Few-Shot Semantic Segmentation

方致远

2022.06.26

- **Remember the Difference: Cross-Domain Few-Shot Semantic Segmentation via Meta-Memory Transfer**
  - Proposing a new setting for few-shot semantic segmentation
- **Integrative Few-Shot Learning for Classification and Segmentation**
  - Proposing a new setting for few-shot semantic segmentation
- **Learning Non-target Knowledge for Few-shot Semantic Segmentation**
  - Eliminating background regions
- Learning What Not to Segment: A New Perspective on Few-Shot Segmentation
  - Eliminating background regions

Remember the Difference: Cross-Domain Few-Shot Semantic Segmentation via Meta-Memory Transfer

Wenjian Wang, Lijuan Duan, Yuxi Wang, Qing En, Junsong Fan, Zhaoxiang Zhang

Cross-Domain Few-Shot Semantic Segmentation via Meta-Memory Transfer

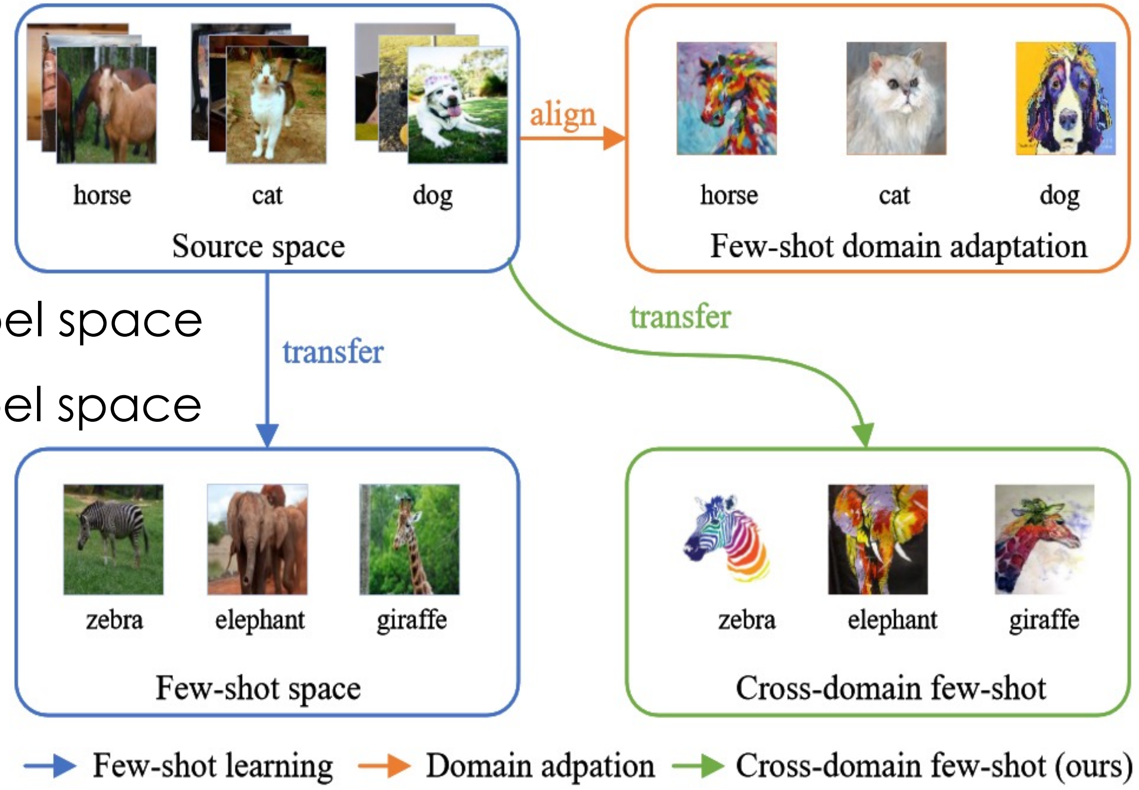
Motivation

Cross domain is a more realistic setting

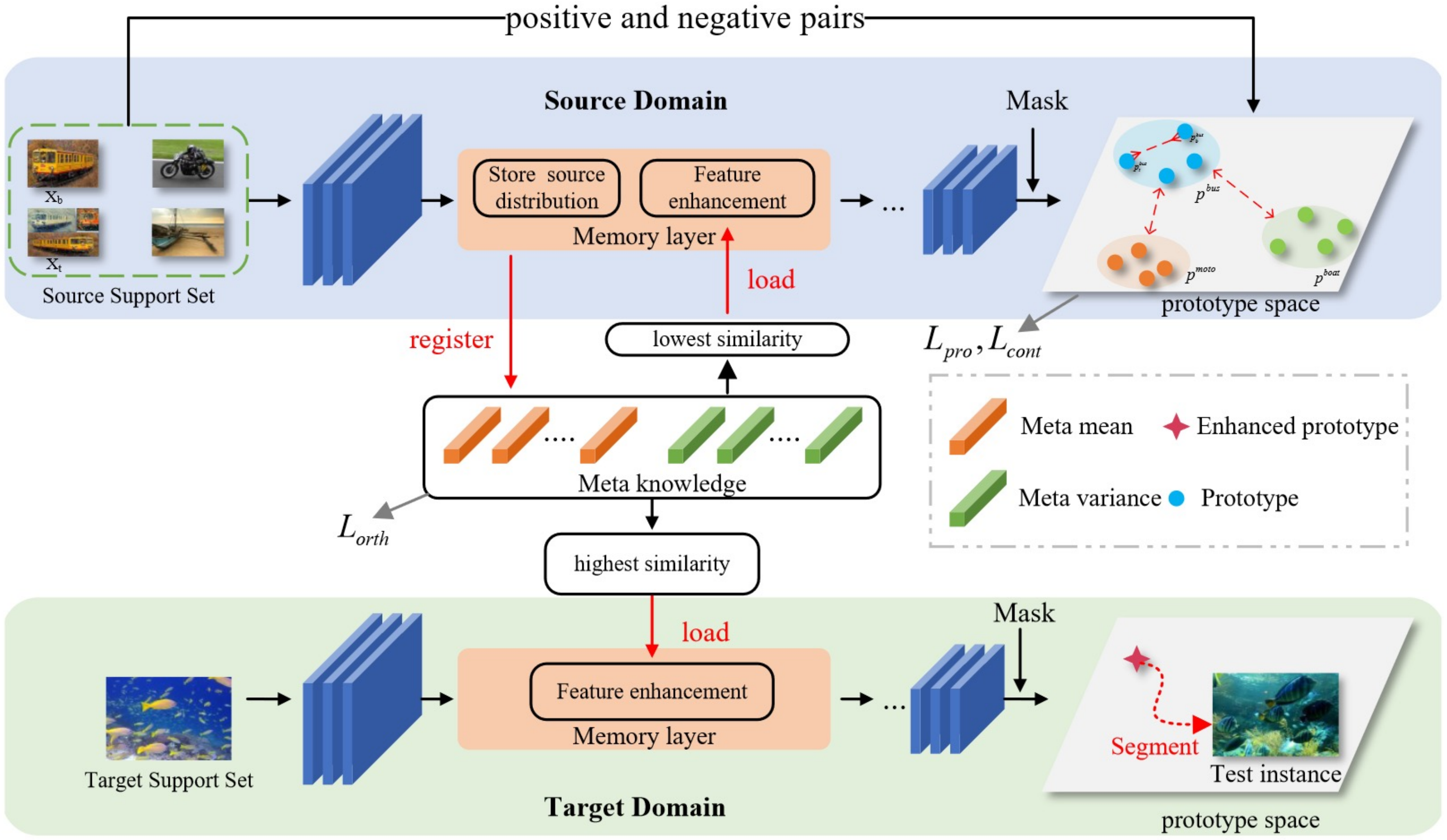
Few-shot: Same domain, Separate label space

Domain adaption: Separate domains, same label space

Cross domain: Separate domains, Separate label space



Cross-Domain Few-Shot Semantic Segmentation via Meta-Memory Transfer



Cross-Domain Few-Shot Semantic Segmentation via Meta-Memory Transfer

Meta memory bank

- $Memory = \{m_j \in R^{1\times C}, e_j \in R^{1\times C}\}_{j=1}^{N}$

$$m_j = \lambda m_j + (1-\lambda)\sum_{b=1}^{B} s_M^{jb}\mu_b$$

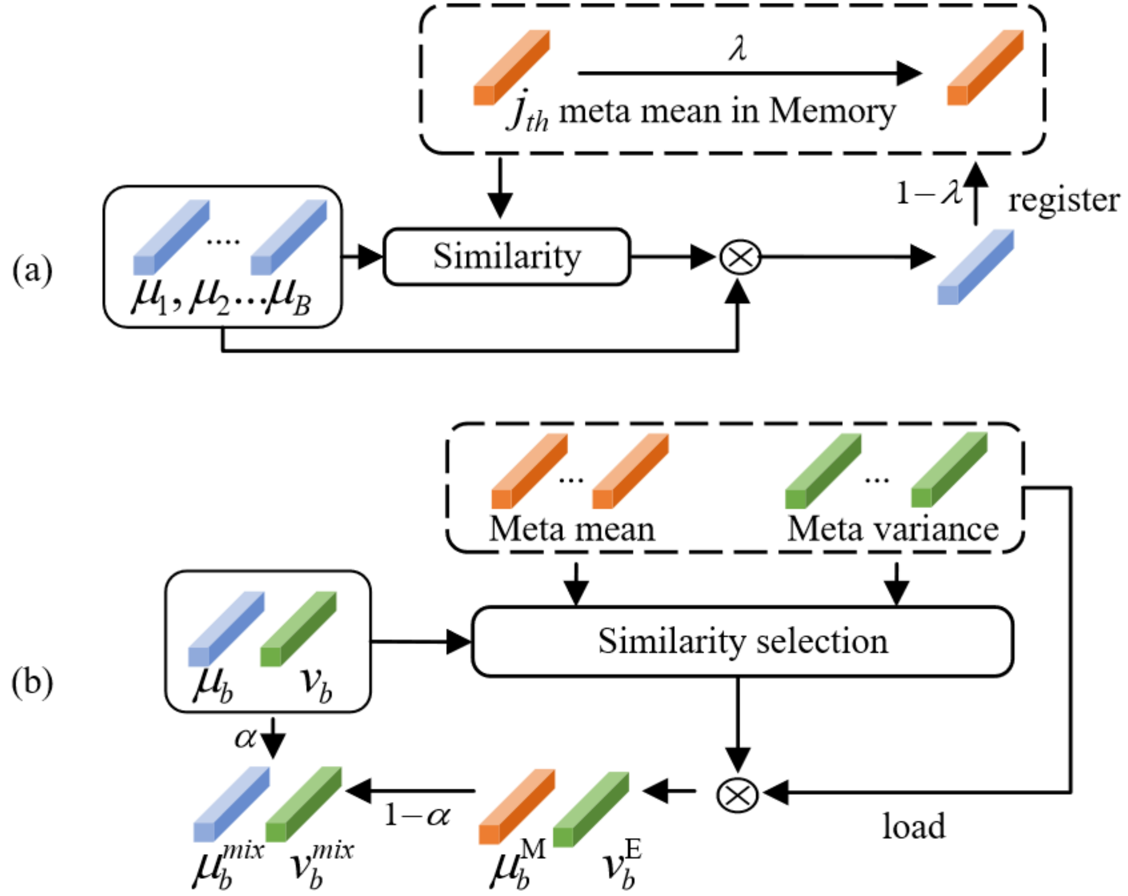
$$e_j = \lambda e_j + (1-\lambda)\sum_{b=1}^{B} s_E^{jb}v_b$$

$$L_{orth} = \frac{1}{2N^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} h_M^{ij} + \sum_{i=1}^{N}\sum_{j=1}^{N} h_E^{ij}\right),$$

- Memory-based Feature Enhancement

$$f_b^{enh} = f_b^{norm} v_b^{mix} + \mu_b^{mix},$$

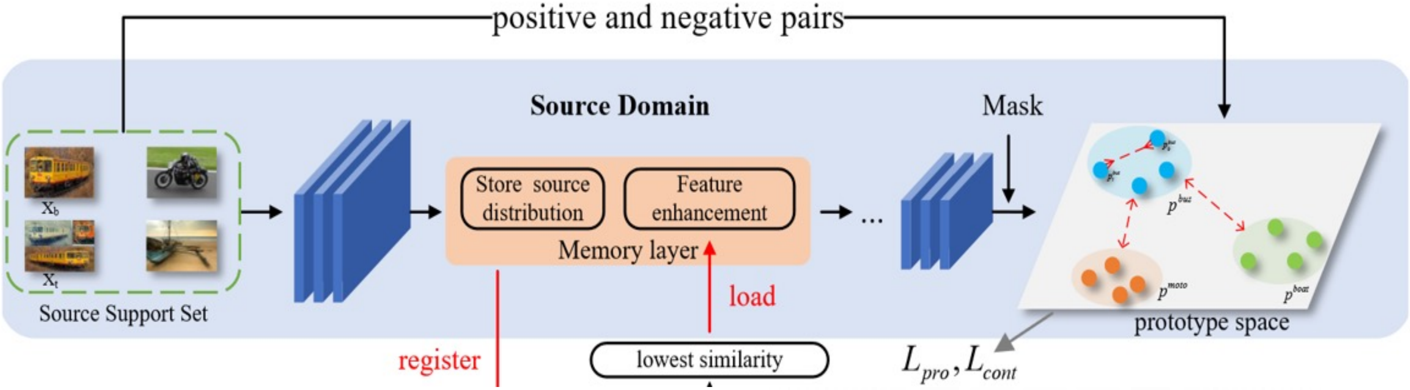
$$\mu_b^{mix} = \alpha\mu_b + (1-\alpha)\mu_b^M, v_b^{mix} = \alpha v_b + (1-\alpha)v_b^E,$$



Cross-Domain Few-Shot Semantic Segmentation via Meta-Memory Transfer

Contrastive loss

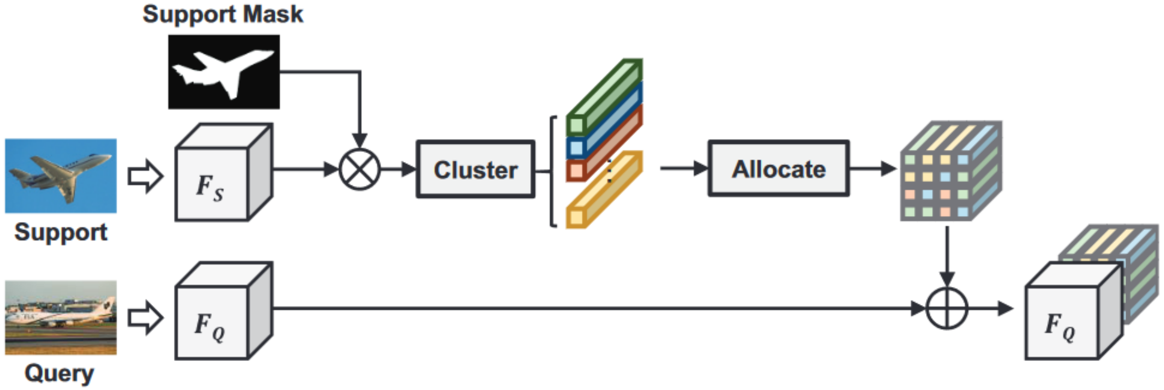
$$L_{cont} = \frac{1}{2|B|} \sum_{i=1}^{|B|} - \log \frac{pos(i)}{pos(i) + neg(i)}$$



Cross-entropy loss

$$L_{pro} = -\frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} Y_q^{ij} \log(R(f_q^{ij}, p_b)),$$

$$L_{all} = L_{orth} + L_{cont} + L_{pro}$$



(b) Adaptive prototype learning and allocation

Adaptive Prototype Learning and Allocation for Few-Shot Segmentation (CVPR 2021)

Cross-Domain Few-Shot Semantic Segmentation via Meta-Memory Transfer

COCO-20^i to SUIM						
Backbone	Methods	split-0	split-1	split-2	split-3	mean
ResNet50	ASGNet [16](CVPR21)	28.1	27.5	26.1	32.3	28.5
	HSNet [24](ICCV21)	**33.8**	35.9	35.3	35.4	35.1
	SCL [47](CVPR21)	27.3	28.8	26.5	25.3	27.0
	Ours	30.5	**38.6**	**42.5**	**36.6**	**37.1**

PASCAL-5^i to SUIM						
Backbone	Methods	split-0	split-1	split-2	split-3	mean
ResNet50	ASGNet [16](CVPR21)	32.4	30.9	28.9	35.2	31.9
	HSNet [24](ICCV21)	30.7	30.0	27.3	27.0	28.8
	SCL [47](CVPR21)	31.3	31.2	32.2	32.5	31.8
	Ours	**35.2**	**33.4**	**34.3**	**36**	**34.7**

5-shot				
split0	split1	split2	split3	Mean
40.2	58.0	55.2	61.8	53.8
57.0	68.0	70.4	76.2	67.9
53.7	**69.8**	67.1	75.9	66.6
-	-	-	-	61.9
60.3	65.8	67.1	72.8	66.5
58.2	65.9	**71.8**	**77.9**	68.4
65.7	69.2	70.8	75.0	**70.1**
43.3	61.2	66.5	70.4	60.3
59.1	69.0	**73.4**	78.7	70.0
67.2	**72.7**	72.0	**78.9**	**72.7**

COCO-20^i to PASCAL-5^i (1-shot)				
Model	Cons	MEnS	MEnT	mean-IoU
BS				57.4
(a)	✓			62.8
(b)	✓	✓		63.5
(c)		✓	✓	62.5
(d)	✓	✓	✓	65.6

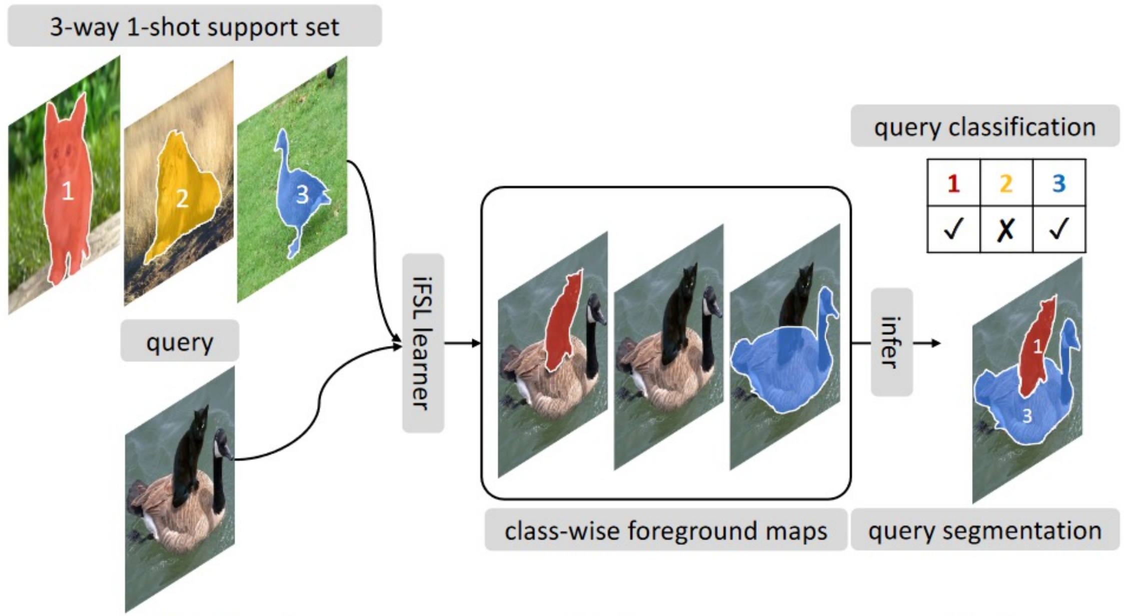
Integrative Few-Shot Learning for Classification and Segmentation

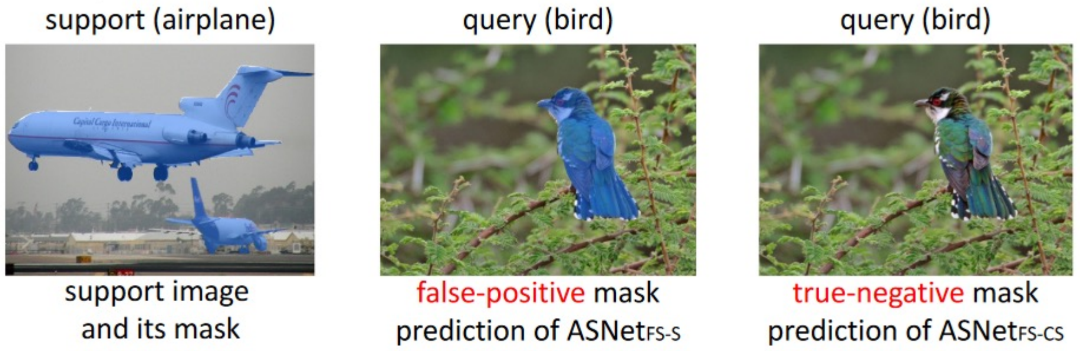
Dahyun Kang, Minsu Cho

Integrative Few-Shot Learning for Classification and Segmentation

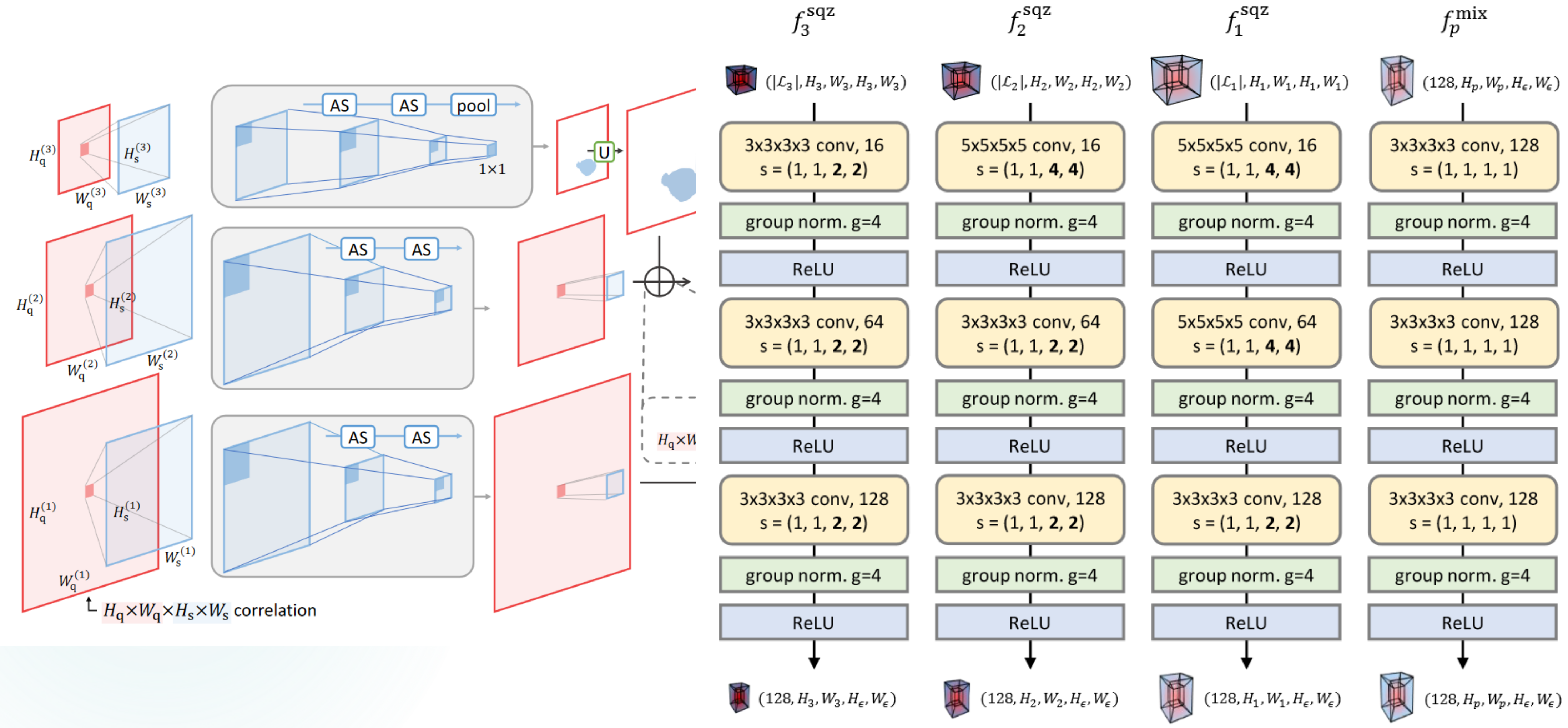
Integrative task of few-shot **classification** and **segmentation**

FS-S learners typically segment out arbitrary **salient objects** in the query, when a query image without any target class is given.





Integrative Few-Shot Learning for Classification and Segmentation



Hypercorrelation Squeeze for Few-Shot Segmentation (ICCV 2021)

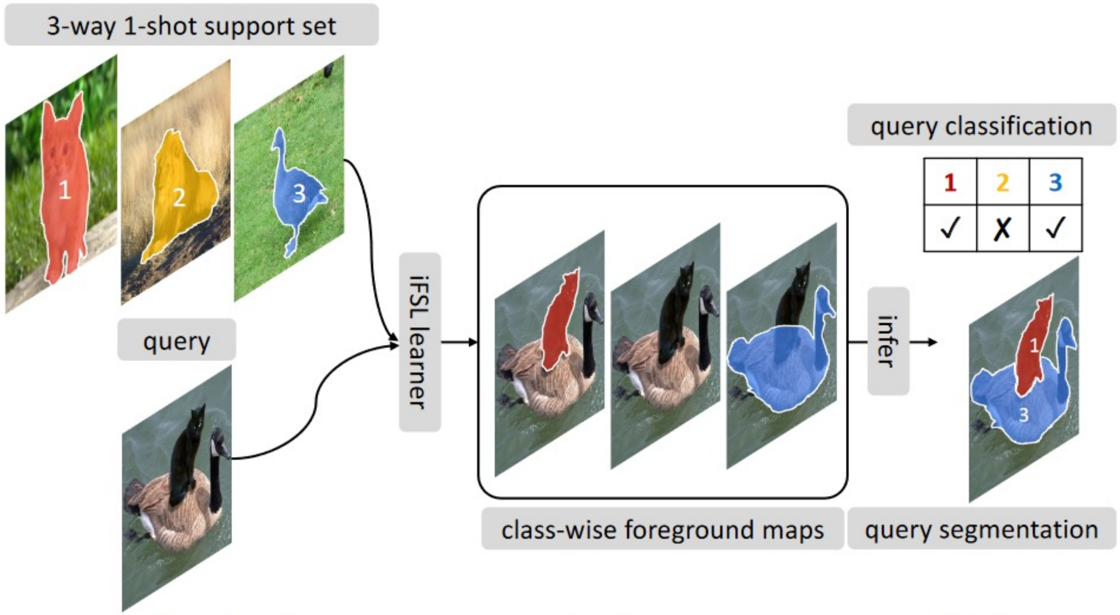
Integrative Few-Shot Learning for Classification and Segmentation

$$\mathcal{L}_{\mathrm{C}} = -\frac{1}{N}\sum_{n=1}^{N} \mathbf{y}_{\mathrm{gt}}^{(n)} \log \frac{1}{HW} \sum_{\mathbf{p}\in[H]\times[W]} \mathbf{Y}^{(n)}(\mathbf{p}),$$

$$\mathbf{Y}_{\mathrm{bg}} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{1} - \mathbf{Y}^{(n)}),$$

$$\mathbf{Y}_{\mathrm{S}} = [\mathbf{Y} || \mathbf{Y}_{\mathrm{bg}}] \in \mathbb{R}^{H\times W\times(N+1)}.$$

$$\mathcal{L}_{\mathrm{S}} = -\frac{1}{(N+1)}\frac{1}{HW}\sum_{n=1}^{N+1}\sum_{\mathbf{p}\in[H]\times[W]} \mathbf{Y}_{\mathrm{gt}}^{(n)}(\mathbf{p}) \log \mathbf{Y}_{\mathrm{S}}^{(n)}(\mathbf{p}),$$



Integrative Few-Shot Learning for Classification and Segmentation

	1-way 1-shot										2-way 1-shot									
	classification 0/1 exact ratio (%)					segmentation mIoU (%)					classification 0/1 exact ratio (%)					segmentation mIoU (%)				
method	5^0	5^1	5^2	5^3	avg.	5^0	5^1	5^2	5^3	avg.	5^0	5^1	5^2	5^3	avg.	5^0	5^1	5^2	5^3	avg.
PANet [78]	69.9	67.7	68.8	69.4	69.0	32.8	45.8	31.0	35.1	36.2	56.2	47.5	44.6	55.4	50.9	33.3	46.0	31.2	38.4	37.2
PFENet [73]	69.8	82.4	68.1	77.9	74.6	38.3	54.7	35.1	43.8	43.0	22.5	61.7	40.3	39.5	41.0	31.1	47.3	30.8	32.2	35.3
HSNet [44]	**86.6**	84.8	76.9	**86.3**	83.7	49.1	59.7	41.0	49.0	49.7	68.0	73.2	57.0	**70.9**	67.3	42.4	53.7	34.0	43.9	43.5
ASNet_w	86.4	86.3	70.9	84.5	82.0	10.8	20.2	13.1	16.1	15.0	**71.6**	72.4	46.4	68.0	64.6	11.4	20.8	12.5	15.9	15.1
ASNet	84.9	**89.6**	**79.0**	86.2	**84.9**	**51.7**	**61.5**	**43.3**	**52.8**	**52.3**	68.5	**76.2**	**58.6**	70.0	**68.3**	**48.5**	**58.3**	**36.3**	**48.3**	**47.8**

Table 1. Performance comparison of ASNet and others on FS-CS and Pascal-5^i [63]. All methods are trained and evaluated under the iFSL framework given strong labels, *i.e.*, class segmentation masks, except for ASNet_w that is trained only with weak labels, *i.e.*, class tags.

	1-way 1-shot		2-way 1-shot	
method	ER	mIoU	ER	mIoU
PANet [78]	66.7	25.2	48.5	23.6
PFENet [73]	71.4	31.9	36.5	22.6
HSNet [44]	77.0	34.3	62.5	29.5
ASNet	**78.6**	**35.8**	**63.1**	**31.6**

method	ER	mIoU
(a) global $\rightarrow$ local	83.9	44.6
(b) w/o masked attention	83.8	50.8
(c) w/o multi-layer fusion	83.1	51.6
ASNet	84.9	52.3

Table 2. Performance comparison of ASNet and others on FS-CS and COCO-20^i [47].

Integrative Few-Shot Learning for Classification and Segmentation

	method	1-way 1-shot 5^0	5^1	5^2	5^3	mIoU	FBIoU	1-way 5-shot 5^0	5^1	5^2	5^3	mIoU	FBIoU	# learn. params.
R50	CANet [98]	52.5	65.9	51.3	51.9	55.4	66.2	55.5	67.8	51.9	53.2	57.1	69.6	-
	PPNet [40]	47.8	58.8	53.8	45.6	51.5	69.2	58.4	67.8	64.9	56.7	62.0	75.8	23.5 M
	PFENet [73]	61.7	69.5	55.4	56.3	60.8	73.3	63.1	70.7	55.8	57.9	61.9	73.9	31.5 M
	SAGNN [87]	64.7	69.6	57.0	57.2	62.1	73.2	64.9	70.0	57.0	59.3	62.8	73.3	-
	MMNet [86]	62.7	70.2	57.3	57.0	61.8	-	62.2	71.5	57.5	62.4	63.4	-	10.4 M
	CMN [88]	64.3	70.0	57.4	59.4	62.8	72.3	65.8	70.4	57.6	60.8	63.7	72.8	-
	MLC [90]	59.2	71.2	**65.6**	52.5	62.1	-	63.5	71.6	**71.2**	58.1	66.1	-	8.7 M
	HSNet [44]	64.3	70.7	60.3	60.5	64.0	76.7	70.3	73.2	67.4	**67.1**	69.5	**80.6**	2.6 M
	ASNet	**68.9**	**71.7**	61.1	**62.7**	**66.1**	**77.7**	**72.6**	**74.3**	65.3	**67.1**	**70.8**	80.4	**1.3 M**

	method	1-way 1-shot mIoU	FBIoU	1-way 5-shot mIoU	FBIoU	# learn. params.
R50	RPMM [89]	30.6	-	35.5	-	38.6 M
	RePRI [3]	34.0	-	42.1	-	-
	MMNet [86]	37.5	-	38.2	-	10.4 M
	MLC [90]	33.9	-	40.6	-	8.7 M
	CMN [88]	39.3	61.7	43.1	63.3	-
	HSNet [44]	39.2	68.2	46.9	70.7	2.6 M
	ASNet	**42.2**	**68.8**	**47.9**	**71.6**	**1.3 M**

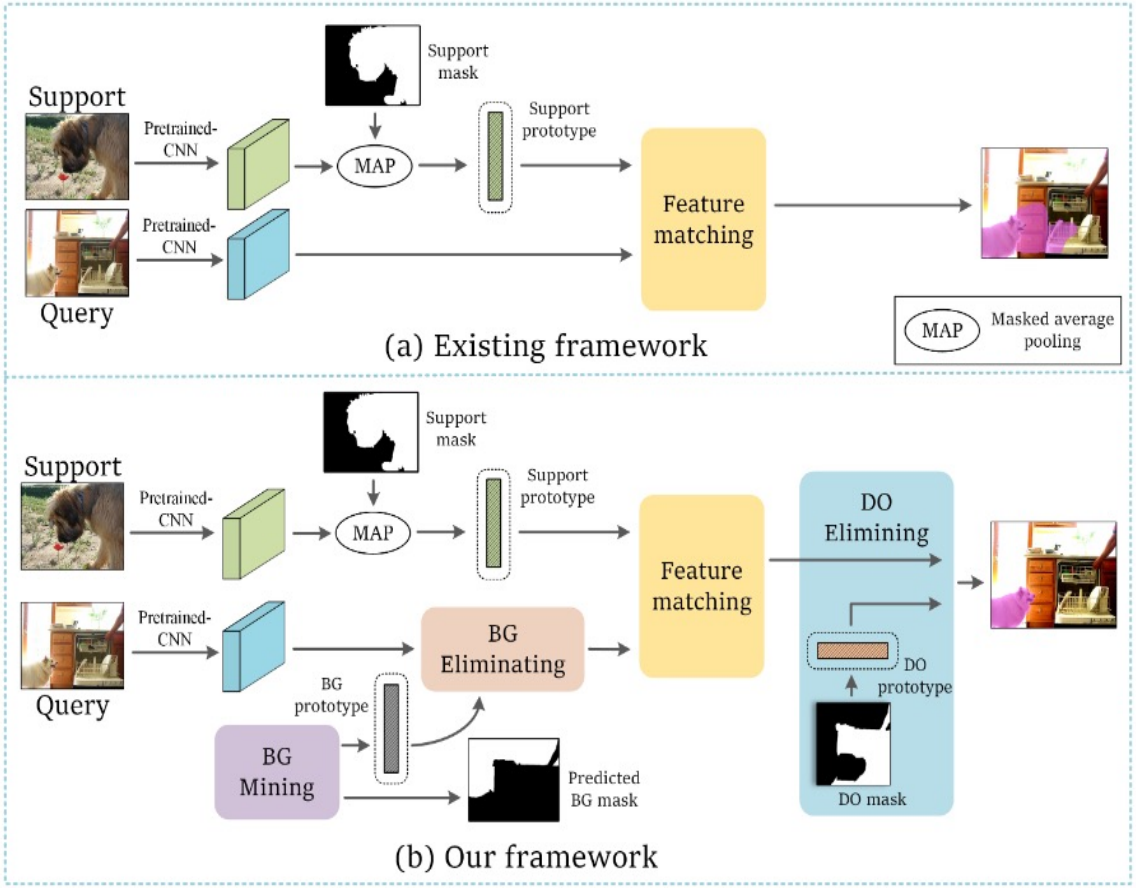
Learning Non-target Knowledge for Few-shot Semantic Segmentation

Yuanwei Liu, Nian Liu, Qinglong Cao, Xiwen Yao, Junwei Han, Ling Shao

Learning Non-target Knowledge for Few-shot Semantic Segmentation

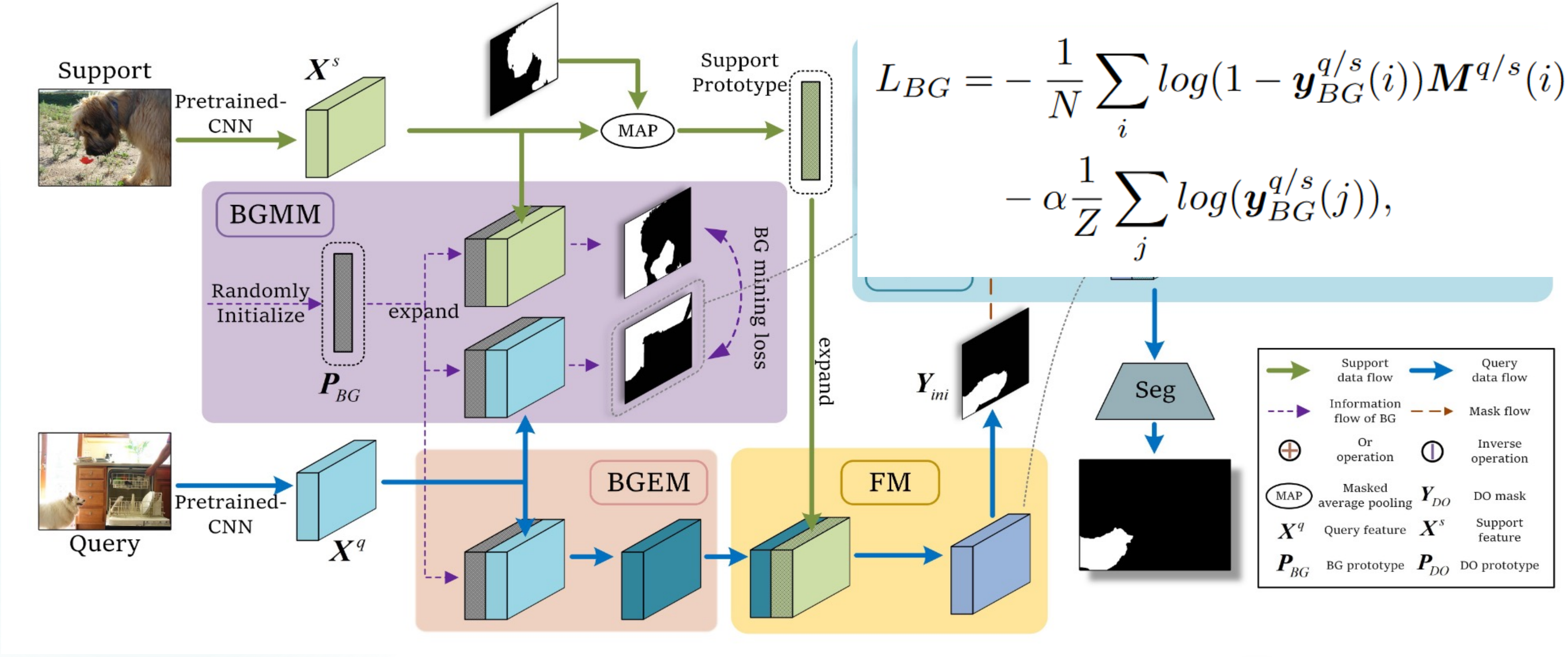
Existing FSSS methods are often hard to tell **non-target** regions

Mine and eliminate **Background (BG)** and **Distracting Objects (DOs)** in the query



Learning Non-target Knowledge for Few-shot Semantic Segmentation

BGMM



$$L_{BG} = -\frac{1}{N}\sum_i log(1 - \boldsymbol{y}_{BG}^{q/s}(i))\boldsymbol{M}^{q/s}(i) - \alpha\frac{1}{Z}\sum_j log(\boldsymbol{y}_{BG}^{q/s}(j)),$$

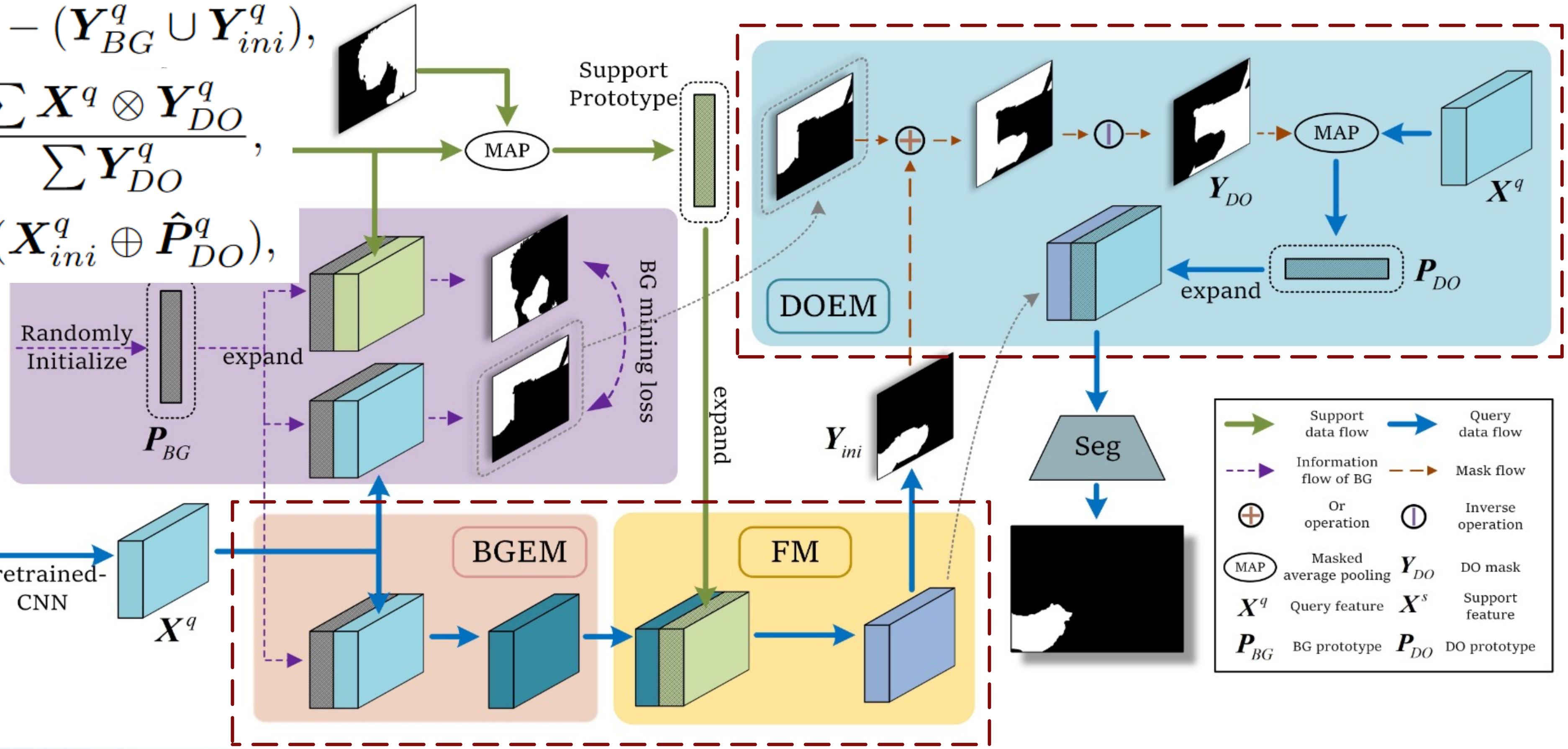
Picanet: Learning pixel-wise contextual attention for saliency detection. (CVPR 2018)

Learning Non-target Knowledge for Few-shot Semantic Segmentation

$$\boldsymbol{Y}_{DO}^{q} = 1 - (\boldsymbol{Y}_{BG}^{q} \cup \boldsymbol{Y}_{ini}^{q}),$$

$$\boldsymbol{P}_{DO}^{q} = \frac{\sum \boldsymbol{X}^{q} \otimes \boldsymbol{Y}_{DO}^{q}}{\sum \boldsymbol{Y}_{DO}^{q}},$$

$$\boldsymbol{y}^{q} = \mathrm{Seg}(\boldsymbol{X}_{ini}^{q} \oplus \hat{\boldsymbol{P}}_{DO}^{q}),$$



$$L_{PCL} = -log\frac{e^{cos(\boldsymbol{P}^{q},\boldsymbol{P}^{s})}}{\sum_{\mathcal{B}}\{e^{cos(\boldsymbol{P}^{q},\boldsymbol{P}_{DO}^{q})} + e^{cos(\boldsymbol{P}^{q},\boldsymbol{P}_{DO}^{s})}\}},$$

Learning Non-target Knowledge for Few-shot Semantic Segmentation

Backbone	Methods	1-Shot Fold-0	1-Shot Fold-1	1-Shot Fold-2	1-Shot Fold-3	1-Shot **Mean**	1-Shot FB-IoU	5-Shot Fold-0	5-Shot Fold-1	5-Shot Fold-2	5-Shot Fold-3	5-Shot **Mean**	5-Shot FB-IoU
	OSLSM [26]	33.6	55.3	40.9	33.5	40.8	61.3	35.9	58.1	42.7	39.1	44.0	61.5
ResNet-101													
	ReRPI [1]	59.6	68.6	**62.2**	47.2	59.4	-	66.2	71.4	67.0	57.7	65.6	-
	MLC [38]	60.8	71.3	61.5	56.9	62.6	-	65.8	**74.9**	**71.4**	63.1	**68.8**	-
	NTRENet	**65.5**	**71.8**	59.1	**58.3**	**63.7**	**75.3**	**67.9**	73.2	60.1	**66.8**	67.0	**78.2**

Backbone	Methods	1-Shot Fold-0	1-Shot Fold-1	1-Shot Fold-2	1-Shot Fold-3	1-Shot **Mean**	1-Shot FB-IoU	5-Shot Fold-0	5-Shot Fold-1	5-Shot Fold-2	5-Shot Fold-3	5-Shot **Mean**	5-Shot FB-IoU
	PPNet [20]	28.1	30.8	29.5	27.7	29.0	-	39.0	40.8	37.1	37.3	38.5	-
	RPMM [37]	29.5	36.8	28.9	27.0	30.6	-	33.8	42.0	33.0	33.3	35.5	-
Re							60.4	-	-	-	-	**42.5**	67.0
							-	37.0	40.3	39.3	36.0	38.2	-
							-	**54.1**	41.2	34.1	33.1	40.6	-
							68.5	38.2	**44.1**	**40.4**	**38.4**	40.3	**69.2**
							62.3	-	-	-	-	29.6	63.9
							-	38.9	40.5	41.5	38.7	39.9	-
ResNet-101	PFENet [31]	34.3	33.0	32.3	30.1	32.4	58.6	38.5	38.6	38.2	34.3	37.4	61.9
	MLC [38]	**50.2**	37.8	27.1	30.4	36.4	-	**57.0**	46.2	37.3	37.2	**44.4**	-
	SAGNN [36]	36.1	**41.0**	38.2	33.5	37.2	60.9	40.9	**48.3**	42.6	38.9	42.7	63.4
	NTRENet	38.3	40.4	**39.5**	**38.1**	**39.1**	**67.5**	42.3	44.4	**44.2**	**41.7**	43.2	**69.6**

BGEM	DOEM	PCL	Fold-0	Fold-1	Fold-2	Fold-3	**Mean**	Precision
			60.8	68.2	55.4	55.3	60.0	61.9
✓			63.2	71.1	57.7	57.4	62.4	62.8
✓	✓		64.7	71.9	58.8	59.0	63.6	63.3
✓	✓	✓	**65.4**	**72.3**	**59.4**	**59.8**	**64.2**	**63.6**