UMB: Understanding Model Behavior for Open-World Object Detection

Xing Xi, Yangyang Huang, Zhijie Zhong, Ronghua Luo*

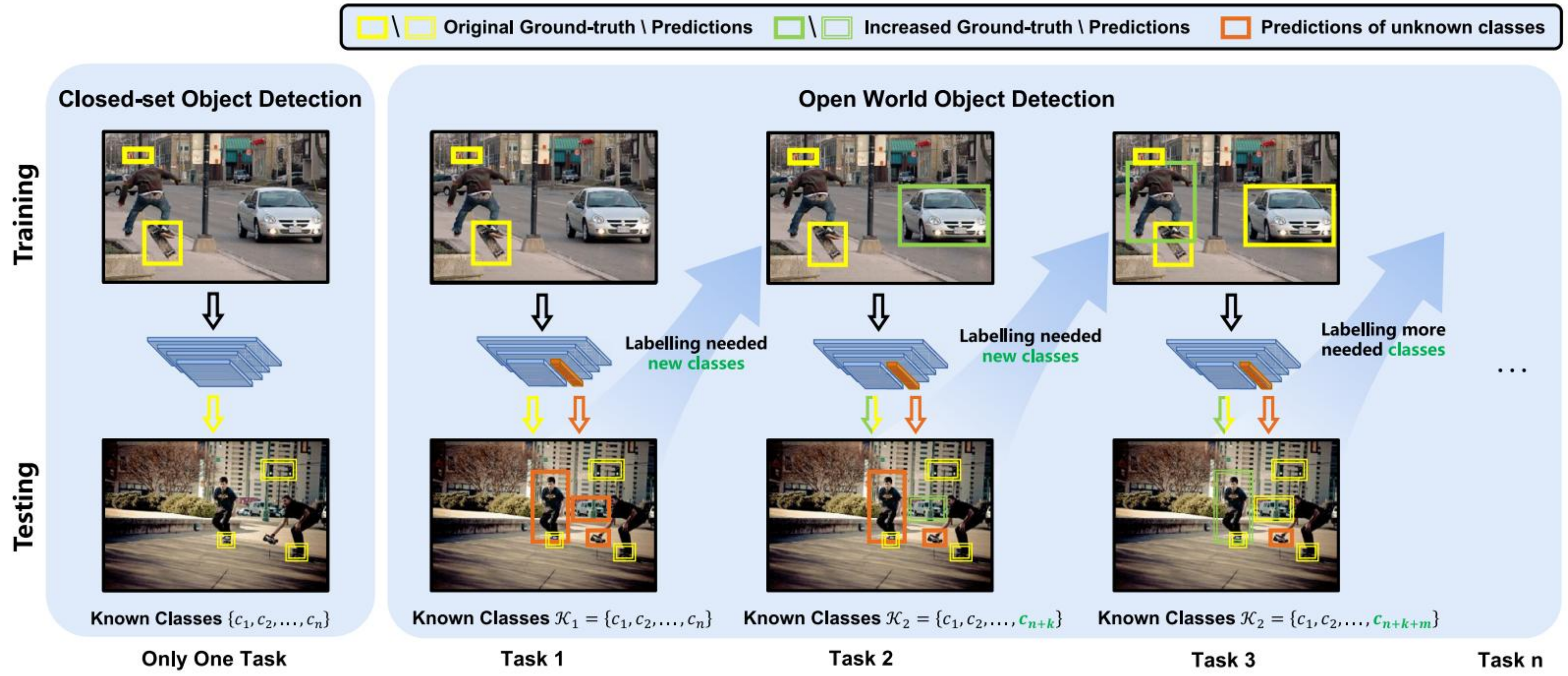
Github: https://github.com/xxyzll

South China University of Technology

UMB: Understanding Model Behavior for Open-World Object Detection
(NeuriPS 2024)

Background

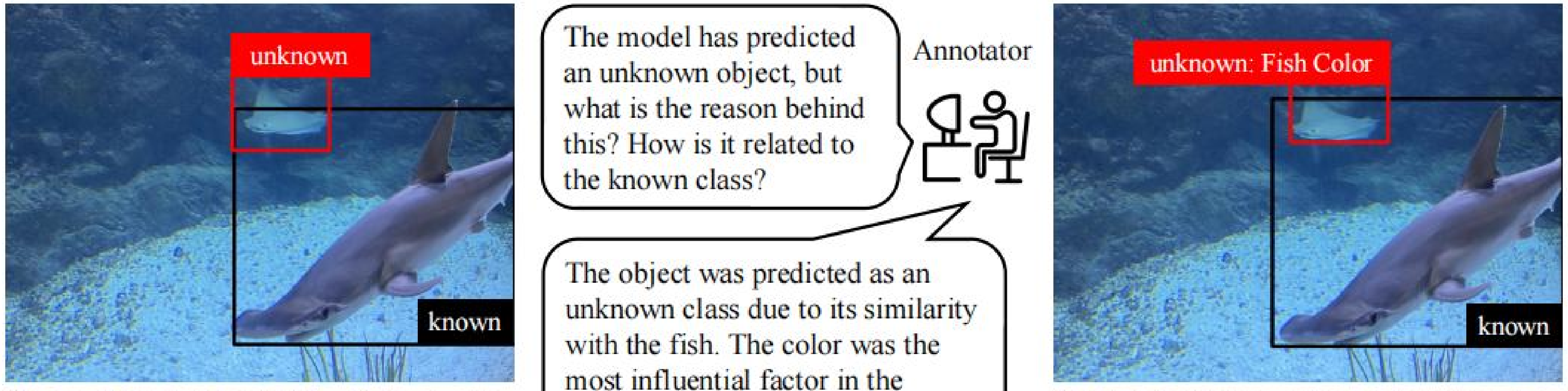
Open World Object Detection (OWOD)



[1] Zhao, X., Ma, Y., Wang, D., Shen, Y., Qiao, Y., & Liu, X. (2023). Revisiting open world object detection. IEEE Transactions on Circuits and Systems for Video Technology.

How to understand the behavior predicted by the model?

Benefits:



Others: Merely identifying unknown objects and providing them to the annotator.

Ours: Identifying unknown targets and comprehending the behavior of the model.

- Infer the similarity of unknown objects to known classes (Helps understand the predictive bias of the model)
- Discover the most significant attributes that influence decision making (Help subsequent data annotation)

How to establish the relationship between attribute similarit y and positive sample probability?

Method:

Known Prediction:

$$p\left(C_j|e_{vis_i}\right) = Sigmoid\left(w_{j,1} \cdot sim\left(e_{vis_i}, e_{att_1}\right) + \ldots + w_{j,n} \cdot sim\left(e_{vis_i}, e_{att_n}\right)\right)$$
$$= Sigmoid\left(\sum_{k=1}^{n} w_{j,k} \cdot sim(e_{vis_i}, e_{att_k})\right),$$

Collecting data: Attribute Similarity->Known Confidence:

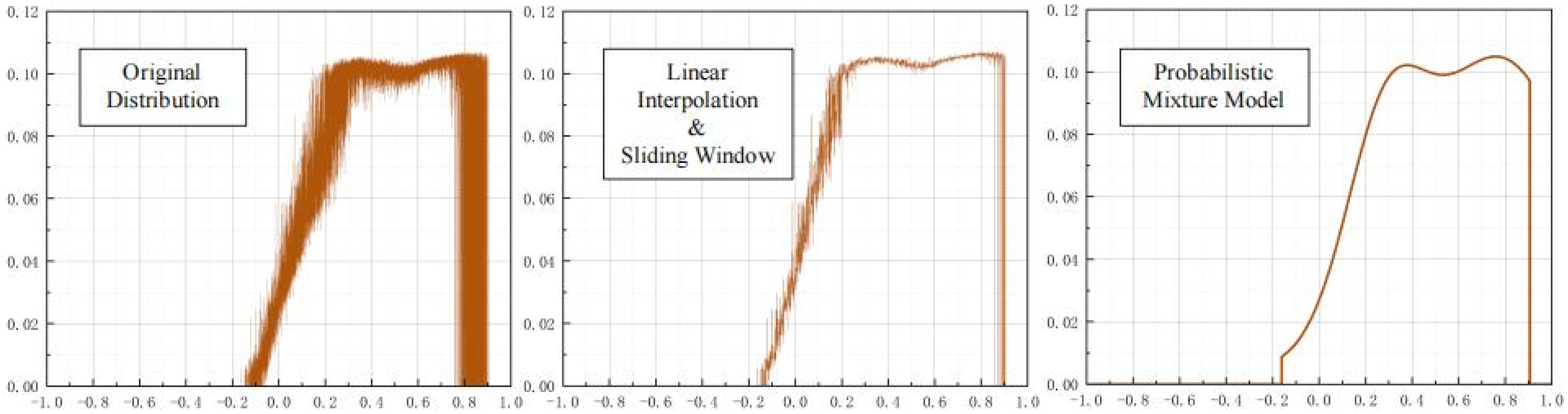
$$\tilde{p}\left(e_{att_i}, C_j|e_{vis_k}\right) = w_{j,i}^{1-\beta} \cdot p\left(C_j|e_{vis_k}\right)^{\beta}, \quad w_{j,i} = W[j,i]$$

What is unknown data? (Maximum Probability)

$$\tilde{p}\left(e_{att_i}, C_u|e_{vis_k}\right) = max\left(\tilde{p}\left(e_{att_i}, C_1|e_{vis_k}\right), \ldots, \tilde{p}\left(e_{att_i}, C_m|e_{vis_k}\right)\right)$$
$$= \underset{j \in [1,m]}{argmax}\left(\tilde{p}\left(e_{att_i}, C_j|e_{vis_k}\right)\right)$$

How to establish the relationship between attribute similarit y and positive sample probability?

Process the data and establish a continuous probability distribution:



Linear Interpolation (Estimating points that do not exist in the data):

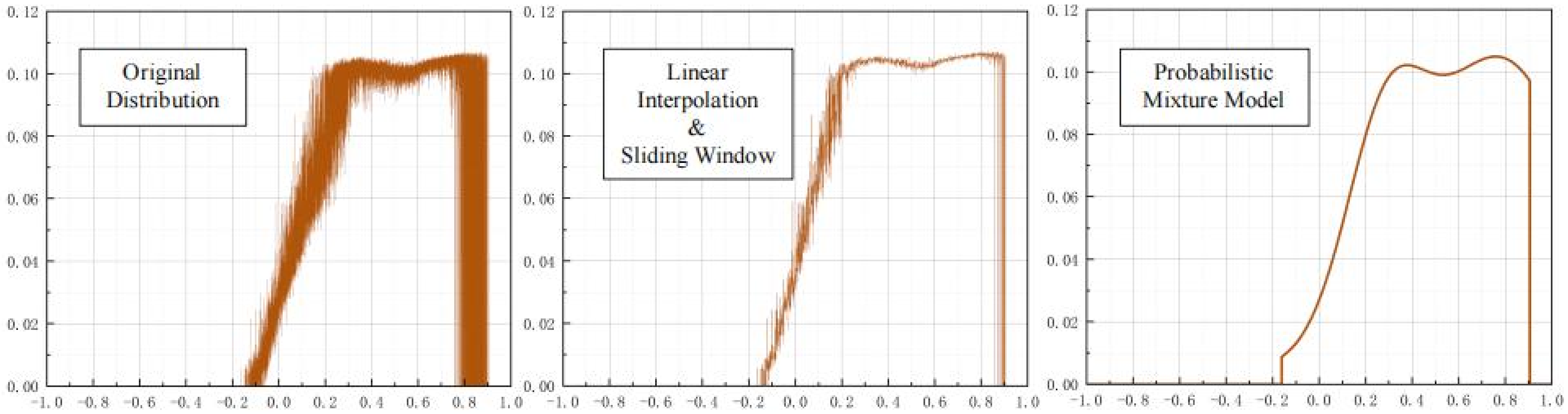
$$f_{i,j}(\underline{x}) = k(\underline{x} - \underline{x}_l) + f_{i,j}(\underline{x}_l), \ \ k = (f_{i,j}(\underline{x}_r) - f_{i,j}(\underline{x}_l))/(\underline{x}_r - \underline{x}_l),$$

Sliding Window (filtering noise):

$$f_{i,j}(sim(e_{vis_k}, e_{att_i})) = \underset{a \in [0, W_{sz}-1]}{argmax} \ f_{i,j}(sim(e_{vis_k}, e_{att_i}) + a).$$

How to establish the relationship between attribute similarity and positive sample probability?

Process the data and establish a continuous probability distribution:



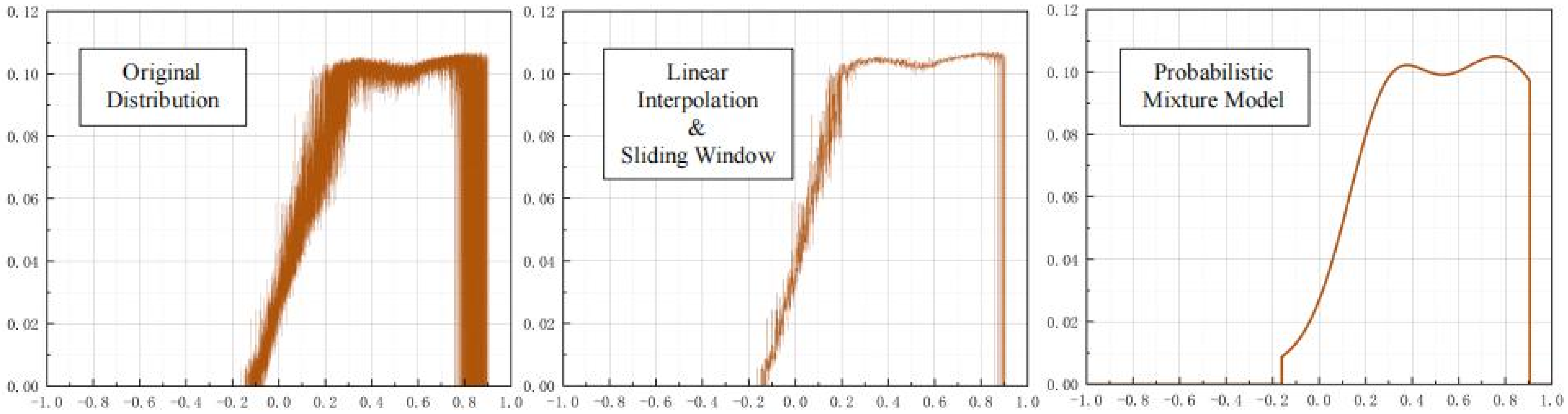
Multiple Gaussian Distributions Mixed:

$$f_{i,u}(sim(e_{vis_k}, e_{att_i})) = \sum_{a=1}^{A} Gm(sim(e_{vis_k}, e_{att_i}) | w_a, \sigma_a, \mu_a),$$

$$Gm(x|w, \sigma, \mu) = w \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}},$$

How to establish the relationship between attribute similarity and positive sample probability?

Process the data and establish a continuous probability distribution:



Multiple Weibull Distributions Mixed (Unsymmetrical):

$$Wb(x|w,\lambda,k) = w \cdot \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{(k-1)} e^{-(\frac{x}{\lambda})^k}.$$

Use Gradient Descent to Find the Optimal Parameters: $Gm(x|w,\sigma,\mu)$ $Wb(x|w,\lambda,k)$

How to predict unknown objects?

Unknown Inference & Additional Information:

Mean Embedding: $e_{att_u} = \frac{1}{m}\sum_{j=1}^{m}\left(\sum_{i=1}^{n} e_{att_i} \cdot w_{j,i}\right) \in \mathbb{R}^d$

Empirical Probability (Empirical Prob):

$$\hat{f}_u(e_{vis_k}) = \sum_{i=1}^{n} f_{i,u}(sim(e_{vis_k}, e_{att_i})) \cdot \bar{w}_i, \ \ \bar{w}_i = \underset{j\in[1,m]}{argmax}\ w_{j,i}.$$

In-Distribution Probability (ID Prob):

$$f_{\mathrm{ID}}(e_{vis_k}) = \sum_{i=1}^{n} Sigmoid(T(e_{vis_k}, e_{att_i})) \cdot \bar{w}_i$$

Out-of-Distribution Probability (OOD Prob):

$$f_{OOD}(e_{vis_k}) = \underset{j\in[1,m]}{argmax}(1 - \underset{j\in[1,m]}{Softmax}\left(T(sim(e_{vis_k}, e_{att_i})) \cdot w_{j,i}\right)).$$

How to infer additional informations?

Unknown Inference & Additional Information:

Unknown Prob:

$$p(C_u|e_{vis_k}) = Sigmoid((\underbrace{\hat{f}_u(e_{vis_k}) \cdot (1-\alpha)}_{Empirical\ Prob} + \underbrace{f_{\mathrm{ID}}(e_{vis_k}) \cdot \alpha}_{ID\ Prob}) \cdot \underbrace{f_{OOD}(e_{vis_k})}_{OOD\ Prob}) \cdot Sigmoid(\underbrace{T(sim(e_{vis_k}, e_{att_u}))}_{Average\ Similarity}),$$

Similarity Between Known and Unknown Classes:

$$S_u(e_{vis_k}) = \underset{j \in [1,m]}{softmax}\left(\sum_{i=1}^{n} f_{i,j}(sim(e_{vis_k}, e_{att_i})) + p(C_j|e_{vis_k})\right)$$

Maximal Attribute Contribution:

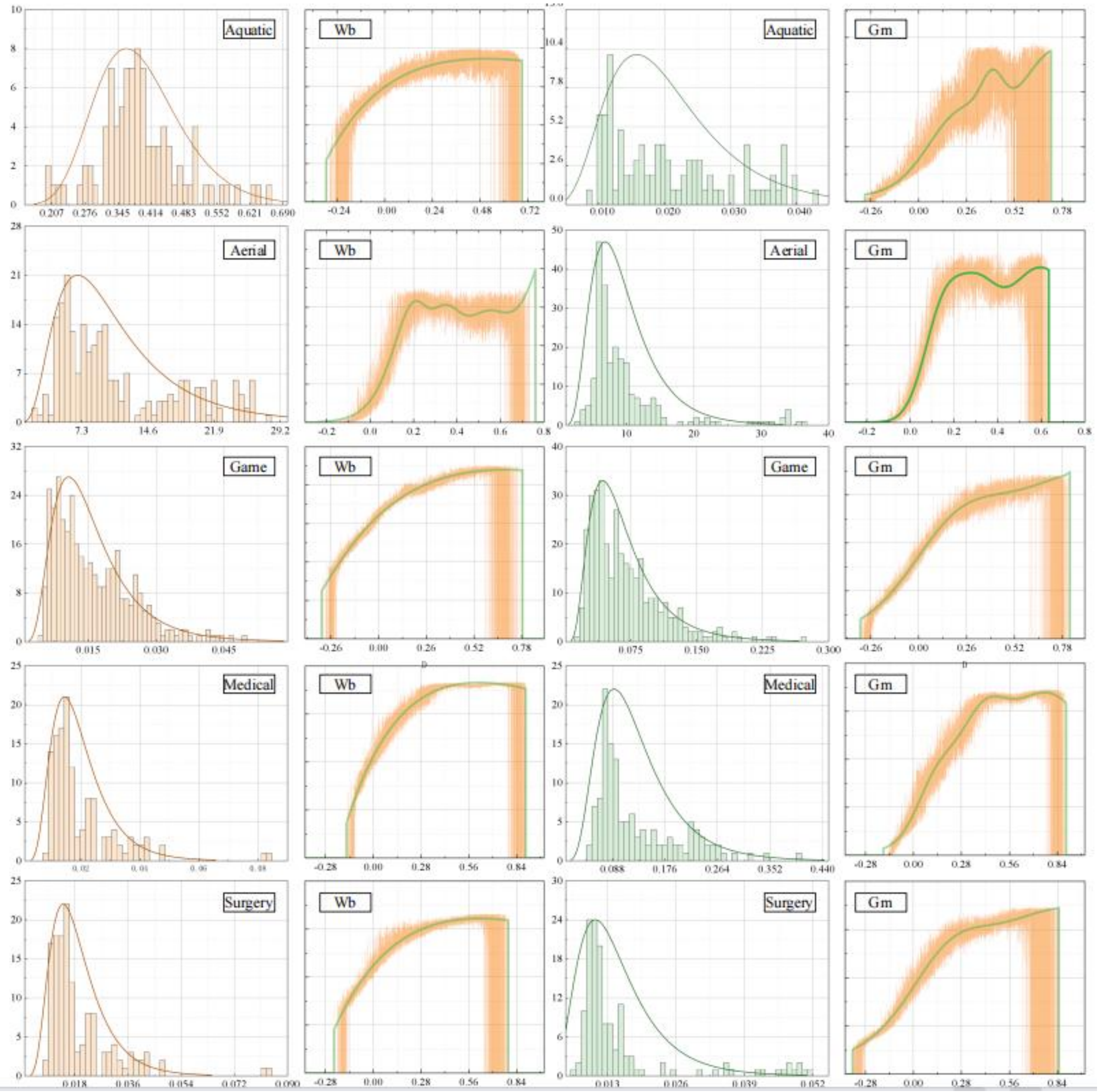
$$Ctr(e_{att_i}) = \overline{w}_i \cdot (Sigmoid(T(e_{vis_k}, e_{att_i})) \cdot \alpha + f_{i,u}(sim(e_{vis_k}, e_{att_i})) \cdot (1-\alpha))$$

State-of-the-art Comparison for Open-World Object Detection

	Aquatic				Aerial				Game				Medical				Surgery				Overall			
Task IDs(->)	Task1		Task2		Task1		Task2		Task1		Task2		Task1		Task2		Task1		Task2		Task1		Task2	
	U	K	PK	CK	U	K	PK	CK	U	K	PK	CK	U	K	PK	CK	U	K	PK	CK	U	K	PK	CK
Base+GT-B	29.8	45.0	45.0	36.7	1.3	5.7	5.7	1.4	15.0	0.4	0.4	0.1	0.5	0.0	0.0	0.1	5.6	1.5	1.4	0.3	10.4	10.5	10.5	7.7
Base-FS-B	7.1	41.1	41.1	31.9	1.2	10.4	10.1	4.0	16.0	4.6	4.8	3.9	0.6	6.1	6.1	3.3	1.3	11.9	11.3	10.9	5.2	14.8	14.7	10.8
FOMO-B	3.5	43.8	44.1	40.8	0.9	12.0	12.6	5.4	13.3	3.8	4.4	4.1	2.1	6.4	5.5	11.5	6.1	12.7	12.9	11.0	5.2	15.7	15.9	14.6
Base+GT-L	34.8	36.0	36.0	42.3	1.0	7.9	7.2	0.8	12.4	0.9	0.8	0.3	2.4	0.2	0.2	0.3	2.4	0.2	2.6	1.3	10.6	9.0	9.4	9.0
Base-FS-L	2.4	43.6	42.9	42.8	9.7	23.7	21.9	13.0	8.2	10.4	10.2	13.4	1.1	23.2	21.7	24.2	3.6	26.0	25.0	7.4	5.0	25.4	24.3	20.2
FOMO-L	18.2	50.1	48.1	47.1	6.0	25.3	23.7	16.0	30.4	10.7	9.9	11.2	9.4	21.8	19.9	34.6	12.0	29.0	28.9	8.5	15.2	27.4	26.1	23.5
Ours:																								
UMB-Gm-B	13.3	43.8	43.0	39.7	**1.5**	18.8	19.0	5.9	15.2	4.1	4.7	4.3	**2.3**	5.4	3.5	11.8	10.1	13.9	14.3	11.1	8.5	17.2	16.9	14.6
UMB-Wb-B	**13.5**	43.8	43.0	39.7	1.4	18.8	19.0	5.9	**16.3**	4.1	4.7	4.3	**2.3**	5.4	3.5	11.8	**14.5**	13.9	14.3	11.1	**9.6**	17.2	16.9	14.6
UMB-Gm-L	**18.6**	50.7	50.5	50.4	**11.2**	42.7	40.4	22.6	**35.1**	11.1	10.7	10.5	**13.2**	22.2	19.1	34.5	24.5	36.6	39.0	17.4	**20.5**	32.7	31.9	27.1
UMB-Wb-L	**18.6**	50.8	50.5	50.4	11.1	42.8	40.4	22.5	32.7	11.1	10.7	10.5	8.6	22.3	17.3	33.2	**25.6**	36.6	39.0	17.4	19.3	32.7	31.6	26.8

UMB: Understanding Model Behavior for Open-World Object Detection
(NeurIPS 2024)

Distribution & Visualization



THANKS