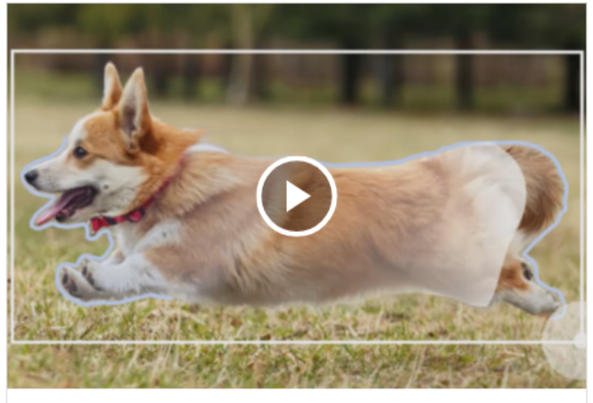
dcc CIENCIAS DE LA COMPUTACIÓN UNIVERSIDAD DE CHILE

Universidad de los Andes > FACULTAD DE INGENIERÍA Y CIENCIAS APLICADAS

Segment Anything Model
SAM



CLASE 8

José M. Saavedra R.
Profesor Asistente

jmsaavedrar@miuandes.cl

Ed. Ingeniería - Oficina 315

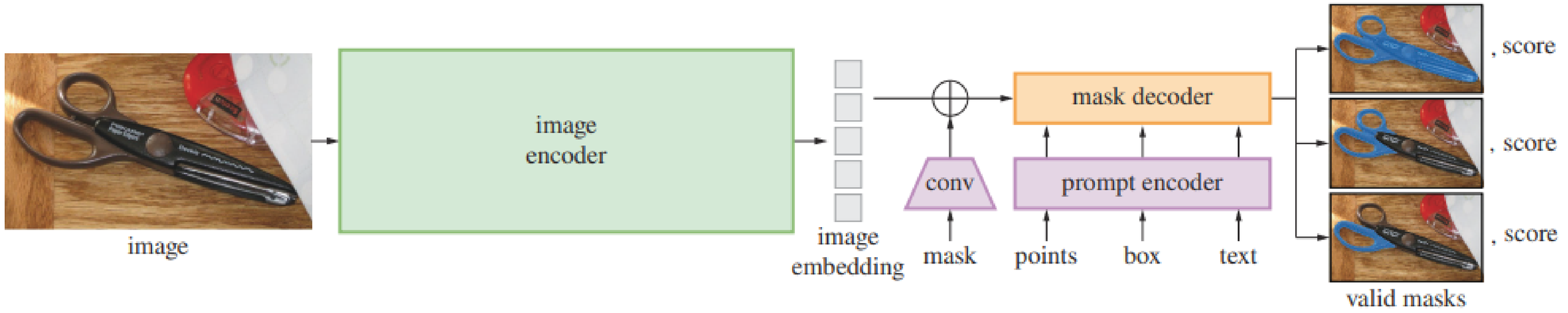
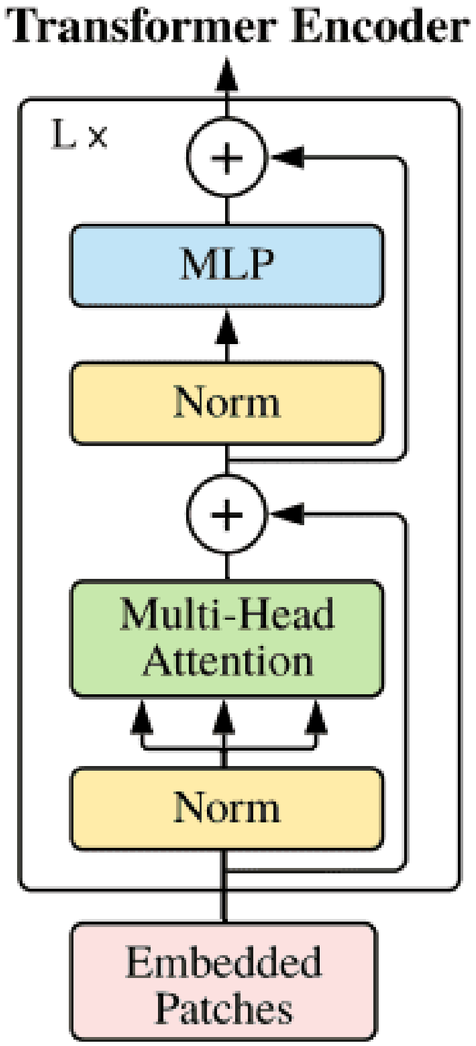
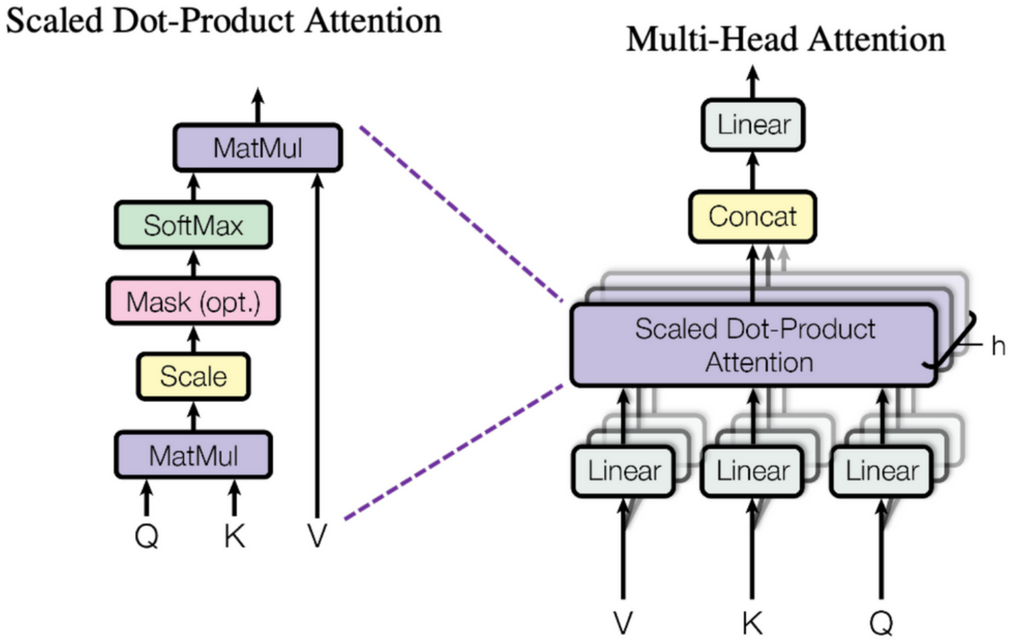


image
image encoder
image embedding
conv
mask
mask decoder
prompt encoder
points
box
text
, score
, score
, score
valid masks

ViT [vision transformer]

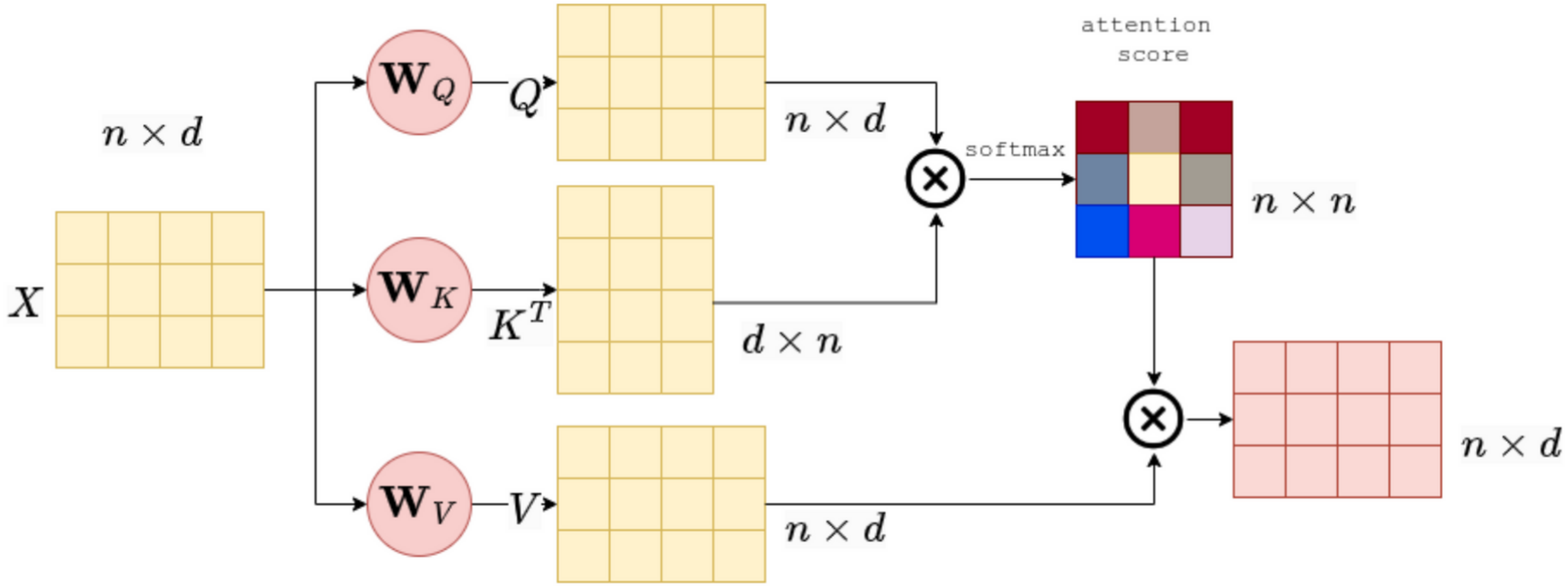
ViT [vision transformer]





$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V,$$

ViT [vision transformer]

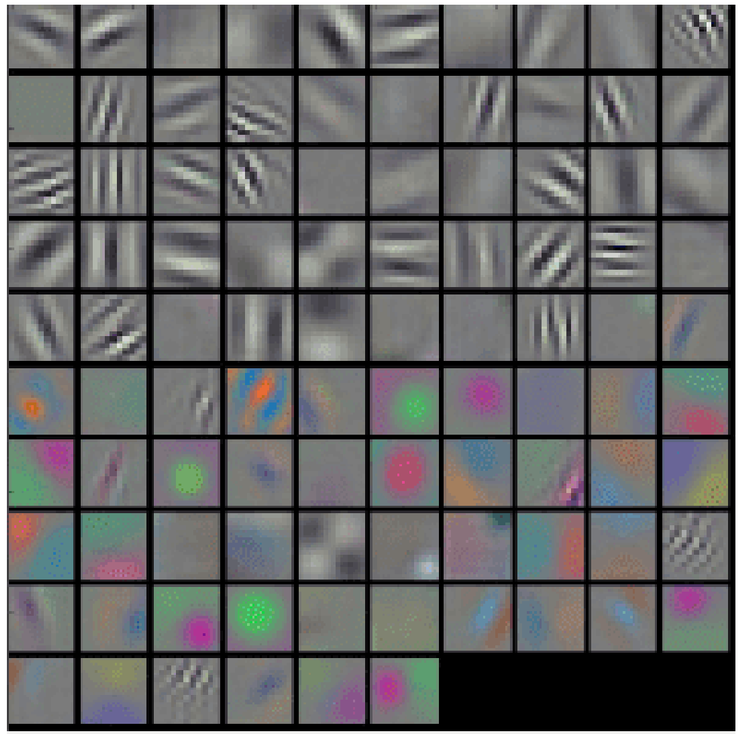


ViT [vision transformer]

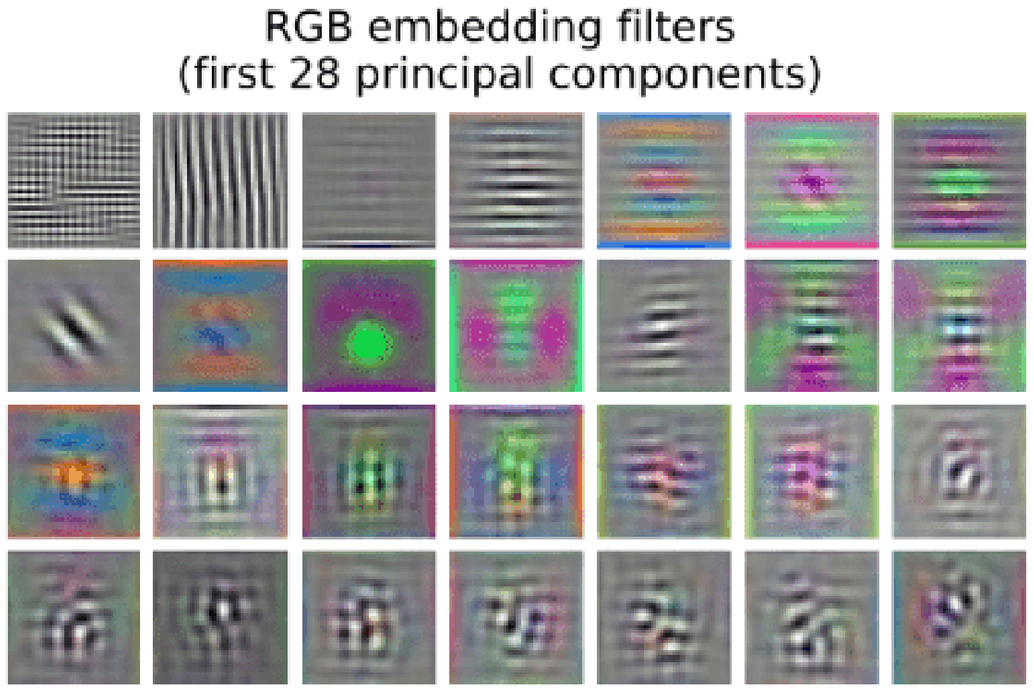
Model	Layers	Hidden size D	MLP size	Heads	Params
ViT-Base	12	768	3072	12	86M
ViT-Large	24	1024	4096	16	307M
ViT-Huge	32	1280	5120	16	632M

ViT [vision transformer]

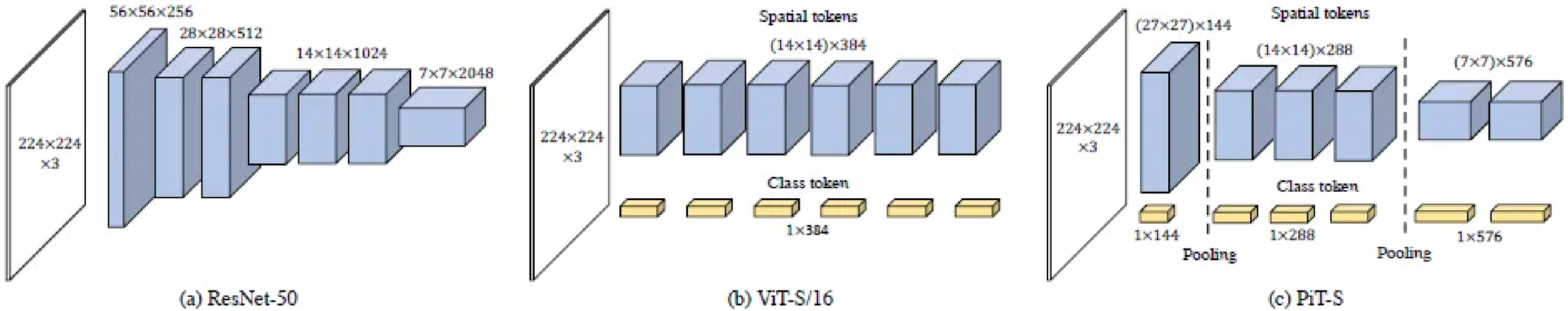
Alexnet 1st conv filters



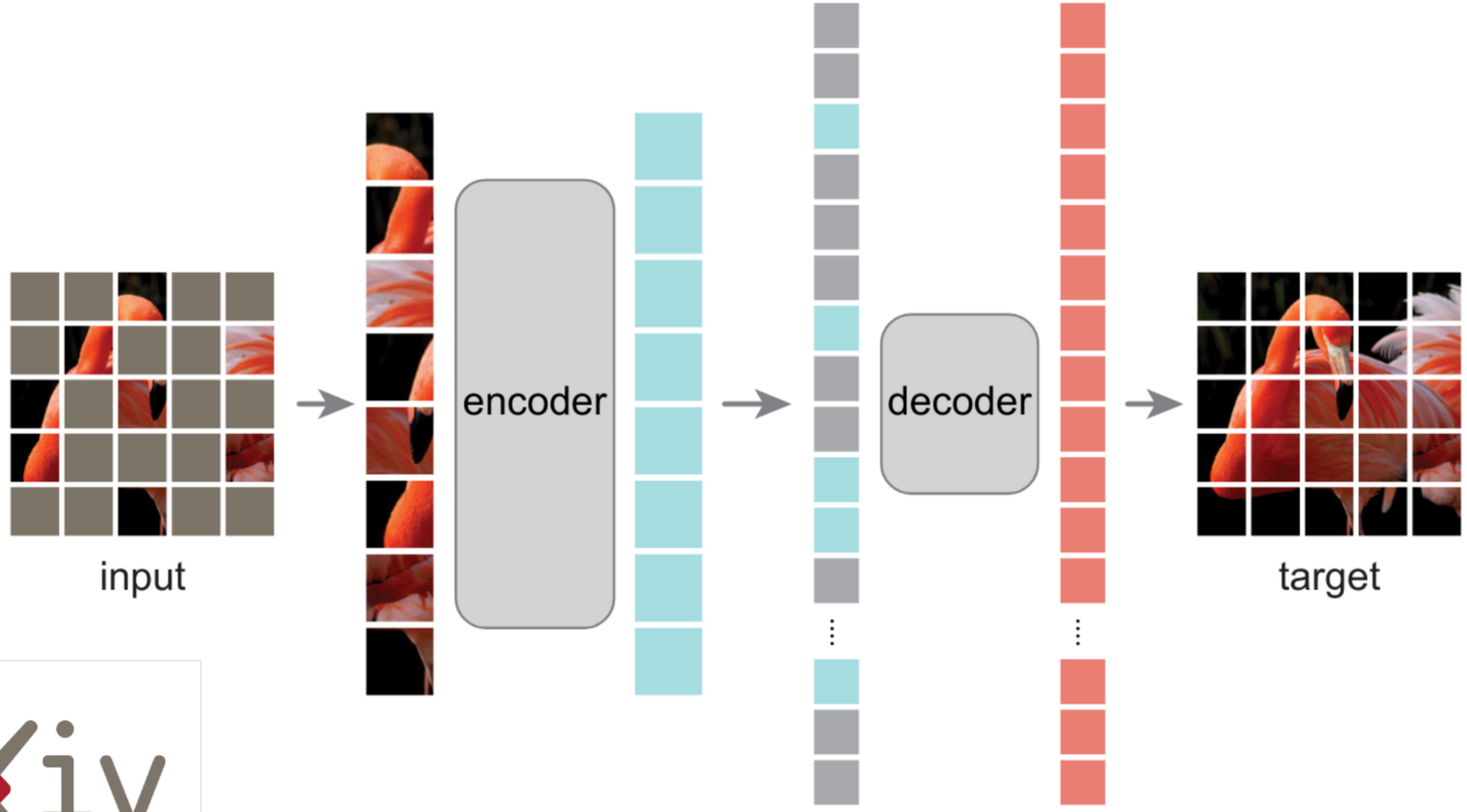
ViT 1st linear embedding filters



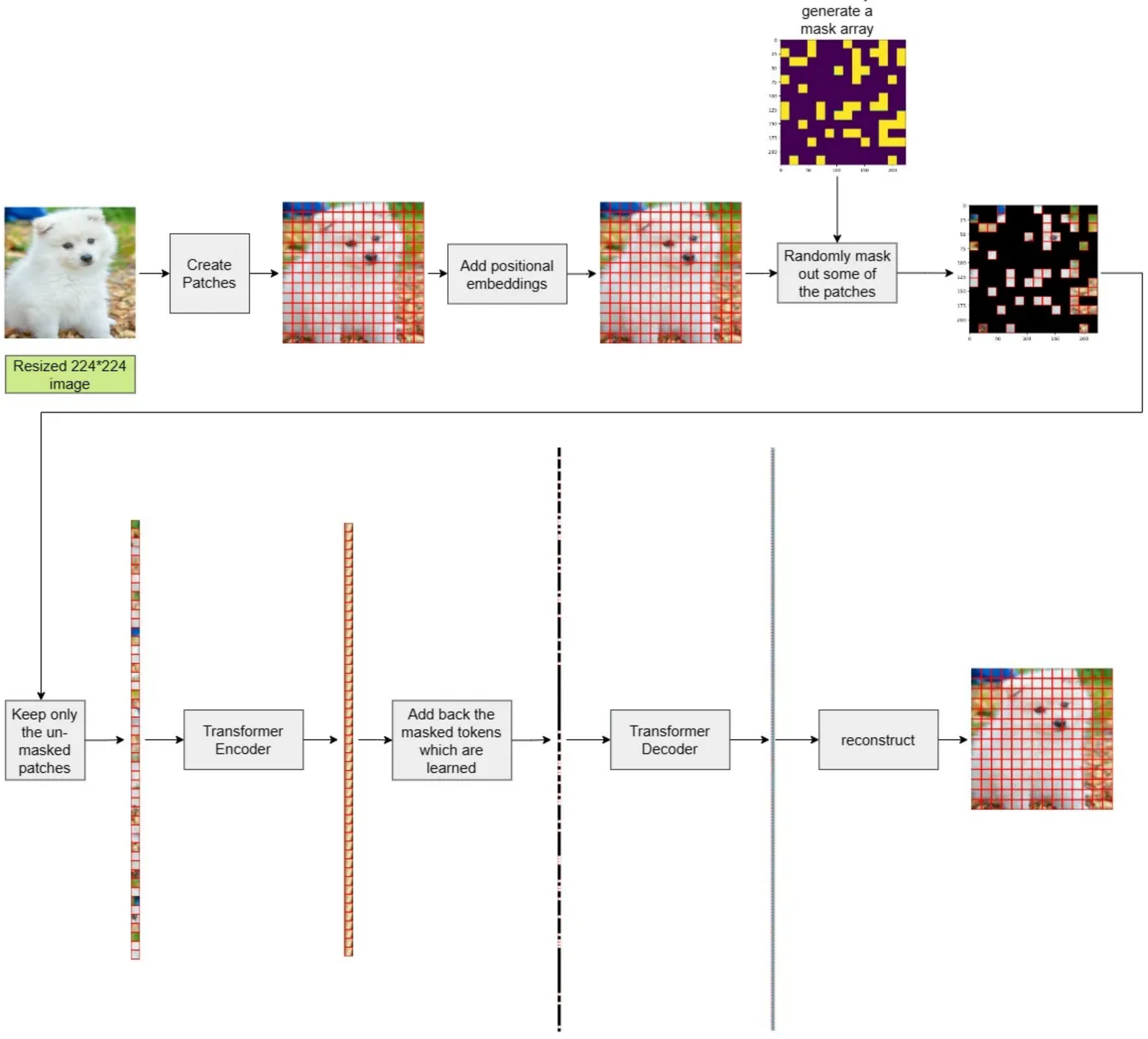
ViT [vision transformer]



MAE [masked autoencoders]

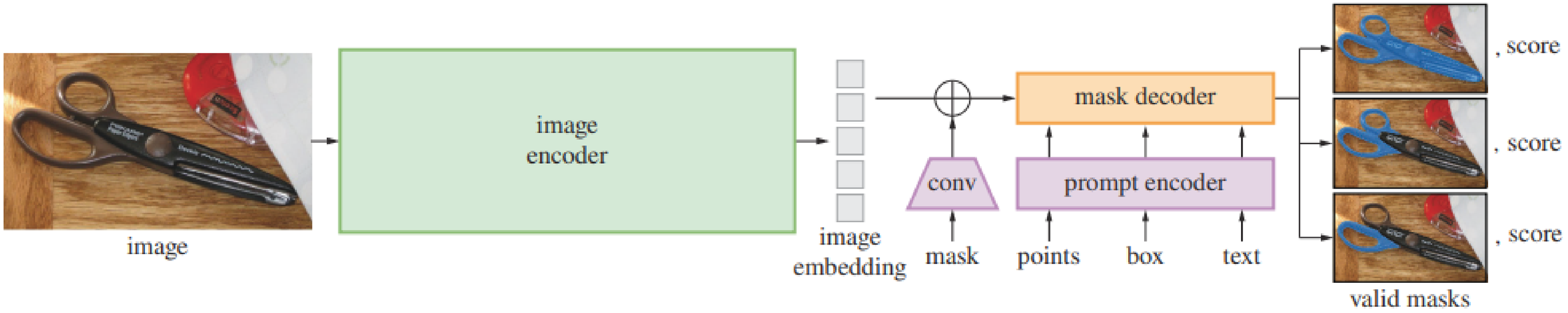


arXiv


generate a mask array
Create Patches
Add positional embeddings
Randomly mask out some of the patches
Resized 224*224 image
Keep only the un-masked patches
Transformer Encoder
Add back the masked tokens which are learned
Transformer Decoder
reconstruct

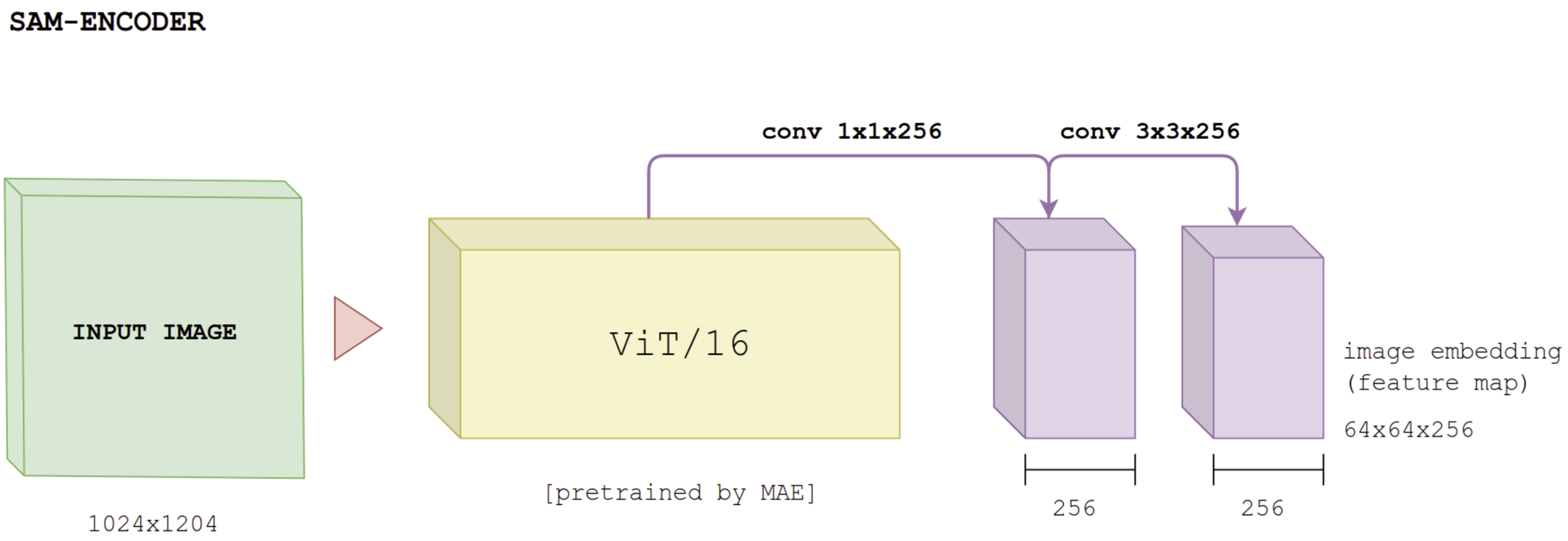
MAE

SAM [segmenting anything]



https://arxiv.org/pdf/2304.02643.pdf

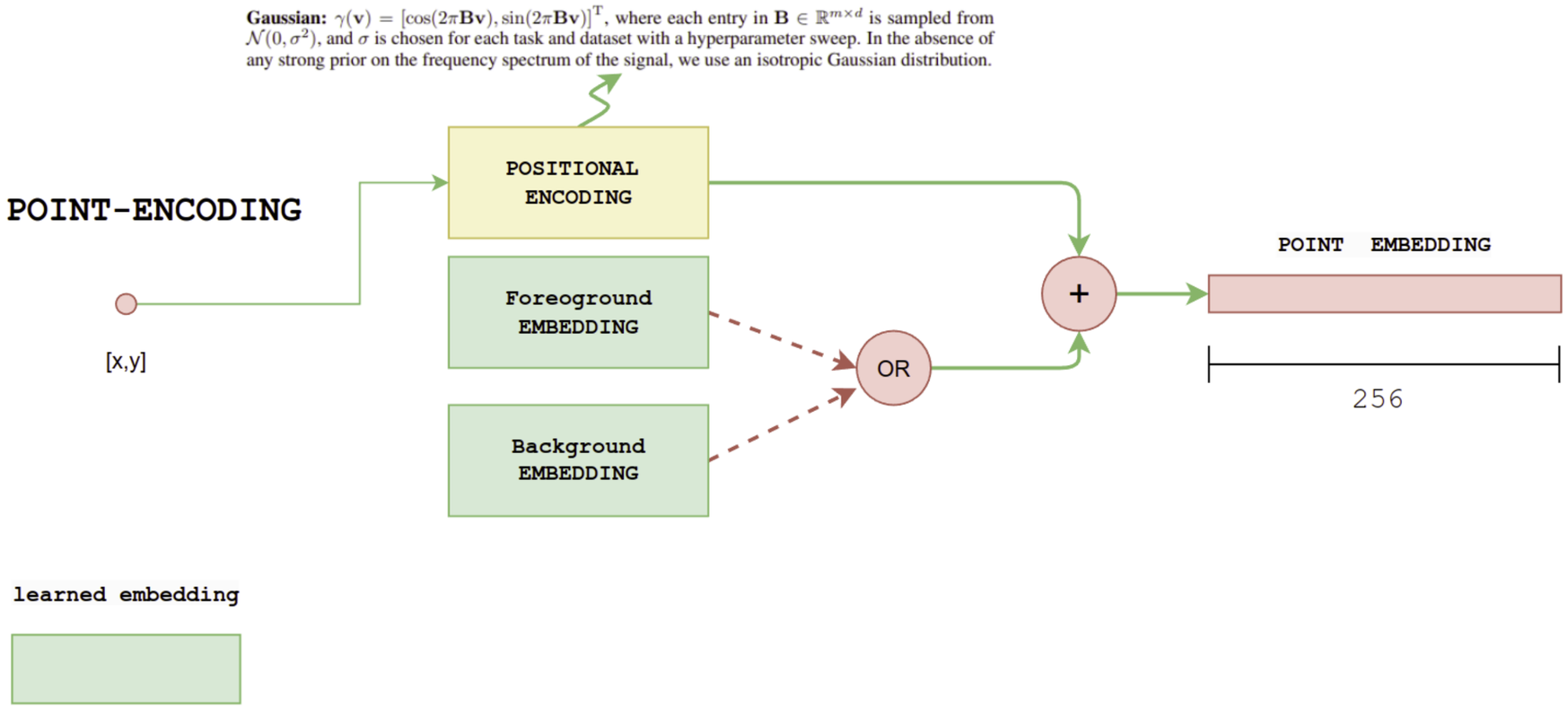
SAM [segmenting anything]



SAM [segmenting anything]

PROMPT ENCODER

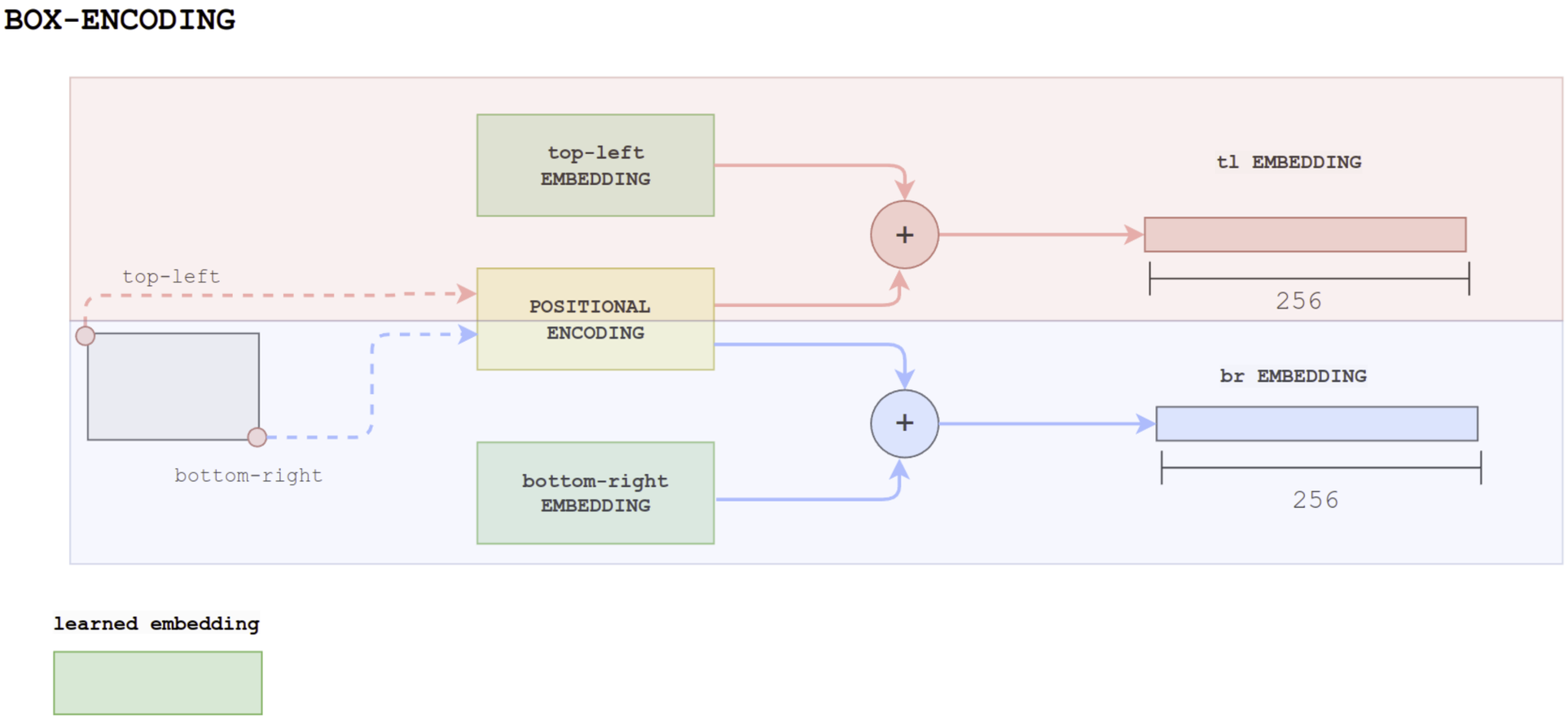
Sparse Encoding



SAM [segmenting anything]

PROMPT ENCODER

Sparse Encoding

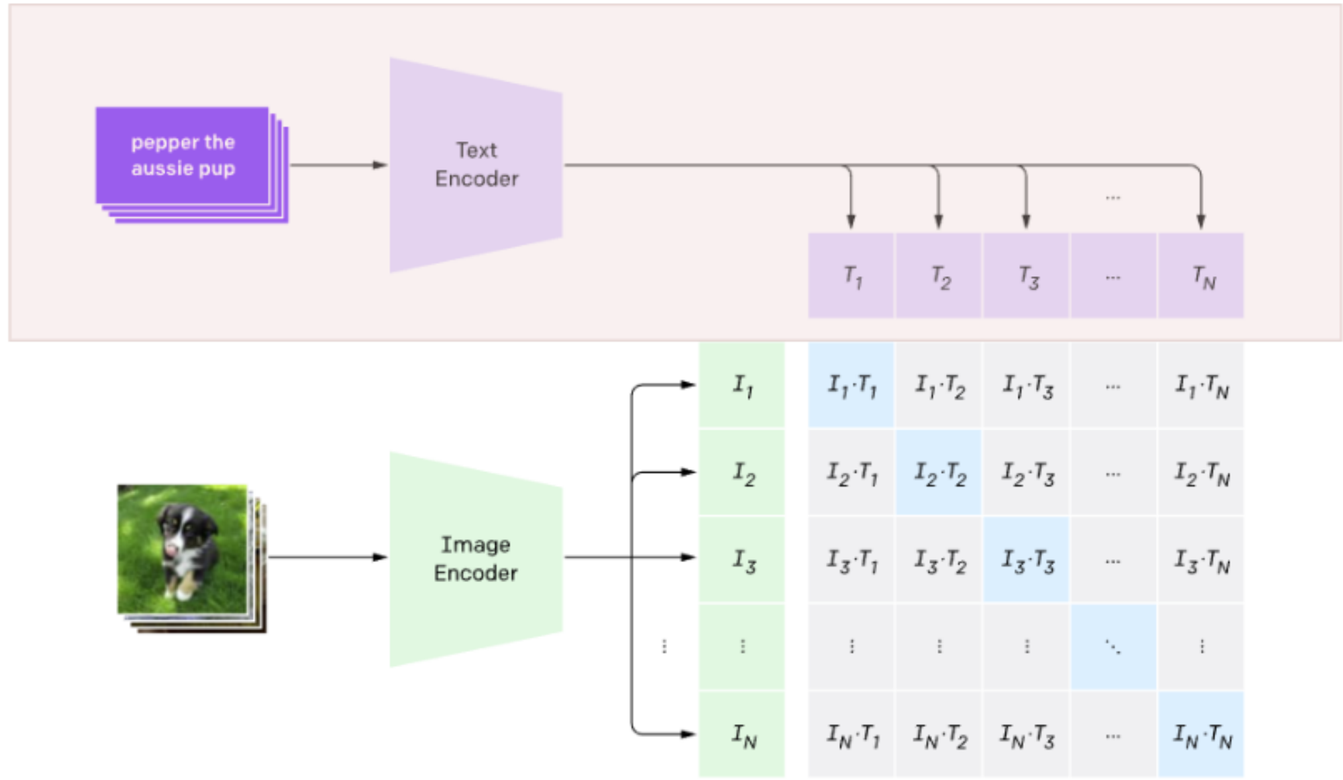


SAM [segmenting anything]

PROMPT ENCODER

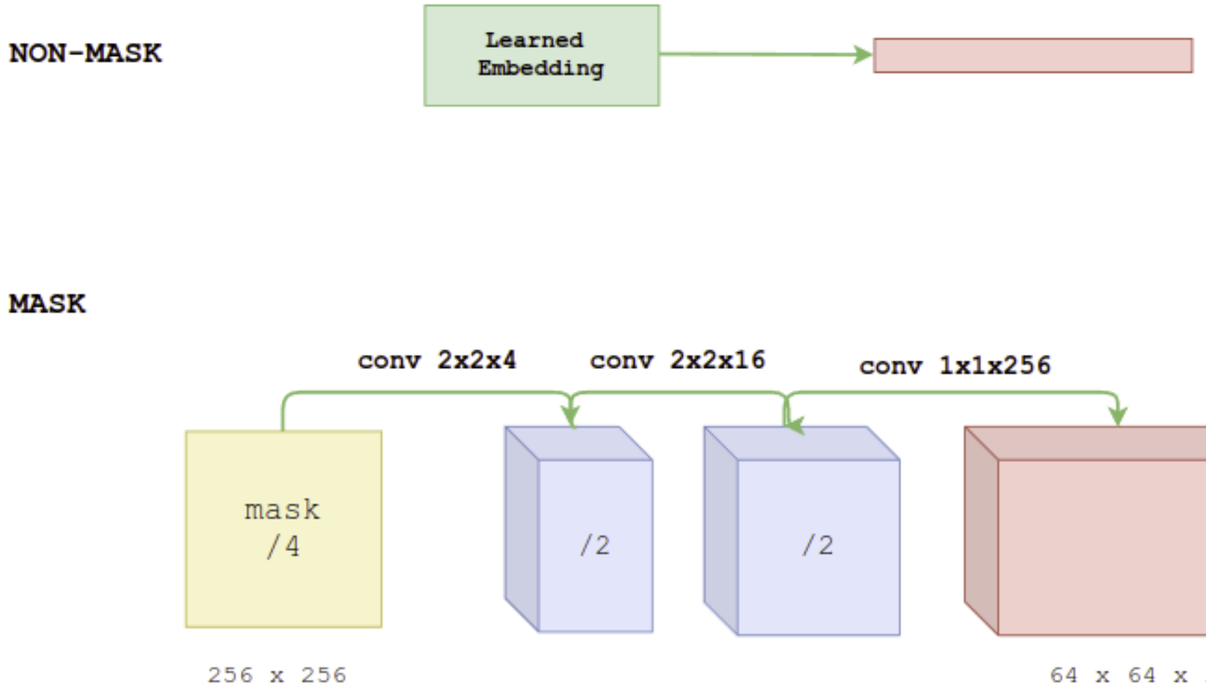
Sparse Encoding

TEXT-ENCODING



SAM [segmenting anything]

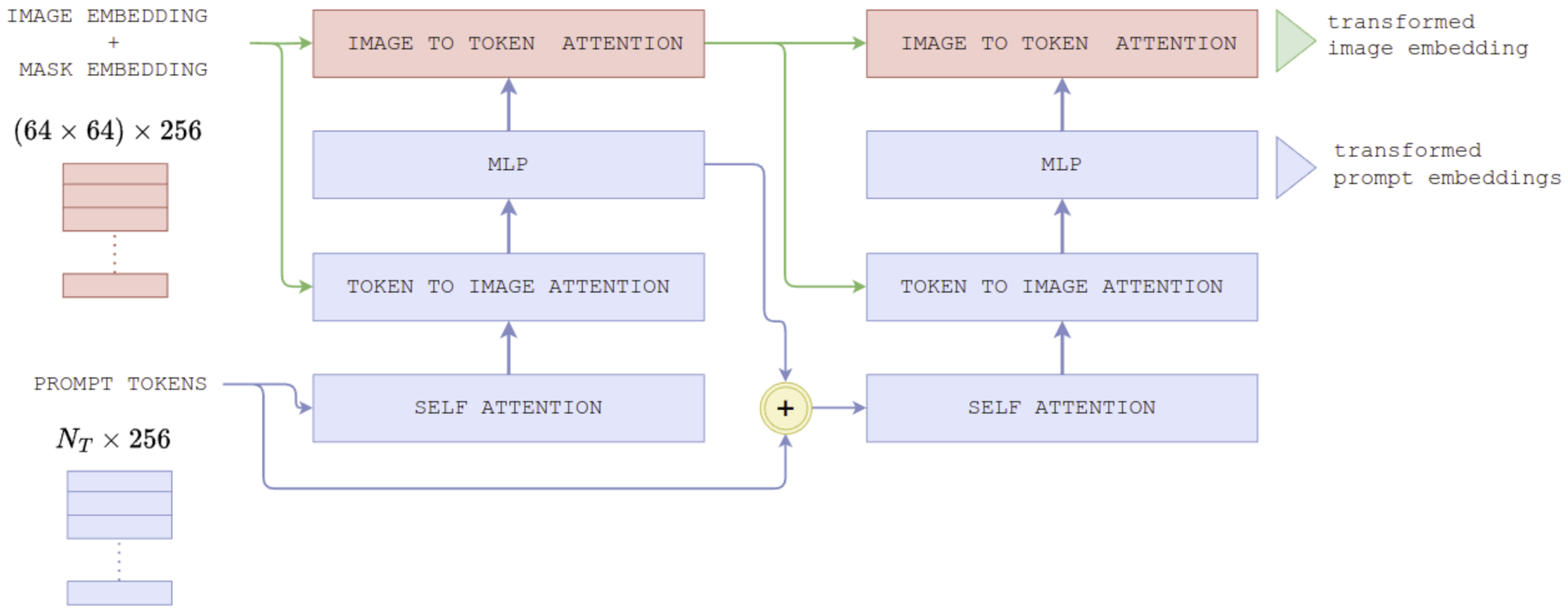
Dense Encoding (mask)




```
self.mask_input_size = (4 * image_embedding_size[0], 4 * image_embedding_size[1])
self.mask_downscaling = nn.Sequential(
    nn.Conv2d(1, mask_in_chans // 4, kernel_size=2, stride=2),
    LayerNorm2d(mask_in_chans // 4),
    activation(),
    nn.Conv2d(mask_in_chans // 4, mask_in_chans, kernel_size=2, stride=2),
    LayerNorm2d(mask_in_chans),
    activation(),
    nn.Conv2d(mask_in_chans, embed_dim, kernel_size=1),
)
```

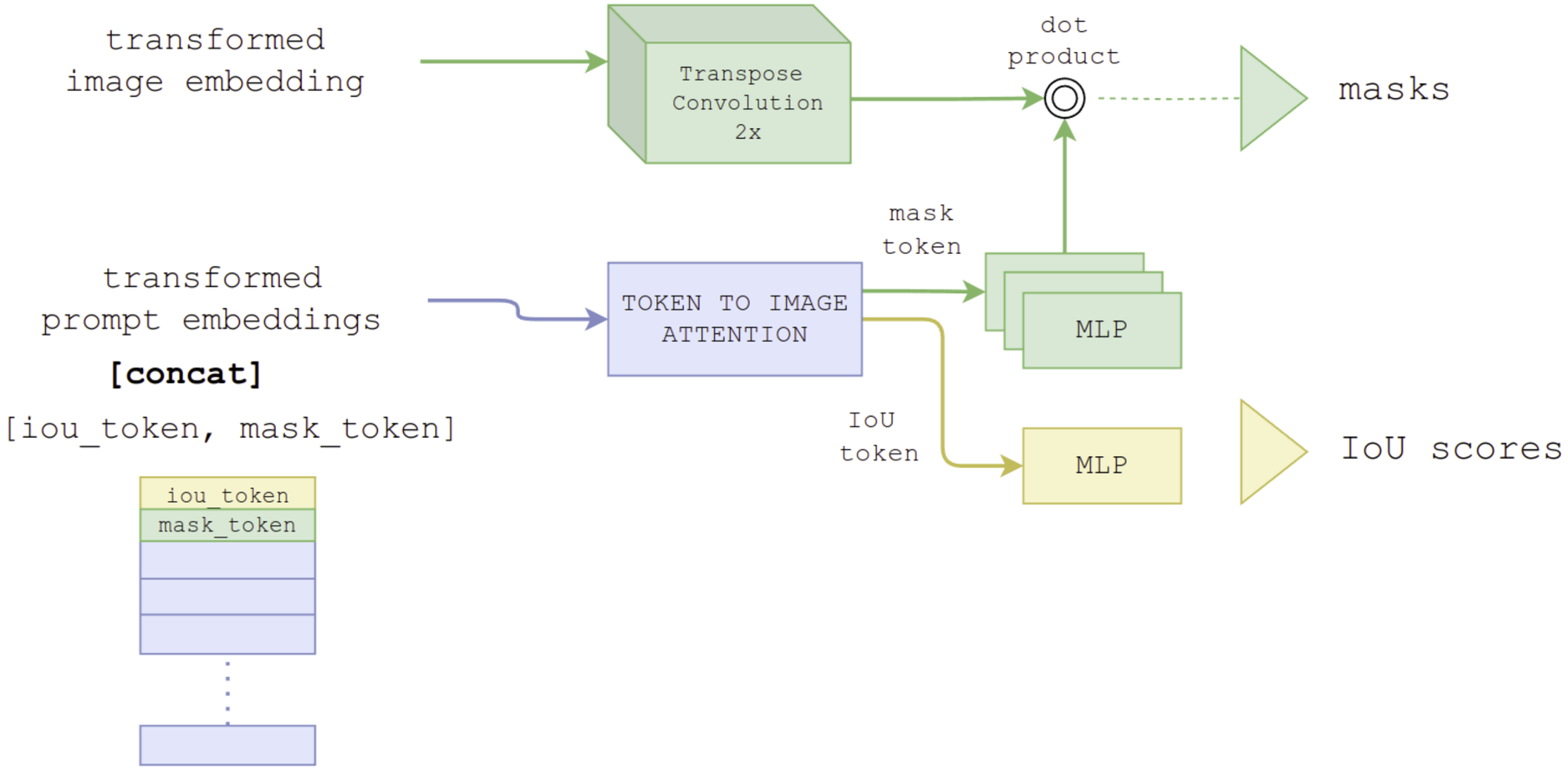
SAM [segmenting anything]

PROMPT-ENCODER



SAM [segmenting anything]

MASK-ENCODER



SAM [segmenting anything]

Muchas Gracias

José Manuel Saavedra Rondo
jmsaavedrar@miuandes.cl