Mask Matching Transformer for Few-Shot Segmentation

Siyu Jiao[1,2]*, **Gengwei Zhang**[3], **Shant Navasardyan**[4], **Ling Chen**[3], **Yao Zhao**[1,2],
Yunchao Wei[1,2], **Humphrey Shi** [4]

[1] Institute of Information Science, Beijing Jiaotong University
[2] Beijing Key Laboratory of Advanced Information Science and Network
[3] AAII, University of Technology Sydney [4] Picsart AI Research (PAIR)
`jiaosiyu@bjtu.edu.cn`

1 Appendix

1.1 Generalization of Feature Alignment Block

We experimentally demonstrate the generalization of Feature Alignment Block (FAB). FAB is applied to CyCTR and HSNet to verify the generalization ability of FAB (Tab. 1).

Table 1: Generalization of **FAB**

	CyCTR	CyCTR-128	HSNet
Original	64.0	63.5	64.0
With FAB	64.3	64.1	64.2

CyCTR+FAB: We directly insert FAB to CyCTR before the cycle-consistent transformer. With our FAB, the well developed CyCTR can still be improved by 0.3% mIoU (from 64.0% to 64.3%). Besides, following Tab.5 in the CyCTR paper, where they ablate the number of cycle-consistent transformer encoders with 128 hidden dimensions. According to their results, using one more cyc-encoder only provides 0.2% improvement (from 63.5% to 63.7%). With our FAB, CyCTR-128 can be improved by 0.6% mIoU (from 63.5% to 64.1%).

HSNet+FAB: HSNet takes 13 middle-level output features from the backbone to the decoder. Performing Cross Alignment (CA) block at all 13 levels is impossible. Thus we only apply SA to HSNet. SA brings 0.2% mIoU improvement to HSNet.

1.2 Ablation on Learnable Parameters

We vary the learnable parameters in Mask Matching (MM) Module by adjusting the hidden dimension of FFN and the number of transformer blocks in CA and show how they affect the final performance (Tab. 2). We conduct experiments on COCO-20^i, where * denotes the default setting of our MM-Former, d is the hidden dimension of FFN and L is the number of transformer layers.

Table 2: Ablation on **Learnable Parameters** (2^{nd} stage)

(d, L)	(256, 1)	(512, 1)	(512, 2)*	(768, 3)	(1024, 4)
mIoU	43.4	43.2	44.2	42.8	42.8
Learnable Parameters	1.4M	1.8M	2.6M	4.6M	7.3M

---

[*] Work done during an internship at Picsart AI Research (PAIR).

36th Conference on Neural Information Processing Systems (NeurIPS 2022).