

Supplementary Materials of α -Former: Local-Feature-Aware (L-FA) Transformer

1 MORE ABLATION STUDIES

1.1 INFLUENCE OF THE ENCODER LAYERS AND DECODER LAYERS

In the table.1, we compare the influence of using different encoder and decoder layers in our architecture. We can see that with the increase of encoder and decoder layers, the performance will first improve and then maintain a similar performance. So, in our architecture, we use six layers of encoder and three layers of decoder.

Table 1: Comparison with the traditional descriptor, the best results are highlighted in **bold**.

encoder	decoder	COD10K			NC4K		
		AP	AP50	AP75	AP	AP50	AP75
1	3	37.256	68.755	37.982	39.453	69.538	40.453
3	1	38.453	70.188	39.423	40.020	70.358	41.168
3	3	40.421	70.861	40.453	41.093	71.592	42.048
3	6	41.424	72.826	40.826	41.726	72.059	42.824
6	3	42.453	72.735	41.758	42.936	72.905	43.278
6	6	42.187	72.682	41.744	42.921	72.723	43.168
6	9	42.424	72.672	41.776	42.876	72.781	43.133

1.2 ABLATION STUDIES OF USING DIFFERENT BACKBONE

In the table.2, we compare the performance of using different backbones in our architecture.

Table 2: Comparison with the traditional descriptor, the best results are highlighted in **bold**.

Backbone	COD10K			NC4K		
	AP	AP50	AP75	AP	AP50	AP75
Resnet-50(Default)	42.453	72.735	41.758	42.936	72.905	43.278
Resnet-18	36.489	67.159	37.188	37.458	68.711	38.950
Resnet-101	43.188	73.725	42.713	43.794	72.313	44.484
Vgg-16	37.148	68.469	37.195	39.948	69.159	40.152

2 MORE IMPLEMENT DETAILS

2.1 MORE DETAILS OF THE FEATURE AGGREGATION ADAPTER

Our feature aggregation adapter uses a tiny initial value to guarantee at the beginning of the training, the output domain is the same as the input image domain. Specifically, we set the mean and the variance value of the convolution weight as 0 and 0.001, and the bias value of the convolution layer as 0. Using the tiny-initialized convolution layer and the skip connection, we can know that the output of the adapter is almost the same as the input at the beginning of the training.

2.2 MORE DETAILS OF THE EDGE-AWARE FEATURE FUSION MODULE

In this section, we provide more details about our edge-aware feature fusion module. Our edge-aware feature fusion module uses multi-scale features to predict the boundary of the target object. As shown in table.3, we provide the input and output shapes of the different edge prediction blocks.

Table 3: Input and output shape of different edge prediction block

Block	Input Shape	Output Shape
block ₅	$\frac{H \times W}{32}$	$\frac{H \times W}{16}$
block ₄	$\frac{H \times W}{16}$	$\frac{H \times W}{8}$
block ₃	$\frac{H \times W}{8}$	$\frac{H \times W}{4}$
block ₂	$\frac{H \times W}{4}$	$\frac{H \times W}{4}$

2.3 MORE DETAILS OF THE PREDICTION HEAD

In this section, we provide more details about our prediction head. We follow the same architecture as OSFormerPei et al. [2022]. As shown in Fig.1. During the training process, we use a fully-connected layer to calculate the location label. At the same time, we use a multi-layer perceptron to calculate the instance-aware parameters. Then we assign positive and negative locations using ground truth. During the testing process, we use a confidence score of the location label to filter ineffective parameters of the instance-aware parameters. Then we use two linear layers to calculate the weight and bias to calculate the segmentation mask. Then we use an up-sampling operation to get the final prediction masks.

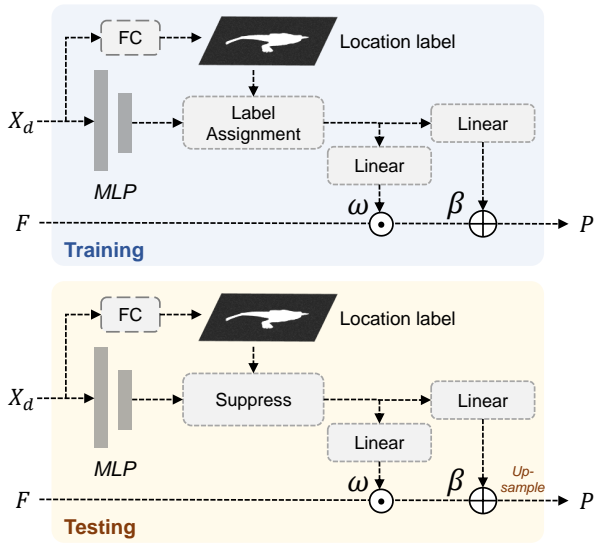


Figure 1: During the training process, our prediction head uses location labels as supervision, and during the testing process, our prediction head uses location labels to filter ineffective parameters.

3 MORE VISUALIZATIONS

As shown in Fig.2, we provide more visualizations in this section.

References

Jialun Pei, Tianyang Cheng, Deng-Ping Fan, He Tang, Chuanbo Chen, and Luc Van Gool. Osformer: One-stage camouflaged instance segmentation with transformers. In *European Conference on Computer Vision*, pages 19–37. Springer, 2022.

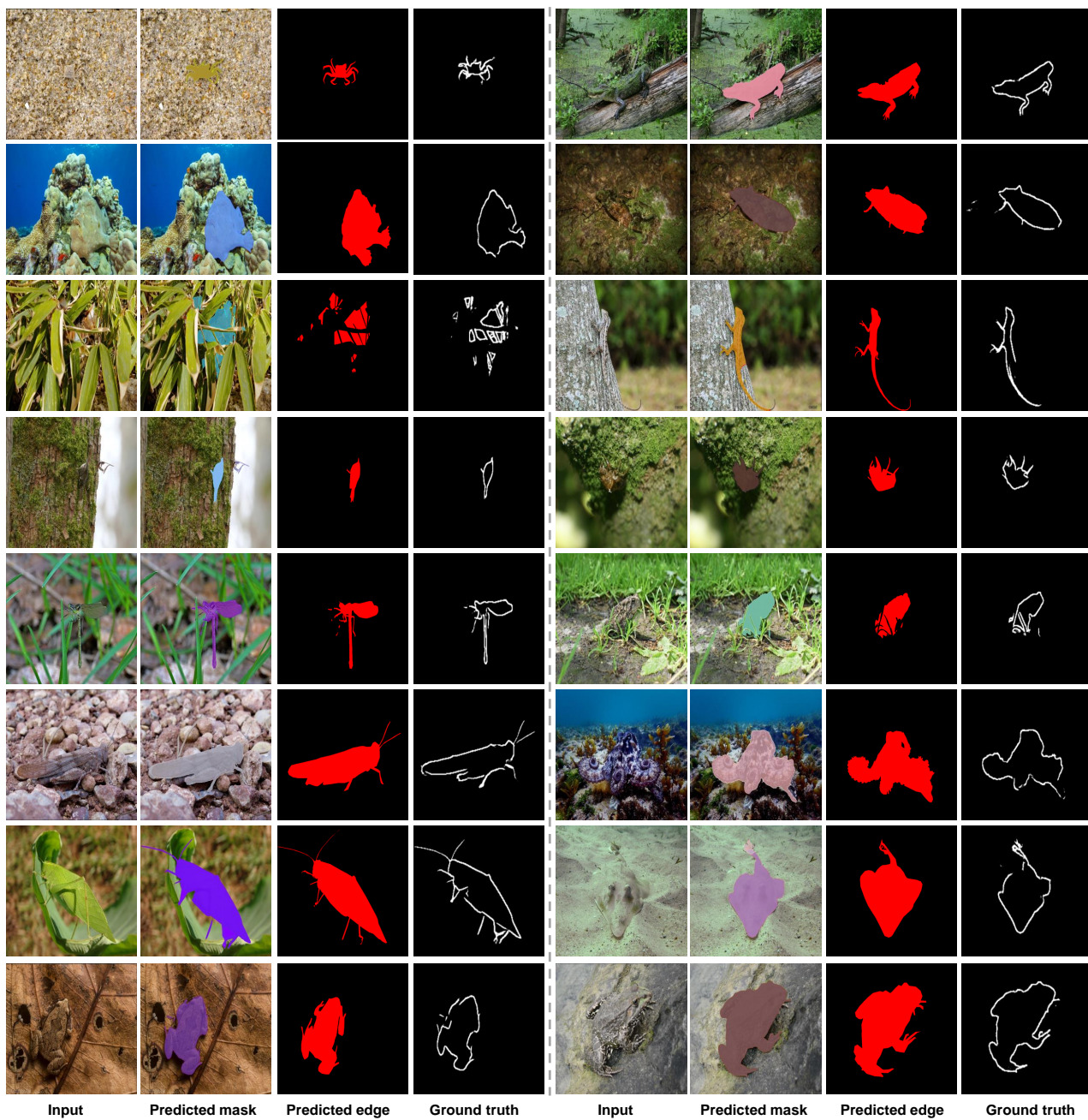


Figure 2: The qualitative results of α -Former.