南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Balanced Classification: A Unified Framework for Long-Tailed Object Detection

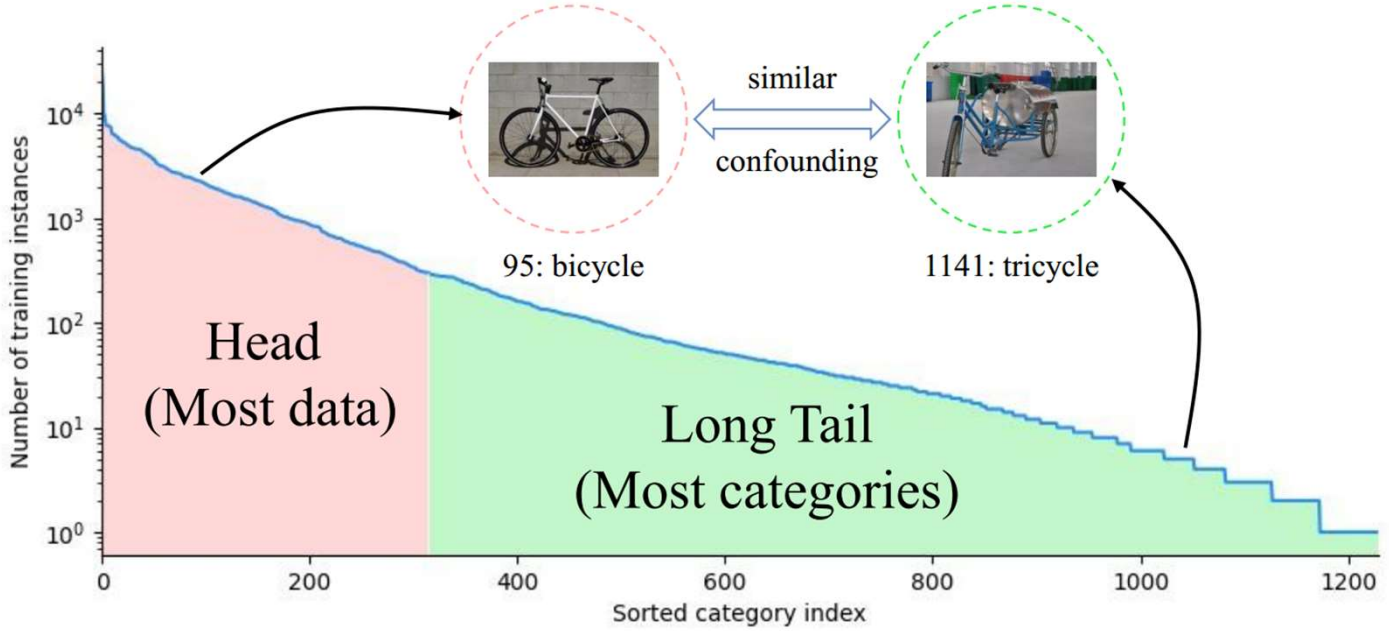
Tianhao Qi, Hongtao Xie*, Pandeng Li, Jiannan Ge and Yongdong Zhang, Senior Member, IEEE

CVPR 2023

Introduction

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Long-tail Problems



a few categories (head categories) occupy the vast majority of samples, while the remaining many categories (tail categories) only correspond to a small proportion of samples

Introduction

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

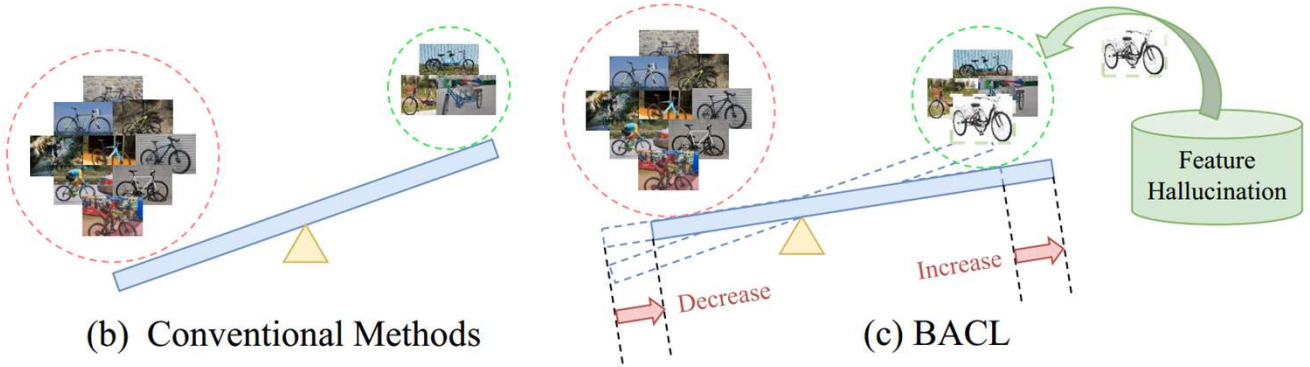
Motivation:

- **The unequal competition among foreground categories arising from extremely imbalanced frequencies;**
- **The latent issue of under-representation in tail categories, characterized by a scarcity of diverse visual instances.**

BAlanced CLassification (BACL)

two-stage

- **a foreground classification balance loss(FBCL)**
- **a dynamic feature hallucination module (FHM)**



Method

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Preliminary:

Following pioneering works [16], [22], we adopt the popular two-stage object detector Faster R-CNN to implement our proposed methods. In the Faster R-CNN pipeline, the backbone network randomly takes an image as input and generates a corresponding feature map. RPN performs convolution on the feature map and produces a fixed number of region proposals. Next, the RoIAlign layer first pools these region proposals into fixed size according to the feature map and then the RoI feature extractor, consisting of two fully connected (FC) layers, encodes them into d-dimensional RoI features $\boldsymbol{h}$. Finally, these RoI features are fed to two separate FC layers dedicated to classification and regression. In the classification branch, the network is tasked with solving a $(C+1)$-category prediction problem (C foreground categories and the background) by applying $\boldsymbol{z} = FC_{cls}(\boldsymbol{h})$, where $\boldsymbol{z} = [z_1, \cdots, z_{C+1}]$ represents the predicted logits.

Traditionally, the classification branch is supervised by the softmax cross-entropy loss during training. For a proposal labeled as i, the gradient passed by the loss for each category j is formulated as follows:

$$\frac{\partial L_{cls}}{\partial z_j} = \begin{cases} p_j - 1, & j = i \\ p_j, & j \neq i \end{cases}$$

Method

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Preliminary:

In order to improve classification accuracy, previous studies [21]–[23] have adopted a sigmoid-based classifier that has demonstrated superiority in large vocabulary datasets. EQLv2 introduces an additional objectness branch to reduce false positives, considering all proposals belonging to the background as positive samples.

a training sample with logits $\boldsymbol{z} = [z_1, \cdots, z_{C+1}]$

one-hot labels $\boldsymbol{y} = [y_1, \cdots, y_{C+1}]$

$$L_{cls} = -\sum_{i=1}^{C+1} \log(\hat{p}_i), \quad \text{where} \quad \hat{p}_i = \begin{cases} p_i, & y_i = 1 \\ 1 - p_i, & y_i = 0 \end{cases}$$

$$p_i = \frac{1}{1 + \exp(-z_i)}.$$

In the inference phase, the estimated probability vector $\tilde{\boldsymbol{p}} = [\tilde{p}_1, \tilde{p}_2, \cdots, \tilde{p}_{C+1}]$ of the $(C+1)$-channel sigmoid-based classifier can be expressed by the following equation:

$$\tilde{p}_i = \begin{cases} (1 - p_{C+1}) \cdot p_i, & 1 \le i \le C \\ p_{C+1}, & i = C+1 \end{cases} \tag{5}$$

Method

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Representation Learning Stage:

TABLE I

GRADUAL PERFORMANCE IMPROVEMENTS ON LVIS V0.5 VAL SET DURING THE REPRESENTATION LEARNING STAGE. AP^b DENOTES THE 101-POINT INTERPOLATED AVERAGE PRECISION FOR BOX PREDICTIONS OVER 10 IOU THRESHOLDS RANGING FROM 0.5 TO 0.95 AND ALL CLASSES, WHILE AP_r, AP_c, AND AP_f REPRESENT THE DETECTION AVERAGE PRECISION FOR RARE, COMMON, AND FREQUENT CATEGORIES, RESPECTIVELY.

Model	Sigmoid	Objectness Branch	Double Proposals	Small Weight Decay	Simple Copy-Paste	AP^b	AP_r	AP_c	AP_f
Faster R-CNN	✗	✗	✗	✗	✗	18.0	1.6	15.0	28.3
	✓	✗	✗	✗	✗	18.7	2.4	16.8	27.7
	✓	✓	✗	✗	✗	20.5	3.6	18.8	29.4
	✓	✓	✓	✗	✗	20.6	4.2	18.8	29.5
	✓	✓	✓	✓	✗	21.2	4.9	20.0	29.2
	✓	✓	✓	✓	✓	**21.7**	**4.9**	**20.4**	**30.1**

Method

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

The long-short-term indicators pair:

TABLE II

DIFFERENT TYPES OF LONG-TERM INDICATORS. THE CONSTANT COLUMN INDICATES WHETHER THIS KIND OF INDICATORS WILL ALTER DURING CALIBRATION. IN THE CRITERION COLUMN, WE ASSUME THAT CATEGORY i IS STRONGER THAN j. IN THE EQUATION COLUMN, $y_{n,i}$ IS THE ONE-HOT LABEL OF CATEGORY i FOR THE n-TH SAMPLE AND $\mathbb{I}[\cdot]$ IS AN INDICATOR FUNCTION THAT OUTPUTS 1 IF THE INPUT CONDITION HOLDS, OTHERWISE 0.

Types	Name	Equation	Constant	Dimension	Criterion
Static Stat.	image frequency	f_i	✓	1	$f_i > f_j$
	instance frequency	F_i	✓	1	$F_i > F_j$
First-Order Dynamic Stat.	cumulative instance number	$N_i = \sum_n \mathbb{I}[y_{n,i} = 1]$	✗	1	$N_i > N_j$
	mean classification score	$s_i^t = \gamma s_i^{t-1} + (1-\gamma)p_i^t$	✗	1	$s_i^t > s_j^t$
	true positive rate	$TPR_i = \frac{\sum_n \mathbb{I}[\operatorname{argmax}_k p_{n,k}=i] \cdot \mathbb{I}[y_{n,i}=1]}{\sum_n \mathbb{I}[y_{n,i}=1]}$	✗	1	$TPR_i > TPR_j$
Second-Order Dynamic Stat.	confusion matrix	$M_{i,j} = \frac{\sum_n \overline{p}_{n,j} \cdot \mathbb{I}[y_{n,i}=1]}{\sum_n \mathbb{I}[y_{n,i}=1]}$, $\overline{p}_{n,j} = \frac{\exp(z_j)}{\sum_{k=1}^{C} \exp(z_k)}$	✗	2	$M_{j,i} > M_{i,j}$

In the inference phase, the estimated probability vector $\tilde{\boldsymbol{p}} = [\tilde{p}_1, \tilde{p}_2, \cdots, \tilde{p}_{C+1}]$ of the $(C+1)$-channel sigmoid-based classifier can be expressed by the following equation:

$$\tilde{p}_i = \begin{cases} (1 - p_{C+1}) \cdot p_i, & 1 \le i \le C \\ p_{C+1}, & i = C+1 \end{cases} \tag{5}$$

Method

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Formulation of Foreground Classification Balance Loss:

$$L_{FCBL} = -\log(p_i) - \log(1-p_{C+1}) - \sum_{j=1,j\neq i}^{C} w_j \log(1-p'_j),$$

Firstly, FCBL introduces an adaptive class-aware margin between any pair of foreground categories, in order to ameliorate the domination of one category over another. The margin is logarithmically proportional to the ratio of the corresponding long-term indicators. For example, considering a sample from foreground category i, the margin δ_{ij} between i and another foreground category j is defined by Eq. (7):

$$\delta_{ij} = \alpha \cdot \log(\frac{l_j}{l_i}) \tag{7}$$

where α controls the range of the margin, and l_i/l_j represents a unified expression for long-term indicators, which can take the form of f_i, F_i, N_i, s_i^t, TPR_i, or $M_{j,i}$ for l_i.

The probability for each non-ground-truth foreground category j is then reformulated according to Eq. (8):

$$p'_j = \frac{1}{1+\exp[-(z_j+\delta_{ij})]} \tag{8}$$

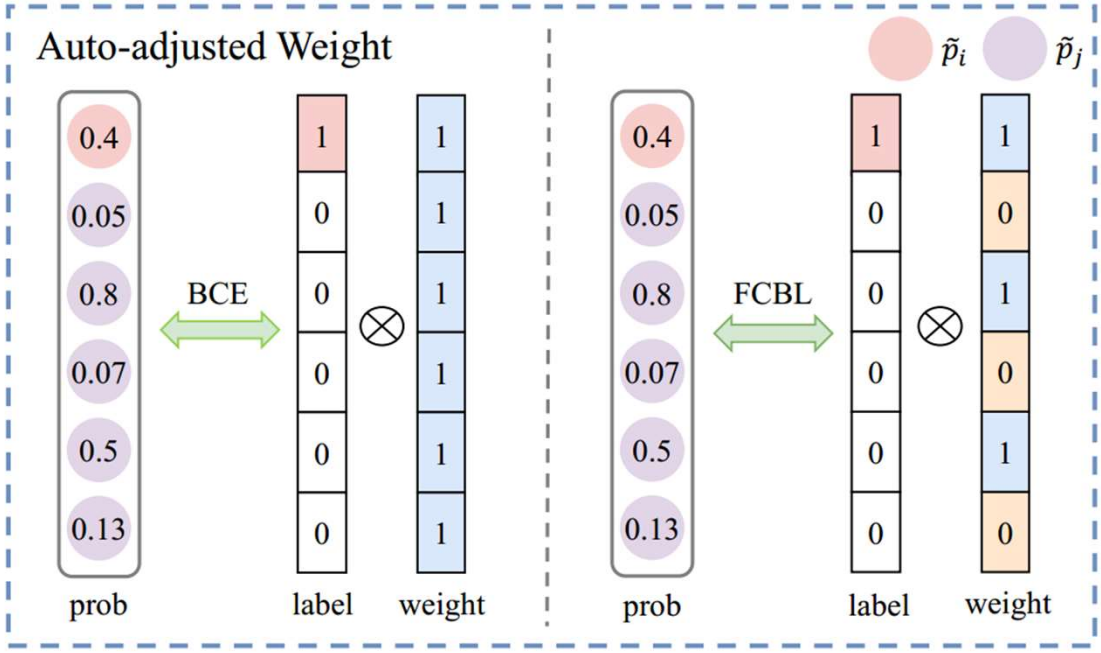
FCBL incorporates an auto-adjusted weight term w_j, defined by Eq. (9), into the binary cross-entropy loss of each non-ground-truth foreground category j (where $1 \leq j \leq C$ and $j \neq i$), where $\tilde{p}_t$ is a pre-defined threshold and $\tilde{p}_i$, $\tilde{p}_j$ represent the corresponding short-term indicators.

$$w_j = \begin{cases} 1, & \tilde{p}_j \geq \tilde{p}_i \\ 1, & \tilde{p}_j \geq \tilde{p}_t \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Method

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Formulation of Foreground Classification Balance Loss:



$$w_j = \begin{cases} 1, & \tilde{p}_j \geq \tilde{p}_i \\ 1, & \tilde{p}_j \geq \tilde{p}_t \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Fig. 3. Visualization of the auto-adjusted weight in FCBL. In default, the tunable threshold $\tilde{p}_t$ is set to 0.7.

Method

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Feature Hallucination Module:

Concretely, FHM generates region proposals that exhibit substantial overlaps with the ground-truth bounding boxes using a non-learnable bounding box generator, as inspired by [18]. In contrast to RPN, the bounding box generator uses coordinate manipulation to stochastically transform the ground-truth bounding boxes in an image into positive proposals. Formally, for a batch of training images I, the designed box generator takes the box coordinates $b_{GT} = [x_1, y_1, x_2, y_2]$ of each instance in I as input, where (x_1, y_1), (x_2, y_2) represent the coordinates of the top-left and bottom-right corners, respectively. It then generates 16 dense proposals for each instance in I, with slight offsets in their locations, as follows:

$$\hat{b} = [x_1 + \frac{\eta_1 w}{6}, y_1 + \frac{\eta_2 h}{6}, x_2 + \frac{\eta_3 w}{6}, y_2 + \frac{\eta_4 h}{6}], \quad (10)$$

Afterward, the subsequent RoIAlign layer and RoI feature extractor encode them as RoI features, not for classification and regression, but for collecting online feature distributions, which include prototypes and variances. In detail, FHM computes the mean u_i and variance v_i of features for each category i that emerges in I and then alters the corresponding prototype μ_i and variance σ_i using the exponential moving average function defined in Eq. (11).

$$\begin{aligned} \mu_i &\leftarrow \beta\mu_i + (1-\beta)u_i, \\ \sigma_i &\leftarrow \beta\sigma_i + (1-\beta)v_i. \end{aligned} \quad (11)$$

Lastly, FHM guarantees the prominence of tail categories by allocating a sampling probability sp_i to each category i, which is inversely proportional to the long-term indicator l_i:

$$sp_i = \frac{1 - l_i}{\sum_{k=1}^{C}(1 - l_k)}, \; i \in 1, \cdots C, \quad (12)$$

where l_i can take the form of f_i, F_i, N_i, s_i^t, TPR_i, or $M_{i,i}$. Utilizing the aforementioned sampling probabilities, FHM randomly selects c categories and generates m hallucinated features for each category i by the constantly renewed feature distribution through the reparametrization trick [28]:

$$f_i = \mu_i + \epsilon \odot \sigma_i, \; \epsilon \in \mathcal{N}(0, 1). \quad (13)$$

Therefore, guided by the long-term indicators, FHM dynamically intensifies the data diversity by introducing novel hallucinated features, particularly for tail categories, thereby mitigating the issue of under-representation.

Method

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Feature Hallucination Module:

Algorithm 1 The pipeline of the classifier learning stage

Input: Batch annotated images $\mathcal{I}$ with N_I objects belonging to C_I classes

1: **Initialize:** frozen feature extractor $\mathcal{H}$ (*i.e.*, backbone, RoIAlign layer and RoI feature extractor) from the representation learning stage
2: **for** $1,\cdots,N_I$ **do**
3: $\hat{b} = [x_1 + \frac{\eta_1 w}{6}, y_1 + \frac{\eta_2 h}{6}, x_2 + \frac{\eta_3 w}{6}, y_2 + \frac{\eta_4 h}{6}]$ ▷ boxes transformation with $b_{GT} = [x_1, y_1, x_2, y_2]$ and label i
4: $h = \mathcal{H}(\hat{b})$
5: **end for**
6: **for** i in $1,\cdots,C_I$ **do**
7: Calculate u_i and v_i of all features h with label i
8: $\mu_i \leftarrow \beta\mu_i + (1-\beta)u_i, \quad \sigma_i \leftarrow \beta\sigma_i + (1-\beta)v_i$ ▷ μ_i, σ_i are the prototype and variance of class i
9: **end for** ▷ feature distributions alteration
10: Select c classes with probabilities sp_i in Eq. (12)
11: **for** i in $1,\cdots,c$ **do**
12: Repeat $f_i = \mu_i + \epsilon \odot \sigma_i, \quad \epsilon \in \mathcal{N}(0,1)$ for m times
13: **end for** ▷ hallucinated features synthesization
14: Update long-short-term indicators pair in Tab. II and Eq. (5) with hallucinated features and RoI features from RPN
15: Calculate classification loss with Eq. (6)

Output: The classification loss L_{FCBL}

$$sp_i = \frac{1 - l_i}{\sum_{k=1}^{C}(1 - l_k)}, \; i \in 1, \cdots C,$$

$$\tilde{p}_i = \begin{cases} (1 - p_{C+1}) \cdot p_i, & 1 \le i \le C \\ p_{C+1}, & i = C+1 \end{cases}$$

$$L_{FCBL} = -\log(p_i) - \log(1 - p_{C+1}) - \sum_{j=1, j\neq i}^{C} w_j \log(1 - p'_j),$$

Experiments

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Ablation Study: The representation learning stage

All related experiments are conducted on LVIS v0.5 by using Faster R-CNN with ResNet-50-FPN backbone.

TABLE I

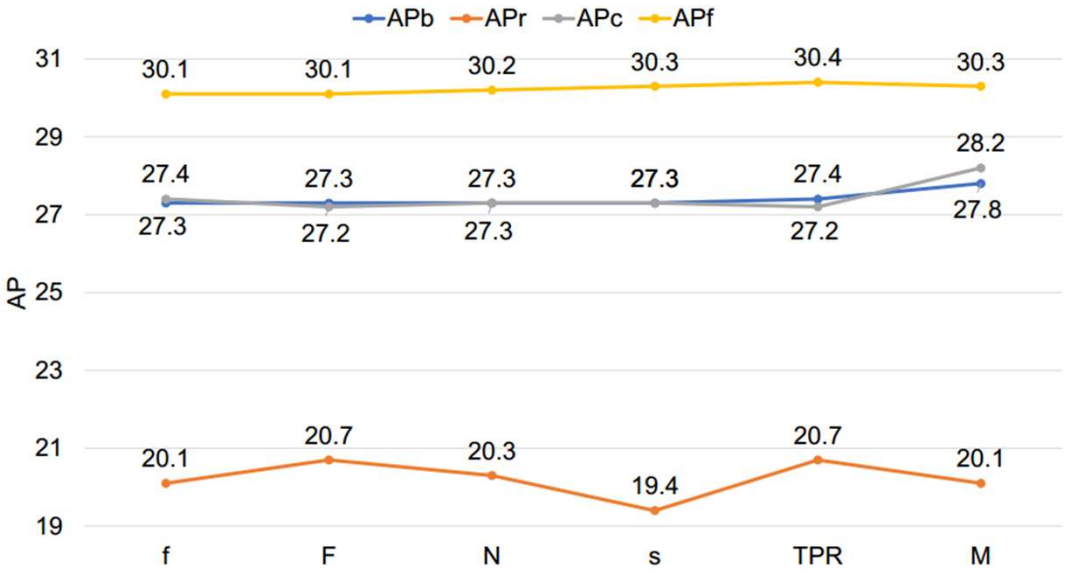
GRADUAL PERFORMANCE IMPROVEMENTS ON LVIS V0.5 VAL SET DURING THE REPRESENTATION LEARNING STAGE. AP^b DENOTES THE 101-POINT INTERPOLATED AVERAGE PRECISION FOR BOX PREDICTIONS OVER 10 IOU THRESHOLDS RANGING FROM 0.5 TO 0.95 AND ALL CLASSES, WHILE AP_r, AP_c, AND AP_f REPRESENT THE DETECTION AVERAGE PRECISION FOR RARE, COMMON, AND FREQUENT CATEGORIES, RESPECTIVELY.

Model	Sigmoid	Objectness Branch	Double Proposals	Small Weight Decay	Simple Copy-Paste	AP^b	AP_r	AP_c	AP_f
Faster R-CNN	✗	✗	✗	✗	✗	18.0	1.6	15.0	28.3
	✓	✗	✗	✗	✗	18.7	2.4	16.8	27.7
	✓	✓	✗	✗	✗	20.5	3.6	18.8	29.4
	✓	✓	✓	✗	✗	20.6	4.2	18.8	29.5
	✓	✓	✓	✓	✗	21.2	4.9	20.0	29.2
	✓	✓	✓	✓	✓	**21.7**	**4.9**	**20.4**	**30.1**

Experiments

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Ablation Study: Comparison of various long-term indicators



Name	Equation
image frequency	f_i
instance frequency	F_i
cumulative instance number	$N_i = \sum_n \mathbb{I}[y_{n,i} = 1]$
mean classification score	$s_i^t = \gamma s_i^{t-1} + (1-\gamma) p_i^t$
true positive rate	$TPR_i = \frac{\sum_n \mathbb{I}[\operatorname{argmax}_k p_{n,k}=i] \cdot \mathbb{I}[y_{n,i}=1]}{\sum_n \mathbb{I}[y_{n,i}=1]}$
confusion matrix	$M_{i,j} = \frac{\sum_n \overline{p}_{n,j} \cdot \mathbb{I}[y_{n,i}=1]}{\sum_n \mathbb{I}[y_{n,i}=1]}$, $\overline{p}_{n,j} = \frac{\exp(z_j)}{\sum_{k=1}^{C} \exp(z_k)}$

Fig. 4. Performance comparison of different long-term indicators as listed in Tab. II on the challenging LVIS v0.5.

Experiments

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Ablation Study: Component Analysis

- an adaptive class-aware margin
- an auto-adjusted weight term
- a dynamic feature hallucination module

TABLE III

RESULTS FOR ALL ARBITRARY COMBINATIONS OF THREE COMPONENTS. MG, WT, AND FHM REPRESENT THE CLASS-AWARE MARGIN, WEIGHT TERM, AND FEATURE HALLUCINATION MODULE, RESPECTIVELY.

Method	MG	WT	FHM	AP^b	AP_r	AP_c	AP_f
Baseline				22.0	4.0	21.2	30.1
-	✓			24.3	8.7	24.7	30.1
-		✓		24.0	8.5	23.8	30.3
FHM			✓	26.5	17.9	26.7	29.8
FCBL	✓	✓		24.2	8.5	24.5	30.3
-	✓		✓	26.7	18.3	27.1	29.6
-		✓	✓	27.2	20.7	26.9	30.2
BACL	✓	✓	✓	**27.8**	**20.1**	**28.2**	**30.3**

Experiments

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Ablation Study: Hyper-parameters

- α IN FCBL
- $\tilde{p}_t$ IN FCBL
- β IN FHM
- c, m IN FHM

TABLE V
RESULTS FOR DIFFERENT VALUES OF $\tilde{p}_t$ IN FCBL.

$\tilde{p}_t$	AP^b	AP_r	AP_c	AP_f
0.9	27.5	20.8	27.4	30.2
0.8	27.6	21.6	27.4	30.2
0.7	**27.8**	**20.1**	**28.2**	**30.3**
0.6	27.6	21.0	27.6	30.2
0.5	27.5	20.9	27.5	30.1

TABLE VI
RESULTS FOR DIFFERENT VALUES OF β IN FHM.

β	AP^b	AP_r	AP_c	AP_f
0.95	27.1	20.1	26.9	30.2
0.9	**27.8**	**20.1**	**28.2**	**30.3**
0.85	27.7	20.2	27.9	30.3

TABLE IV
RESULTS FOR DIFFERENT VALUES OF α IN FCBL.

α	AP^b	AP_r	AP_c	AP_f
0	27.2	20.7	26.9	30.2
0.7	27.4	21.0	27.2	30.2
0.8	27.5	21.0	27.6	30.1
0.85	**27.8**	**20.1**	**28.2**	**30.3**
0.9	27.6	21.3	27.5	30.2
1.0	27.5	21.3	27.4	30.2

TABLE VII
RESULTS FOR DIFFERENT VALUES OF c, m IN FHM.

c	m	AP^b	AP_r	AP_c	AP_f
12	12	27.3	20.4	27.2	30.3
8	16	27.6	20.5	27.7	30.2
8	**12**	**27.8**	**20.1**	**28.2**	**30.3**
8	8	27.3	20.5	27.2	30.2
4	12	27.5	20.8	27.5	30.2

Experiments

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Ablation Study: Formulation of the confusion matrix

$$M_{i,j} = \frac{\sum_n \mathbb{I}[\operatorname{argmax}_k p_{n,k} = j] \cdot \mathbb{I}[y_{n,i} = 1]}{\sum_n \mathbb{I}[y_{n,i} = 1]}, \quad 1 \le i, j \le C. \qquad \times$$

$$M_{i,j} = \frac{\sum_n \overline{p}_{n,j} \cdot \mathbb{I}[y_{n,i}=1]}{\sum_n \mathbb{I}[y_{n,i}=1]}, \qquad \surd$$

$$\overline{p}_{n,j} = \frac{\exp(z_j)}{\sum_{k=1}^{C} \exp(z_k)}$$

TABLE VIII

RESULTS FOR TWO FORMULAS OF THE CONFUSION MATRIX.

$M_{i,j}$	AP^b	AP_r	AP_c	AP_f
Eq in Last Row of Tab. II	**27.8**	**20.1**	**28.2**	**30.3**
Eq. (14)	27.3	19.9	27.2	30.3

Experiments

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Ablation Study: Formulation of the weight term

$$w_j = \begin{cases} 1, & \tilde{p}_j \geq \tilde{p}_i \\ 1, & \tilde{p}_j \geq \tilde{p}_t \\ 0, & \text{otherwise} \end{cases}$$

TABLE IX

RESULTS FOR DIFFERENT FORMULATIONS OF THE WEIGHT TERM.

$\tilde{p}_j \geq \tilde{p}_i$	$\tilde{p}_j \geq \tilde{p}_t$	AP^b	AP_r	AP_c	AP_f
✓		27.5	20.8	27.5	30.2
	✓	20.1	11.1	20.5	23.0
✓	✓	**27.8**	**20.1**	**28.2**	**30.3**

Experiments

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Ablation Study： Training pipelines

TABLE X
RESULTS FOR TWO KINDS OF TRAINING PIPELINES.

Strategy	AP^b	AP_r	AP_c	AP_f
End-to-end	22.0	6.3	21.5	28.9
Decoupled	**27.8**	**20.1**	**28.2**	**30.3**

Experiments

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Effectiveness of the Overall Framework:

TABLE XII
RESULTS FOR THE BASELINE MODEL AND BACL ON VARIOUS ARCHITECTURES AND BACKBONES ON LVIS V1.0 VAL SET.

Architecture	Backbone	BACL	AP^b	AP_r	AP_c	AP_f
Faster R-CNN [13]	R-50-FPN	✗	19.3	1.1	16.1	**30.9**
		✓	**26.1**	**16.0**	**25.6**	**30.9**
	R-101-FPN	✗	20.9	1.0	18.2	**32.7**
		✓	**27.8**	**18.1**	**27.3**	32.6
Cascade Faster R-CNN [34]	R-50-FPN	✗	22.7	1.5	20.6	**34.4**
		✓	**28.6**	**21.3**	**27.7**	32.8
	R-101-FPN	✗	24.5	2.6	23.1	**35.8**
		✓	**29.8**	**22.0**	**28.8**	34.3

Experiments

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Comparison with State-of-the-Arts:

TABLE XIII
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART METHODS ON LVIS *val* SET. THE RESNET-50-FPN AND RESNET-101-FPN ARE ADOPTED AS BACKBONES FOR FASTER R-CNN. ALL METHODS ARE TRAINED WITH A 2X SCHEDULE, *i.e.*, 24 EPOCHS IN TOTAL. BCE* DENOTES SIGMOID CROSS-ENTROPY LOSS WITH AN OBJECTNESS BRANCH. **BOLD** NUMBERS DENOTE THE BEST RESULTS.

Strategy	Methods	LVIS v0.5								LVIS v1.0							
		ResNet-50-FPN				ResNet-101-FPN				ResNet-50-FPN				ResNet-101-FPN			
		AP_b	AP_r	AP_c	AP_f	AP_b	AP_r	AP_c	AP_f	AP_b	AP_r	AP_c	AP_f	AP_b	AP_r	AP_c	AP_f
End-to-end	SCE [13]	22.0	4.0	21.2	30.1	23.3	2.8	23.2	31.5	19.3	1.1	16.1	30.9	20.9	1.0	18.2	32.7
	BCE [13]	21.8	5.3	21.0	29.5	23.7	5.7	23.3	31.3	19.5	1.6	16.6	30.6	21.4	2.0	19.3	32.3
	BCE* [13]	23.9	6.7	23.8	31.0	25.2	7.8	25.2	32.0	21.6	2.6	19.7	32.1	23.1	3.7	21.4	**33.5**
	RFS [14]	25.0	14.1	24.8	29.6	25.9	14.8	25.5	30.8	24.2	14.2	22.3	30.6	25.7	15.9	23.7	32.2
	EQL [21]	24.0	9.4	24.4	29.2	25.5	9.9	26.1	31.1	21.8	3.6	21.1	30.5	23.4	4.5	22.9	32.3
	DropLoss [65]	23.3	9.7	24.7	27.1	26.1	11.2	28.5	29.0	21.8	5.2	21.8	29.1	23.5	5.9	23.9	30.7
	RIO [17]	24.4	15.7	24.0	28.4	26.4	16.4	26.4	30.5	23.4	15.3	21.2	29.4	25.5	17.2	23.7	31.2
	Forest R-CNN [26]	26.0	16.6	26.3	29.4	26.9	15.2	27.6	30.0	-	-	-	-	-	-	-	-
	BALMS [20]	25.5	17.6	25.0	29.3	27.2	17.3	27.3	31.0	24.1	15.2	23.0	29.4	26.9	**18.5**	25.2	32.4
	De-confound-TDE [66]	25.3	13.2	25.4	30.0	27.3	14.3	28.0	31.5	23.7	10.0	22.4	31.2	-	-	-	-
	EQLv2 [23]	26.5	17.2	26.2	30.7	27.4	18.3	26.7	31.8	25.4	15.8	23.5	31.7	26.8	17.1	24.9	33.1
	Seesaw [25]	26.3	16.3	26.1	30.6	27.3	16.8	26.8	32.0	24.8	14.8	22.7	31.6	26.6	14.9	25.2	33.3
	FASA [56]	23.6	10.7	22.8	29.6	24.3	11.3	23.2	30.9	21.5	7.4	19.2	30.2	22.9	9.0	20.6	31.6
	PCB [62]	23.9	9.1	22.9	**31.2**	26.5	11.4	26.2	**32.9**	23.0	6.2	21.5	**32.2**	24.6	8.0	23.1	**33.5**
Decoupled	SimCal [15]	22.5	13.6	20.3	29.0	-	-	-	-	-	-	-	-	-	-	-	-
	BAGS [16]	25.5	16.8	25.6	28.8	26.6	16.2	26.7	30.7	23.7	14.2	22.2	29.6	25.4	14.9	25.2	31.4
	ACSL [22]	23.7	14.8	23.5	27.5	25.7	16.5	25.8	29.1	22.2	9.9	21.3	28.5	23.7	11.0	23.0	30.2
	DisAlign [24]	25.2	14.1	25.2	29.5	27.4	15.9	27.8	31.5	20.9	3.9	20.4	29.0	25.5	13.3	24.5	32.0
	LOCE [18]	26.7	18.3	27.5	28.9	27.9	21.9	27.7	30.5	25.1	15.7	24.2	30.1	26.7	18.4	25.5	31.7
	BACL	**27.8**	**20.1**	**28.2**	30.3	**29.4**	**22.1**	**30.1**	31.3	**26.1**	**16.0**	**25.7**	30.9	**27.8**	18.1	**27.3**	32.6

Experiments

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Extension to Long-Tailed Instance Segmentation:

TABLE XIV
COMPARISON RESULTS FOR INSTANCE SEGMENTATION ON LVIS V1.0 VAL SET. AP^m, AP_r, AP_c, AP_f ARE MASK APS.

Backbone	BACL	AP^m	AP^b	AP_r	AP_c	AP_f
R-50-FPN	✗	19.4	19.9	1.3	17.2	**29.9**
	✓	**25.4**	**26.1**	**17.7**	**25.0**	29.1
R-101-FPN	✗	20.8	21.8	1.5	19.5	30.8
	✓	**27.2**	**28.4**	**19.3**	**27.0**	**30.9**

南京航空航天大學
Nanjing University of Aeronautics and Astronautics

Thanks