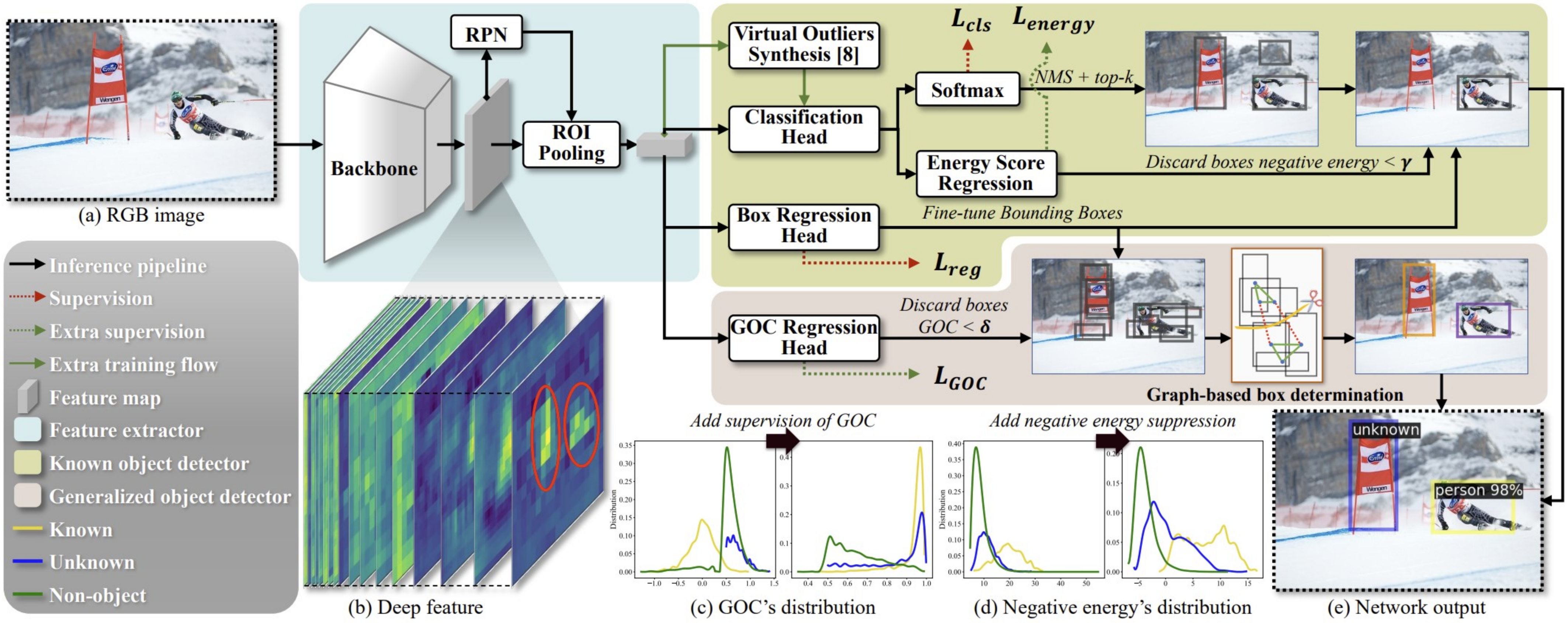
Unknown Sniffer for Object Detection: Don’t Turn a Blind Eye to Unknown Objects

Wenteng Liang, Feng Xue, Yihao Liu, Guofeng Zhong, Anlong Ming*

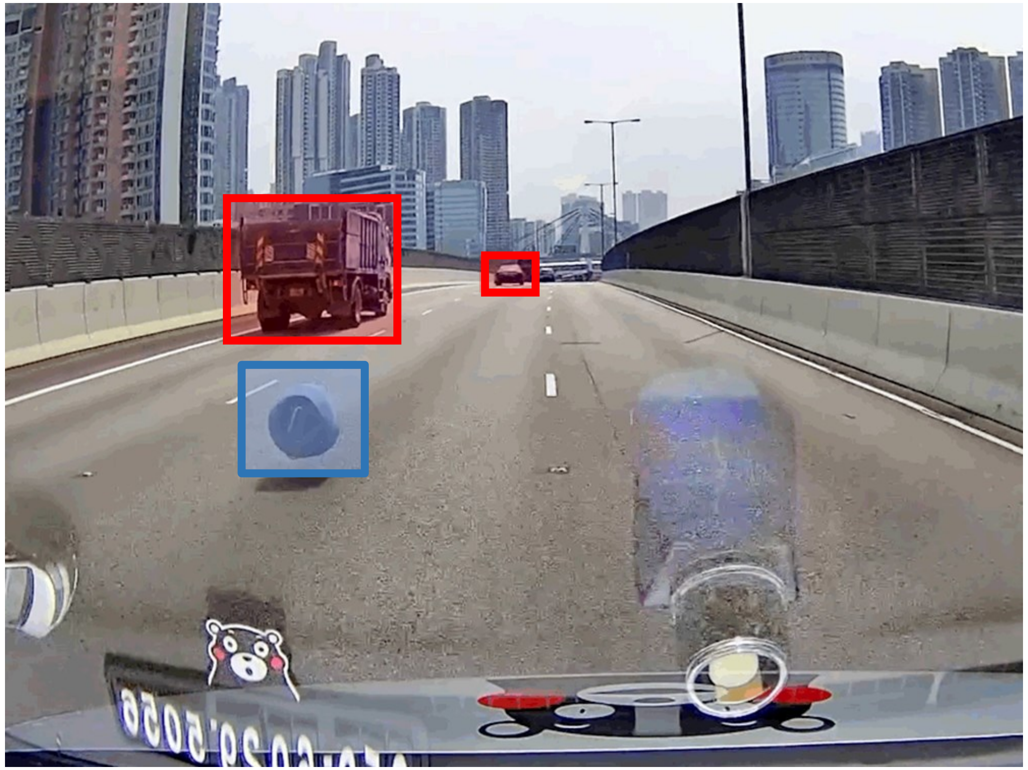
Beijing University of Posts and Telecommunications

TUE-AM-307

Overview


RPN
Backbone
ROI Pooling
Virtual Outliers Synthesis [8]
Classification Head
Softmax
Energy Score Regression
Box Regression Head
GOC Regression Head
L_{cls}
L_{energy}
L_{reg}
L_{GOC}
NMS + top-k
Discard boxes negative energy < γ
Fine-tune Bounding Boxes
Discard boxes GOC < δ
Graph-based box determination
Add supervision of GOC
Add negative energy suppression
Inference pipeline
Supervision
Extra supervision
Extra training flow
Feature map
Feature extractor
Known object detector
Generalized object detector
Known
Unknown
Non-object
unknown
person 98%
(a) RGB image
(b) Deep feature
(c) GOC's distribution
(d) Negative energy's distribution
(e) Network output

Breaking through the closed-world setting



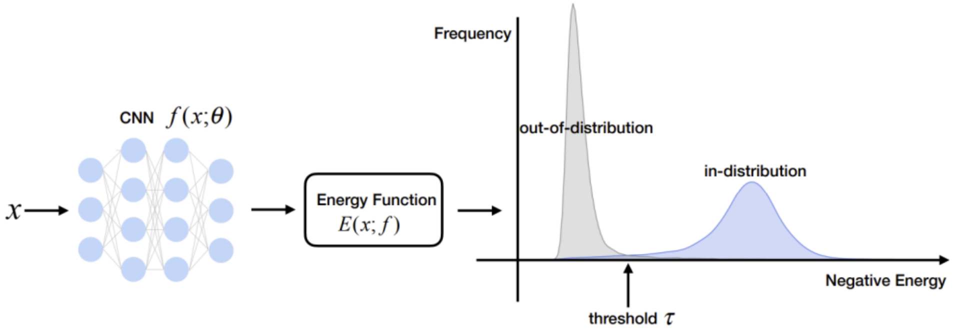
Known object   Unknown object

- In a **closed-world** with a limited number of categories, deep learning has achieved great success in detection tasks.

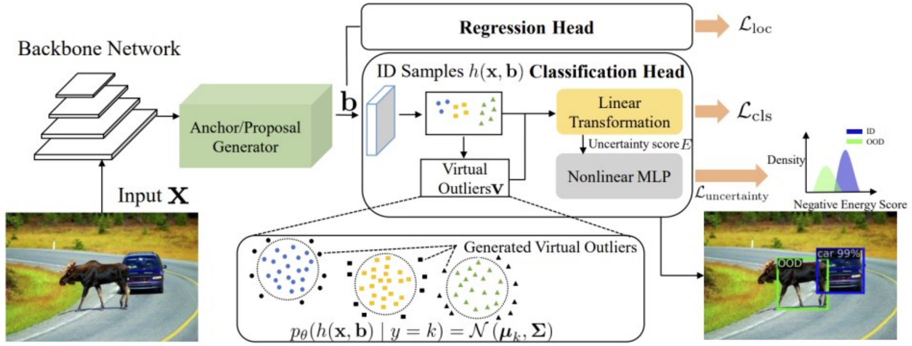
- However, models based on the idealized assumption have been unable to meet **complex real-world** needs.

Previous Methods

- **Open-set classification & detection**

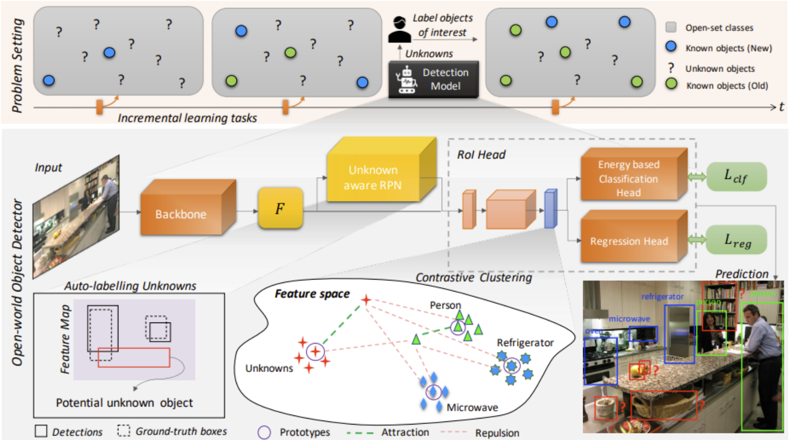


[Energy. NIPS 2020]

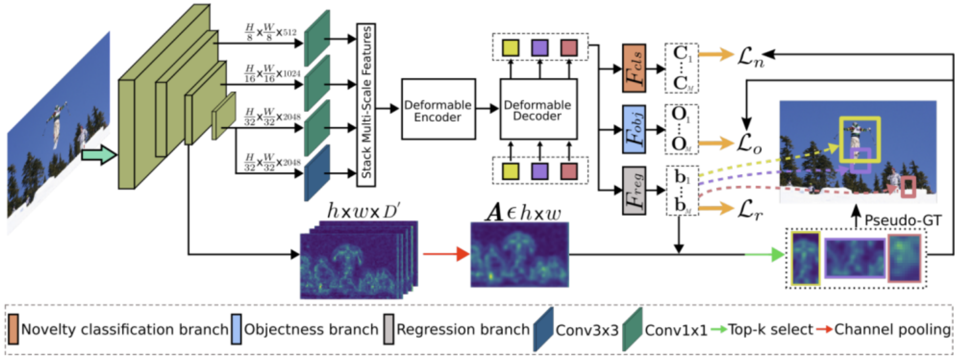


[VOS. ICLR 2022]

- **Open-world object detection**



[ORE. CVPR 2021]



[OW-DETR. CVPR 2022]

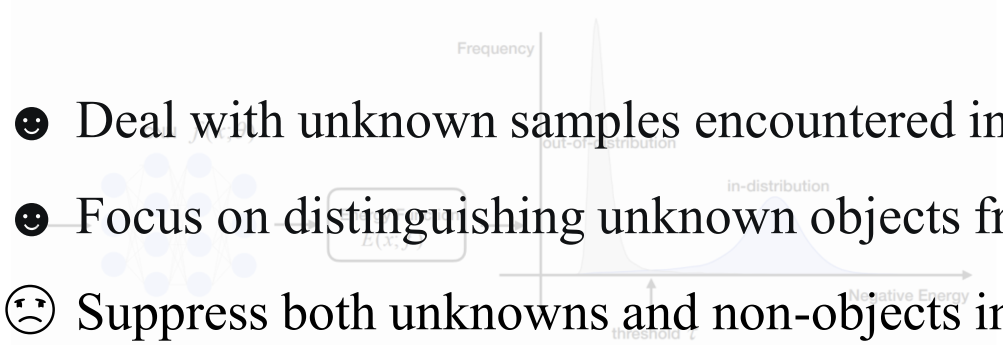
Previous Methods

■ **Open-set classification & detection**

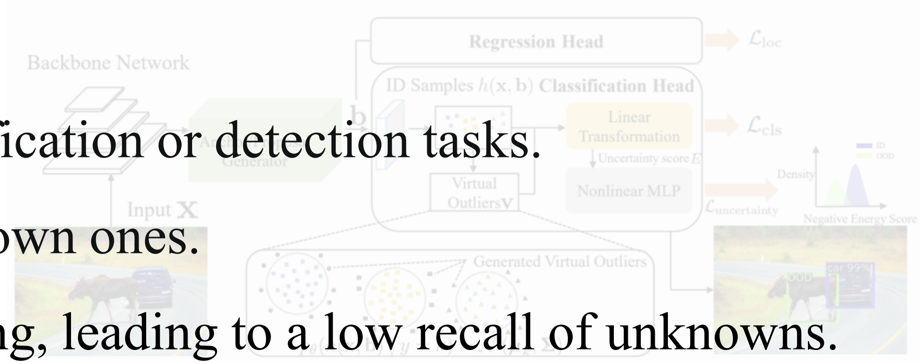
☻ Deal with unknown samples encountered in classification or detection tasks.

☻ Focus on distinguishing unknown objects from known ones.

☹ Suppress both unknowns and non-objects in training, leading to a low recall of unknowns.



[Energy. NIPS 2020]



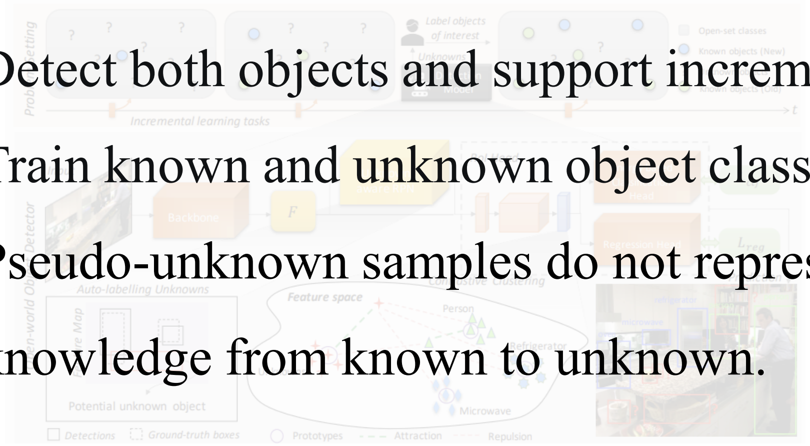
[VOS. ICLR 2022]

■ **Open-world object detection**

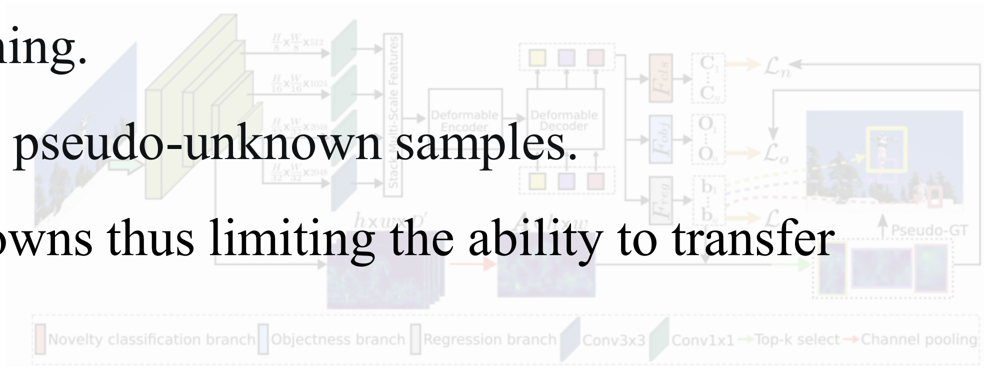
☻ Detect both objects and support incremental learning.

☻ Train known and unknown object classifiers with pseudo-unknown samples.

☹ Pseudo-unknown samples do not represent unknowns thus limiting the ability to transfer knowledge from known to unknown.

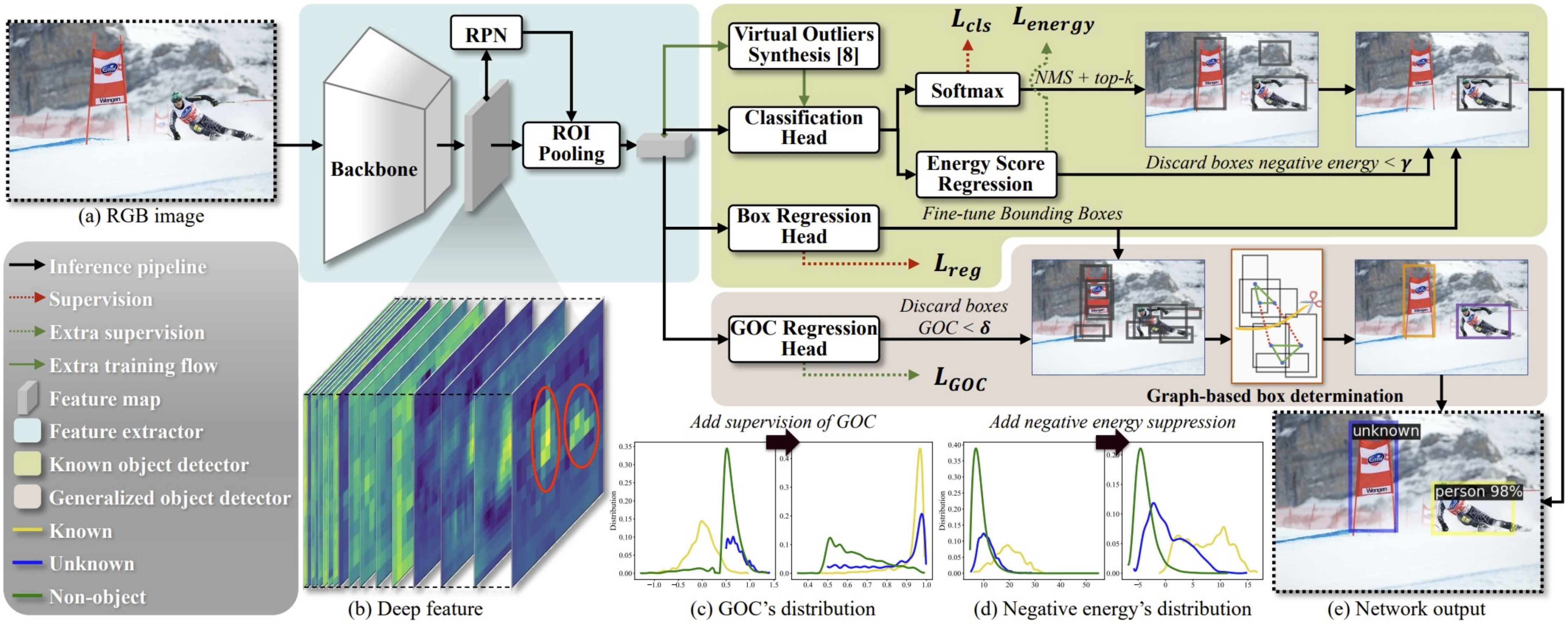


[ORE. CVPR 2021]



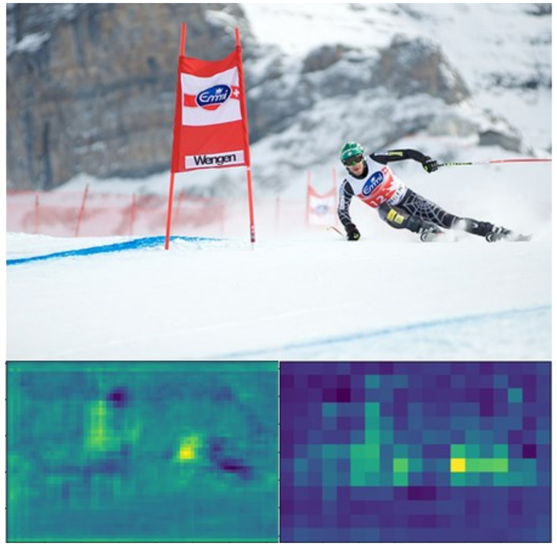
[OW-DETR. CVPR 2022]

Overview

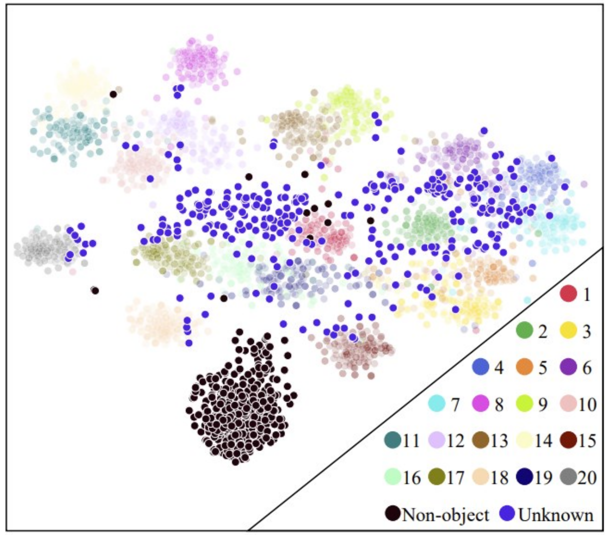


- Generalized Object Confidence (GOC)
- Negative Energy Suppression
- Graph-Based Box Determination

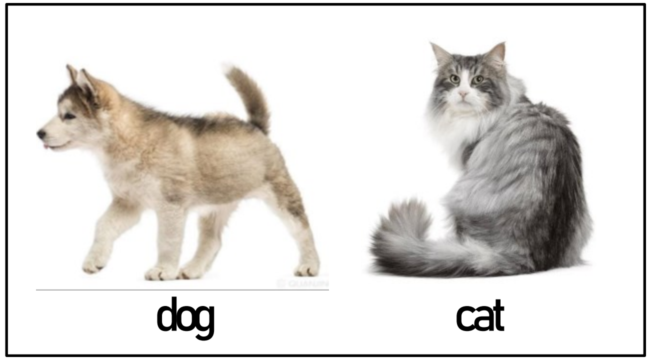
Generalized Object Confidence

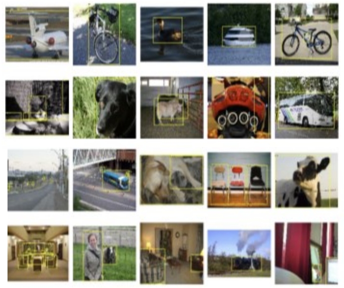


Many unknowns are encoded by deep feature.



Similarity between “objects”

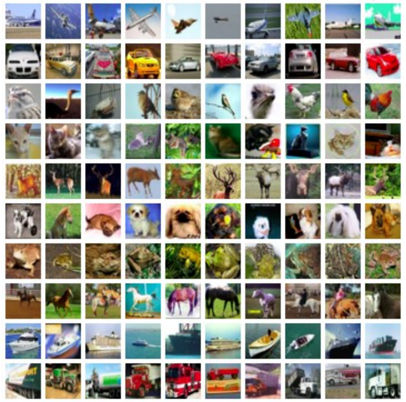




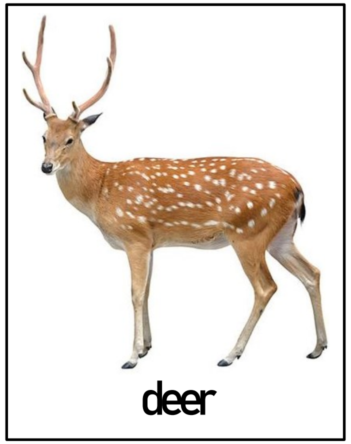
Known Objects

GOC

“Analogy”



Unknown Objects



Generalized Object Confidence

- Complete-object bounding boxes should have high GOC:

$$L_{pos}=\frac{1}{K}\sum_{k\in[1,K]}\frac{1}{|B_n^{k,\mathbf{c}}|}\sum_{b_i\in\mathbf{B}_n^{k,\mathbf{c}}}\left(\Phi(f_i)-1\right)^2$$

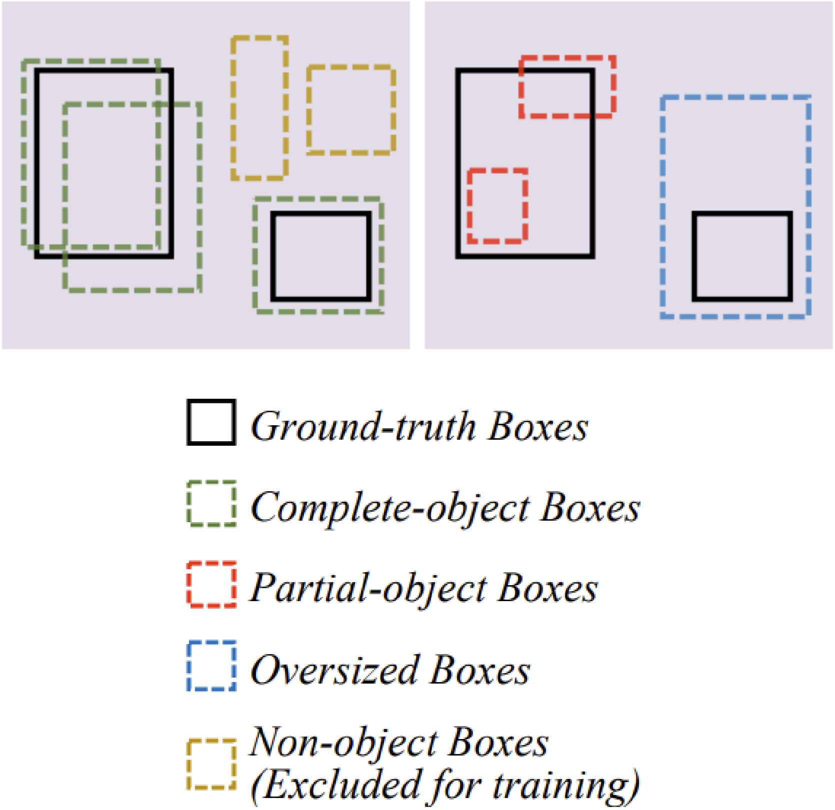
- More complete boxes have higher GOC:

$$L_{con}=\frac{1}{K}\sum_{k\in[1,K]}\left\lfloor\frac{2}{|\mathbf{B}_n^{k,\mathbf{c}}|}\right\rfloor\sum_{b_i,b_j\in\mathbf{B}_n^{k,\mathbf{c}}}\max\left(0,\frac{\Phi(f_i)-\Phi(f_j)}{\alpha}+\zeta\right)$$

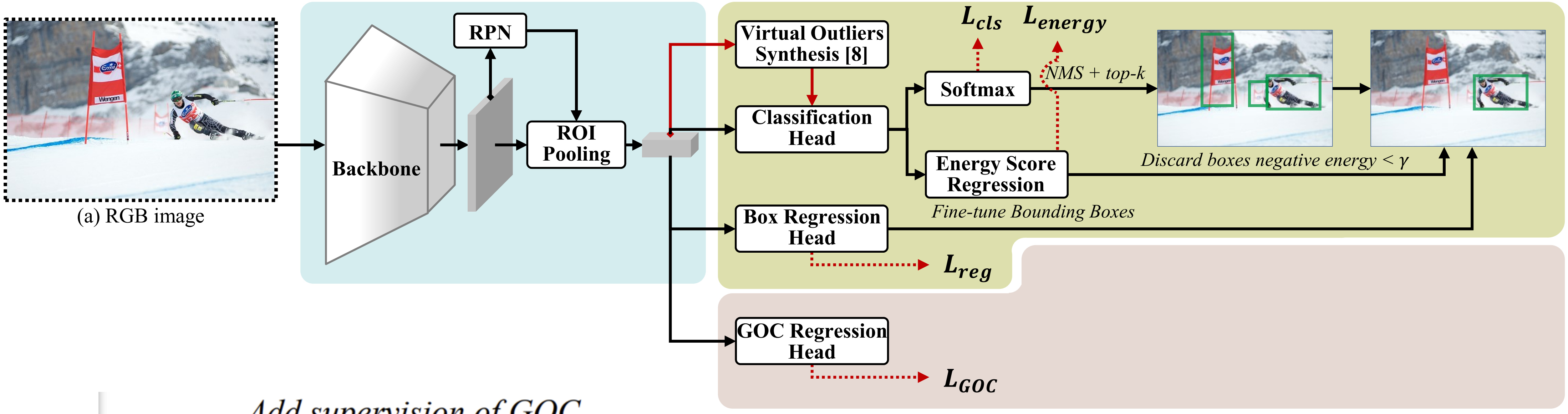
- Partial-object or oversized boxes should have low GOC:

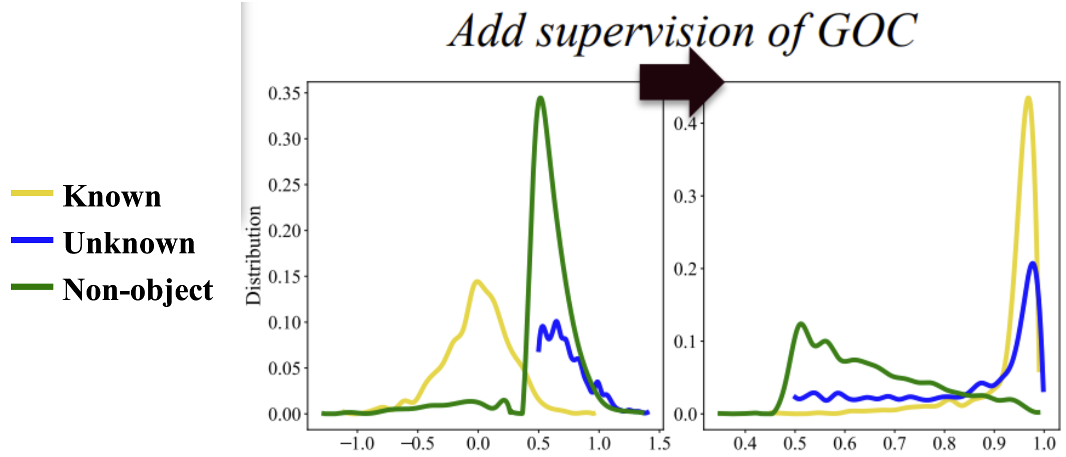
$$L_{neg}=\frac{1}{K}\sum_{k\in[1,K]}\frac{1}{|B_n^{k,\mathbf{po}}|}\sum_{b_i\in\mathbf{B}_n^{k,\mathbf{po}}}\max\left(0,\Phi(f_i)-\delta\right)$$

Total GOC loss $$L_{GOC}=L_{neg}+L_{pos}+L_{con}$$



Generalized Object Confidence



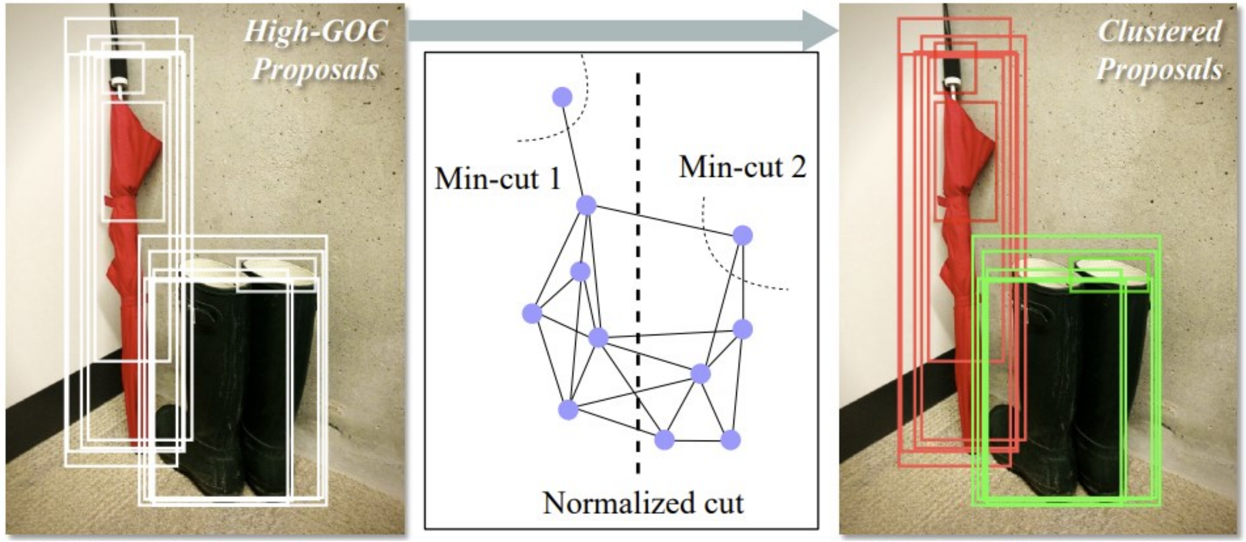


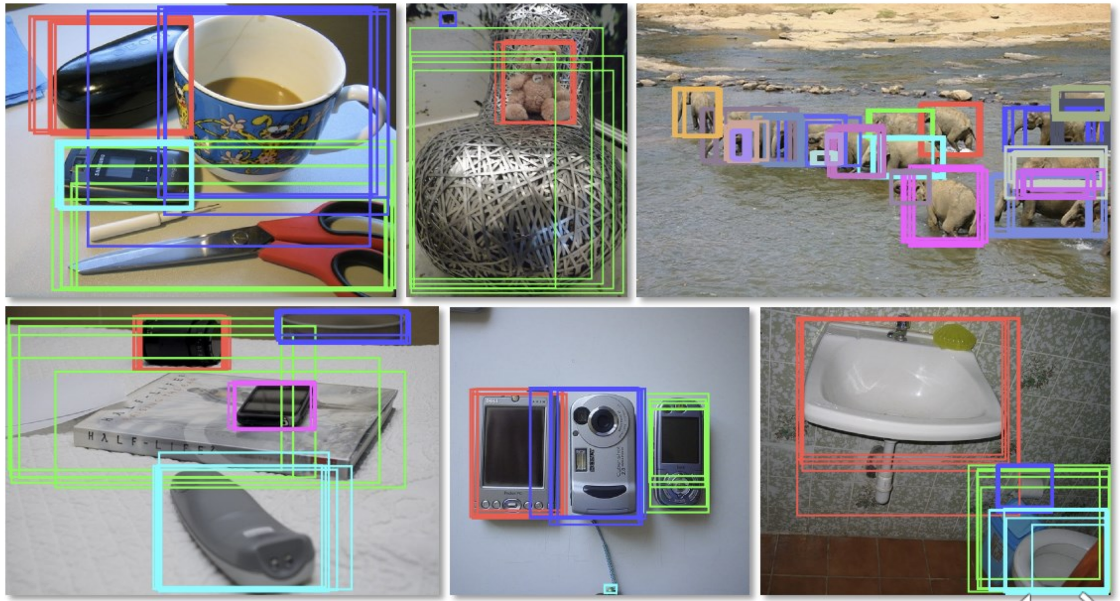
GOC's distribution

- Without L_{GOC}, unknown boxes have the same output as non-object.
- With L_{GOC}, unknown boxes have the same output as known objects.

Graph-based Top-scoring Box Determination

The top-scoring box determination $\longrightarrow$ Graph partitioning $\begin{cases} \text{Node represents a box} \\ \text{Edge represents the IoU} \end{cases}$



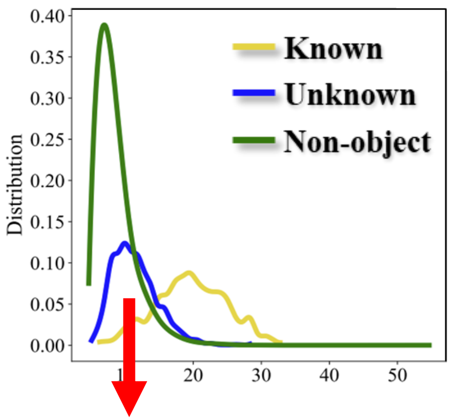


Negative Energy Suppression for Non-object

We follow the energy value of VOS to distinguish unknown objects from known ones：

$$E(b_i) = -\log \sum_{c\in[1,C]} \mathbf{w}_c \cdot \exp^{\mathbf{f}_c}$$

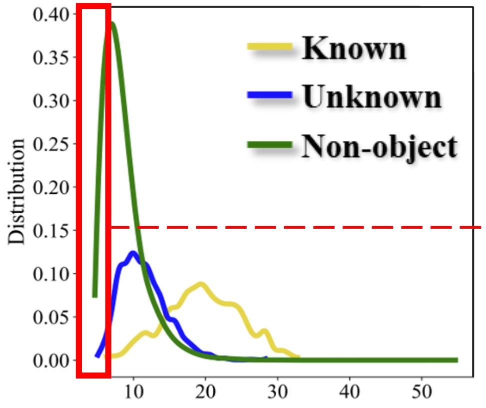
Negative energy's distribution



Indistinguishable

How to separate non-object from objects without the aid of non-object annotations ?

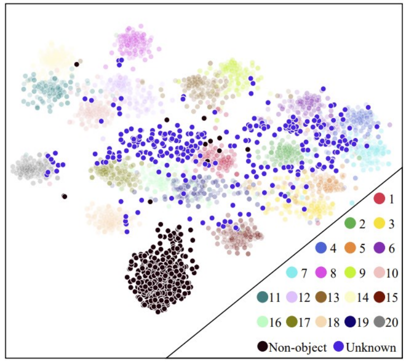
Negative Energy Suppression for Non-object



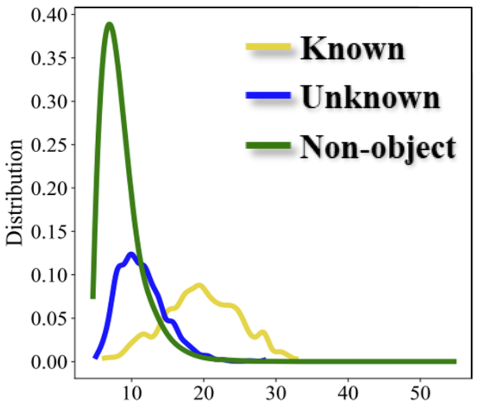
- Suppression loss is applied to the top-T proposals with the lowest negative energy scores.

$$L_{suppression} = \frac{1}{T} \sum\nolimits_{i \in [1,T]} \max(0, -E(b_i))$$

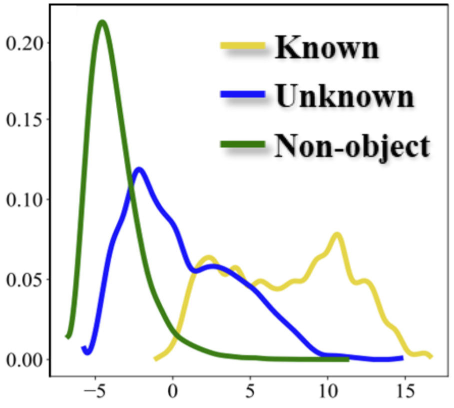
T-SNE



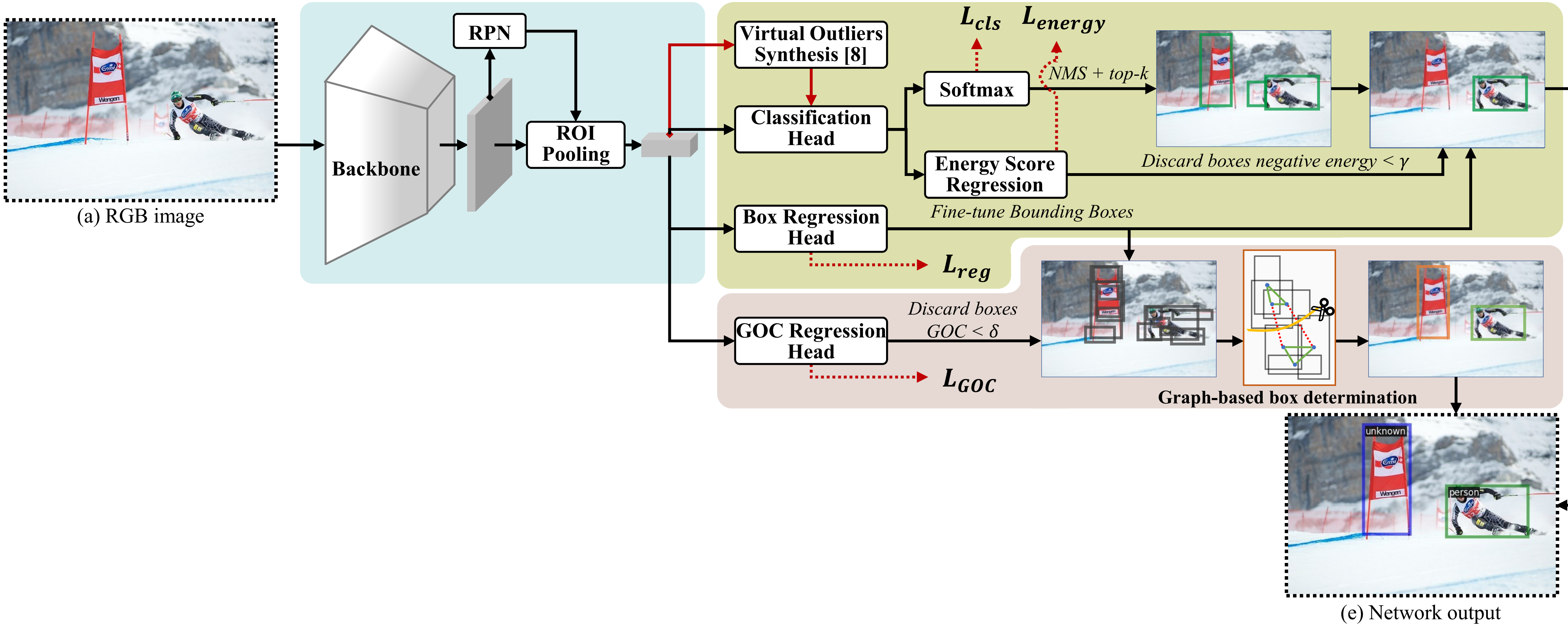
- Non-objects also have feature similarities, so **suppression will be transmitted.**



Indirectly widening the gap



Negative Energy Suppression for Non-object



The overall energy loss consists of our proposed $L_{suppression}$ and $L_{uncertainty}$ defined by VOS :

$$L_{energy} = L_{suppression} + L_{uncertainty}$$

Unknown Object Detection Benchmark

- **Training data**:

  Pascal VOC dataset that contains 20 categories

- **Testing data**:

Datasets	Images	Known	Unknown
VOC-Pretest	200	5.09	0
VOC-Test	4952	3.02	0
COCO-OOD♣	504	0	3.28
COCO-Mixed♣	897	2.96	2.82

♣ denotes the augmented datasets

Annotated samples in COCO-OOD and COCO-Mix.

Unknown Object Detection Benchmark

Evaluation Metrics

For known objects:

- mAP

For unknown objects:

- Unknown Average Precision (U-AP)
- Unknown F1-Score(U-F1)
- Unknown Recall Rate (U-REC)
- Precision Rate of Unknown (U-PRE)

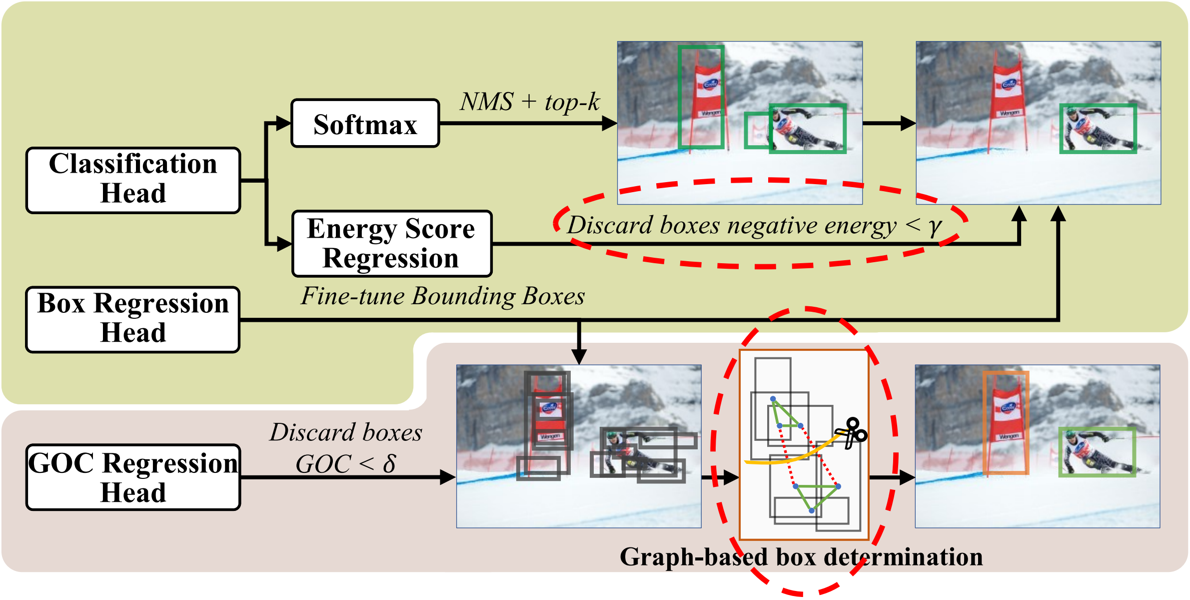
For mixed data:

- Absolute Open-Set Error (A-OSE)

Pretest mode in the Benchmark

Datasets	Images	Known	Unknown
VOC-Pretest	200	5.09	0

- Select 200 images in the training set that do not contain any potential unknown objects.



1. The threshold γ is set by making 95% of predicted proposals have a negative energy score greater than it.

2. The threshold of NCut is determined when the AP of known objects is the largest.

Experiment

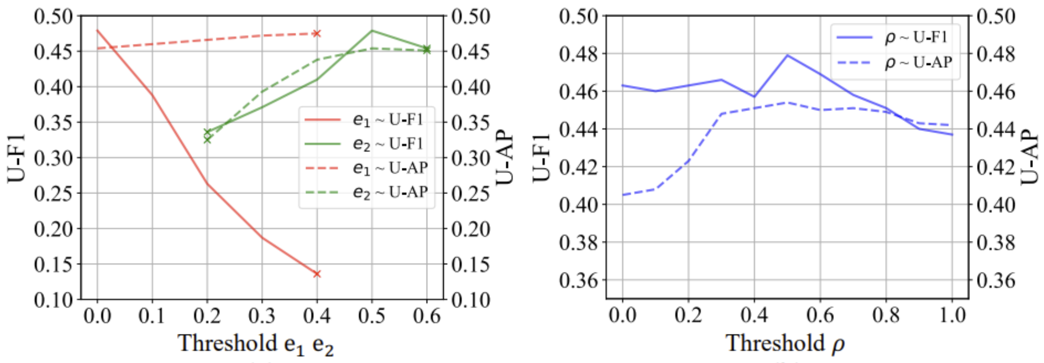
Groups	Methods	VOC-Test	COCO-OOD				COCO-Mix					
		mAP	U-AP	U-F1	U-PRE	U-REC	mAP	U-AP	U-F1	U-PRE	U-REC	AOSE
①	MSP [15]	0.470	0.213	0.314	0.279	0.359	0.364	0.055	0.169	0.190	0.153	588
①	Mahalanobis [5]	0.447	0.129	0.271	0.309	0.241	0.351	0.051	0.149	0.207	0.116	604
①	Energy score [23]	0.474	0.213	0.308	0.260	0.377	0.364	0.048	0.169	0.167	0.171	470
②	OW-DETR [11]	0.420	0.033	0.056	0.030	0.380	**0.414**	0.007	0.025	0.014	0.161	569
②	ORE [17]	0.243	0.214	0.255	0.153	**0.782**	0.213	0.140	0.175	0.103	**0.592**	485
③	VOS[1] [8]	**0.485**	0.135	0.196	0.342	0.137	0.377	0.040	0.101	**0.262**	0.062	640
③	VOS[2] [8]	0.469	0.205	0.317	0.291	0.348	0.364	0.051	0.172	0.184	0.163	409
④	Ours	0.464	**0.454**	**0.479**	**0.433**	0.535	0.359	**0.150**	**0.287**	0.222	0.409	**398**

Table 2. Comparisons with the detector using open-set classification ①, open-world object detection ②, and open-set detection ③ methods. VOS[1] denotes the model with the threshold given by the official repository[1], which is calculated on the BDD100K dataset [39]. And VOS[2] utilizes the threshold computed on the COCO-OOD dataset by using the official code[2]. Best results are in bold, second best are underlined.

Experiment

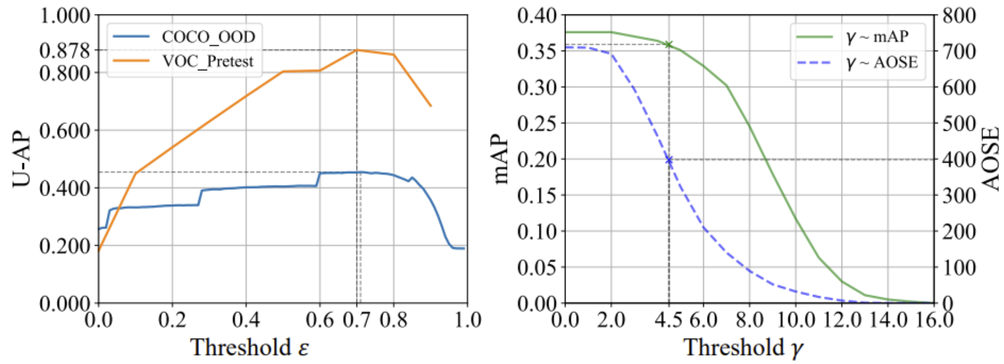
Row	GOC	NES	GBD	U-AP	U-F1	U-PRE	U-REC
1	×	×	×	0.066	0.050	0.026	0.808
2	×	×	✓	0.250	0.434	0.395	0.481
3	×	✓	×	0.442	0.054	0.028	0.861
4	✓	×	×	0.479	0.323	0.215	0.646
5	×	✓	✓	0.409	0.467	**0.437**	0.502
6	✓	×	✓	**0.455**	0.454	0.399	0.528
7	✓	✓	✓	0.454	**0.479**	0.433	**0.535**

Table 3. **Ablation studies on COCO-OOD**. GOC, NES and GBD refer to 'generalized object confidence', 'negative energy suppression' and 'graph-based box determination', respectively. When 'GBD' is ×, we use NMS and top-k as post-processing with the same thresholds with the known detector for a fair comparison.



(a) (b)

Figure 6. **Sensitivity analysis on (a) thresholds e_1, e_2, and (b) threshold ρ.** × indicates the failed training outside this threshold.

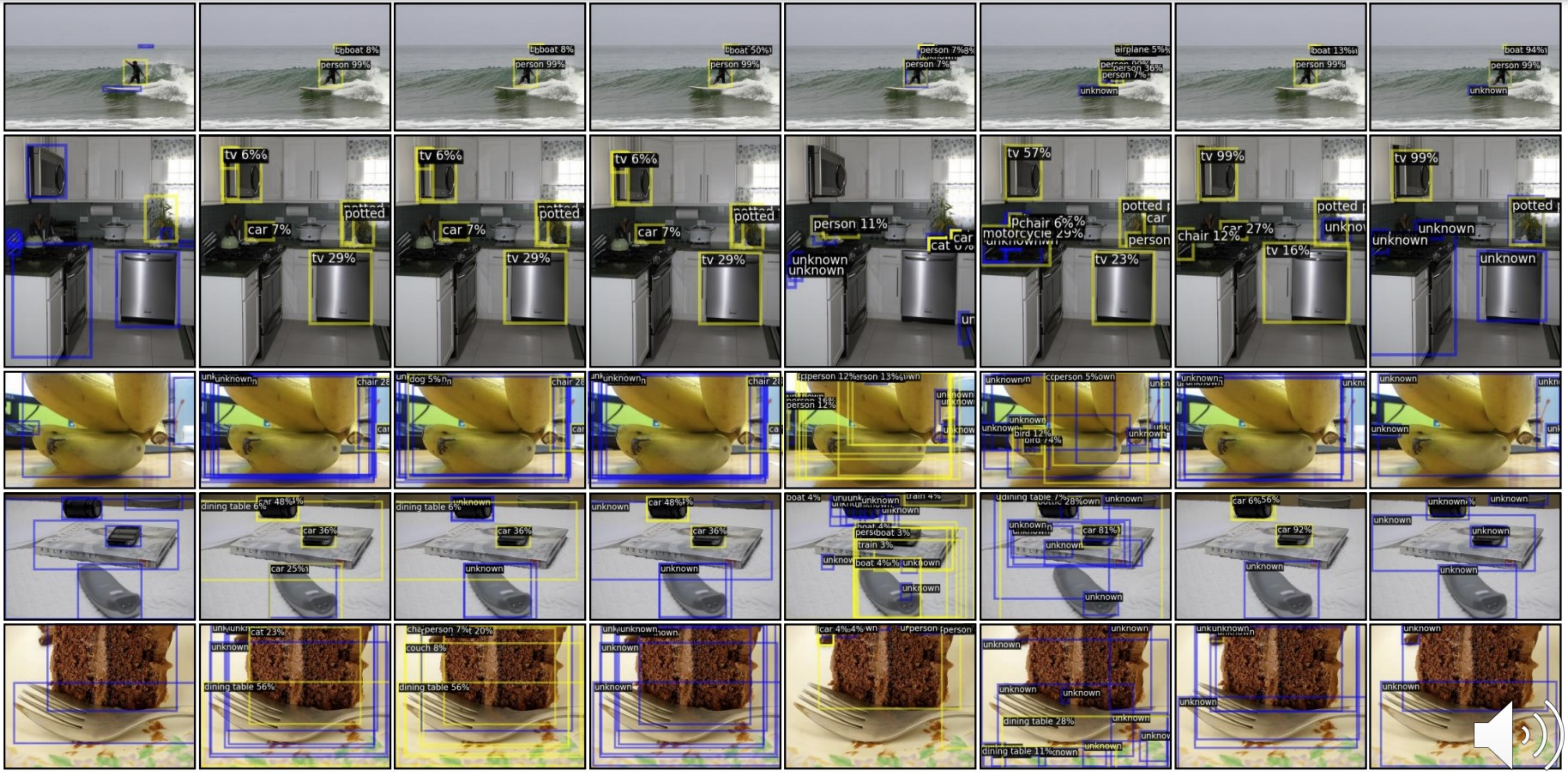


(a) (b)

Figure 8. (a) Comparison between the thresholds ε determined in the pretest set and the COCO-OOD dataset. (b) Comparison of mAP and AOSE between the thresholds determined in the pretest set (dot) and the COCO-Mix dataset (line).

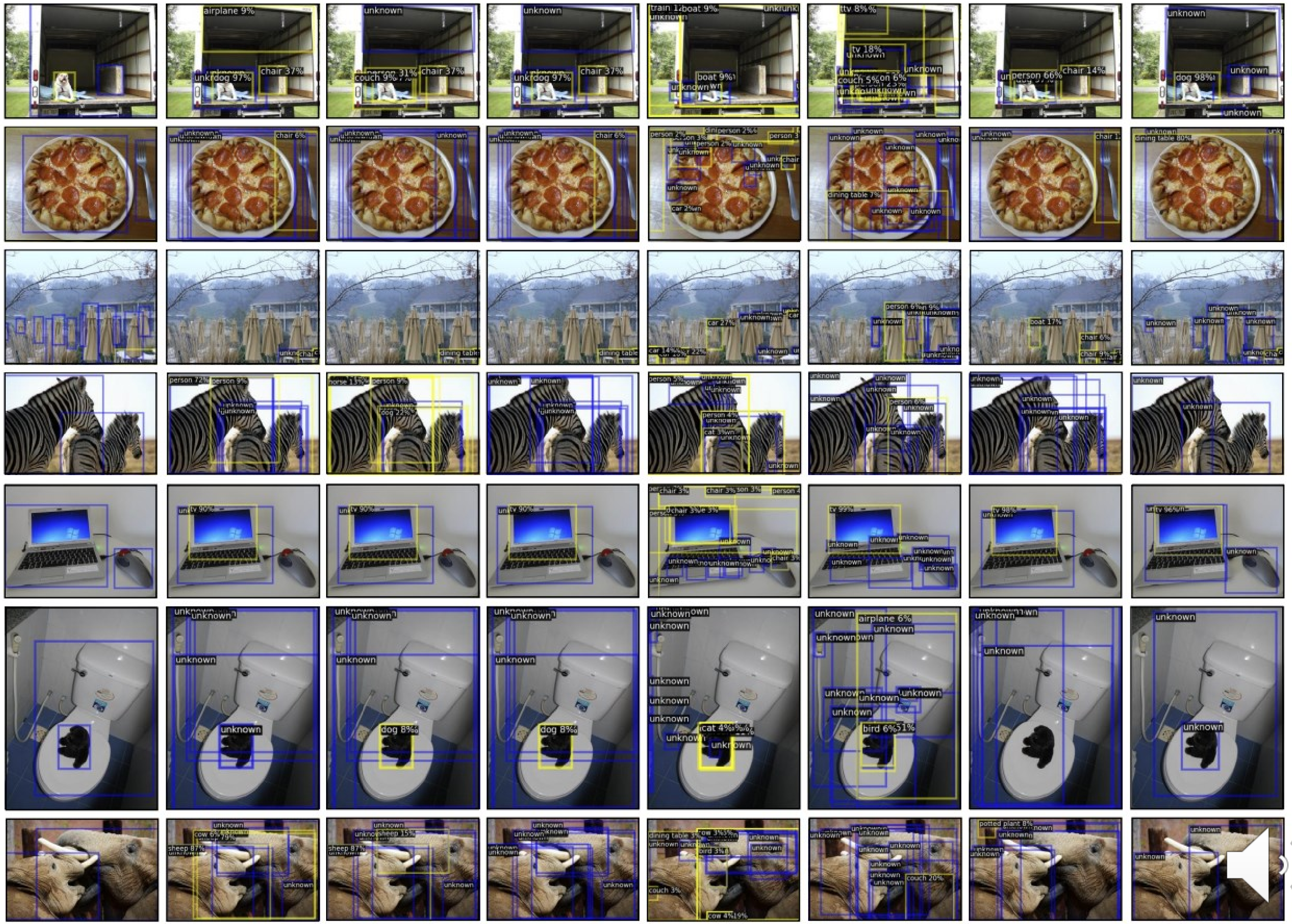
Detection Results and Video Demos

Visualization



Ground Truth | MSP | Mahalanobis | Energy score | OW-DETR | ORE | VOS | Ours

Visualization

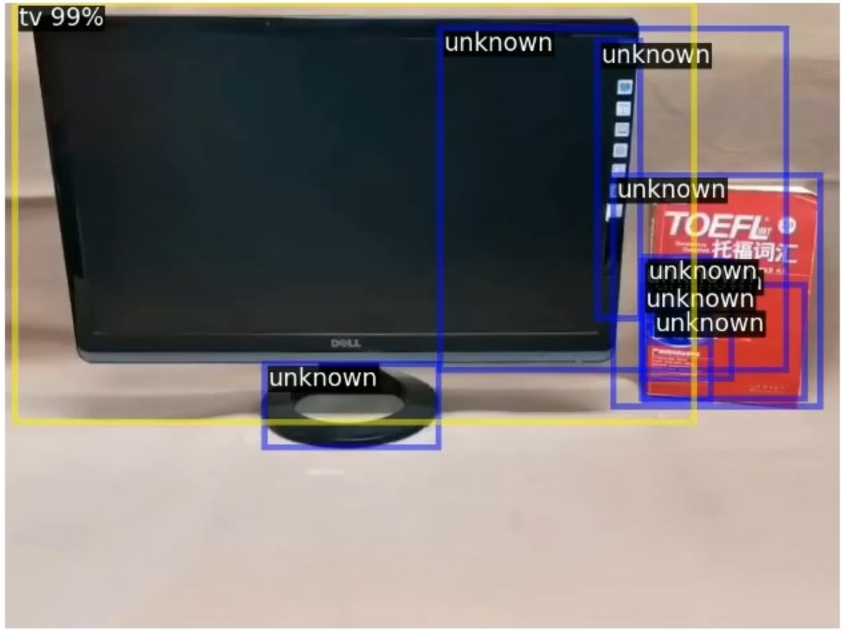


Ground Truth | MSP | Mahalanobis | Energy score | OW-DETR | ORE | VOS | Ours

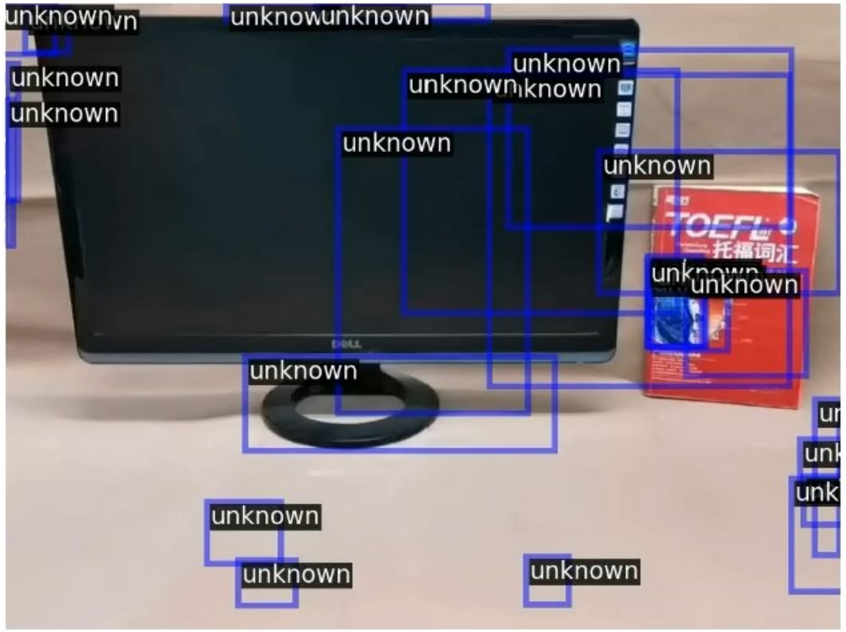
Video Demo1

UnSniffer

VOS

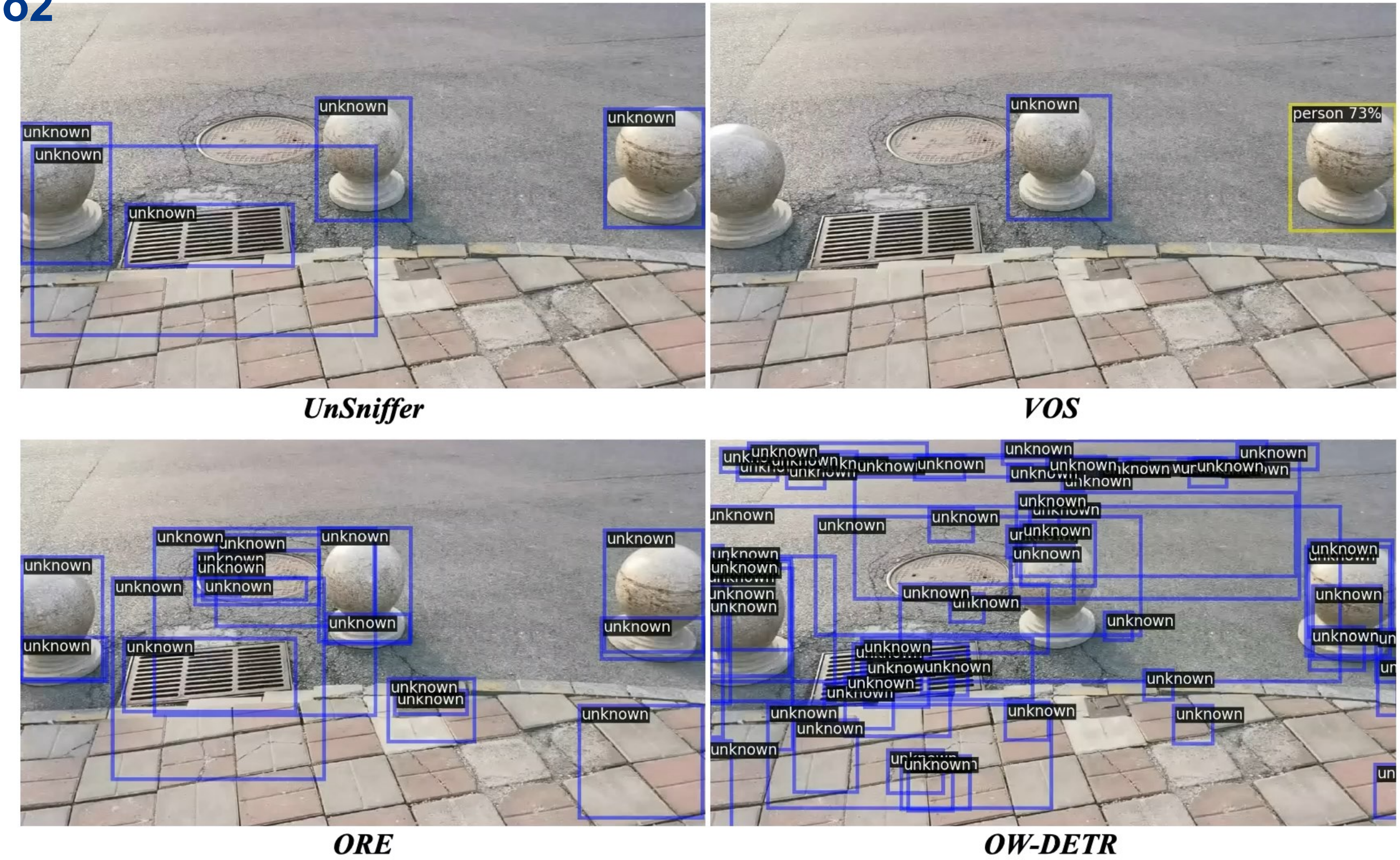


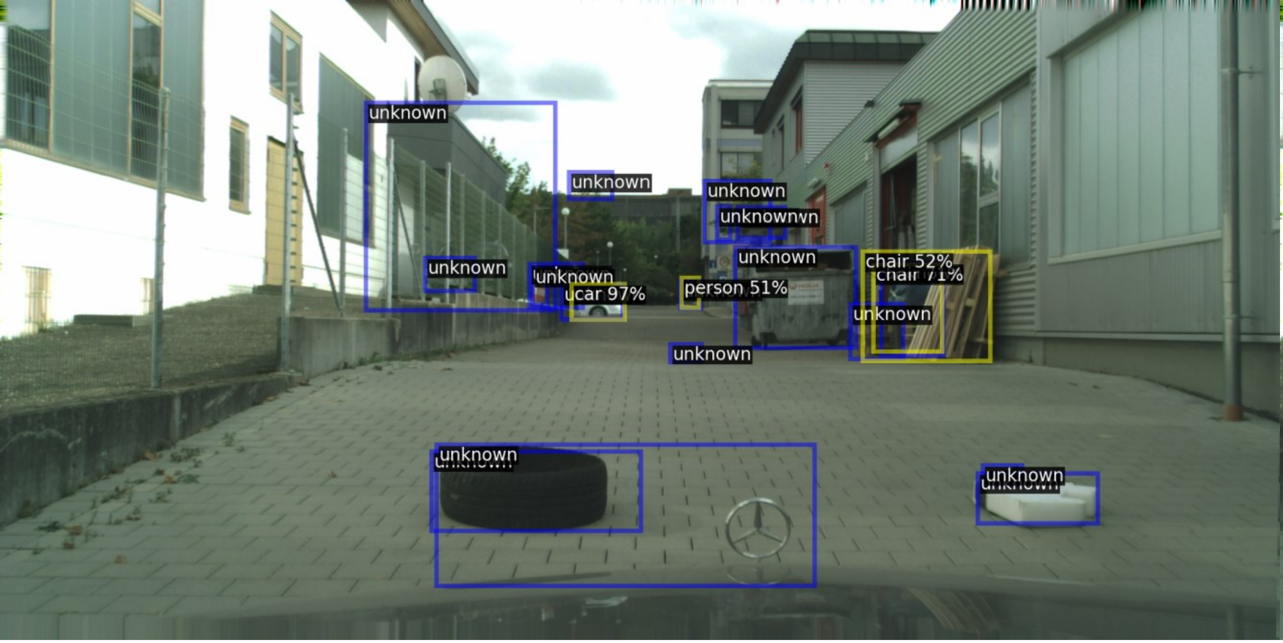
ORE



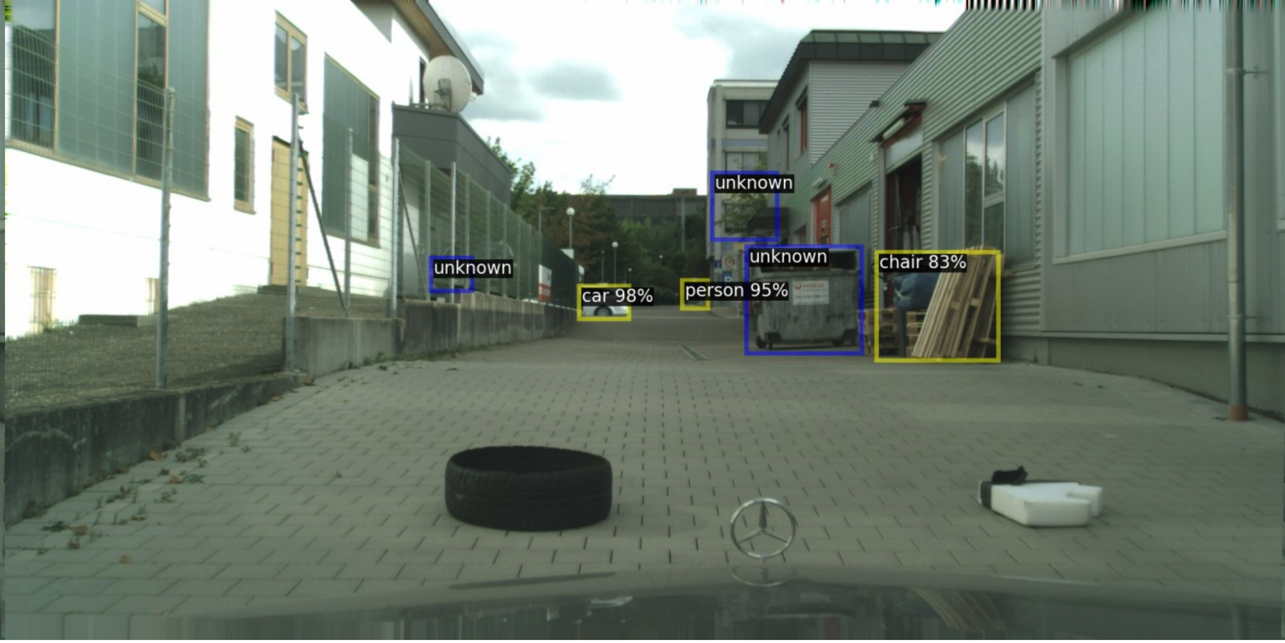
OW-DETR

Video Demo2

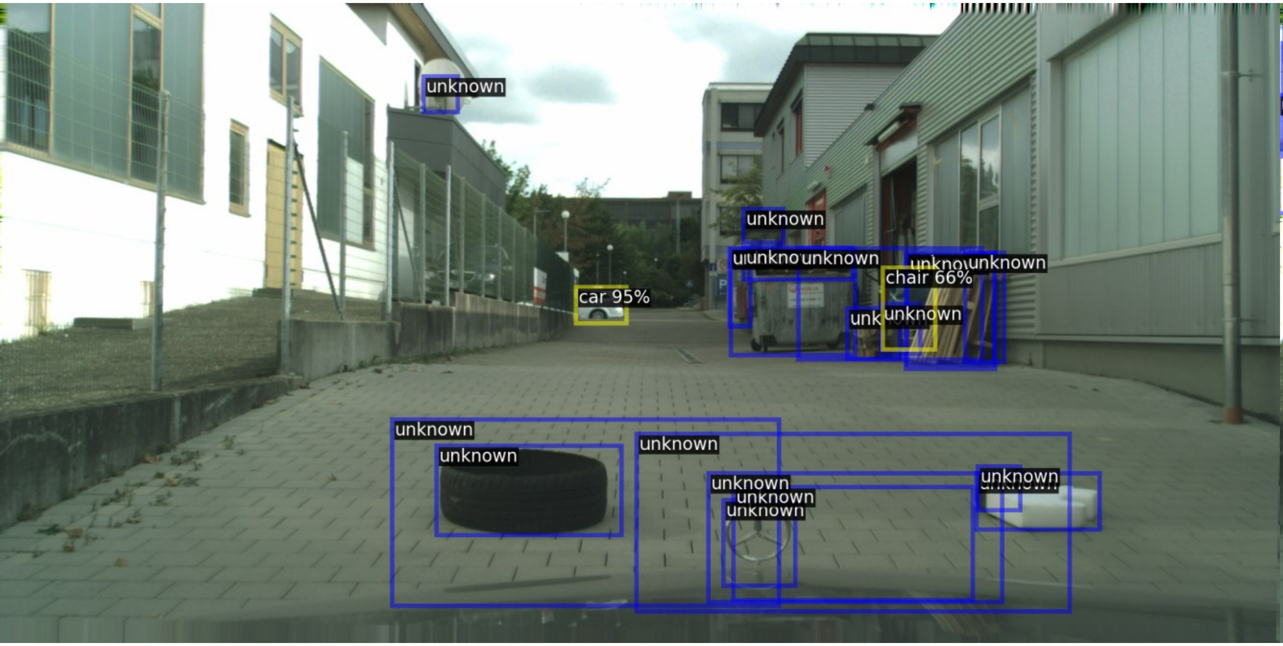




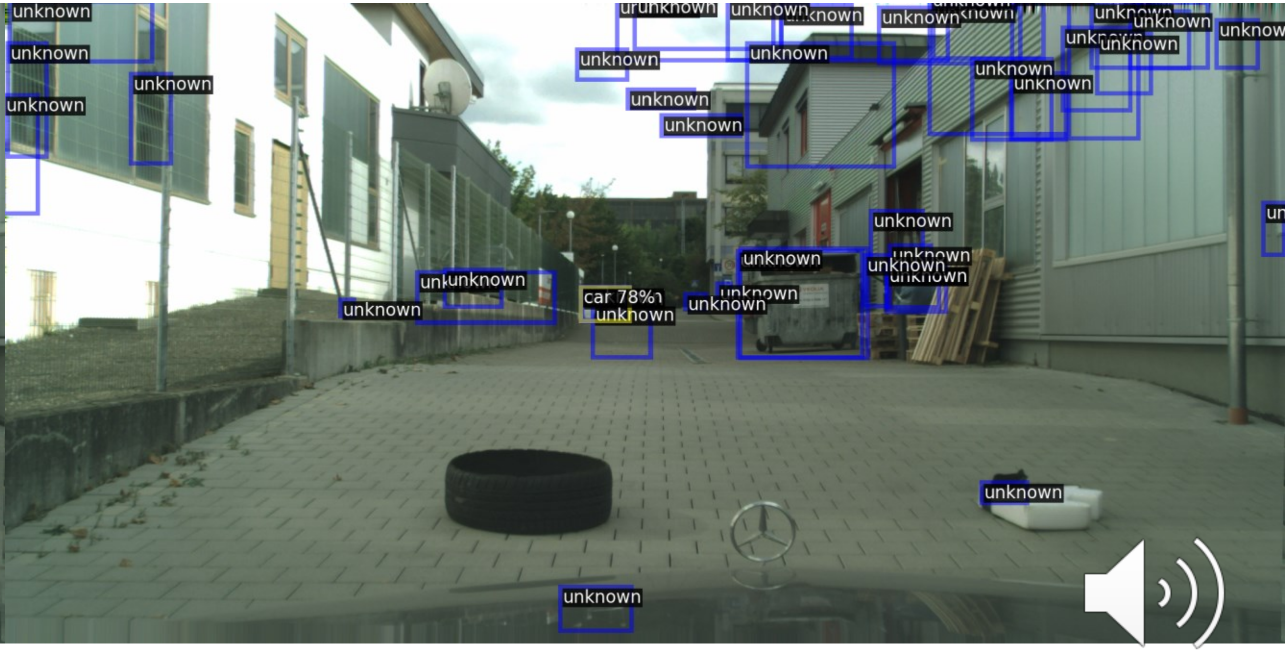
UnSniffer



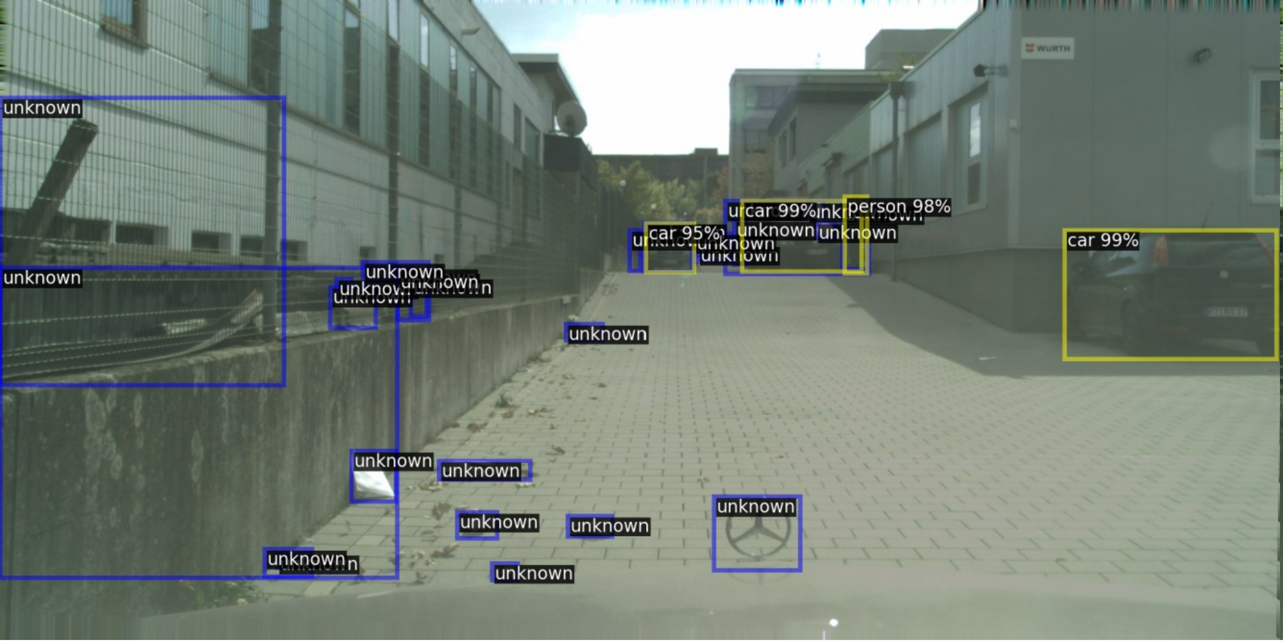
VOS



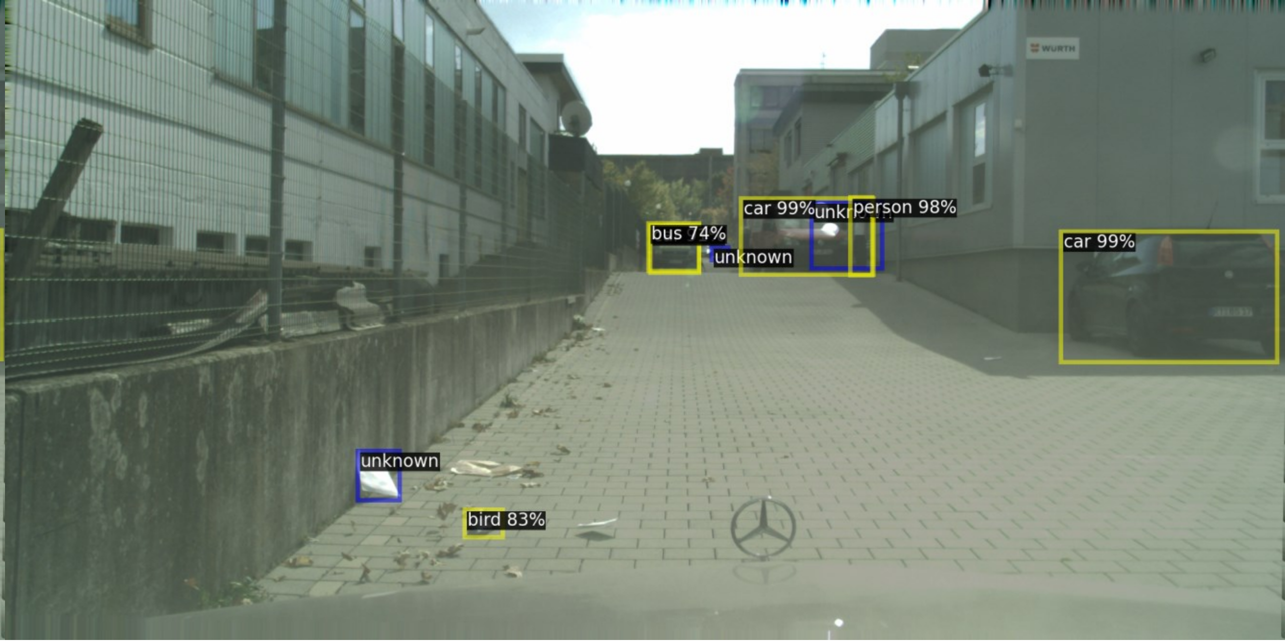
ORE



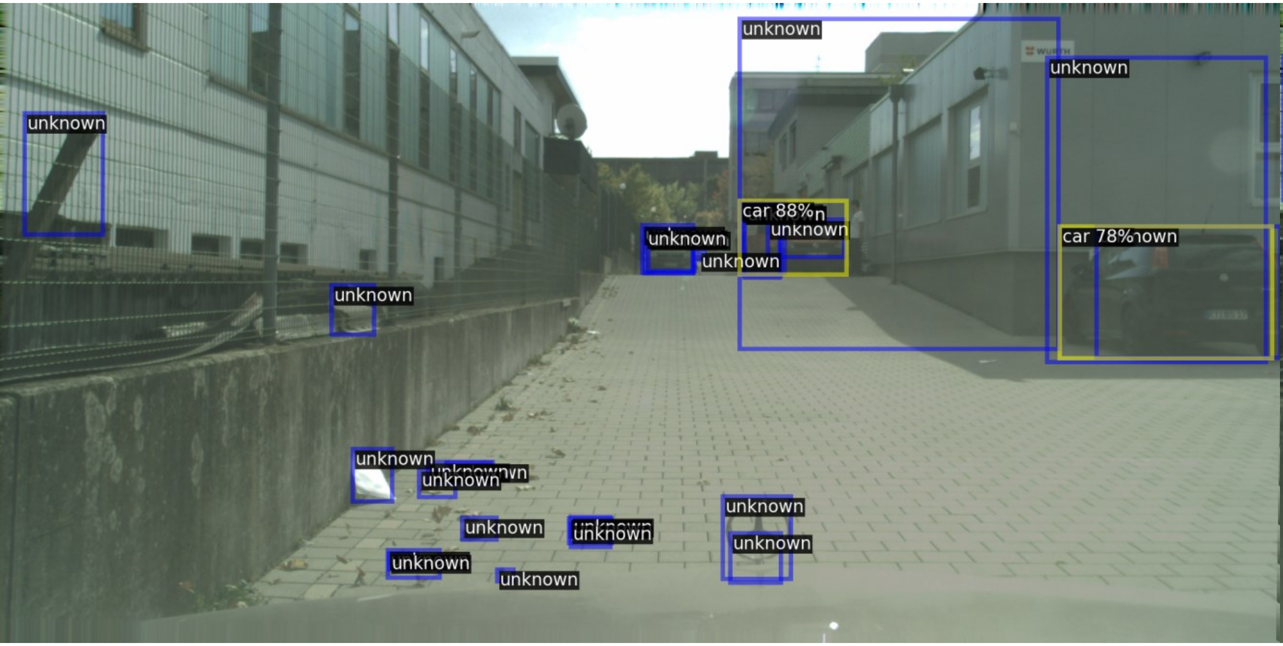
OW-DETR



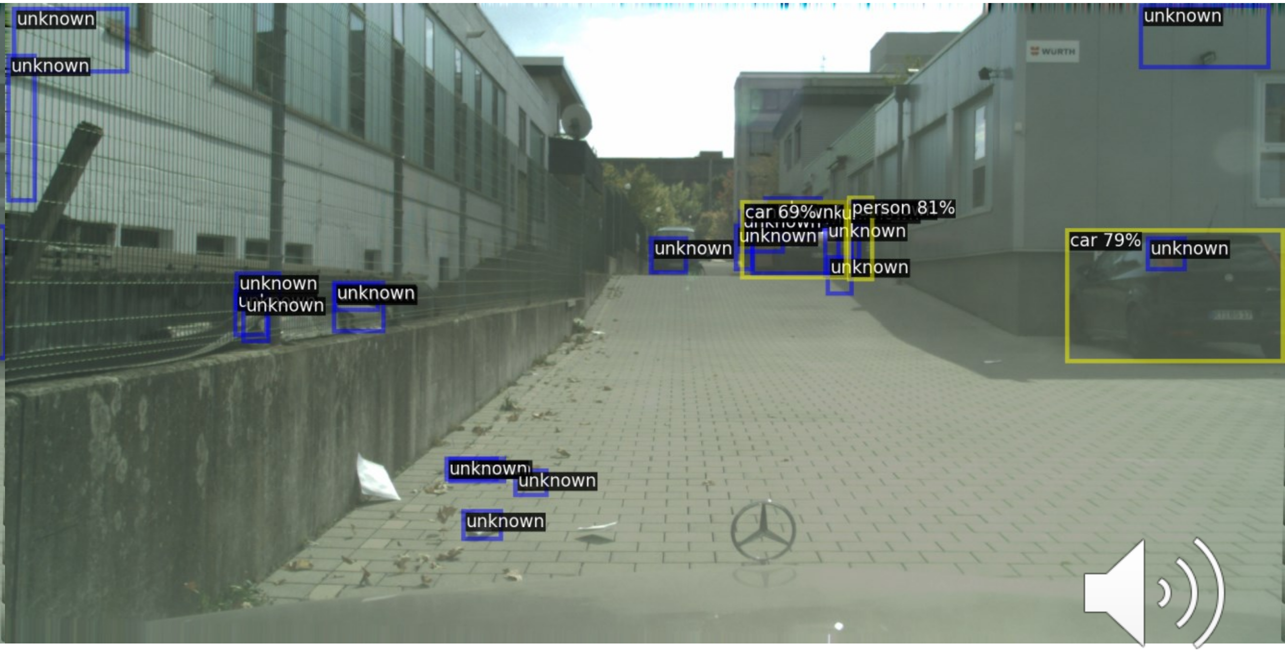
UnSniffer



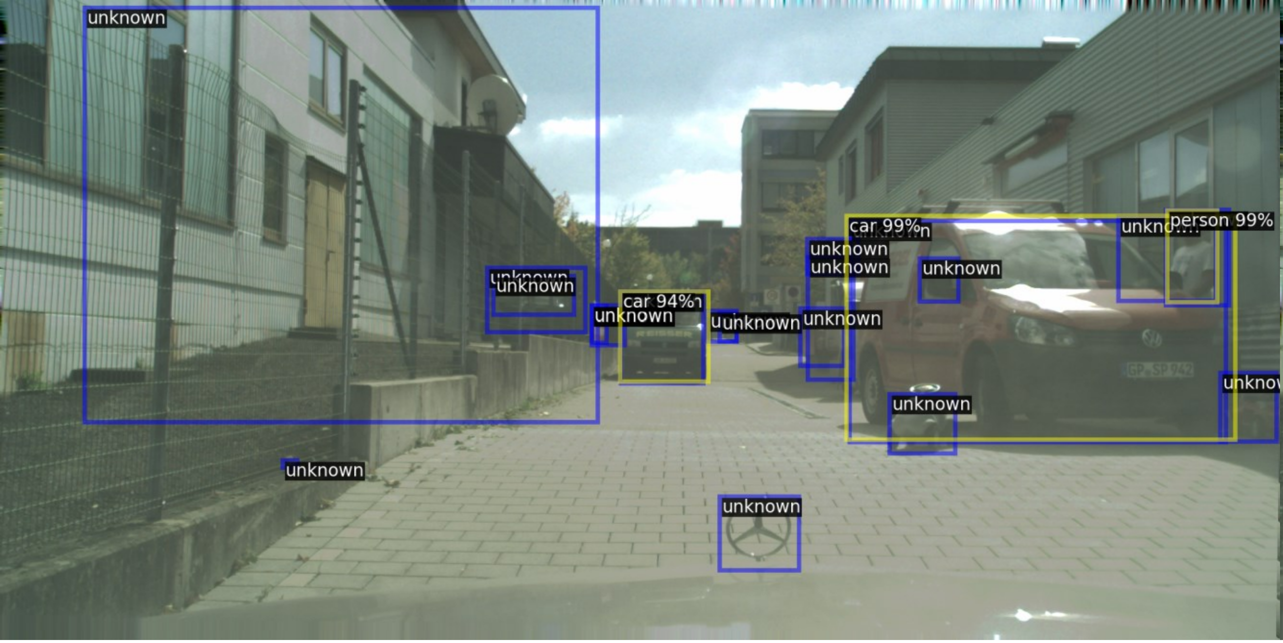
VOS



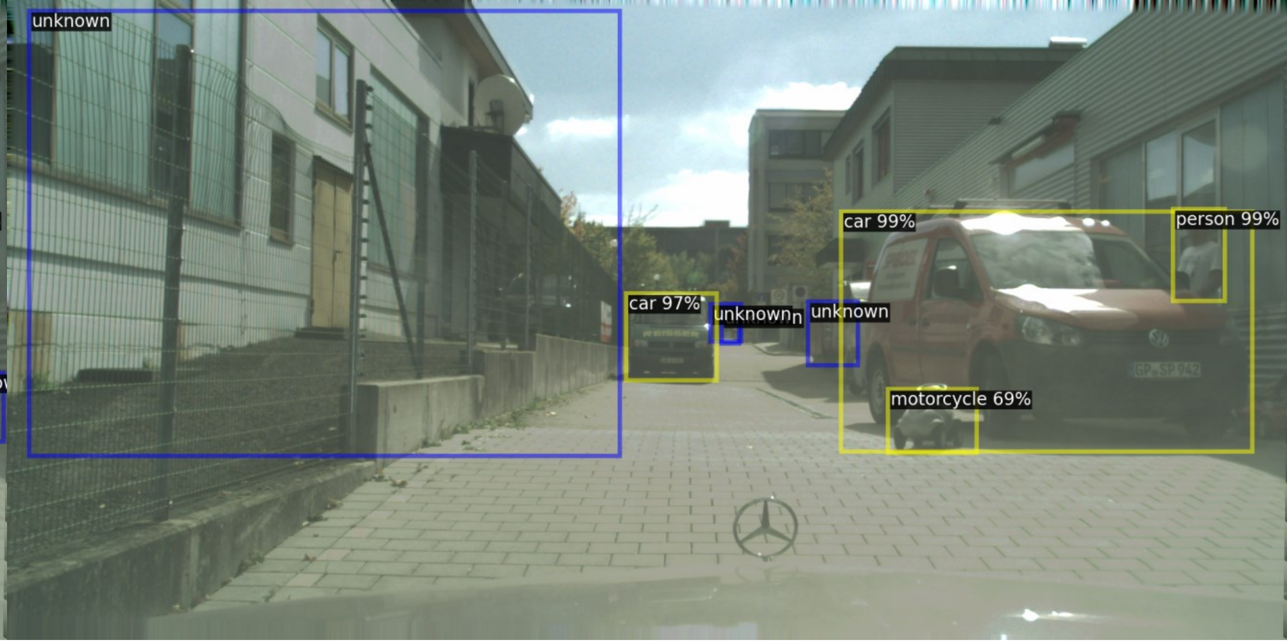
ORE



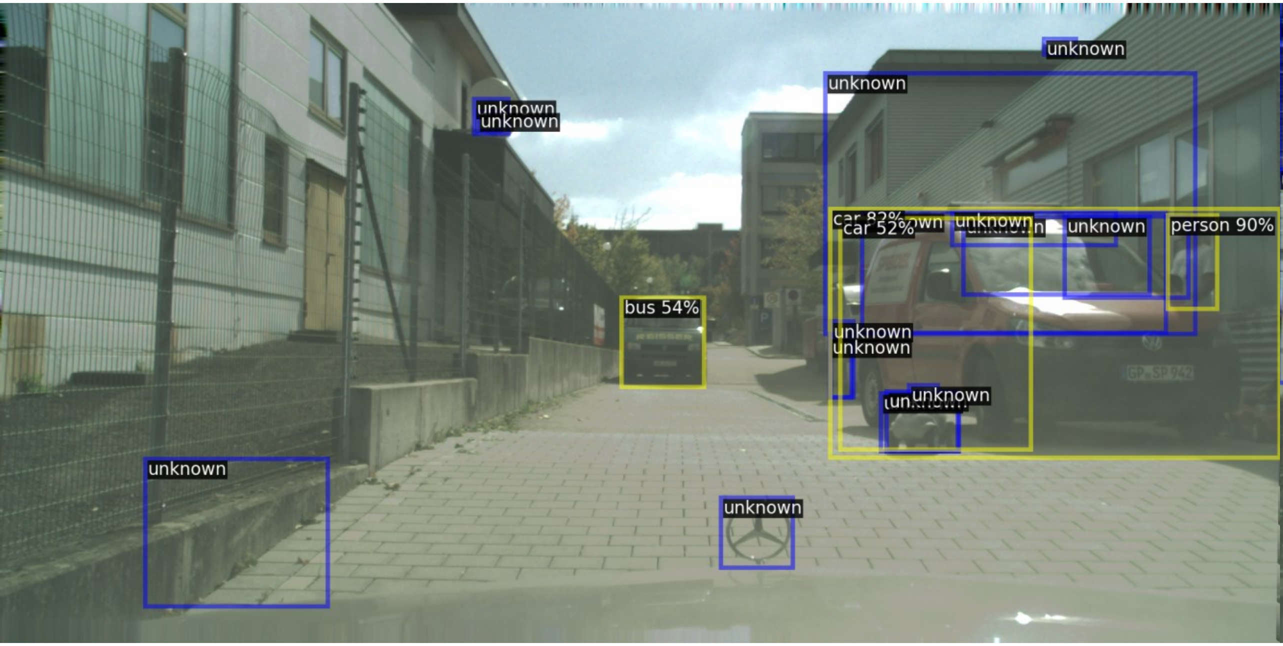
OW-DETR



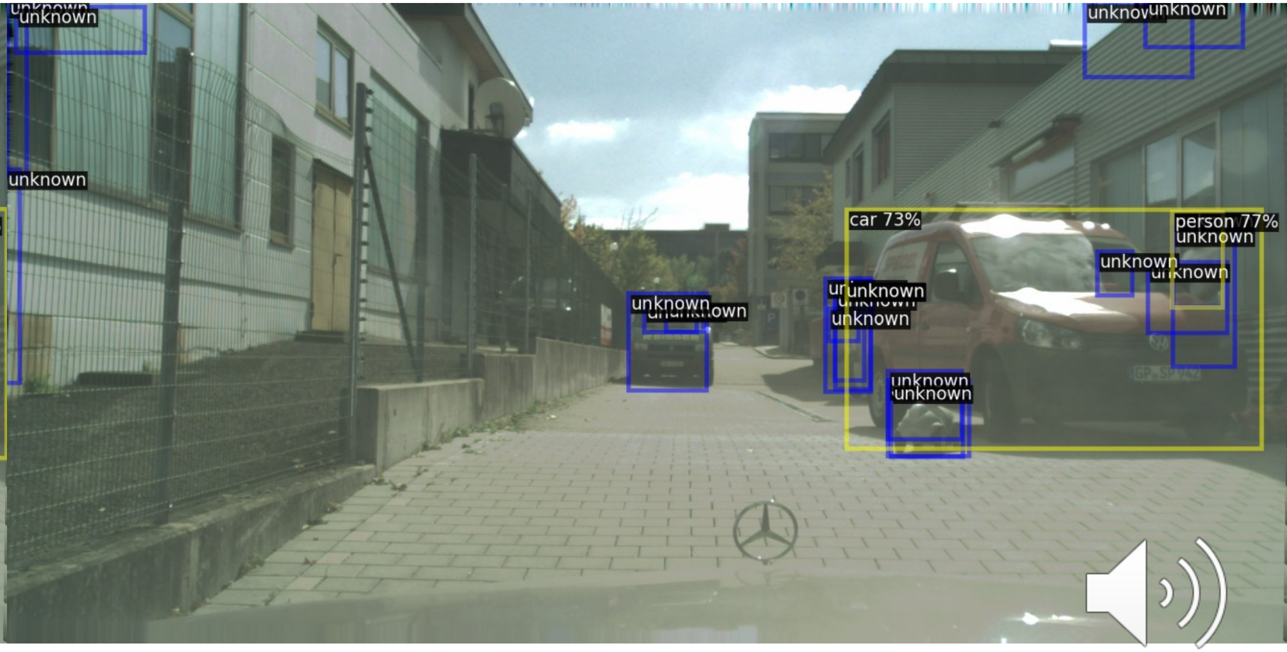
UnSniffer



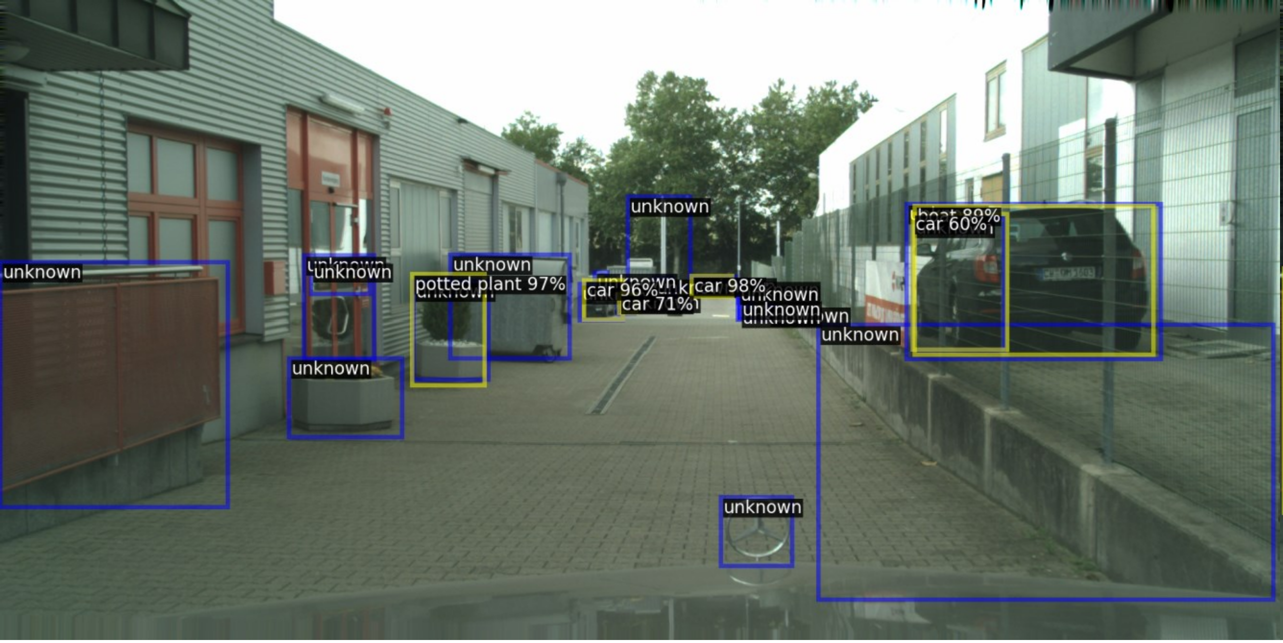
VOS



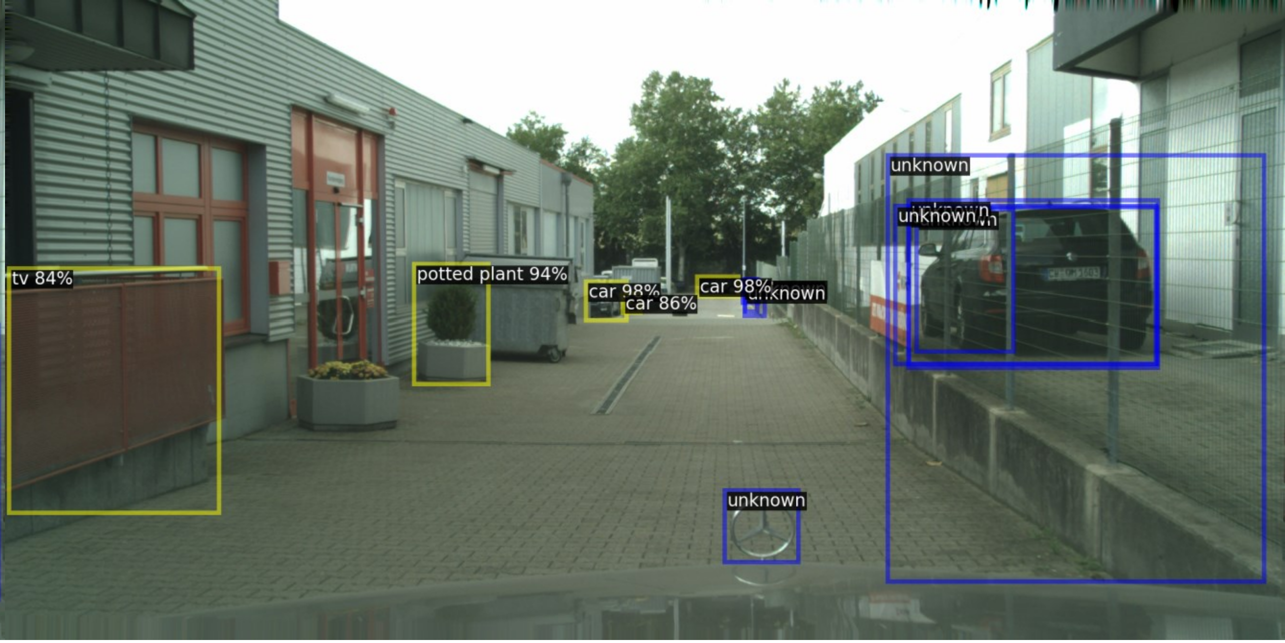
ORE



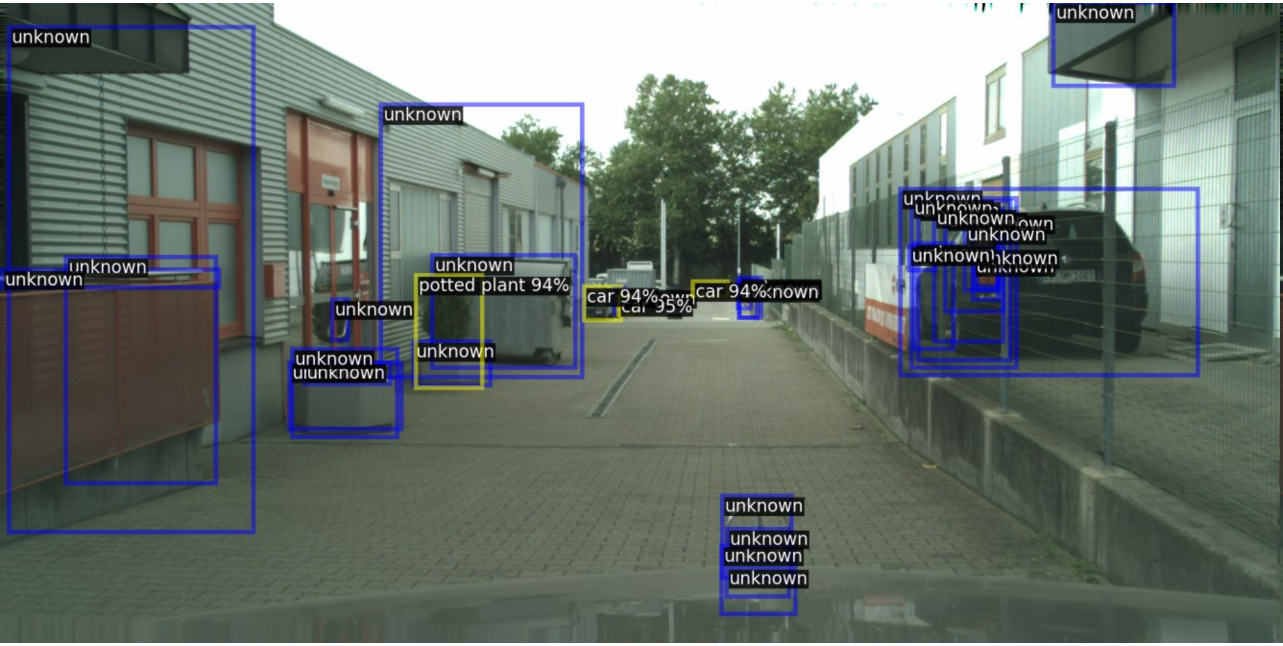
OW-DETR



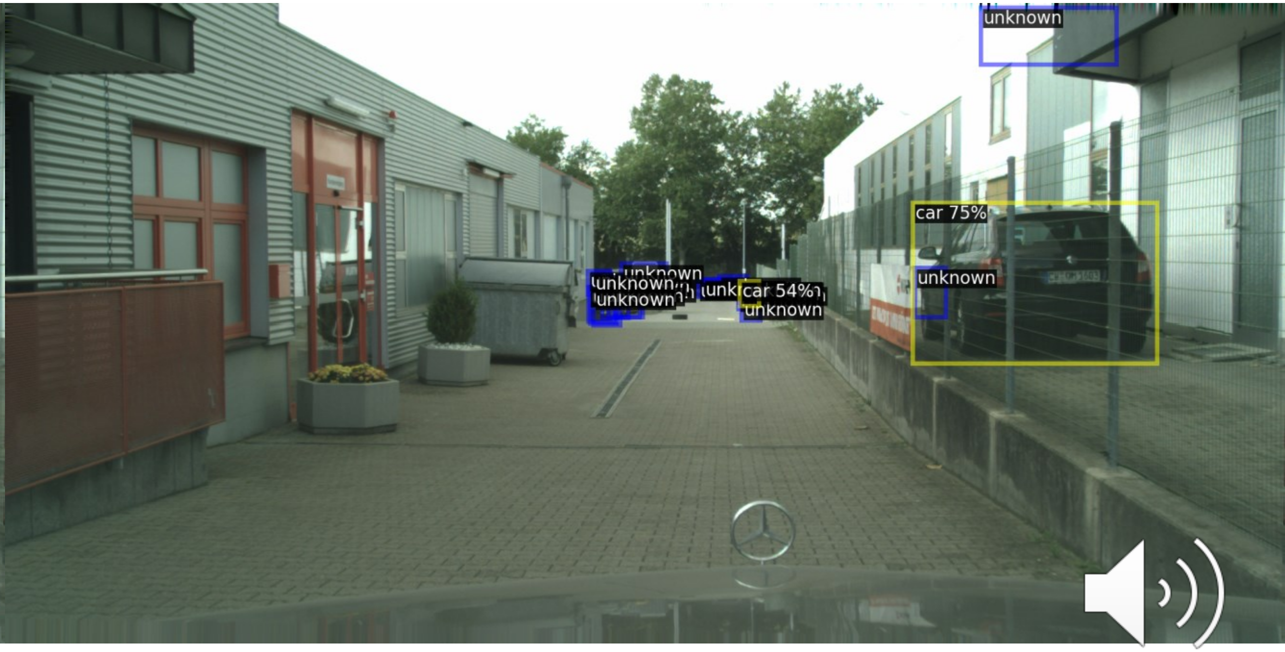
UnSniffer



VOS



ORE



OW-DETR

Thanks !!!