北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

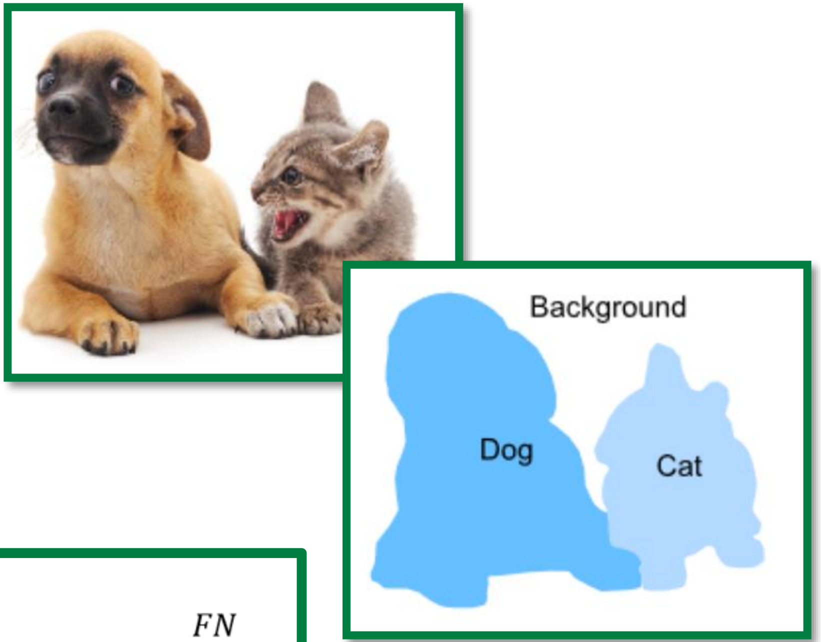
Scale-Aware Graph Neural Network for Few-Shot Semantic Segmentation

Guo-Sen Xie[1], Jie Liu[1], Huan Xiong[1,3*], Ling Shao[2]
[1]Mohamed bin Zayed University of AI, UAE [2]Inception Institute of Artificial Intelligence, UAE
[3]Harbin Institute of Technology, China

德以明理 学以精工

Semantic Segmentation

- Category-label prediction for each pixel
- A pixel-wise classification
- mask

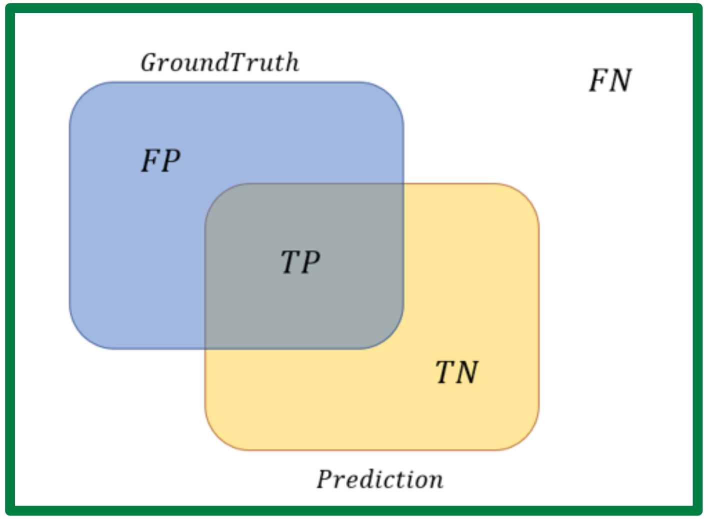


IoU

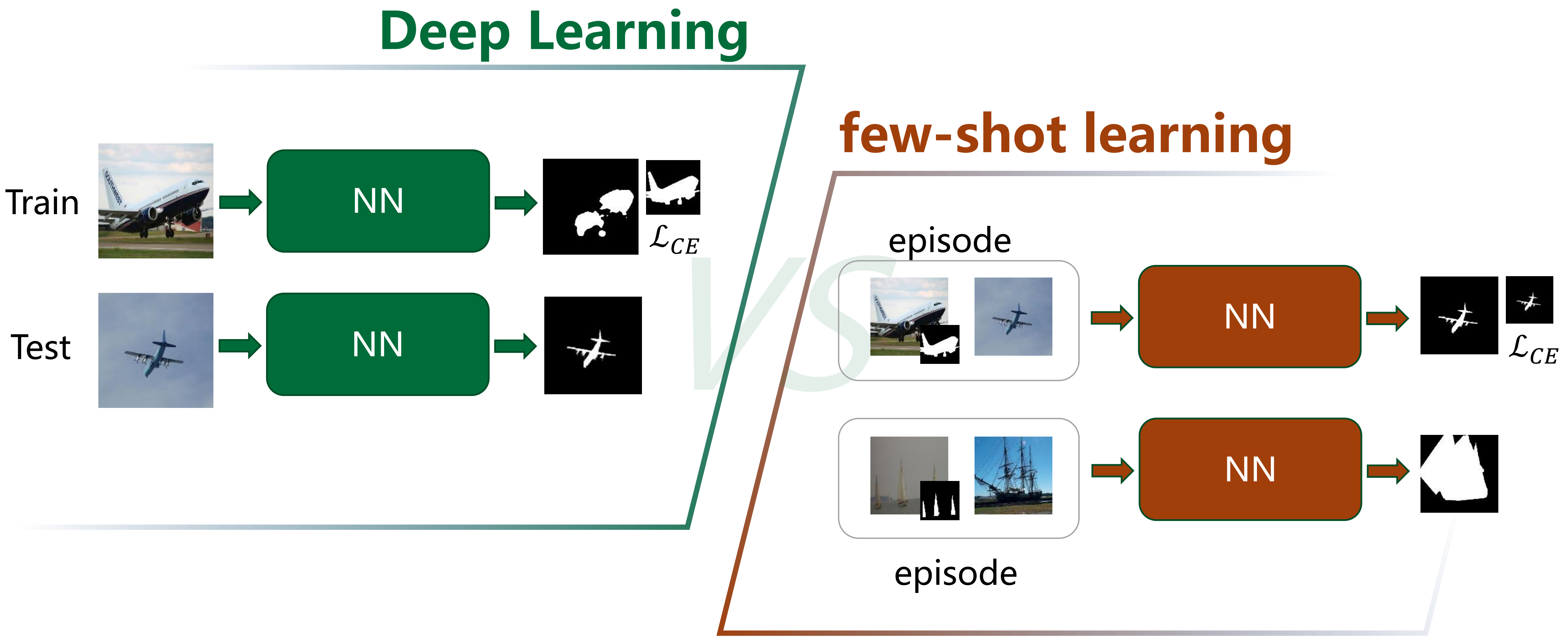
- $\mathrm{IoU} = \dfrac{\mathrm{Prediction} \cap \mathrm{GroundTruth}}{\mathrm{Prediction} \cup \mathrm{GroundTruth}}$

$$= \frac{\mathrm{T}P}{\mathrm{TP} + \mathrm{T}N + FP}$$

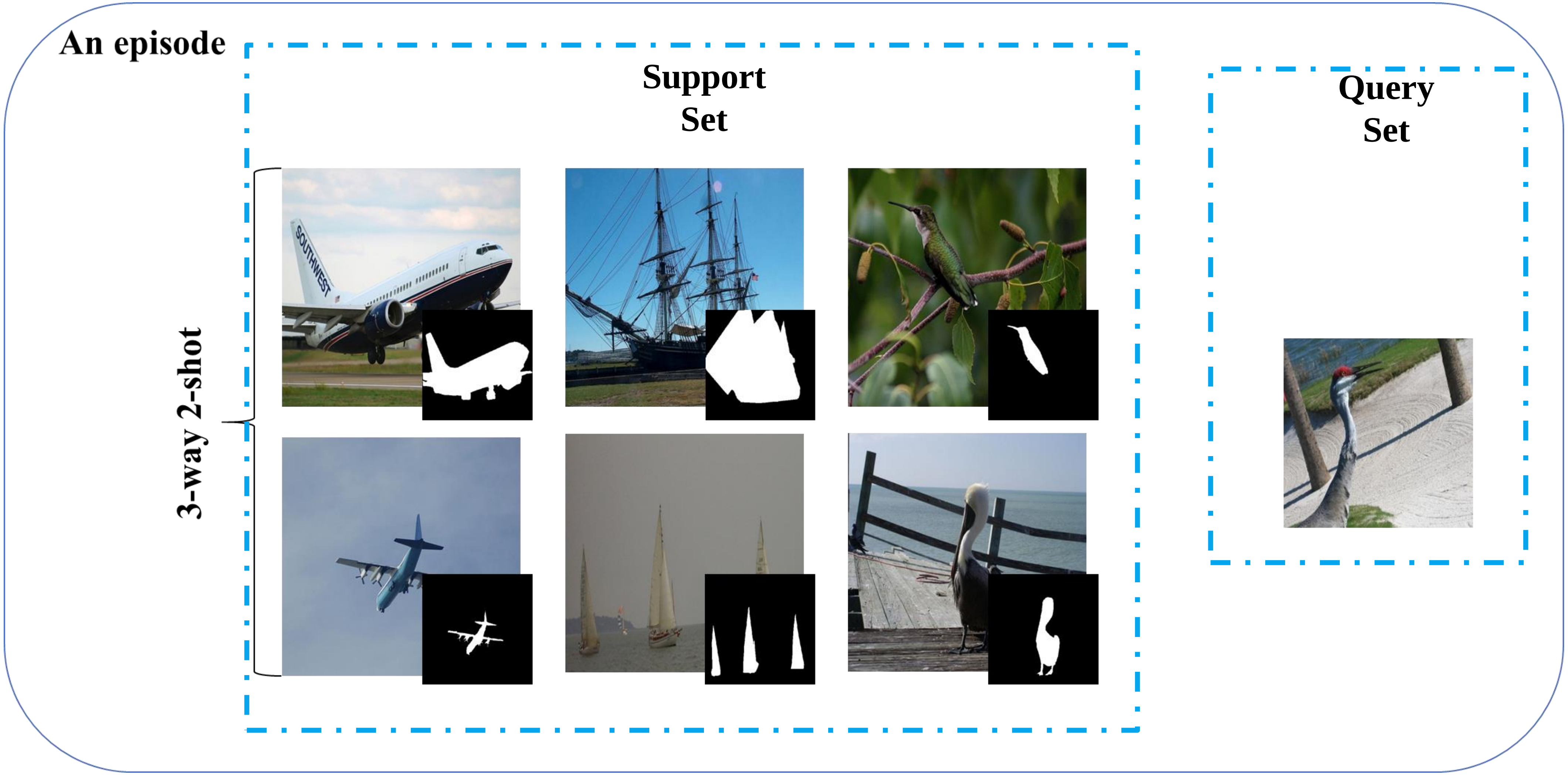
- $\mathrm{mIoU}(\mathrm{mean\ IoU}), \mathrm{FBIoU}(\mathrm{binary\ IoU})$



Few-shot learning



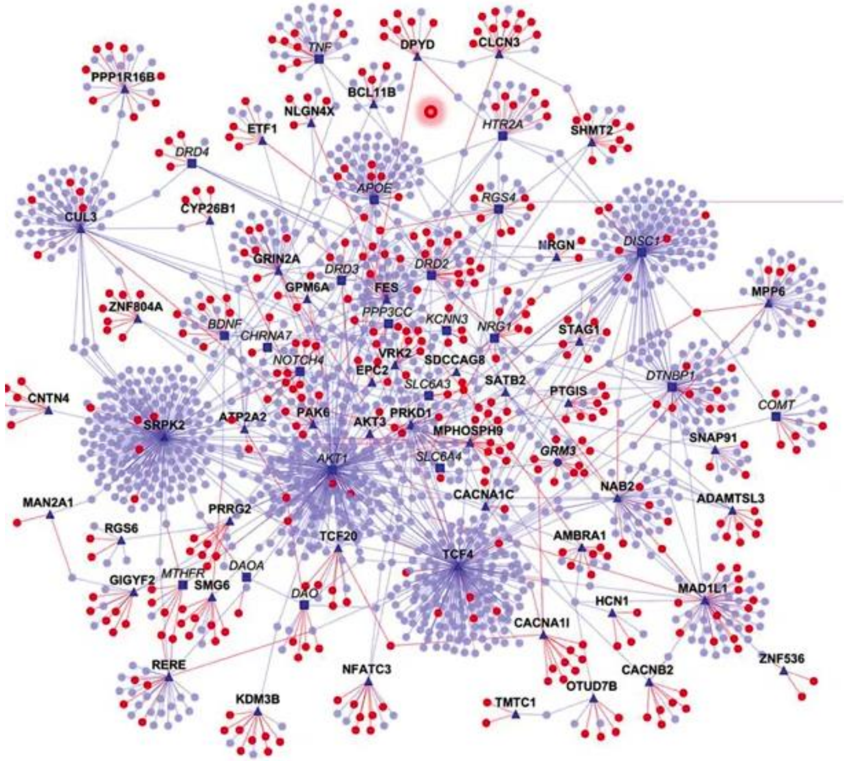
Few-shot Semantic Segmentation

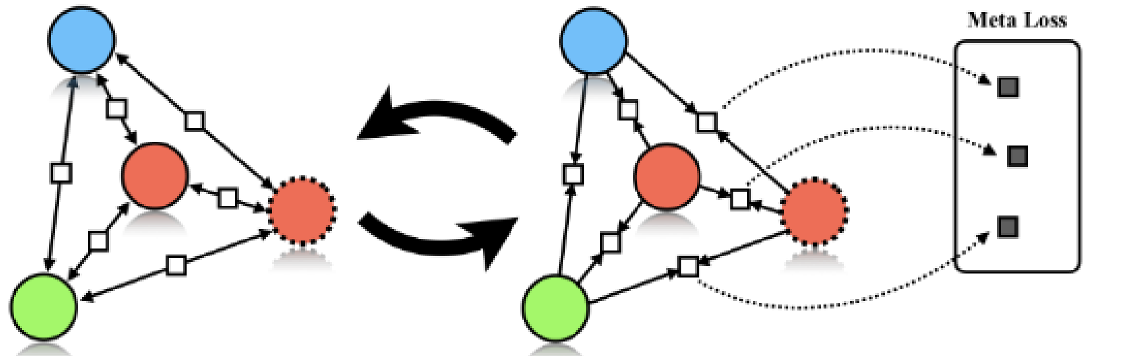


Graph Neural Network(GNN)

$$\mathbf{H} = F(\mathbf{H}, \mathbf{X})$$
$$\mathbf{O} = G(\mathbf{H}, \mathbf{X}_N)$$

Update Node (and edge) with Neural Networks

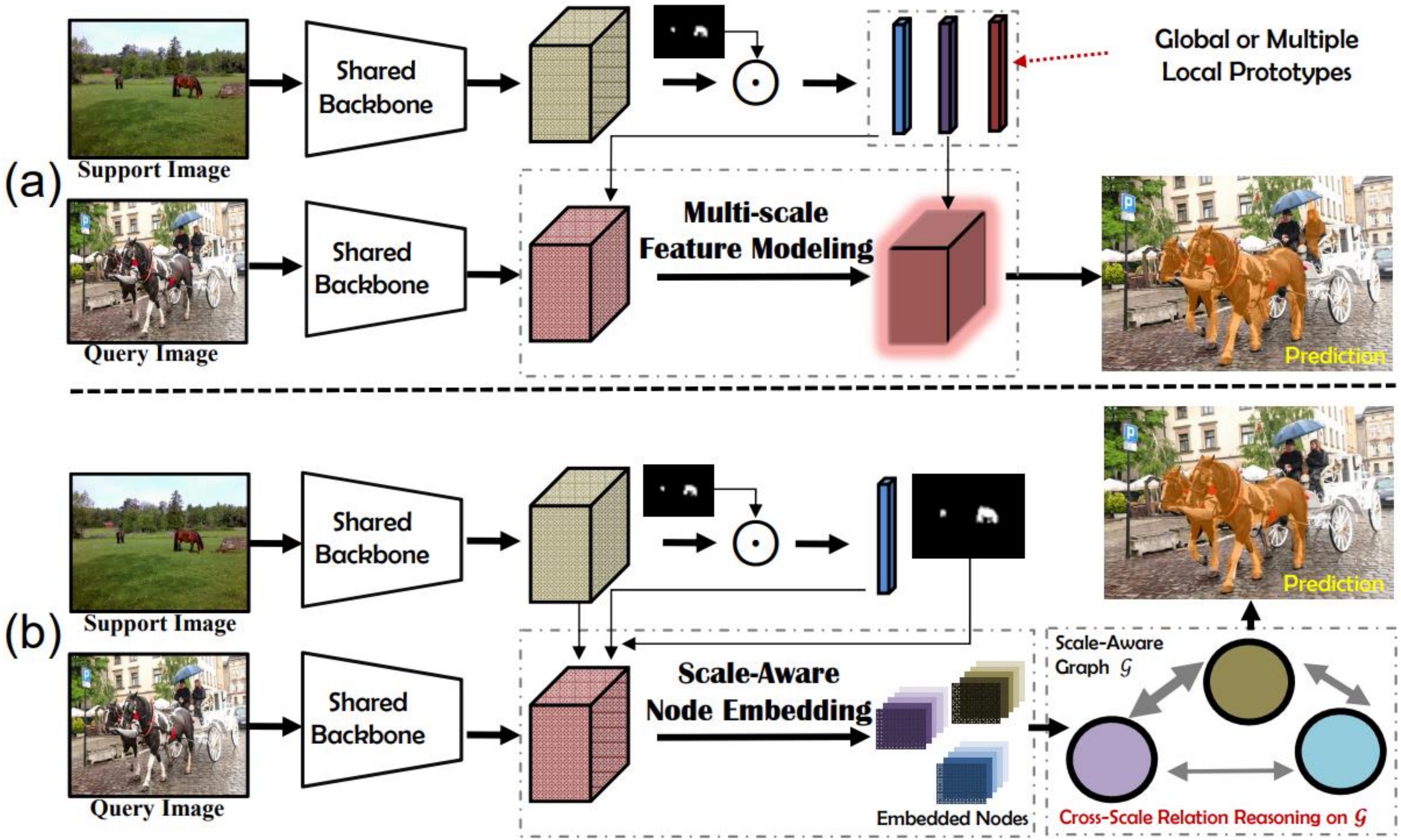




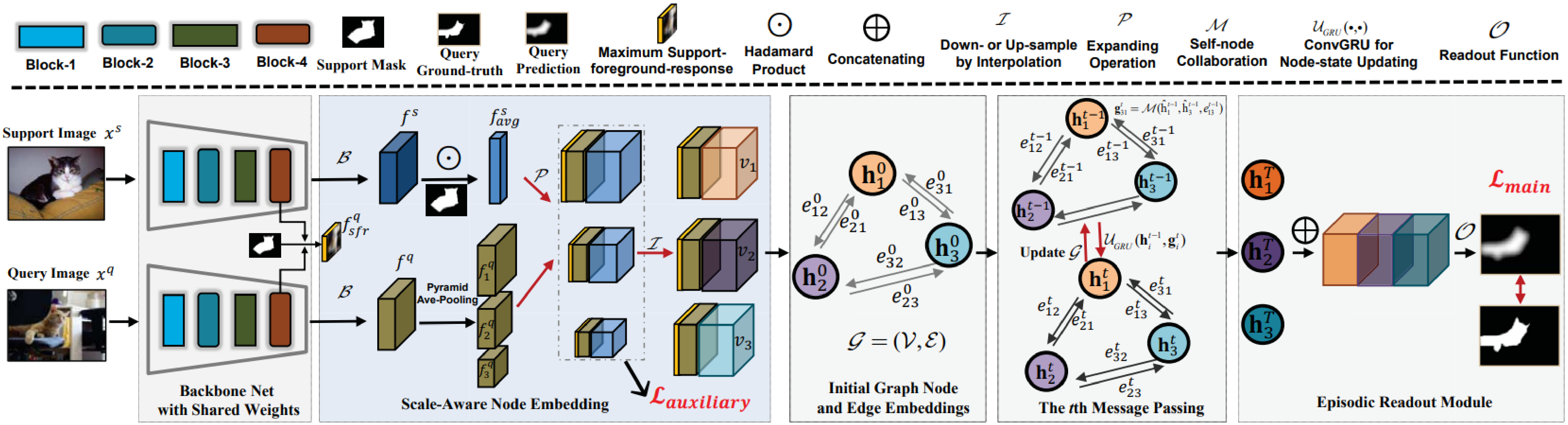
(a) Node Feature Update (b) Edge Feature Update (c) Meta Update

Not suitable for few-shot learning

Motivation


(a)
Support Image
Query Image
Shared Backbone
Shared Backbone
Global or Multiple Local Prototypes
Multi-scale Feature Modeling
Prediction
(b)
Support Image
Query Image
Shared Backbone
Shared Backbone
Scale-Aware Node Embedding
Embedded Nodes
Scale-Aware Graph $\mathcal{G}$
Prediction
Cross-Scale Relation Reasoning on $\mathcal{G}$

Method



Initial:

Node: Multi-scale concatenated features

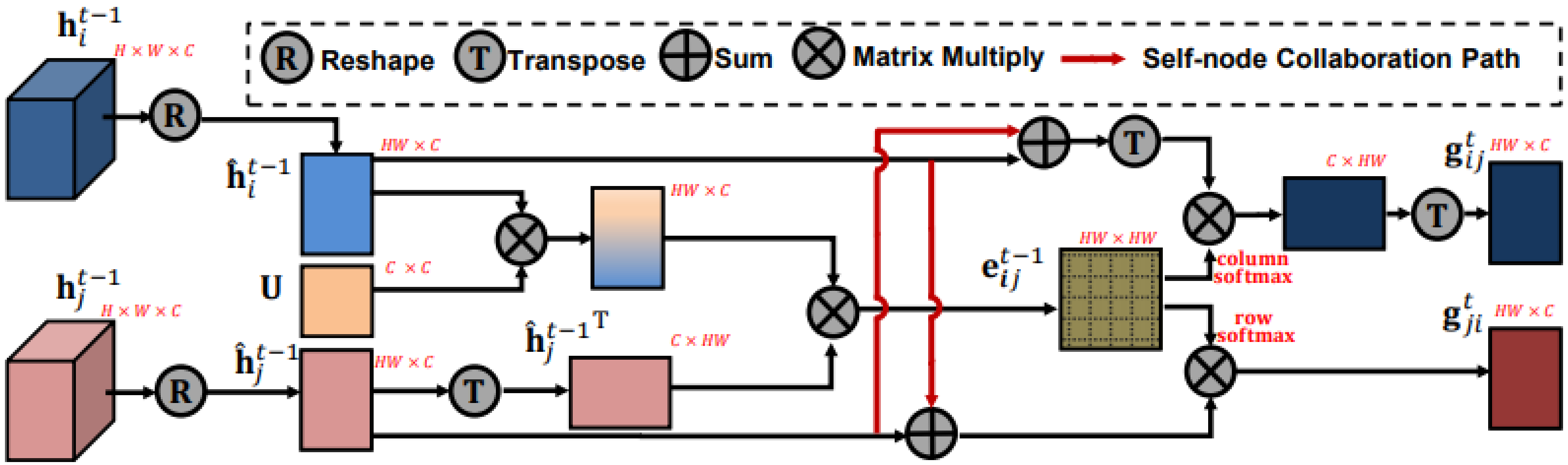
Edge: Calculated with a learnable matrix

$$\mathbf{h}_i^0 = \mathcal{I}_{H\times W}(\mathcal{C}(\mathbf{f}_i^q \oplus \mathcal{P}_{H_i\times W_i}(f_{avg}^s) \oplus \mathcal{I}_{H_i\times W_i}(f_{sfr}^q))),$$

$$\mathbf{e}_{ij}^t = \hat{\mathbf{h}}_i^t \mathbf{U} \hat{\mathbf{h}}_j^{t^\top} \in \mathbb{R}^{HW\times HW},$$

Graph Update

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

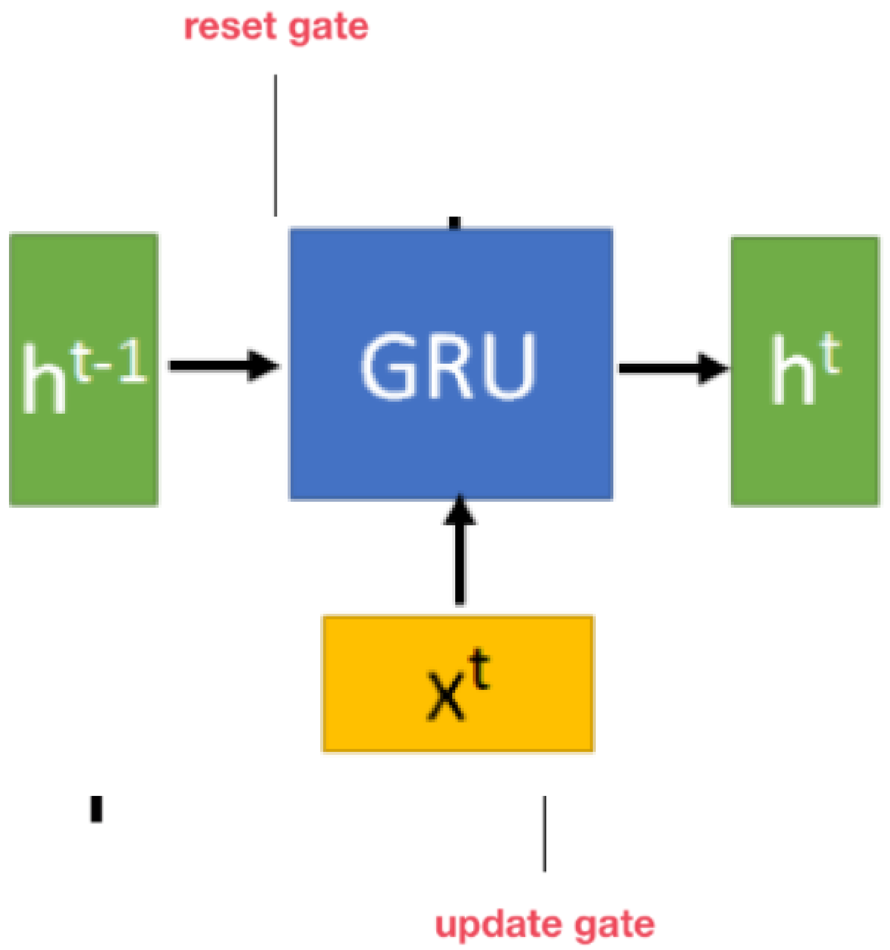


Update:

$$
\begin{aligned}
\mathbf{g}_{ji}^t =& \mathcal{M}(\hat{\mathbf{h}}_i^{t-1}, \hat{\mathbf{h}}_j^{t-1}, \mathbf{e}_{ij}^{t-1}) \\
=& \text{softmax}(\mathbf{e}_{ij}^{t-1})(\hat{\mathbf{h}}_j^{t-1} + \hat{\mathbf{h}}_i^{t-1}) \in \mathbb{R}^{HW \times C}.
\end{aligned} \quad (6)
$$

Note: both Edges and Nodes are updated in each iteration

ConvGRU



$$
\begin{aligned}
\mathbf{z}_l^l &= \sigma(\mathbf{W}_z^l * \mathbf{x}_l^l + \mathbf{U}_z^l * \mathbf{h}_{l-1}^l),\\
\mathbf{r}_l^l &= \sigma(\mathbf{W}_r^l * \mathbf{x}_l^l + \mathbf{U}_r^l * \mathbf{h}_{l-1}^l),\\
\tilde{\mathbf{h}}_l^l &= \tanh(\mathbf{W}^l * \mathbf{x}_l^l + \mathbf{U} * (\mathbf{r}_l^l \odot \mathbf{h}_{l-1}^l),\\
\mathbf{h}_l^l &= (1 - \mathbf{z}_l^l)\mathbf{h}_{l-1}^l + \mathbf{z}_l^l \tilde{\mathbf{h}}_l^l,
\end{aligned}
$$

z : update gate
r : reset gate
h : hidden state

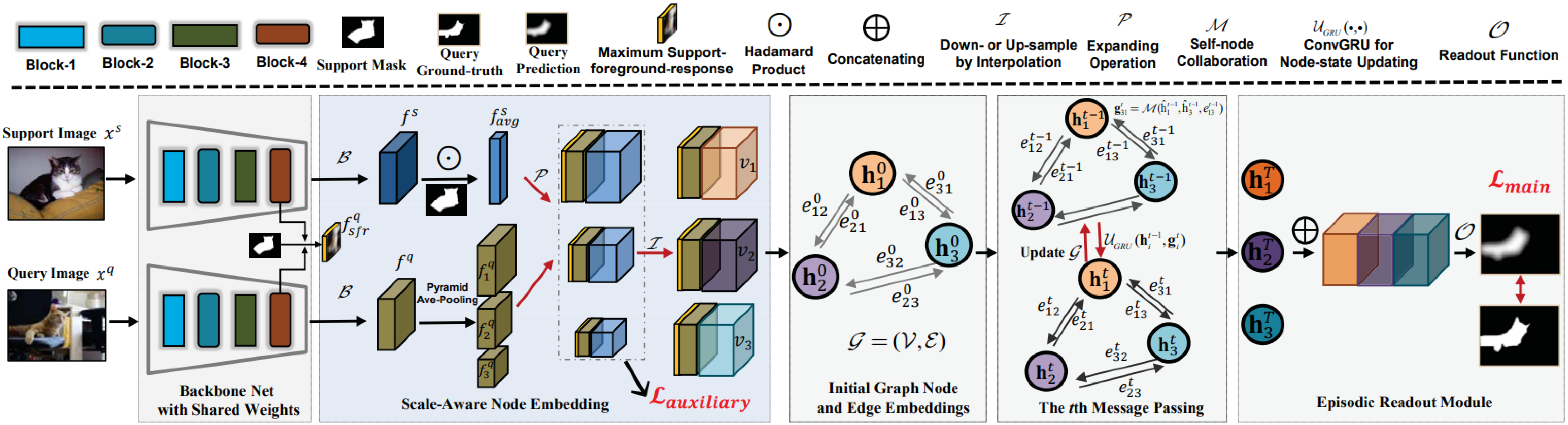
$$\mathbf{g}_i^t = \mathcal{F}_{\text{reshape}}\Big(\sum_{v_j \in \mathcal{V}(i)} \mathbf{g}_{ji}^t\Big) \in \mathbb{R}^{H \times W \times C},$$

$$\mathbf{h}_i^t = \mathcal{U}_{\text{GRU}}(\mathbf{h}_i^{t-1}, \mathbf{g}_i^t) \in \mathbb{R}^{H \times W \times C}.$$

[1]Nicolas Ballas, Li Yao, Chris Pal, and Aaron Courville. Delving deeper into convolutional networks for learning video representations. In arXiv:1511.06432, 2015. 5

Method

北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY



Initial:

Node: Multi-scale concatenated features

Edge: Calculated with a learnable matrix

$$\mathbf{h}_i^0 = \mathcal{I}_{H\times W}(\mathcal{C}(\mathbf{f}_i^q \oplus \mathcal{P}_{H_i\times W_i}(f_{avg}^s) \oplus \mathcal{I}_{H_i\times W_i}(f_{sfr}^q))),$$

$$\mathbf{e}_{ij}^t = \hat{\mathbf{h}}_i^t \mathbf{U} \hat{\mathbf{h}}_j^{t^\top} \in \mathbb{R}^{HW\times HW},$$

Experiment

Methods	Backbone	mean-IoU (1-shot)					FB-IoU	mean-IoU (5-shot)					FB-IoU
		Fold-0	Fold-1	Fold-2	Fold-3	Mean	(1-shot)	Fold-0	Fold-1	Fold-2	Fold-3	Mean	(5-shot)
OSLSM (BMVC'17) [27]	VGG-16	33.6	55.3	40.9	33.5	40.8	61.3	35.9	58.1	42.7	39.1	43.9	61.5
co-FCN (ICLRW'18) [24]	VGG-16	31.7	50.6	44.9	32.4	41.1	60.1	37.5	50.0	44.1	33.9	41.4	60.2
AMP (ICCV'19) [28]	VGG-16	41.9	50.2	46.7	34.7	43.4	62.2	41.8	55.5	50.3	39.9	46.9	63.8
SG-One (TCYB'19) [48]	VGG-16	40.2	58.4	48.4	38.4	46.3	63.1	41.9	58.6	48.6	39.4	47.1	65.9
PANet (ICCV'19) [34]	VGG-16	42.3	58.0	51.1	41.2	48.1	66.5	51.8	64.6	59.8	46.5	55.7	70.7
CANet (CVPR'19) [47]	ResNet-50	52.5	65.9	51.3	51.9	55.4	66.2	55.5	67.8	51.9	53.2	57.1	69.6
PGNet (ICCV'19) [46]	ResNet-50	56.0	66.9	50.6	50.4	56.0	69.9	57.7	68.7	52.9	54.6	58.5	70.5
FWB (ICCV'19) [22]	ResNet-101	51.3	64.5	56.7	52.2	56.2	-	54.8	67.4	62.2	55.3	59.9	-
PMMs (ECCV'20) [41]	ResNet-50	52.0	67.5	51.5	49.8	55.2	-	55.0	68.2	52.9	51.1	56.8	-
PPNet (ECCV'20) [20]	ResNet-50	47.8	58.8	53.8	45.6	51.5	-	58.4	67.8	**64.9**	56.7	62.0	-
DAN (ECCV'20) [33]	ResNet-101	54.7	68.6	**57.8**	51.6	58.2	71.9	57.9	69.0	60.1	54.9	60.5	72.3
PFENet (TPAMI'20) [31]	ResNet-50	61.7	69.5	55.4	56.3	60.8	73.3	63.1	**70.7**	55.8	57.9	61.9	73.9
BriNet (BMVC'20) [42]	ResNet-50	56.5	67.2	51.6	53.0	57.1	-	-	-	-	-	-	-
SimPropNet (IJCAI'20) [10]	ResNet-50	54.9	67.3	54.5	52.0	57.2	73.0	57.2	68.5	58.4	56.1	60.0	72.9
Baseline	ResNet-50	62.1	68.2	55.3	53.8	59.9	71.7	63.3	68.7	55.1	55.3	60.6	71.8
SAGNN	ResNet-50	**64.7**	**69.6**	57.0	**57.2**	**62.1**	73.2	**64.9**	70.0	57.0	59.3	**62.8**	73.3

ASGNet (ours)	59.84	67.43	55.59	54.39	59.31	**64.55**	**71.32**	**64.24**	57.33	**64.36**	**5.05**

*ResNet101

ours-SCL (PFENet)	resnet50	**63.0**	**70.0**	56.5	**57.7**	**61.8**	**64.5**	**70.9**	57.3	**58.7**	**62.9**

Experiment

Backbone	mean-IoU	
	1-shot	5-shot
VGG-16	58.4	59.3
ResNet-50	**62.1**	**62.8**
ResNet-101	60.8	61.5

Table 3. Effects of backbones.

5-shot testing	mean-IoU	FB-IoU
1-shot baseline	62.1	73.2
Feature-Avg	**62.8**	**73.3**
Mask-Avg	62.5	72.4
Mask-OR	61.8	72.0

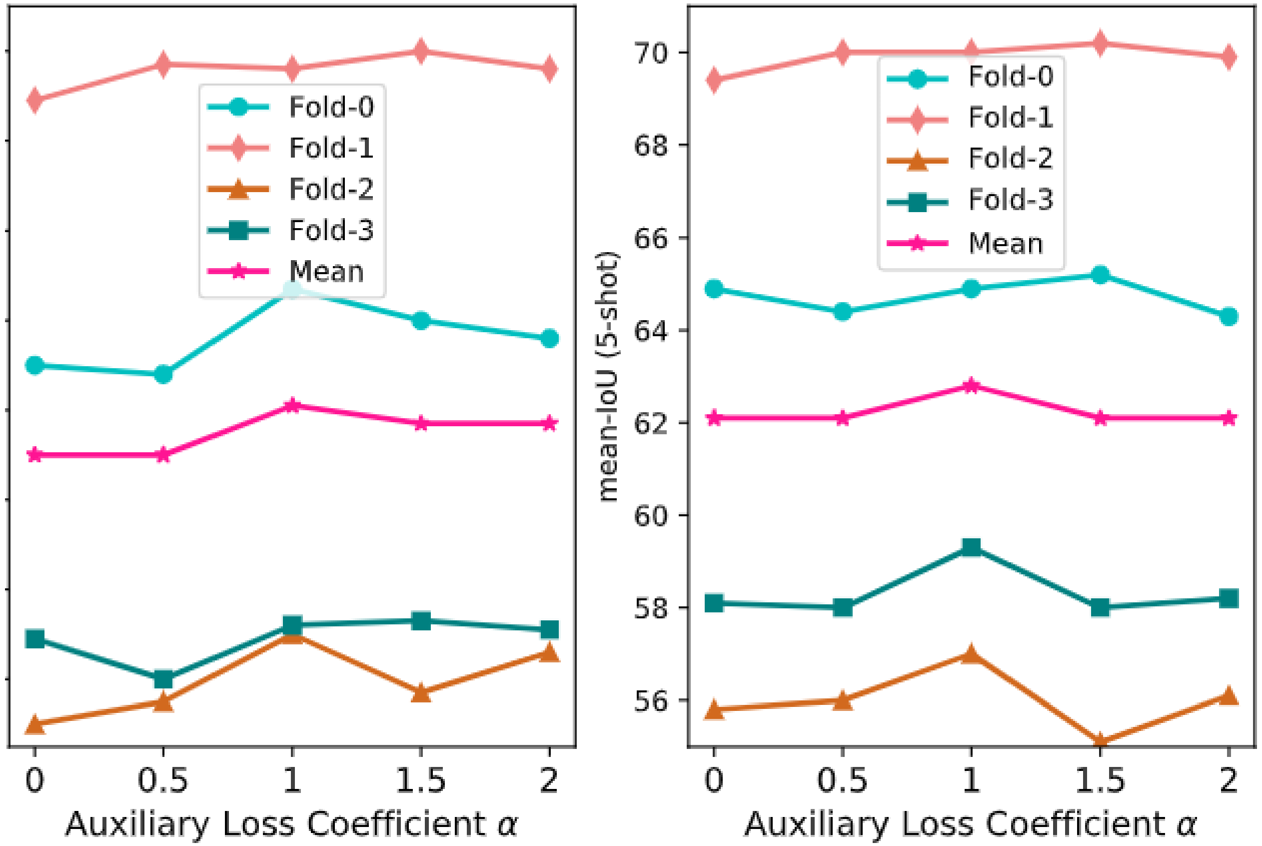
Table 4. Feature fusion under 5-shot setting.

Setting		mean-IoU	
$\lvert\mathcal{V}\rvert$	T	1-shot	5-shot
1	1	n.a.	n.a.
2	1	61.0	61.7
3	1	**62.1**	**62.8**
4	1	61.0	61.9
3	1	62.1	62.8
3	2	**62.4**	**62.9**
3	3	62.1	62.5

Table 5. Effects of $\lvert\mathcal{V}\rvert$ and T.

Models	mean-IoU	
	1-shot	5-shot
SAGNN *w* SC	61.2	62.7
SAGNN *w* OI	61.1	62.4
Full SAGNN	**62.1**	**62.8**

Table 6. Effects of self-node collaboration.



5. mean-IoUs under different auxiliary loss coefficients α.