Detecting Open World Objects via Partial Attribute Assignment

Supplementary Material

Contents

A. Implementation Details

A.1. POT Optimization

We follow [S2] to solve the Partial Optimal Transport (POT) problem. In particular, the POT problem with entropic regularization in Eq. (6) can be rewritten as a Kullback-Leibler projection:

$$\min_{\mathbf{T}\in\Pi(\mathcal{V},\mathcal{A})} \langle \mathbf{T}, \mathbf{C}\rangle_F - \epsilon h(\mathbf{T}) = \epsilon \min_{\mathbf{T}\in\Pi(\mathcal{V},\mathcal{A})} \mathrm{KL}(\mathbf{T}|\mathbf{D})\,, \quad \text{(A)}$$

in which $\mathbf{D} = \exp\left(-\frac{\mathbf{C}}{\epsilon}\right)$ and $\mathrm{KL}(\cdot|\cdot)$ denotes the Kullback-Leibler divergence:

$$\mathrm{KL}(\mathbf{T}|\mathbf{D}) \triangleq \sum_{m,n=1}^{M,N} T_{m,n}\left(\log\left(\frac{T_{m,n}}{D_{m,n}}\right) - 1\right)\,. \quad \text{(B)}$$

According to the definition in Eq. (5), $\Pi(\mathcal{V},\mathcal{A})$ constitutes an intersection of two convex sets: $\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2$, where

$$\mathcal{C}_1 = \{\mathbf{T} \in \mathbb{R}_+^{M\times N}; \mathbf{T}\mathbf{1}_N = \mathcal{V}\}\,, \quad \text{(C)}$$

$$\mathcal{C}_2 = \{\mathbf{T} \in \mathbb{R}_+^{M\times N}; \mathbf{T}^\top\mathbf{1}_M \leqslant \mathcal{A}\}\,. \quad \text{(D)}$$

We follow Dykstra's algorithm [S3] to iteratively solve the above convex optimization problem in the Kullback-Leibler setting [S1]. To begin with, we assume

$$\forall\, i \in \mathbb{N}, \quad \mathcal{C}_{i+2} = \mathcal{C}_i\,. \quad \text{(E)}$$

With the initialization:

$$\mathbf{T}^{(0)} = \mathbf{D}\,, \quad \boldsymbol{q}^{(0)} = \boldsymbol{q}^{(-1)} = \mathbf{1}\,, \quad \text{(F)}$$

the iterative calculations in step i are defined as

$$\mathbf{T}^{(i)} = \arg\min_{\mathbf{T}^{(i)}\in\mathcal{C}_i} \mathrm{KL}\left(\mathbf{T}^{(i)}|\mathbf{T}^{(i-1)} \odot \boldsymbol{q}^{(i-2)}\right)\,, \quad \text{(G)}$$

$$\boldsymbol{q}^{(i)} = \boldsymbol{q}^{(i-2)} \odot \frac{\mathbf{T}^{(i-1)}}{\mathbf{T}^{(i)}}\,, \quad \text{(H)}$$

where "$\odot$" denotes the Hadamard product. The above calculation converges to

$$\mathbf{T}^{(i)} \to \arg\min_{\mathbf{T}^{(i)}\in\mathcal{C}} \mathrm{KL}\big(\mathbf{T}^{(i)}|\mathbf{D}\big) \quad \text{as } i \to \infty\,. \quad \text{(I)}$$

We use a threshold $\gamma = 1\times10^{-6}$ to stop the iteration, *i.e.*, if $\|\mathbf{T}^{(i)} - \mathbf{T}^{(i-1)}\| < \gamma$, then $\mathbf{T}^{(i)}$ is returned as the final POT solution for Eq. (6).

A.2. Dataset Statistics

Tab. A shows the statistics (*i.e.*, the numbers of training/test images, classes, and attributes in the attribute pool) of the five datasets used in our experiments. These datasets are designed for the few-shot or low-data setting, recognizing that most real-world applications cannot gather datasets at the scale of traditional benchmarks. As shown in Tab. B, 77 out of 111 object classes were never seen during pretraining of the vision foundation model we use. Tab. C provides the performance comparison on different few-shot settings.

Following former works [S4, S7, S8], attributes were generated by prompting a Large Language Model, *i.e.*, GPT-3.5, with known class names. These attributes are

Dataset (↓)	Train Images	Test Images	Classes	Attributes
Aquatic	318	319	7 (4+3)	385
Aerial	5000	5000	20 (10+10)	1229
Game	788	787	59 (30+29)	390
Medical	93	89	12 (6+6)	390
Surgery	912	917	13 (6+7)	808

Table A. Statistics of the five datasets. "7 (4+3)" in the "Classes" column means that there are a total of 7 classes in the Aquatic dataset, of which 4 classes are *known* (K) in Task 1 (or *previously known* (PK) in Task 2), 3 classes are *unknown* (U) in Task 1 (or *currently known* (CK) in Task 2), and so on.

Dataset (→)	Aquatic	Aerial	Game	Medical	Surgery
Known Classes	4 / 4	8 / 10	5 / 30	0 / 6	0 / 6
Unknown Classes	3 / 3	8 / 10	3 / 29	1 / 6	2 / 7

Table B. Statistics of the class overlap with the pretraining datasets of OWL-ViT [S5]. "5 / 30" in the "Game" column means that there are 30 known classes in the Game dataset, of which 5 classes are also in the pretraining datasets of OWL-ViT, and so on.

Dataset (→)	Aquatic				Aerial				Game				Medical				Surgery				Overall			
Task ID (→)	Task 1		Task 2		Task 1		Task 2		Task 1		Task 2		Task 1		Task 2		Task 1		Task 2		Task 1		Task 2	
	U	K	PK	CK	U	K	PK	CK	U	K	PK	CK	U	K	PK	CK	U	K	PK	CK	U	K	PK	CK
B/16 Backbone, 1-Shot:																								
BASE-FS	7.1	25.7	22.8	20.4	1.2	6.4	6.8	6.7	16.0	2.1	1.6	1.4	0.6	3.7	3.8	3.7	1.3	4.7	5.7	5.3	5.2	**8.5**	**8.1**	7.5
FOMO [S8]	14.2	22.2	18.4	19.9	2.0	6.1	6.0	6.3	5.9	1.7	1.2	1.1	0.8	3.9	2.9	2.8	2.3	5.4	5.3	5.3	5.0	7.9	6.8	7.1
PASS (Ours)	14.9	13.2	12.7	18.8	3.5	11.5	10.6	8.5	10.4	1.8	1.2	1.9	9.5	3.0	3.0	6.5	14.2	7.9	7.3	12.0	**10.5**	7.5	6.9	**9.5**
B/16 Backbone, 10-Shot:																								
BASE-FS	7.1	37.8	37.9	28.1	1.2	8.6	8.7	1.8	16.0	4.1	4.2	3.1	0.6	5.9	5.9	1.7	1.3	11.9	11.3	9.6	5.2	13.7	14.0	8.9
FOMO [S8]	10.0	37.4	36.2	29.8	1.3	9.5	9.8	2.0	11.0	3.8	4.1	3.3	4.8	5.9	6.0	1.2	13.1	12.9	14.4	10.7	8.0	13.9	14.1	9.4
PASS (Ours)	17.0	31.7	31.4	29.1	2.3	15.5	15.8	2.5	11.5	7.1	8.6	10.5	9.4	5.5	4.7	4.6	14.6	12.8	13.2	16.7	**11.0**	**14.5**	**14.8**	**12.7**
B/16 Backbone, 100-Shot:																								
BASE-FS	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 [S8]	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
PASS (Ours)	5.2	43.4	43.2	46.6	1.9	14.0	16.0	7.0	21.5	10.0	7.7	9.0	4.9	8.4	6.8	12.1	14.3	15.6	13.1	14.7	**9.6**	**18.3**	**17.4**	**17.9**
L/14 Backbone, 1-Shot:																								
BASE-FS	2.4	18.1	17.4	16.9	9.7	15.8	15.9	13.2	8.2	9.0	8.7	5.8	1.1	20.8	20.2	21.3	3.6	25.0	24.2	11.1	5.0	17.7	17.3	13.7
FOMO [S8]	18.0	18.1	17.4	17.0	3.1	15.6	15.7	12.7	28.3	7.2	5.2	4.6	6.1	20.5	14.8	22.7	11.5	25.0	24.3	11.4	13.4	17.3	15.5	13.7
PASS (Ours)	18.5	22.6	20.3	19.2	5.5	27.4	26.8	17.4	28.7	9.4	7.3	7.1	7.5	20.6	14.9	24.2	13.4	27.1	23.6	21.1	**14.7**	**21.5**	**18.6**	**17.8**
L/14 Backbone, 10-Shot:																								
BASE-FS	2.4	37.0	36.5	27.6	9.7	21.8	21.1	6.8	8.2	11.4	11.2	12.7	1.1	27.3	25.8	27.5	3.6	24.1	23.7	7.6	5.0	24.3	23.7	16.4
FOMO [S8]	12.8	37.2	36.5	27.6	5.6	22.0	21.5	7.9	30.3	11.6	10.6	11.2	13.6	25.4	24.0	33.0	11.3	26.8	28.0	11.3	14.7	24.6	24.1	18.2
PASS (Ours)	19.3	36.5	35.1	30.5	6.2	33.8	33.1	8.3	33.5	22.7	22.8	24.4	12.9	25.2	20.8	35.4	16.4	37.7	39.3	36.3	**17.6**	**31.2**	**30.2**	**27.0**
L/14 Backbone, 100-Shot:																								
BASE-FS	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 [S8]	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
PASS (Ours)	21.7	53.9	56.6	58.3	8.4	34.2	36.1	20.2	36.0	24.3	23.7	26.3	13.1	34.3	30.0	32.0	16.6	45.6	47.9	43.3	**19.1**	**38.5**	**38.9**	**36.0**

Table C. OWOD results with different few-shot regimes on the five real-world object detection datasets. The evaluation on each dataset is divided into two tasks, and we report U-, K-mAP for Task 1, and PK-, CK-mAP for Task 2, which are introduced in Sec. 5.1. **Best** overall results are highlighted in each column.

grouped into 10 different types: *shape*, *color*, *texture*, *size*, *context*, *features*, *appearance*, *behavior*, *environment*, and *material*. Accordingly, each attribute is described in the template of "`object which (is/has/etc) <TYPE> is <ATTRIBUTE>`", *e.g.*, `object which (is/has/etc) shape is straight`. Tab. G further illustrates the top selected attributes for each dataset using our proposed PASS.

A.3. Model Implementations

For few-shot training, we follow [S8] to feed an image and its corresponding ground truth bounding box into the pre-trained OWL-ViT [S5] model to generate predicted bounding boxes and class embeddings. The class embeddings were filtered based on their associated bounding boxes, ensuring that only those with an intersection over union (IoU) of at least 0.8 with the ground-truth object were retained. From the filtered class embeddings, we selected the one farthest from the mean of all the filtered embeddings to produce the final image output. During training, we also follow [S8] to optimize attributes with an additional adaptation loss that minimizes the domain gap between attributes and the extracted class embeddings. Prompt ensembling is also used to produce the final attribute embeddings, by averaging the text embeddings obtained from the 7 most effective CLIP prompt templates [S5, S6]. The trade-off parameter λ in Eq. (10) is set to 5, whose sensitivity is discussed in Fig. B.

B. Additional Results and Analysis

B.1. Results in More Few-Shot Settings

In real-world applications, low-data scenarios are often encountered, making it important to evaluate the performance of the proposed PASS and existing baselines under different few-shot settings. Tab. C presents the results for the 1-, 10-, and 100-shot regimes. In the 1-shot regime, the improvement of PASS over baselines is less pronounced, likely due to the extremely limited training data. However, in the 10- and 100-shot regimes, PASS demonstrates stronger performance with greater improvements over the baselines. This highlights the efficiency of our method in effectively utilizing additional training data, enabled by our proposed attribute curation strategy grounded in the POT theory.

B.2. Effect of Attribute Selection

B.2.1. Effect of Target Attribute Number

In our experiments, the target attribute number N' to be selected for each dataset is determined so that the average num-

Dataset (→)	**Aquatic**		**Surgery**	
	U	K	U	K
B/16 Backbone:				
PASS ($\bar{N}' = 10$)	3.7	42.8	**15.7**	12.8
PASS ($\bar{N}' = 25$, default)	**5.2**	**43.4**	14.3	**15.6**
PASS ($\bar{N}' = 40$)	4.8	40.3	10.9	12.4
L/14 Backbone:				
PASS ($\bar{N}' = 10$)	18.2	47.2	15.8	40.6
PASS ($\bar{N}' = 25$, default)	**21.7**	**53.9**	16.6	45.6
PASS ($\bar{N}' = 40$)	19.4	44.0	**17.7**	**46.5**

Table D. Effect of target attribute number. We report U-mAP (U) and K-mAP (K) on two representative datasets: Aquatic and Surgery, with Task 1 evaluation in the 100-shot regime. **Best** results are highlighted in each column.

p_{ID} (→)	**Max**	**Mean**	**Mah.**	**Max+Mean**	**Max+Mah.**
Known-mAP	45.6	45.4	45.8	45.6	45.9
Unknown-mAP	16.6	16.6	16.7	16.6	17.0

Table E. OWOD results using different choices of p_{ID}. Performance on Surgery (100-shot, Task 1) with L/14 backbone is reported.

Method (→)	**FOMO** [S8]				**PASS**
Training Stage (→)	Stage 1	Stage 2	Stage 3	Total	–
B/16 Backbone	01:22	00:01	04:36	05:59	**03:26**
L/14 Backbone	03:18	00:01	21:18	24:37	**05:35**

Table F. Training time comparison between FOMO [S8] and our proposed PASS on the Aquatic dataset (100-shot, Task 1).

ber of selected attributes per known class (denoted as $\bar{N}'$) is 25, following [S8]. Here, we analyze the impact of varying the number of target attributes. As presented in Tab. D, setting $\bar{N}'$ to 25 generally delivers strong performance in most scenarios. However, in certain cases, increasing the number of target attributes further enhances the results, highlighting the ability of our proposed PASS to effectively leverage a greater number of available attributes.

B.2.2. Effect of Curriculum Steps

As we incorporate a curriculum schedule into the attribute selection process during training, here we evaluate the impact of varying the curriculum step size (η). Fig. A illustrates the performance across different values of η. Notably, larger η results in a smoother selection process, as fewer attributes are filtered out at each step, and we can observe that a moderate η often achieves favorable results. Importantly, our method demonstrates relatively low sensitivity to this hyperparameter, highlighting the robustness of the proposed approach in effective attribute selection and optimization.

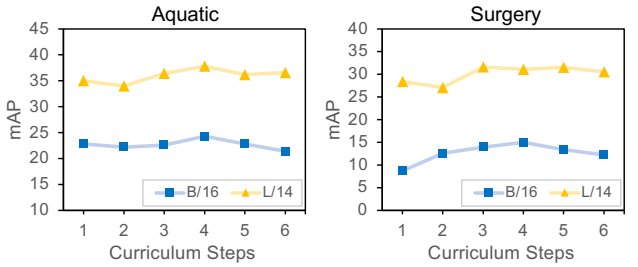

Figure A. Effect of curriculum steps. We report the average of U- and K-mAP on two representative datasets: Aquatic and Surgery, with Task 1 evaluation in the 100-shot regime.

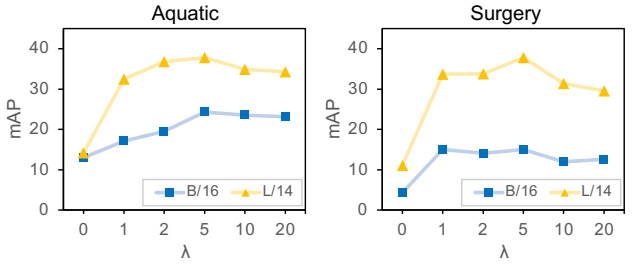

Figure B. Effect of λ in Eq. (10). We report the average of U- and K-mAP on two representative datasets: Aquatic and Surgery, with Task 1 evaluation in the 100-shot regime.

B.3. Additional Hyperparameter Analysis

We introduce an additional hyperparameter, λ, in Eq. (10) to balance the contributions of the classification loss and the POT loss. As shown in Fig. B, the performance varies with different λ values, with the best results generally achieved when λ is set to 5—ensuring that the POT loss is of a similar magnitude to the classification loss. Notably, our method demonstrates relatively low sensitivity to changes in λ, emphasizing its robustness in accurately delivering attributes to the need for specific detection tasks.

B.4. Choices of p_{ID}

While we adopt the choice of p_{ID} in Eq. (11) from FOMO [S8] for fair and direct comparisons, there are still alternative choices. Tab. E presents experimental results using *max/mean of attribute relevance* (Max/Mean), *Mahalanobis distance* (Mah.), and their combination (*average*). While Max and Mean yield similar mAPs, Mean often achieves higher Recalls but lower Precisions, probably due to its broader attribute dependence. Since these methods excel in different facets of the distribution, their combination can sometimes produce better overall results. We leave the exploration of more sophisticated p_{ID} designs in future work.

B.5. Training Time Comparison

We provide training time ([minutes]:[seconds]) comparison between FOMO [S8] and our proposed PASS in Tab. F using the same NVIDIA RTX A6000 GPU. We use FOMO's best

hyperparameters: 5K and 780 iterations for stage 1 and stage 3 training. For fair comparisons, PASS is also trained for 5K iterations, which likewise yields the best performance. While being efficient in stage 1 (*attribute selection*) and stage 2 (*attribute adaptation*), FOMO's stage 3 (*attribute refinement*) requires iteratively performing forward passes through the foundation model, significantly increasing the training time. Grounded in POT theory, our proposed PASS integrates all three stages into a single, efficient end-to-end training process, eliminating the need for redundant forward passes and significantly reducing training time while improving detection results.

B.6. More Qualitative Results

B.6.1. Detection Results

Fig. C presents additional visualizations of the detection results, complementing those in Fig. 3, with all results produced using the 100-shot L/14 models. Our proposed PASS consistently demonstrates improved detection performance compared to FOMO [S8], producing results that more closely align with ground-truth annotations. This highlights the effectiveness of PASS in selecting and optimizing relevant and useful attributes. Additionally, we include examples of failure cases in which neither FOMO nor PASS achieves satisfactory results. For instance, in the Medical dataset, accurately identifying finger bones remains challenging, likely due to the limited and ambiguous nature of the training data. These observations underscore the complexity of OWOD in real-world scenarios, highlighting that significant challenges persist and warrant further research in this domain.

B.6.2. Selected Attributes and ID Scores

Tab. G presents the top selected attributes identified by our proposed PASS method, with all results generated using the 100-shot L/14 models. Most of the selected attributes are found to be highly relevant and beneficial for specific object detection tasks. For instance, in the Surgery dataset, the shape attribute "`pointed tips`" can be effective in detecting surgical tools. These findings further demonstrate the effectiveness and interpretability of PASS in identifying both known and unknown object classes through the use of selected and optimized attributes.

References

[S1] Heinz H Bauschke and Adrian S Lewis. Dykstras algorithm with bregman projections: A convergence proof. *Optimization*, 48(4):409–427, 2000. 1

[S2] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015. 1

[S3] Richard L Dykstra. An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, 78(384):837–842, 1983. 1

[S4] Sachit Menon and Carl Vondrick. Visual classification via description from large language models. In *International Conference on Learning Representations (ICLR)*, 2023. 1

[S5] Matthias Minderer, Alexey Gritsenko, Austin Stone, Maxim Neumann, Dirk Weissenborn, Alexey Dosovitskiy, Aravindh Mahendran, Anurag Arnab, Mostafa Dehghani, Zhuoran Shen, et al. Simple open-vocabulary object detection with vision transformers. In *European Conference on Computer Vision (ECCV)*, pages 728–755, 2022. 1, 2

[S6] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning (ICML)*, pages 8748–8763, 2021. 2

[S7] Orr Zohar, Shih-Cheng Huang, Kuan-Chieh Wang, and Serena Yeung. Lovm: Language-only vision model selection. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2023. 1

[S8] Orr Zohar, Alejandro Lozano, Shelly Goel, Serena Yeung, and Kuan-Chieh Wang. Open world object detection in the era of foundation models. *arXiv preprint arXiv:2312.05745*, 2023. 1, 2, 3, 4, 5

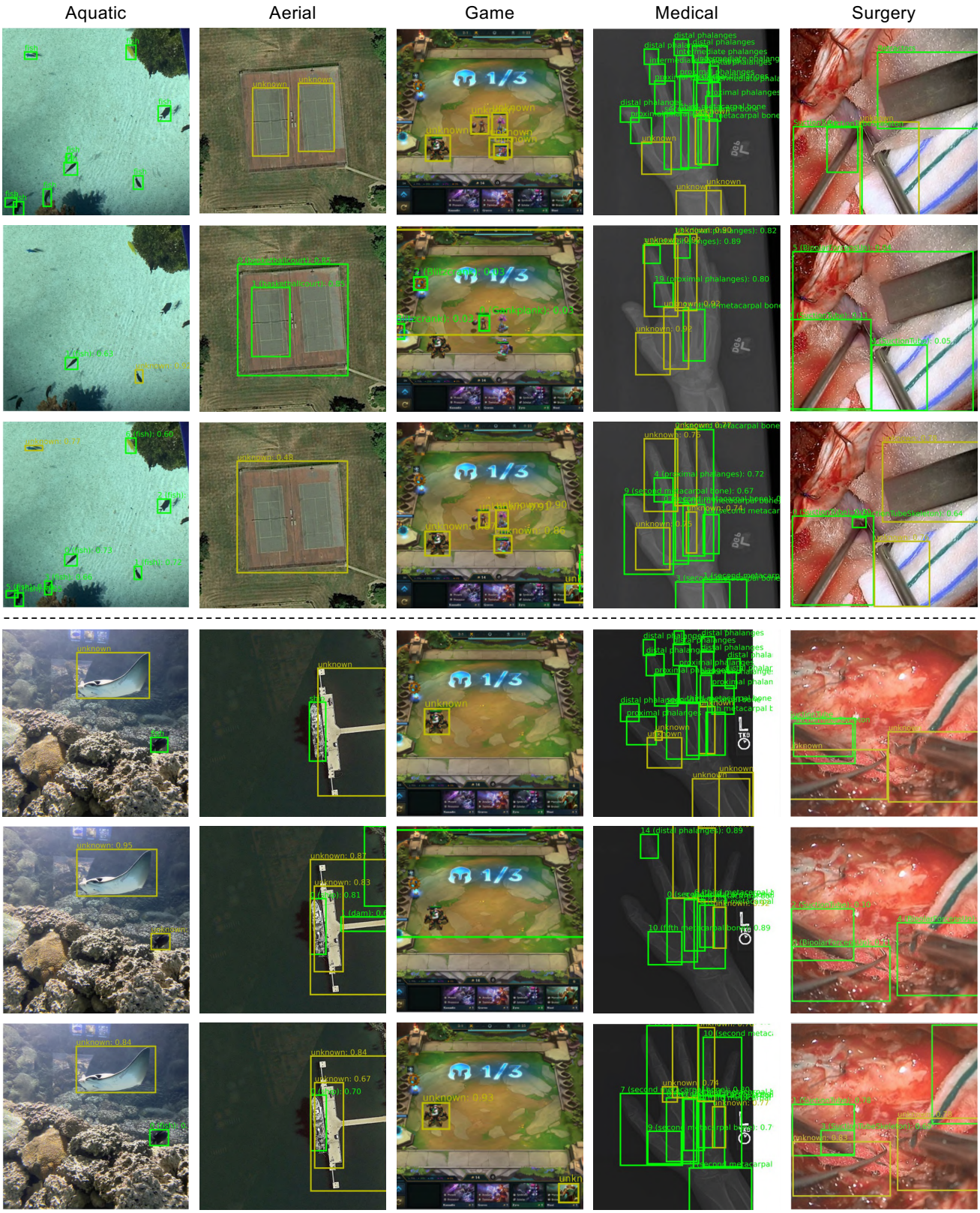

Figure C. Visualization of detection results on the five real-world datasets. **First Row:** images in each dataset with ground truth bounding boxes and class names. **Second Row:** Detection results using FOMO [S8]. **Third Row:** Detection results using our proposed PASS. The **Fourth** to **Sixth Rows** follow the same pattern. We use green and yellow boxes to indicate known and unknown objects, respectively.

Attribute Type (↓)	ID Score 1	Selected Attribute 1	ID Score 2	Selected Attribute 2	Unselected Attribute
Aquatic:					
Shape	1.00	*caudal fin shape*	1.00	*dorsal fin shape*	*straight*
Color	0.76	*yellow*	0.41	*black*	*pink*
Texture	1.00	*ridged*	1.00	*jelly-like*	*joint texture*
Size	1.00	*asymmetric*	1.00	*pectoral fins*	*oval*
Context	1.00	*aquatic*	1.00	*underwater*	*cliffs*
Features	1.00	*rounded snout*	1.00	*diving ability*	*ability to fly*
Appearance	1.00	*skeletal structure*	1.00	*lack of wheels*	*oral surface*
Behavior	0.61	*disjointed*	–	–	*flying*
Environment	1.00	*artificial reef*	1.00	*continental shelf*	*cave*
Material	1.00	*bony*	0.61	*protein*	*magnesium*
Aerial:					
Shape	1.00	*equilateral*	1.00	*slanted*	*elastic*
Color	1.00	*turquoise*	0.67	*maroon*	*pink*
Texture	1.00	*turf*	1.00	*spotted*	*marbled pattern*
Size	0.17	*tracks*	0.17	*smoke emitting*	*tiny*
Context	1.00	*warning track*	1.00	*construction sites*	*sky*
Features	1.00	*spires*	1.00	*theater*	*kitchen*
Appearance	1.00	*presence of parking lots*	1.00	*circular base*	*lanterns*
Behavior	1.00	*tournament*	0.34	*docking*	*crawling*
Environment	1.00	*downtown*	1.00	*town*	*arctic*
Material	0.84	*soil*	0.17	*gravel*	*pvc*
Game:					
Shape	1.00	*thin*	1.00	*fat*	*oval*
Color	1.00	*gray*	1.00	*shadow*	*hand*
Texture	1.00	*patterned*	1.00	*grooved*	*lined*
Size	1.00	*small*	1.00	*medium*	*width of phalanges*
Context	1.00	*torso*	1.00	*folds*	*presence of artifacts*
Features	1.00	*stripes*	1.00	*lunate*	*tentacles*
Appearance	1.00	*edges*	1.00	*contour*	*presence of arthritis*
Behavior	1.00	*healed*	1.00	*extension*	*contracture*
Environment	1.00	*dangerous*	1.00	*deteriorated*	–
Material	1.00	*iron*	1.00	*opaque*	*lamellae*
Medical:					
Shape	1.00	*knobby*	1.00	*straight*	*spiral*
Color	1.00	*contrast*	1.00	*shadow*	*green*
Texture	1.00	*lined*	1.00	*fissured*	*scaly*
Size	1.00	*proportionality*	1.00	*large*	*volume*
Context	1.00	*wrist bones*	1.00	*adjacent bones*	*wrinkles*
Features	1.00	*index middle phalanx*	1.00	*middle proximal phalanx*	*paws*
Appearance	1.00	*presence of fractures*	1.00	*presence of bone deformities*	*presence of wheels*
Behavior	1.00	*hypersupination*	1.00	*extension*	*inflamed*
Environment	1.00	*sparse*	–	–	*deteriorated*
Material	1.00	*hydroxyapatite*	1.00	*transparent*	*ligament*
Surgery:					
Shape	1.00	*pointed tips*	1.00	*microscopic*	*wavy*
Color	1.00	*silver color*	0.97	*stainless steel*	*pink*
Texture	1.00	*knurled*	1.00	*textured grip*	*bumpy*
Size	–	–	–	–	*narrow*
Context	1.00	*surgical bone mallet*	1.00	*surgical bone clips*	*monitor screens*
Features	1.00	*tension adjustment*	1.00	*adjustable*	*fiber optic*
Appearance	1.00	*sharp*	1.00	*shiny*	*transparent*
Behavior	1.00	*stabilizing*	1.00	*twistable*	*coagulating*
Environment	1.00	*surgical tools*	0.99	*mobile surgical unit*	*non-sterile environment*
Material	1.00	*cutting edge*	1.00	*precise*	*paper*

Table G. Selected attributes with in-distribution (ID) scores in the five datasets. Attributes are grouped into 10 different types, and we show top-2 selected attributes per each type with the highest ID scores, and 1 representative unselected attribute with zero ID score.