OpenAUC: Towards AUC-Oriented Open-Set Recognition

Zitai Wang, Qianqian Xu*, Zhiyong Yang,

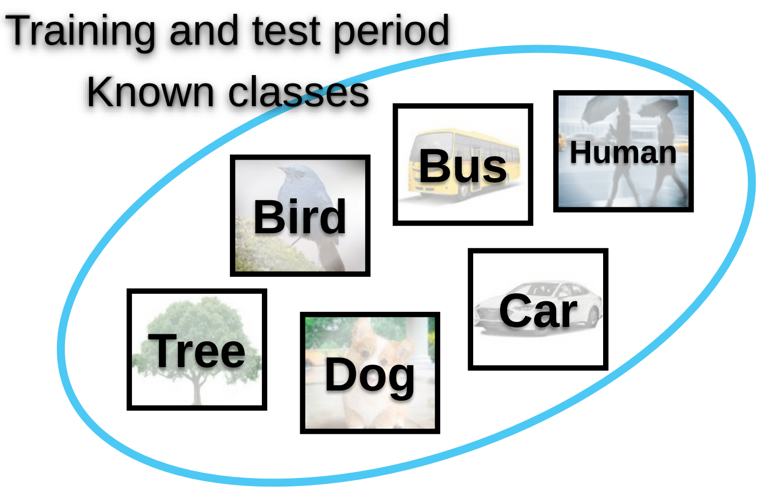
Yuan He, Xiaochun Cao, Qingming Huang*

Zitai Wang

2022.10

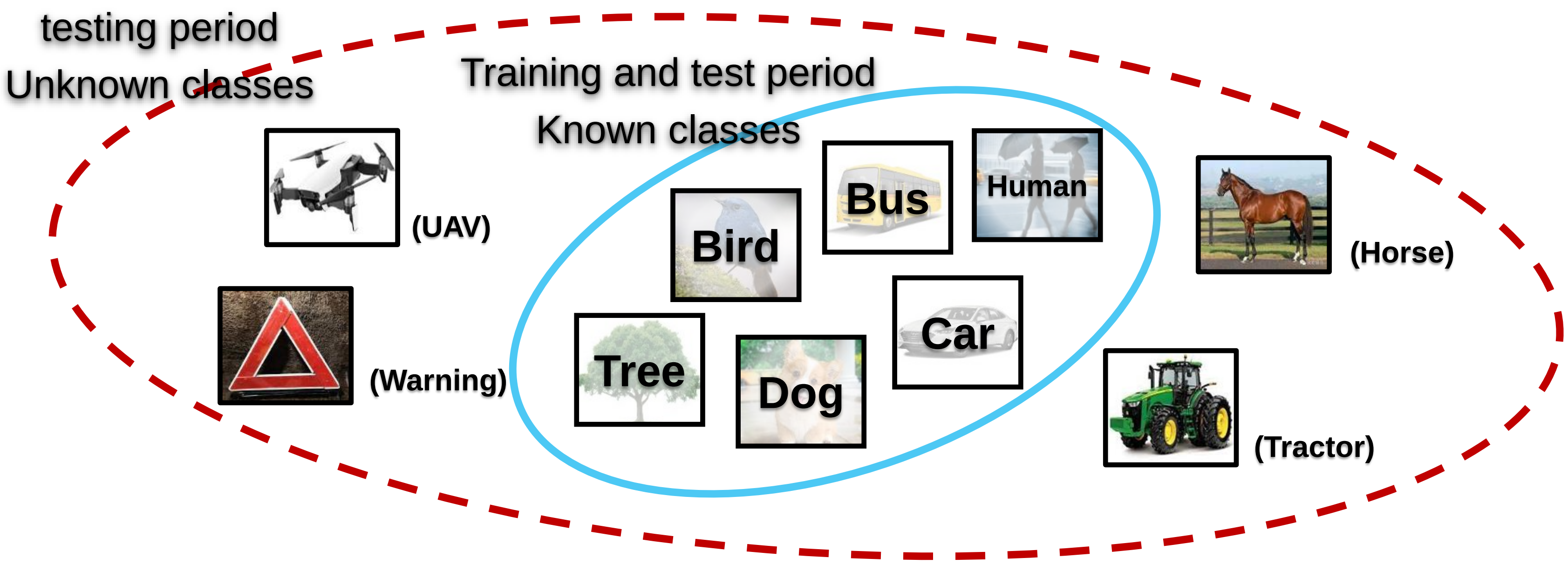
Open-set Recognition (OSR)

- Traditional machine learning implicitly follows a close-set assumption



Open-set Recognition (OSR)

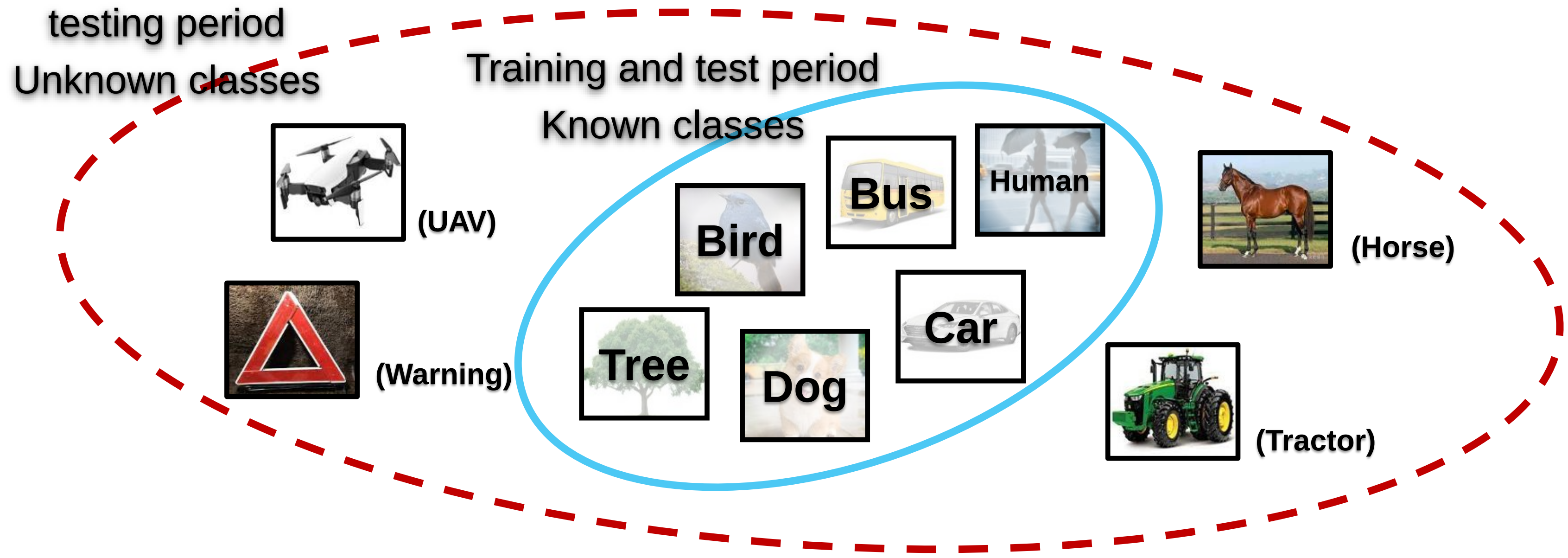
- In many practical scenarios, some test samples inevitably belong to none of the known classes



Close-set models will classify novel samples into known classes!

Open-set Recognition (OSR)

- In many practical scenarios, some test samples inevitably belong to none of the known classes



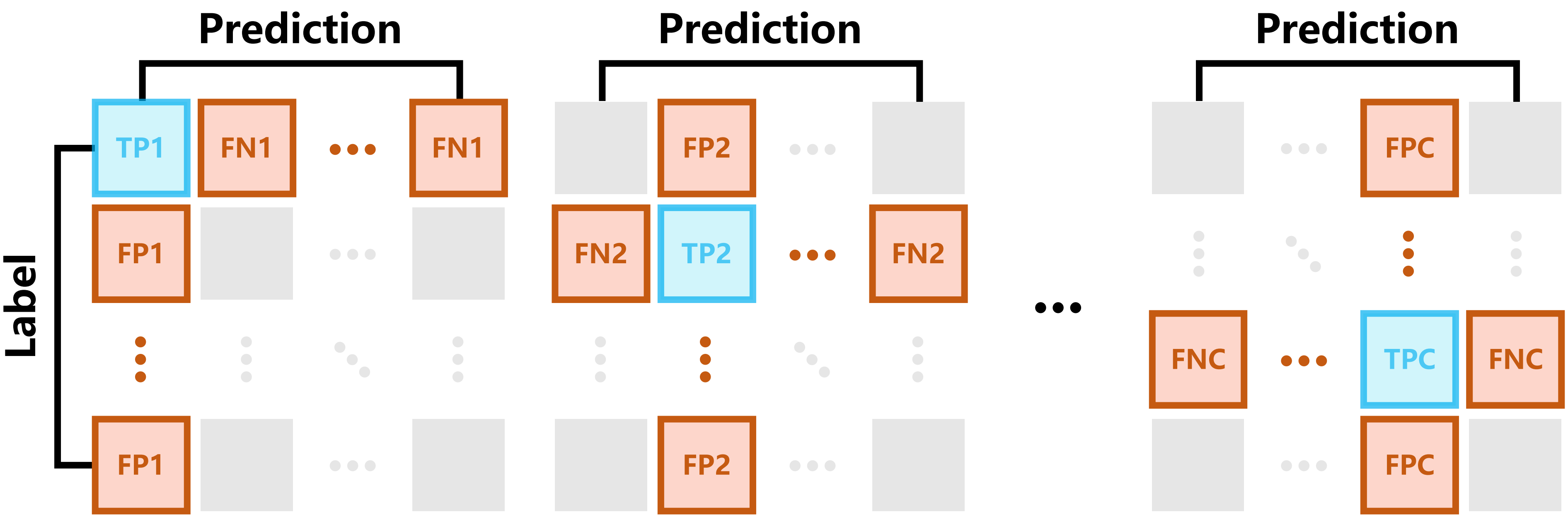
Goal 1
- ✓ Correctly classify close-set samples

Goal 2
- ✓ Discriminate open-set samples from close-set ones

How to evaluate model performance in this complicated setting?

Traditional metrics for OSR

☐ Classification-based ones



TP_i / (TP_i + FP_i), i ∈ close-set

F-score ∝ **Precision**, **Recall**

TP_i / (TP_i + FN_i), i ∈ close-set

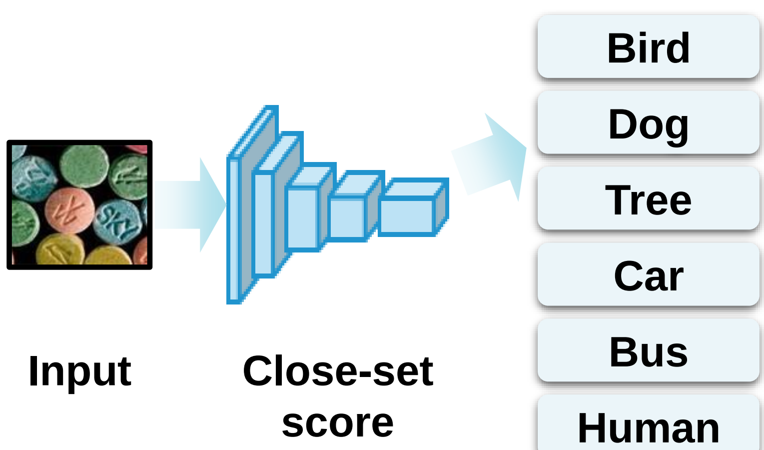
Traditional metrics for OSR

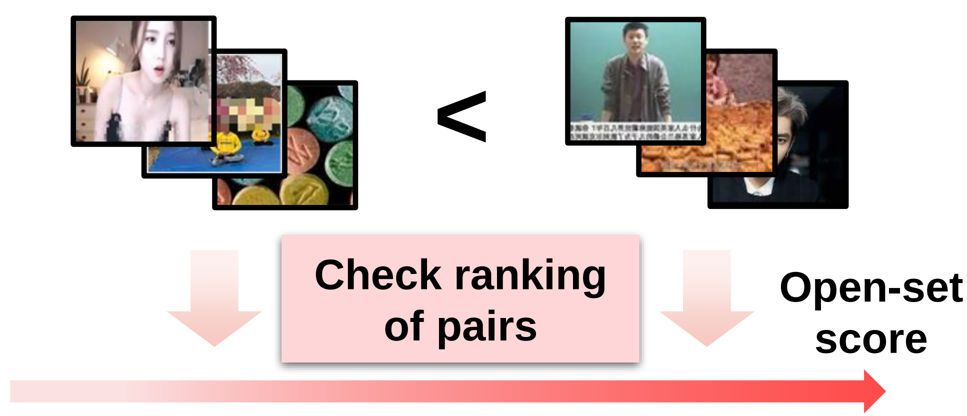
◻ Novelty-detection-based ones

Goal 1
✓ Correctly classify close-set samples

Goal 2
✓ Discriminate open-set samples from close-set ones

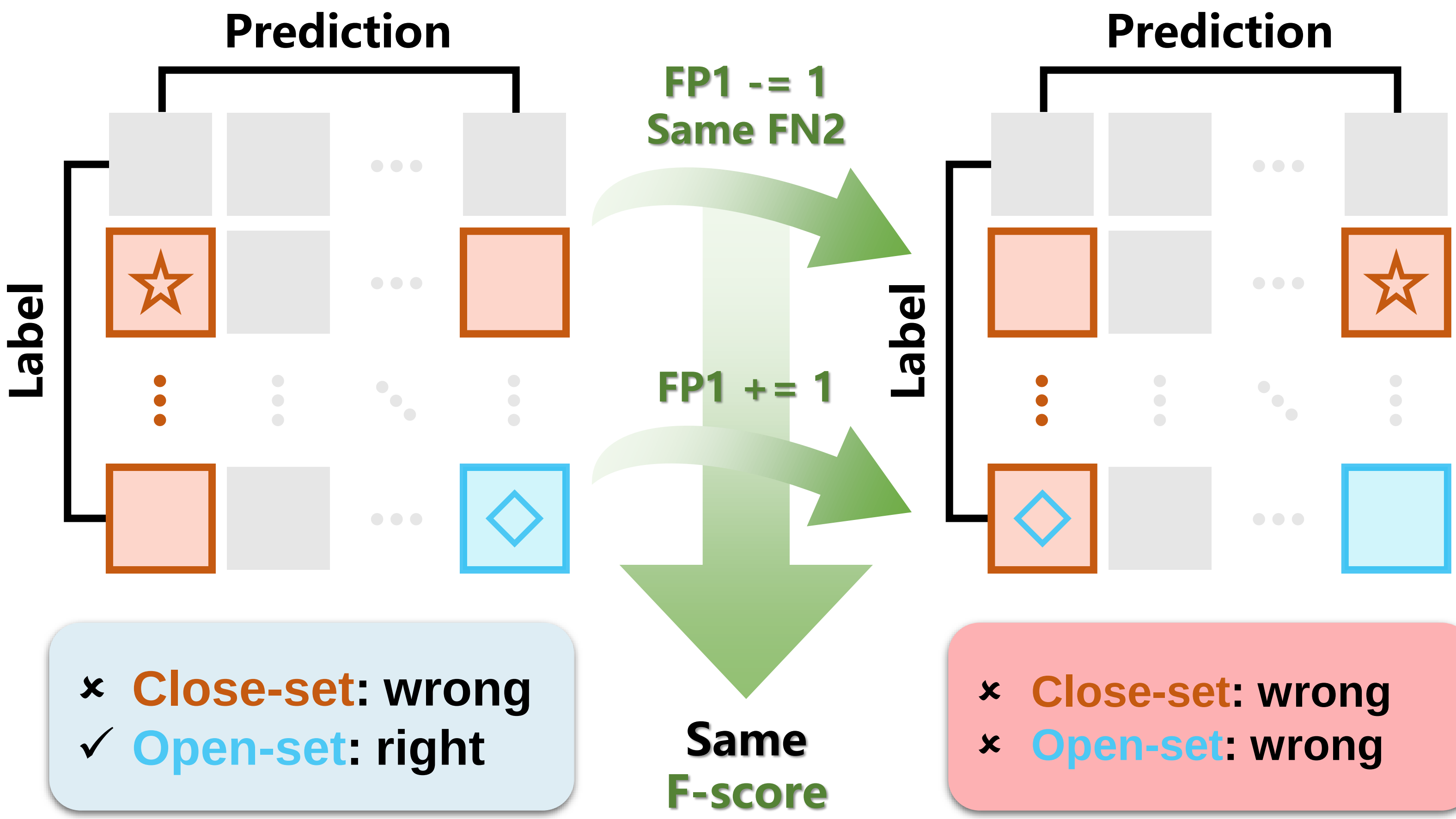
Multiclass classification (Accuracy) + **Novelty detection (AUROC)**





Limitation of traditional metrics

◻ Inconsistency property of F-score



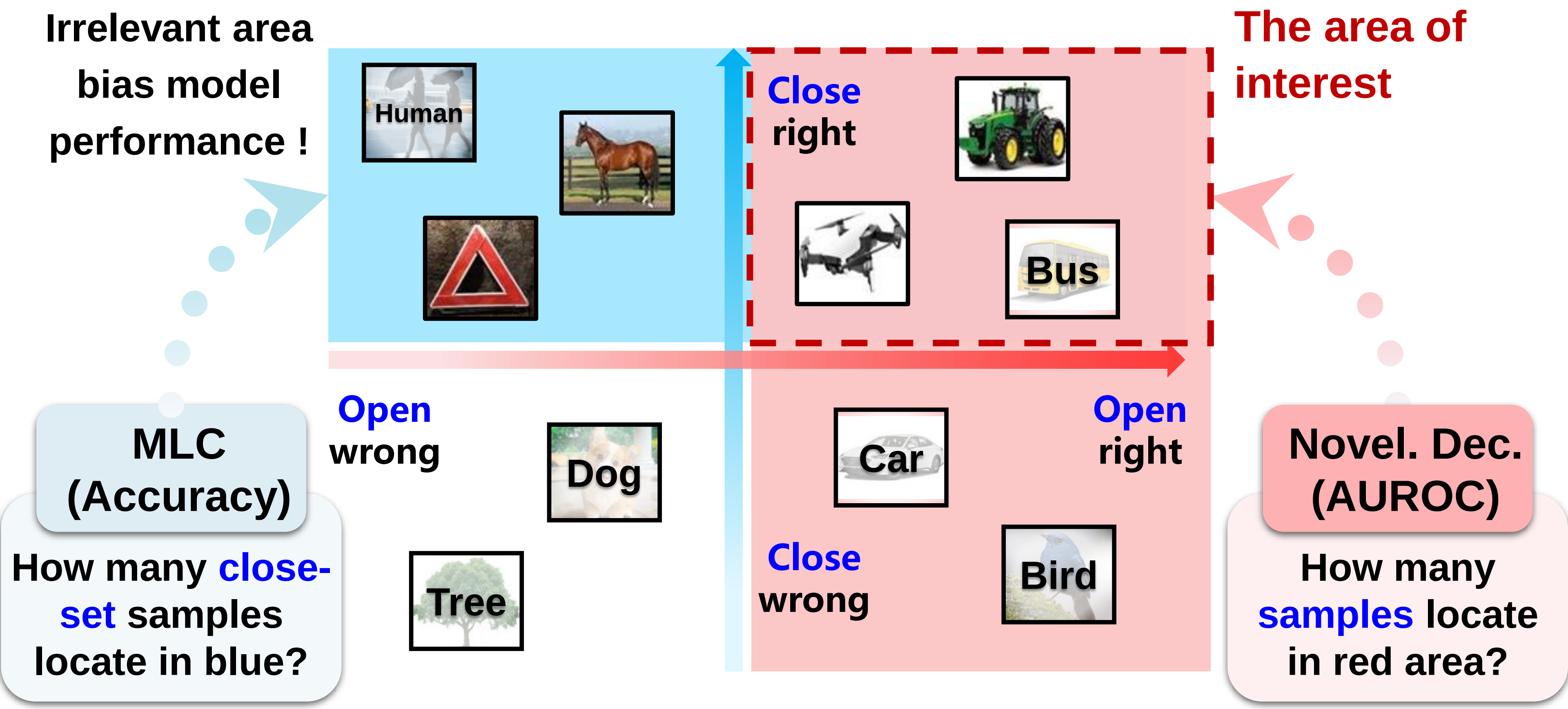
× Close-set: wrong
✓ Open-set: right

× Close-set: wrong
× Open-set: wrong

Metric value is inconsistent with Model performance!

Limitation of traditional metrics

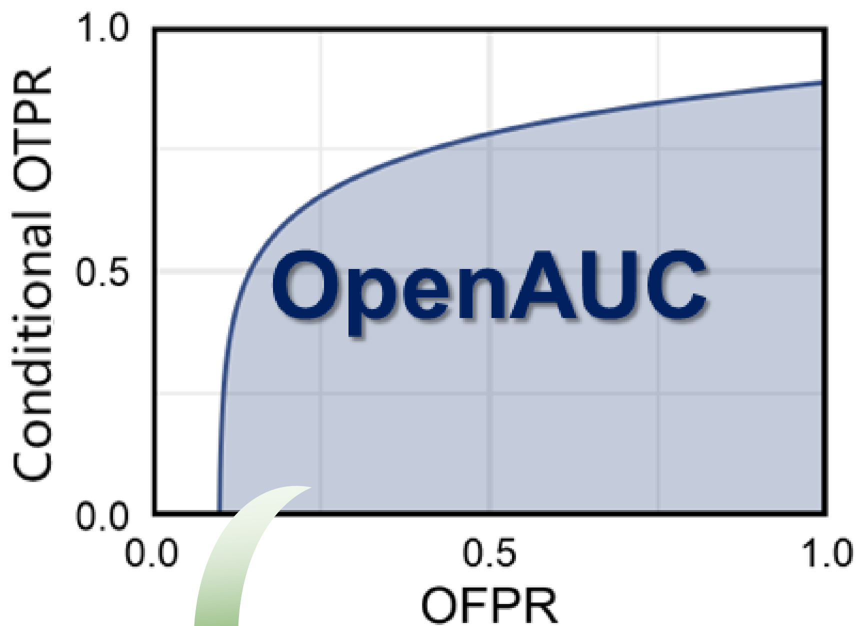
☐ Inconsistency property of close-set Acc and AUROC



Local metrics are inconsistent with global performance, 1 + 1 ≠ 2 !

The definition of OpenAUC

□ Aggregating close-set and open-set performances under different thresholds



OFPR:
Pr(Open-set sample is wrong)

- ✗ Open-set score less than t

Conditional OTPR:
Pr(Close-set sample is right)

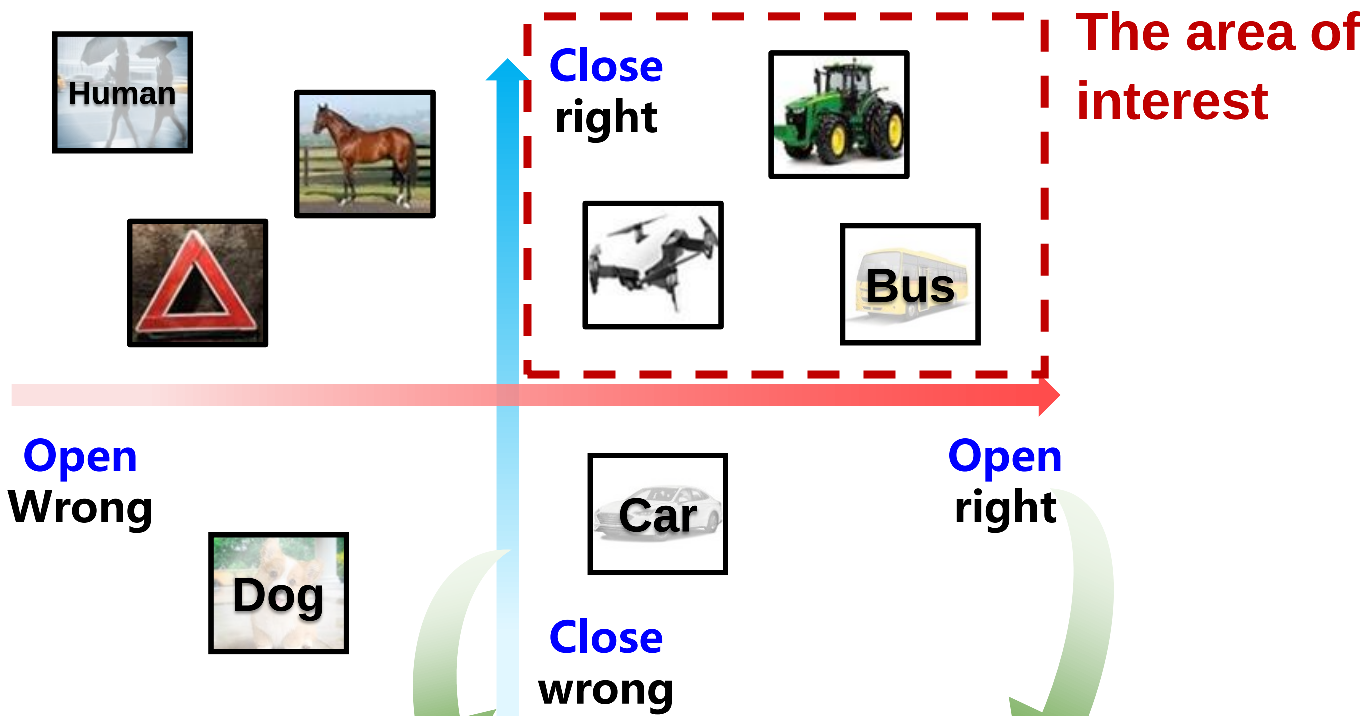
- ✓ Open-set score less than t
- ✓ Correct prediction on close-set

$$\mathtt{OpenAUC} := \int_{-\infty}^{+\infty} \mathtt{COTPR}(\mathtt{OFPR}^{-1}(t))\, dt.$$

The integral formulation is hard to calculate

The definition of OpenAUC

◻ A concise pairwise formulation



$$\mathit{OpenAUC} = \mathop{\mathbb{E}}_{\substack{\boldsymbol{z}_k \sim D_k \\ \boldsymbol{z}_u \sim D_u}} \Big[\underbrace{\mathbf{1}\left[y_k = h(\boldsymbol{x}_k)\right]}_{(a)} \cdot \underbrace{\mathbf{1}\left[r(\boldsymbol{x}_u) > r(\boldsymbol{x}_k)\right]}_{(b)} \Big].$$

OpenAUC considers the samples located in the area of interest !

The advantage of OpenAUC

□ Further theoretical results

Proposition 5. *Given a sample pair $((\boldsymbol{x}_1, C+1), (\boldsymbol{x}_2, y_2))$, where $y_2 \neq C+1$, for any (h, r) such that $R(\boldsymbol{x}_1) = 1, R(\boldsymbol{x}_2) = 0, h(\boldsymbol{x}_2) \neq y_2$, if $(\tilde{h}, \tilde{r})$ makes the same predictions as (h, r) expect that $\tilde{R}(\boldsymbol{x}_1) = 0, \tilde{h}(\boldsymbol{x}_1) = h(\boldsymbol{x}_2)$ and $\tilde{R}(\boldsymbol{x}_2) = 1$, we have $\texttt{OpenAUC}(\tilde{h}, \tilde{r}) < \texttt{OpenAUC}(h, r)$.*

OpenAUC is consistent when the inconstant property of F-score happens

Proposition 6. *Given a dataset $\mathcal{S}$, for any (f, r) such that $\texttt{OpenAUC} = k$ and any threshold t_{C+1} such that $\texttt{FPR}_{C+1} = a \neq 0$, we have $\texttt{TPR}_{C+1} \geq 1 - (1-k)/a$.*

Optimizing OpenAUC increases the lower bound of open-set performance

Proposition 7. *Given two close-set samples $(\boldsymbol{x}_1, y_1)$ and $(\boldsymbol{x}_2, y_2)$ and an open-set sample $(\boldsymbol{x}_3, C+1)$, if $(\tilde{h}, \tilde{r})$ makes the same predictions as (h, r) expect that $h(\boldsymbol{x}_1) = \tilde{h}(\boldsymbol{x}_1) = y_1, h(\boldsymbol{x}_2), \tilde{h}(\boldsymbol{x}_2) \neq y_2, r(\boldsymbol{x}_2) > r(\boldsymbol{x}_3) > r(\boldsymbol{x}_1), r(\boldsymbol{x}_3) = \tilde{r}(\boldsymbol{x}_3)$ and $\tilde{r}(\boldsymbol{x}_2) = r(\boldsymbol{x}_1), \tilde{r}(\boldsymbol{x}_1) = r(\boldsymbol{x}_2)$, we have $\texttt{OpenAUC}(\tilde{h}, \tilde{r}) < \texttt{OpenAUC}(h, r)$.*

OpenAUC is consistent when the inconstant property of AUROC happens

OpenAUC overcomes the limitations of traditional metrics

OpenAUC optimization

□ Empirical minimization objective

Goal 1

✓ Correctly classify close-set samples

Goal 2

✓ Discriminate open-set samples from close-set ones

$$\hat{\mathcal{R}}_{L,\ell}(f,r) := \frac{1}{N_k}\sum_{i=1}^{N_k} L(h(\boldsymbol{x}_i), y_i) + \frac{1}{N_k N_u}\sum_{i=i}^{N_k}\sum_{j=1}^{N_u}\left[\mathbf{1}\left[y_i = h(\boldsymbol{x}_i)\right]\cdot \ell(r(\boldsymbol{x}_j) - r(\boldsymbol{x}_i))\right],$$

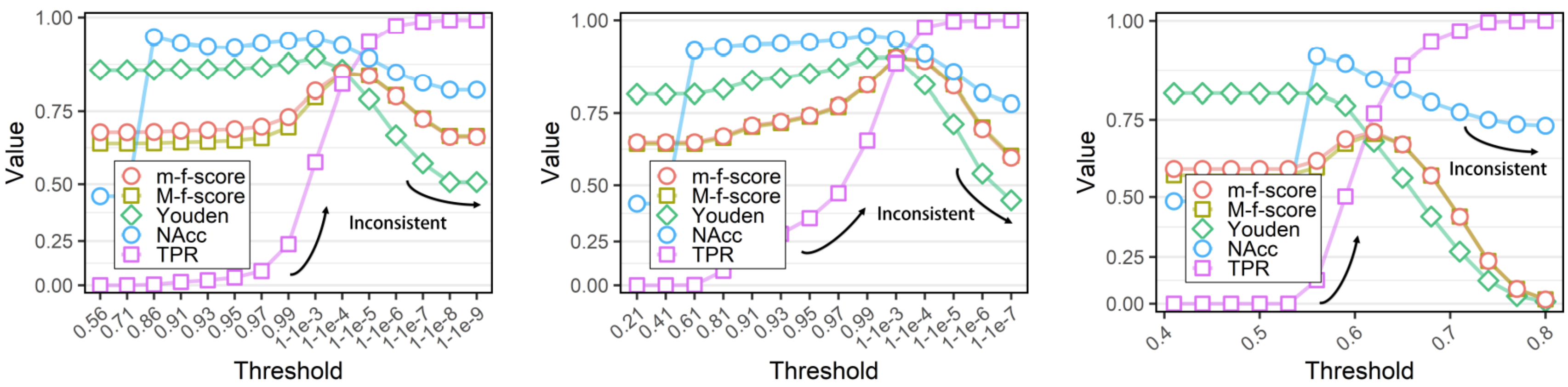
- Common MLC loss function such as CE

- Optimize the AUC loss only if close-set sample has been correctly classified

Optimizing OpenAUC is consistent with the decision process

Empirical results

□ Inconsistency property of classification-based metrics



(a) CIFAR+10 (b) CIFAR+50 (c) TinyImageNet

Figure 1: The inconsistency property of classification-based metrics. We can find that all these metrics decrease rapidly as the TPR performance of unknown classes increases.

Empirical results

Optimizing OpenAUC help boost model performances

Table 4: Empirical results on CUB, where E/M/H corresponds to the results on the Easy/Medium/Hard split of open-set samples. The best and the runner-up method on each metric are marked with red and blue, respectively.

	Close-set Accuracy	AUC (E/M/H)	OpenAUC (E/M/H)
Softmax	78.1	79.7 / 73.8 / 66.9	67.2 / 63.0 / 57.8
GCPL [31]	82.5	85.0 / 78.7 / 73.4	74.7 / 70.3 / 66.7
RPL [26]	82.6	85.5 / 78.1 / 69.6	74.5 / 69.0 / 62.4
ARPL [13]	82.1	85.4 / 78.0 / 70.0	74.4 / 68.9 / 62.7
CE+ [14]	86.2	88.3 / 82.3 / 76.3	79.8 / 75.4 / 70.8
ARPL+ [14]	85.9	83.5 / 78.9 / 72.1	76.0 / 72.4 / 66.8
Ours	86.2	88.8 / 83.2 / 78.1	80.2 / 76.1 / 72.5

	Error@95%TPR (E/M/H)	macro F-score (E/M/H)	micro F-score (E/M/H)
Softmax	46.6 / 55.9 / 62.8	67.4 / 66.5 / 66.6	69.0 / 68.9 / 70.8
GCPL [31]	37.0 / 46.8 / 51.3	77.6 / 75.4 / 74.0	78.4 / 76.8 / 77.4
RPL [26]	39.5 / 53.5 / 64.0	75.4 / 73.3 / 72.4	76.7 / 75.2 / 76.6
ARPL [13]	37.6 / 49.9 / 62.7	75.3 / 73.1 / 72.2	76.6 / 75.0 / 76.5
CE+ [14]	28.4 / 42.1 / 52.3	82.6 / 80.3 / 78.3	83.3 / 81.6 / 81.4
ARPL+ [14]	48.7 / 60.6 / 67.8	80.8 / 79.0 / 77.3	81.7 / 80.4 / 80.4
Ours	28.1 / 39.7 / 47.6	82.2 / 79.7 / 78.1	83.0 / 81.2 / 81.1

Thanks for your listening!