CoOVD: Continual Open-Vocabulary Object Detection with Novel Concept Injection

Yupeng Zhang[1,2] Ruize Han [3] Yuzhong Feng[1] Zixin Ren[1] Yuntong Tian[4,1] Liang Wan[1,2] *

[1] College of Intelligence and Computing, Tianjin University. [2] Key Research Center for Surface Monitoring and Analysis of Relics, State Administration of Cultural Heritage.

[3] Faculty of Computer Science and Artificial Intelligence, Shenzhen University of Advanced Technology. [4] School of Computer Science and Software Engineering, Southwest Petroleum University.

{zhangyupeng, yzfeng, renzixin, lwan} @tju.edu.cn; hanruize@suat-sz.edu.cn; tianyuntong@swpu.edu.cn

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

Motivation. Open-vocabulary detectors can recognize categories beyond their supervised vocabulary, but still struggle with newly introduced concepts and their associations with object regions. As these concepts arrive continually, full retraining on accumulated data becomes costly and depends on access to historical training images. A practical detector should therefore acquire new concepts while retaining previously learned knowledge and its pretrained open-vocabulary capability.



Figure 1: Pretrained visual features produce coherent responses within object regions.

Task. We introduce CoOVD, which sequentially learns new concept groups without historical image replay. Evaluation jointly considers new-concept acquisition, historical-concept retention, and preservation of original open-vocabulary knowledge.

Key insight. Fig. 1 shows coherent feature responses within novel-object regions, suggesting that pretrained visual features remain informative. Fig. 2 illustrates how novel concepts can share visual attributes with familiar concepts. This motivates NoInDet to freeze the visual encoder and selectively update textual parameters, learning new concept–region associations while preserving existing semantic knowledge.

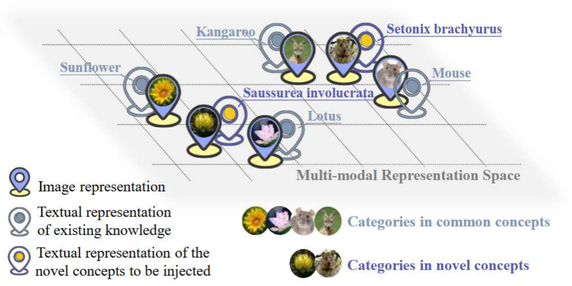

Figure 2: Transferable visual representations motivate text-side adaptation for novel-concept learning.

Benchmark: Novel-114

114 concepts · 7 sequential tasks · 8,891 images

Novel-114 targets concepts poorly recognized by the initial detector, with 4,424 training images and 4,467 test images. Evaluation tracks new-concept acquisition, historical-concept retention, and original OVD preservation without historical image replay.

Table 1: Statistics of each task in the Novel-114.

Task	# Concepts	Subset	# Images	# Objects
Animals-25	25	train	1049	1334
		test	1057	1360
Clothing-6	6	train	184	211
		test	187	231
Cold Weapons-4	4	train	150	272
		test	151	247
Food-28	28	train	1407	3093
		test	1420	3188
Instruments-8	8	train	134	170
		test	139	250
Plants25	25	train	873	3456
		test	882	3370
Vessels-18	18	train	627	757
		test	631	759
All Categories	**114**	**train**	**4424**	**9293**
		test	**4467**	**9405**

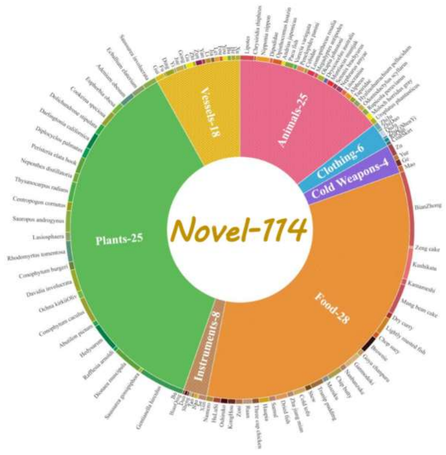

Figure 3: Task-wise composition and data distribution of Novel-114.

Method: NoInDet

As shown in Fig. 4, text-side adaptation. Built on LLMDet, NoInDet freezes the visual encoder and updates a small, fixed subset of parameters in the text encoder and projector. Current-task region–text contrastive learning establishes associations between novel concept names and object regions, while textual regularization helps preserve existing knowledge without historical image replay.RSSD (Representation Space Stability Distillation). RSSD uses randomly sampled common-concept texts and stored historical-concept texts as lightweight semantic anchors. With the preceding task's frozen text branch as a reference, it aligns individual concept representations and pairwise semantic relations at both the text-encoder and projector levels, stabilizing the semantic space used for open-vocabulary matching.KPD (Knowledge-aware Parameter Decoupling). KPD estimates parameter sensitivity through probe gradients from RSSD's absolute semantic distillation terms. It selects less sensitive parameter elements for adaptation and masks gradients on more sensitive elements, coupling representation preservation with selective updates to accommodate new concepts while limiting disruption to existing knowledge.

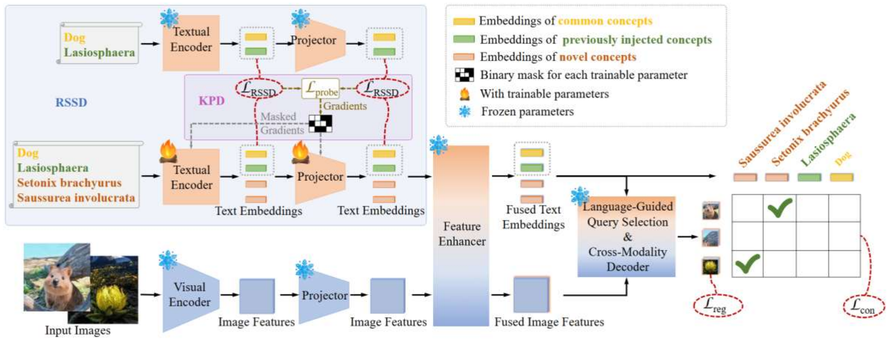

Figure 4: Continual concept injection without additional deployment parameters or historical image replay.

Results

NoInDet achieves the highest final performance among the compared methods, reaching 44.3 AP and exceeding the strongest baseline, SGVF, by 5.0 AP. Across seven sequential tasks, it maintains 47.1 mean COCO AP and achieves 68.8 mean current-task AP, demonstrating a favorable balance between concept acquisition and knowledge preservation. These results support the effectiveness of text-side adaptation with RSSD and KPD for continual concept injection, without additional deployment parameters or historical image replay.

Table 2: The main results on MS COCO and Novel-114 (%).

Method	Train \ Test	MS COCO	Animals-25	Clothing-6	Cold Weapons-4	Food-28	Instruments-8	Plants-25	Vessels-18	Avg.
LLMDet	–	54.4/70.8	5.9/7.3	10.7/22.4	27.1/32.0	9.5/13.5	31.3/32.7	3.5/5.1	4.1/4.8	18.3/23.6
ZiRa	Animals-25	46.4/60.7	86.6/94.5	–	–	–	–	–	–	–
	Clothing-6	43.7/57.2	85.2/93.2	43.9/49.1	–	–	–	–	–	–
	Cold Weapons-4	43.0/56.4	69.1/75.5	37.0/40.5	30.0/36.0	–	–	–	–	–
	Food-28	34.5/45.5	58.9/66.3	14.8/22.7	20.2/24.8	70.5/80.7	–	–	–	–
	Instruments-8	35.8/47.2	61.6/69.1	13.8/20.0	18.6/23.3	68.7/78.8	44.0/51.0	–	–	–
	Plants-25	19.7/26.4	27.0/30.4	11.3/17.6	29.2/34.0	1.3/1.5	13.5/15.0	66.3/85.8	–	–
	Vessels-18	17.7/23.2	31.3/35.1	11.8/15.8	20.8/24.1	38.2/45.8	21.5/24.3	61.0/78.8	55.9/59.4	32.3/38.3
MoE-Adapters	Animals-25	39.4/52.3	84.5/91.6	–	–	–	–	–	–	–
	Clothing-6	36.4/48.7	64.9/70.7	33.5/37.4	–	–	–	–	–	–
	Cold Weapons-4	31.3/42.5	58.6/64.6	32.0/34.0	59.0/64.8	–	–	–	–	–
	Food-28	31.3/42.5	65.4/72.3	25.7/29.0	34.7/40.4	64.8/74.2	–	–	–	–
	Instruments-8	38.1/52.0	49.5/55.1	28.0/31.2	30.6/35.4	62.5/72.0	31.1/32.2	–	–	–
	Plants-25	33.5/45.7	56.0/63.2	27.2/33.0	20.7/23.9	48.7/57.5	12.4/17.6	42.7/55.0	–	–
	Vessels-18	30.7/42.0	58.4/65.1	28.8/36.8	23.2/26.8	47.1/55.4	16.2/19.8	38.5/50.0	60.3/61.3	37.9/44.7
LADA	Animals-25	35.7/47.2	87.9/95.0	–	–	–	–	–	–	–
	Clothing-6	17.6/22.6	69.7/80.9	64.2/68.2	–	–	–	–	–	–
	Cold Weapons-4	20.2/26.7	63.6/71.1	39.7/44.9	75.3/81.8	–	–	–	–	–
	Food-28	32.4/43.8	34.8/41.1	13.8/19.3	26.1/28.7	74.3/84.9	–	–	–	–
	Instruments-8	18.1/24.4	15.4/20.6	17.9/23.5	23.9/28.1	56.5/66.7	57.9/61.9	–	–	–
	Plants-25	23.4/31.8	20.2/25.0	5.6/10.6	20.2/25.0	27.5/35.0	34.2/38.9	70.1/89.3	–	–
	Vessels-18	20.8/27.9	16.0/20.5	6.0/10.6	20.6/23.6	9.3/13.1	38.4/52.6	12.3/14.1	70.9/71.9	24.3/29.3
C-CLIP	Animals-25	48.3/64.0	64.4/71.7	–	–	–	–	–	–	–
	Clothing-6	43.3/58.0	59.0/66.0	42.0/45.7	–	–	–	–	–	–
	Cold Weapons-4	45.2/61.5	27.6/32.5	30.4/32.8	51.4/60.6	–	–	–	–	–
	Food-28	45.4/63.1	17.6/20.9	11.3/19.3	34.2/40.4	46.8/55.9	–	–	–	–
	Instruments-8	46.6/62.6	14.8/17.1	17.9/26.3	30.7/35.3	31.5/37.4	41.9/45.5	–	–	–
	Plants-25	32.7/45.4	9.0/11.1	17.4/25.4	26.8/31.1	12.0/14.7	12.5/17.5	32.6/43.0	–	–
	Vessels-18	32.7/45.6	4.4/5.9	16.3/24.2	25.7/31.6	7.1/10.0	10.8/16.5	18.6/24.3	23.7/26.3	17.4/23.1
SGVF	Animals-25	26.6/35.9	83.7/92.0	–	–	–	–	–	–	–
	Clothing-6	29.5/39.1	82.2/91.6	61.1/67.4	–	–	–	–	–	–
	Cold Weapons-4	26.3/35.1	78.8/88.4	39.1/45.6	73.4/79.3	–	–	–	–	–
	Food-28	24.7/33.7	60.1/76.5	24.3/31.9	44.8/51.0	71.2/82.0	–	–	–	–
	Instruments-8	24.0/32.6	52.0/70.5	16.9/20.4	40.0/46.8	64.3/74.1	64.2/68.6	–	–	–
	Plants-25	24.7/33.7	59.8/70.9	16.5/20.4	41.1/46.8	52.3/62.7	47.2/50.4	65.9/85.9	–	–
	Vessels-18	19.3/26.4	52.1/63.1	11.9/15.1	39.4/47.0	42.7/52.0	36.3/39.1	39.3/54.5	73.4/74.7	39.3/46.5
KeepLoRa	Animals-25	50.2/67.6	14.9/19.8	–	–	–	–	–	–	–
	Clothing-6	29.5/40.6	2.6/3.8	8.7/17.7	–	–	–	–	–	–
	Cold Weapons-4	34.5/47.1	5.4/7.0	9.8/22.5	26.7/32.0	–	–	–	–	–
	Food-28	45.1/60.3	7.1/8.8	6.7/18.1	29.6/36.8	17.3/23.6	–	–	–	–
	Instruments-8	31.9/46.2	4.9/6.2	9.4/17.9	21.0/25.9	6.9/10.4	23.8/26.8	–	–	–
	Plants-25	44.7/59.9	7.0/9.2	13.2/29.0	26.5/32.7	9.3/14.0	26.7/30.4	5.5/8.0	–	–
	Vessels-18	43.8/59.7	6.1/7.6	10.1/18.8	25.2/31.0	8.4/13.3	28.2/32.1	3.7/5.2	6.8/8.9	16.5/22.1
NoInDet	Animals-25	49.8/65.1	87.9/95.2	–	–	–	–	–	–	–
	Clothing-6	49.5/64.3	80.7/88.6	47.1/54.6	–	–	–	–	–	–
	Cold Weapons-4	49.1/63.7	81.9/89.0	40.5/53.5	71.0/79.6	–	–	–	–	–
	Food-28	47.0/62.0	73.3/81.0	24.3/32.7	55.8/62.7	72.8/82.9	–	–	–	–
	Instruments-8	47.1/61.9	75.9/83.4	14.9/25.0	51.0/61.3	68.8/78.7	62.5/67.1	–	–	–
	Plants-25	45.7/60.2	48.7/54.5	20.6/29.6	44.7/50.1	55.3/64.8	63.3/69.5	69.6/89.1	–	–
	Vessels-18	41.7/54.8	50.1/57.3	18.2/26.7	31.1/38.8	43.2/54.1	35.7/39.2	63.7/83.3	70.4/73.7	44.3/53.5

Visualization

Selected examples show more complete object coverage and fewer missed or redundant detections, complementing the quantitative results across diverse concept groups.

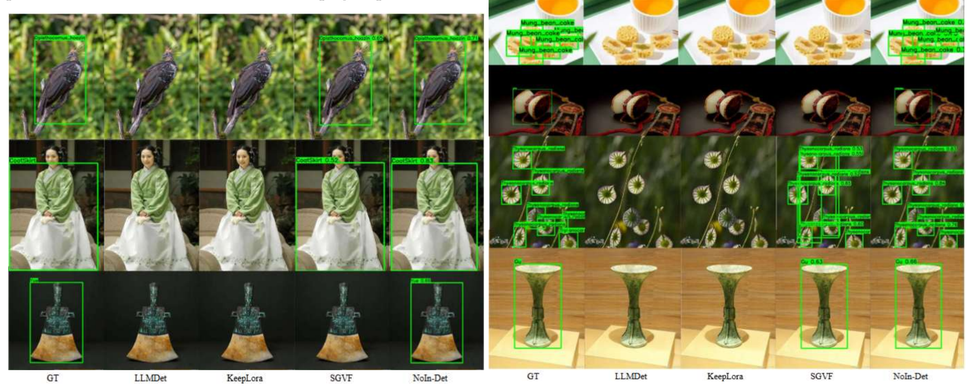

Figure 5: Qualitative comparison on Novel-114.