Learning What Not to Segment: A New Perspective on Few-Shot Segmentation

—— Chunbo Lang, Gong Cheng, Binfei Tu, CVPR 2022 oral

方致远

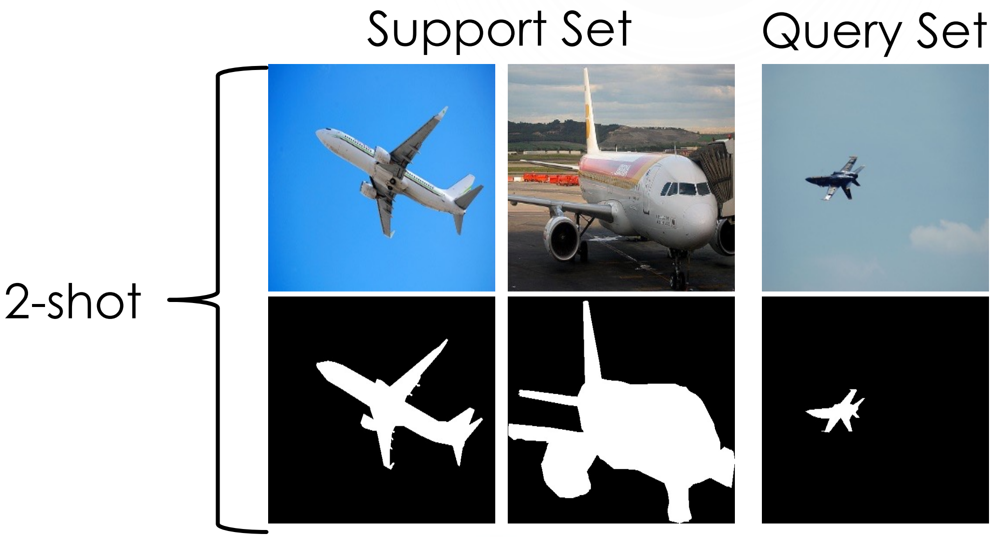
2022.05.08

Few-shot Segmentation

- **Semantic Segmentation**
  - Segment the targets of semantic categories (seen)
  - Required a large amount of labeled data
  - Can not handle the unseen categories
- **Few-shot Segmentation:**
  - Segment the targets of a specific semantic category (unseen)
  - leveraging few labeled data

Few-shot Segmentation

- Train set D_{train} with categories C_{base}, test set D_{test} with categories C_{novel}
  - $C_{base} \cap C_{novel} = \emptyset$
- Input construction: episode $=\{S, Q\}^N$
  - Support set $S = \{(x_i^s, m_i^s)\}_{i=1}^K$
  - Query set $Q = \{(x_i^q, m_i^q)\}$
  - The categories of S and Q are the **same**
- $Prediction = f(Q \mid S)$



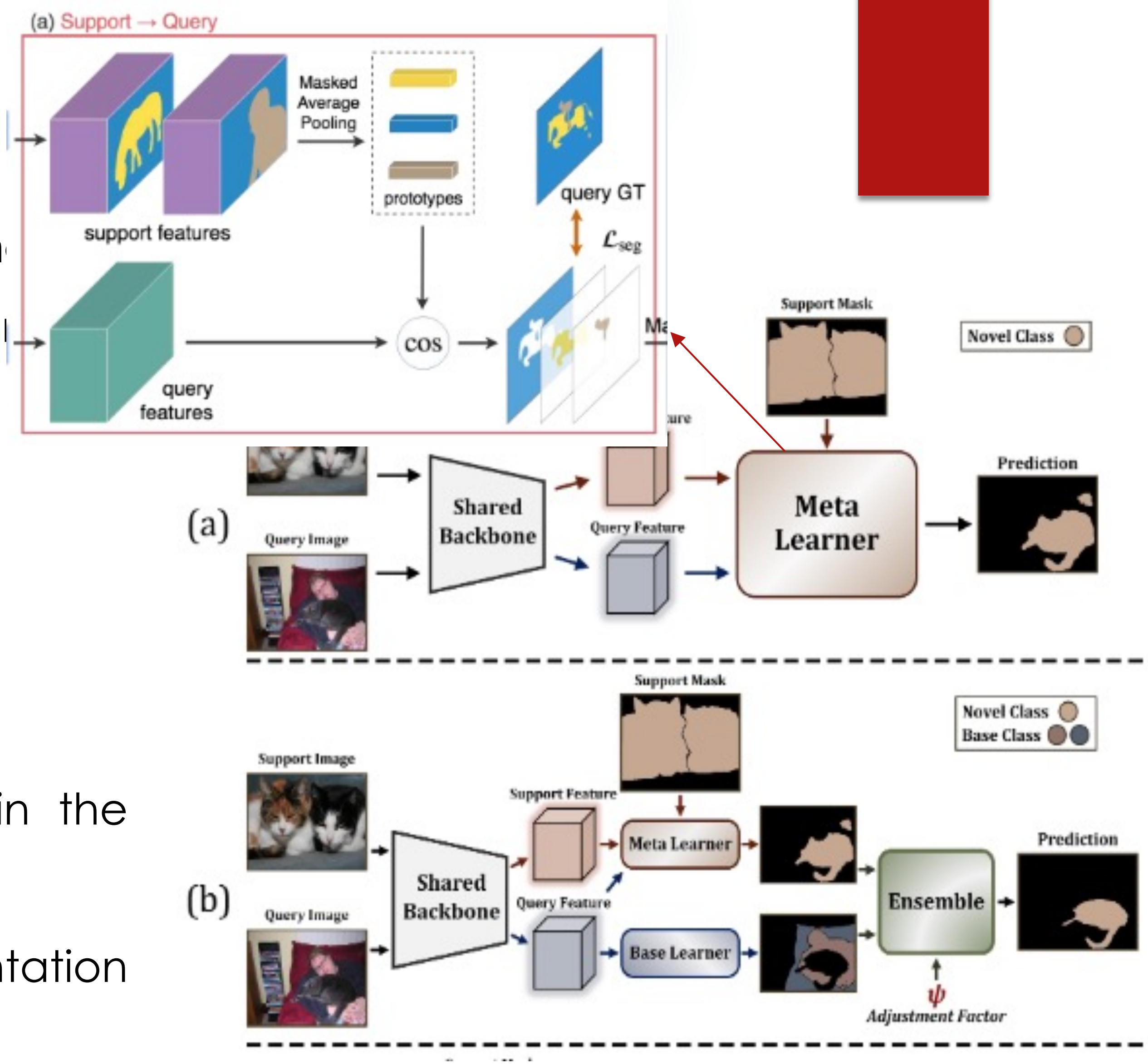
Motivation

(a) Conventional approaches to train the FSS m

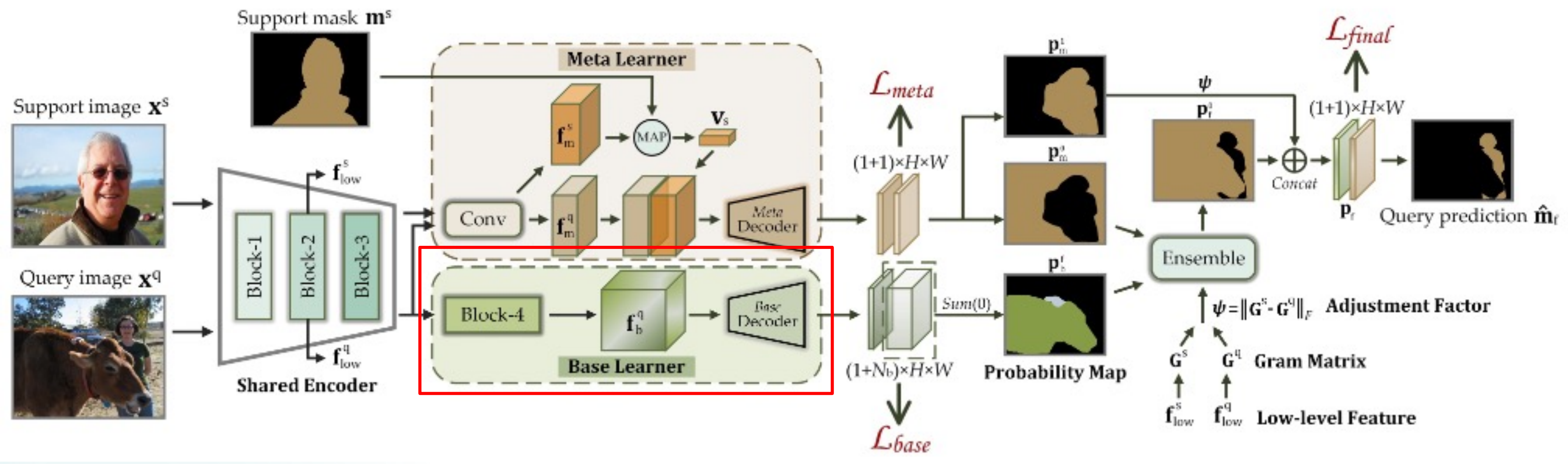
- introduce a bias towards the seen classes than being ideally class-agnostic
- sensitive to the quality of support images

(b) Proposed BAM

- Base learner identify confusable regions in the query image
- base learner provide highly reliable segmentation results



Method - Base learner

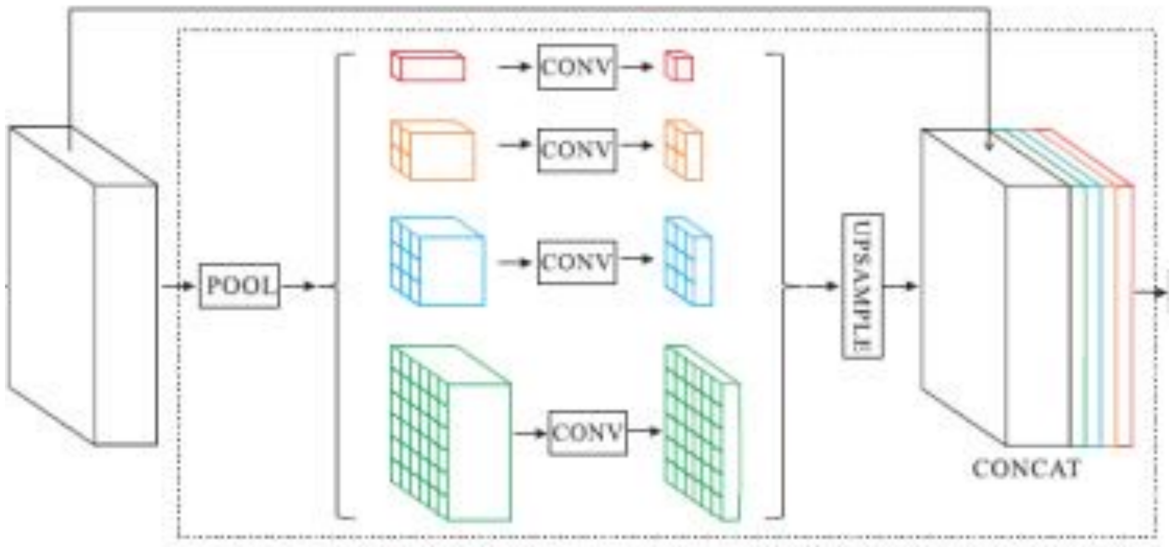


- **Base learner:** PSPNet trained on D_{base}

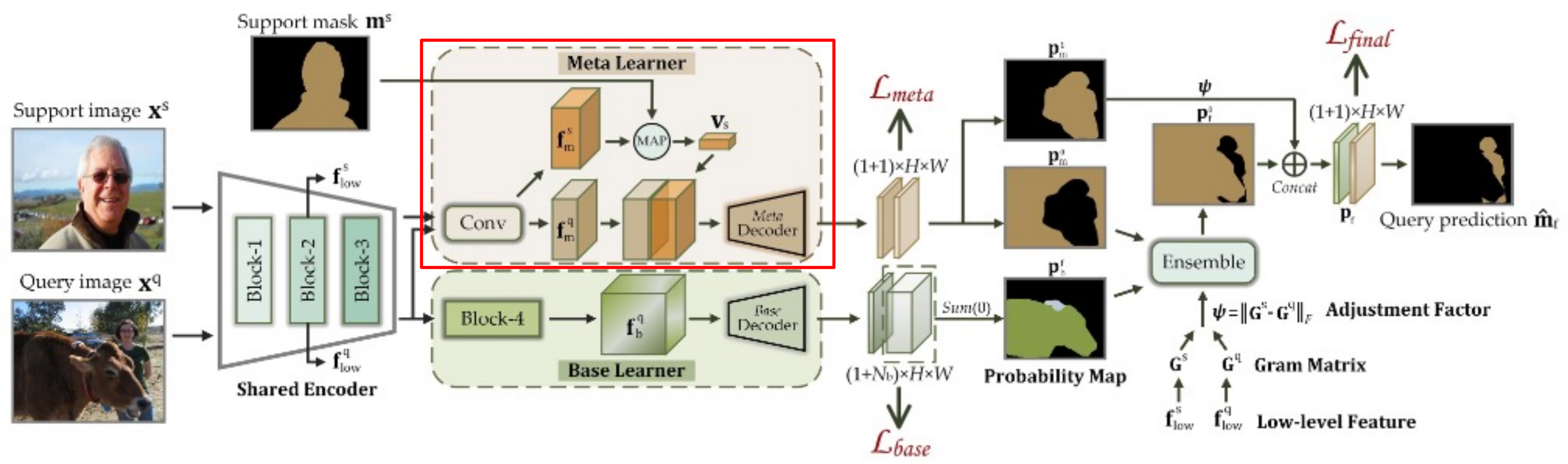
$$\mathbf{f}_b^q = \mathcal{F}_{\text{conv}}\left(\mathcal{E}\left(\mathbf{x}^q\right)\right) \in \mathbb{R}^{c\times h\times w},$$

$$\mathbf{p}_b = \operatorname{softmax}\left(\mathcal{D}_b\left(\mathbf{f}_b^q\right)\right) \in \mathbb{R}^{(1+N_b)\times H\times W}$$

$$\mathcal{L}_{\text{base}} = \frac{1}{n_{\text{bs}}}\sum_{i=1}^{n_{\text{bs}}} \text{CE}\left(\mathbf{p}_{b;i}, \mathbf{m}_{b;i}^q\right),$$

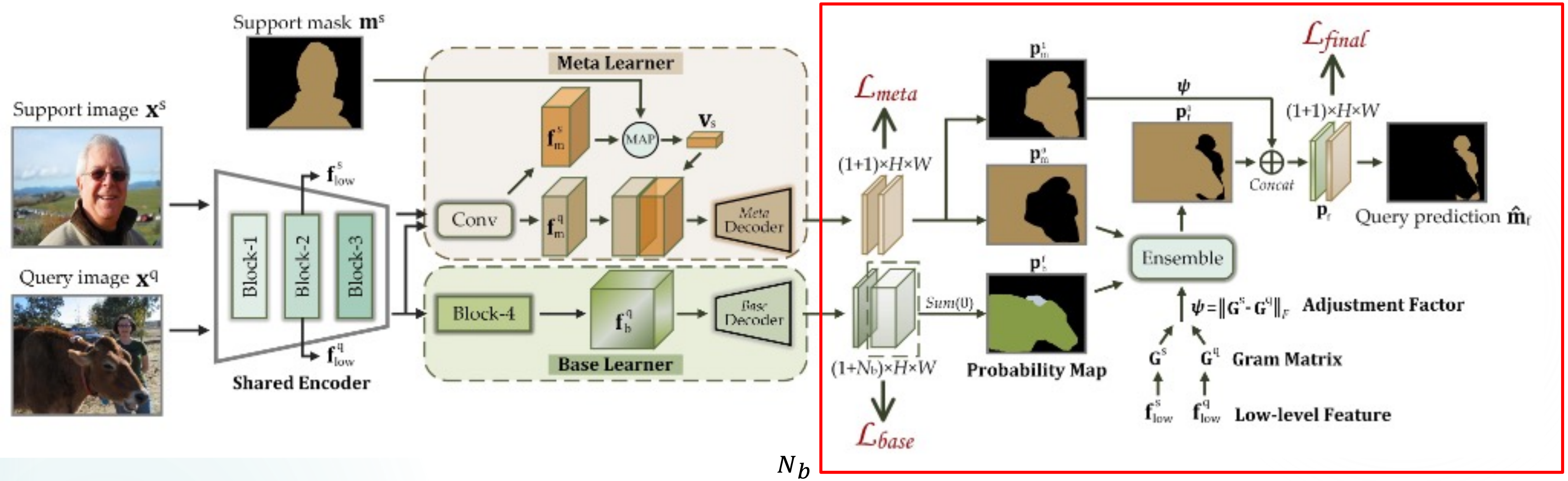


Method - Meta learner



- **Meta learner:** segment the object in query image under the guidance of support images
- $v_s = \mathcal{F}_{pool}\left(f_m^s \odot \mathcal{I}(m^s)\right) \in \mathbb{R}^c$
- $p_m = softmax\left(\mathcal{D}_m\left(\mathcal{F}_{guidance}(v_s, f_m^q)\right)\right) \in \mathbb{R}^{2\times H\times W}$
- $\mathcal{D}_m$: ASPP

Method - Ensemble



- Integrate the prediction of base learner $p_b^f = \sum_{i=1}^{N_b} p_b^i$
- Obtain **Gram matrices** $G^{s/q}$

$$A_s = \mathcal{F}_{reshape}(f_{low}^s) \in \mathbb{R}^{C_1\times N}$$

$$G^s = A_s A_s^T \in \mathbb{R}^{C_1\times C_1}$$

$$\psi = ||G^s - G^q||_F$$

$$p_f^0 = \mathcal{F}_{ensemble}\left(\mathcal{F}_\psi(p_m^0), p_b^f\right)$$

$$p_f = p_f^0 \oplus \mathcal{F}_\psi(p_m^1)$$

Method - K-shot

- **Conventional method:** average the support feature vectors
  - Suboptimal: low quality support images contains less guidance

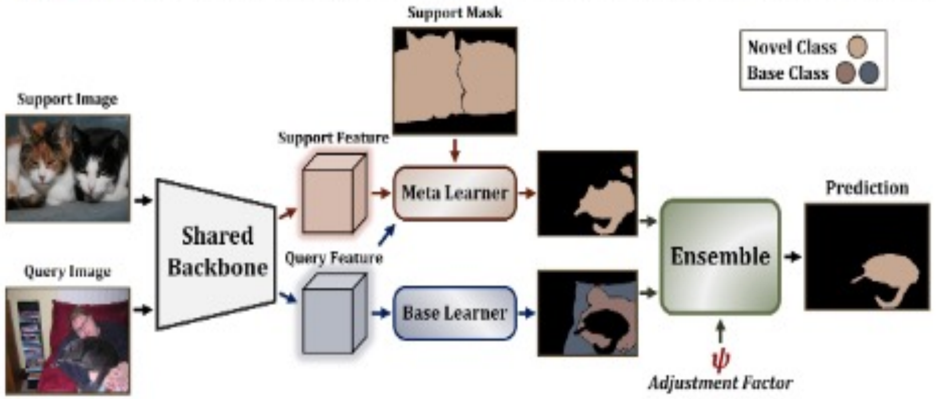
- **Weighted fusion:** a smaller value of ψ_i indicates a greater contribution
  - $\psi_i = \left||G_i^s - G^q|\right|_F$
  - Sort ψ_k, $\quad \psi_k = \text{concate}(\psi_i)$
  - $\eta = softmax\left(w_2^T ReLU(w_1^T \psi_t)\right) \in \mathbb{R}^K$
  - η reverts to the original order
  - Fusing support feature vectors and ψ with η

实验

Backbone	Method	1-shot Fold-0	1-shot Fold-1	1-shot Fold-2	1-shot Fold-3	1-shot Mean	5-shot Fold-0	5-shot Fold-1	5-shot Fold-2	5-shot Fold-3	5-shot Mean
VGG16	SG-One (TCYB'19) [67]	40.20	58.40	48.40	38.40	46.30	41.90	58.60	48.60	39.40	47.10
	PANet (ICCV'19) [56]	42.30	58.00	51.10	41.20	48.10	51.80	64.60	59.80	46.50	55.70
	FWB (ICCV'19) [56]	47.00	59.60	52.60	48.30	51.90	50.90	62.90	56.50	50.10	55.10
	CRNet (CVPR'20) [33]	-	-	-	-	55.20	-	-	-	-	58.50
	PFENet (TPAMI'20) [51]	56.90	68.20	54.40	52.40	58.00	59.00	69.10	54.80	52.90	59.00
	HSNet (ICCV'21) [37]	59.60	65.70	59.60	54.00	59.70	64.90	69.00	64.10	58.60	64.10
	Baseline	59.90	67.51	64.93	55.72	62.02	64.02	71.51	69.39	63.55	67.12
	BAM (ours)	**63.18**	**70.77**	**66.14**	**57.53**	**64.41**	**67.36**	**73.05**	**70.61**	**64.00**	**68.76**
ResNet50	CANet (ICCV'19) [66]	52.50	65.90	51.30	51.90	55.40	55.50	67.80	51.90	53.20	57.10
	PGNet (ICCV'19) [65]	56.00	66.90	50.60	50.40	56.00	57.70	68.70	52.90	54.60	58.50
	CRNet (CVPR'20) [33]	-	-	-	-	55.70	-	-	-	-	58.80
	PPNet (ECCV'20) [34]	48.58	60.58	55.71	46.47	52.84	58.85	68.28	66.77	57.98	62.97
	PFENet (TPAMI'20) [51]	61.70	69.50	55.40	56.30	60.80	63.10	70.70	55.80	57.90	61.90
	HSNet (ICCV'21) [37]	64.30	70.70	60.30	60.50	64.00	70.30	73.20	67.40	67.10	69.50
	Baseline	65.68	71.41	65.56	58.93	65.40	67.28	72.38	69.16	66.25	68.77
	BAM (ours)	**68.97**	**73.59**	**67.55**	**61.13**	**67.81**	**70.59**	**75.05**	**70.79**	**67.20**	**70.91**

Backbone	Method	1-shot Fold-0	1-shot Fold-1	1-shot Fold-2	1-shot Fold-3	1-shot Mean	5-shot Fold-0	5-shot Fold-1	5-shot Fold-2	5-shot Fold-3	5-shot Mean
VGG16	FWB [38]	18.35	16.72	19.59	25.43	20.02	20.94	19.24	21.94	28.39	22.63
	PFENet [51]	35.40	38.10	36.80	34.70	36.30	38.20	42.50	41.80	38.90	40.40
	PRNet [32]	27.46	32.99	26.70	28.98	29.03	31.18	36.54	31.54	32.00	32.82
	Baseline	38.42	43.75	44.32	39.84	41.58	45.93	48.88	47.87	46.96	47.41
	BAM (ours)	**38.96**	**47.04**	**46.41**	**41.57**	**43.50**	**47.02**	**52.62**	**48.59**	**49.11**	**49.34**
ResNet50	HFA [31]	28.65	36.02	30.16	33.28	32.03	32.69	42.12	30.35	36.19	35.34
	ASGNet [23]	-	-	-	-	34.56	-	-	-	-	42.48
	HSNet [37]	36.30	43.10	38.70	38.70	39.20	43.30	51.30	48.20	45.00	46.90
	Baseline	41.92	45.35	43.86	41.24	43.09	46.98	51.87	49.49	47.81	49.04
	BAM (ours)	**43.41**	**50.59**	**47.49**	**43.42**	**46.23**	**49.26**	**54.20**	**51.63**	**49.55**	**51.16**

(b)



MS COCO

消融实验

PT	$\mathcal{L}_{meta}$	Init.	ψ	mIoU	FB-IoU
				57.61	70.75
✓				59.12	71.94
✓	✓			59.76	72.79
✓	✓	✓		62.49	75.43
✓	✓	✓	✓	**64.41**	**77.26**

$$p_f^0 = \mathcal{F}_{ensemble}\left(\mathcal{F}_{\psi}(p_m^0), p_b^f\right)$$

$$\mathcal{F}(p_m^0 \oplus \psi)$$

```
nn.Parameter(torch.tensor([[1.0],[0.0]]))
```