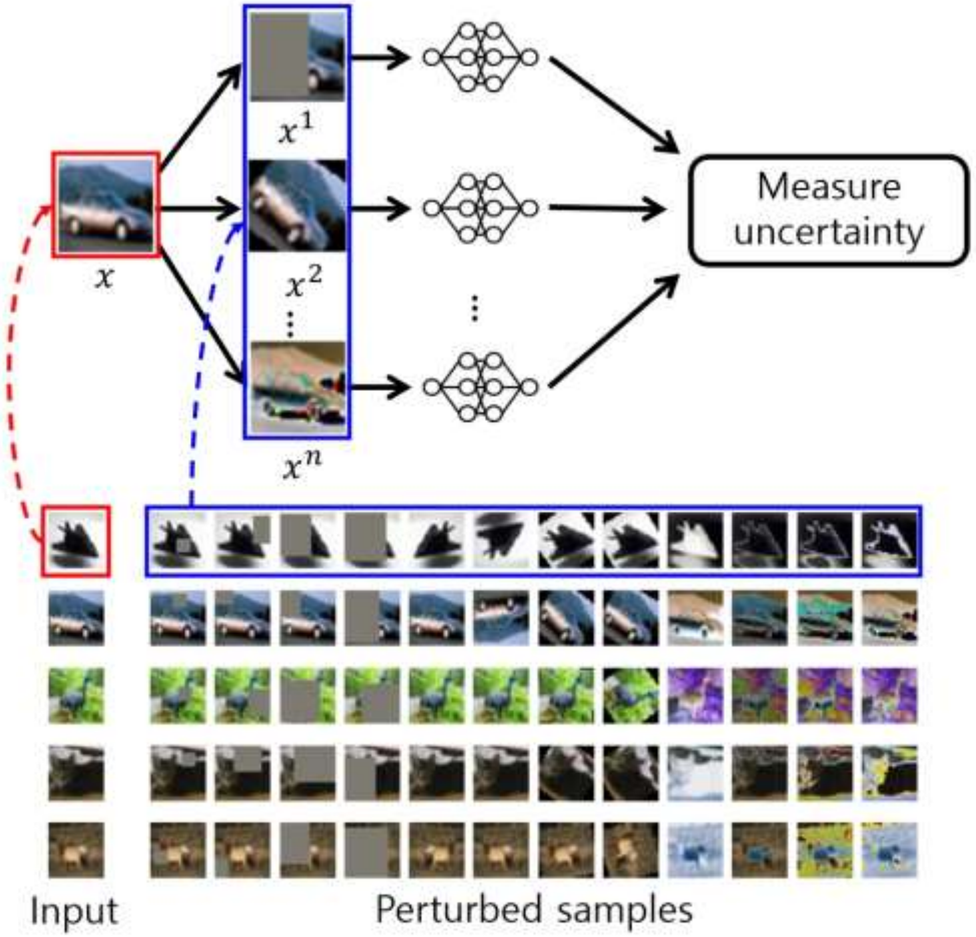
Rainbow Memory: Continual Learning with a Memory of Diverse Samples

- Representative for corresponding class
- Discriminative to other classes



$$p(y=c|x) = \int_{\tilde{\mathscr{D}}} p(y=c|\tilde{x}_t) p(\tilde{x}_t|x) d\tilde{x}_t$$

$$\approx \frac{1}{A} \sum_{t=1}^{A} p(y=c|\tilde{x}_t),$$

$$S_c = \sum_{t=1}^{T} \mathbb{1}_c \operatorname*{argmax}_{\hat{c}} p(y=\hat{c}|\tilde{x}_t),$$

$$u(x) = 1 - \frac{1}{T} \max_c S_c,$$

Algorithm 1 Diversity-Aware Memory Update

Input: K denotes memory size, N_t denotes the number of seen classes until task t, $\mathfrak{D}_t^S$ denotes stream data at task t, $\mathfrak{D}_{t-1}^M$ denotes exemplars stored in a episodic memory after task $t-1$.

Output: $\mathfrak{D}_t^M$ exemplars after learning task t.

$\mathfrak{D}_t^M = \{\}$ ▷ New exemplars from scratch

$k_c = floor(K/N_t)$ ▷ Class-balanced sampling

for $c = 1, 2, \ldots, N_t$ **do**

    $\mathfrak{D}_c = \{(x,y) | y = c, (x,y) \in \mathfrak{D}_t^S \cup \mathfrak{D}_{t-1}^M\}$

    Sort $\mathfrak{D}_c$ by $u(x)$ computed by (4)

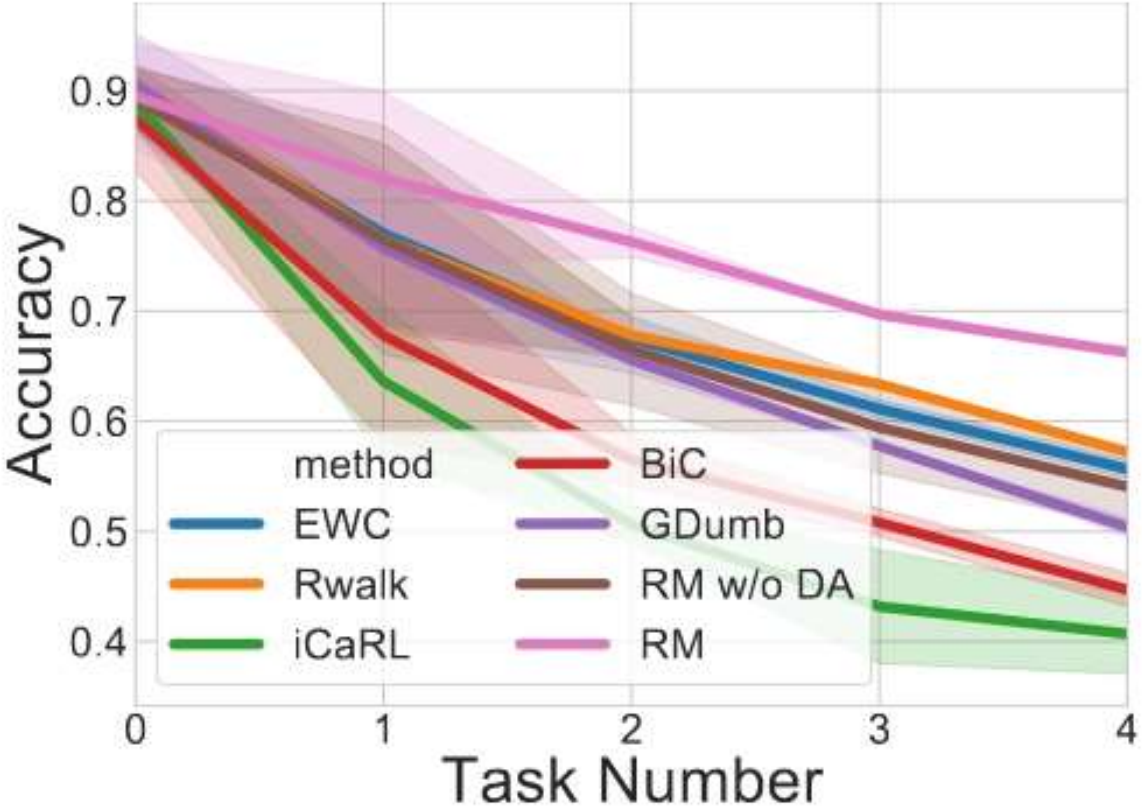
    for $j = 1, 2, \ldots, k_c$ **do**

        $i = j * |\mathfrak{D}_c| / k_c$ ▷ $|\mathfrak{D}_c|/k_c$ step-size indexing

        $\mathfrak{D}_t^M \mathrel{+}= \mathfrak{D}_c[i]$

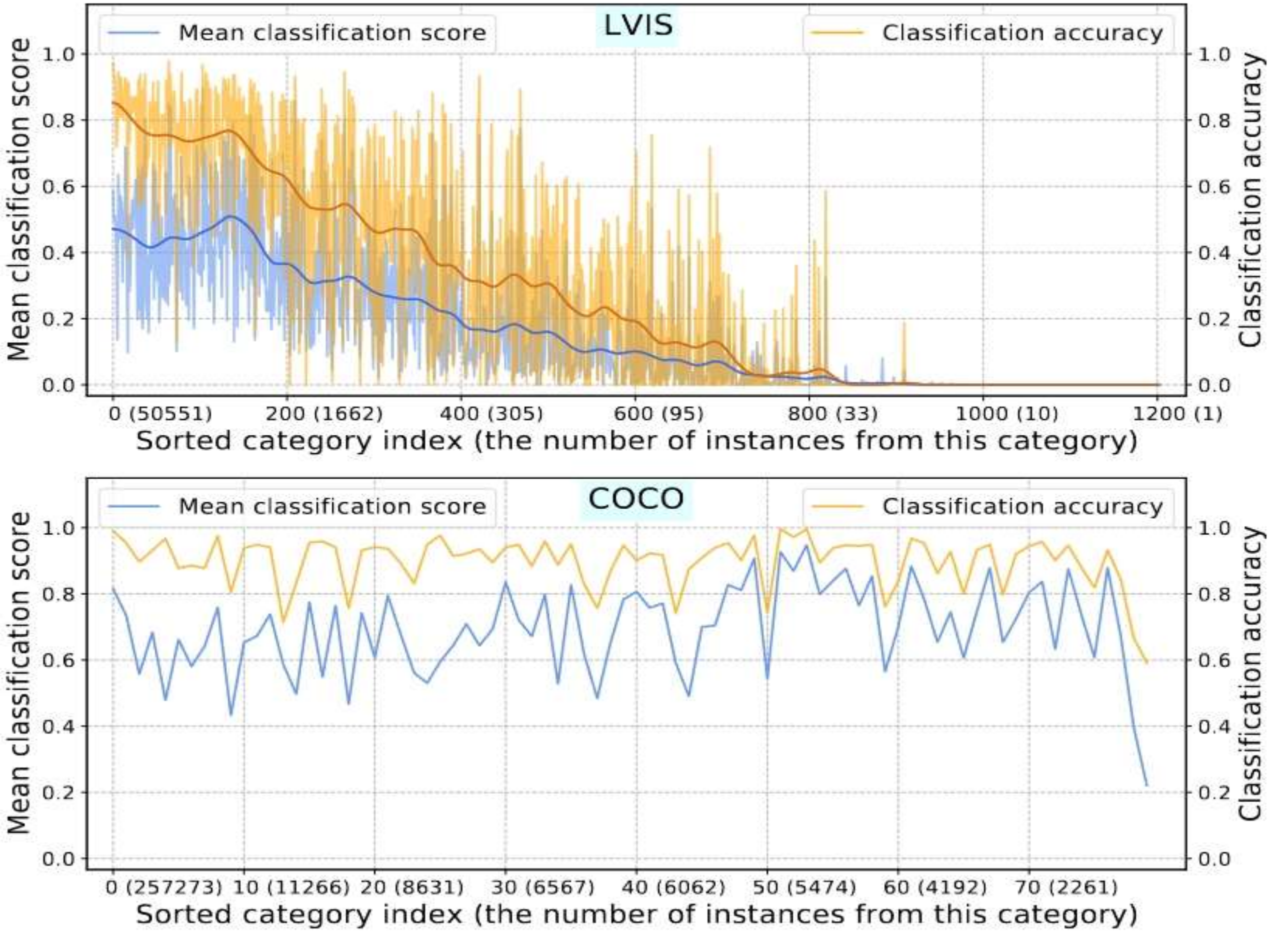
    end for

end for



Exploring Classification Equilibrium in Long-Tailed Object Detection

- Data re-sampling is prone to over-fit tail data and under-represent the head data
- Loss reweighting may cause unstable training



$$s_y^i = \alpha s_y^{i-1} + (1-\alpha) p_y^i,$$

Equilibrium Loss

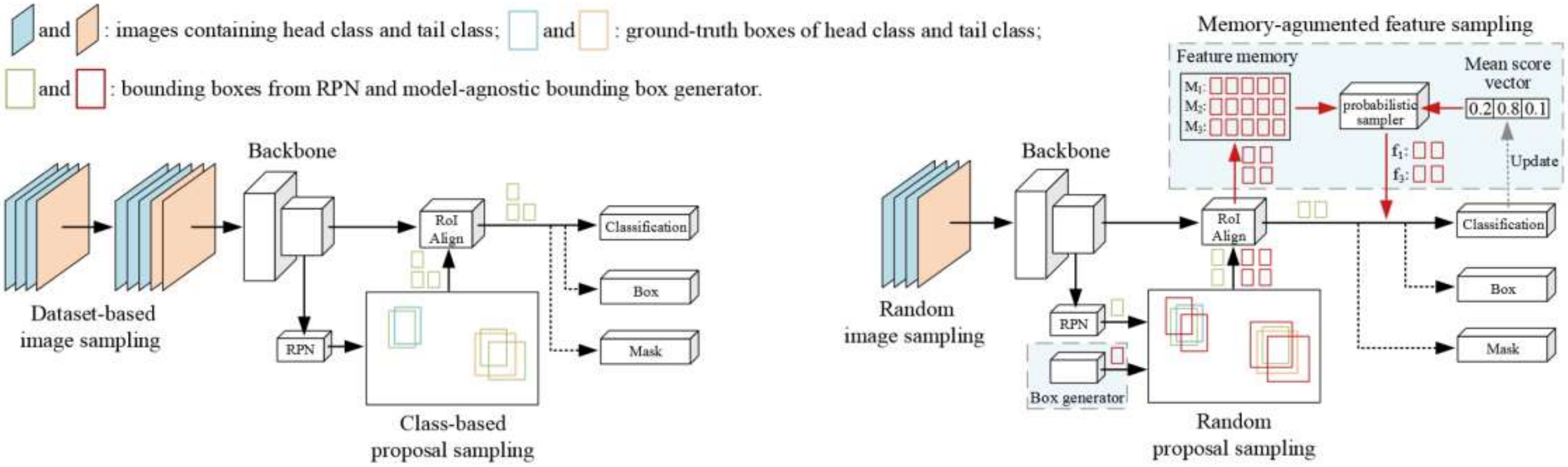
$$L(z, y) = -log\frac{e^{z_y}}{\sum_{y' \in \{1,2,\ldots,C+1\}} e^{z_{y'}}} = log[1 + \sum_{y' \neq y} e^{z_{y'} - z_y}].$$

$$L(z, y) = log[1 + \sum_{y' \neq y} e^{z_{y'} - z_y + \delta_{yy'}}],$$

$$\delta_{yy'} = log(\frac{s_{y'}}{s_y}).$$

- reduce the suppression of dominant classes (i.e., having high mean classification score) over weak classes, by reducing the margin between dominant positive classes and weak negative classes
- enlarge the suppression of weak classes over dominant classes, by increasing the margin between weak positive classes and dominant negative classes

Memory-augmented Feature Sampling



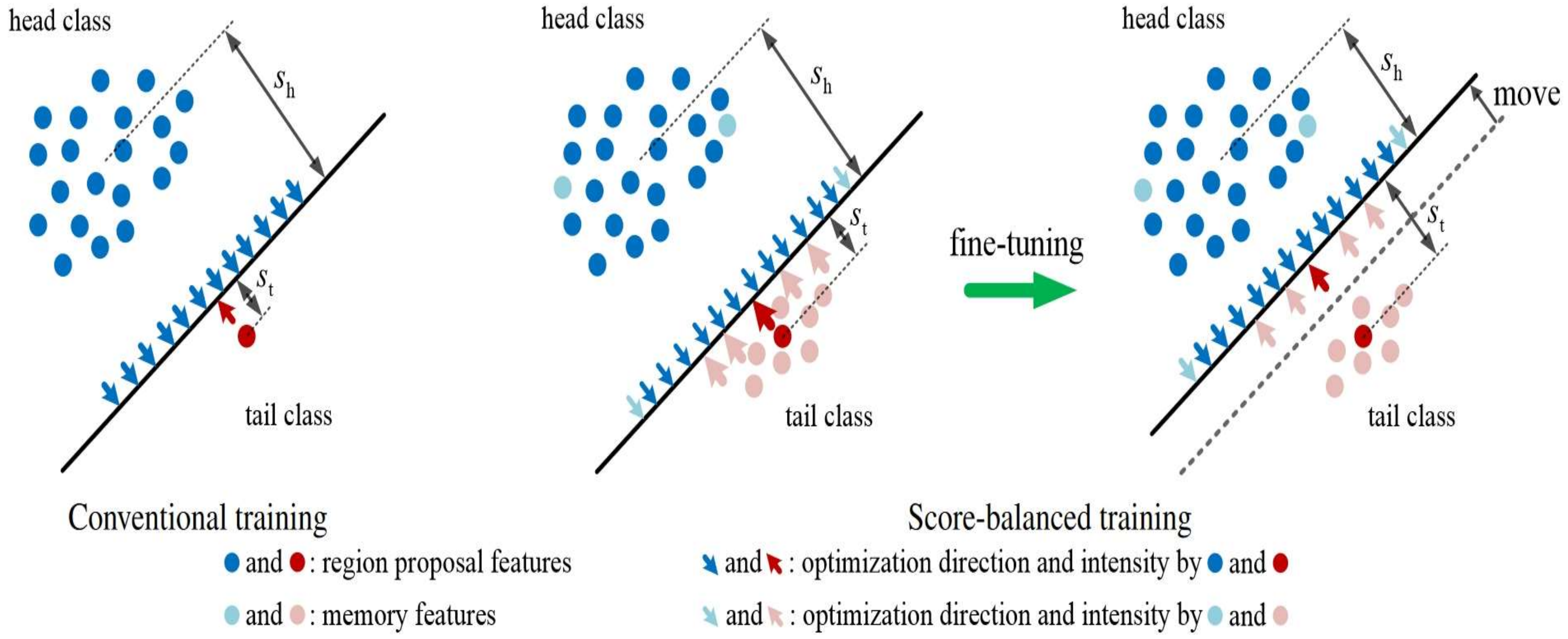
(a) Existing sampling methods

(b) The proposed method

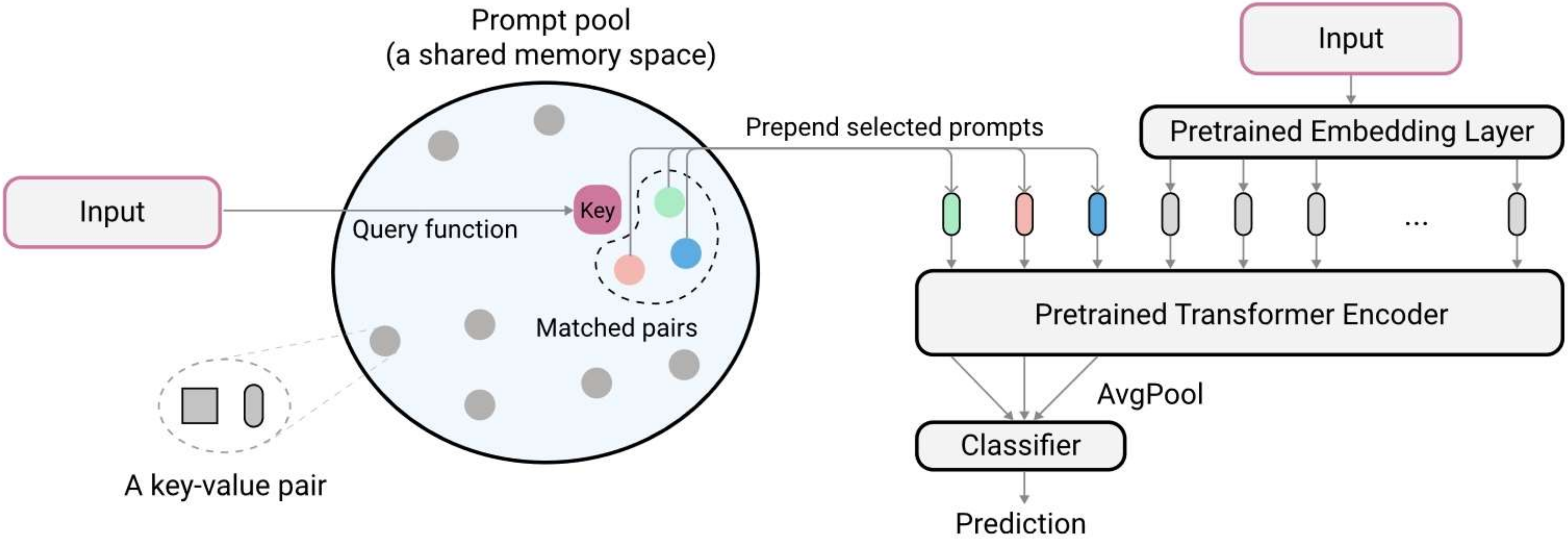
$$p_y = \frac{f(s_y)}{\sum_{y'} f(s_{y'})}, \qquad f(s_y) = \frac{1}{s_y}.$$

- Proposal resampling depend on RPN
- Tailed data overfitting

$$\hat{b} = [x_1 \pm \frac{\eta_1 w}{6}, y_1 \pm \frac{\eta_2 h}{6}, x_2 \pm \frac{\eta_3 w}{6}, y_2 \pm \frac{\eta_4 h}{6}], \qquad \mathbb{M}_y = [f_y^1, f_y^2, ..., f_y^M],$$


head class
s_h
s_t
tail class
Conventional training
head class
s_h
s_t
tail class
fine-tuning
head class
s_h
move
s_t
tail class
Score-balanced training
and : region proposal features
and : memory features
and : optimization direction and intensity by and
and : optimization direction and intensity by and

Method	Framework	Backbone	Dataset	AP^b	AP	AP_r	AP_c	AP_f
RFS [7]	Mask R-CNN	R-50-FPN	LVIS v0.5	26.1	25.9	17.8	26.2	28.8
EQL [22]				24.1	25.2	14.6	24.4	26.8
Forest R-CNN [27]				25.9	25.6	18.3	26.4	27.6
BAGS [13]				25.8	26.3	18.0	26.9	28.7
BALMS [19]				27.6	27.0	19.6	28.9	27.5
EQL v2 [21]†				27.0	27.1	18.6	27.6	29.9
LOCE (ours)				**28.2**	**28.4**	**22.0**	**29.0**	**30.2**
RFS [7]	Mask R-CNN	R-50-FPN	LVIS v1.0	24.7	23.7	13.5	22.8	29.3
EQL [22]				22.5	21.6	3.8	21.7	29.2
Seesaw Loss [24]$^{*\,\dagger}$				24.3	23.3	13.0	22.9	28.2
EQL v2 [21]†				26.1	25.5	17.7	24.3	30.2
LOCE (ours)				**27.4**	**26.6**	**18.5**	**26.2**	**30.7**
RFS [7]	Mask R-CNN	R-101-FPN	LVIS v1.0	26.6	25.5	16.6	24.5	30.6
EQL [22]				24.0	22.7	3.7	23.3	30.4
BAGS [13]				26.4	25.6	17.3	25.0	30.1
Seesaw Loss [24]†				27.4	27.1	18.7	26.3	31.7
EQL v2 [21]†				27.9	27.2	**20.6**	25.9	31.4
LOCE (ours)				**29.0**	**28.0**	19.5	**27.8**	**32.0**



$$\min_{\mathbf{P},\mathbf{K},\phi} \quad \mathcal{L}(g_\phi(f_r^{\text{avg}}(\boldsymbol{x}_p)), y) + \lambda \sum_{\mathbf{K}_{\boldsymbol{x}}} \gamma\left(q(\boldsymbol{x}), \boldsymbol{k}_{s_i}\right),$$