

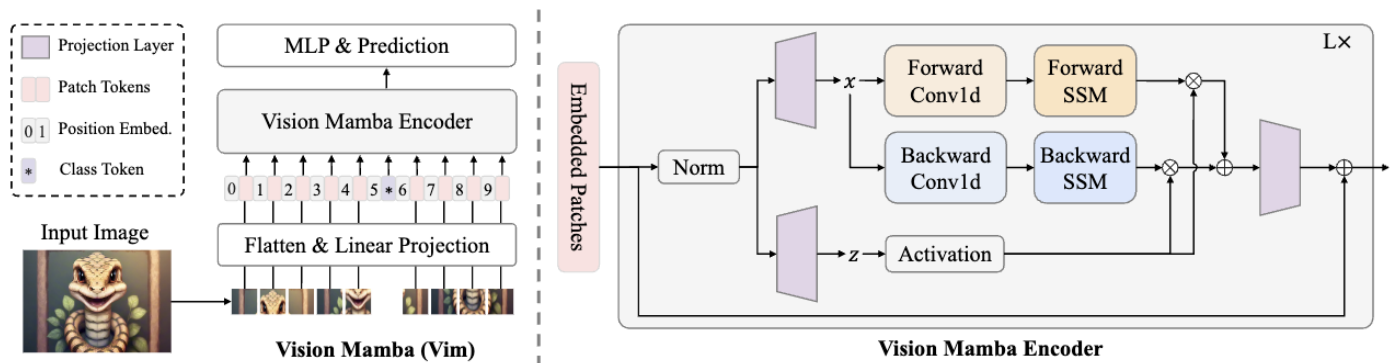
# Transformer in CV

## MEGALODON

<https://arxiv.org/pdf/2404.08801.pdf>

## Vision Mamba

<https://github.com/hustvl/Vim/>



## Vision Transformer ViT

[https://github.com/huggingface/pytorch-image-models/blob/main/timm/models/vision\\_transformer.py](https://github.com/huggingface/pytorch-image-models/blob/main/timm/models/vision_transformer.py)

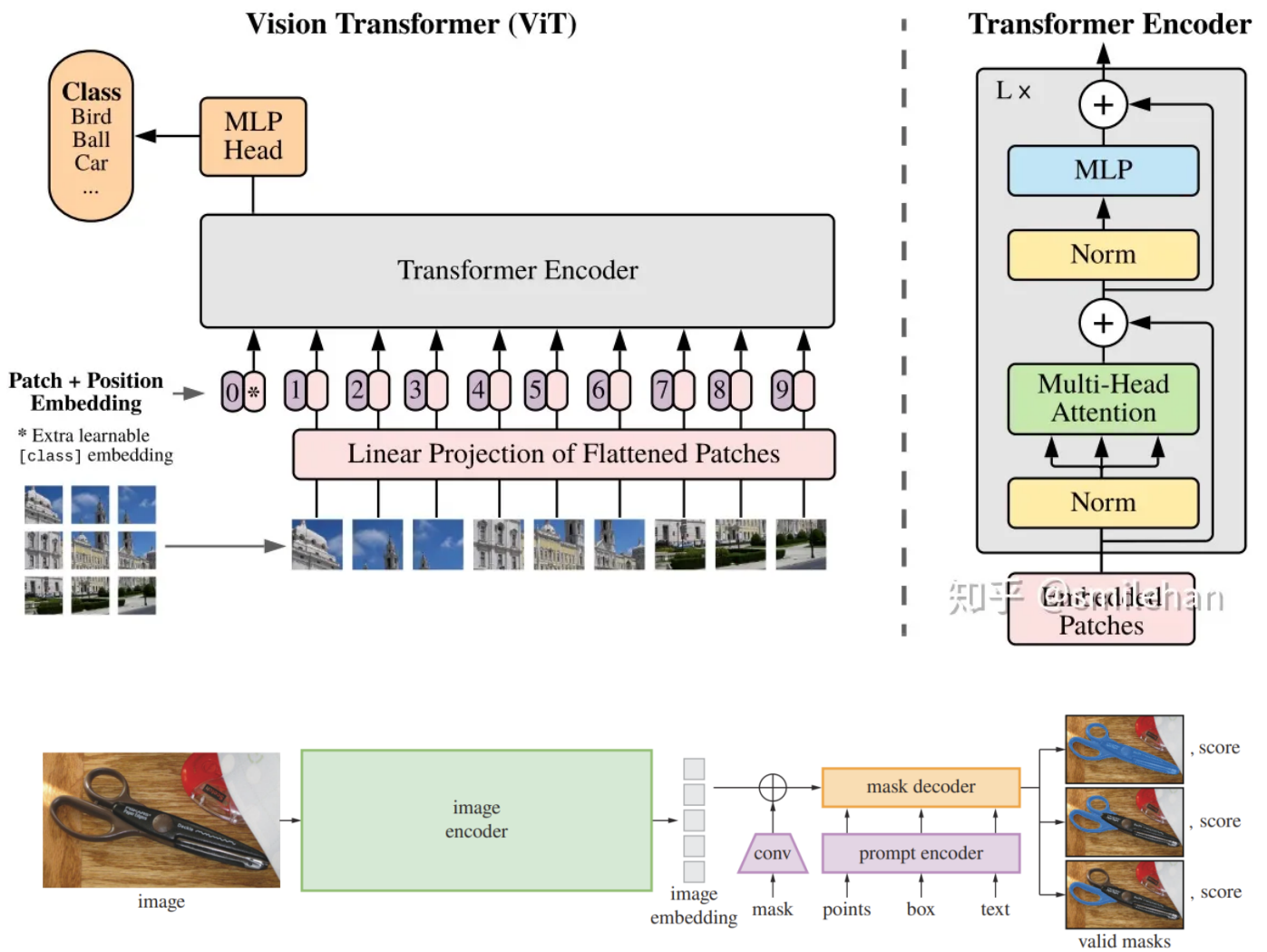
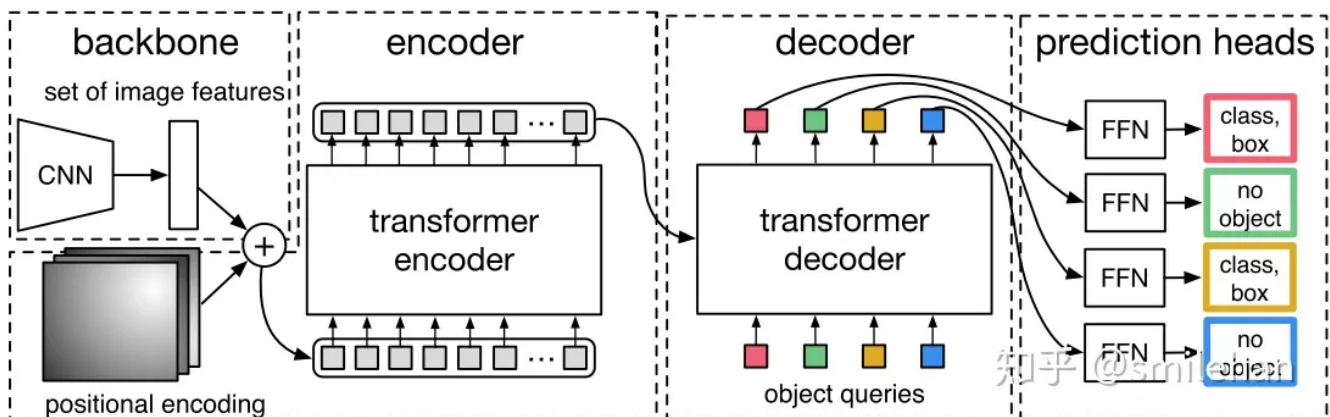
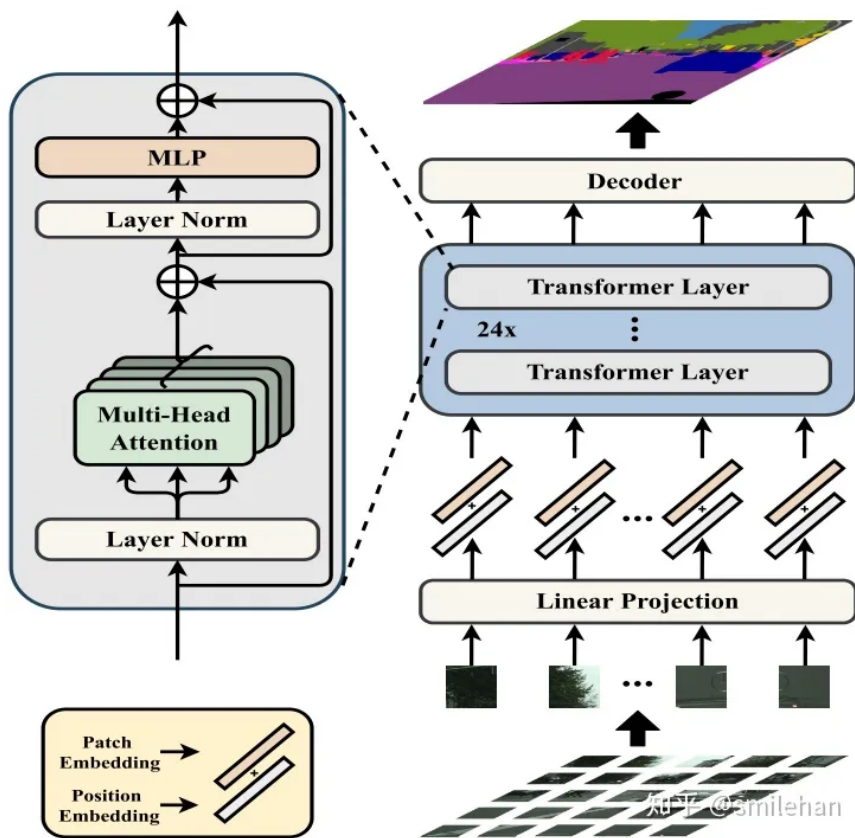


Figure 4: Segment Anything Model (SAM) overview. A heavyweight image encoder outputs an image embedding that can then be efficiently queried by a variety of input prompts to produce object masks at amortized real-time speed. For ambiguous prompts corresponding to more than one object, SAM can output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

## DEtection TRansformer □ DETR



## SEgmentation TRansformer □ SETR



Revision #1

Created 11 January 2025 09:44:04 by Colin

Updated 11 January 2025 09:44:06 by Colin