Supplementary Material for
MIANet: Aggregating Unbiased Instance and General Information for Few-Shot Semantic Segmentation

1. Experiments

1.1. Implement details

(1) In the hierarchical prior module (HPM) of MIANet, the size of M_{ins} is $\{(60, 60), (30, 30), (15, 15), (8, 8)\}$, which is consistent with PFENet [6].

(2) In the general information module (GIM), the *middle-level features* are obtained by concatenating the intermediate features of backbone. For instance, we get the middle-level features of ResNet50 through concatenating the features from block 2 and block 3 [7]. The middle-level feature dimension c is 256.

1.2. Comparison with State-of-the-art Methods

First, we list the FB-IoU results in Table 1, where the proposed method can gain great improvement, especially in the case of using the VGG16.

Then we report the results in Table 2 when the ResNet101 is used as the backbone under 1-shot settings. It can be seen that our approach achieves new state-of-the-art performance and outperforms previous state-of-the-art result by 1.43%.

Table 1. Performance comparison in terms of FB-IoU. The results are the averaged FB-IoU scores of all the four folds. "VGG" means the backbone of VGG16, and "ResNet" means ResNet50.

Datasets	Methods	1-shot		5-shot	
		VGG	ResNet	VGG	ResNet
PASCAL-5^i	PFENet [6]	72.00	73.30	72.30	73.90
	HSNet [5]	73.40	76.70	76.60	80.60
	DPCN [2]	73.70	78.00	77.20	80.70
	BAM [1]	77.26	**81.10**	79.71	82.18
	NTRENet [3]	73.10	77.00	74.20	78.40
	MIANet	**79.22**	79.54	**82.69**	**82.20**
COCO-20^i	HSNet [5]	-	68.20	-	70.70
	DPCN [2]	62.50	63.20	66.10	67.40
	NTRENet [3]	-	68.50	-	69.20
	MIANet	**71.01**	**71.51**	**73.81**	**73.13**

Table 2. Performance comparison on PASCAL-5^i when using ResNet101.

Margin	Fold-0	Fold-1	Fold-2	Fold-3	mIoU
PFENet [6]	60.50	69.40	54.40	55.90	60.10
HSNet [5]	67.30	72.30	62.00	**63.10**	66.20
NTRENet [3]	65.50	71.80	59.10	58.30	63.70
MIANet	**68.54**	**76.34**	**64.92**	60.70	**67.63**

Table 3. Ablation studies of the averaging strategy.

Average	Fold-0	Fold-1	Fold-2	Fold-3	mIoU
	63.84	72.75	67.44	60.38	66.10
✓	65.42	73.58	67.76	61.65	**67.10**

Table 4. Ablation studies of the pretrained strategy.

Pretrained	Fold-0	Fold-1	Fold-2	Fold-3	mIoU
	63.56	72.92	65.48	58.18	65.03
✓	65.42	73.58	67.76	61.65	**67.10**

1.3. Ablation study

We conduct extra ablation studies to validate the impact of our designs. Note that the experiments in this section are performed on PASCAL-5^i dataset using the VGG16 backbone unless specified otherwise. And the evaluation metric is mean-IoU.

Effect of the averaging strategy. In MIANet, we average the negative set since the elements in the background of the support images are very complex. We show the result in Table 3 if the averaging strategy is not implemented. Averaging the background elements brings a 1% performance gain.

Effect of the pretrained strategy. Current s-o-t-a methods [1, 4] usually adopt the pretrained strategy to pretrain the backbone before meta-training. We conduct the experiment in Table 4 which demonstrates the effectiveness of the strategy.

Table 5. Ablation studies of the margin in triplet loss on PASCAL-5^i when using ResNet50.

Margin	Fold-0	Fold-1	Fold-2	Fold-3	mIoU
0.1	67.69	**76.30**	67.09	61.84	68.23
0.2	66.75	75.32	**67.82**	**63.20**	68.27
0.5	**68.51**	75.76	67.46	63.15	**68.72**
1	68.32	75.23	66.72	62.47	68.19

Table 6. Ablation studies of the metric tools.

Methods	Fold-0	Fold-1	Fold-2	Fold-3	mIoU
cosine distance	62.65	72.51	68.72	56.67	65.14
euclidean distance	65.42	73.58	67.76	61.65	**67.10**

Effect of the margin. We report the ablation study about how to choose the margin in our proposed triplet loss, whose results are listed in 5. The best result is achieved when the margin is 0.5.

Effect of the metric tools in the triplet loss. In the triplet loss, euclidean distance is used as our metric tool to calculate the distance of triplets. We investigate two types of metric tools, i.e. euclidean distance and cosine distance. The results are listed in Table 6. The euclidean distance leads the performance by 1.96%. As Figure 1 shows, euclidean distance makes MIANet learn better from the hard triplets. When using the cosine distance, the value of the triplet loss is maintained around 0.5 (**margin**), which means that the triplet loss cannot distinguish the positive samples and negative samples well.

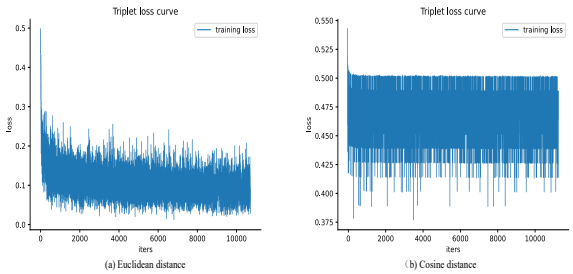

Figure 1. Visual display of the triplet loss in training when using different metric tools.

1.4. More Visualizations

We demonstrate more qualitative results in Figure 2. Moreover, some **failure cases** are also provided in Figure 3. As the Figure 3 shows, we can conclude that (1) intra-class differences seriously affect the segmentation performance, especially the cases of perspective distortion (2nd, 3rd, and 7th columns). (2) The segmentation of small objects is also unsatisfactory (1st and 2nd columns). (3) The bias to the base classes is still an urgent problem in few-shot segmentation (5th and 6th columns). How to more effectively deal with these problems requires better modeling of changes in views, pose and occlusion.

References

[1] Chunbo Lang, Gong Cheng, Binfei Tu, and Junwei Han. Learning what not to segment: A new perspective on few-shot segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8057–8067, 2022. 1

[2] Jie Liu, Yanqi Bao, Guo-Sen Xie, Huan Xiong, Jan-Jakob Sonke, and Efstratios Gavves. Dynamic prototype convolution network for few-shot semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11553–11562, 2022. 1

[3] Yuanwei Liu, Nian Liu, Qinglong Cao, Xiwen Yao, Junwei Han, and Ling Shao. Learning non-target knowledge for few-shot semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11573–11582, 2022. 1

[4] Zhihe Lu, Sen He, Xiatian Zhu, Li Zhang, Yi-Zhe Song, and Tao Xiang. Simpler is better: Few-shot semantic segmentation with classifier weight transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8741–8750, 2021. 1

[5] Juhong Min, Dahyun Kang, and Minsu Cho. Hypercorrelation squeeze for few-shot segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6941–6952, 2021. 1

[6] Zhuotao Tian, Hengshuang Zhao, Michelle Shu, Zhicheng Yang, Ruiyu Li, and Jiaya Jia. Prior guided feature enrichment network for few-shot segmentation. *IEEE Annals of the History of Computing*, (01):1–1, 2020. 1

[7] Chi Zhang, Guosheng Lin, Fayao Liu, Rui Yao, and Chunhua Shen. Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5217–5226, 2019. 1

Figure 2. Qualitative results of our method MIANet and baseline on PASCAL-5^i and COCO-20^i benchmarks. Zoom in for details.

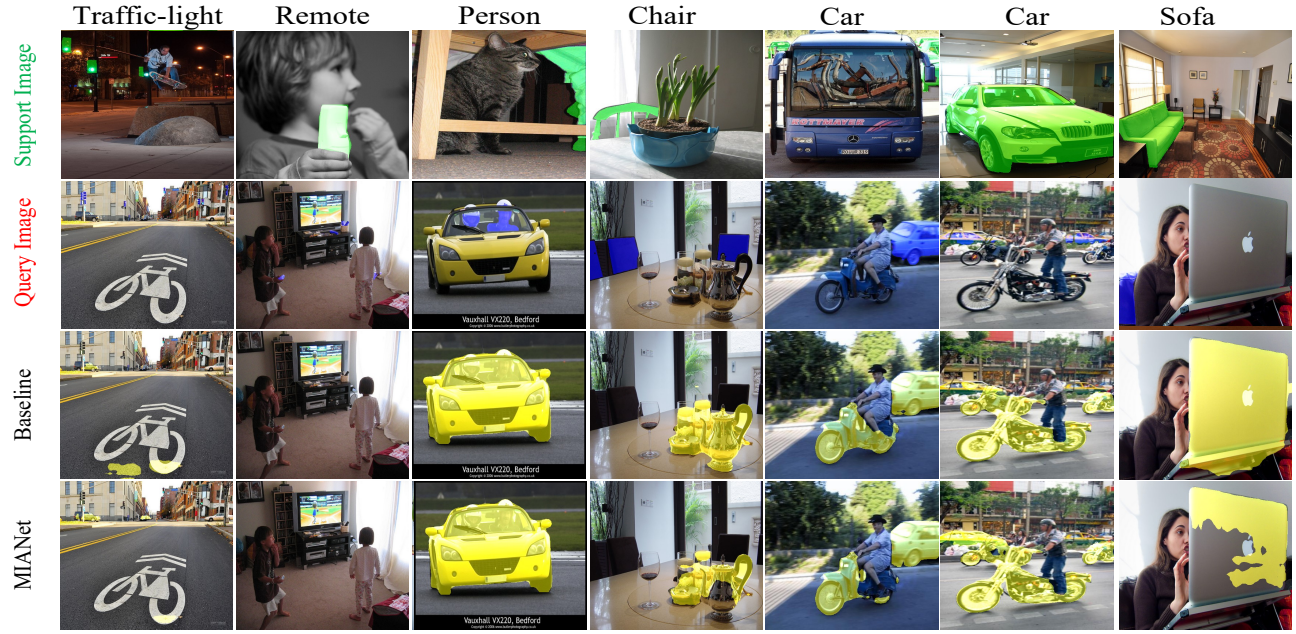

Figure 3. **Failure results** of our method MIANet and baseline on PASCAL-5^i and COCO-20^i benchmarks. Zoom in for details.