南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Learning in Imperfect Environment: Multi-Label Classificationwith Long-Tailed Distribution and Partial Labels

Wenqiao Zhang [1] Changshuo Liu [2] Lingze Zeng [2] Bengchin Ooi [2] Siliang Tang [1]
Yueting Zhuang [1]

[1] Zhejiang University, China, [2] National University of Singapore, Singapore,

wenqiaozhang@zju.edu.cn, liu717@comp.nus.edu.sg, Zenglz_pro@163.com, ooibc@comp.nus.edu.sg, siliang@zju.edu.cn, yzhuang@zju.edu.cn

ICCV 2023

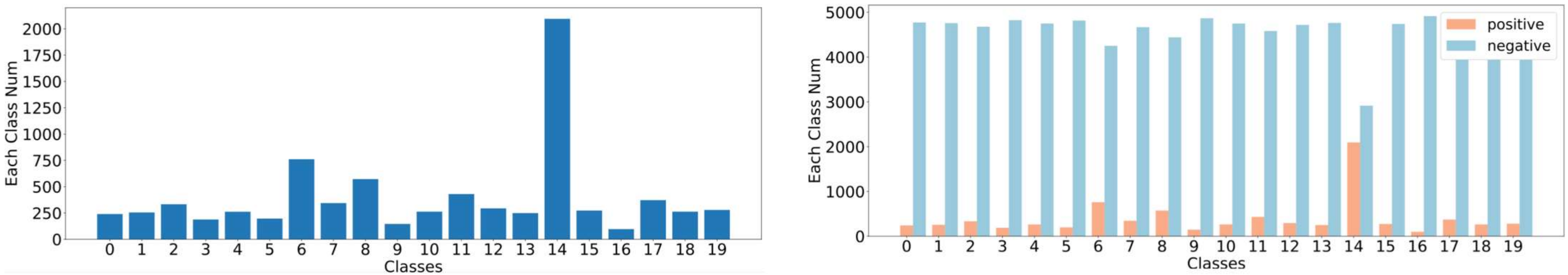
Background

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Imbalance between different classes

Imbalance exists in ratio of positives to negatives for each class

- Head class have a large ratio
- Tail classes have a small ratio

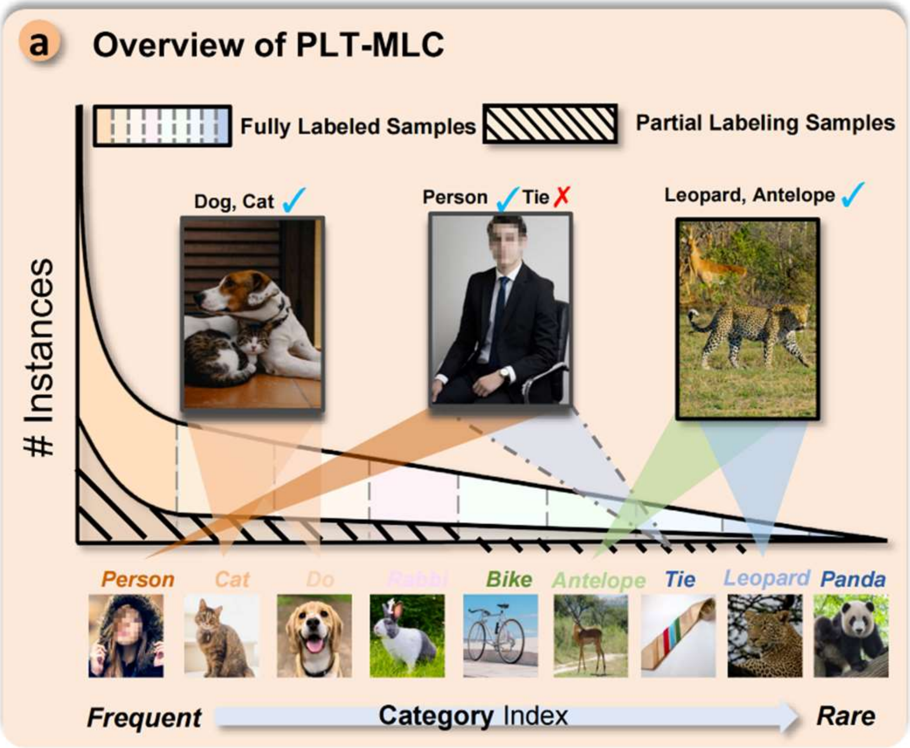


(a) (b)

Background

Complexities that typically arise in real-world applications:

i) Long-Tailed (LT) Class Distribution.
ii) Partial Labels (PL)of Instances.



illustrates an overview of the proposed PLT-MLC task.

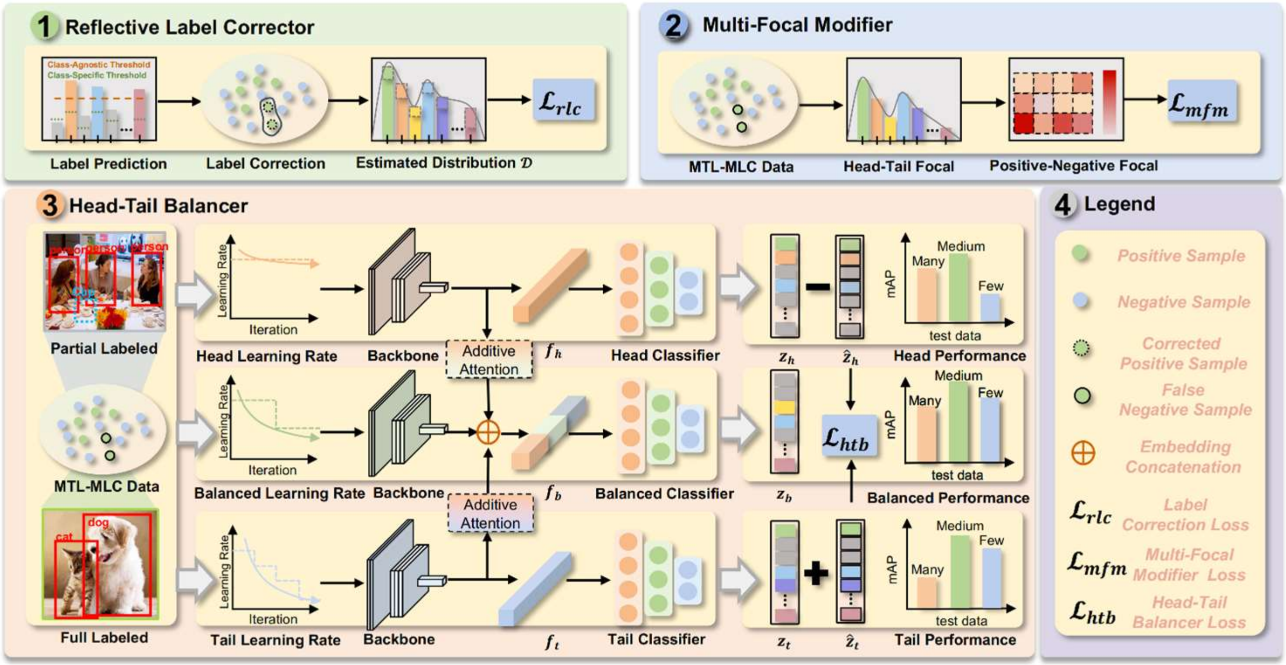
i) False Negative Training.
ii) Head-Tail and Positive-Negative Imbalance.
iii) Head Overfitting and Tail Underfitting.

Consequently,a robust PLT-MLC model should address the co-occurringimbalances simultaneously.

Methods

Correction→Modification→Balance

Reflective Label Corrector(RLC), Multi-Focal Modifier(MFM) and Head-Tail Balancer(HTB)



RLC module (Correction) corrects the missing labels along with the training and dynamically re-weights the sample weight according to the estimated class distribution.

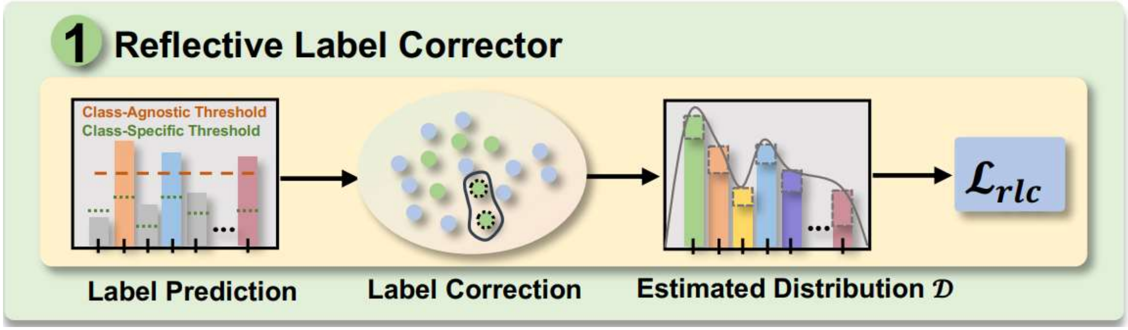
MFM module (Modification) adjusts the focal of different instances according tohead-tail and positive-negative imbalance under the extreme LT distribution.

HTB module(Balance) measures the model's optimization direction and correspondingly develops a balanced learning scheme to producestable PLT-MLC performance.

$$\underbrace{\mathcal{L}((\mathcal{S});\Theta_b)}_{\text{COMIC Loss}} = \underbrace{\lambda_c \cdot \mathcal{L}_{rlc}}_{\text{RLC Loss}} + \underbrace{\lambda_m \cdot \mathcal{L}_{mfm}}_{\text{MFM Loss}} + \underbrace{\lambda_b \cdot \mathcal{L}_{htb}}_{\text{HTB Loss}}$$

Methods

Reflective Label Corrector (RLC):

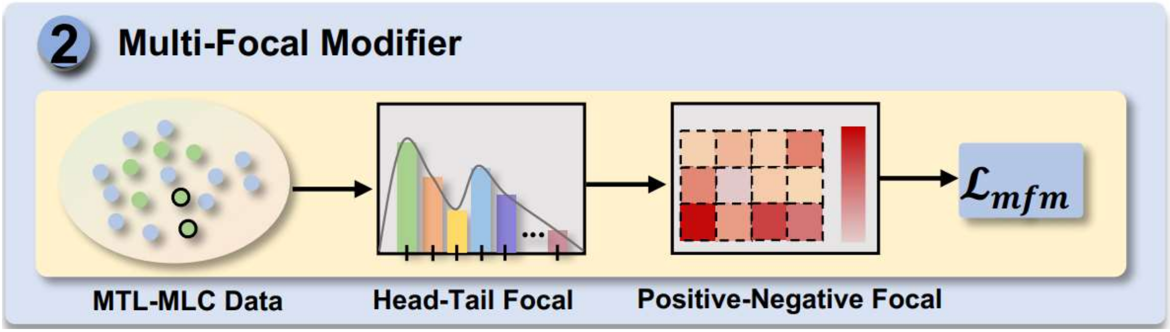


the average category possibility of past trained data with class c

$$\hat{y}_c = \begin{cases} 1, & \text{if } p_c > \max\{\tau, \boxed{P_c}\}, y_c = 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\mathcal{L}_{rlc}(p) = \begin{cases} \mathcal{L}^{+}_{mfm}(p), & \text{if } \hat{y} = 1 \\ \mathbb{1}_{(y=1)}\, \mathcal{L}^{+}_{mfm}(p) + \mathbb{1}_{(\hat{y}=0)}\, \mathcal{L}^{-}_{mfm}(p), & \text{otherwise} \end{cases}$$

Multi-Focal Modifier（MFM）:



$$\mathcal{L}_{fl}(p) = \begin{cases} \mathcal{L}^{+}_{fl} = (1-p)^{\gamma}\log(p), & \text{if } y = 1 \\ \mathcal{L}^{-}_{fl} = p^{\gamma}\log(1-p), & \text{if } y = 0 \end{cases}$$

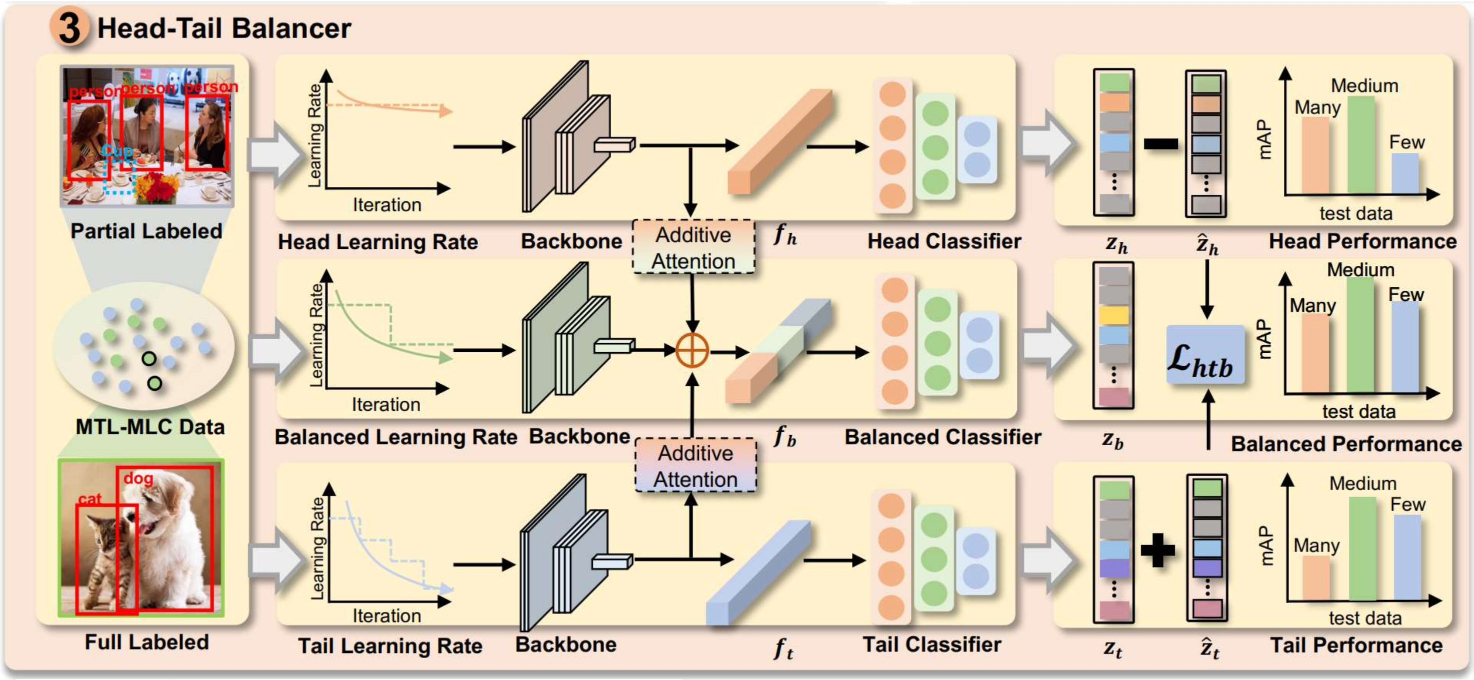
$$\gamma^{(i)} = \begin{cases} \gamma^{(i)+} = \gamma^{+}_{pn} + w^{+} \cdot \gamma^{(i)}_{ht}, & \text{if } y = 1 \\ \gamma^{(i)-} = \gamma^{-}_{pn} + w^{-} \cdot \boxed{\gamma^{(i)}_{ht}}, & \text{if } y = 0 \end{cases}$$

the static class distribution D of training set
with max normalization function

$$\mathcal{L}_{mfm}(p) = \begin{cases} \mathcal{L}^{+}_{mfm} = \sum_{i=1}^{C} (1-p)^{\gamma^{(i)+}} \log(p), & \text{if } y = 1 \\ \mathcal{L}^{-}_{mfm} = \sum_{i=1}^{C} p^{\gamma^{(i)-}} \log(1-p), & \text{if } y = 0 \end{cases}$$

Methods

Head-Tail Balancer



the accumulated gradient

$$\mathbf{e}_t = \mu \cdot \mathbf{e}_{t-1} + \boxed{\mathrm{sum}}(g_t), \forall t = 1, \cdots, T$$

$$\mathbf{f}_b = \mathrm{Attn}(\hat{\mathbf{f}}_b, [\mathbf{f}_h, \mathbf{f}_t])) + \hat{\mathbf{f}}_b$$

$$\mathbf{z}_x = \frac{\rho}{N_g} \sum_{k=1}^{N_g} \frac{w_k^\top \mathbf{f}_x}{(||w_k|| + \eta)||\mathbf{f}_x||}, x \in \{h, t, b\}$$

ρ: a scaling factor akin to the inverse temperature in Gibbs distribution
η: a class-agnostic baseline energy
W_k: the k-th learned parameter matrix.

$$\hat{\mathbf{z}}_x = \mathbf{z}_x \pm \frac{\rho}{N_g} \sum_{k=1}^{N_g} \frac{\mathrm{sim}(\mathbf{z}_x, \mathbf{e}_t) \cdot (w_j)^\top \mathbf{e}_t}{||w_k|| + \eta}, x \in \{h, t\}$$

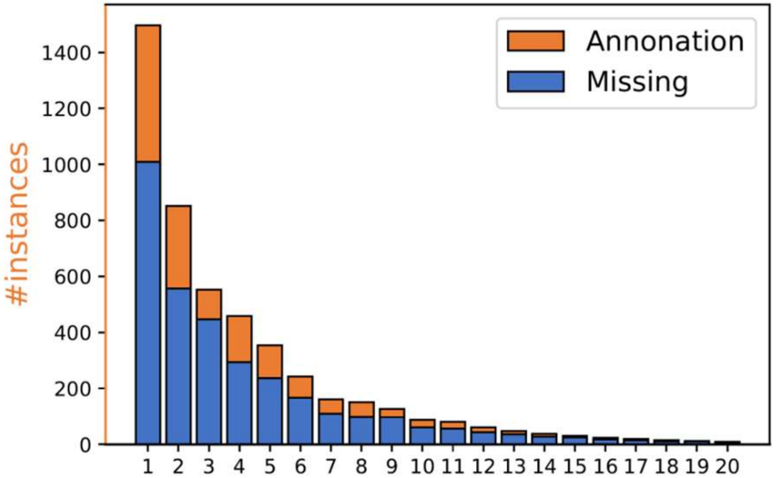
$\mathcal{L}_{mfm}$

softmax

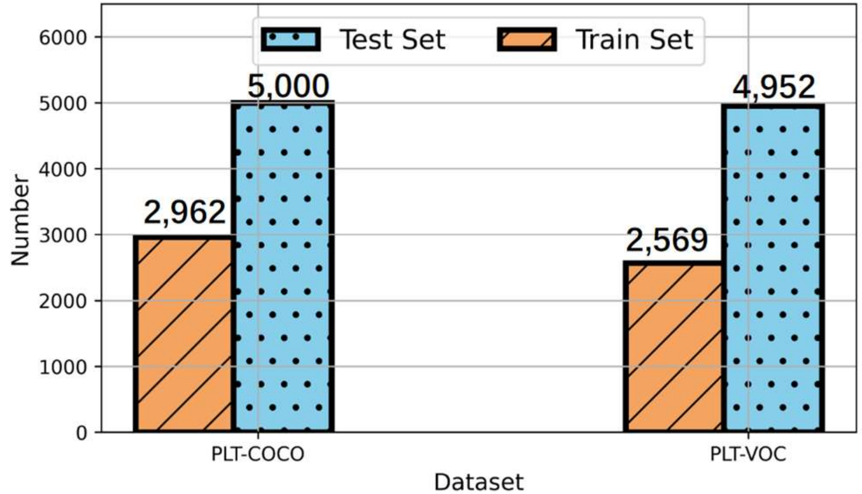
$$\mathcal{L}_{htb} = \kappa_h \cdot \boxed{\mathcal{L}}(\phi(\hat{\mathbf{z}_h}) \cdot \phi(\mathbf{z}_b)) + \kappa_t \cdot \mathcal{L}(\boxed{\phi}(\hat{\mathbf{z}}_t) \cdot \phi(\mathbf{z}_b))$$

$$\kappa_h = \frac{(\mathcal{L}(\hat{z_h}))^\alpha}{(\mathcal{L}(\hat{z_t}))^\alpha + (\mathcal{L}(\hat{z_h}))^\alpha} \text{ and } \kappa_t = \frac{(\mathcal{L}(\hat{z_t}))^\alpha}{(\mathcal{L}(\hat{z_t}))^\alpha + (\mathcal{L}(\hat{z_h}))^\alpha}$$

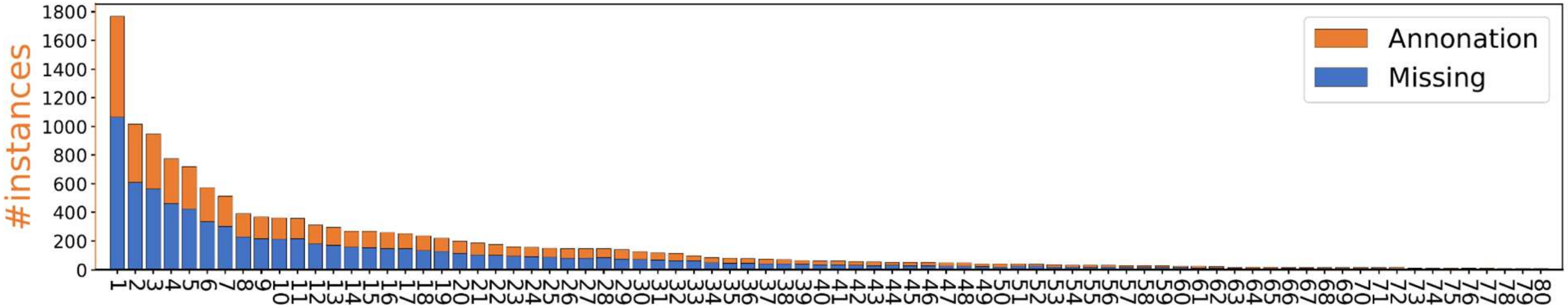
Experiments



(a) PLT-VOC Dataset



(b) Statistics of Two Datasets



Experiments

南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

$E2E^*$ indicates that the PLT model is learned in an end-to-end manner.
We color each row as the **best**, **second best** and **lowest score**

Category	Methods	$E2E^*$	PLT-COCO Dataset				PLT-VOC Dataset			
			Many Shot	*Medium Shot*	*Few Shot*	*Total Shot*	*Many Shot*	*Medium Shot*	*Few Shot*	*Total Shot*
MLC	BCE [45]	✔	42.57±0.11	56.67±0.19	46.40±0.60	48.92±0.23	67.37±0.18	88.27±0.39	83.79 ±0.41	78.79±0.14
	Focal [26]	✔	41.05±0.07	58.33±0.12	53.58±0.31	51.39±0.15	67.02±0.11	87.49±0.18	82.82±0.78	78.13±0.23
	ASL [3]	✔	41.60±0.17	58.15±0.15	52.67±0.17	51.20±0.08	67.67±0.10	87.79±0.13	82.23±0.55	78.35±0.11
LT-MLC	DB [37]	✔	44.83±0.31	58.96±0.24	53.82±0.47	52.16±0.36	69.22±0.28	88.56±0.42	83.72±0.35	78.86±0.23
	DB-Focal [37]	✔	45.76±0.25	59.74±0.21	53.85±0.16	52.57±0.27	68.96±0.22	88.89±0.18	83.42±0.20	78.90±0.26
	LWS [13]	-	44.86±0.58	58.79±0.63	53.48±0.51	52.86±0.60	69.08±0.44	88.24±0.55	83.46±0.47	78.28±0.49
PL-MLC	Pseudo-Label [15]	-	41.41±0.41	57.46±0.35	53.12±0.33	51.67±0.37	67.38±0.24	87.58±0.35	83.26±0.42	78.32±0.30
	ML-GCN [5]	✔	43.43±0.53	58.46±0.61	53.74±0.48	52.14±0.55	68.46±0.44	88.17±0.61	82.46±0.38	79.02±0.56
	Hill [44]	✔	42.50±0.16	56.89±0.19	47.31±0.37	49.28±0.09	68.79±0.15	86.70±0.17	78.15±0.99	77.40±0.22
	P-ASL [2]	✔	43.09±0.05	57.67±0.07	53.46±0.22	51.75±0.17	68.95±0.22	87.24±0.13	83.37±0.33	78.96±0.16
PLT-MLC	Head Model (Ours)	✔	47.59±0.09	59.07±0.12	52.35±0.28	53.30±0.19	72.91±0.28	88.59±0.31	82.12±0.27	80.70±0.30
	Tail Model (Ours)	✔	46.30±0.25	58.76±0.29	53.38±0.14	53.09±0.27	71.65±0.34	88.68±0.41	83.51±0.24	80.58±0.36
	COMIC (Ours)	✔	**49.21±0.22**	**60.08±0.13**	**55.36±0.21**	**55.08±0.14**	**73.10±0.35**	**89.18±0.45**	**84.53±0.48**	**81.53±0.35**
Improv. ↑	-	-	1.62 ∼ 8.16	0.34 ∼ 3.41	0.53 ∼ 8.96	1.78 ∼ 6.16	0.19 ∼ 6.08	0.29 ∼ 2.48	-0.3 ∼ 6.38	0.83 ∼ 4.13

Experiments

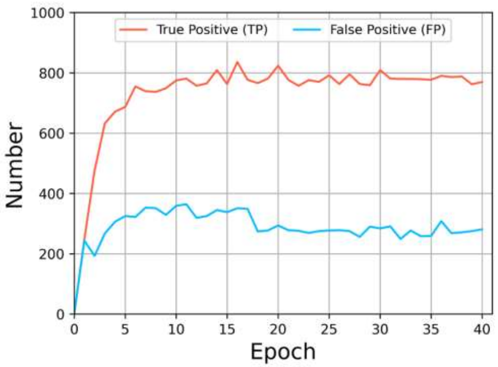
Missing Ratio	PLT-COCO Dataset			
	Total Shot	*Many Shot*	*Medium Shot*	*Low Shot*
0%	57.07±0.09	52.21±0.11	59.98±0.12	61.12±0.24
30%	55.80±0.17	49.97±0.11	62.59±0.15	54.56±0.17
40%	54.75±0.19	48.93±0.24	60.31±0.21	54.14±0.21
50%	54.69±0.15	48.74±0.12	56.68±0.16	57.25±0.24

Performance comparison under different missing labeled settings. 0% indicates an LT dataset that is fully labeled.

$$\mathcal{L}_{htb} = \kappa_h \cdot \mathcal{L}(\phi(\hat{\mathbf{z}_h}) \cdot \phi(\mathbf{z}_b)) + \kappa_t \cdot \mathcal{L}(\phi(\hat{\mathbf{z}}_t) \cdot \phi(\mathbf{z}_b))$$

$$\kappa_h = \frac{(\mathcal{L}(\hat{z_h}))^\alpha}{(\mathcal{L}(\hat{z}_t))^\alpha + (\mathcal{L}(\hat{z_h}))^\alpha} \text{ and } \kappa_t = \frac{(\mathcal{L}(\hat{z}_t))^\alpha}{(\mathcal{L}(\hat{z}_t))^\alpha + (\mathcal{L}(\hat{z_h}))^\alpha}$$

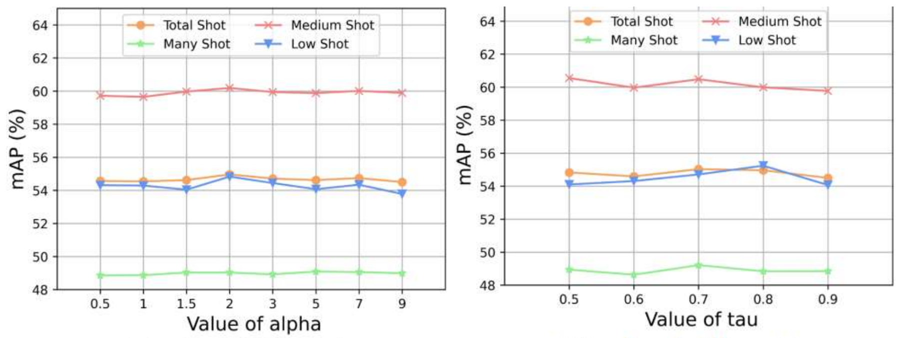
In-depth analysis of label correction.



(a) Number of TP and FP

$$\hat{y}_c = \begin{cases} 1, & \text{if } p_c > \max\{\tau, P_c\}, y_c = 0 \\ 0, & \text{otherwise} \end{cases}$$

Ablations with respect to coefficient α and τ .



(a) mAp of different α

(b) mAp of different tau

Experiments

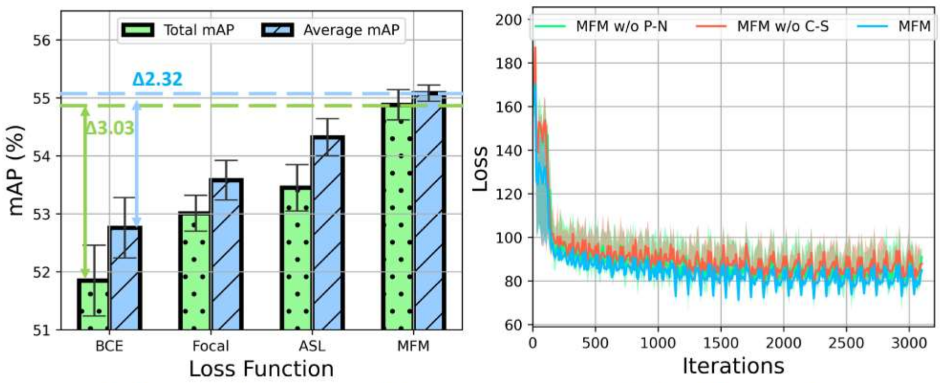
Ablation study of different modules.
M,C,B represent modification, correction and balance learning

Models	Setting			PLT-COCO Dataset		
	M	C	B	*Total mAP*	*Average mAP*	*Recall*
-RLC		✓	✓	54.70±0.13	54.42±0.15	85.26±0.08
-MFM	✓		✓	54.60±0.13	54.33±0.13	84.59±0.19
-HTB	✓	✓		53.65±0.31	53.36±0.31	84.19±0.23
COMIC	✓	✓	✓	**55.08±0.14**	**54.88±0.19**	**88.19±0.22**

Ablation of MFM. ↓ indicates the mAP decay

MFM Factor		PLT-COCO Dataset			
P-N	H-T	*Total Shot*	*Many Shot*	*Medium Shot*	*Low Shot*
	✓	54.44 (↓ 0.64)	48.65 (↓ 0.56)	60.00 (↓ 0.08)	53.81 (↓ 1.55)
✓		53.70 (↓ 1.38)	48.38 (↓ 0.83)	58.99 (↓ 1.09)	52.91 (↓ 2.45)
✓	✓	55.08	49.21	60.08	55.36

MLT-MLC results using different losses



(a) Performance Comparison of Losses **(b) Loss Visualization**

Thanks