

UMB: Understanding Model Behavior for Open-World Object Detection

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

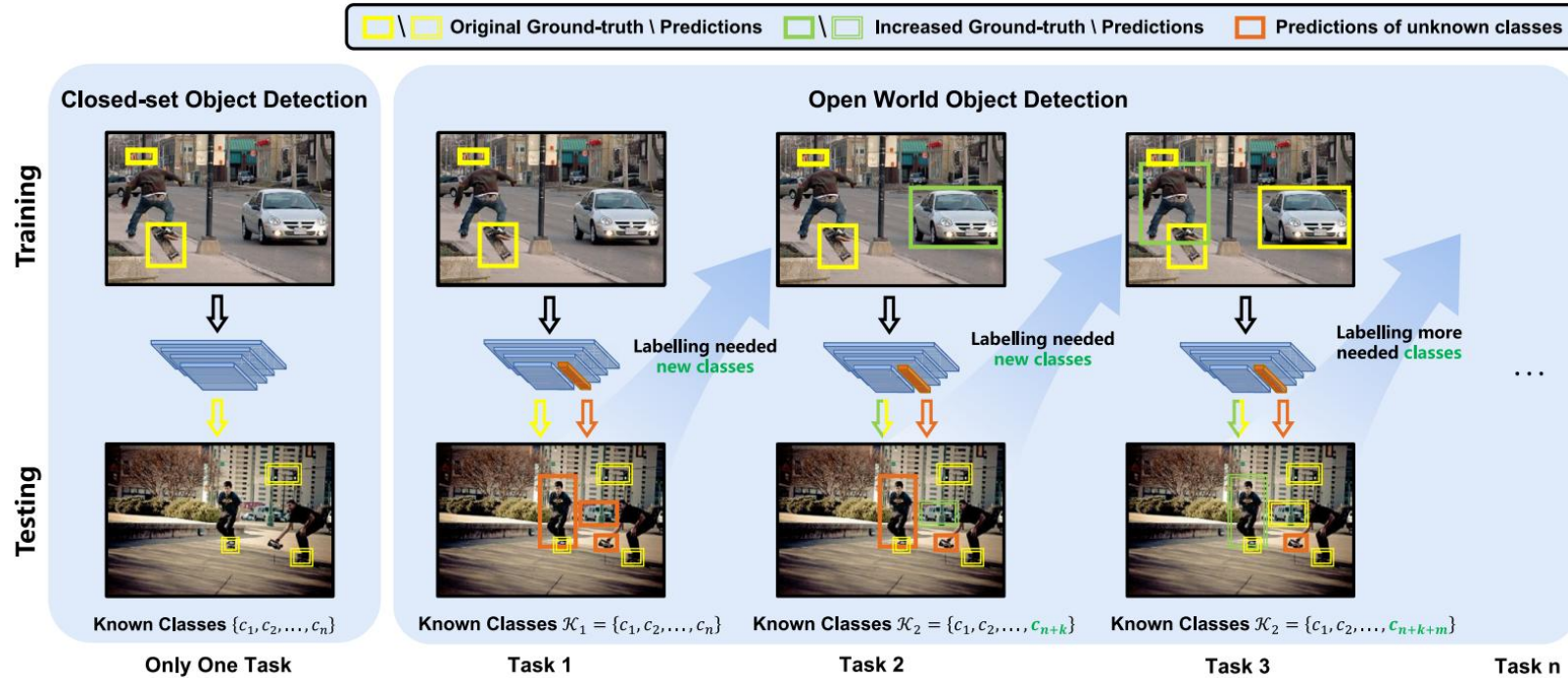
Github: <https://github.com/xxyzll>



华南理工大学
South China University of Technology

Background

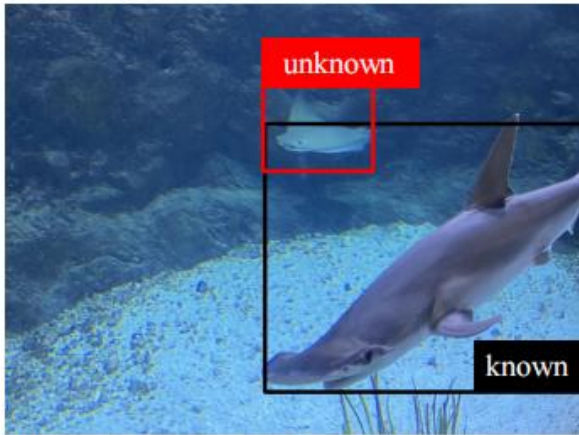
Open World Object Detection (OWOD)



[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:

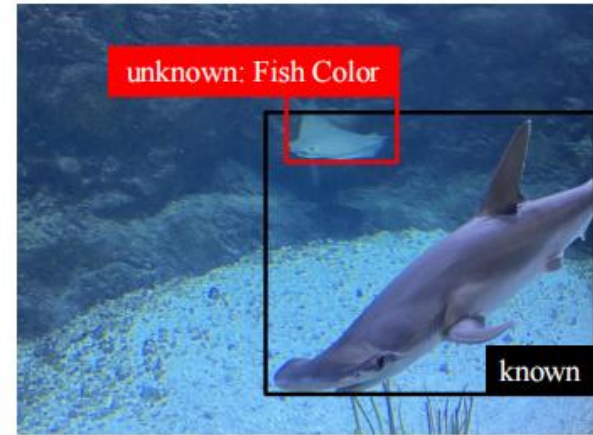


Others: Merely identifying unknown objects and providing them to the annotator.

The model has predicted an unknown object, but what is the reason behind this? How is it related to the known class?



The object was predicted as an unknown class due to its similarity with the fish. The color was the most influential factor in the model's decision-making process.



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 similarity and positive sample probability?

Method:

Known Prediction:

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

Collecting data: Attribute Similarity->Known Confidence:

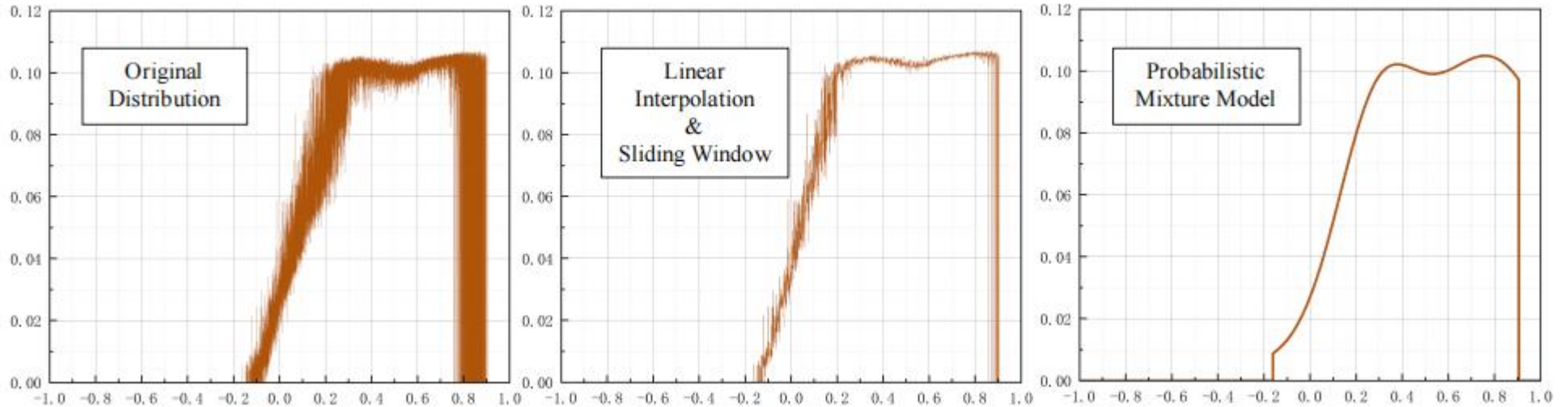
$$\tilde{p}(e_{att_i}, C_j|e_{vis_k}) = w_{j,i}^{1-\beta} \cdot p(C_j|e_{vis_k})^\beta, \quad w_{j,i} = W[j, i]$$

What is unknown data? (Maximum Probability)

$$\begin{aligned} \tilde{p}(e_{att_i}, C_u|e_{vis_k}) &= \max(\tilde{p}(e_{att_i}, C_1|e_{vis_k}), \dots, \tilde{p}(e_{att_i}, C_m|e_{vis_k})) \\ &= \underset{j \in [1, m]}{\text{argmax}}(\tilde{p}(e_{att_i}, C_j|e_{vis_k})) \end{aligned}$$

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

Process the data and establish a continuous probability distribution:



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

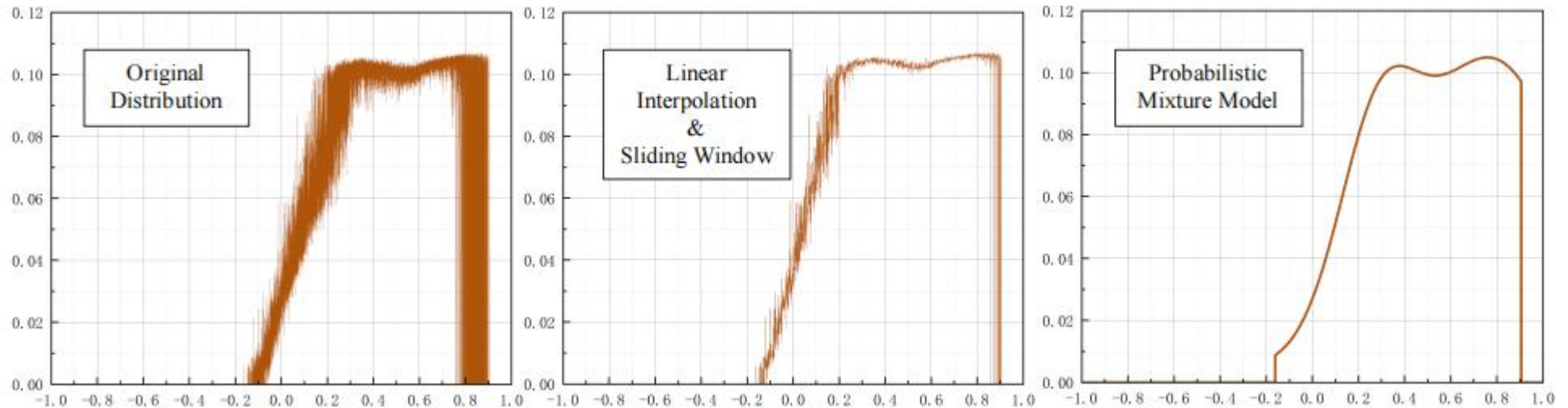
$$f_{i,j}(\underline{x}) = k(\underline{x} - \underline{x}_l) + f_{i,j}(\underline{x}_l), \quad k = (f_{i,j}(\underline{x}_r) - f_{i,j}(\underline{x}_l)) / (\underline{x}_r - \underline{x}_l),$$

Sliding Window (filtering noise):

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

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

Process the data and establish a continuous probability distribution:



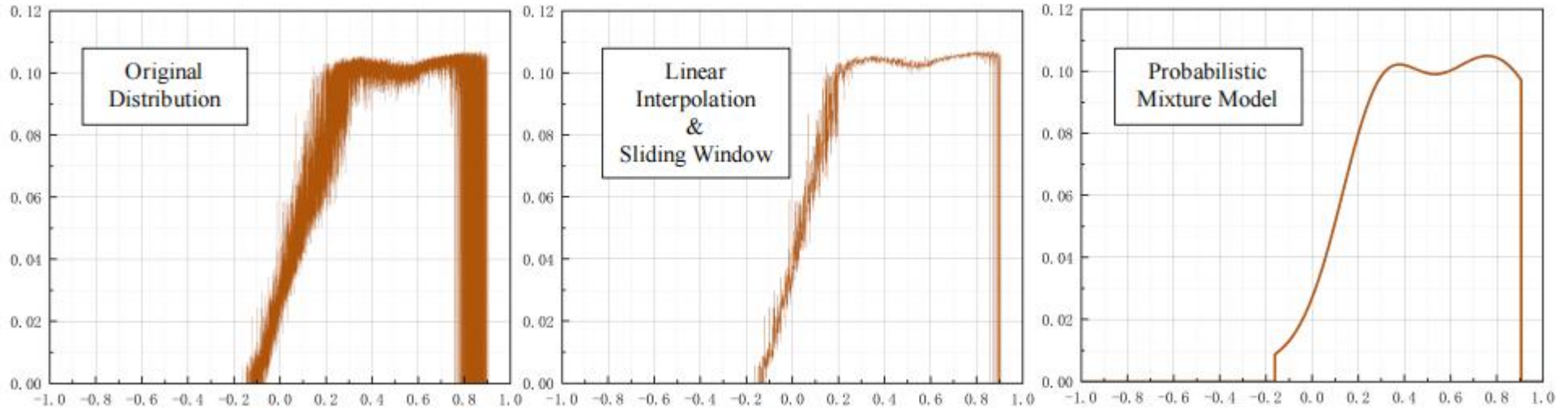
Multiple Gaussian Distributions Mixed:

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

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

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):

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

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):

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

In-Distribution Probability (ID Prob):

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

Out-of-Distribution Probability (OOD Prob):

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

How to infer additional informations?

Unknown Inference & Additional Information:

Unknown Prob:

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

Similarity Between Known and Unknown Classes:

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

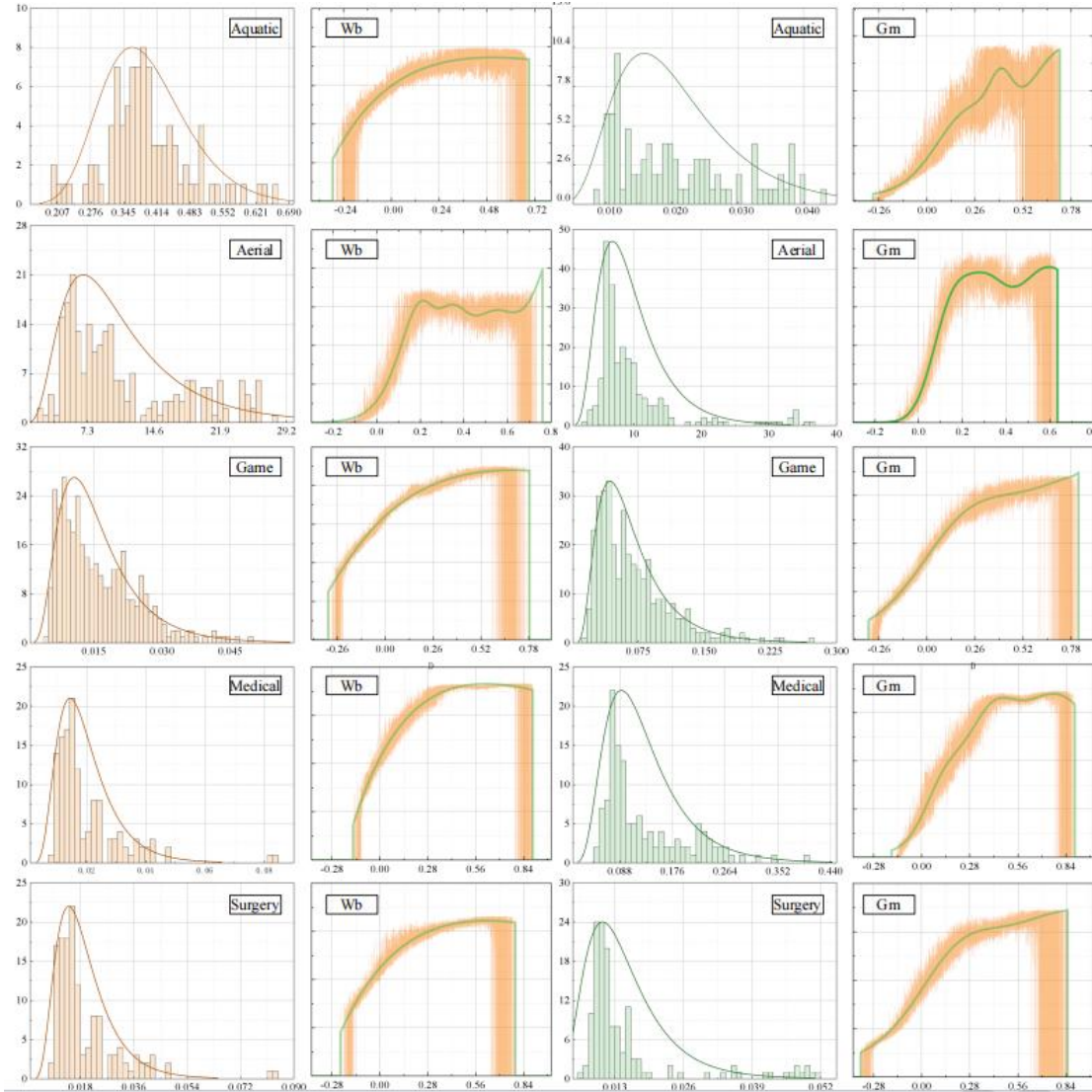
Maximal Attribute Contribution:

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

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

Task IDs(->)	Aquatic				Aerial				Game				Medical				Surgery				Overall			
	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	<u>7.1</u>	41.1	41.1	31.9	<u>1.2</u>	10.4	10.1	4.0	<u>16.0</u>	4.6	4.8	3.9	0.6	6.1	6.1	3.3	1.3	11.9	11.3	10.9	<u>5.2</u>	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	<u>2.1</u>	6.4	5.5	11.5	<u>6.1</u>	12.7	12.9	11.0	<u>5.2</u>	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	<u>9.7</u>	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	<u>18.2</u>	50.1	48.1	47.1	6.0	25.3	23.7	16.0	<u>30.4</u>	10.7	9.9	11.2	<u>9.4</u>	21.8	19.9	34.6	<u>12.0</u>	29.0	28.9	8.5	<u>15.2</u>	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

Distribution & Visualization



This section illustrates model behavior through visualizations and tables for various categories. Each category is shown with a ground truth image, a model prediction image with bounding boxes and confidence scores, and a table of detected objects with their categories and attribute texts.

Aquatic

ID	Category	Attribute Text
1	Shark	Texture is feathers. (16.3 %)
2	Shark	Appearance is suckers. (17.1 %)
3	Penguin	Context is sand-colored. (16.7 %)
4	Penguin	Context is sand-colored. (17.5 %)

Aerial

ID	Category	Attribute Text
1	Chimney	Shape is round. (23.6 %)
2	Chimney	Shape is round. (22.7 %)
3	Expressway-toll-station	Context is concessio n stands. (23.0 %)

Game

ID	Category	Attribute Text
1	Jhin	object which (is/as/ etc) Features is horns. (14.8 %)
2	Lissandra	object which (is/as/ etc) Shape is cuboid. (14.8 %)
3	Jhin	object which (is/as/ etc) Texture is fine. (15.1 %)

Medical

ID	Category	Attribute Text
1	Soft Tissue Calcification	Material is cortical shell. (67.8 %)
2	Distal Phalanges	Features is triquetrum. (69.5 %)
3	Third Metacarpal Bone	Context is tomo. (68.9 %)

Surgery

ID	Category	Attribute Text
1	Scalpel	Behavior is repositi- oning. (6.8 %)
2	Scalpel	Shape is thin. (6.8 %)
3	Hook	Shape is thin. (6.9 %)

THANKS