Multiple Instance Active Learning for Object Detection

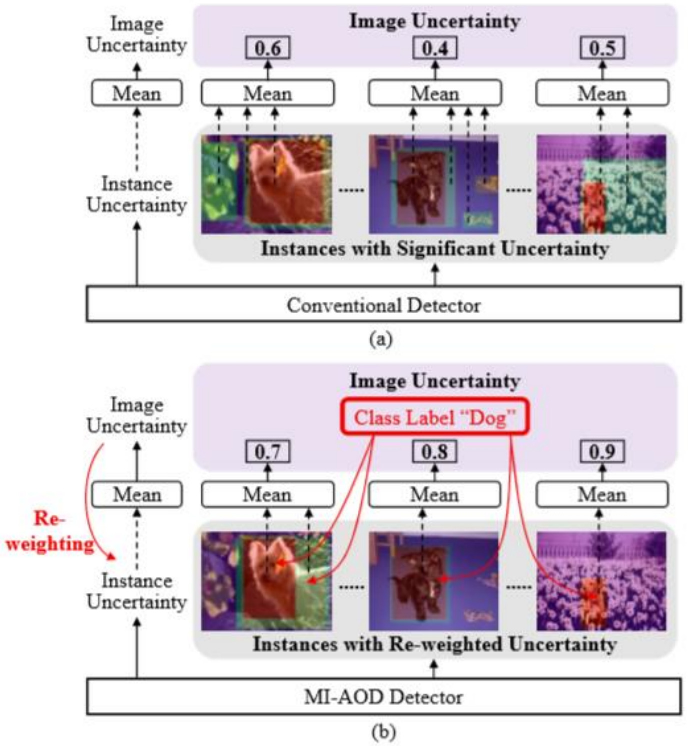
Tianning Yuan, Fang Wan, Mengying Fu, Jianzhuang Liu,
Songcen Xu, Xiangyang Ji and Qixiang Ye

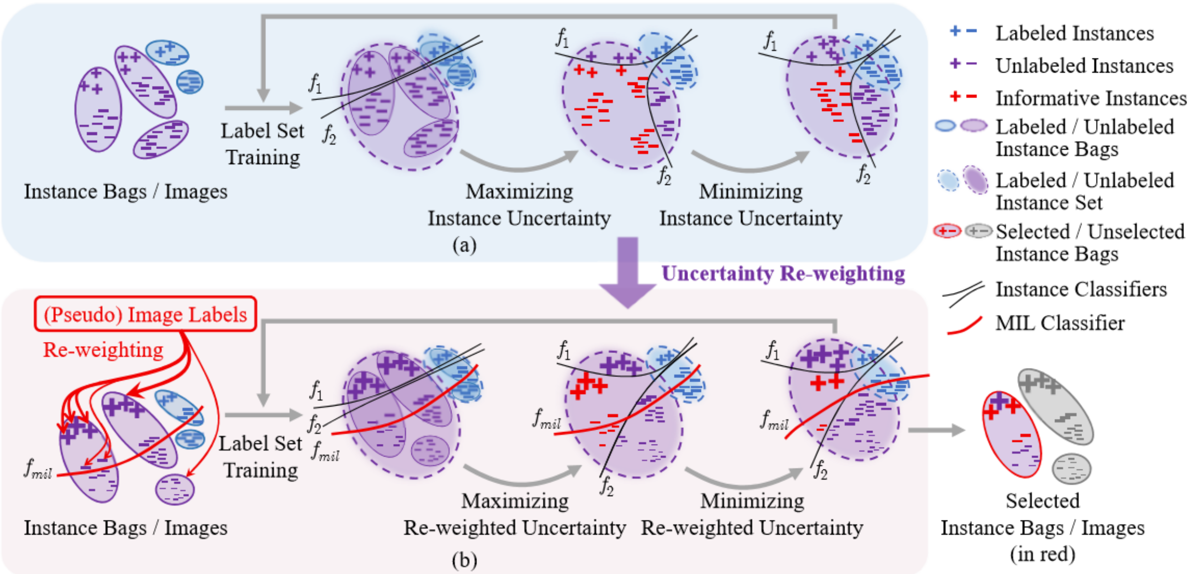
Outline

- Contribution
  - ➢ propose Multiple Instance Active Object Detection (MI-AOD)
  - ➢ design instance uncertainty learning (IUL) and instance uncertainty re-weighting (IUR) modules.
  - ➢ apply MI-AOD to object detection on commonly used datasets.
- Experiment Performance
- Ablation Study
- Model Analysis

MI-AOD

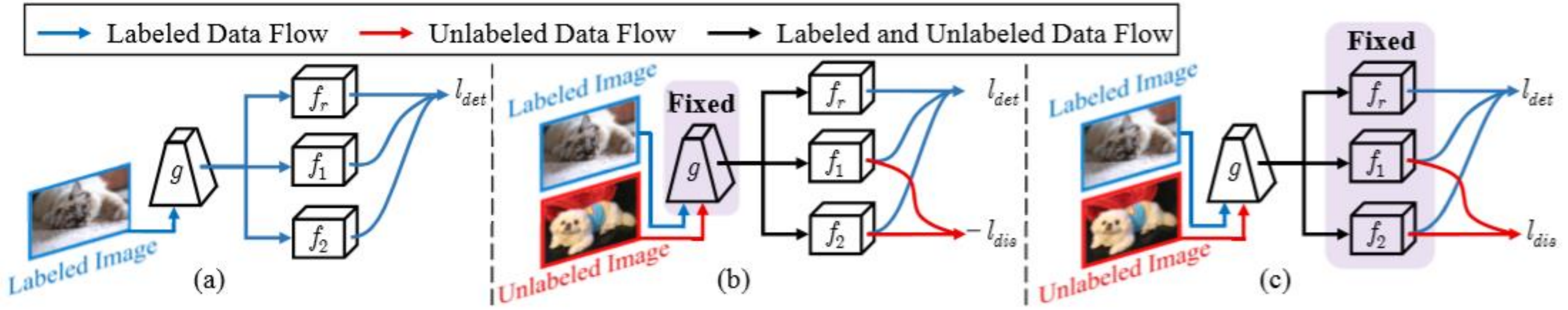
How to evaluate the uncertainty of the unlabeled instances using the detectortrained on the labeled set. → Instance Uncertainty Learning，IUL





how to precisely estimate the image uncertainty while filtering out noisy instances. → Instance Uncertainty Re-weighting，IUR

Instance Uncertainty Learning



g denote the feature extractor parameterized by θg

$$\Theta = \{\theta_{f_1}, \theta_{f_2}, \theta_{f_r}, \theta_g\}$$

*Using the RetinaNet as the baseline construct a detector

(a) Label Set Training.

Focal loss

the ground-truth class label

$$\arg\min_{\Theta} l_{det}(x) = \sum_i \Big(\mathrm{FL}(\hat{y}_i^{f_1}, y_i^{cls}) + \mathrm{FL}(\hat{y}_i^{f_2}, y_i^{cls}) + \mathrm{SmoothL1}(\hat{y}_i^{f_r}, y_i^{loc}) \Big), \quad (1)$$

Smooth L1 loss

bounding box label

$$\hat{y}_i^{f_1} = f_1(g(x_i))$$
$$\hat{y}_i^{f_2} = f_2(g(x_i))$$
$$\hat{y}_i^{f_r} = f_r(g(x_i))$$

(b) Maximizing Instance Uncertainty

$$\arg\min_{\Theta \backslash \theta_g} \mathcal{L}_{max} = \sum_{x \in \mathcal{X}_L} l_{det}(x) - \sum_{x \in \mathcal{X}_U} \lambda \cdot l_{dis}(x),$$

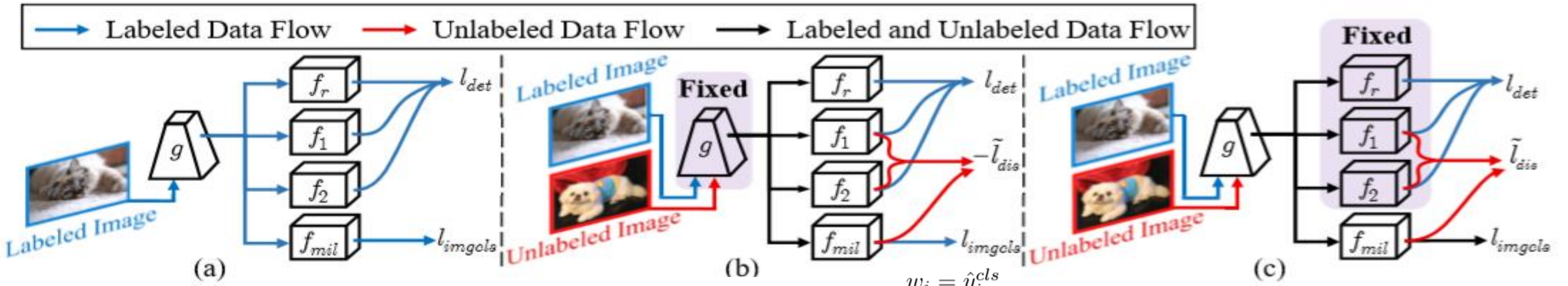
prediction discrepancy loss

$$l_{dis}(x) = \sum_i |\hat{y}_i^{f_1} - \hat{y}_i^{f_2}|$$

(c) Minimizing Instance Uncertainty.

$$\arg\min_{\theta_g} \mathcal{L}_{min} = \sum_{x \in \mathcal{X}_L} l_{det}(x) + \sum_{x \in \mathcal{X}_U} \lambda \cdot l_{dis}(x).$$

Instance Uncertainty Re-weighting



$$w_i = \hat{y}_i^{cls}$$

$$\tilde{l}_{dis}(x) = \sum_i |w_i \cdot (\hat{y}_i^{f_1} - \hat{y}_i^{f_2})|,$$

$$\hat{y}^{f_{mil}} = f_{mil}(g(x)) \rightarrow N \times C \text{ score matrix.}$$

$$\hat{y}_{i,c}^{cls} = \frac{\exp(\hat{y}_{i,c}^{f_{mil}})}{\sum_c \exp(\hat{y}_{i,c}^{f_{mil}})} \cdot \frac{\exp\left((\hat{y}_{i,c}^{f_1} + \hat{y}_{i,c}^{f_2})/2\right)}{\sum_i \exp\left((\hat{y}_{i,c}^{f_1} + \hat{y}_{i,c}^{f_2})/2\right)},$$

$$y_c^{cls} \in \{0,1\}$$

$$l_{imgcls}(x) = -\sum_c \left(y_c^{cls} \log \sum_i \hat{y}_{i,c}^{cls} + (1 - y_c^{cls}) \log(1 - \sum_i \hat{y}_{i,c}^{cls}) \right),$$

image classification loss

Maximizing Instance Uncertainty

$$\operatorname*{argmin}_{\Theta \backslash \theta_g} \mathcal{L}_{max} = \sum_{x \in \mathcal{X}_L} l_{det}(x) - \sum_{x \in \mathcal{X}_U} \lambda \cdot l_{dis}(x), \rightarrow \operatorname*{argmin}_{\tilde{\Theta} \backslash \theta_g} \tilde{\mathcal{L}}_{max} = \sum_{x \in \mathcal{X}_L} \left(l_{det}(x) + l_{imgcls}(x) \right) - \sum_{x \in \mathcal{X}_U} \lambda \cdot \tilde{l}_{dis}(x),$$

Minimizing Instance Uncertainty.

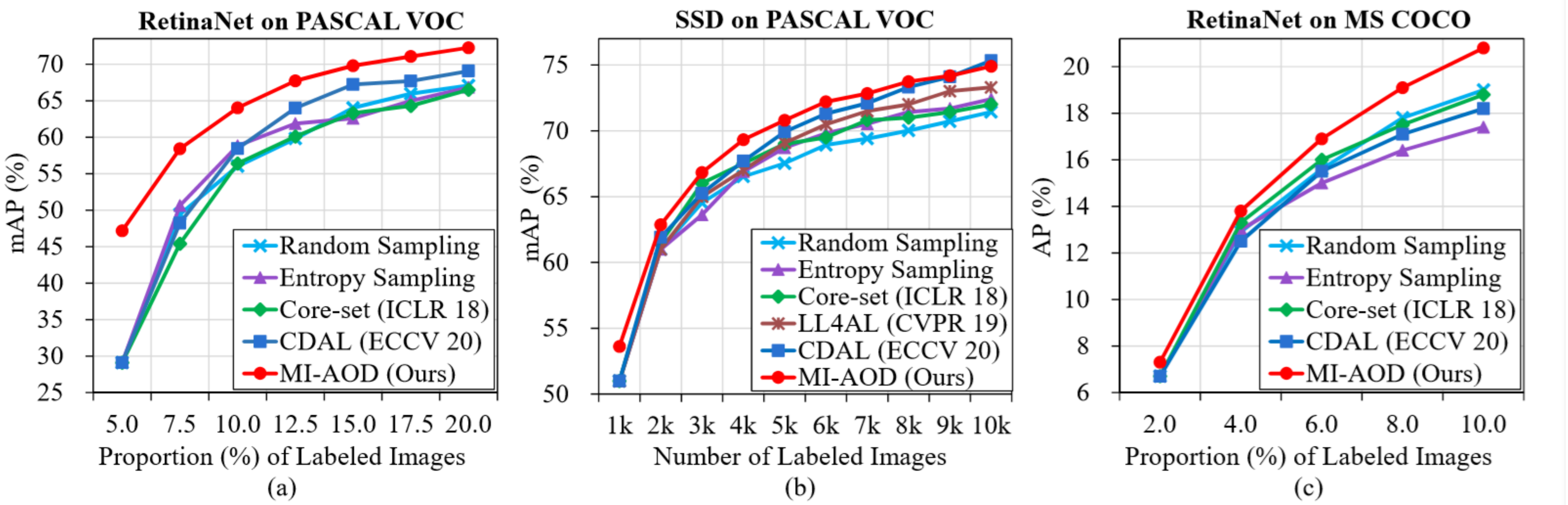
$$\operatorname*{argmin}_{\theta_g} \mathcal{L}_{min} = \sum_{x \in \mathcal{X}_L} l_{det}(x) + \sum_{x \in \mathcal{X}_U} \lambda \cdot l_{dis}(x). \rightarrow \operatorname*{argmin}_{\theta_g, \theta_{f_{mil}}} \tilde{\mathcal{L}}_{min} = \sum_{x \in \mathcal{X}_L} \left(l_{det}(x) + l_{imgcls}(x) \right) + \sum_{x \in \mathcal{X}_U} \left(\lambda \cdot \tilde{l}_{dis}(x) + l_{imgcls}(x) \right).$$

$$\boxed{y_c^{pseudo} = \mathbb{1}\left(\max_i \left(\frac{\hat{y}_{i,c}^{f_1} + \hat{y}_{i,c}^{f_2}}{2} \right), 0.5 \right),}$$

where $\mathbb{1}(a, b)$ is a binarization function. When a > b, it returns 1; otherwise 0.

Experiment

Performance



Ablation Study

Training		Sample Selection			mAP (%) on Proportion (%) of Labeled Images							
IUL	IUR	Rand.	Max Unc.	Mean Unc.	5.0	7.5	10.0	12.5	15.0	17.5	20.0	100.0
		✓			28.31	49.42	56.03	59.81	64.02	65.95	67.09	77.28
✓		✓			30.09	49.17	55.64	60.93	64.10	65.77	67.20	
✓			✓		30.09	49.79	58.94	63.11	65.61	67.84	69.01	
✓				✓	30.09	49.74	60.60	64.29	67.13	68.76	70.06	
	✓	✓			47.18	57.12	60.68	63.72	66.10	67.59	68.48	**78.37**
	✓		✓		47.18	57.58	61.74	64.58	66.98	68.79	70.33	
	✓			✓	**47.18**	**58.03**	**63.98**	**66.58**	**69.57**	**70.96**	**72.03**	

Training	mAP (%) on Proportion (%) of Labeled Imgs.				
IUL	2.0	4.0	6.0	8.0	10.0
	51.01	61.48	69.14	75.14	79.77
✓	**58.07**	**67.75**	**74.91**	**78.88**	**80.96**

w_i	Set	mAP (%) on Proportion (%) of Labeled Imgs.						
		5.0	7.5	10.0	12.5	15.0	17.5	20.0
1	$\varnothing$	30.09	49.17	55.64	60.93	64.10	65.77	67.20
$\hat{y}_i^{f_1}$	$\varnothing$	31.67	50.67	55.93	60.78	64.17	66.22	67.30
1	$\mathcal{X}_L$	42.52	54.08	57.18	63.43	65.04	66.74	68.32
$\hat{y}_i^{cls}$	$\mathcal{X}$	**47.18**	**57.12**	**60.68**	**63.72**	**66.10**	**67.59**	**68.48**

Ablation Study

λ	k	mAP (%) on Proportion (%) of Labeled Imgs.						
		5.0	7.5	10.0	12.5	15.0	17.5	20.0
2	10k	47.18	56.94	64.44	67.70	69.58	70.67	72.12
1	10k	47.18	57.30	**64.93**	67.40	69.63	70.53	71.62
0.5	10k	47.18	**58.41**	64.02	**67.72**	**69.79**	**71.07**	**72.27**
0.2	10k	47.18	58.02	64.44	67.67	69.42	70.98	72.06
0.5	N	47.18	58.03	63.98	66.58	69.57	70.96	72.03
0.5	10k	47.18	58.41	**64.02**	**67.72**	**69.79**	**71.07**	**72.27**
0.5	100	47.18	**58.74**	63.62	67.03	68.63	70.26	71.47
0.5	1	47.18	57.58	61.74	64.58	66.98	68.79	70.33

$$\underset{\Theta\backslash\theta_g}{\operatorname{argmin}}\,\mathcal{L}_{max} = \sum_{x\in\mathcal{X}_L} l_{det}(x) - \sum_{x\in\mathcal{X}_U} \lambda\cdot l_{dis}(x), \quad (2)$$

$$\underset{\theta_g}{\operatorname{argmin}}\,\mathcal{L}_{min} = \sum_{x\in\mathcal{X}_L} l_{det}(x) + \sum_{x\in\mathcal{X}_U} \lambda\cdot l_{dis}(x). \quad (4)$$

$$\underset{\tilde{\Theta}\backslash\theta_g}{\operatorname{argmin}}\,\tilde{\mathcal{L}}_{max} = \sum_{x\in\mathcal{X}_L} \Big(l_{det}(x)+l_{imgcls}(x)\Big) - \sum_{x\in\mathcal{X}_U} \lambda\cdot\tilde{l}_{dis}(x), \quad (8)$$

$$\underset{\theta_g,\theta_{f_{mil}}}{\operatorname{argmin}}\,\tilde{\mathcal{L}}_{min} = \sum_{x\in\mathcal{X}_L} \Big(l_{det}(x) + l_{imgcls}(x)\Big) + \sum_{x\in\mathcal{X}_U} \Big(\lambda\cdot\tilde{l}_{dis}(x) + l_{imgcls}(x)\Big). \quad (9)$$

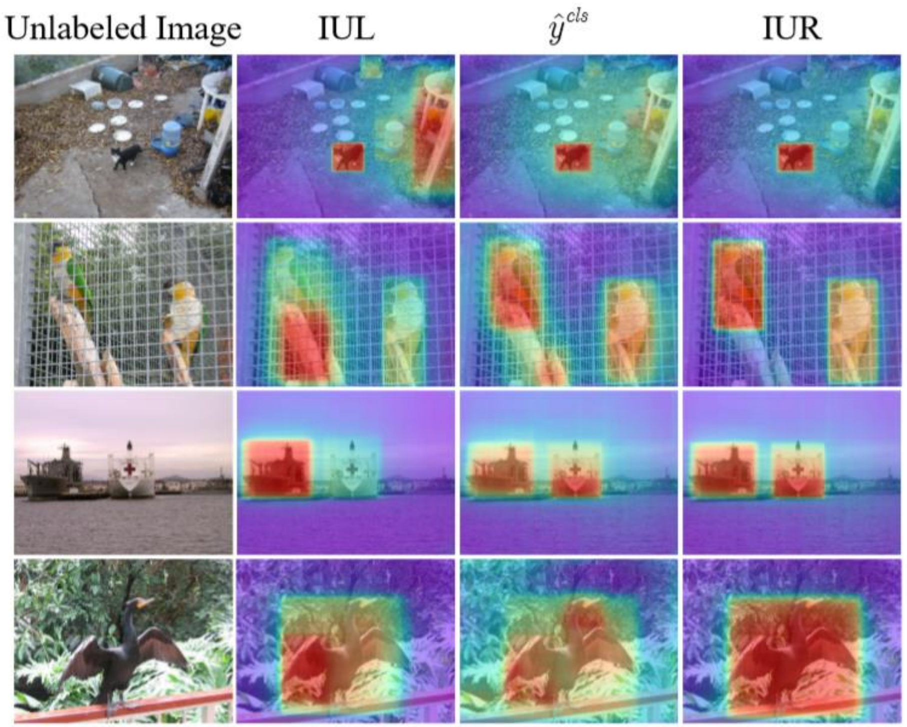
MI-AOD has the best performance when λ is set to 0.5 and k is set to 10k (for ~100k instances/anchorsineachimage).

Method	Time (h) on Proportion (%) of Labeled Imgs.						
	5.0	7.5	10.0	12.5	15.0	17.5	20.0
Random	0.77	1.12	1.45	1.78	2.12	2.45	2.78
CDAL [1]	1.18	1.50	1.87	2.19	2.68	**2.83**	**2.82**
MI-AOD	**1.03**	**1.42**	**1.78**	**2.18**	**2.55**	2.93	3.12

MI-AOD costs less time at early cycles than CDAL.

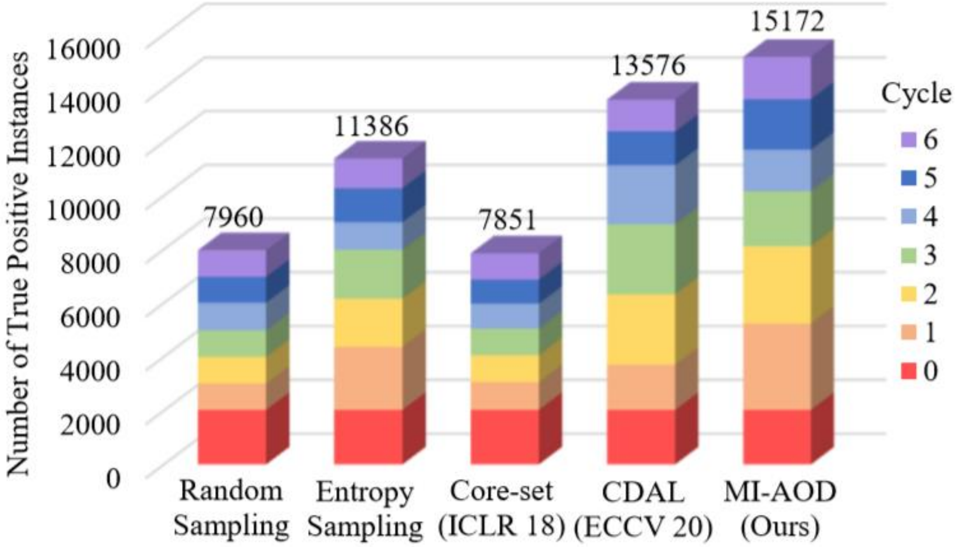
Model Analysis

Visualization Analysis.



IUR leverages the image classification scores to re-weight instances towards accurate instance uncertainty prediction.

Statistical Analysis.



MI-AOD approach can activate true positive objects better while filtering out interfering instances