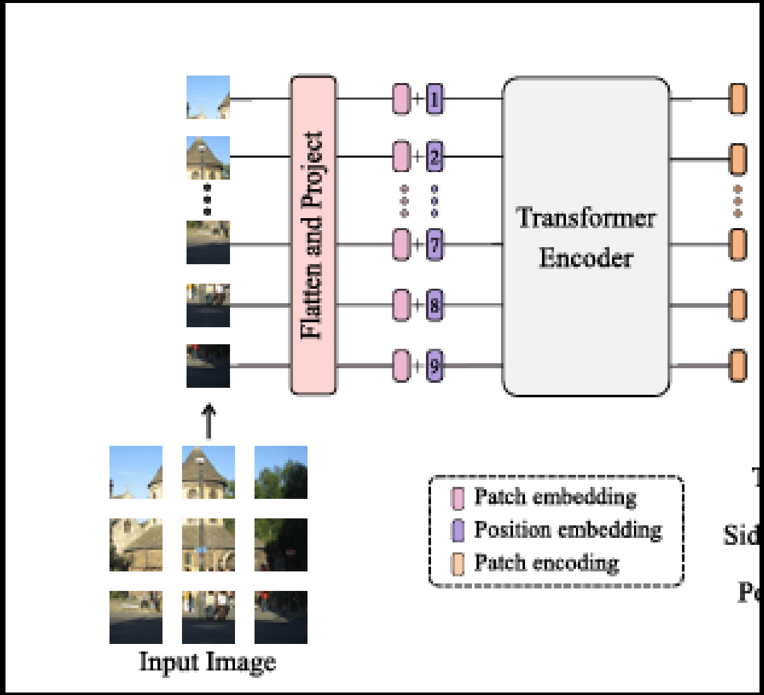
AIL 862

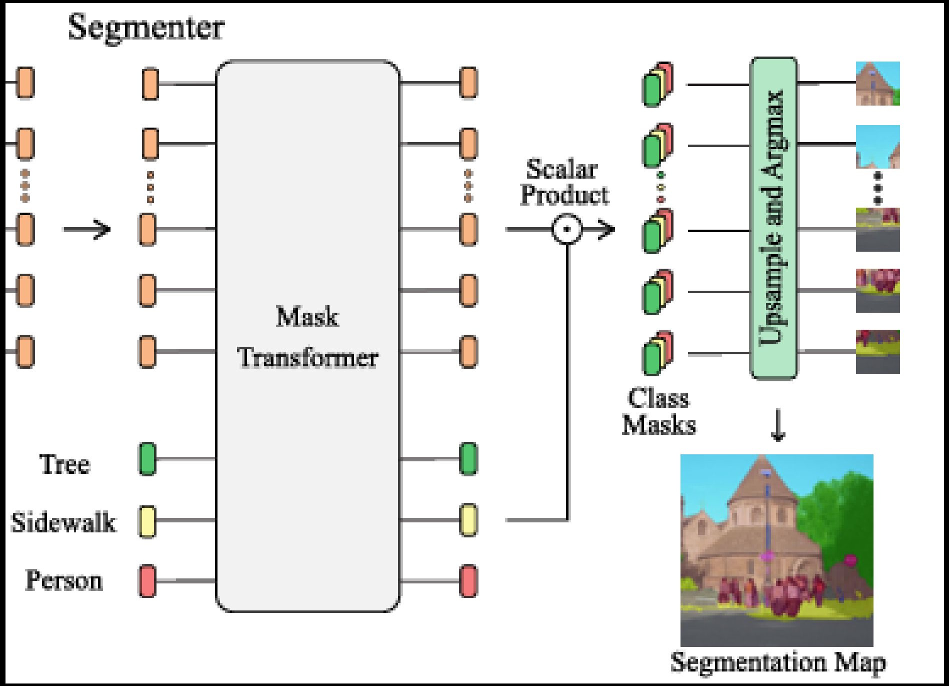
Lecture 20

Segmenter



Segmenter: Transformer for Semantic Segmentation, 2021

Segmenter

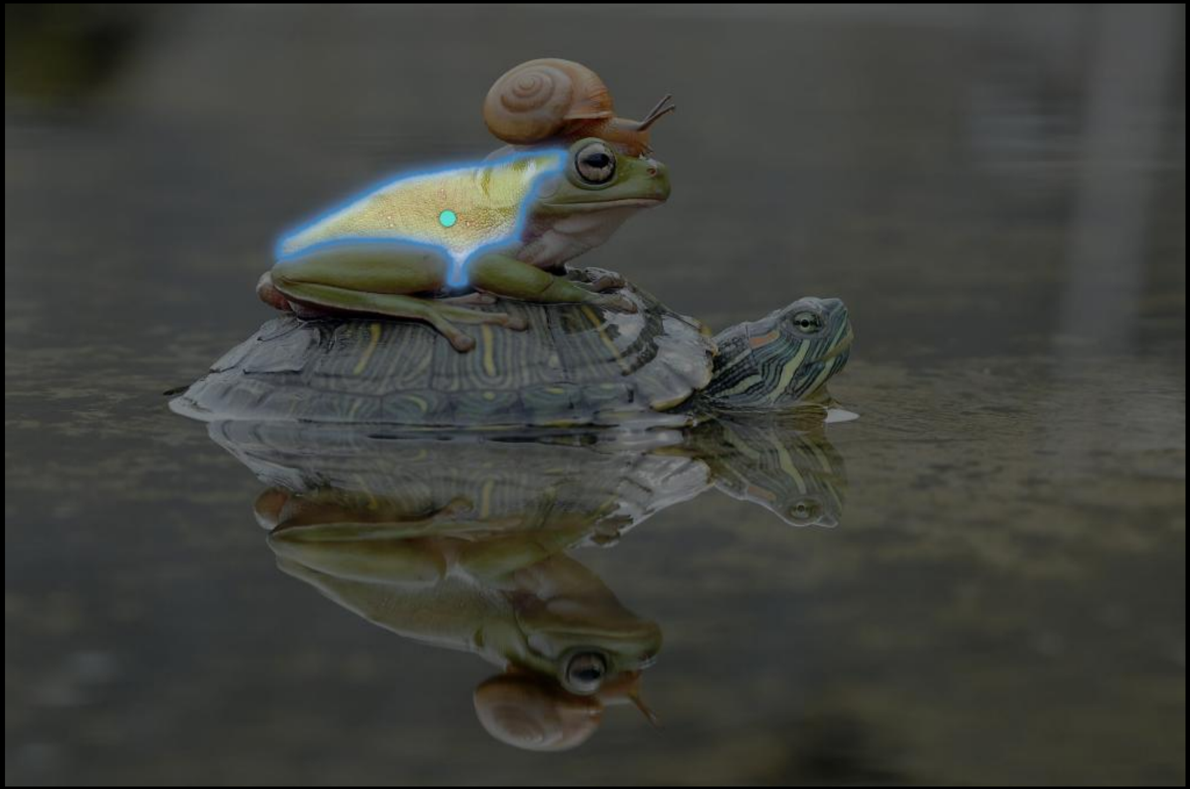


Segmenter: Transformer for Semantic Segmentation, 2021

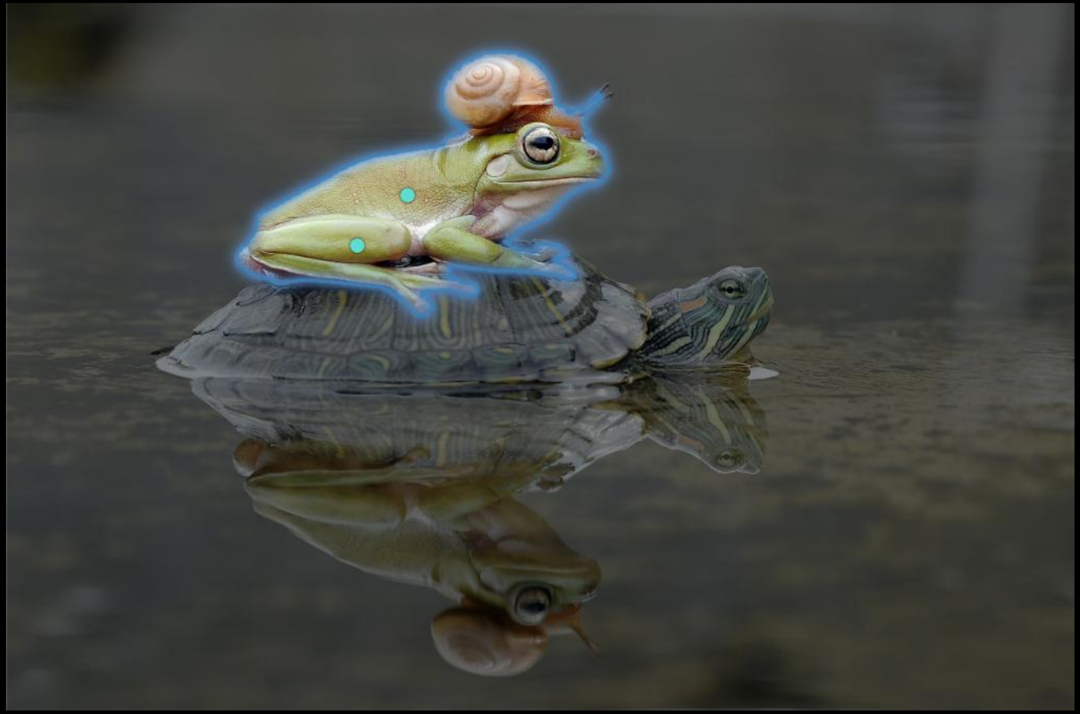
CNN-Transformer cross teaching

Using Clip

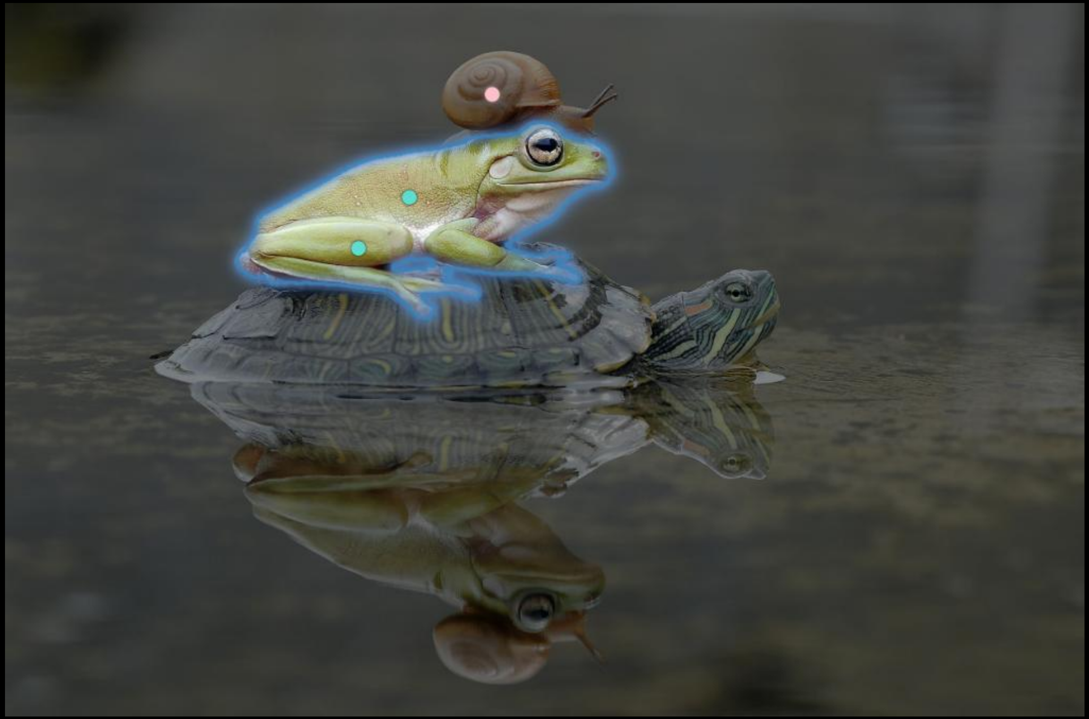
Segmentation with prompts



Segmentation with prompts



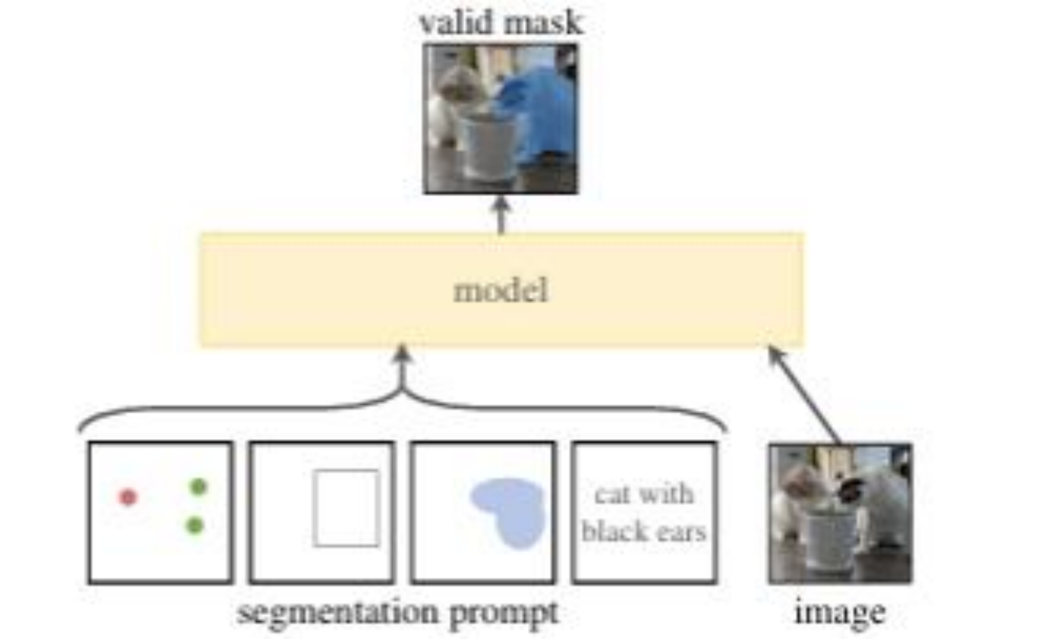
Segmentation with prompts



Many uses

- Human supervision + AI augmentation

Segment Anything Model



SAM

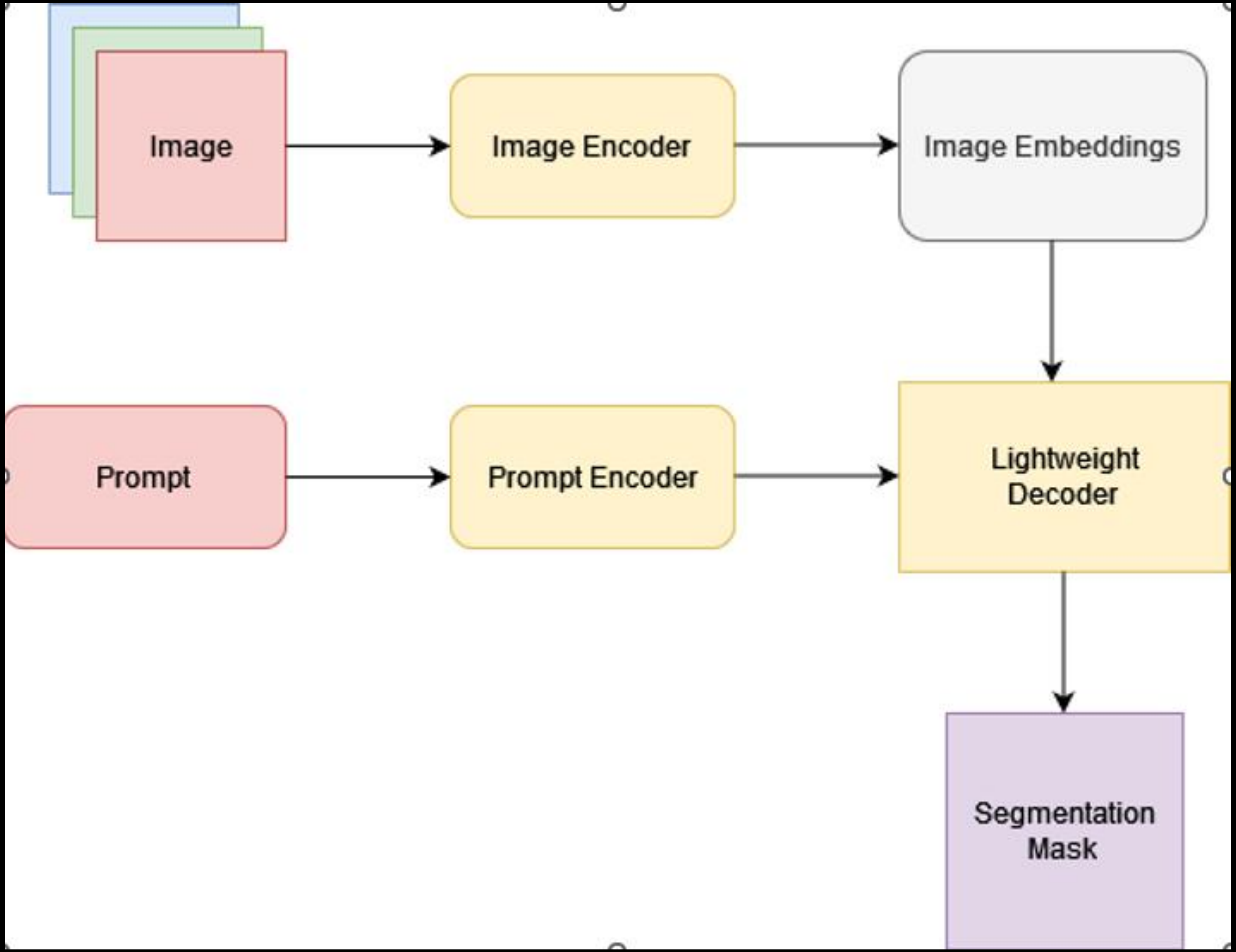


Image
Image Encoder
Image Embeddings
Prompt
Prompt Encoder
Lightweight Decoder
Segmentation Mask

Image Encoder

Image Encoder

- MAE pre-trained ViT

Image Encoder

- MAE pre-trained ViT

- Image encoder runs once per image

- Can be applied prior to prompting the model

Prompt encoder

- Two sets of prompts: sparse (points, boxes, text) and dense (masks).

- Points and boxes are represented by positional encodings summed with learned embeddings for each prompt type

Prompt encoder

- Dense prompts (i.e., masks) are embedded using convolutions and summed with the image embedding.

Decoder

- The decoder maps the image embedding, prompt embeddings, and an output token to a mask.

Ambiguity in prediction

- Model may predict multiple output masks for a single prompt

- During training only the minimum loss over different output masks is considered

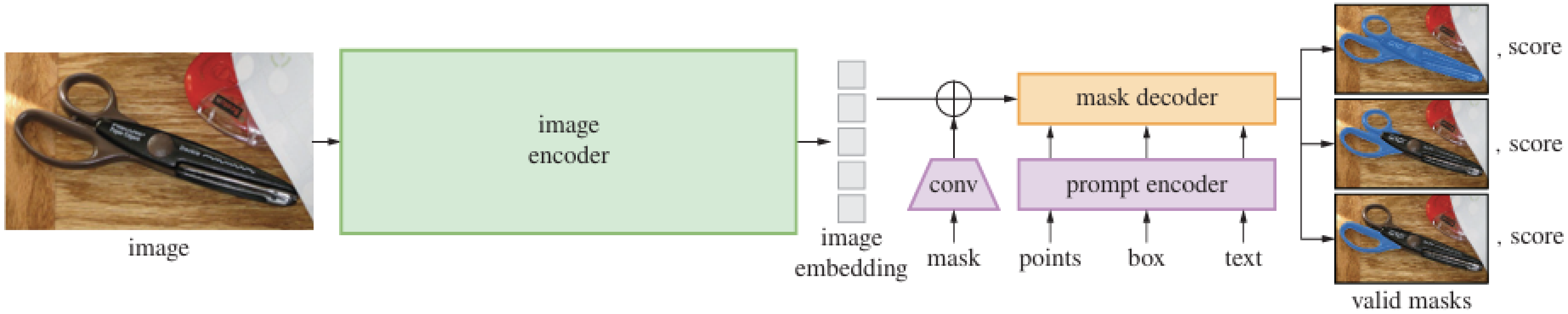
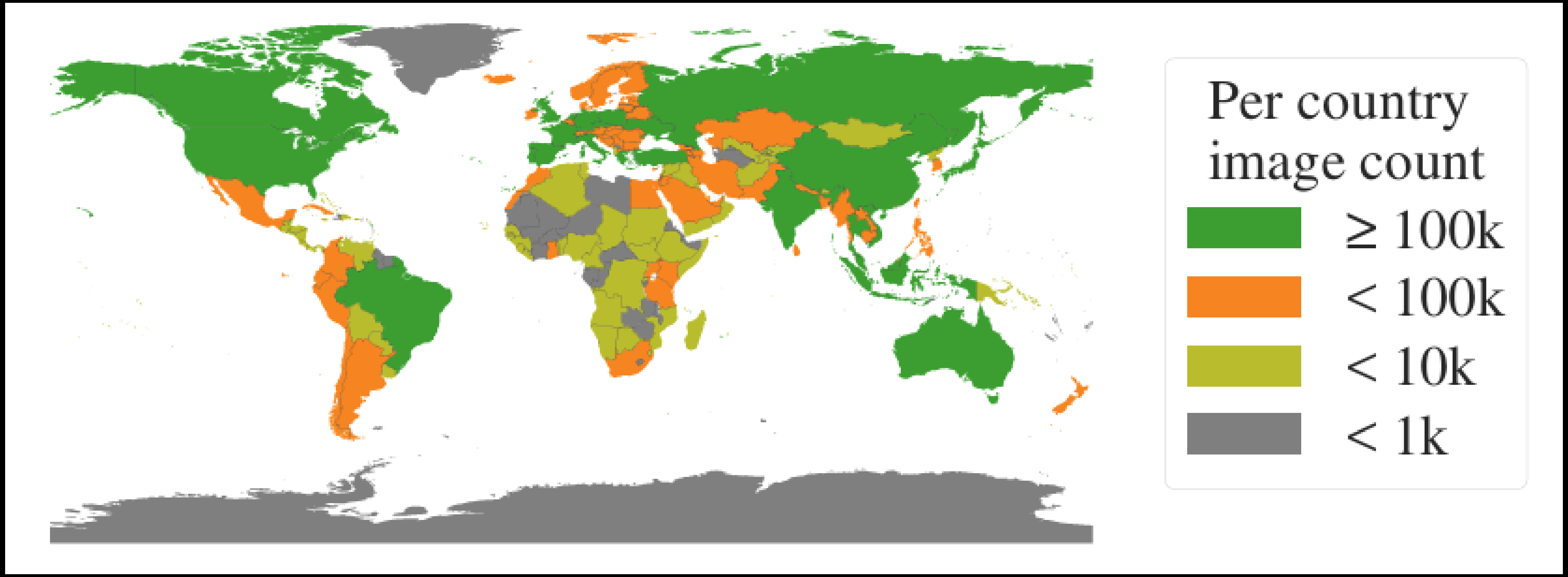


image
image encoder
image embedding
mask
conv
mask decoder
prompt encoder
points
box
text
, score
, score
, score
valid masks

SAM Data Engine

SAM Dataset Fairness



Fairness in segmenting people