SGT: SELF-GUIDED TRANSFORMER FOR FEW-SHOT SEMANTIC SEGMENTATION

Kangkang Ai, Haigen Hu*, Qianwei Zhou, Qiu Guan
Zhejiang University of Technology,Hangzhou, PR China.
***Corresponding author: hghu@zjut.edu.cn**

Summary

This paper proposes a novel method for few-shot segmentation.

A Self-Guided Transformer (SGT) is proposed by leveraging intra-image similarity to improve intra-object inconsistencies. The proposed SGT can selectively guide segmentation, emphasizing the regions that are easily distinguishable while adapting to the challenges caused by less discriminative regions within objects.

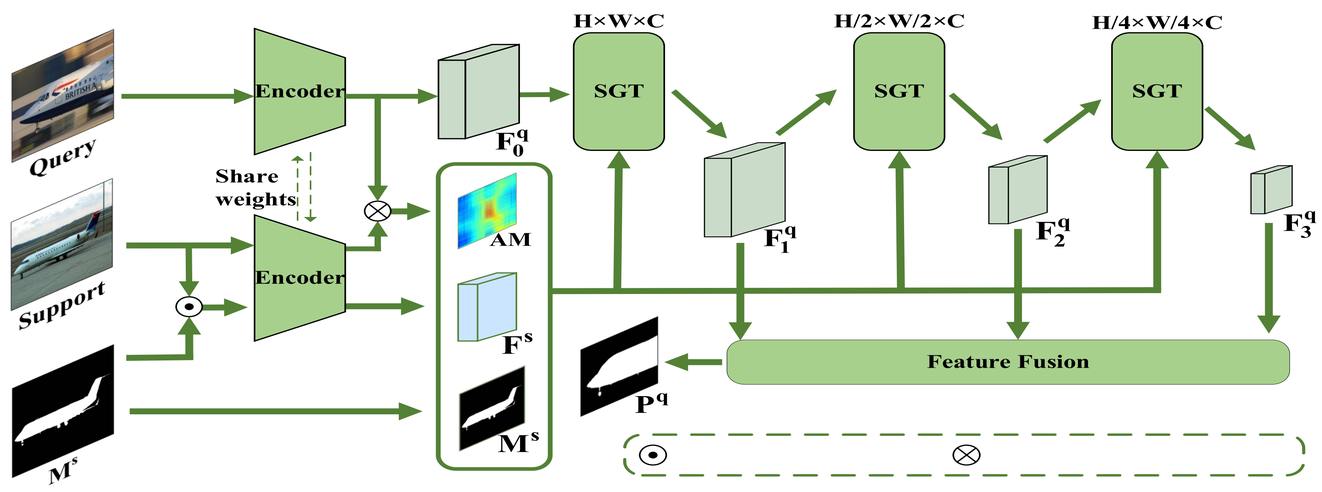
- A novel self-guided transformer module is proposed to solve the problem of feature differences within objects in query images, thereby achieving alignment between hard-to-distinguish and easy-to-distinguish features.
- A series of experiments are conducted to verify the effectiveness of the proposed method, showing that the proposed SGT can achieve state-of-the-art results on several FSS datasets.

Motivation

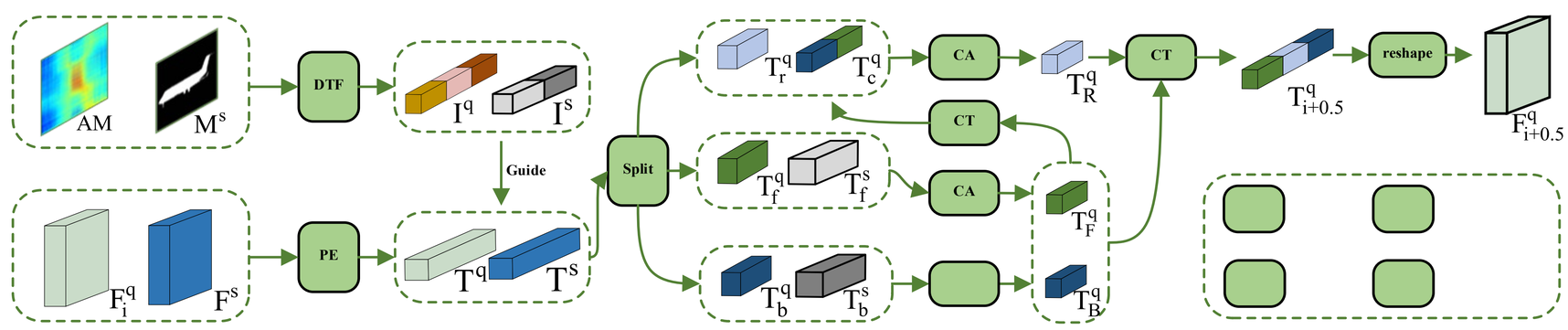
(a) Support **(b) Query** **(c) QS** **(d) QQ**

The figure above shows that the query prototype can activate more regions than the support prototype, but due to the inconsistency of the internal features of the object, the query prototype can also only activate parts of the target.

Method



Overview: The overall framework adopts a two-branch structure with a shared backbone network, where the query features are augmented at different scales by the SGT module, the features at different scales are fused by a feature fusion module, and the segmentation of the query image is finally realized by a simple classifier.



SGT: Based on the information provided by the affinity matrix (AM), SGT will first divide the query features according to the segmentation difficulty and perform different operations on different types of features.

Experiments and Results

Methods	1-Shot					5-shot				
	Fold-0	Fold-1	Fold-2	Fold-3	Mean	Fold-0	Fold-1	Fold-2	Fold-3	Mean
VGG-16 Backbone										
PANet [5]	42.3	58.0	51.1	41.2	48.2	51.8	64.6	59.8	46.5	55.7
FWB [17]	47.0	59.6	52.6	48.3	51.9	50.9	62.9	56.5	50.1	55.1
HSNet [20]	59.6	65.7	59.6	54.0	59.7	64.9	69.0	64.1	58.6	64.2
PFENet [21]	56.9	68.2	54.4	52.4	58.0	59.0	69.1	54.8	52.9	59.0
BAM [15]	64.4	71.3	67.0	58.6	65.3	67.8	73.8	72.1	**64.6**	**69.6**
HDMNet [14]	64.8	71.4	67.7	56.4	65.1	**68.1**	73.1	**71.8**	64.0	69.3
SGT(Ours)	**66.2**	**72.9**	**67.8**	**58.8**	**66.4**	68.0	**74.6**	71.6	62.7	69.2
ResNet-50 Backbone										
HSNet [20]	64.3	70.7	60.3	60.5	64.0	70.3	73.2	67.4	67.1	69.5
CyCTR [10]	65.7	71.0	59.5	59.7	64.0	69.3	73.5	63.8	63.5	67.5
SSP [12]	60.5	67.8	66.4	51.0	61.4	67.5	72.3	75.2	62.1	69.3
DCAMA [22]	67.5	72.3	59.6	59.0	64.6	70.5	73.9	63.7	65.8	68.5
BAM [15]	69.2	74.7	67.8	61.7	68.4	71.8	75.7	72.0	67.5	71.8
HDMNet [14]	**71.0**	75.4	69.0	62.1	69.4	71.3	76.2	71.3	**68.5**	71.8
SGT(Ours)	70.1	**76.1**	**69.2**	**64.3**	**69.9**	**73.7**	**77.5**	**73.0**	66.3	**72.6**
ResNet-101 Backbone										
PFENet [21]	60.5	69.4	54.4	55.9	60.1	62.8	70.4	54.9	57.6	61.4
CyCTR [10]	69.3	72.7	56.5	58.6	64.3	73.5	74.0	58.6	60.2	66.6
HSNet [20]	67.3	72.3	62.0	63.1	66.2	71.8	74.4	67.0	68.3	70.4
DCAMA [22]	65.4	71.4	63.2	58.3	64.6	70.7	73.7	66.8	61.9	68.3
BAM [15]	69.9	75.4	67.1	62.1	68.6	72.6	77.1	70.7	**69.8**	72.6
SGT(Ours)	**70.3**	**76.6**	**70.4**	**64.8**	**70.5**	**73.2**	**78.0**	**73.7**	66.2	**72.8**

Results: As can be seen from the tables, the performance of our method on competitive with the state-of-the-art methods on all three backbones. Under the 1-shot setting, the average mIoU scores of our method in PASCAL-5i on ResNet-101 backbones are 70.5%, which outperform the state-of-the-art method by 1.9%.

base	CA	$SGT_{K=1}$	$SGT_{K=5}$	triloss	mIoU	FB-IoU
✓					66.4	78.3
✓	✓				67.5	78.9
✓				✓	65.6	77.6
✓			✓		69.3	80.2
✓		✓		✓	69.6	80.6
✓			✓	✓	**69.9**	**80.8**

Backbone	Methods	FB-IoU 1-shot	FB-IoU 5-shot	learnable params
ResNet-50	ASGNet [26]	60.4	67.0	10.4M
	BAM [15]	71.1	73.3	4.9M
	HDMnet* [14]	72.2	74.3	4.2M
	SGT(Ours)	**73.1**	**76.5**	**3.9M**

Results: The ablation experiments show that all the proposed modules help to improve the performance of the few-shot segmentation. And our method can obtain the optimal FB-mIoU while using a smaller number of parameters.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (Grant Nos. 62373324, 62271448 and U20A20171), in part by Zhejiang Provincial Natural Science Foundation of China(Grant Nos. LGF22F030016 and LY21F020027), and in part Key Programs for Science and Technology Development of Zhejiang Province (2022C03113).