EVIDENCE-DRIVEN PROGRESSIVE KNOWLEDGE FORMATION AND SELECTIVE ADAPTATION: A NEUROSCIENCE-INSPIRED CONTINUAL LEARNING FRAMEWORK

Anonymous authors
Paper under double-blind review

ABSTRACT

Current object detectors suffer from catastrophic forgetting in open-world continual learning, where novel objects emerge progressively from incomplete, noisy, or ambiguous observations. Existing methods typically adapt representations directly to incoming observations, causing immature novel concepts to be prematurely assimilated into established knowledge and destabilizing previously learned representations. Neuroscience suggests that lifelong learning relies on complementary memory systems, where specific experiences are progressively consolidated into generalized knowledge while established representations remain relatively stable. Inspired by this principle, we propose Evidence-driven Progressive Knowledge formation and selective Adaptation (EPKA), a new continual learning framework that transforms uncertain observations into stable knowledge before allowing them to reshape existing representations. EPKA comprises two coupled mechanisms. Progressive Structural Evidence Evolution (PSEE) progressively accumulates and consolidates compatible observations into persistent, structurally coherent novel representations. Adaptive Representation Competition Regulation (ARCR) selectively adapts representations based on their stability, while protecting established knowledge from unnecessary changes. Together, they establish an evidence-knowledge-adaptation process that balances plasticity for emerging concepts with stability of established knowledge. Extensive experiments across three open-world continual learning scenarios and five benchmarks demonstrate that EPKA consistently improves novel knowledge acquisition while mitigating catastrophic forgetting, highlighting its adaptability and plasticity in continuously evolving visual environments.

1 INTRODUCTION

Open-world object detection aims to enable detectors to continuously acquire knowledge about emerging object categories without compromising previously learned representations. Existing open-world continual detection methods, like OpenMax (Bendale & Boult, 2016), ORE (Joseph et al., 2021), OW-DETR (Gupta et al., 2022), PROB (Zohar et al., 2023), OW-VAP (Xi et al., 2025a), and recent open-world continual learning methods (Li et al., 2025) have made substantial progress in discovering unknown objects and adapting detectors to emerging categories. However, these approaches generally assume that newly observed instances can directly serve as reliable signals for model adaptation, leaving the formation of novel knowledge largely implicit. In open-world environments, this assumption can be problematic because early observations of an emerging category are often sparse, noisy, incomplete, or visually ambiguous. Rather than representing a well-established concept, such observations provide only partial evidence about its underlying representation. Incorporating them directly into the model can therefore prematurely commit the learner to an immature concept, causing unreliable knowledge to propagate through subsequent learning and increasing interference with previously established representations. This leads to a fundamental question for open-world continual detection: *when is the accumulated evidence sufficient to establish reliable knowledge, and how should such knowledge be selectively incorporated without unnecessarily altering established representations?*

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

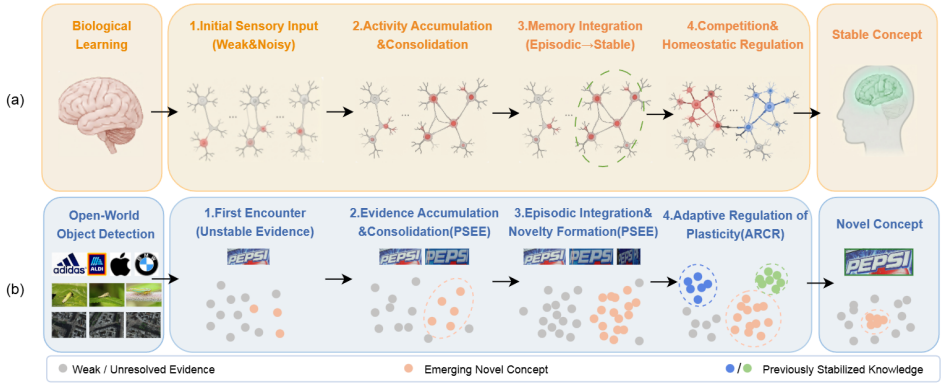


Figure 1: (a) Biological systems progressively form stable concepts from uncertain sensory inputs through accumulation, consolidation, and regulation. (b) EPKA mimics this process by transforming unstable observations into reliable novel knowledge via PSEE and ARCR.

Neuroscience offers a complementary perspective on this problem (McClelland et al., 1995; Sun et al., 2023). Rather than treating every new experience as immediately established knowledge, biological learning supports a gradual transition from newly encoded experiences to more stable representations through consolidation and organization (Figure 1). This perspective suggests that knowledge formation should be progressive: individual observations provide evidence for emerging concepts, while their accumulation and consolidation gradually determine which representations become stable knowledge. Inspired by this principle, we rethink continual learning as a process of progressive knowledge formation followed by selective adaptation, rather than immediate adaptation to every incoming observation.

Progressive knowledge formation, however, addresses only *when* a novel concept becomes sufficiently reliable; it does not determine *how strongly* the established representation should change once new knowledge is formed. Neuroscientific studies further suggest that competition for limited synaptic resources can regulate neural plasticity, constraining how strongly new information modifies existing representations (Triesch et al., 2018). This provides a complementary principle for continual learning adaptation: newly formed knowledge should not be incorporated uniformly, but should adapt the representation space according to its evidence and its interaction with previously acquired knowledge. Consistent with this perspective, Shi et al. (Shi et al., 2025) show that coordinating episodic and generalized representations can facilitate continual learning, highlighting the importance of allowing new experiences to interact with, rather than simply overwrite existing representations. Together, these observations suggest a two-stage learning principle: **new knowledge should first be progressively established from accumulated evidence and then selectively integrated into the existing representation space**.

Based on this principle, we introduce Evidence-driven Progressive Knowledge formation and selective Adaptation (EPKA), a neuroscience-inspired continual learning framework. Unlike conventional continual learning approaches that treat an unseen category as a set of independent unknown instances, EPKA models its representation as an evolving state formed from accumulated evidence and adapted within an existing representation space. Rather than treating every newly observed instance as an immediate update signal, EPKA accumulates evidence across sequential observations to characterize an emerging concept. As evidence accumulates, the representation of the novel concept is progressively refined, evolving from an uncertain and immature state toward a more stable representation. Only when the accumulated evidence provides sufficient support, EPKA selectively adapts the model to incorporate the emerging knowledge, while established representations remain protected from unnecessary modification. In this way, the framework learns not only *what is new*, but also *when a novel concept is sufficiently mature to become knowledge and how strongly it should influence the existing representation space*.

To realize this learning principle, EPKA consists of two complementary mechanisms. *Progressive Structural Evidence Evolution (PSEE)* accumulates and consolidates structural evidence across sequential observations, progressively transforming uncertain observations into a more reliable rep-

resentation of an emerging category. Building upon the resulting knowledge state, *Adaptive Representation Competition Regulation (ARCR)* dynamically regulates representation plasticity according to the novelty and temporal stability of emerging patterns, allocating greater adaptation capacity to sufficiently supported new knowledge while suppressing unnecessary changes to established representations. Through the coordinated operation of these two mechanisms, EPKA addresses two coupled sources of instability in continual learning: *premature assimilation of immature knowledge* and *excessive adaptation of established knowledge*.

Our main contributions are as follows. First, we adopt **a new continual learning perspective** by identifying premature knowledge assimilation as an overlooked source of instability in open-world continual learning and reframing continual learning as a process of progressive knowledge formation rather than immediate adaptation to newly observed instances. Second, we develop **an evidence-driven progressive knowledge formation mechanism** that progressively accumulates and evaluates evidence across observations, enabling emerging concepts to evolve from uncertain representations into stable knowledge before consolidation. Third, we introduce **a selective adaptation mechanism** that incorporates reliable novel knowledge while minimizing unnecessary modifications to established representations, thereby mitigating both premature assimilation of immature knowledge and catastrophic interference. Finally, we conduct **comprehensive evaluatio**n through extensive experiments on three open-world scenarios and five benchmarks, together with ablation evaluation and mechanism-level analysis. These results demonstrate that explicitly modeling knowledge formation and adaptation as distinct but coupled processes provides consistent benefits for learning emerging concepts while preserving previously acquired knowledge.

2 Related Work

Open-World Object Detection. Open-world object detection (OWOD) aims to discover unknown objects and incrementally learn emerging categories while preserving previously acquired knowledge. ORE (Joseph et al., 2021) established the OWOD setting by jointly considering unknown discovery and incremental learning. Subsequent methods explored transformer-based detection and uncertainty-aware modeling, including OW-DETR (Gupta et al., 2022) and PROB (Zohar et al., 2023). Recent studies further investigate opens-world detection through improved class-agnostic objectness modeling and semantic-aware unknown recognition, such as Open-World Objectness Modeling Unifies Novel Object Detection (Zhang et al., 2025) and OW-OVD (Xi et al., 2025b), and OW-DAR (Ye et al., 2026). However, these approaches mainly focus on unknown discovery, objectness estimation, or recognition under open-world uncertainty, while the progressive process of transforming uncertain observations into stable novel knowledge remains largely implicit. In contrast, EPKA explicitly models evidence accumulation before knowledge adaptation, enabling emerging concepts to be progressively consolidated.

Continual Learning and Knowledge Consolidation. Recent continual learning studies investigate stable knowledge adaptation through pretrained model adaptation, replay-based consolidation, and parameter-efficient updating (McDonnell et al., 2023; Zhang et al., 2023). Recent efforts further explore adaptive knowledge regulation by addressing negative transfer and interference during continual updates (Wang et al., 2026; Qiao & Mahdavi, 2026). However, existing approaches mainly focus on how acquired information is retained, transferred, or adapted after training signals become available, while the reliability of emerging knowledge is usually assumed. In open-world scenarios, observations of new categories are often sparse and ambiguous. EPKA addresses this overlooked stage by explicitly separating progressive knowledge formation from subsequent representation adaptation.

3 Method

3.1 Overview

As illustrated in Figure 2, EPKA follows an *evidence-knowledge-adaptation* paradigm. PSEE progressively aggregates and refines structural evidence from new observations into stable novel representations, while ARCR regulates representation plasticity to selectively incorporate the resulting knowledge into the existing representation space. PSEE determines *when* emerging evidence be-

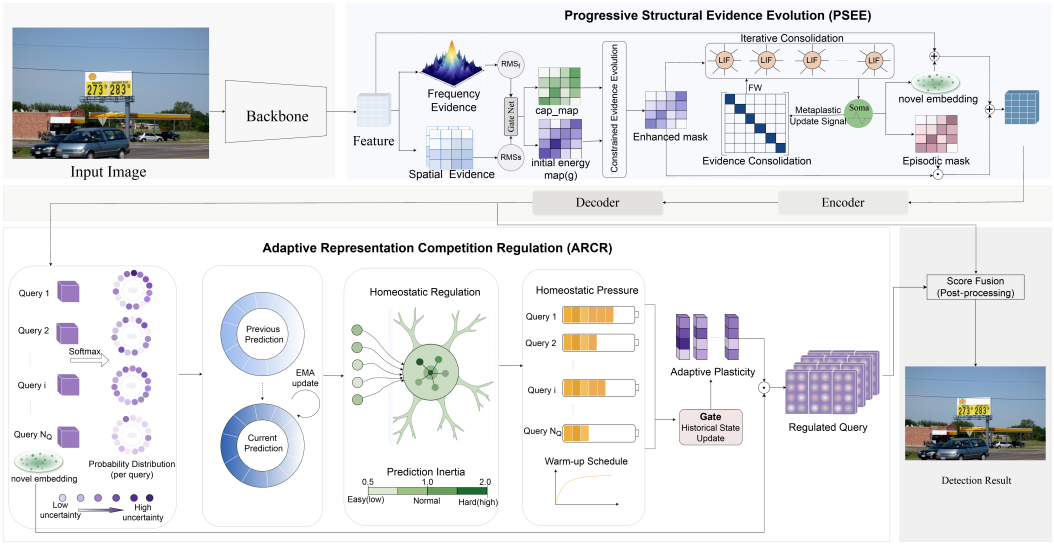


Figure 2: Overview of the proposed EPKA framework. PSEE enables progressive knowledge formation from accumulated evidence, while ARCR selectively regulates representation adaptation based on temporal stability.

comes reliable knowledge, whereas ARCR determines *how strongly* that knowledge should reshape existing representations. Together, they separate knowledge reliability from adaptation strength, enabling progressive knowledge formation and selective integration of novel concepts.

3.2 Progressive Structural Evidence Evolution

Given a backbone feature $z_{i,t}$ from latent representation pathway i at observation step t, PSEE progressively transforms observation-level responses into a consolidated representation of an emerging concept. It comprises three stages: *structural evidence construction*, *evidence refinement*, and *temporal consolidation*. The first two determine which observation-level responses constitute reliable structural evidence, while the last accumulates recurrent evidence into a persistent representation.

Structural Evidence Construction. We project the backbone feature into complementary spatial and frequency response spaces, where the frequency response is generated through adaptive Fourier-domain refinement to capture spectral characteristics. Their responses are RMS-aligned and adaptively fused as:

$$E^0_{i,t} = G_{i,t} \odot \bar{E}^s_{i,t} + (1 - G_{i,t}) \odot \bar{E}^f_{i,t}, \qquad G_{i,t} = \sigma\Big(\phi_g([\bar{E}^s_{i,t}; \bar{E}^f_{i,t}])\Big), \tag{1}$$

where $\bar{E}^s$ and $\bar{E}^f$ denote the RMS-aligned spatial and Fourier-refined frequency responses, respectively, and $\phi_g(\cdot)$ is a lightweight gating function.

Constrained Evidence Refinement. The initial evidence may contain noisy or unstable responses. We therefore formulate evidence refinement as:

$$\begin{aligned} E^*_{i,t} = \arg\min_{E} \quad & \|E - E^0_{i,t}\|_2^2 + \alpha\|\nabla E\|_2^2 + \beta\|E\|_1 \\ \text{s.t.} \quad & 0 \le E \le C_t, \qquad C_t = C_{\min} + (C_{\max} - C_{\min})(1 - \hat{e}_t), \end{aligned} \tag{2}$$

where $\hat{e}_t \in [0, 1]$ denotes the normalized feature energy, and $C_{\min}$ and $C_{\max}$ define the response amplitude range. The operator ∇ denotes the Laplacian smoothing operator with a fixed local convolution kernel. The fidelity, smoothness, sparsity, and amplitude constraints preserve observation consistency, suppress unstable responses, promote selective structures, and limit individual observation dominance. Collectively, they favor structured evidence that is sufficiently reliable to persist.

We solve the constrained problem through unfolded proximal-gradient iterations:

$$E^{k+1}_{i,t} = \Pi_{[0,C_{i,t}]}\left[\mathrm{Prox}_{\eta\beta\|\cdot\|_1}\left(E^k_{i,t} - \eta\nabla F(E^k_{i,t})\right)\right], \qquad F(E) = \|E - E^0_{i,t}\|_2^2 + \alpha\|\nabla E\|_2^2. \tag{3}$$

Each unfolding step progressively refines the evidence under the structural constraints, with the depth K controlling the extent of refinement. We treat K as a mechanism parameter and analyze its sensitivity separately.

Temporal Consolidation. Although $E^*_{i,t}$ captures refined evidence from the current observation, a single observation may be insufficient to establish a persistent concept. PSEE therefore accumulates recurrent evidence through a LIF-style temporal state and a metaplastic Hebbian memory update:

$$v_t = \lambda v_{t-1} + W x_t, \qquad s_t = \mathcal{H}(v_t - \theta), \qquad W_f^{t+1} = (1-\eta_t) W_f^t + \eta_t s_t s_t^T, \tag{4}$$

where v_t is the temporal state, λ is the decay factor, W is the input projection, θ is the firing threshold, s_t is the thresholded response, and W_f^t denotes the fast memory state. The update rate η_t is adaptively regulated by metaplasticity. The resulting consolidated representation $S_{i,t}$ is obtained by integrating the current thresholded response with the accumulated fast-memory state W_f^t.

This mechanism provides temporal selectivity complementary to structural refinement: structural constraints determine *what evidence is retained*, whereas temporal consolidation determines *which retained evidence persists across observations*. Recurrent evidence is progressively reinforced, while transient responses have limited influence on the consolidated representation.

$$z_{i,t}^{\mathrm{PSEE}} = z_{i,t} + W_A(A_{i,t} \odot E^*_{i,t}), \qquad N_{i,t} = \phi_N\left([\mathrm{Pool}(E^*); \mathrm{Pool}(S_{i,t-1}); \mathrm{Pool}((E^*) - S_{i,t-1})]\right), \tag{5}$$

where Pool denotes global average pooling and $A_{i,t} = \sigma(f_A(E^*_{i,t}))$ denotes the episodic integration gate. $z_{i,t}^{\mathrm{PSEE}}$ is forwarded to the detector, while $N_{i,t}$ serves as the consolidated novelty representation for subsequent representation regulation in ARCR.

3.3 Adaptive Representation Competition Regulation

The novel representation produced by PSEE must be integrated into a representation space shaped by previously acquired knowledge. We model the decoder as a set of latent representation pathways, each with its own prediction history. ARCR allocates plasticity across these pathways according to their prediction stability and the novelty of incoming evidence.

We first estimate the historical stability of each representation pathway from its prediction dynamics. Let p_i^t denote the class prediction distribution produced by the i-th pathway at observation step t. It prediction uncertainty and longer-term state are given by:

$$H_i^t = -\sum_c p_{ic}^t \log p_{ic}^t, \qquad U_i^t = \mu U_i^{t-1} + (1-\mu) H_i^t, \tag{6}$$

respectively, where μ controls the temporal smoothing. However, uncertainty alone does not distinguish transient ambiguity from persistent instability. We therefore further measure the consistency between consecutive predictions:

$$I_i^t = \frac{1 + \cos(p_i^t, p_i^{t-1})}{2}, \qquad \widehat{U}_i^t = I_i^t U_i^t, \tag{7}$$

where I_i^t reflects the temporal consistency of the pathway prediction and U_i^t denotes the temporal state that provides the stability signal for adaptive plasticity.

Stability alone, however, can not determine whether a representation should adapt: a stable pathway should resist change unless sufficiently strong evidence indicates the emergence of a novel pattern. We therefore introduce the novel embedding Ni, t from PSEE as the complementary signal of emerging knowledge. This establishes a homeostatic balance between two competing forces: historical stability constrains plasticity to protect established knowledge, while emerging novelty increases the pressure for adaptation. We combine them to define a adaptive pressure:

$$P_i^t = \sigma\left(\gamma g_N(N_{i,t}) - \lambda \widehat{U}_i^t\right), \qquad \bar{P}_i^t = \min\left(1, \frac{t}{T_{\mathrm{warm}}}\right) P_i^t, \tag{8}$$

where $g_N(\cdot)$ maps the novelty representation into a scalar novelty intensity, γ and λ control the contributions of novelty and historical stability, respectively. T_{warm} denotes the warm-up duration. The resulting pressure therefore reflects both sides of the adaptation decision: emerging evidence

provides a reason to change, whereas persistent prediction behavior provides a reason to preserve. The warm-up factor further prevents unreliable early prediction statistics from inducing excessive plasticity.

The pressure finally controls both the current representation update and the retention of the historical state:

$$z_{i,t}^{\mathrm{ARCR}} = z_{i,t}^{\mathrm{PSEE}} + \bar{P}_i^t W_\Delta N_{i,t}, \qquad h_{i,t} = G_{i,t}^h \odot z_{i,t}^{\mathrm{ARCR}} + (1 - G_{i,t}^h) \odot h_{i,t-1}, \tag{9}$$

where $G_{i,t}^h$ denotes the historical retention gate, and W_Δ projects the novelty representation into the adaptation space.

4 Experiments

4.1 Experimental Setup

Object Detection Across Diverse Open-world scenarios: To assess the generality of the proposed evidence-driven learning principle, we evaluate EPKA across three open-world scenarios and five benchmarks, covering diverse object categories and visual domains. Specifically, for *logo detection*, we evaluate on FlickrLogos-32 (32 classes)(Romberg et al., 2011), QMUL-OpenLogo (352 classes)(Su et al., 2018), and FoodLogoDet-1500 (1,500 classes)(Hou et al., 2021). Beyond logos, we further evaluate *fine-grained insect pest detection* using IP102 (Wu et al., 2019), which contains 102 pest categories with substantial intra-class variation and a long-tailed distribution, and *remote-sensing object detection* using SIMD (Haroon et al., 2020), which covers 15 vehicle categories characterized by large variations in object scale, viewpoint, and scene context.

Evaluation Metrics: We report mean Average Precision (mAP) for known-category detection and Unknown Recall (U-Recall) for unknown object discovery. In incremental task, mAP is further decomposed into current known (CK), previously known (PK), and their union (Both).

Implementation Details: We build EPKA upon PROB (Zohar et al., 2023). Each dataset is divided into four incremental tasks. Current-task categories are treated as known, while future-task categories remain unknown. Unless otherwise specified, we use a DINO-pretrained ResNet-50 backbone with 100 latent decoder representation slots and 256-dimensional representation embeddings. All models are optimized with AdamW using a weight decay of 10^{-4}, an initial learning rate of 8×10^{-5}, and a batch size of 1 on a single NVIDIA RTX 3090 GPU.

4.2 Main Results

Results Across Diverse Open-World Scenarios: We evaluate EPKA on five benchmarks covering logo, fine-grained pest, and remote-sensing object detection. As shown in Table 1, EPKA consistently achieves superior unknown discovery and known-category retention across all three logo benchmarks, improving U-Recall by up to 17.3 points and CK by up to 17.8 points over competing methods. For IP102 and SIMD, where open-world protocols and public implementations remain limited, we compare EPKA with representative recent methods under reproducible settings. EPKA improves U-Recall/CK/Both-mAP by up to 1.0/3.2/2.8 points on IP102 and 3.8/8.3/4.9 points on SIMD. These consistent improvements across heterogeneous categories and visual domains demonstrate that EPKA is not tied to a specific detection scenario, but provides a general evidence-driven mechanism for continual knowledge formation: PSEE progressively consolidates reliable evidence, while ARCR selectively adapts emerging representations and protects established knowledge.

We further provide qualitative comparisons between EPKA and competing methods, showing the effectiveness of EPKA. Additional qualitative results are presented in Appendix A.

4.3 Ablation and Mechanism-level Analysis

Ablation of PSEE and ARCR: We first investigate the complementary roles of PSEE and ARCR. Table 2 reports component-level ablations under the same protocol. Removing either component consistently degrades performance across all three benchmarks, while their combination generally achieves the best results. On FlickrLogos-32, for example, adding PSEE alone improves T1 U-Recall from 59.6 to 69.4, while adding ARCR alone further increases it to 74.5. Combining both

Table 1: Comparison with state-of-the-art methods on five detection benchmarks (Best in bold, second-best underlined).

Dataset	Method	Task1 U-recall(%)	Task1 mAP(%) CK	Task2 U-recall(%)	Task2 mAP(%) CK	Task2 mAP(%) PK	Task2 mAP(%) Both	Task3 U-recall(%)	Task3 mAP(%) CK	Task3 mAP(%) PK	Task3 mAP(%) Both	Task4 mAP(%) CK	Task4 mAP(%) PK	Task4 mAP(%) Both
FlickrLogos-32	ORE[CVPR'21]	48.8	56.9	45.2	41.3	53.6	47.4	66.8	27.2	50.2	38.7	43.7	44.0	43.8
	OW-DETR[CVPR'22]	57.4	57.2	57.4	35.1	45.9	40.5	73.9	23.3	45.4	38.0	43.4	43.9	43.7
	CAT[CVPR'23]	57.3	58.1	58.0	44.2	55.2	49.7	70.8	26.7	53.3	44.5	47.9	44.8	45.5
	ALLOW[CVPR'23]	51.2	52.2	50.9	40.0	52.4	46.7	65.3	24.7	50.6	41.6	43.9	40.2	43.3
	PROB[CVPR'23]	59.6	60.0	59.0	43.4	52.0	47.7	71.7	27.5	41.5	36.8	47.7	36.3	39.2
	RandomBox[ICCV'23]	54.0	61.4	57.2	–	–	49.1	47.0	–	–	40.1	–	–	46.7
	EAGLE[ACMMM CL'24]	62.3	64.9	61.0	46.8	58.3	52.5	76.0	33.3	51.5	45.4	57.2	44.8	47.4
	Hyp-OW[AAAI'24]	59.6	57.3	59.1	44.1	50.1	47.0	71.3	27.5	41.4	36.6	47.8	36.3	39.1
	OWOBJ[CVPR'25]	61.5	63.2	60.8	46.0	57.6	53.4	75.2	33.1	51.1	45.3	58.4	46.0	47.3
	Ours	**78.0**	**65.8**	**78.3**	**57.1**	**59.5**	**58.3**	**91.1**	**41.0**	**55.6**	**50.8**	**68.8**	**47.1**	**52.5**
QMUL-OpenLogo	ORE[CVPR'21]	38.3	28.2	39.8	13.2	24.5	18.8	48.8	9.8	19.8	16.5	12.7	17.5	16.8
	OW-DETR[CVPR'22]	43.5	32.1	45.0	12.7	23.9	18.3	56.1	9.3	19.7	16.2	14.1	17.7	16.8
	CAT[CVPR'23]	34.5	32.9	42.5	15.7	**26.2**	21.0	55.9	9.4	19.2	16.3	14.6	17.7	17.0
	ALLOW[CVPR'23]	37.2	32.5	43.7	14.6	24.2	21.0	56.2	10.3	19.9	16.5	14.3	17.0	16.6
	PROB[CVPR'23]	43.8	33.3	48.0	17.2	24.3	20.8	60.6	12.0	18.2	16.1	13.6	15.9	15.3
	RandomBox[ICCV'23]	40.2	34.6	46.0	–	–	21.3	59.9	–	–	14.0	–	–	15.6
	EAGLE[ACMMM CL'24]	53.0	36.1	53.3	17.5	25.3	21.4	63.2	12.3	20.3	17.6	16.4	**17.9**	17.8
	Hyp-OW[AAAI'24]	44.7	34.2	43.3	16.7	25.2	21.0	61.4	11.5	19.2	16.5	14.5	16.2	15.6
	OWOBJ[CVPR'25]	52.0	35.2	53.1	18.1	25.3	21.2	62.6	14.0	**20.9**	18.6	15.6	17.0	16.6
	Ours	**57.0**	**39.6**	**56.5**	**21.1**	25.2	**23.2**	**64.6**	**21.4**	20.0	**20.5**	**26.7**	16.1	**18.2**
FoodLogoDet-1500	ORE[CVPR'21]	27.4	53.2	25.9	39.7	35.8	38.4	17.2	46.8	45.7	46.2	44.6	45.7	45.2
	OW-DETR[CVPR'22]	30.4	55.6	35.7	41.2	34.9	35.7	53.4	40.0	46.5	43.2	48.8	47.2	48.0
	CAT[CVPR'23]	37.4	58.6	44.3	38.8	39.7	41.0	57.2	45.2	52.7	50.2	49.7	51.2	50.8
	ALLOW[CVPR'23]	33.5	56.5	39.9	33.6	31.3	32.6	44.4	34.4	42.5	32.7	34.3	32.7	32.3
	PROB[CVPR'23]	73.9	59.3	73.9	42.1	37.1	40.7	81.7	51.6	49.4	50.2	55.9	46.2	48.6
	RandomBox[ICCV'23]	62.4	57.3	76.2	–	–	41.5	77.7	–	–	48.2	–	–	40.1
	EAGLE[ACMMM CL'24]	75.3	60.2	80.0	43.3	39.0	42.0	82.4	54.2	50.8	51.7	56.3	48.9	49.9
	Hyp-OW[AAAI'24]	75.6	60.4	78.6	41.7	38.0	41.2	82.0	55.0	52.8	51.7	55.1	48.1	48.4
	OWOBJ[CVPR'25]	79.1	62.7	80.0	42.4	38.4	41.9	82.6	56.6	52.6	51.9	56.4	49.0	49.6
	Ours	**84.7**	**69.6**	**88.2**	**55.8**	**54.3**	**55.0**	**86.7**	**57.4**	**55.0**	**55.8**	**74.2**	**54.2**	**59.2**
SIMD	PROB[CVPR'23]	9.3	36.3	8.7	20.2	47.7	34.0	10.1	14.4	33.3	27.0	5.4	26.6	23.6
	RandomBox[ICCV'23]	6.9	-	20	-	-	21.8	9.7	-	-	18.8	-	-	11.8
	EAGLE[ACMMM CL'24]	10.1	38.4	12.1	27.9	49.5	38.7	**11.3**	12.9	28.5	35.3	4.5	30.7	26.9
	Hyp-OW[AAAI'24]	8.8	12.7	19.3	21.5	29.8	27.2	9.9	15.2	23.5	21.2	4.9	28.9	20.5
	OWOBJ[CVPR'25]	8.5	32.7	19.2	20.6	46.7	33.6	10.2	22.9	40.6	34.7	5.3	36.6	31.7
	Ours	**12.0**	**39.4**	**23.1**	**36.2**	**50.9**	**43.6**	11.1	**23.5**	**41.7**	**35.6**	**7.2**	**37.7**	**33.3**
IP102	PROB[CVPR'23]	89.9	55.3	93.1	35.4	55.6	45.5	85.6	36.1	45.4	45.3	41.1	44.3	43.5
	EAGLE[ACMMM CL'24]	90.9	66.6	89.3	38.1	55.4	46.8	84.6	40.0	48.1	45.4	42.1	44.3	43.8
	OWOBJ[CVPR'25]	85.7	51.1	84.6	40.0	48.1	45.4	83.2	31.9	43.4	39.5	39.9	43.8	42.9
	Ours	**91.9**	**68.8**	**94.1**	**39.9**	**59.1**	**49.5**	**86.1**	**43.2**	**50.7**	**48.2**	**44.3**	**47.0**	**46.4**

Table 2: Ablation study of PSEE and ARCR on three open-world logo detection benchmarks.

Dataset	Components Baseline	PSEE	ARCR	Task1 U-recall(%)	Task1 mAP(%) CK	Task2 U-recall(%)	Task2 mAP(%) CK	PK	Both	Task3 U-recall(%)	Task3 mAP(%) CK	PK	Both	Task4 mAP(%) CK	PK	Both
FlickrLogos-32	✓			59.6	60.0	59.0	43.4	52.0	47.7	71.7	27.5	41.5	36.8	47.7	36.3	39.2
	✓	✓		69.4	63.7	71.0	49.4	57.4	53.4	82.2	36.6	54.0	48.2	53.9	43.5	46.1
	✓		✓	74.5	65.1	74.7	40.1	54.4	47.3	90.7	37.1	49.1	45.1	53.9	40.1	43.5
	✓	✓	✓	78.0	65.8	78.3	57.1	59.5	58.3	91.1	41.0	55.6	50.8	68.8	47.1	52.5
QMUL-OpenLogo	✓			43.8	33.3	48.0	17.2	24.3	20.8	60.6	12.0	18.2	16.1	13.6	15.9	15.3
	✓	✓		55.2	38.7	56.4	21.9	25.1	23.5	69.9	25.8	20.4	22.2	27.3	17.3	19.8
	✓		✓	55.2	36.4	54.8	20.9	25.1	23.0	64.5	20.1	17.2	18.1	26.1	16.0	17.8
	✓	✓	✓	57.0	39.6	56.5	21.1	25.2	23.2	64.6	21.4	20.0	20.5	26.7	16.1	18.2
FoodLogoDet-1500	✓			73.9	59.3	73.9	42.1	37.1	40.7	81.7	51.6	49.4	50.2	55.9	46.2	48.6
	✓	✓		80.5	64.8	85.5	46.0	47.8	46.9	84.6	54.7	51.3	52.5	68.6	47.7	52.9
	✓		✓	82.5	65.2	86.9	55.2	55.7	55.5	84.8	55.1	52.0	53.0	68.0	50.9	55.1
	✓	✓	✓	84.7	69.6	88.2	55.8	54.3	55.0	86.7	57.4	55.0	55.8	74.2	54.2	59.2

yields 78.0, demonstrating that the two mechanisms provide complementary benefits. PSEE progressively stabilizes emerging evidence before it enters the representation space, whereas ARCR regulates how strongly the resulting knowledge modifies existing representations. Their combination therefore explicitly couples *knowledge formation* with *selective adaptation*.

Evidence Analysis: We next examine the individual design choices within PSEE on FlickrLogos-32. As shown in Table 3 , removing any evidence constraint consistently reduces performance. In particular, removing the smoothness constraint causes the largest drop in T1 U-Recall, indicating that suppressing isolated and unstable responses is important for reliable evidence formation. The fidelity, sparsity, and amplitude constraints further contribute to Both mAP, suggesting that effective evidence refinement requires jointly preserving observation-supported structures, suppressing irrelevant responses, and preventing individual observations from dominating the evolving knowledge state.

Table 3: Mechanism-level analysis on FlickrLogos-32. T1 U-R, T4 Both mAP, and Avg. U-R denote Task1 U-Recall, Task4 mAP, and average U-Recall, respectively.

Design	Variant	T1 U-R(%) ↑	T4 Both mAP(%) ↑	Avg. U-R(%) ↑
Evidence	w/o Fidelity	76.9	49.0	80.7
	w/o Smoothness	74.9	48.7	80.1
	w/o Sparsity	76.9	50.1	81.0
	w/o Amplitude Bound	76.6	49.7	82.1
Temporal	No Temporal Integration	77.9	51.7	81.9
	Fixed EMA	77.5	51.0	80.9
	EMA + Adaptive Gate	77.9	50.0	82.0
Novelty	w/o Novelty Guidance	77.4	49.2	80.0
Ours	**Full**	**78.0**	**52.5**	**82.5**

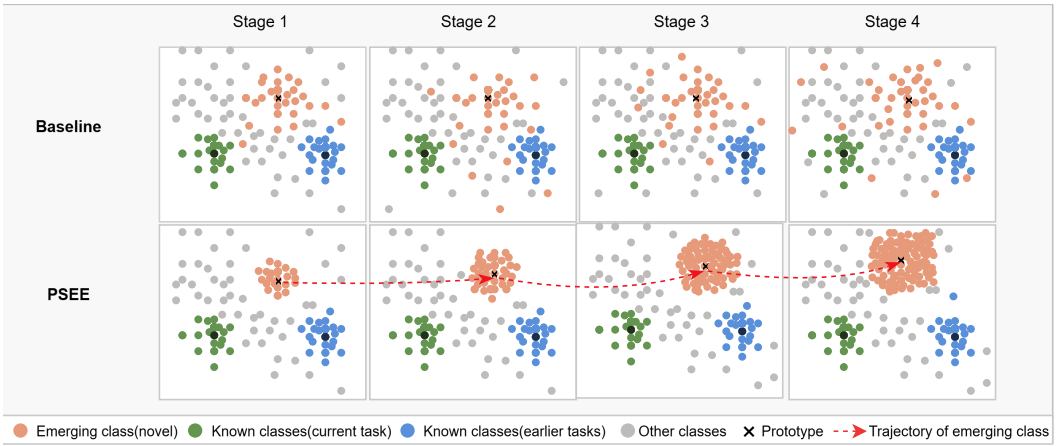


Figure 3: Progressive evolution of an emerging object representation within a single incremental task.

Temporal Consolidation Analysis: Temporal consolidation is essential for progressive knowledge formation: removing it or replacing adaptive updates with fixed EMA consistently degrades performance, showing that evidence accumulation alone cannot preserve persistent structural patterns. EMA with an adaptive gate also remains inferior to the full model, indicating that adaptive accumulation cannot replace explicit novel representation. Removing novelty guidance further reduces Avg. U-Recall from 82.5 to 80.0, confirming the benefit of coupling representation plasticity with evolving novelty. Overall, the results validate the complementary roles of evidence refinement, temporal consolidation, and novelty-aware regulation.

Hyperparameter Sensitivity Analysis: We examine the unfolding depth K, proximal-gradient step size η, spatial regularization weight α and β. Detailed analyses are provided in Appendix A.2.1, Appendix A.2.2, Appendix A.2.3 and Appendix A.2.4, respectively.

4.4 Dynamic Analysis

Progressive Evidence Stabilization: First, we examine how an emerging logo evolves as compatible observations accumulate. For a representative unseen category, we track its evidence and representation states over successive observation windows within a single incremental task, and compare their evolution with the baseline. This analysis reveals whether PSEE progressively transforms unstable observations into a more coherent and stable knowledge representation. Figure 3 shows that the emerging representation becomes progressively more compact as observations accumulate, whereas the baseline remains more sensitive to individual observations and exhibits greater dispersion. This supports the proposed evidence-evolution process, which suppresses transient responses and progressively consolidates compatible evidence.

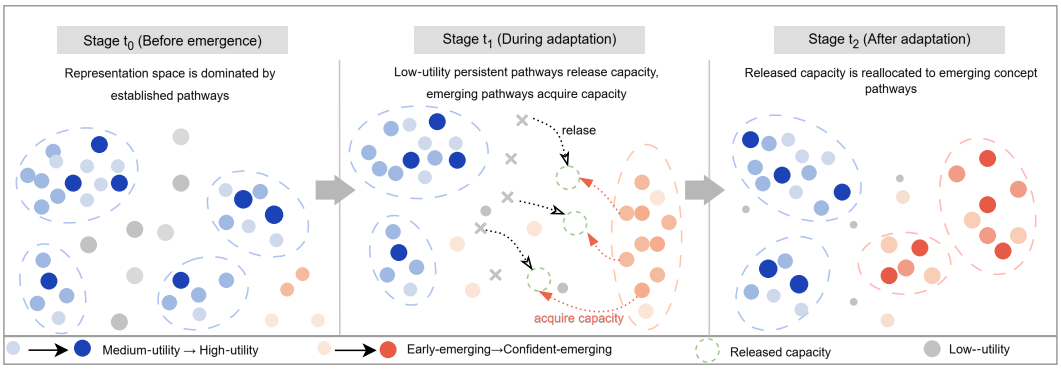


Figure 4: Adaptive representation plasticity during within-task adaptation.

Adaptive Representation Regulation: We next examine how ARCR regulates the incorporation of emerging concepts. We track the adaptation of the emerging representation and the changes in established representations over successive observations, illustrating how EPKA selectively incorporates sufficiently supported novel knowledge while limiting unnecessary changes to established representations. Figure 4 shows that stable representation pathways receive increasingly smaller updates, while pathways associated with emerging concepts retain greater plasticity. This selective adaptation protects established representations from unnecessary changes while allocating sufficient plasticity to accommodate new evidence.

4.5 Discussion on the Applicability of EPKA

The gains of EPKA depend not only on the emergence of novel categories, but also on the uncertainty of representation formation and the degree of interference between emerging and established knowledge. PSEE progressively consolidates reliable evidence when individual observations are insufficient to form stable representations, while ARCR regulates plasticity when new knowledge interacts with existing representations. EPKA is therefore particularly suited to continual learning scenarios characterized by ambiguous visual evidence, high intra-class variation, strong inter-class similarity, or substantial representation interference, as exemplified by the Logo, IP102, and SIMD benchmarks.

The negligible gains on COCO (Lin et al., 2014) further support this observation. Unlike these domain-specific benchmarks, COCO contains a broad range of common object categories with relatively well-established and discriminative visual structures. For many categories, effective representations can be formed from individual observations with limited evidence accumulation and interference, reducing the marginal benefit of PSEE and ARCR. Thus, the results suggest that EPKA is most beneficial when continual adaptation requires both *progressive knowledge formation* and *selective representation adaptation*, rather than merely accommodating newly introduced categories. A promising direction is to make these mechanisms difficulty-aware by incorporating feature separability, inter-class similarity, and sample ambiguity to dynamically allocate evidence accumulation and representation plasticity.

5 Conclusion

We presented EPKA, an evidence-driven continual learning framework for open-world object detection that decouples progressive knowledge formation from selective representation adaptation. PSEE consolidates compatible evidence into stable novel representations, while ARCR regulates representation plasticity to incorporate new knowledge without unnecessarily altering established representations. Experiments across five benchmarks and three open-world detection scenarios demonstrate consistent gains in novel object discovery and known-category recognition. EPKA thus emphasizes the importance of separating *when knowledge is formed* from *how strongly representations adapt*, providing an evidence-knowledge-adaptation perspective for open-world continual learning.

USE OF GENERATIVE AI

In this work, we used generative AI solely as auxiliary tools for improving the clarity and readability of the manuscript, including grammar checking, language polishing, and refinement of academic expressions. We have not used generative AI tools for research ideas, methodology design, experiments, data analysis, interpretation of results, and conclusions. All AI-assisted content was carefully reviewed and revised by the authors, and we take full responsibility for the final content of this work, including all text, claims, and research contributions.

REFERENCES

Abhijit Bendale and Terrance E. Boult. Towards open set deep networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1563–1572, 2016. doi: 10.1109/CVPR.2016.173.

Akshita Gupta, Sanath Narayan, K J Joseph, Salman Khan, Fahad Shahbaz Khan, and Mubarak Shah. Ow-detr: Open-world detection transformer. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9225–9234, 2022. doi: 10.1109/CVPR52688.2022.00902.

Muhammad Haroon, Muhammad Shahzad, and Muhammad Moazam Fraz. Multisized object detection using spaceborne optical imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:3032–3046, 2020. doi: 10.1109/JSTARS.2020.3000317.

Qiang Hou, Weiqing Min, Jing Wang, Sujuan Hou, Yuanjie Zheng, and Shuqiang Jiang. Foodlogodet-1500: A dataset for large-scale food logo detection via multi-scale feature decoupling network. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM '21, pp. 4670–4679, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450386517. doi: 10.1145/3474085.3475289. URL https://doi.org/10.1145/3474085.3475289.

K J Joseph, Salman Khan, Fahad Shahbaz Khan, and Vineeth N Balasubramanian. Towards open world object detection. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5826–5836, 2021. doi: 10.1109/CVPR46437.2021.00577.

Yujie Li, Xiangkun Wang, Xin Yang, Marcello Bonsangue, Junbo Zhang, and Tianrui Li. Improving open-world continual learning under the constraints of scarce labeled data. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, pp. 1647–1658, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714542. doi: 10.1145/3711896.3737004. URL https://doi.org/10.1145/3711896.3737004.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (eds.), *Computer Vision – ECCV 2014*, pp. 740–755, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10602-1.

James L McClelland, Bruce L McNaughton, and Randall C O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995. doi: 10.1037/0033-295X.102.3.419.

Mark D. McDonnell, Dong Gong, Amin Parveneh, Ehsan Abbasnejad, and Anton van den Hengel. Ranpac: random projections and pre-trained models for continual learning. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc.

Fuli Qiao and Mehrdad Mahdavi. Merge before forget: A single loRA continual learning via continual merging. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=i1Rj7yU6eF.

Stefan Romberg, Lluis Garcia Pueyo, Rainer Lienhart, and Roelof van Zwol. Scalable logo recognition in real-world images. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, ICMR '11, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450303361. doi: 10.1145/1991996.1992021. URL https://doi.org/10.1145/1991996.1992021.

Qianqian Shi, Faqiang Liu, Hongyi Li, Guangyu Li, Luping Shi, and Rong Zhao. Hybrid neural networks for continual learning inspired by corticohippocampal circuits. *Nature Communications*, 16:1272, 2025. doi: 10.1038/s41467-025-56405-9.

Hang Su, Xiatian Zhu, and Shaogang Gong. Open logo detection challenge. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2018.

Weinan Sun, Madhu Advani, Nelson Spruston, Andrew Saxe, and James E Fitzgerald. Organizing memories for generalization in complementary learning systems. *Nature Neuroscience*, 26(8): 1438–1448, 2023. doi: 10.1038/s41593-023-01382-9.

Jochen Triesch, Anh Duong Vo, and Anne-Sophie Hafner. Competition for synaptic building blocks shapes synaptic plasticity. *eLife*, 7:e37836, 2018. doi: 10.7554/eLife.37836.

Yiwen Wang, Diana Benavides-Prado, and Yun Sing Koh. Detect, decide, unlearn: A transfer-aware framework for continual learning. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=Lej4WvdpFE.

Xiaoping Wu, Chi Zhan, Yu-Kun Lai, Ming-Ming Cheng, and Jufeng Yang. Ip102: A large-scale benchmark dataset for insect pest recognition. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8779–8788, 2019. doi: 10.1109/CVPR.2019.00899.

Xing Xi, Xing Fu, Weiqiang Wang, and Ronghua Luo. OW-VAP: Visual attribute parsing for open world object detection. In *Forty-second International Conference on Machine Learning*, 2025a. URL https://openreview.net/forum?id=OXIIRxwJwx.

Xing Xi, Yangyang Huang, Ronghua Luo, and Yu Qiu. Ow-ovd: Unified open world and open vocabulary object detection. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25454–25464, 2025b. doi: 10.1109/CVPR52734.2025.02370.

Linhua Ye, Xing Xi, and Ronghua Luo. Ow-dar: Dual-granularity adaptive reconstruction-error modeling for open-world object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 11946–11954, 2026. doi: 10.1609/aaai.v40i14.38182. URL https://ojs.aaai.org/index.php/AAAI/article/view/38182.

Zhongming Yuan, Hao Xiong, and Sujuan Hou. Eagle network: A novel incremental learning framework for detecting unknown logos in open-world environments. In *Proceedings of the 1st on Continual Learning Meets Multimodal Foundation Models: Fundamentals and Advances*, ACMMM CL'24, pp. 23–30, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400711886. doi: 10.1145/3688859.3690081. URL https://doi.org/10.1145/3688859.3690081.

Gengwei Zhang, Liyuan Wang, Guoliang Kang, Ling Chen, and Yunchao Wei. Slca: Slow learner with classifier alignment for continual learning on a pre-trained model. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19091–19101, 2023. doi: 10.1109/ICCV51070.2023.01754.

Shan Zhang, Yao Ni, Jinhao Du, Yuan Xue, Philip Torr, Piotr Koniusz, and Anton van den Hengel. Open-world objectness modeling unifies novel object detection. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30332–30342, 2025. doi: 10.1109/CVPR52734.2025.02824.

Orr Zohar, Kuan-Chieh Wang, and Serena Yeung. Prob: Probabilistic objectness for open world object detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11444–11453, 2023. doi: 10.1109/CVPR52729.2023.01101.

A Appendix

A.1 Qualitative Results

We further provide qualitative comparisons between EPKA and competing methods. As shown in Figure 5, EPKA effectively discovers previously unseen objects while maintaining accurate localization of previously learned categories. These examples qualitatively corroborate the complementary roles of progressive evidence formation and selective adaptation in balancing novel-category discovery with the retention of established knowledge.

A.2 Hyperparameter Sensitivity Analysis

A.2.1 Sensitivity to Unfolding Depth

We investigate the effect of the proximal-gradient iteration count K on the constrained evidence solver using FlickrLogos-32. As shown in Table 4, performance improves rapidly with increasing K and reaches its best overall performance at $K = 5$, after which additional iterations yield little or no benefit. This diminishing-return behavior suggests that five iterations provide sufficient evidence refinement while avoiding unnecessary computational overhead. We therefore set $K = 5$ as the default in all experiments.

A.2.2 Sensitivity to Optimization Step Size

We investigate the effect of the proximal-gradient step size η on the constrained evidence solver using FlickrLogos-32. As shown in Table 5, performance improves as η increases from small values and reaches its best overall performance at $\eta = 0.6$, while larger step sizes provide little or no further benefit. This behavior indicates that a moderate step size enables effective evidence refinement without causing overly aggressive updates. We therefore set $\eta = 0.6$ as the default in all experiments.

A.2.3 Sensitivity to Spatial Regularization

We further examine the sensitivity to the spatial regularization weight α, which controls the strength of the Laplacian-based smoothness constraint in the evidence solver. Using FlickrLogos-32, we vary α while keeping the other hyperparameters fixed and report the results in Table 6. The results show that moderate values of α strike a favorable balance between local structural consistency and the preservation of informative evidence, whereas overly small or large values lead to less effective refinement. The best overall performance is achieved at $\alpha = 0.16$, which we therefore use as the default setting in all experiments.

A.2.4 Sensitivity to Sparsity Regularization

We also study the effect of the sparsity regularization weight β, which determines the strength of the soft-thresholding operation in the proximal evidence refinement. We vary β on FlickrLogos-32 while fixing the remaining hyperparameters.Table 7 reports the quantitative results across tested values of β. The results show that a moderate sparsity constraint effectively suppresses weak or noisy responses while preserving informative structural evidence. In contrast, an overly weak constraint provides insufficient suppression, whereas an overly strong constraint may remove useful evidence. The best overall performance is obtained at $\beta = 0.015$, which is adopted as the default setting throughout our experiments.

Table 4: Sensitivity to the number of unfolded proximal-gradient iterations K on FlickrLogos-32. Performance is measured by average U-Recall and Task4 Both mAP.

Iterations K	T1 U-R ↑	T4 Both mAP ↑	Avg. U-R ↑
$K = 1$	71.5	51.0	79.2
$K = 2$	74.0	51.6	79.6
$K = 3$	76.5	51.9	81.4
$K = 4$	77.6	52.0	81.6
$\mathbf{K = 5}$	**78.0**	**52.5**	**82.5**
$K = 6$	77.2	51.7	81.2

Table 5: Sensitivity to the proximal-gradient step size η on FlickrLogos-32. Performance is measured by average U-Recall and Task4 Both mAP.

Step Size η	T1 U-R ↑	T4 Both mAP ↑	Avg. U-R ↑
$\eta = 0.2$	76.5	50.9	80.7
$\eta = 0.4$	77.7	51.5	81.9
$\eta = \mathbf{0.6}$	**78.0**	**52.5**	**82.5**
$\eta = 0.8$	77.6	52.2	82.2

Table 6: Sensitivity to the spatial regularization weight α on FlickrLogos-32. Performance is measured by average U-Recall and Task4 Both mAP.

Smoothness Weight α	T1 U-R ↑	T4 Both mAP ↑	Avg. U-R ↑
$\alpha = 0.12$	76.9	51.7	81.3
$\alpha = \mathbf{0.16}$	**78.0**	**52.5**	**82.5**
$\alpha = 0.20$	77.8	51.3	81.0

Table 7: Sensitivity to the sparsity regularization weight β on FlickrLogos-32. Performance is measured by average U-Recall and Task4 Both mAP.

Sparsity Weight β	T1 U-R ↑	T4 Both mAP ↑	Avg. U-R ↑
$\beta = 0.010$	77.3	51.1	82.2
$\beta = \mathbf{0.015}$	**78.0**	**52.5**	**82.5**
$\beta = 0.020$	76.1	52.0	81.9

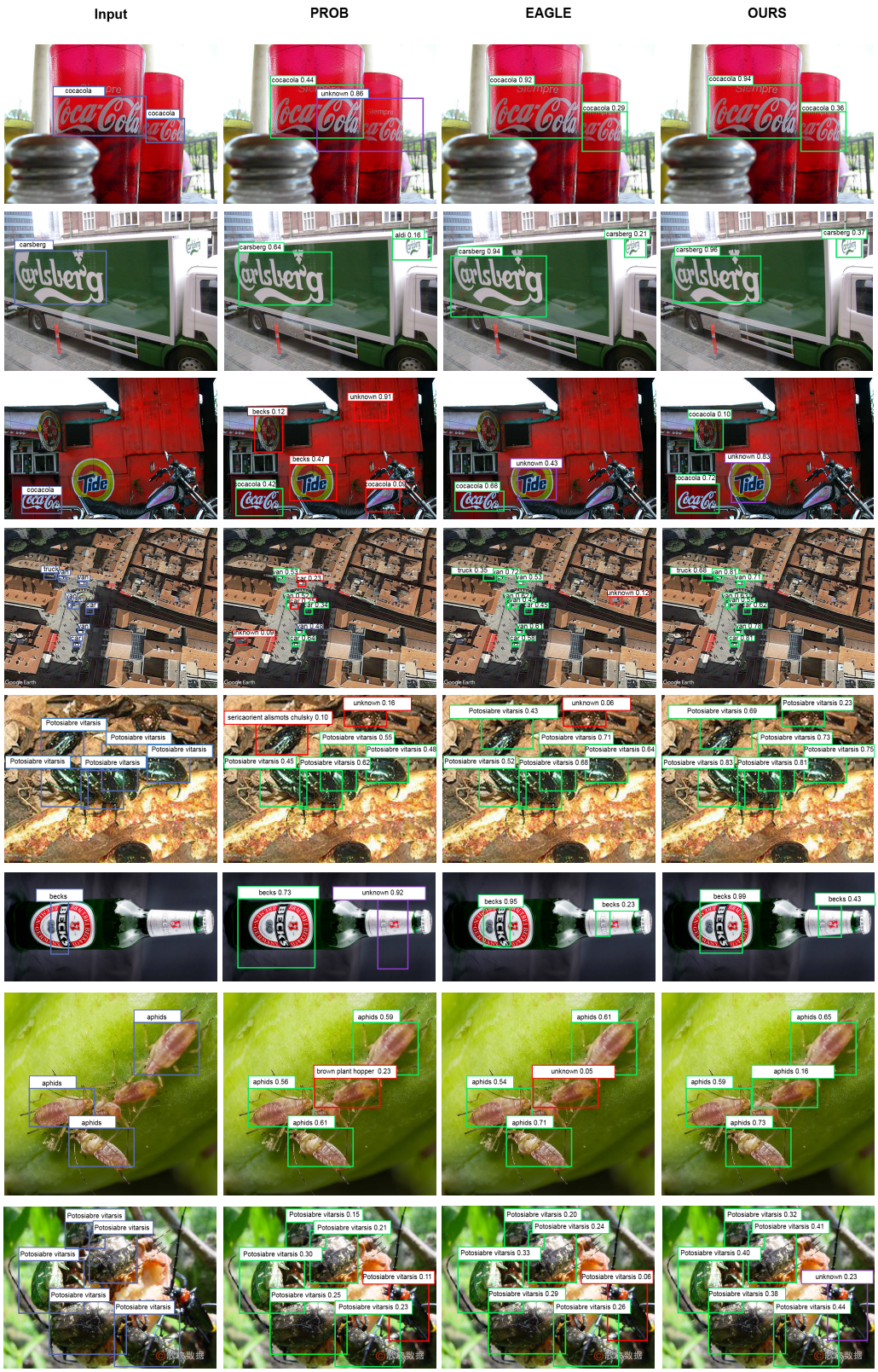


Figure 5: Qualitative comparison of open-world object detection results among PROB (Zohar et al., 2023), EAGLE (Yuan et al., 2024), and our approach. Blue boxes indicate annotated object instances, green boxes represent correctly detected known categories, red boxes denote false positives, and purple boxes highlight discovered unknown objects.