Image/video matting

Defination

Image. -> background + foreground   (alpha matte)

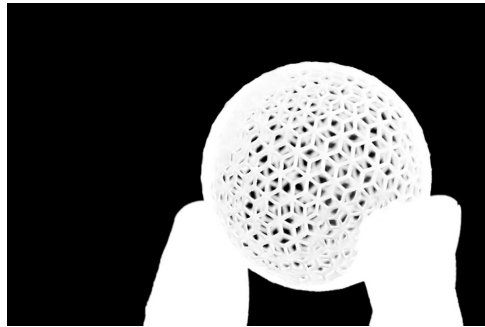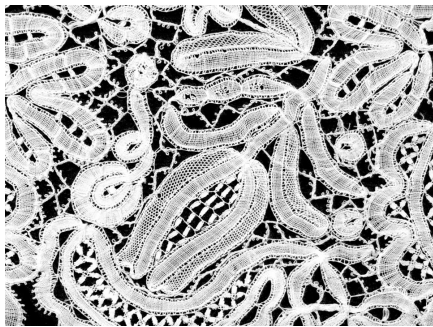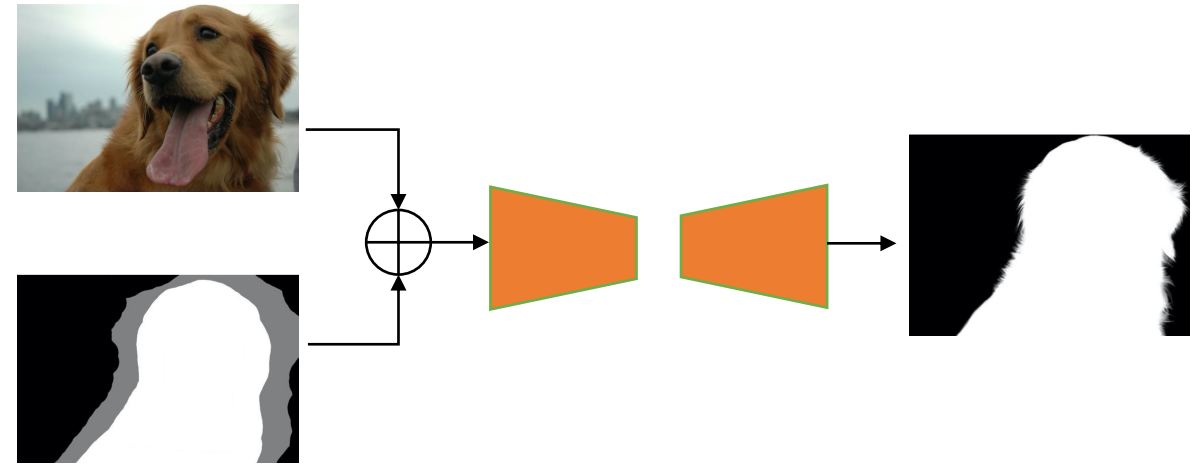$$I_i = \alpha_i F_i + (1 - \alpha_i)B_i, \quad \alpha_i \in [0, 1],$$

Reference:

Trimap (bg/fg/unknown) (DIM, GCA…)

background. (bg matting, bg matting v2)

Mask (mg matting)

Semantic (semantic image matting)

MatteFormer

1. Transformer-based model / reducing computational cost -> Swin transformer (backbone)

2. Prior-token (represents global feature of fg/bg/unknown)

3. Design PAST block (prior-attentive swin transformer)

   PA-WSA (self attention with spatial-tokens and prior-tokens)

   Prior memory at each block

Swin Transformer

1. high-resolution computation -> non-overlap small-size patches
2. Shift windows

(a) Architecture

(b) Two Successive Swin Transformer Blocks
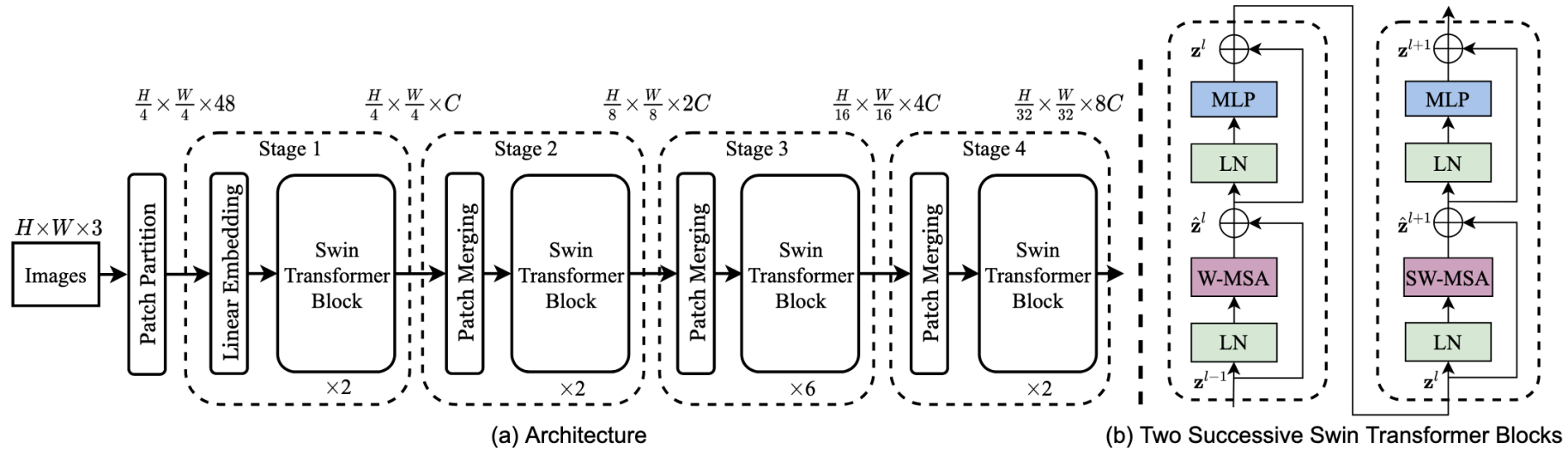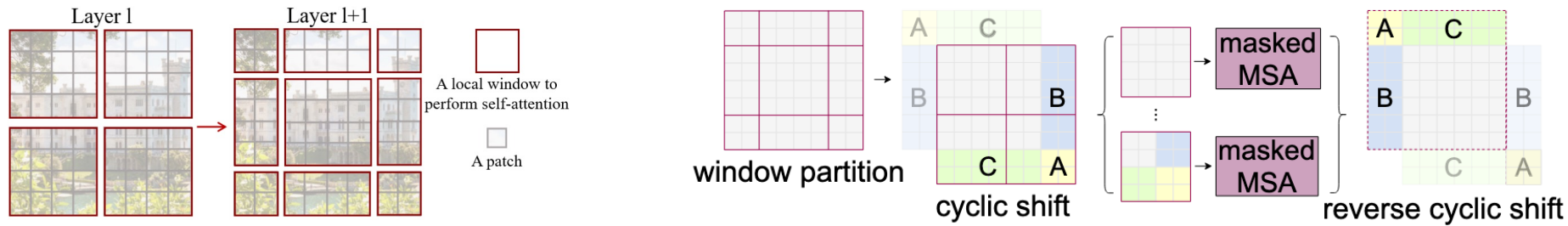
MatteFormer

1. prior-tokens

$$\mathbf{p}^q = \frac{1}{N_q} \sum_{i=1}^{N} r_i^q \cdot \mathbf{z}_i, \quad q \in \{\mathbf{fg}, \mathbf{bg}, \mathbf{uk}\},$$
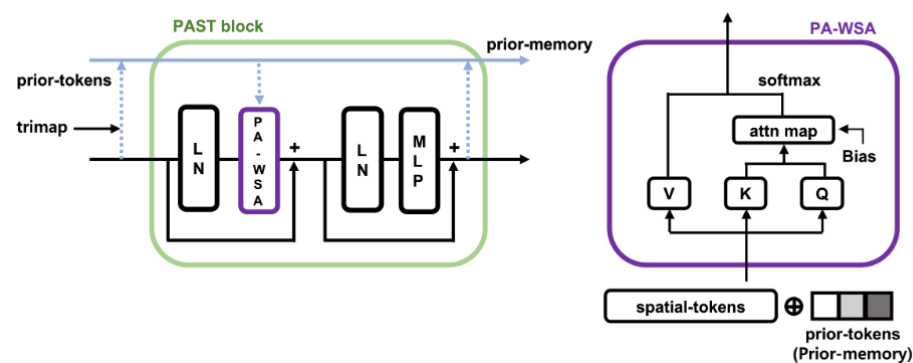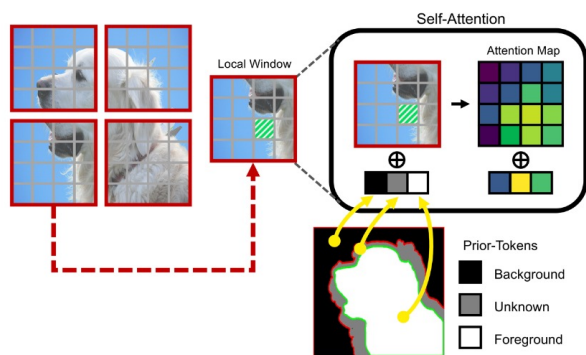
2. Network
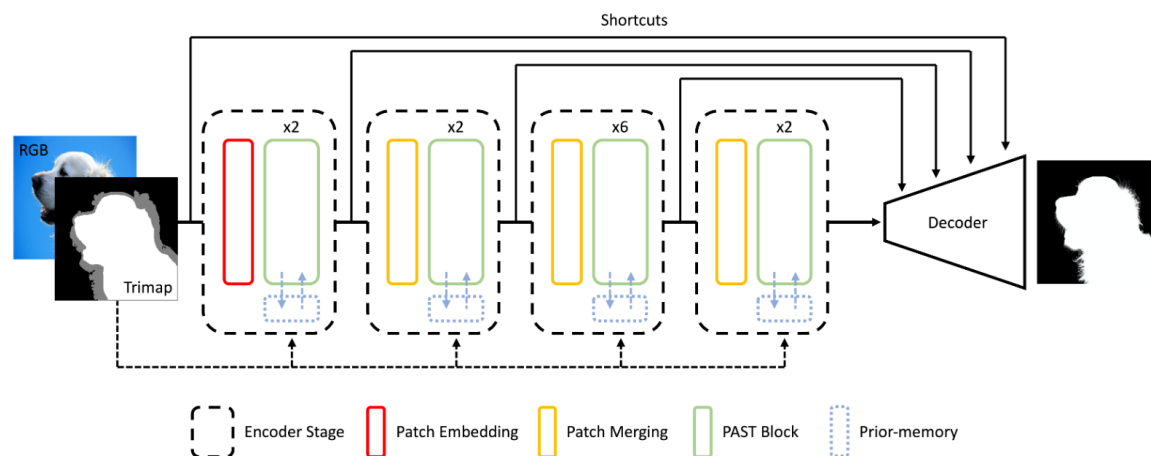
MatteFormer

1. Encoder: swin transformer

   Decoder: Progressive Refinement Module

$$\alpha_l = \alpha_l{}' \odot g_l + \alpha_{l-1} \odot (1 - g_l),$$

$$g_l(x, y) = \begin{cases} 1, & \text{if } 0 < \alpha_{l-1}(x, y) < 1 \\ 0, & \text{otherwise} \end{cases},$$

2. Loss

$$L_{total} = L_{l1} + L_{comp} + L_{lap}$$

up   up

os8

os4

os1

Pred alpha

MatteFormer

Results. Based composition-1k

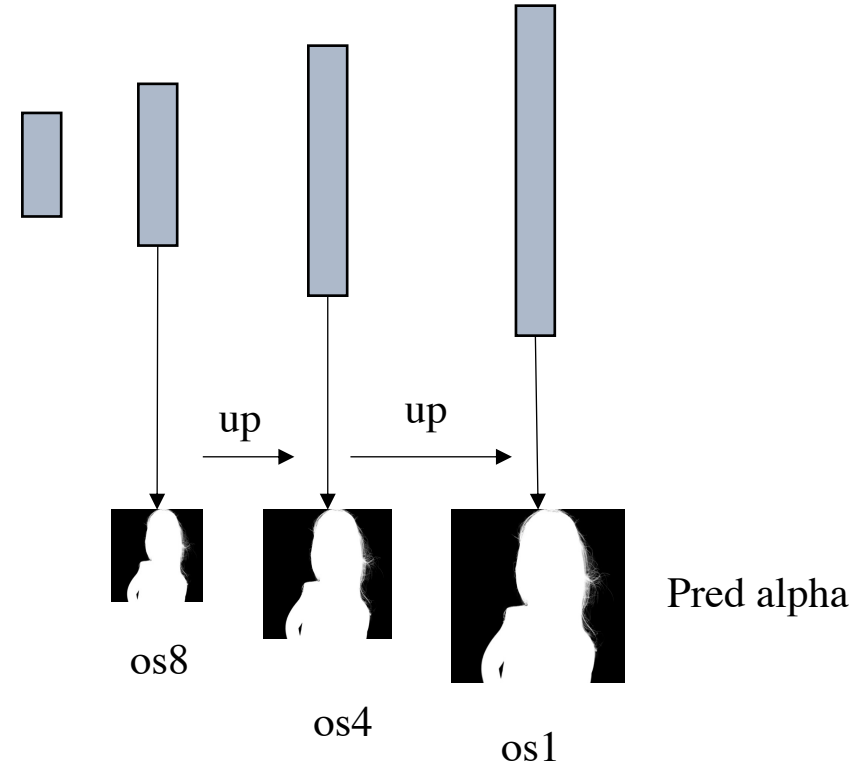| Method	SAD	MSE (10^{-3})	Grad	Conn
Learning Based Matting [59]	113.9	48	91.6	122.2
Closed-Form Matting [20]	168.1	91	126.9	167.9
KNN Matting [6]	175.4	103	124.1	176.4
Deep Image Matting [51]	50.4	14	31.0	50.8
AlphaGan [33]	52.4	30	38.0	-
IndexNet [31]	45.8	13	25.9	43.7
HAttMatting [34]	44.0	7.0	29.3	46.4
AdaMatting [2]	41.7	10.0	16.8	-
SampleNet [40]	40.4	9.9	-	-
Fine-Grained Matting [27]	37.6	9.0	18.3	35.4
Context-Aware Matting [18]	35.8	8.2	17.3	33.2
GCA Matting [22]	35.3	9.1	16.9	32.5
HDMatt [53]	33.5	7.3	14.5	29.9
MG Matting [54]	31.5	6.8	13.5	27.3
MG Matting-trimap*	28.9	5.7	11.4	24.9
MG Matting-trimap,res50*	28.4	5.4	11.1	24.3
TIMINet [29]	29.1	6.0	11.5	25.4
SIM [39]	28.0	5.8	10.8	24.8
Ours (MatteFormer)	**23.8**	**4.0**	**8.7**	**18.9**

Method	SAD	MSE (10^{-3})	Grad	Conn
Baseline (no prior-token)	26.43	5.20	9.57	21.89
Baseline + GAP prior-token	25.30	4.72	9.63	20.61
Baseline + uk prior-token	24.70	4.46	9.10	19.73
+ uk/fg/bg prior-token	24.19	4.05	8.72	19.19
+ prior-memory (MatteFormer)	23.80	4.03	8.68	18.90

Methods	w ASPP		w/o ASPP	
	SAD	MSE (10^{-3})	SAD	MSE (10^{-3})
Baseline (no prior-token)	26.97	5.35	26.43	5.20
+ u.k prior-token	25.60	4.68	24.71	4.46
+ u.k/f.g/b.g prior-token	25.52	4.38	24.19	4.05
+ prior-memory (MatteFormer)	25.15	4.30	23.80	4.03

MatteFormer

visualization

| Image | Trimap | IndexNet [31] | GCA Matting [22] | MG Matting [54] | MatteFormer (Ours) | GT |

Paper link

Matteformer:

https://arxiv.org/pdf/2203.15662.pdf

Rmat:

https://arxiv.org/pdf/2201.06889.pdf