模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

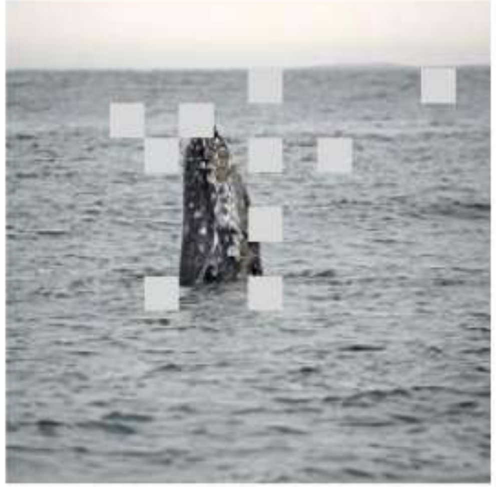
ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

LoCoOp: Few-Shot Out-of-Distribution Detection via Prompt Learning

Atsuyuki Miyai[1] **Qing Yu**[1,2] **Go Irie**[3] **Kiyoharu Aizawa**[1]
[1]The University of Tokyo [2]LY Corporation [3]Tokyo University of Science
{miyai,yu,aizawa}@hal.t.u-tokyo.ac.jp goirie@ieee.org

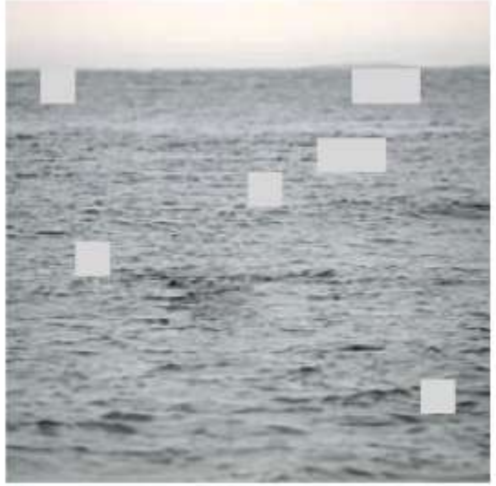
NeurIPS 2023

Motivation

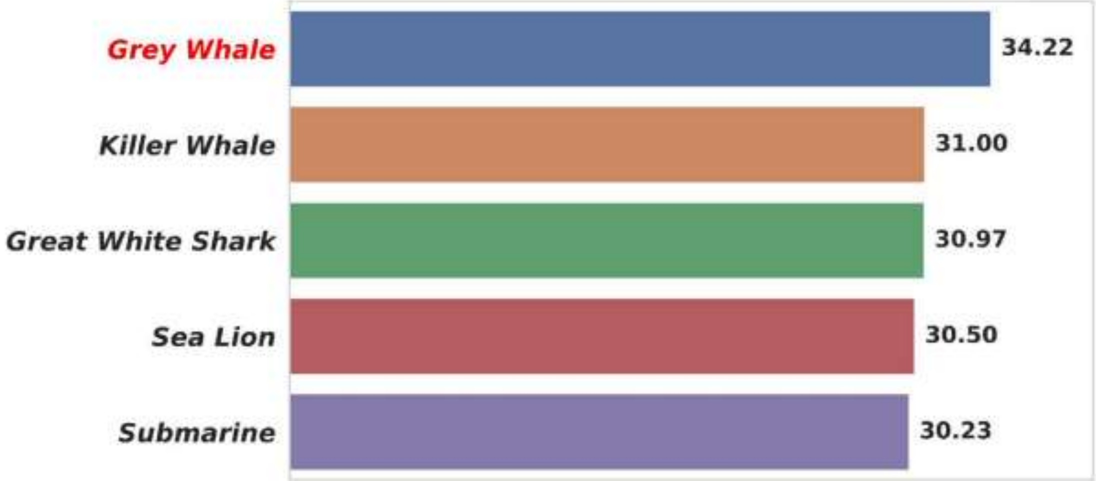


ID class: grey whale

Top 5 logits on ImageNet-1K classes

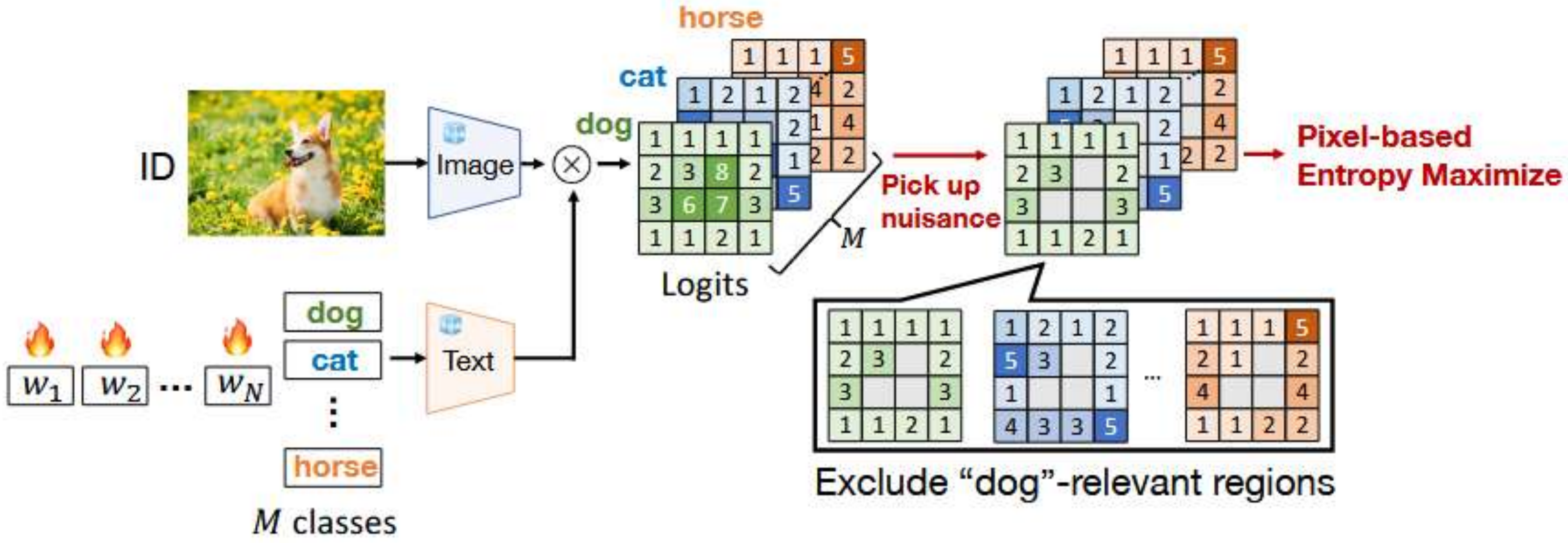


Background interference



Top 5 logits on ImageNet-1K classes

LoCoOp



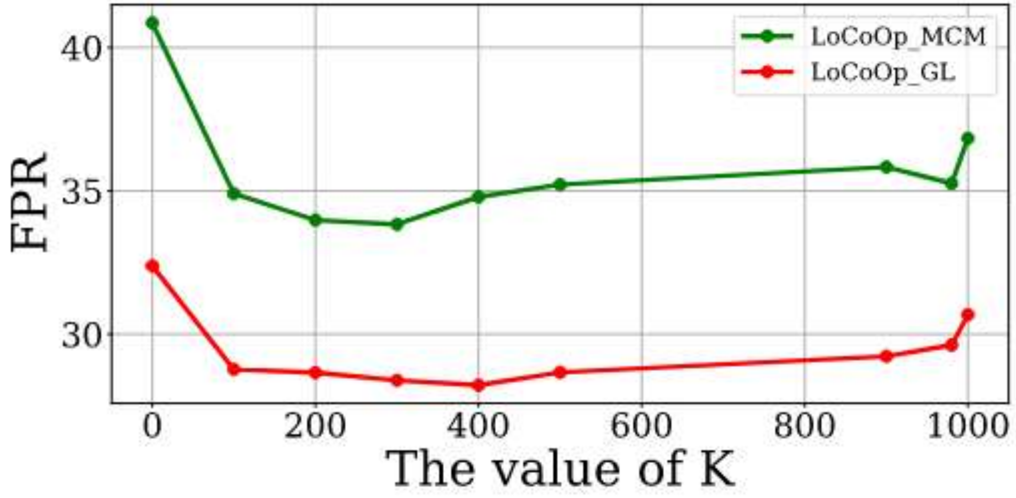
- 计算每个区域的特征 $\boldsymbol{f}_i{}^{\text{in}} = \text{Proj}_{v \to t}(v\left(\boldsymbol{f}'_i{}^{\text{in}}\right))$
- 计算每个区域的类别概率 $p_i(y=m \mid \boldsymbol{x}^{\text{in}}) = \frac{\exp\left(\text{sim}\left(\boldsymbol{f}_i{}^{\text{in}}, \boldsymbol{g}_m\right)/\tau\right)}{\sum_{m'=1}^{M}\exp\left(\text{sim}\left(\boldsymbol{f}_i{}^{\text{in}}, \boldsymbol{g}_{m'}\right)/\tau\right)}.$
- 获得top-K的ID无关区域 $J = \{i \in I : \text{rank}(p_i(y = y^{\text{in}} | \boldsymbol{x}^{\text{in}})) > K\}$
- ID无关区域的熵最大化 $\mathcal{L}_{\text{ood}} = -H(p_j),\quad \mathcal{L} = \mathcal{L}_{\text{coop}} + \lambda \mathcal{L}_{\text{ood}}$
- OOD分数 $S_{\text{MCM}} = \max_m \frac{\exp(\text{sim}(\boldsymbol{f}, \boldsymbol{g}_m)/\tau)}{\sum_{m'=1}^{M}\exp(\text{sim}(\boldsymbol{f}, \boldsymbol{g}_{m'})/\tau)},$

$$S_{\text{GL-MCM}} = S_{\text{MCM}} + \max_{m,i} \frac{\exp(\text{sim}(\boldsymbol{f}_i, \boldsymbol{g}_m)/\tau)}{\sum_{m'=1}^{M}\exp(\text{sim}(\boldsymbol{f}_i, \boldsymbol{g}_{m'})/\tau)}.$$

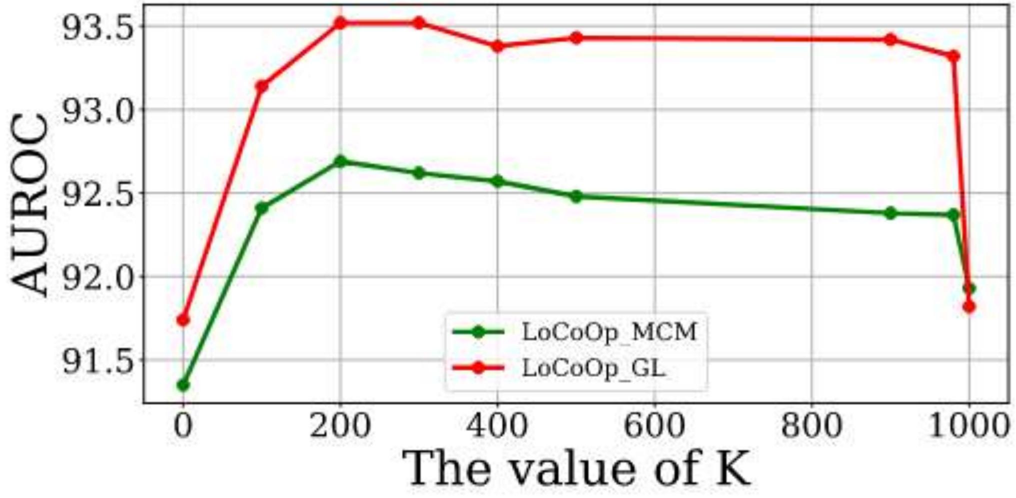
Experiment

Method	iNaturalist FPR95↓	iNaturalist AUROC↑	SUN FPR95↓	SUN AUROC↑	Places FPR95↓	Places AUROC↑	Texture FPR95↓	Texture AUROC↑	**Average** FPR95↓	**Average** AUROC↑
Zero-shot										
MCM [30]*	30.94	94.61	37.67	92.56	44.76	89.76	57.91	86.10	42.82	90.76
GL-MCM [33]*	15.18	96.71	30.42	93.09	38.85	89.90	57.93	83.63	35.47	90.83
Fine-tuned										
ODIN [28]†	30.22	94.65	54.04	87.17	55.06	85.54	51.67	87.85	47.75	88.80
ViM [50]†	32.19	93.16	54.01	87.19	60.67	83.75	53.94	87.18	50.20	87.82
KNN [45]†	29.17	94.52	35.62	92.67	39.61	91.02	64.35	85.67	42.19	90.97
NPOS$_{MCM}$ [46]†	16.58	96.19	43.77	90.44	45.27	89.44	46.12	88.80	37.93	91.22
NPOS$_{MCM}$ [46]*	19.59	95.68	48.26	89.70	49.82	88.77	51.12	87.58	42.20	90.43
NPOS$_{GL}$*	18.70	95.36	38.99	90.33	41.86	89.36	47.89	86.44	36.86	90.37
Prompt learning				*one-shot (one label per class)*						
CoOp$_{MCM}$	43.38	91.26	38.53	91.95	46.68	89.09	50.64	87.83	44.81	90.03
CoOp$_{GL}$	21.30	95.27	31.66	92.16	40.44	89.31	52.93	84.25	36.58	90.25
LoCoOp$_{MCM}$ (ours)	38.49	92.49	33.27	93.67	39.23	91.07	49.25	89.13	40.17	91.53
LoCoOp$_{GL}$ (ours)	24.61	94.89	25.62	94.59	34.00	**92.12**	49.86	87.49	33.52	92.14
				16-shot (16 labels per class)						
CoOp$_{MCM}$	28.00	94.43	36.95	92.29	43.03	89.74	**39.33**	91.24	36.83	91.93
CoOp$_{GL}$	**14.60**	96.62	28.48	92.65	36.49	89.98	43.13	88.03	30.67	91.82
LoCoOp$_{MCM}$ (ours)	23.06	95.45	32.70	93.35	39.92	90.64	40.23	**91.32**	33.98	92.69
LoCoOp$_{GL}$ (ours)	16.05	**96.86**	**23.44**	**95.07**	**32.87**	91.98	42.28	90.19	**28.66**	**93.52**

Experiment

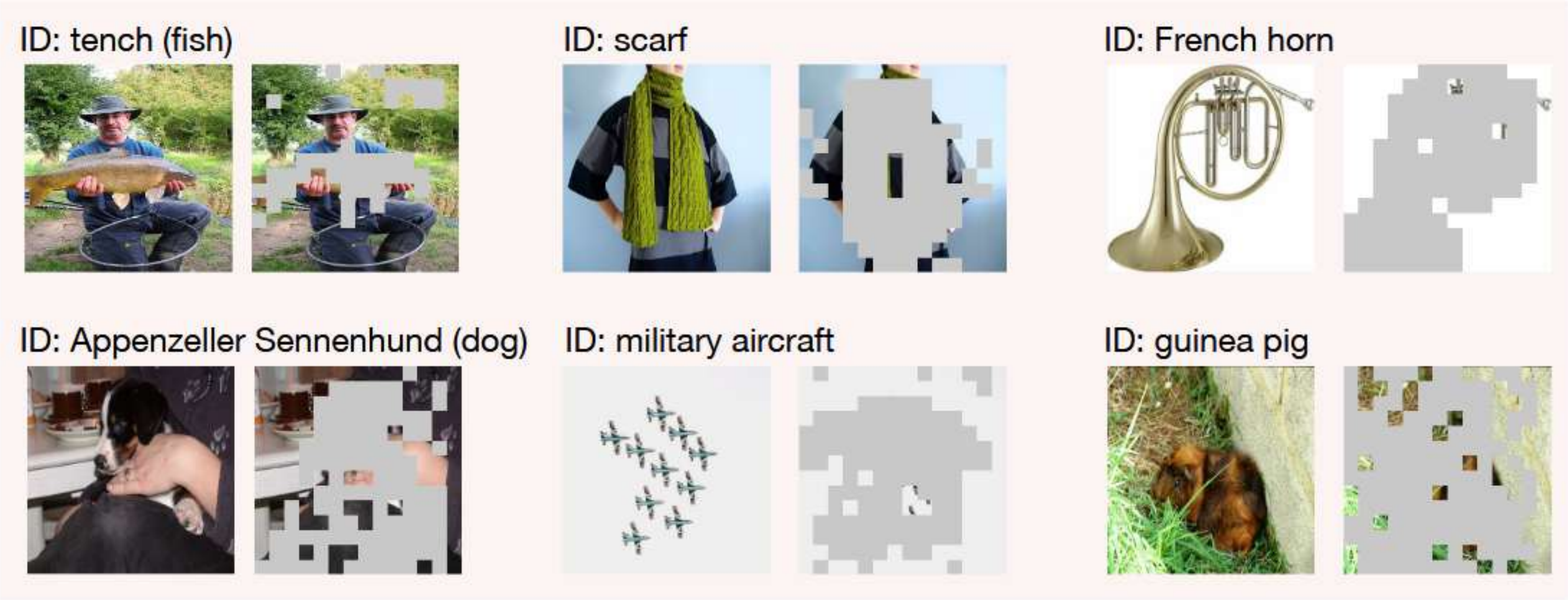


(a) FPR95



(b) AUROC

Experiment


ID: tench (fish)
ID: scarf
ID: French horn
ID: Appenzeller Sennenhund (dog)
ID: military aircraft
ID: guinea pig

模式分析与机器智能 工业和信息化部重点实验室
MIIT Key Laboratory of Pattern Analysis & Machine Intelligence

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

Automated Essential Concept Discovery for Few-Shot Out-of-Distribution Detection

Guangyao Chen[1,2], Kai Horstmann[2], Zhilong Wang[1,2], Fengqi You[1,2,3]*

[1] AI for Science Institute (CUAISci), Cornell University, Ithaca, NY 14853, USA

[2] College of Engineering, Cornell University, Ithaca, NY 14853, USA

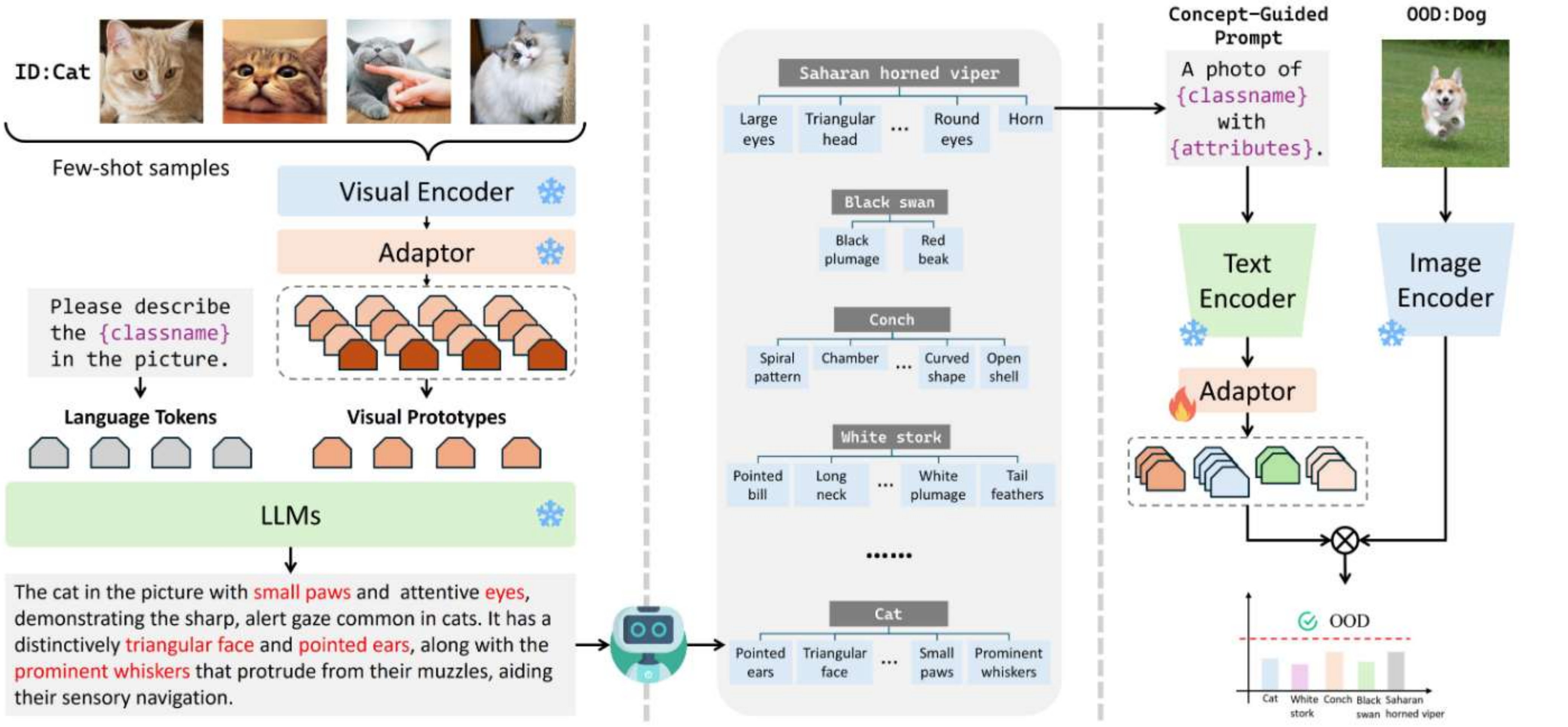
[3] Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14853, USA

{gy.chen, kah288, zhilongwang, fengqi.you}@cornell.edu

CVPR 2025

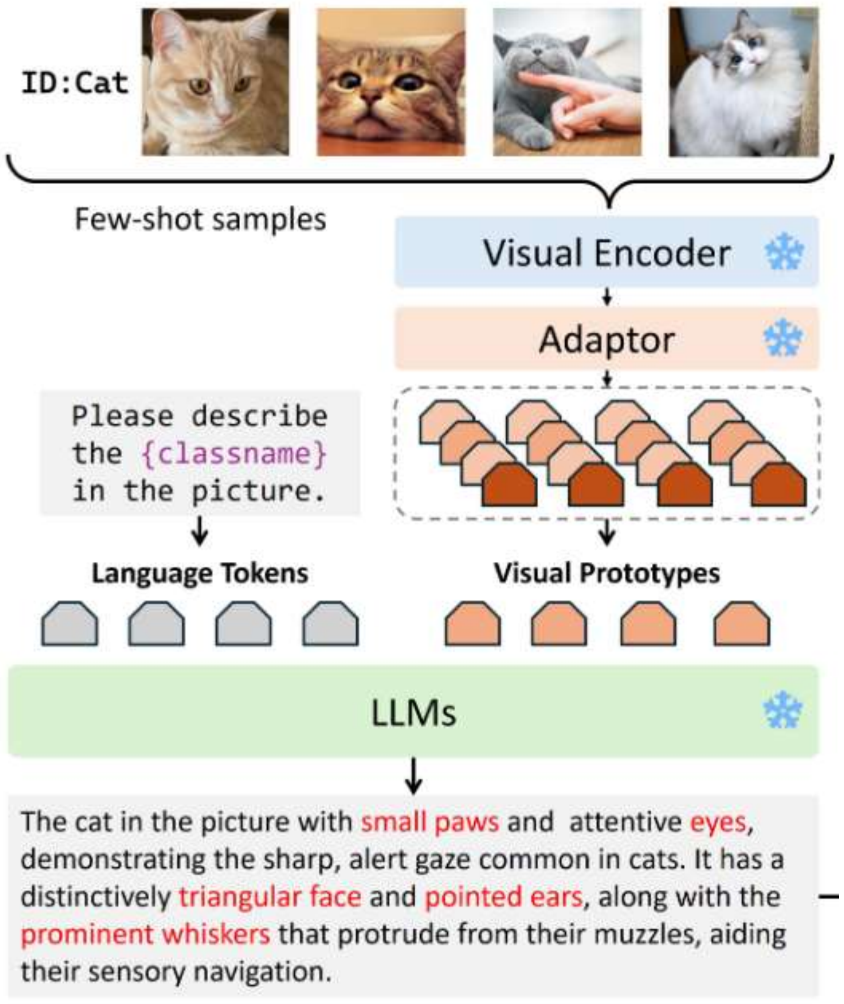
Automated Essential Concept Discovery

Automated Essential Concept Discovery



(a) Automated Essential Concept Discovery (b) Essential Concept Memory (c) Essential Concept Matching

Essential Concept Extraction

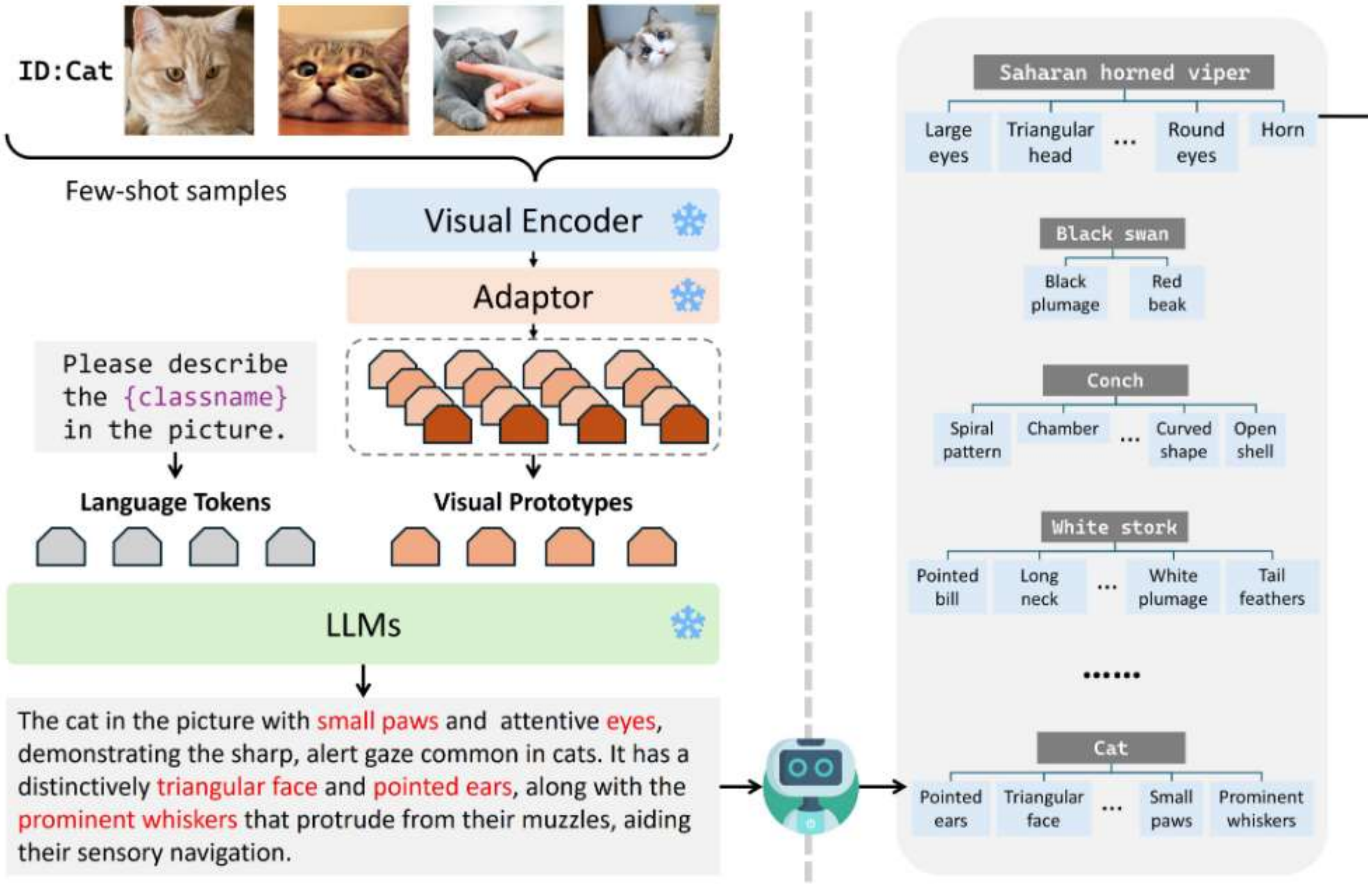


(a) Automated Essential Concept Discovery

- 构建K-shot数据集 $\mathcal{S}^c = \{(\boldsymbol{x}_i^c, y_i^c)\}_{i=1}^K$
- 计算类别特征原型 $\boldsymbol{P}_c = \frac{1}{K}\sum_{i=1}^{K}\mathcal{A}(\mathcal{V}(\boldsymbol{x}_i^c))$.
- 多模态模型输出类别描述

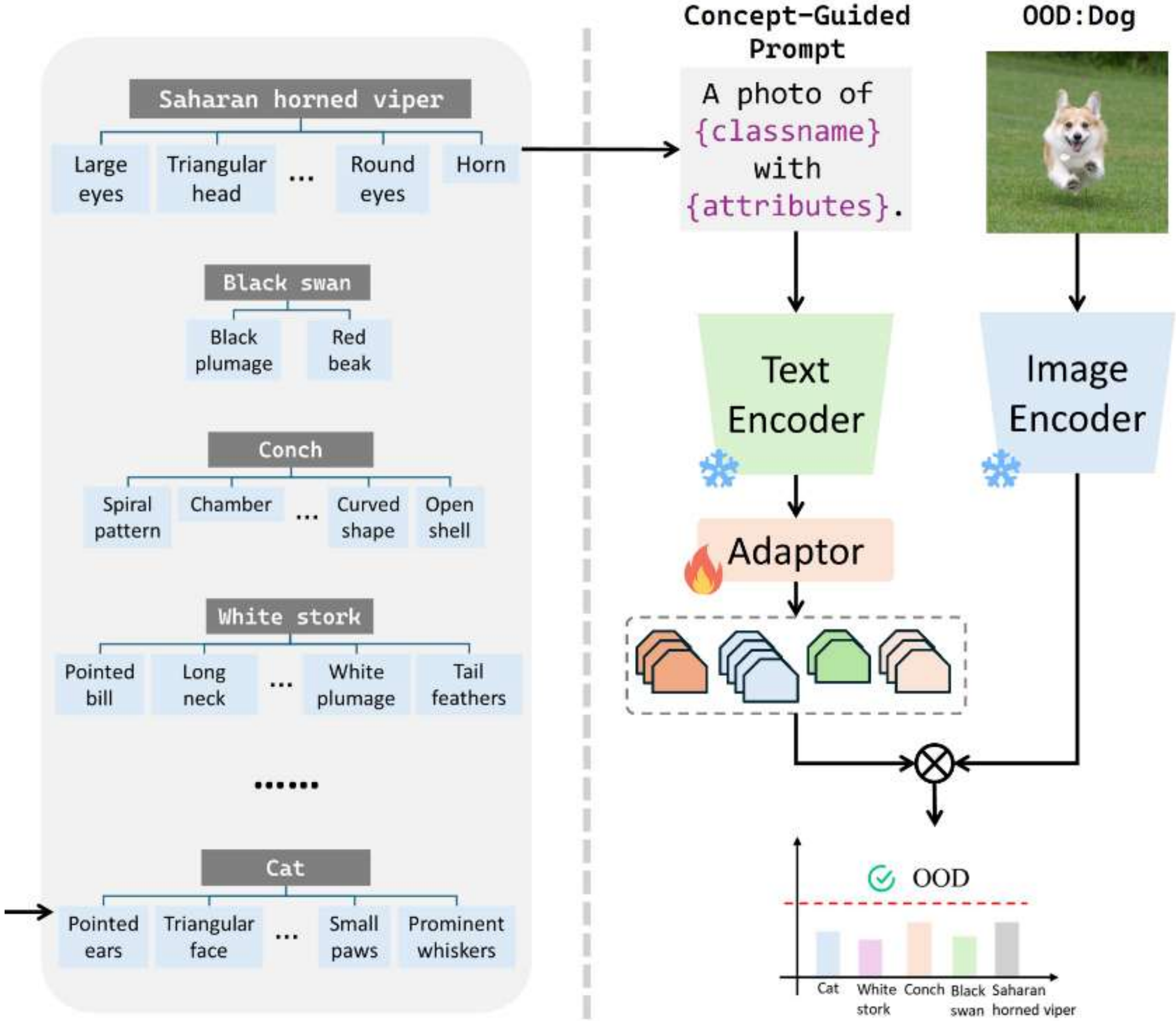
"Please describe the {classname} in the picture."

Concept Documentation



(a) Automated Essential Concept Discovery (b) Essential Concept Memory

Concept Documentation



- 文本特征 $\boldsymbol{F}_t = (1-\alpha)\cdot a(g(\boldsymbol{t})) + \alpha \cdot g(\boldsymbol{t})$
- 视觉特征 $\boldsymbol{F}_v = f(\boldsymbol{x}_i)$
- 特征对齐 $\hat{\boldsymbol{F}}_t = \mathcal{W}\cdot\mathcal{M}\cdot\boldsymbol{F}_t$
- 预测概率

$$p(y=y_i|\boldsymbol{x}_i) = \frac{\exp\left(\mathrm{sim}\left(\boldsymbol{F}_v, \hat{\boldsymbol{F}}_t^i\right)/\tau\right)}{\sum_{c=1}^{C}\exp\left(\mathrm{sim}\left(\boldsymbol{F}_v, \hat{\boldsymbol{F}}_t^c\right)/\tau\right)}$$

- OOD分数

$$\mathrm{Score}_{\mathrm{OOD}}(\boldsymbol{x}) = -\max_{c}\, p(y=c|\boldsymbol{x}).$$

Experiment

模式识别与神经计算研究组
PAttern Recognition and NEural Computing

Method	Texture		iNaturalist		Places		SUN		Avg	
	AUC ↑	FPR95 ↓	AUC ↑	FPR95 ↓	AUC ↑	FPR95 ↓	AUC ↑	FPR95 ↓	AUC ↑	FPR95 ↓
Zero-shot methods										
MCM [32]	86.11	57.77	94.61	30.91	89.77	44.69	92.57	34.59	90.76	42.74
CLIP-based posthoc methods										
MSP [11]	74.84	73.66	77.74	74.57	72.18	79.12	73.97	76.95	74.98	76.22
MaxLogit [13]	88.63	48.72	88.03	60.88	87.45	55.54	91.16	44.83	88.82	52.49
Energy [27]	88.22	50.39	87.18	64.98	87.33	57.40	91.17	46.42	88.48	54.80
ReAct [40]	88.13	49.88	86.87	65.57	87.42	56.85	91.04	46.17	88.37	54.62
ODIN [24]	87.85	51.67	94.65	30.22	85.54	55.06	87.17	54.04	88.80	47.75
Prompt learning methods										
CoOp [55]	89.47	45.00	93.77	29.81	90.58	40.11	93.29	40.83	91.78	51.68
LoCoOp [33]	90.19	42.28	96.86	16.05	91.98	32.87	95.07	23.44	93.52	28.66
LoCoOp+AECD	92.01	38.40	97.32	14.34	92.00	32.07	94.73	25.18	94.02	27.50
Open-vocabulary OOD detection										
CoOp (10%) [23]	87.58	50.55	91.08	42.53	89.56	46.12	91.52	41.92	89.94	45.28
LoCoOp (10%) [23]	88.21	47.32	94.47	34.90	91.64	39.85	92.54	26.30	91.72	37.09
LoCoOp+AECD (10%)	90.11	44.42	95.23	32.75	92.33	41.34	93.63	24.3	92.83	35.70

模式分析与机器智能工业和信息化部重点实验室
MIIT Key Laboratory of Pattern Analysis & Machine Intelligence

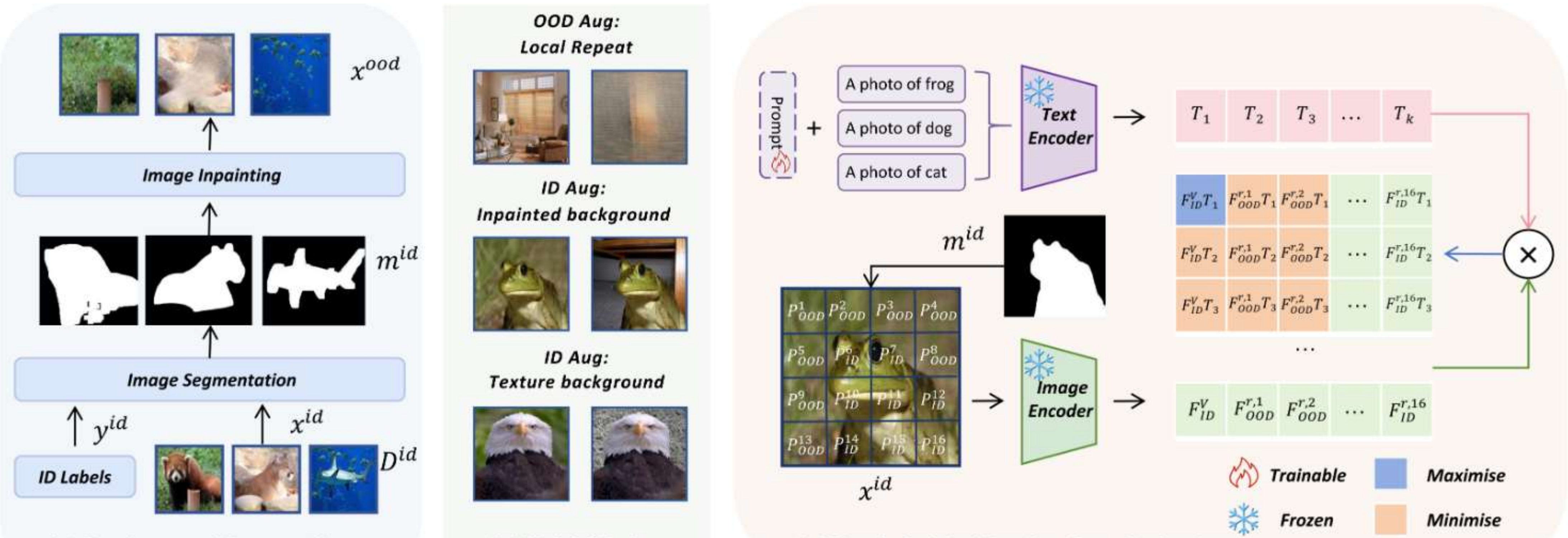
ParNeC 模式识别与神经计算研究组 PAttern Recognition and NEural Computing

Overcoming Shortcut Problem in VLM for Robust Out-of-Distribution Detection

Zhuo Xu[1]*, Xiang Xiang[1,2] * †, Yifan Liang[1]
[1] National Key Lab of Multi-Spectral Information Intelligent Processing Technology,
School of Artificial Intelligence and Automation,
Huazhong University of Science and Technology, Wuhan, China
[2] Peng Cheng National Laboratory, Shenzhen, China

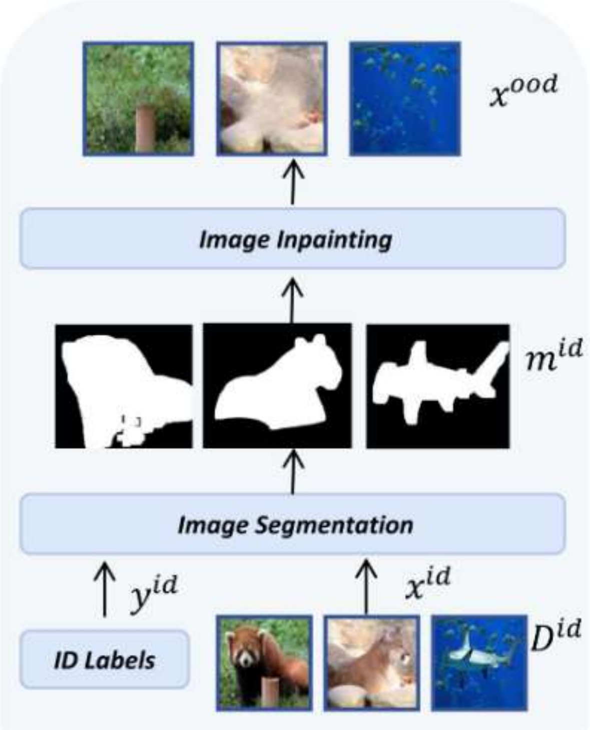
CVPR 2025

OSPCoOp



(a) Background Decoupling (b) ID&OOD Aug (c) Mask-Guided Region Regularization

Background Decoupling



(a) Background Decoupling

- 语义分割 $m^{id} = \mathrm{Seg}(x^{id}, y^{id})$
- Inpainting $x^{ood} = \mathrm{Inp}(x^{id} \odot (1 - m^{id}), m^{id} \oplus B)$
- CLIP模型进一步筛选

$$x^{ood} = \{x^{ood} \mid \epsilon < z_k^{id} - z_k^{ood}\},\ z_k = \mathrm{sim}(\mathcal{I}(x), \mathcal{T}(t_k))$$

Background Decoupling

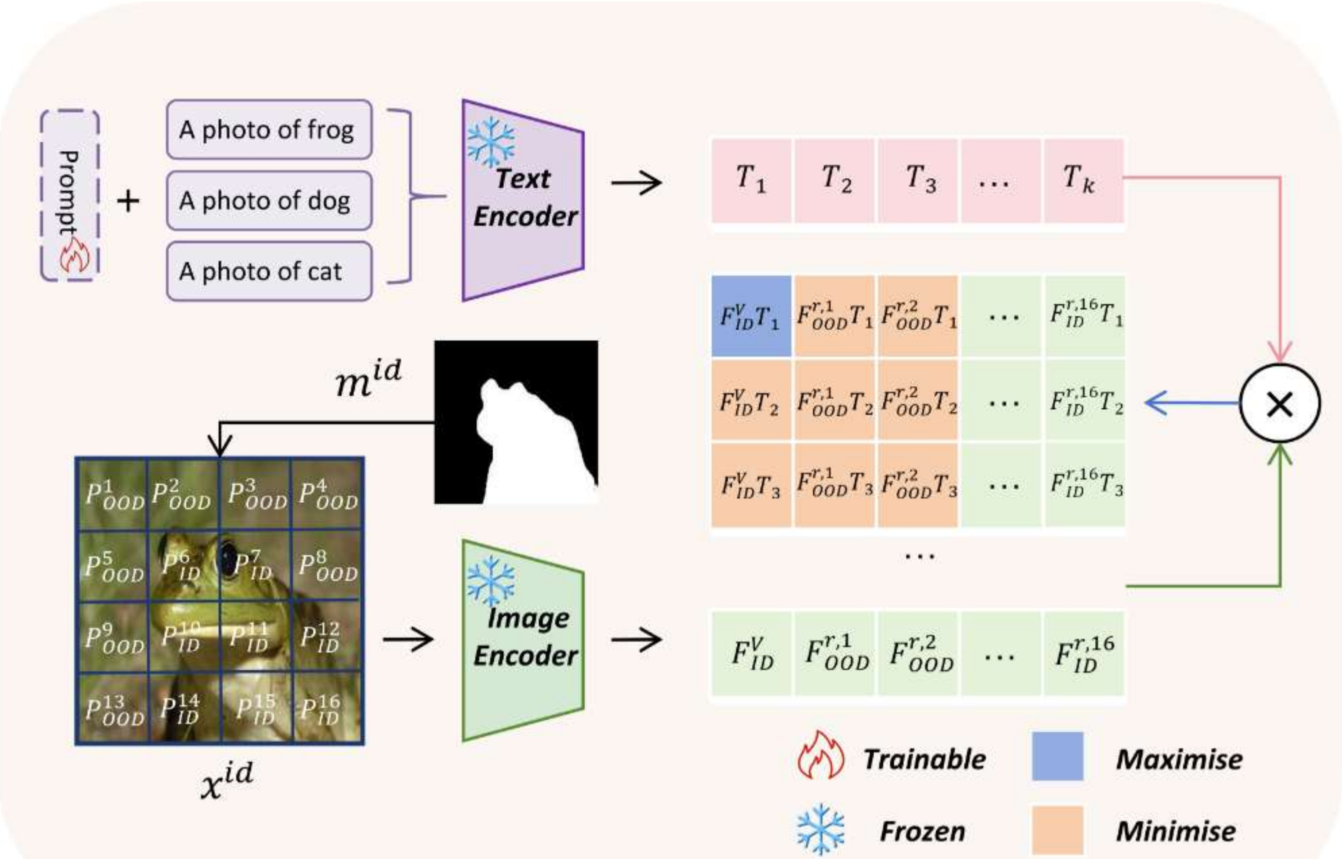
Figure 3. Some images from ImageNet-Bg (right) and the ImageNet validation set (left).

Image Augmentation

(b) ID&OOD Aug

- OOD图像增强：背景inpainting + ID区域纹理重复
- ID图像增强：ID区域和OOD背景组合

Mask-Guided Region Regularization



(c) Mask-Guided Region Regularization

- ID-global relevance loss

$$\mathcal{L}_{id} = -\frac{1}{|D^{id}_{train}|} \sum_{x^{id} \in D^{id}_{train}} \log \frac{\exp(z_k/\tau)}{\sum_{j=1}^{K} \exp(z_j/\tau)}$$

- OOD-global irrelevance loss

$$\mathcal{L}^{g}_{ood} = -\frac{1}{|D^{ood}_{train}|} \sum_{x \in D^{ood}_{train}} H[p(y \mid x)]$$

- OOD-region irrelevance loss

$$\mathcal{L}^{r}_{ood} = -\frac{1}{|D^{all}_{train}|} \sum_{x_i \in D^{all}_{train}} \frac{1}{N_i} \sum_{n=1}^{N_i} H[p(y \mid P^{ood}_n)]$$

$$P \in \begin{cases} P^{id} & \text{if } \sum_{j=0}^{n} m^{id}_j > \theta \\ P^{ood} & \text{if } \sum_{j=0}^{n} m^{id}_j \le \theta \end{cases}$$

- OOD inference

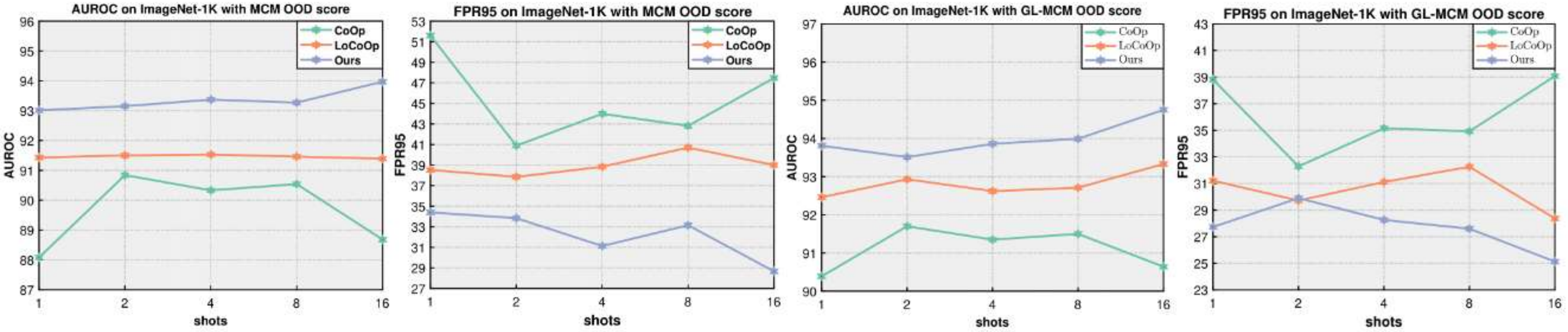
$$S_{OOD} = \max_k p(y_k \mid \mathbf{x}) + \beta \max_{k,i} p(y_k \mid \mathbf{P}_i)$$

Experiment

ParNeC 模式识别与神经计算研究组 PAttern Recognition and NEural Computing

Method	iNaturalist		SUN		Places		Texture		Avg	
	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓
Training-free methods										
ZOC [10]	86.09	87.30	81.20	81.51	83.39	73.06	76.46	98.90	81.79	85.19
MCM [33]	94.61	30.94	92.56	37.67	89.76	44.76	86.10	57.91	90.76	42.82
GL-MCM [39]	96.71	15.18	93.09	30.42	89.90	38.85	83.63	57.93	90.83	35.47
CLIPN-A [49]	95.27	23.94	93.93	26.17	90.93	40.83	92.28	33.45	93.10	31.10
DPM-F [57] †	96.84	15.26	91.78	42.58	89.60	45.99	85.74	57.55	90.99	40.35
Outlier-label exposure methods										
NegLabel [21]	99.48	1.99	95.43	21.05	91.95	34.95	90.90	44.79	94.25	25.69
LAPT [55]	99.63	1.16	96.01	19.12	92.01	33.01	91.06	40.32	94.68	23.40
EOE [4]	97.52	12.29	95.73	20.40	92.94	30.16	85.64	57.53	92.96	30.09
OLE [8]	98.33	7.61	94.87	22.44	92.45	31.73	92.40	34.70	94.51	24.12
Requires few-shot training (or w/ fine-tuning)										
CoOp [60]	96.62	14.60	92.65	28.48	89.98	36.49	88.03	43.13	91.82	30.67
LoCoOp [35]	96.86	16.05	95.07	23.44	91.98	32.87	90.19	42.28	93.52	28.66
SCT [51]	95.86	13.94	95.33	20.55	92.24	29.86	89.06	41.51	93.37	26.47
DPM-T [57] †	97.04	14.47	93.19	33.06	89.78	39.46	87.49	49.73	91.88	34.18
ID-like [2]	**98.19**	**8.98**	91.64	42.03	91.15	41.74	**94.38**	**26.77**	93.84	29.88
NegPrompt [27] †	90.69	45.97	92.18	39.43	91.65	37.49	90.01	44.84	91.13	41.93
OSPCoOp (Ours)	97.13	15.25	**96.74**	**18.26**	**94.01**	**25.74**	91.13	41.26	**94.75**	**25.13**

Experiment



OOD Aug	Places		Texture		Avg	
	AUR ↑	FPR ↓	AUR ↑	FPR ↓	AUR ↑	FPR ↓
None	92.02	34.71	89.27	46.38	92.97	32.54
Rep	92.58	32.17	90.08	44.50	93.43	29.82
Bg	93.85	25.95	90.90	41.35	94.65	24.77
Rep+Bg	**94.01**	**25.74**	**91.13**	**41.26**	**94.75**	**25.13**

Table 4. Ablation study of OOD augmentation. ’Rep’ stands for the data generated by repeating local ID regions, and ’Bg’ stands for using the decoupled OOD content with ID regions inpainted. ’Avg’ stands for the results on the four traditional OOD datasets.

Thanks