模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

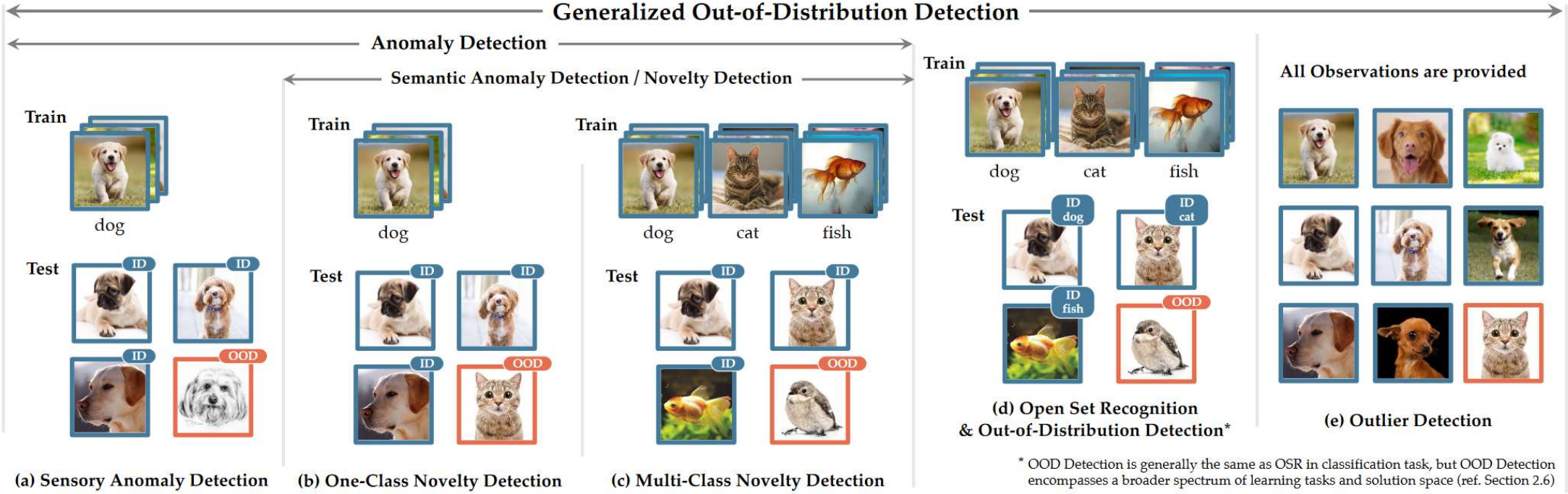
ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

YolOOD: Utilizing Object Detection Concepts for Multi-Label Out-of-Distribution Detection

Alon Zolfi[1], Guy Amit[1], Amit Baras[1], Satoru Koda[2], Ikuya Morikawa[2], Yuval Elovici[1], Asaf Shabtai[1]

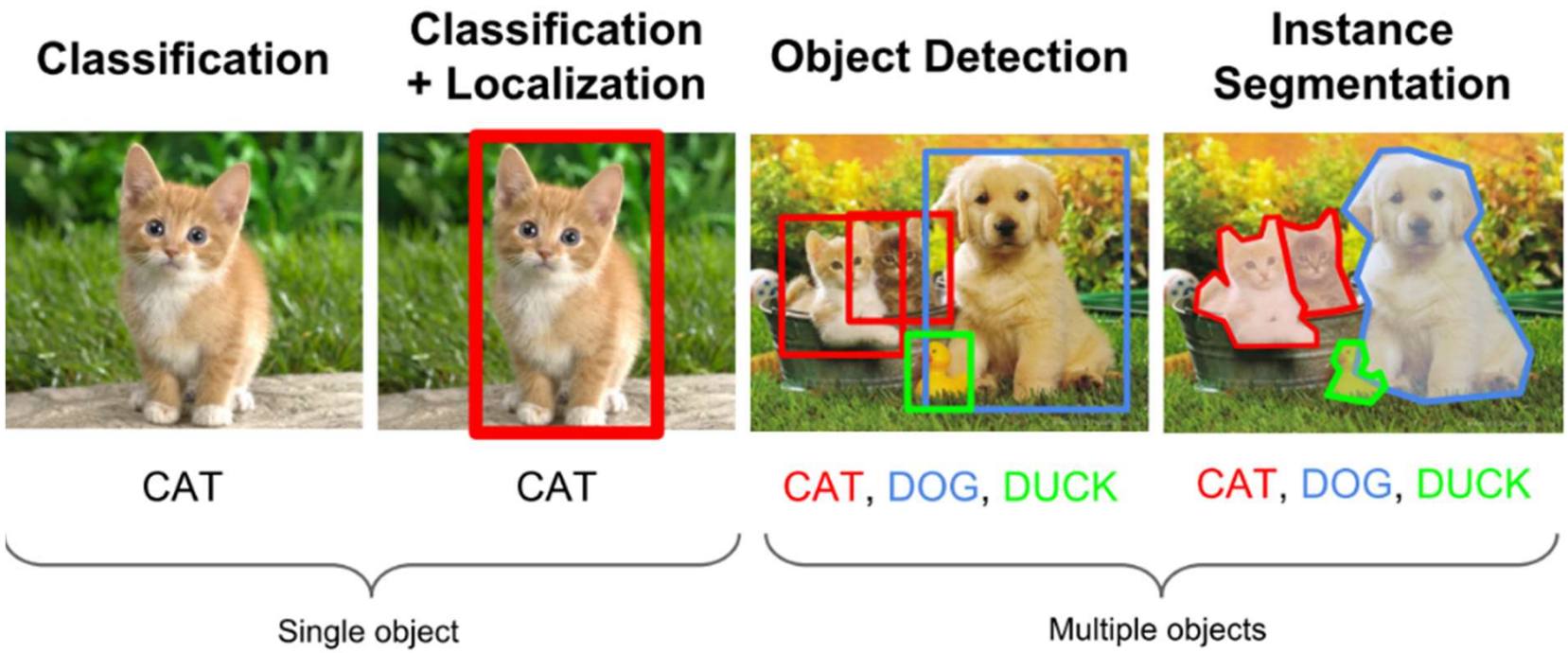
CVPR 2024

Out-Of-Distribution Detection



* OOD Detection is generally the same as OSR in classification task, but OOD Detection encompasses a broader spectrum of learning tasks and solution space (ref. Section 2.6)

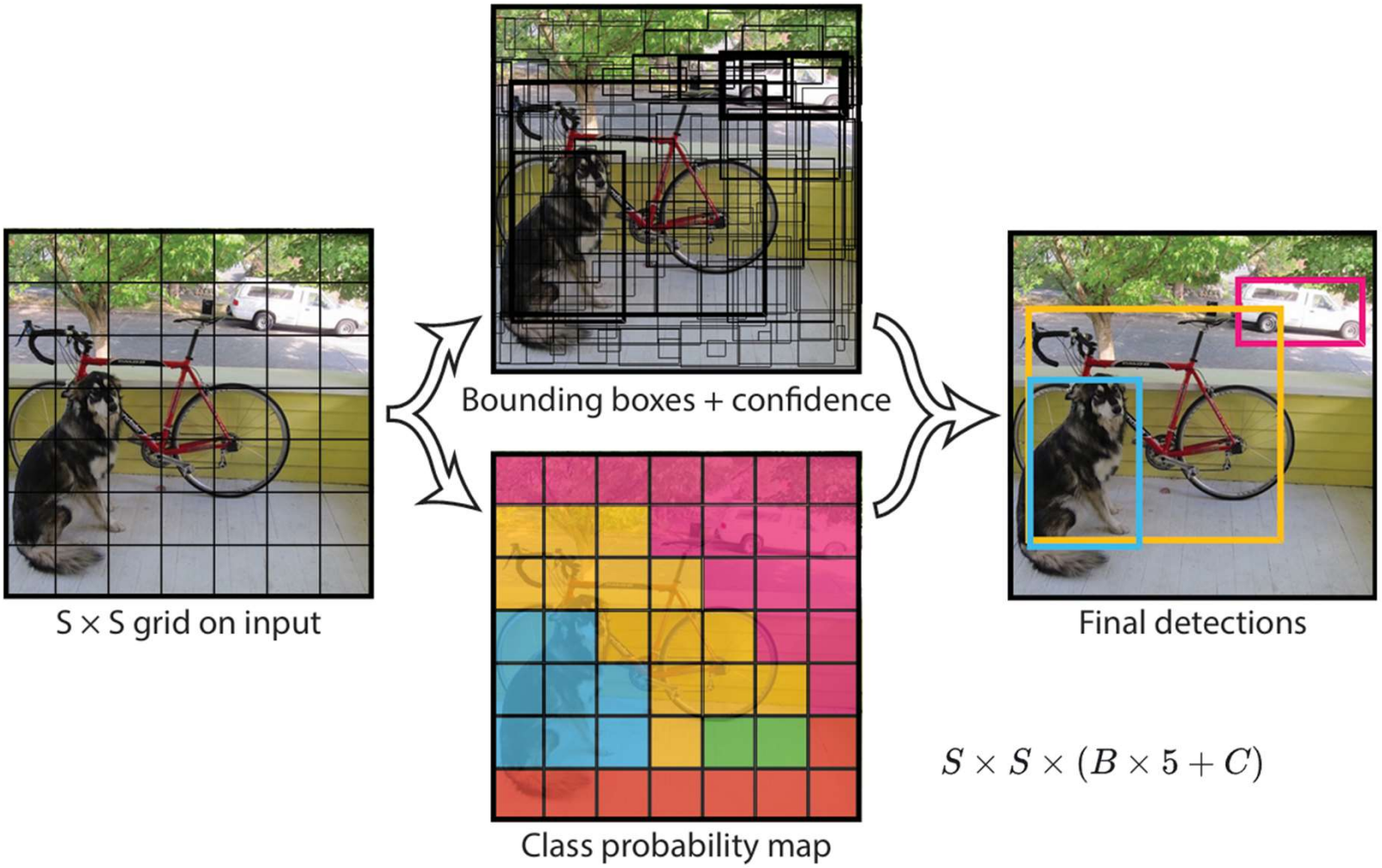
Object Detection



Motivation

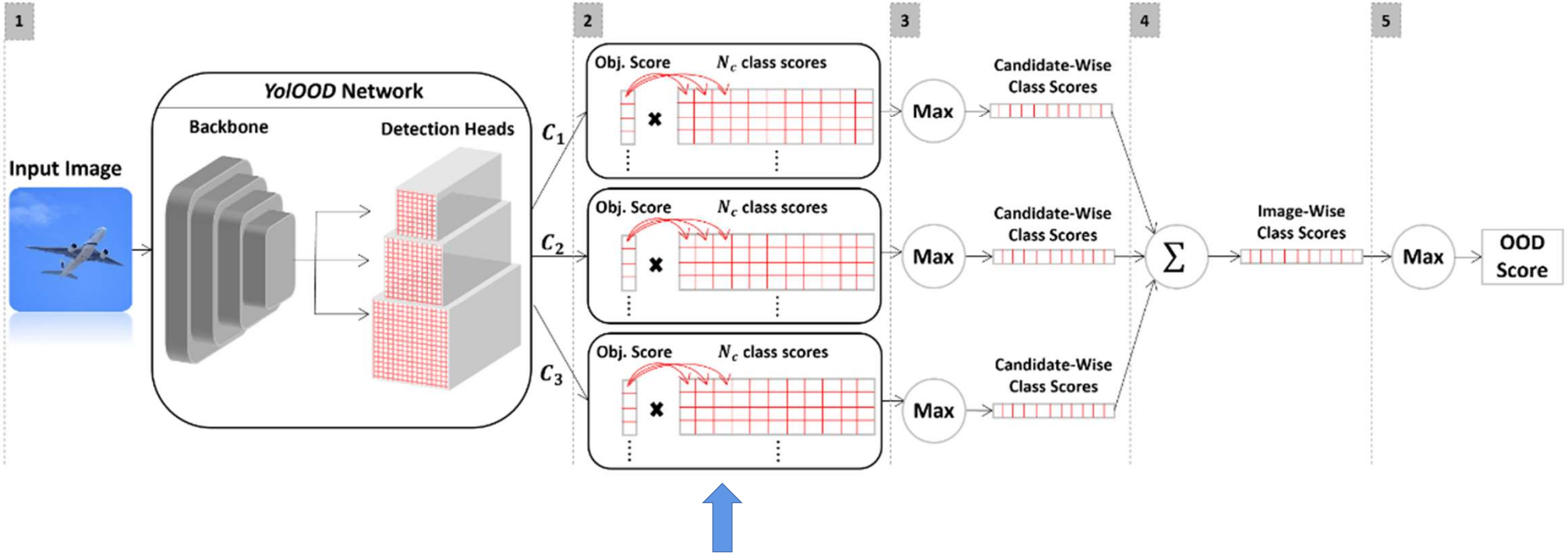
Multi-label image classification and object detection are two closely related tasks. The former involves assigning multiple labels to an image, while the latter goes a step further by not only recognizing the objects present in an image but also localizing them with bounding boxes. Therefore, object detectors have the inherent ability to distinguish between objects of interest and irrelevant objects. This ability, along with object detection similarity to multilabel classification, can be leveraged to create an OOD detection mechanism for the multi-label setting.

YOLO



$$S \times S \times (B \times 5 + C)$$

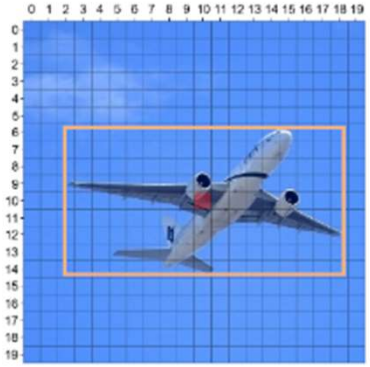
YolOOD



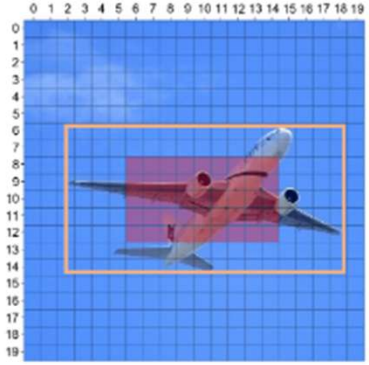
$$W_k \times H_k \times (1 + N_c) \ , \ k \in \{1, 2, 3\}$$

Objectness Score

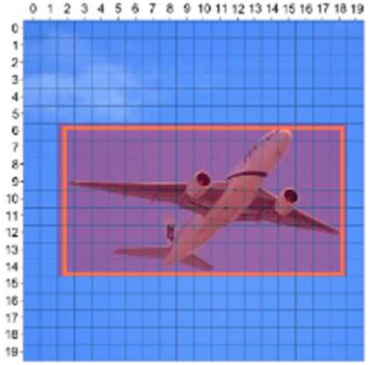
(a) Clean



(b) $p = 0$



(c) $p = 0.5$



(d) $p = 1$

$$\text{let } (W_o, H_o) \in [0,1]^2 \subset \mathbb{R}^2$$

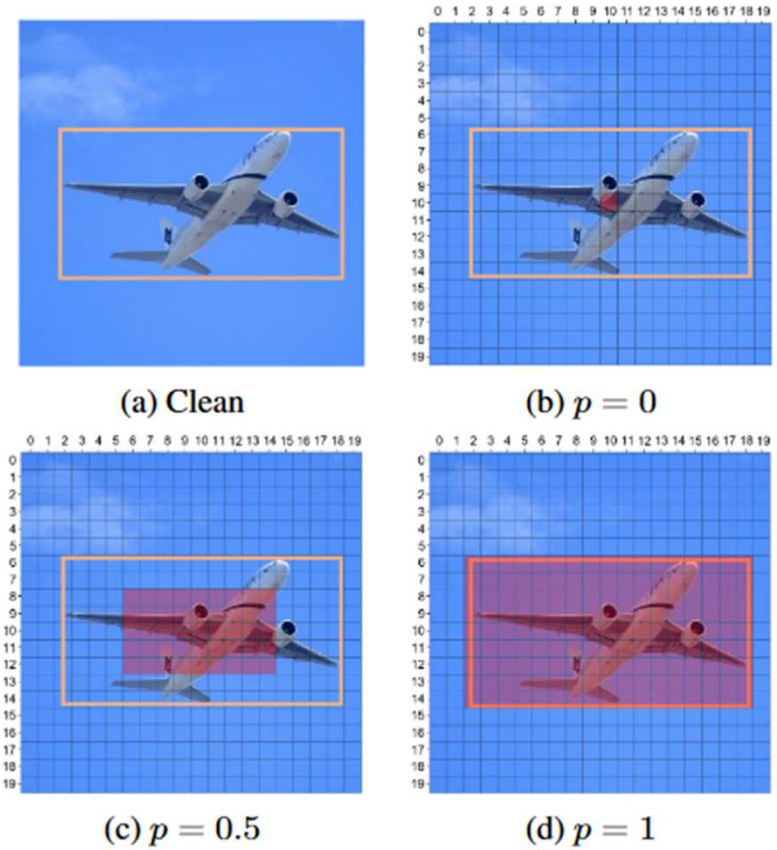
$$(W_r, H_r) = (W_o \cdot W_k, H_o \cdot H_k)$$

$$\varphi_u = i \geq x_{\text{k,center}} - p_k \cdot \frac{W_r}{2}, \varphi_l = j \geq y_{\text{k,center}} - p_k \cdot \frac{H_r}{2}$$

$$\varphi_d = i \leq x_{\text{k,center}} + p_k \cdot \frac{W_r}{2}, \varphi_r = j \leq y_{\text{k,center}} + p_k \cdot \frac{H_r}{2}$$

$$\hat{c}_{\text{obj}}(i,j) = \begin{cases} 1 & \varphi_u \wedge \varphi_d \wedge \varphi_l \wedge \varphi_r \\ 0 & \text{else} \end{cases}$$

Class Scores

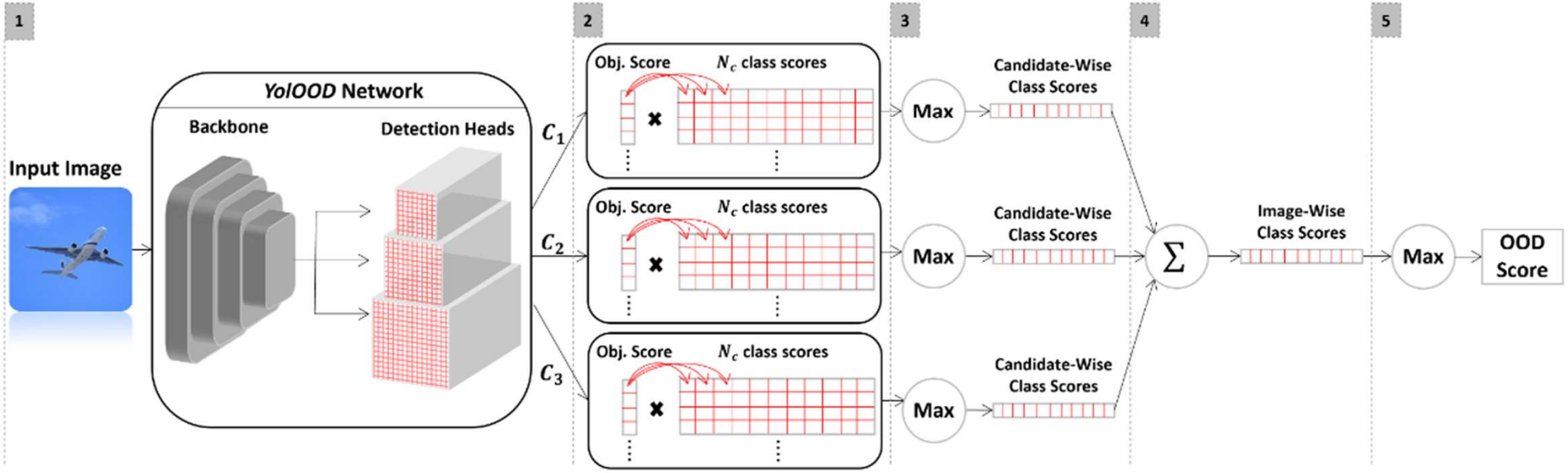


(a) Clean (b) $p = 0$

(c) $p = 0.5$ (d) $p = 1$

$$\hat{c}_{\text{cls } n}(i,j) = \begin{cases} 1 & \text{class } n \text{ is in cell } (i,j) \\ 0 & \text{else} \end{cases}$$

YolOOD Loss Function



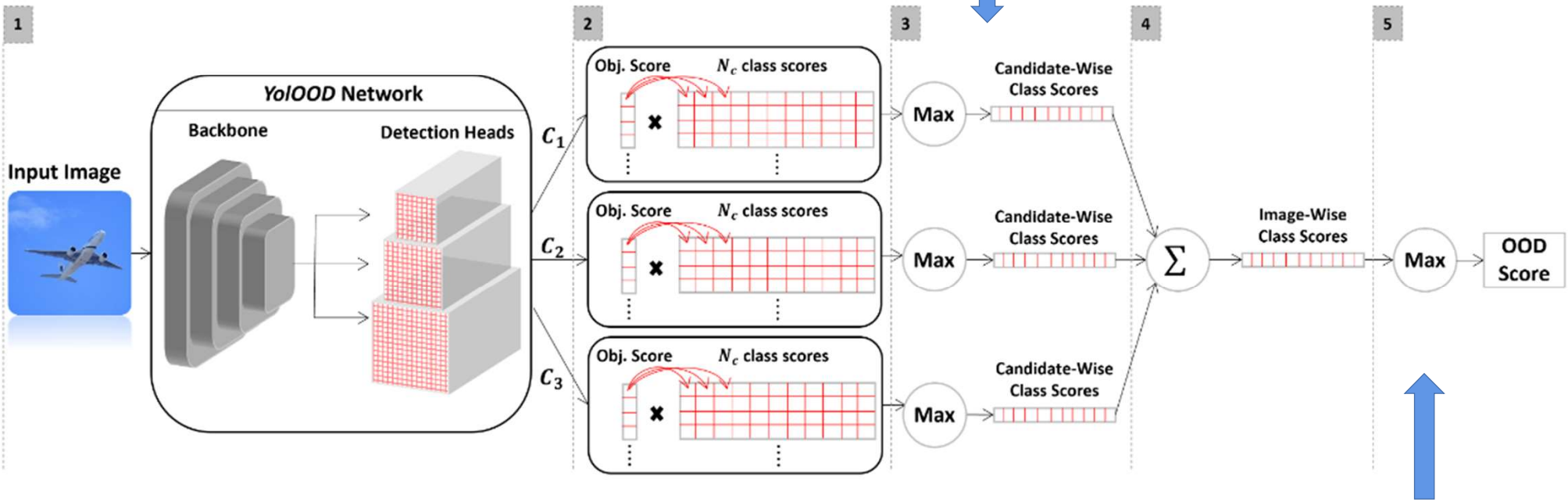
$$\mathcal{L}_{\text{obj}} = \sum_{c \in \mathcal{C}} \mathcal{L}_{\text{BCE}}(c_{\text{obj}}, \hat{c}_{\text{obj}})$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{obj}} + \mathcal{L}_{\text{cls}}$$

$$\mathcal{L}_{\text{cls}} = \sum_{c \in \mathcal{C}} \sum_{n \in \{1,..,N_c\}} \hat{c}_{\text{obj}} \cdot \mathcal{L}_{\text{BCE}}(c_{\text{cls n}}, \hat{c}_{\text{cls n}})$$

Inference

$$y_n = \max_{c \in \mathcal{C}} \{\sigma(c_{\text{obj}}) \cdot \sigma(c_{\text{cls}\,n})\}$$



$$\text{YolOOD}(x) = \max_{n \in \{1,..,N_c\}} \sum_{\mathcal{C}_k \in f_{\text{YolOOD}}(x)} \max_{c \in \mathcal{C}_k} \{\sigma(c_{\text{obj}}) \cdot \sigma(c_{\text{cls}\,n})\}$$

$$G(x,\tau) = \begin{cases} 1 & \text{YolOOD}(x) \geq \tau \\ 0 & \text{YolOOD}(x) < \tau \end{cases}$$

Experiment

Method	$\mathcal{D}_{out}$: Objects365$_{out}$ / $\mathcal{D}_{in}$: PASCAL-VOC	Objects365$_{out}$ / MS-COCO	$\mathcal{D}_{in}$: Objects365$_{in}$	NUS-WIDE / PASCAL-VOC	NUS-WIDE / MS-COCO	NUS-WIDE / Objects365$_{in}$
	FPR95 ↓ / AUROC ↑ / AUPR ↑					
MaxLogit [11]	28.91 / 94.96 / 95.32	16.39 / 96.90 / 99.17	29.95 / 94.33 / 94.38	23.60 / 95.99 / 96.05	12.16 / 97.53 / 99.24	38.07 / 92.62 / 91.48
MSP [10]	50.78 / 88.36 / 88.61	46.26 / 86.78 / 95.63	65.20 / 83.99 / 84.13	47.34 / 89.34 / 88.71	40.89 / 88.33 / 95.53	78.08 / 78.42 / 76.91
Mahalanobis [17]	73.34 / 73.90 / 70.94	88.01 / 48.45 / 75.08	83.32 / 63.19 / 56.67	77.23 / 73.76 / 67.76	90.48 / 52.71 / 75.58	88.46 / 62.47 / 54.29
ODIN [18]	28.91 / 94.96 / 95.32	16.39 / 96.90 / 99.17	29.95 / 94.33 / 94.38	23.60 / 95.99 / 96.05	12.16 / 97.53 / 99.24	38.07 / 92.62 / 91.48
JointEnergy [34]	27.90 / 95.37 / 96.04	14.80 / 97.16 / **99.28**	23.13 / 95.84 / **96.20**	20.19 / 96.53 / 96.76	8.29 / 97.90 / 99.39	24.46 / 95.34 / 94.96
YolOOD-a[1]	18.37 / 96.10 / 95.85	11.70 / 97.21 / 99.19	18.40 / 95.76 / 95.15	21.24 / 96.29 / 96.08	7.62 / 98.13 / 99.43	12.19 / 97.64 / 97.29
YolOOD-o[2]	**16.38 / 96.60 / 96.49**	**11.53 / 97.29** / 99.23	**17.24 / 95.97** / 95.42	**18.47 / 96.85 / 96.77**	**4.40 / 98.56 / 99.57**	**9.54 / 97.99 / 97.61**

[1]trained using the auto-generated annotations. [2]trained using the original annotations.

Experiment

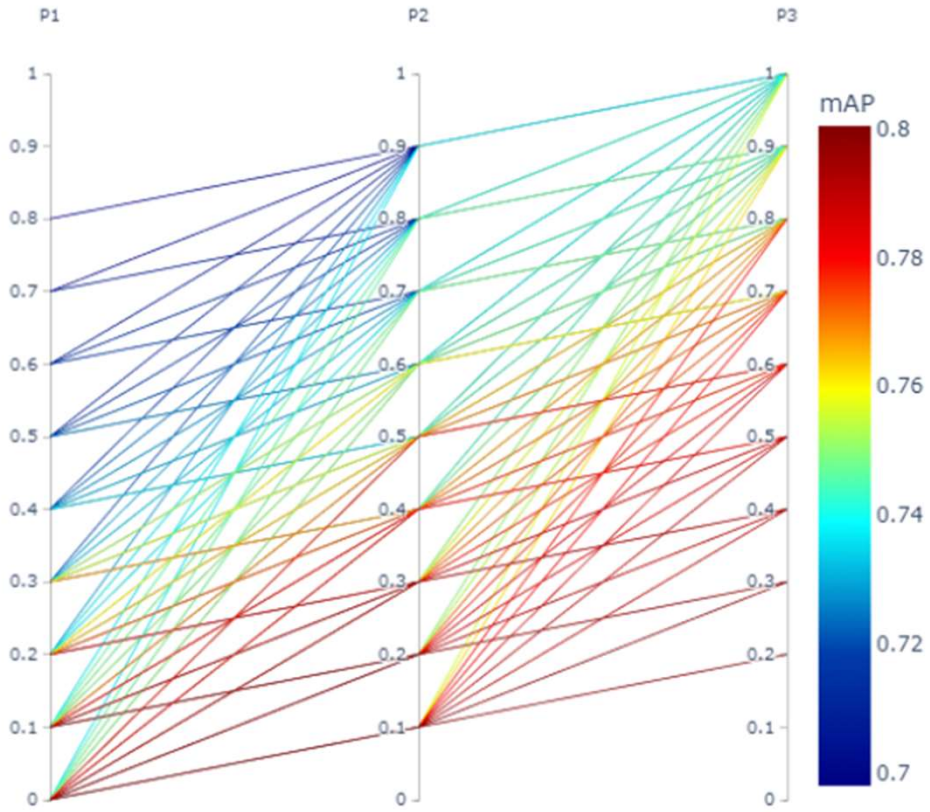


Figure 4. Models' mAP for different p_k combinations on the COCO dataset.

Experiment

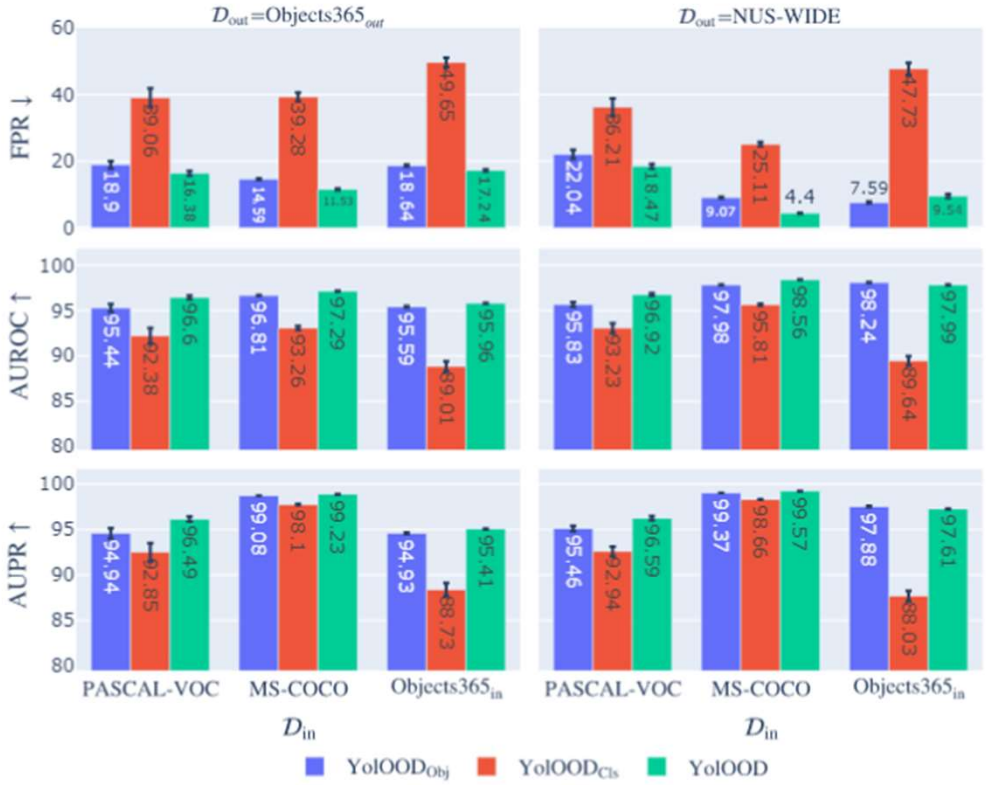


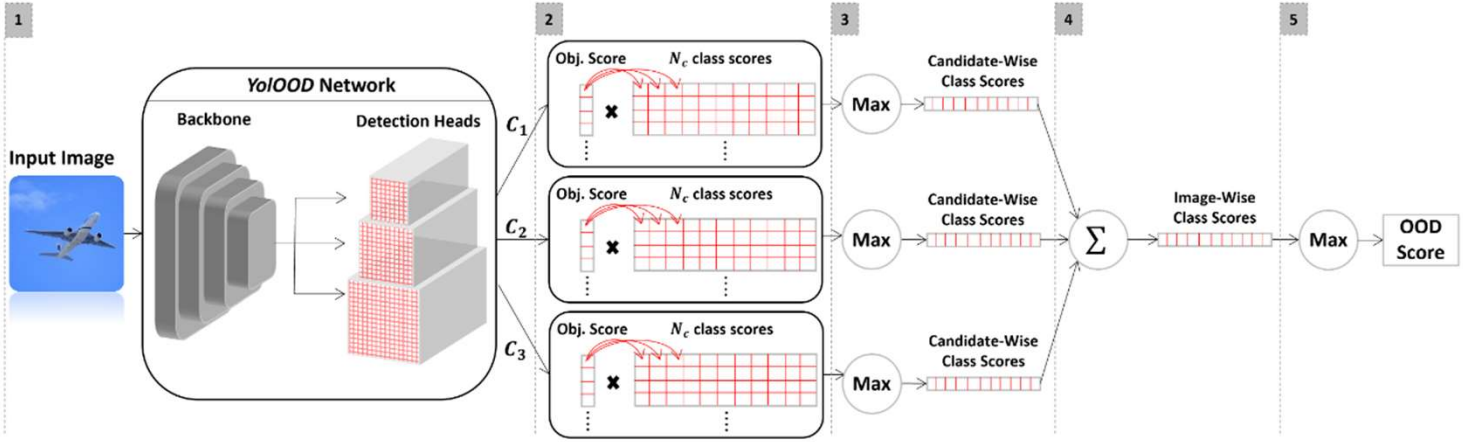
Figure 5. A comparison of the impact of different aggregation methods on YolOOD candidates' output scores (described in Section 4.2).

$$\text{YolOOD}(x) = \max_{n\in\{1,..,N_c\}} \sum_{\mathcal{C}_k \in f_{\text{YolOOD}}(x)} \max_{c\in\mathcal{C}_k} \{\sigma(c_{\text{obj}}) \cdot \sigma(c_{\text{cls}\,n})\}$$

$$\text{YolOOD}_{\text{Cls}}(x) = \max_{n\in\{1,..,N_c\}} \sum_{\mathcal{C}_k \in f_{\text{YolOOD}}(x)} \max_{c\in\mathcal{C}_k} \{\sigma(c_{\text{cls}\,n})\}$$

$$\text{YolOOD}_{\text{Obj}}(x) = \sum_{\mathcal{C}_k \in f_{\text{YolOOD}}(x)} \max_{c\in\mathcal{C}_k} \{\sigma(c_{obj})\}$$

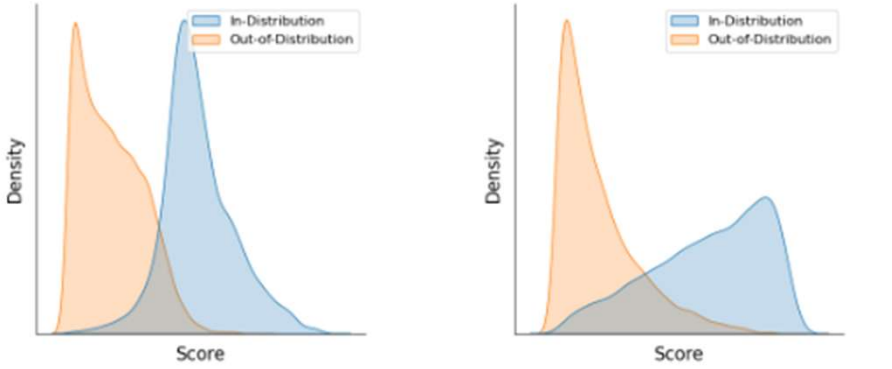
Experiment



Class Agg.	Head Agg.	$\mathcal{D}_{\text{in}}$ PASCAL-VOC	MS-COCO	Objects365$_{\text{in}}$
		FPR95 ↓ / AUROC ↑ / AUPR ↑		
Max	Max	24.10 / 95.63 / 95.61	8.55 / **98.01** / **99.43**	21.01 / 95.70 / 95.08
	Multiply	19.76 / 96.02 / 95.74	10.55 / 97.59 / 99.30	18.22 / 96.23 / 95.68
	Sum	**17.43** / **96.73** / **96.63**	**7.97** / 97.93 / 99.41	**13.39** / **96.98** / **96.52**
Sum	Max	44.56 / 78.31 / 70.89	54.70 / 80.59 / 92.21	35.26 / 85.67 / 79.28
	Multiply	**19.58** / **96.04** / **95.80**	**9.61** / **97.81** / **99.38**	**17.55** / **96.39** / **95.94**
	Sum	40.11 / 84.91 / 81.83	39.50 / 90.20 / 96.71	26.18 / 93.68 / 92.39

Table 2. OOD detection performance when using different combinations of aggregation functions for YolOOD's detection heads and class scores output vector. The results are averaged across the OOD datasets.

Experiment



(a) YolOOD (FPR95=16.38%) (b) YOLO (FPR95=40.57%)

Figure 6. Score distribution when using PASCAL-VOC as the in-distribution dataset and $\text{Objects365}_{\text{out}}$ as the OOD dataset.

Thanks