Reference Image & Prompt

Target Images

In-context query

Inter-features

inter-similarity

S

average

⊙ Dot product

Self-attention maps

S

$\{\mathcal{S}'_l\}^L$

flatten

H*W

H*W × H × W

intra-similarity

⊗ element-wise multiplication

element-wise multiplication

Matting Head

alpha

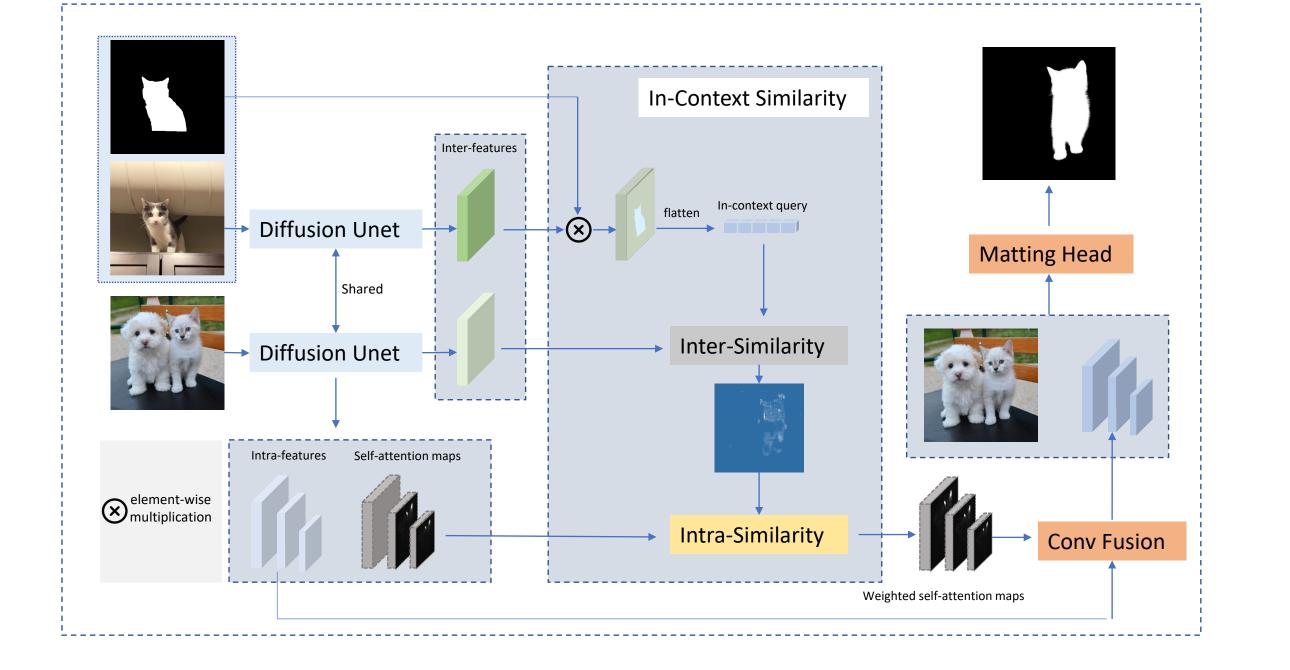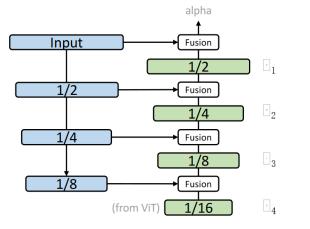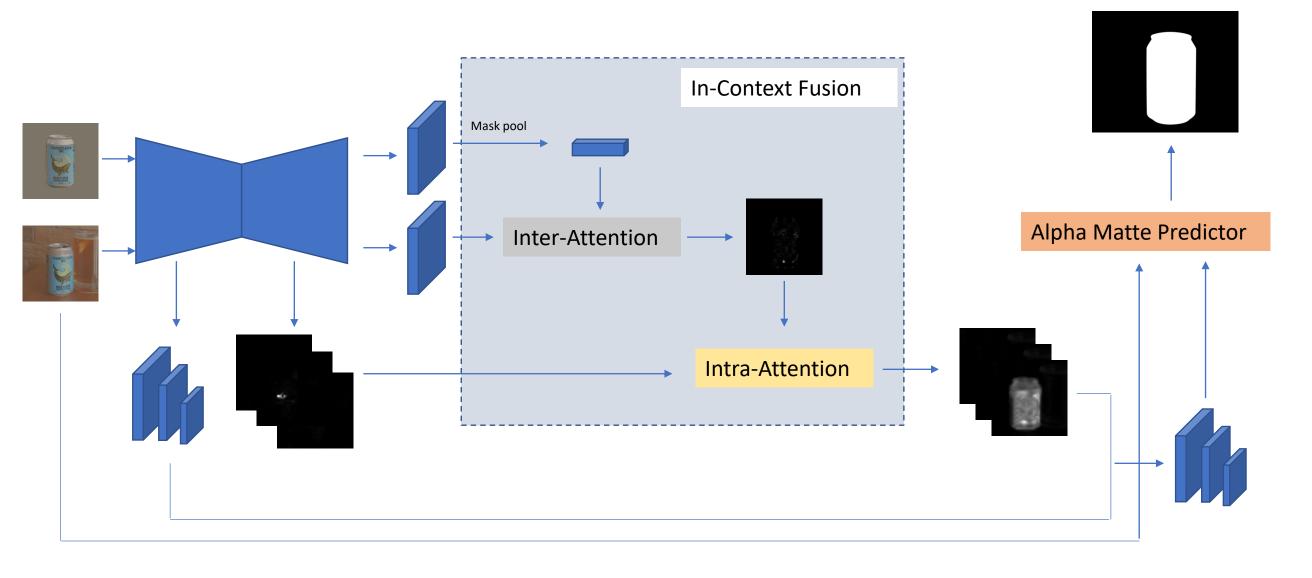| Input | Fusion | 1/2 |
1/2	Fusion	1/4
1/4	Fusion	1/8
1/8	Fusion	1/16 (from ViT)

$$\mathcal{A}_{\text{Inter}} = Softmax\left(\frac{Q \cdot K^{\top}}{\sqrt{d}}\right) \qquad \mathcal{A}_{\text{Intra}} = \mathcal{S} \cdot \mathcal{A}_{\text{Inter}}$$