SEMANTIC ANCHOR LEARNING WITH CONFLICT-AWARE EVIDENTIAL INFERENCE FOR CLIP-BASED FEW-SHOT OPEN-SET RECOGNITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Prompt learning for vision-language models has shown strong performance in few-shot classification and has recently been extended to few-shot open-set recognition through unknown-text modeling. Existing methods construct surrogate unknown semantics using manually selected open words, generated pseudo samples, or statistically derived unknown directions, but they provide limited control over the geometry between known and unknown prompts and typically make open-set decisions from a single response score. We propose *Semantic Anchor Learning with Conflict-Aware Evidential Inference* (SAEC), a CLIP-based framework that couples unknown semantic construction, boundary regularization, and evidential decision making. SAEC first factorizes auxiliary token embeddings into a sparse dictionary and composes its primitives into synthesized unknown semantic anchors. It then regularizes the shared embedding space with unknown-aware probability regularization, explicit known–unknown separation, and intra-unknown dispersion. At inference time, SAEC converts competing known and unknown similarities into a Dempster–Shafer-inspired mass representation with a residual ignorance term. Across eight standard few-shot open-set benchmarks, SAEC improves AUROC over the directly comparable UTL baseline on all eight settings, with an average gain of 1.35 points and a maximum gain of 3.58 points, while achieving the best closed-set accuracy on six settings among the reported baselines.

1 INTRODUCTION

Few-shot open-set recognition (FSOR) requires a model to recognize known classes from only a few labeled examples while rejecting test samples from classes absent during training (Boudiaf et al., 2023; Huang et al., 2022; Jeong et al., 2021; Wang et al., 2023; Zhang et al., 2024; Deng et al., 2023). Vision-language models such as CLIP (Radford et al., 2021) provide a strong foundation for this setting because textual prompts can transfer semantic information across visual domains. Prompt learning methods such as CoOp and CoCoOp (Zhou et al., 2022b;a) adapt CLIP by optimizing continuous context tokens, but they are usually designed for closed-set classification where every test class has a known class name.

The central difficulty in CLIP-based FSOR is therefore not only how to tune prompts for known classes, but also how to construct a useful representation of the unknown side of the decision boundary. Prior CLIP-based open-set methods approximate unknown semantics through manually selected open words, generated pseudo samples, or universal unknown prompts (Liao et al., 2025; Qu et al., 2023; Ma et al., 2025). Among them, UTL (Ma et al., 2025) is the most direct baseline: it constructs a universal unknown-word space from auxiliary labels through PCA and trains unknown text with unknown label smoothing. This design is effective, but it still leaves three limitations. First, a PCA-derived unknown basis is statistical rather than explicitly compositional, so its unknown directions are not parameterized through reusable semantic components. Second, its unknown-learning objective mainly operates on output probabilities and does not explicitly regularize the geometry between known and unknown prompts in the CLIP embedding space. Third, its open-set inference is driven by a response score that does not separately represent known support, unknown support, and residual ignorance.

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.

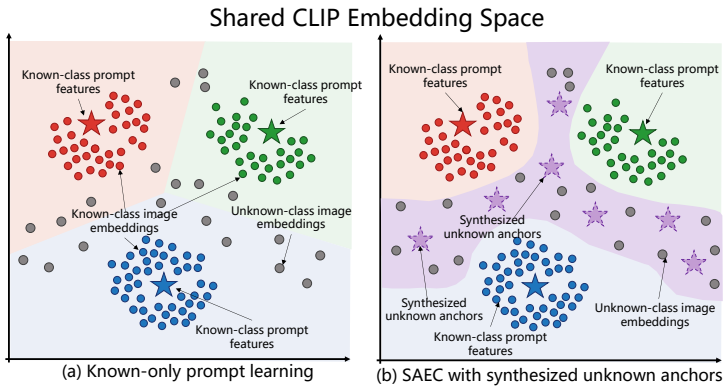


Figure 1: Conceptual illustration in the shared CLIP embedding space. (a) In known-only prompt learning, the model is guided only by known-class prompt features, so unknown-class image embeddings in inter-class or open-space regions are not explicitly modeled. (b) SAEC introduces synthesized unknown anchors and explicitly regularizes their geometry relative to known prompts, leading to more structured unknown-region coverage.

This paper addresses these limitations by coupling unknown prompt construction, boundary shaping, and evidential inference. The key distinction from UTL is not a single replacement of PCA with dictionary learning. Instead, SAEC changes the full modeling chain: unknown semantics are parameterized through sparse dictionary compositions, these anchors are explicitly regularized in the shared embedding space, and open-set decisions are made from competing known and unknown evidence. The methodological contribution therefore lies in coupling structured unknown-text parameterization, prompt-space regularization, and competing-evidence inference under a frozen CLIP backbone, rather than in sparse coding in isolation. The resulting framework keeps both CLIP encoders frozen and optimizes only the shared context vectors and unknown-anchor composition weights for each target task.

Our contributions are as follows:

- We propose Syntax-Decoupled Semantic Anchor Synthesis (SDSA), which factorizes auxiliary CLIP token embeddings into sparse dictionary atoms and composes them into synthesized unknown anchors.
- We introduce Unknown-Aware Semantic Anchor Regularization (USAR), which combines unknown-aware probability regularization with explicit known–unknown separation and intra-unknown dispersion.
- We develop Conflict-Aware Evidential Inference (CEI), which maps competing known and unknown support into a Dempster–Shafer-inspired mass representation with an explicit residual ignorance term for open-set scoring.
- We evaluate the framework on eight FSOR benchmark settings and provide ablations for the main components and hyperparameters.

Figure 1 illustrates the motivation of SAEC in the shared CLIP embedding space. With only known-class prompt features, unknown samples in inter-class regions are not explicitly modeled. SAEC addresses this issue by introducing synthesized unknown anchors and explicitly regularizing their geometry relative to known prompts.

2 Related Work

Prompt learning for vision-language models. CLIP (Radford et al., 2021) aligns visual and textual representations through large-scale contrastive pretraining. Prompt learning methods such as CoOp (Zhou et al., 2022b) and CoCoOp (Zhou et al., 2022a) adapt CLIP to downstream tasks by replacing hand-crafted templates with learnable context tokens. These methods are effective under few-shot supervision, but they assume that the test label space is known. Directly applying closed-

set prompt learning to FSOR is insufficient because unknown classes have no explicit class-name prompts during training.

Few-shot open-set recognition. Classical FSOR methods extend metric learning, prototype learning, and open-space modeling to data-limited settings (Nag et al., 2023; Boudiaf et al., 2023; Huang et al., 2022; Jeong et al., 2021; Wang et al., 2023; Zhang et al., 2024; Deng et al., 2023). Recent CLIP-based methods exploit textual semantics to model open-set boundaries. M-Tuning (Liao et al., 2025) uses manually selected open words, LMC (Qu et al., 2023) relies on large model collaboration to synthesize pseudo unknown image-text pairs, and UTL (Ma et al., 2025) constructs universal unknown text from auxiliary labels without explicit unknown samples. SAEC follows the same FSOR protocol as UTL but reformulates unknown-text construction through semantic anchors and augments open-set inference with an evidential score derived from known, unknown, and residual-ignorance masses.

Evidential uncertainty and open-set inference. Dempster–Shafer theory represents belief masses over hypotheses and over the full frame of discernment, making it useful for describing insufficient evidence (Dempster, 2008; Shafer, 1976). Recent evidential deep learning methods introduce such mass or opinion representations for uncertainty modeling and open-set recognition (Sensoy et al., 2018; Tong et al., 2021; Zhao et al., 2023; Qu et al., 2024). In this paper, we use the DST mass representation conservatively: CEI does not claim calibrated posterior epistemic uncertainty or perform Dempster's combination rule. Instead, it constructs an evidential decision score from known and unknown support while reserving an explicit mass for residual ignorance.

3 METHOD

3.1 PROBLEM SETTING AND OVERVIEW

Let $\mathcal{Y}^C = \{1, \ldots, N_C\}$ denote the known classes available during training. In the 16-shot setting, each known class provides only a small labeled support set, while test data may contain both known-class samples and unknown-class samples. We use CLIP (Radford et al., 2021) with a frozen image encoder $\mathcal{I}(\cdot)$ and text encoder $\mathcal{T}(\cdot)$. For image x and prompt P_j, we obtain the encoder features $\mathbf{f}_x = \mathcal{I}(x)$ and $\mathbf{t}_j = \mathcal{T}(P_j)$, and compute their cosine similarity as

$$s_j(x) = \frac{\mathbf{t}_j^\top \mathbf{f}_x}{\|\mathbf{t}_j\|_2 \|\mathbf{f}_x\|_2}. \tag{1}$$

SAEC consists of three stages. SDSA constructs unknown prompts from sparse dictionary-based anchors. USAR trains known and unknown prompts with probability and embedding-space constraints. CEI performs open-set inference through belief–plausibility scoring. Figure 2 provides the overall training and inference pipeline and shows how the three components are connected.

3.2 SYNTAX-DECOUPLED SEMANTIC ANCHOR SYNTHESIS

SDSA separates the shared prompt context from unknown-specific semantic content. Here, *syntax-decoupled* refers operationally to separating the shared prompt structure from the unknown semantic slot; it does not imply explicit syntactic parsing. The shared context is reused across prompts, while the class-name slot of an unknown prompt is replaced by a synthesized semantic anchor. The anchor is not an unconstrained free vector; it is a learnable composition over a sparse dictionary obtained from auxiliary token embeddings. We use *semantic* to denote that these atoms are learned in CLIP's token-embedding space, without requiring each individual atom to be human-interpretable.

Let $\mathcal{Y}^a = \{y_i^a\}_{i=1}^M$ denote an auxiliary vocabulary filtered to avoid direct overlap with the target known and unknown labels. For each auxiliary name, we obtain token embeddings from CLIP and remove special tokens. The remaining token embeddings form

$$Z = [z_1, z_2, \ldots, z_N]^\top \in \mathbb{R}^{N \times d}. \tag{2}$$

We learn a dictionary $D = [d_1, \ldots, d_K]^\top \in \mathbb{R}^{K \times d}$ and sparse codes $A \in \mathbb{R}^{N \times K}$ by

$$\min_{D,A} \frac{1}{2}\|Z - AD\|_F^2 + \alpha \sum_{i=1}^{N} \|\mathbf{a}_i\|_1, \quad \text{s.t. } \|d_k\|_2 \le 1. \tag{3}$$

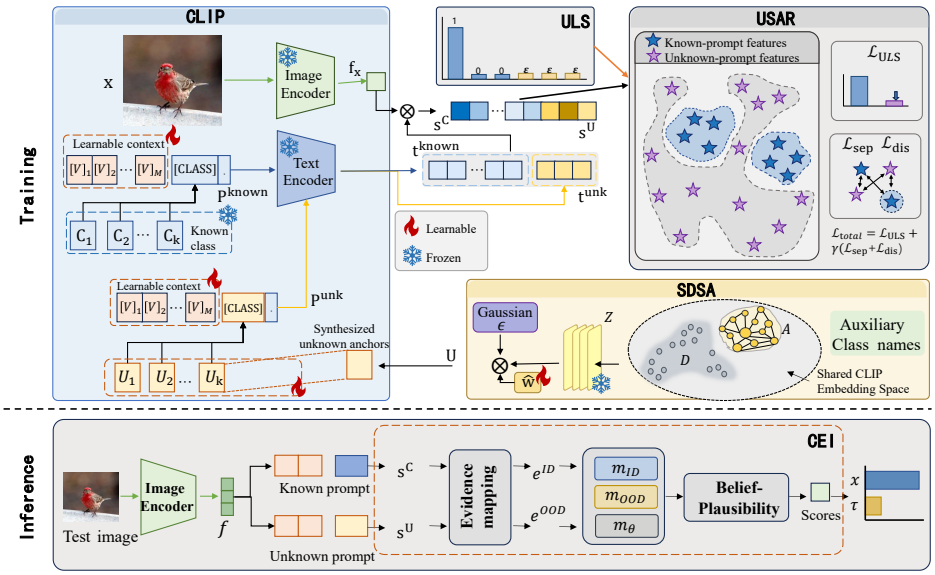


Figure 2: SAEC framework. During training, SDSA synthesizes unknown anchors and USAR optimizes known prompts and unknown-anchor compositions. During inference, CEI converts competing known and unknown similarities into evidential masses, forms a belief–plausibility interval for the known hypothesis, and produces the final open-set score.

The linear sparse model is used deliberately at the token-embedding level. It does not assume that unknown categories lie on a globally linear manifold. Instead, it factorizes frozen CLIP token embeddings into reusable embedding directions with sparse activations. The final unknown text features remain nonlinear in the synthesized anchor because the anchor is later processed by the frozen CLIP text Transformer. Equation 3 is solved once on the auxiliary token matrix. Then D is normalized and fixed during target-task training, which prevents the auxiliary dictionary from drifting toward the few observed known classes.

Given the normalized dictionary $\tilde{D}$, the k-th unknown anchor is generated from a learnable composition vector $\mathbf{w}_k \in \mathbb{R}^K$. During training, we add a Gaussian perturbation $\boldsymbol{\xi}_k \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$:

$$\hat{\mathbf{w}}_k = \mathbf{w}_k + \boldsymbol{\xi}_k, \qquad \mathbf{u}_k = \hat{\mathbf{w}}_k \tilde{D}, \qquad U_k = \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|_2}. \tag{4}$$

For known class y_j and synthesized unknown anchor U_k, the corresponding prompts are

$$P_j^C = [\texttt{SOS}, V, \mathbf{v}(y_j), \texttt{EOS}], \qquad P_k^U = [\texttt{SOS}, V, U_k, \texttt{EOS}], \tag{5}$$

where $V = [\mathbf{v}_1, \ldots, \mathbf{v}_L]$ denotes learnable context tokens. The text features are $\mathbf{t}_j^C = \mathcal{T}(P_j^C)$ and $\mathbf{t}_k^U = \mathcal{T}(P_k^U)$. Gaussian perturbations are used only during training; at inference, $\boldsymbol{\xi}_k = \mathbf{0}$ and the learned composition weights generate deterministic unknown anchors.

3.3 Unknown-aware semantic anchor regularization

USAR jointly regularizes the response to unknown prompts and constrains prompt geometry. Let $s_j^C(x)$ and $s_k^U(x)$ denote the similarities between image x and the j-th known prompt or k-th unknown prompt. The unknown-prompt probability under the joint known/unknown normalization is

$$p_k^U(x) = \frac{\exp(s_k^U(x)/\tau)}{\sum_{j=1}^{N_C} \exp(s_j^C(x)/\tau) + \sum_{r=1}^{N_u} \exp(s_r^U(x)/\tau)}. \tag{6}$$

For a known sample with label y_t, the unknown-aware auxiliary loss is

$$\ell_O(x) = -\log p_{y_t}^C(x) - \varepsilon \sum_{k=1}^{N_u} \log p_k^U(x), \tag{7}$$

where ε is the smoothing factor and τ denotes the temperature parameter used to scale cosine similarities before the joint known–unknown softmax, following UTL (Ma et al., 2025). The unknown-prompt weights constitute an auxiliary weighting scheme rather than a normalized target distribution; their overall contribution is controlled by λ. The training loss combines this auxiliary term with the standard known-class cross entropy:

$$\mathcal{L}_{\mathrm{ULS}} = (1-\lambda)(-\log p^C_{y_t}(x)) + \lambda \ell_O(x). \tag{8}$$

To explicitly shape the embedding space, we penalize excessive similarity between unknown anchors and known prompts,

$$\mathcal{L}_{\mathrm{sep}} = \frac{1}{N_u N_C} \sum_{i=1}^{N_u} \sum_{j=1}^{N_C} \mathrm{ReLU}(\mathrm{sim}(\mathbf{t}^U_i, \mathbf{t}^C_j) - m), \tag{9}$$

and prevent unknown anchors from collapsing into a narrow region,

$$\mathcal{L}_{\mathrm{dis}} = \frac{2}{N_u(N_u-1)} \sum_{1 \le i < j \le N_u} \mathrm{ReLU}(\mathrm{sim}(\mathbf{t}^U_i, \mathbf{t}^U_j) - \delta)^2. \tag{10}$$

The overall objective is

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{ULS}} + \gamma(\mathcal{L}_{\mathrm{sep}} + \mathcal{L}_{\mathrm{dis}}). \tag{11}$$

3.4 Conflict-aware evidential inference

At inference time, CEI converts known and unknown similarities into nonnegative evidence values:

$$e^{\mathrm{ID}}_j = \mathrm{Softplus}(s^C_j(x)), \qquad e^{\mathrm{OOD}}_k = \mathrm{Softplus}(s^U_k(x)). \tag{12}$$

We retain the strongest support from each branch,

$$e^{\mathrm{ID}} = \max_j e^{\mathrm{ID}}_j, \qquad e^{\mathrm{OOD}} = \max_k e^{\mathrm{OOD}}_k. \tag{13}$$

Under the binary frame $\Theta = \{\mathrm{ID}, \mathrm{OOD}\}$, a positive prior mass C_0 reserves support for unresolved evidence. With $S = e^{\mathrm{ID}} + e^{\mathrm{OOD}} + C_0$, we assign

$$m_{\mathrm{ID}} = \frac{e^{\mathrm{ID}}}{S}, \qquad m_{\mathrm{OOD}} = \frac{e^{\mathrm{OOD}}}{S}, \qquad m_\Theta = \frac{C_0}{S}. \tag{14}$$

The masses satisfy $m_{\mathrm{ID}} + m_{\mathrm{OOD}} + m_\Theta = 1$. In the resulting belief-function representation (Shafer, 1976), the lower and upper support bounds for the known hypothesis are

$$\mathrm{Bel}(\mathrm{ID}) = m_{\mathrm{ID}}, \qquad \mathrm{Pl}(\mathrm{ID}) = m_{\mathrm{ID}} + m_\Theta = 1 - m_{\mathrm{OOD}}. \tag{15}$$

CEI scalarizes this belief interval multiplicatively:

$$\mathrm{Score}(x) = \mathrm{Bel}(\mathrm{ID})\,\mathrm{Pl}(\mathrm{ID}) = m_{\mathrm{ID}}(1 - m_{\mathrm{OOD}}) = \frac{e^{\mathrm{ID}}(e^{\mathrm{ID}} + C_0)}{(e^{\mathrm{ID}} + e^{\mathrm{OOD}} + C_0)^2}. \tag{16}$$

This score jointly requires a strong lower support bound and a high plausibility bound. Its square root is the geometric mean of the two bounds and is rank-equivalent to Equation 16, so the two produce identical AUROC rankings.

Monotonicity. Let $a = e^{\mathrm{ID}}$, $b = e^{\mathrm{OOD}}$, and $c = C_0 > 0$. Then

$$\frac{\partial\,\mathrm{Score}}{\partial a} = \frac{(2a+c)b + ac + c^2}{(a+b+c)^3} > 0, \qquad \frac{\partial\,\mathrm{Score}}{\partial b} = -\frac{2a(a+c)}{(a+b+c)^3} \le 0. \tag{17}$$

Thus, CEI is strictly increasing in known evidence and non-increasing in competing unknown evidence, with strict decrease in e^{OOD} whenever $e^{\mathrm{ID}} > 0$. Moreover, when $m_{\mathrm{OOD}} = 0$, Equation 16 reduces to $\mathrm{Bel}(\mathrm{ID})$; otherwise, competing unknown mass discounts the known belief. We therefore use *conflict-aware* to mean explicit discounting by competing unknown support, not the Dempster conflict coefficient or Dempster's combination rule.

For a hard open-set decision, a sample is classified as known if $\mathrm{Score}(x) > \eta$ and rejected otherwise:

$$\hat{y} = \begin{cases} \arg\max_j s^C_j(x), & \mathrm{Score}(x) > \eta, \\ \text{unknown}, & \text{otherwise}. \end{cases} \tag{18}$$

For the reported AUROC, no operating threshold is selected: AUROC is obtained by sweeping the continuous CEI score. The threshold η is required only when a hard known/unknown decision is desired.

Table 1: Open-set performance under the target 16-shot protocol measured by AUROC (%). Training-free baselines use no target shots; "–" denotes unreported results.

Method	C10	C10+10	C10+50	C100	T200	T200-H	M100	IN200
CLIP (Radford et al., 2021)	91.12	81.33	90.20	82.40	81.61	80.07	84.73	90.73
CLIP-PE (Radford et al., 2021)	92.36	84.95	91.27	84.27	84.74	81.87	85.00	88.91
LMC (training-free) (Qu et al., 2023)	**96.60**	**98.90**	**98.50**	–	86.70	–	–	–
CoOp (Zhou et al., 2022b)	92.49	87.66	91.75	87.82	88.03	84.28	80.04	88.92
CoCoOp (Zhou et al., 2022a)	92.88	84.61	90.87	86.32	84.15	84.01	80.01	89.01
ID-like (Bai et al., 2024)	90.56	91.51	90.73	71.58	47.88	49.85	53.95	92.91
NegPrompt (Li et al., 2024)	83.43	88.13	84.74	69.51	83.21	77.53	–	–
M-Tuning (Liao et al., 2025)	92.82	97.83	96.04	81.14	86.61	82.96	88.29	97.27
UTL (Ma et al., 2025)	94.18	89.71	93.81	86.82	89.06	85.30	95.78	96.31
SAEC	96.44	93.29	94.45	**87.85**	**90.03**	**85.62**	**96.27**	**97.79**

4 EXPERIMENTS

4.1 SETUP

We evaluate SAEC on eight standard FSOR benchmark settings following established FSOR/OSR evaluation protocols (Boudiaf et al., 2023; Ma et al., 2025; Yang et al., 2022; Lu et al., 2022): CIFAR10 (C10), CIFAR10+10 (C10+10), CIFAR10+50 (C10+50), CIFAR100 (C100) (Krizhevsky & Hinton, 2009), TinyImageNet (T200), TinyImageNet-Hard (T200-H) (Le et al., 2015), MiniImageNet (M100) (Vinyals et al., 2016), and ImageNet-200 (IN200) (Yang et al., 2022; Deng et al., 2009). C10 uses a 6/4 known/unknown split; C10+10 and C10+50 use four CIFAR10 known classes with 10 or 50 CIFAR100 unknown classes; C100 uses 20/80; T200 and T200-H use 20/180 and 40/160; M100 uses 64/36; and IN200 uses 200/800 known/unknown classes. We report closed-set accuracy (ACC) and open-set AUROC.

All experiments use CLIP ViT-B/16 (Radford et al., 2021) as the pretrained backbone and keep both CLIP encoders frozen. The number of context tokens is $L = 16$, the number of dictionary atoms is $K = 100$, and the number of synthesized unknown prompts is $N_u = 50$. The dictionary sparsity coefficient is $\alpha = 1$, and the Gaussian perturbation scale for stochastic anchor synthesis is $\sigma = 0.1$. The composition matrix W is initialized from a Gaussian distribution with mean 0 and standard deviation 0.02. We set $\lambda = 0.1$, $\varepsilon = 0.1$, $\gamma = 0.5$, $m = \delta = 0.2$, and $C_0 = 10$. Training uses SGD with batch size 32, initial learning rate 0.002, and cosine decay for up to 50 epochs on one NVIDIA RTX 3090 GPU. We follow UTL (Ma et al., 2025) for the temperature-scaled similarity normalization used in unknown label smoothing.

The auxiliary vocabulary follows the UTL protocol (Ma et al., 2025) and uses 1,000 label names drawn from ImageNet-21K (Ridnik et al., 2021); labels associated with the target unknown classes are filtered using WordNet (Miller, 1995) synset IDs. This synset-level filtering guarantees label-level disjointness, but it does not imply complete semantic independence because hypernyms, hyponyms, neighboring concepts, or compound-token overlaps may remain. We therefore interpret the auxiliary vocabulary as a controlled external semantic source rather than as a proof of strict semantic independence.

4.2 MAIN COMPARISON UNDER THE FEW-SHOT SETTING

Table 1 summarizes open-set ranking performance. Compared with the directly comparable UTL baseline, SAEC improves AUROC on all eight settings—C10, C10+10, C10+50, C100, T200, T200-H, M100, and IN200—by $+2.26$, $+3.58$, $+0.64$, $+1.03$, $+0.97$, $+0.32$, $+0.49$, and $+1.48$ points, respectively. Among all reported methods, SAEC achieves the best AUROC on C100, T200, T200-H, M100, and IN200.

The compared methods span training-free CLIP variants (Radford et al., 2021), the training-free LMC method (Qu et al., 2023), closed-set prompt learning methods (Zhou et al., 2022b;a), few-shot OOD methods (Bai et al., 2024; Li et al., 2024), and CLIP-based open-set methods evaluated under the few-shot protocol (Liao et al., 2025; Ma et al., 2025). Table 2 reports closed-set recognition: SAEC obtains the best ACC on six of the eight settings and remains competitive on the remaining

Table 2: Closed-set performance under the 16-shot setting measured by ACC (%).

Method	C10	C10+10	C10+50	C100	T200	T200-H	M100	IN200
CLIP (Radford et al., 2021)	90.72	91.36	91.72	82.98	83.72	83.01	85.97	75.31
CLIP-PE (Radford et al., 2021)	91.87	92.75	92.75	84.47	84.47	83.04	86.37	76.11
CoOp (Zhou et al., 2022b)	94.48	96.30	96.30	88.39	88.77	85.42	92.97	76.77
CoCoOp (Zhou et al., 2022a)	93.10	95.78	95.78	88.05	87.60	85.42	90.68	77.31
M-Tuning (Liao et al., 2025)	95.50	**97.27**	97.27	85.66	**90.69**	85.57	94.33	77.33
UTL (Ma et al., 2025)	94.18	96.58	95.62	87.85	88.18	85.22	93.21	80.01
SAEC	**97.92**	96.88	**98.44**	**88.42**	90.31	**85.77**	**95.51**	**82.34**

Table 3: Contextual comparison with representative conventional and CLIP-based OSR methods. Rows marked "all data" use all available known-class training images and are not directly comparable to the 16-shot setting.

Method	C10		C10+10		C10+50		T200		IN200	
	AUROC	ACC	AUROC	ACC	AUROC	ACC	AUROC	ACC	AUROC	ACC
CPN (all data) (Yang et al., 2022)	82.80	92.90	88.11	94.80	87.90	–	63.90	81.40	79.56	82.20
RPL (all data) (Chen et al., 2020)	86.10	95.10	85.60	95.50	85.00	–	70.20	81.70	91.70	66.20
ARPL (all data) (Chen et al., 2022)	91.00	94.50	97.10	94.70	95.10	–	78.20	76.10	94.90	82.30
PMAL (all data) (Lu et al., 2022)	95.10	97.50	97.80	97.80	96.90	–	83.10	84.70	93.90	84.10
All You Need (all data) (Vaze et al., 2022)	93.60	94.10	97.90	95.92	96.50	–	83.00	71.20	95.71	71.30
ARPL+CS (all data) (Vaze et al., 2022)	93.90	94.25	**98.10**	97.07	96.70	–	82.50	76.84	96.16	79.02
M-Tuning (all data) (Liao et al., 2025)	96.29	96.30	96.28	96.23	96.17	–	87.30	90.70	96.47	81.90
UTL (all data) (Ma et al., 2025)	96.33	97.93	94.15	97.88	95.47	98.21	90.31	92.53	96.73	81.30
SAEC (16-shot)	96.44	97.92	93.29	96.88	94.45	98.44	90.03	90.31	97.79	82.34
SAEC (all data)	**96.92**	**98.14**	97.96	**98.32**	**97.53**	**99.37**	**91.07**	**94.11**	**98.69**	**84.56**

Table 4: Progressive component ablation. The table evaluates the cumulative effect of replacing the PCA basis with dictionary learning, adding USAR, and adding CEI.

Method	C10+10		T200		M100		IN200	
	AUROC	ACC	AUROC	ACC	AUROC	ACC	AUROC	ACC
UTL baseline	89.71	96.58	89.06	88.18	95.78	93.21	96.31	80.01
+ dictionary learning	90.76	96.77	89.13	89.87	96.01	**96.01**	96.73	81.23
+ USAR	91.15	96.78	89.91	89.03	96.17	94.79	97.33	81.97
+ CEI (full SAEC)	**93.29**	**96.88**	**90.03**	**90.31**	**96.27**	95.51	**97.79**	**82.34**

two. Taken together, the two tables show improved known/unknown discrimination while preserving strong known-class recognition.

For broader context, Table 3 additionally compares SAEC with representative conventional and CLIP-based OSR methods under their reported training regimes. Because rows marked "all data" use substantially more known-class supervision than the 16-shot setting, this table is intended as a contextual comparison rather than a direct few-shot ranking. Table 3 shows that the 16-shot SAEC model remains competitive with several all-data OSR systems, while the all-data SAEC variant further improves performance when full known-class supervision is available.

4.3 Progressive component ablation

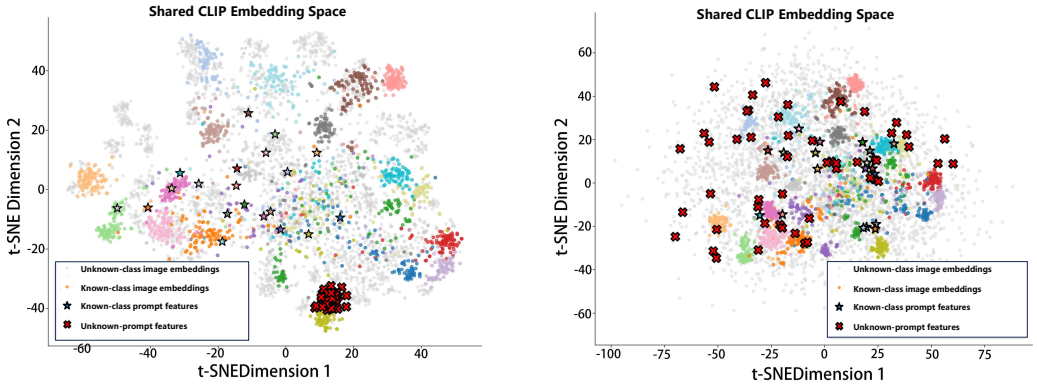
Table 4 shows a consistent cumulative improvement in AUROC across the four evaluated benchmarks. Replacing the PCA-based unknown basis with sparse dictionary learning improves AUROC in all four cases, supporting the use of a compositional dictionary for unknown-anchor construction in the evaluated configuration. Adding USAR provides further AUROC gains on all four benchmarks, consistent with the intended role of explicit known–unknown separation and unknown-anchor dispersion. Incorporating CEI further improves AUROC on all four benchmarks and yields the strongest final configuration. Because CEI operates only at inference, the transition from '+ USAR' to '+ CEI' changes the open-set scoring rule while retaining the learned representation, providing a direct test of the proposed evidential score over the preceding response-based decision. The

Table 5: Shot sensitivity on T200 and T200-H.

Dataset	T200				T200-H			
Shot	2	4	8	16	2	4	8	16
AUROC	88.51	88.82	89.49	**90.03**	84.75	85.68	**85.74**	85.62
ACC	85.63	86.74	87.83	**90.31**	85.02	85.42	**86.02**	85.77

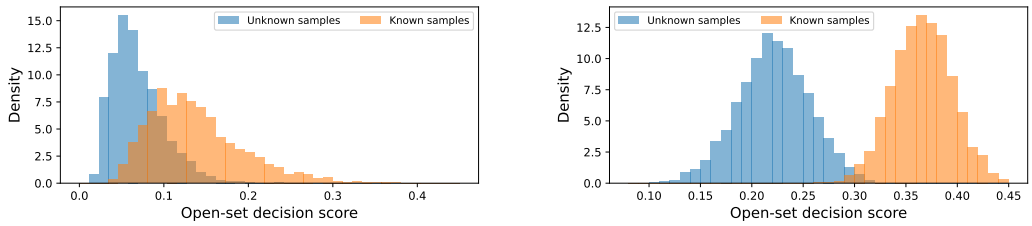
Table 6: Effect of ε on C10+50. Each cell reports AUROC/ACC.

Method	0.001	0.01	0.1	0.3	0.5
UTL (Ma et al., 2025)	91.85/95.07	92.76/95.95	93.84/96.50	92.27/94.77	92.48/94.80
SAEC	**93.21/95.53**	**93.08/97.47**	**94.45/98.44**	**93.68/96.60**	**92.98/95.70**



(a) UTL feature space (b) SAEC feature space

Figure 3: Feature-space visualization on C100. The two panels compare the distributions produced by UTL and SAEC.



(a) UTL decision scores (b) SAEC decision scores

Figure 4: Open-set decision-score distributions on C100. The two panels compare the score distributions produced by UTL and SAEC.

table is otherwise progressive rather than factorial, so it supports the cumulative design path but does not isolate every interaction among SDSA and USAR.

4.4 Sensitivity, Efficiency, and Qualitative Evidence

Shot sensitivity is reported directly in Table 5. On T200, both AUROC and ACC generally improve as the number of shots increases, while T200-H is comparatively stable from 4 to 16 shots, indicating that the method does not rely on a single-shot configuration.

The effect of the unknown-label smoothing weight is shown in Table 6. SAEC performs best around $\varepsilon = 0.1$ on C10+50, while both smaller and larger values reduce the balance between open-set AUROC and closed-set ACC.

Figures 3 and 4 provide complementary qualitative views of the behavior of UTL and SAEC on C100. Figure 3 visualizes the representation structure, while Figure 4 shows the corresponding open-set decision-score distributions. These visualizations provide qualitative support, whereas the quantitative conclusions are based on the reported AUROC and ACC results.

Tables 5 and 6 provide the most directly relevant robustness analyses in the main text, while the remaining sensitivity studies for N_u, γ, C_0, and (m, δ) are reported in Appendix B. Overall, performance is stable around the selected settings, whereas several more extreme choices reduce either AUROC, ACC, or both. In terms of efficiency, SAEC keeps both CLIP encoders frozen. With ViT-B/16, $d = 512$, $L = 16$, $N_u = 50$, and $K = 100$, the target-task trainable parameters in the shared context V and anchor-composition matrix W total $Ld+N_uK = 13{,}192$. CEI itself uses only scalar operations on existing similarity scores, so it does not introduce an additional neural inference branch.

5 Conclusion

This paper presents SAEC, a CLIP-based FSOR framework that links unknown semantic construction, embedding-space regularization, and evidential open-set inference. SDSA constructs unknown prompts through sparse dictionary compositions, USAR explicitly regularizes known–unknown separation and unknown-anchor dispersion, and CEI scores samples by combining the belief and plausibility bounds induced by competing known, unknown, and residual-ignorance masses. Across eight standard FSOR settings, SAEC improves AUROC over the directly comparable UTL baseline on all eight while retaining strong closed-set recognition. The current evidence supports this coupled design under the evaluated CLIP ViT-B/16 protocol; cross-backbone generalization and formal uncertainty calibration remain open directions.

Reproducibility Statement

We provide the method definition in Section 3, including the dictionary objective, anchor synthesis process, training loss, and evidential inference rule. Section 4 reports the benchmark splits, optimization settings, and all primary hyperparameters. Appendix A gives pseudocode for dictionary construction, SAEC training, and CEI inference. Anonymous code and configuration files can be submitted as supplementary material during review.

AI Use Statement

Generative AI tools were used to assist with language editing, LaTeX formatting, structural reorganization, and revision planning. They were not used to generate experimental results, fabricate data, or replace the authors' technical verification. The authors reviewed and take responsibility for all AI-assisted text, mathematical notation, claims, and artifacts included in this submission.

REFERENCES

Yichen Bai, Zongbo Han, Bing Cao, Xiaoheng Jiang, Qinghua Hu, and Changqing Zhang. Id-like prompt learning for few-shot out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17480–17489, June 2024.

Malik Boudiaf, Etienne Bennequin, Myriam Tami, Antoine Toubhans, Pablo Piantanida, Celine Hudelot, and Ismail Ben Ayed. Open-set likelihood maximization for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24007–24016, June 2023.

Guangyao Chen, Limeng Qiao, Yemin Shi, Peixi Peng, Jia Li, Tiejun Huang, Shiliang Pu, and Yonghong Tian. Learning open set network with discriminative reciprocal points. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (eds.), *Computer Vision – ECCV 2020*, pp. 507–522, Cham, 2020. Springer International Publishing. ISBN 978-3-030-58580-8.

Guangyao Chen, Peixi Peng, Xiangqian Wang, and Yonghong Tian. Adversarial reciprocal points learning for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8065–8081, 2022. doi: 10.1109/TPAMI.2021.3106743.

Arthur P. Dempster. *Upper and Lower Probabilities Induced by a Multivalued Mapping*, pp. 57–72. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-44792-4. doi: 10.1007/978-3-540-44792-4_3. URL https://doi.org/10.1007/978-3-540-44792-4_3.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

Shule Deng, Jin-Gang Yu, Zihao Wu, Hongxia Gao, Yansheng Li, and Yang Yang. Learning relative feature displacement for few-shot open-set recognition. *IEEE Transactions on Multimedia*, 25: 5763–5774, 2023. doi: 10.1109/TMM.2022.3198880.

Shiyuan Huang, Jiawei Ma, Guangxing Han, and Shih-Fu Chang. Task-adaptive negative envision for few-shot open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7171–7180, June 2022.

Minki Jeong, Seokeon Choi, and Changick Kim. Few-shot open-set recognition by transformation consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12566–12575, June 2021.

Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical Report 0, University of Toronto, Toronto, Ontario, 2009. URL https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Yann Le, Xuan Yang, et al. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015.

Tianqi Li, Guansong Pang, Xiao Bai, Wenjun Miao, and Jin Zheng. Learning transferable negative prompts for out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17584–17594, June 2024.

Ning Liao, Xiaopeng Zhang, Min Cao, and Junchi Yan. M-tuning: Prompt tuning with mitigated label bias in open-set scenarios. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(6):5885–5899, 2025. doi: 10.1109/TCSVT.2024.3524181.

Jing Lu, Yunlu Xu, Hao Li, Zhanzhan Cheng, and Yi Niu. Pmal: Open set recognition via robust prototype mining. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 1872–1880, 2022.

Rui Ma, Qilong Wang, Bing Cao, Qinghua Hu, and Yahong Han. Unknown text learning for clip-based few-shot open-set recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 657–667, October 2025.

George A. Miller. Wordnet: a lexical database for english. *Commun. ACM*, 38(11):39–41, November 1995. ISSN 0001-0782. doi: 10.1145/219717.219748. URL https://doi.org/10.1145/219717.219748.

Sayak Nag, Dripta S. Raychaudhuri, Sujoy Paul, and Amit K. Roy-Chowdhury. Reconstruction guided meta-learning for few shot open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12):15394–15405, 2023. doi: 10.1109/TPAMI.2023.3320731.

Haoxuan Qu, Xiaofei Hui, Yujun Cai, and Jun Liu. Lmc: Large model collaboration with cross-assessment for training-free open-set object recognition. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 46491–46504. Curran Associates, Inc., 2023. doi: 10.52202/075280-2016. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/91813e5ddd9658b99be4c532e274b49c-Paper-Conference.pdf.

Jingen Qu, Yufei Chen, Xiaodong Yue, Wei Fu, and Qiguang Huang. Hyper-opinion evidential deep learning for out-of-distribution detection. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 84645–84668. Curran Associates, Inc., 2024. doi: 10.52202/079017-2687. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/99d4ceebdf75b64e8ed608a245b63416-Paper-Conference.pdf.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Tal Ridnik, Emanuel Ben-Baruch, Asaf Noy, and Lihi Zelnik. Imagenet-21k pretraining for the masses. In J. Vanschoren and S. Yeung (eds.), *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021. URL https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/98f13708210194c475687be6106a3b84-Paper-round1.pdf.

Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/a981f2b708044d6fb4a71a1463242520-Paper.pdf.

Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976. ISBN 9780691214696. doi: 10.1515/9780691214696. URL https://doi.org/10.1515/9780691214696.

Zheng Tong, Philippe Xu, and Thierry Denœux. An evidential classifier based on dempster-shafer theory and deep learning. *Neurocomputing*, 450:275–293, 2021. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2021.03.066. URL https://www.sciencedirect.com/science/article/pii/S0925231221004525.

Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=5hLP5JY9S2d.

Oriol Vinyals, Charles Blundell, Timothy Lillicrap, koray kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/90e1357833654983612fb05e3ec9148c-Paper.pdf.

Haoyu Wang, Guansong Pang, Peng Wang, Lei Zhang, Wei Wei, and Yanning Zhang. Glocal energy-based learning for few-shot open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7507–7516, June 2023.

Hong-Ming Yang, Xu-Yao Zhang, Fei Yin, Qing Yang, and Cheng-Lin Liu. Convolutional prototype network for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2358–2370, 2022. doi: 10.1109/TPAMI.2020.3045079.

Zhenyu Zhang, Guangyao Chen, Yixiong Zou, Yuhua Li, and Ruixuan Li. Learning unknowns from unknowns: Diversified negative prototypes generator for few-shot open-set recognition. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, pp. 6053–6062, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400706868. doi: 10.1145/3664647.3680646. URL https://doi.org/10.1145/3664647.3680646.

Chen Zhao, Dawei Du, Anthony Hoogs, and Christopher Funk. Open set action recognition via multi-label evidential learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22982–22991, June 2023.

Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Conditional prompt learning for vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16816–16825, June 2022a.

Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9):2337–2348, 2022b. ISSN 1573-1405. doi: 10.1007/s11263-022-01653-1. URL https://doi.org/10.1007/s11263-022-01653-1.

A PSEUDOCODE

Algorithm 1 Sparse dictionary construction

Require: Auxiliary label vocabulary $\mathcal{Y}^a$, CLIP tokenizer and token embedding layer, number of atoms K, sparsity coefficient α.
Ensure: Normalized dictionary $\tilde{D}$.

Initialize an empty token matrix Z.
for each auxiliary label $y_i^a \in \mathcal{Y}^a$ **do**
    Tokenize y_i^a with the CLIP tokenizer.
    Obtain token embeddings and remove SOS/EOS embeddings.
    Append all valid token embeddings to Z.
end for
Solve $\min_{D,A} \frac{1}{2}\|Z - AD\|_F^2 + \alpha \sum_i \|\mathbf{a}_i\|_1$ subject to $\|d_k\|_2 \le 1$.
Normalize each atom: $\tilde{d}_k = d_k / \|d_k\|_2$.
return $\tilde{D} = [\tilde{d}_1, \ldots, \tilde{d}_K]^\top$.

Algorithm 2 SAEC training

Require: Frozen CLIP encoders $\mathcal{I}, \mathcal{T}$, known training set $\mathcal{D}_C$, known labels $\mathcal{Y}^C$, dictionary $\tilde{D}$, hyperparameters $N_u, \sigma, \lambda, \varepsilon, \gamma, m, \delta$.
Ensure: Learned context tokens V and unknown-anchor composition weights W.

Initialize shared context tokens V and composition weights $W = \{\mathbf{w}_k\}_{k=1}^{N_u}$.
for each training iteration **do**
    Sample a mini-batch (x, y) from $\mathcal{D}_C$.
    for $k = 1$ to N_u **do**
        Sample $\boldsymbol{\xi}_k \sim \mathcal{N}(0, \sigma^2 I)$.
        Compute $U_k = ((\mathbf{w}_k + \boldsymbol{\xi}_k)\tilde{D}) / \|(\mathbf{w}_k + \boldsymbol{\xi}_k)\tilde{D}\|_2$.
        Build unknown prompt $P_k^U = [\texttt{SOS}, V, U_k, \texttt{EOS}]$.
    end for
    Build known prompts $P_j^C = [\texttt{SOS}, V, \mathbf{v}(y_j), \texttt{EOS}]$ for all known classes.
    Encode image and text features with frozen CLIP encoders.
    Compute $\mathcal{L}_{\text{ULS}}$, $\mathcal{L}_{\text{sep}}$, and $\mathcal{L}_{\text{dis}}$.
    Update V and W by minimizing $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{ULS}} + \gamma(\mathcal{L}_{\text{sep}} + \mathcal{L}_{\text{dis}})$.
end for
return V, W.

Algorithm 3 Conflict-aware evidential inference

Require: Test image x, known prompts $\{P_j^C\}_{j=1}^{N_C}$, unknown prompts $\{P_k^U\}_{k=1}^{N_u}$, prior mass C_0, threshold η.
Ensure: Prediction $\hat{y} \in \mathcal{Y}^C \cup \{\text{unknown}\}$.

Compute known similarities $\{s_j^C(x)\}_{j=1}^{N_C}$ and unknown similarities $\{s_k^U(x)\}_{k=1}^{N_u}$.
Convert similarities into evidence: $e_j^{\text{ID}} = \text{Softplus}(s_j^C(x))$, $e_k^{\text{OOD}} = \text{Softplus}(s_k^U(x))$.
Retain branch-level evidence: $e^{\text{ID}} = \max_j e_j^{\text{ID}}$, $e^{\text{OOD}} = \max_k e_k^{\text{OOD}}$.
Set $S = e^{\text{ID}} + e^{\text{OOD}} + C_0$.
Compute masses $m_{\text{ID}} = e^{\text{ID}}/S$, $m_{\text{OOD}} = e^{\text{OOD}}/S$, $m_\Theta = C_0/S$.
Compute $\text{Score} = m_{\text{ID}}(1 - m_{\text{OOD}})$.
if $\text{Score} > \eta$ **then**
    return $\hat{y} = \arg\max_j s_j^C(x)$.
else
    return $\hat{y} = \text{unknown}$.
end if

B ADDITIONAL EXPERIMENTAL RESULTS

The primary qualitative visualizations, the contextual OSR comparison, shot sensitivity, and ε sensitivity are included in the main paper. This appendix reports the remaining hyperparameter analyses for N_u, γ, C_0, and (m, δ).

Table 7: Effect of the number of synthesized unknown prompts N_u on C100.

N_u	1	5	25	50	100
AUROC	87.84	87.89	87.79	**87.85**	87.61
ACC	87.72	87.90	87.67	**88.42**	88.10

Table 8: Effect of γ on IN200.

γ	0.1	0.3	0.5	0.7	1.0
AUROC	95.17	97.05	**97.79**	95.25	94.29
ACC	81.08	81.05	**82.34**	80.98	80.25

Table 9: Effect of the ignorance prior C_0 on C10+50.

C_0	1	5	10	20
AUROC	92.87	93.38	**94.45**	93.03
ACC	96.65	97.65	**98.44**	97.67

Table 10: Effect of jointly varying m and δ with $m = \delta$ on C10+50.

$m = \delta$	0.1	0.2	0.5	0.8
AUROC	93.27	**94.45**	92.93	93.69
ACC	97.83	**98.44**	97.47	96.33