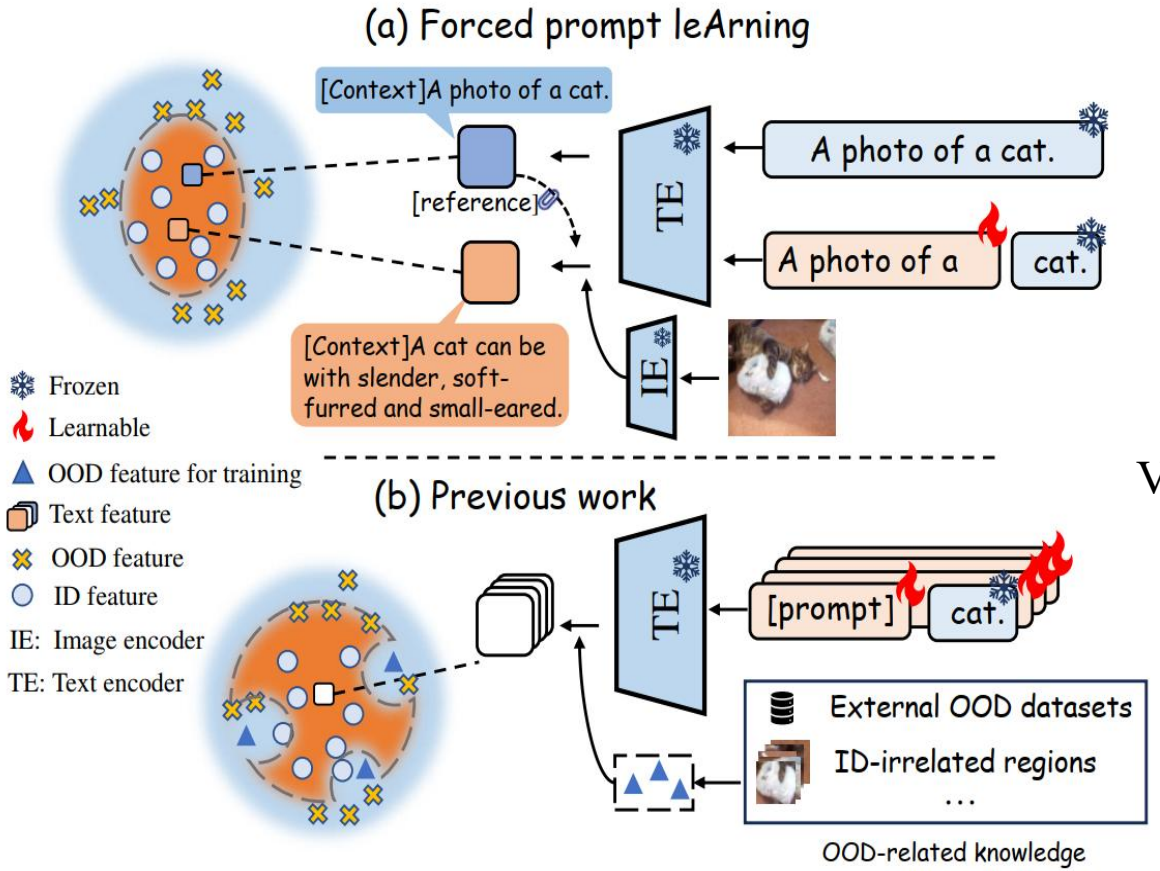
FA: Forced Prompt Learning of Vision-Language Models for Out-of-Distribution Detection

Xinhua Lu[1,2,4], Runhe Lai[1,2,4], Yanqi Wu[1,2,4], Kanghao Chen[3†], Wei-Shi Zheng[1,2,4], Ruixuan Wang[1,2,4†]
[1]Sun Yat-sen University [2]Peng Cheng Laboratory
[3]Hong Kong University of Science and Technology (Guangzhou)
[4]Key Laboratory of Machine Intelligence and Advanced Computing
{luxh55, lairh5, wuyq268}@mail2.sysu.edu.cn, kchen879@connect.hkust-gz.edu.cn,
wszheng@ieee.org, wangruix5@mail.sysu.edu.cn

Background and Motivation



Leveraging OOD-related knowledge

Limitation

Existing VLM-based Methods

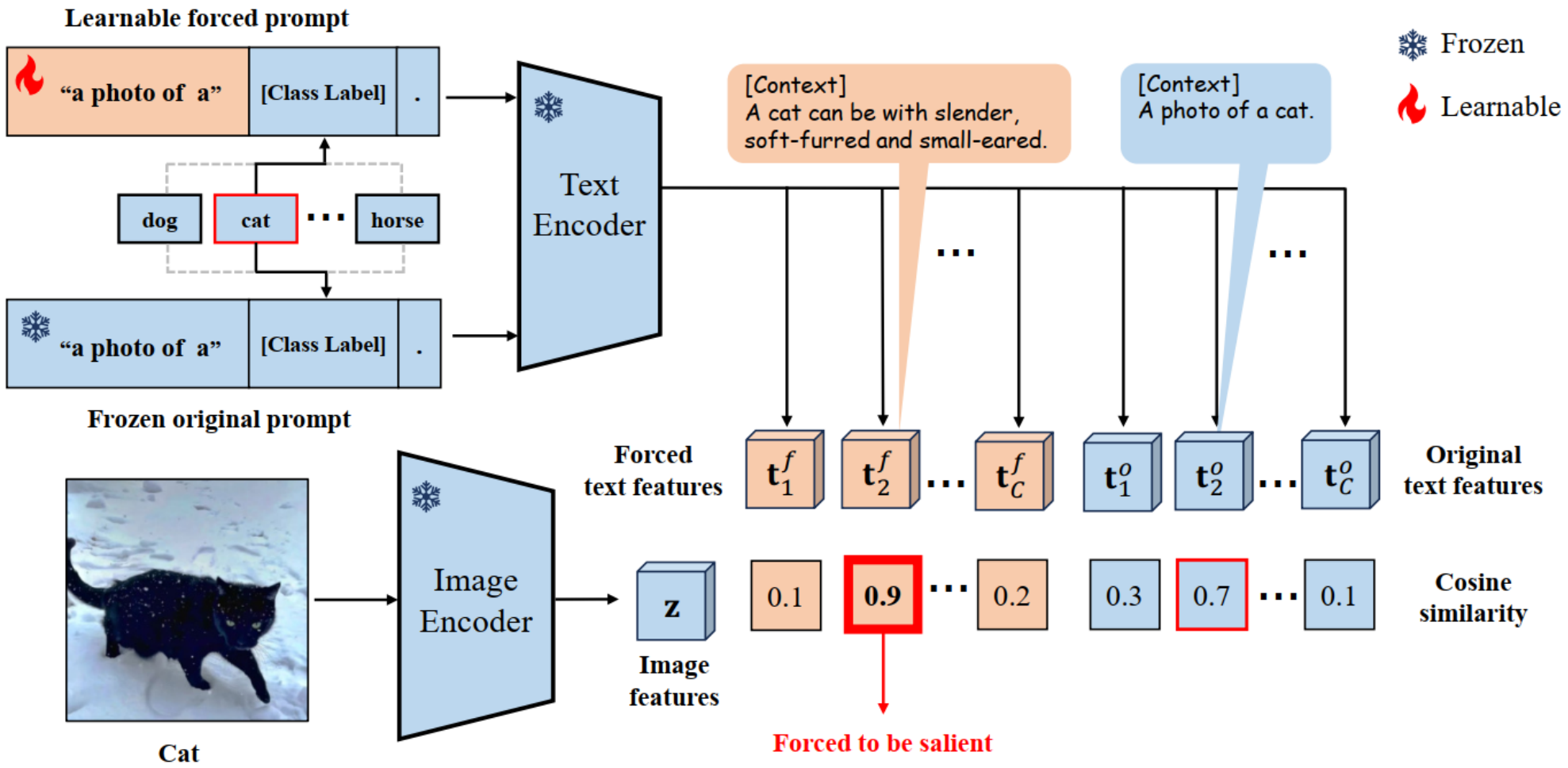
- Relying on external auxiliary datasets → Requires substantial computational resources as well as labor costs
- Using OOD data extracted from ID data → Difficult to match the infinite real-world outliers encountered in practice

Rather than focusing on the intricate OOD-related knowledge, our FA aims to fully exploit the ID knowledge and improve OOD detection.

Contribution

- We propose a simple yet effective framework, which fully exploits the ID knowledge to improve OOD detection without focusing on intricate OOD-related knowledge.
- We present a Forced prompt leArning (FA) strategy to exploit richer knowledge beyond the textual semantics of class labels.
- We evaluate our method on diverse OOD benchmarks, showing that our model consistently outperforms current SOTA methods.

Method


Learnable forced prompt
"a photo of a"
[Class Label]
.
dog
cat
horse
"a photo of a"
[Class Label]
.
Frozen original prompt
Text Encoder
[Context]
A cat can be with slender, soft-furred and small-eared.
[Context]
A photo of a cat.
Frozen
Learnable
Forced text features
Original text features
Image Encoder
Cat
z
Image features
0.1
0.9
0.2
0.3
0.7
0.1
Cosine similarity
Forced to be salient

Method

Training:

$$\mathcal{L}_{FCE} = \mathbb{E}_{(\mathbf{x},y_c)\sim\mathcal{D}^{ID}_{train}} \left[-\log \frac{e^{s^f_c/\tau}}{\sum_{j=1}^{C} e^{s^f_j/\tau} + \sum_{j=1}^{C} e^{s^o_j/\tau}} \right]$$

$$s^f_j = \cos(\mathbf{z}, \mathbf{t}^f_j) \text{ and } s^o_j = \cos(\mathbf{z}, \mathbf{t}^o_j)$$

The text features of the forced prompt achieve a higher cosine similarity with the image features compared to those of the original prompt.

$$\mathcal{L}_{FCE-K} = \mathbb{E}_{(\mathbf{x},y_c)\sim\mathcal{D}^{ID}_{train}} \left[-\log \frac{e^{s^f_c/\tau}}{\sum_{j=1}^{C} e^{s^f_j/\tau} + K\sum_{j=1}^{C} e^{s^o_j/\tau}} \right]$$

Method

Inference:

$$S_{\text{MCM}}(\mathbf{x}) = \max_{c} \frac{e^{\cos(\mathbf{z}^g, \mathbf{t}_c^a)/\tau_0}}{\sum_{j=1}^{C} e^{\cos(\mathbf{z}^g, \mathbf{t}_j^f)/\tau_0} + K e^{\cos(\mathbf{z}^g, \mathbf{t}_j^o)/\tau_0}},$$

$$S_{\text{L-MCM}}(\mathbf{x}) = \max_{i,c} \frac{e^{\cos(\mathbf{z}_i^l, \mathbf{t}_c^a)/\tau_0}}{\sum_{j=1}^{C} e^{\cos(\mathbf{z}_i^l, \mathbf{t}_j^f)/\tau_0} + K e^{\cos(\mathbf{z}_i^l, \mathbf{t}_j^o)/\tau_0}}$$

$$S_{\text{GL-MCM}}(\mathbf{x}) = S_{\text{MCM}}(\mathbf{x}) + S_{\text{L-MCM}}(\mathbf{x})$$

The difference in semantic similarity with the two types of prompts **:**

ID > OOD

MCM score
GL-MCM score **:**

ID > OOD

Experiments

Method	iNaturalist		SUN		Places		Textures		Average	
	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑
Zero-shot methods										
MCM	31.95	94.16	37.22	92.55	42.98	90.10	58.35	85.83	42.63	90.66
GL-MCM	15.09	96.72	29.08	93.41	37.07	90.37	58.94	83.11	35.04	90.90
CLIPN	19.17	96.17	26.43	94.02	32.26	92.62	41.23	90.12	30.21	93.19
CLIP-based post-hoc methods										
MSP†	74.57	77.74	76.95	73.97	79.72	72.18	73.66	74.84	76.22	74.68
ODIN†	98.93	57.73	88.72	78.42	87.80	76.88	85.47	71.49	90.23	71.13
Energy†	64.98	87.18	46.42	91.17	57.40	87.33	50.39	88.22	54.80	88.48
ReAct†	65.57	86.87	46.17	91.04	56.85	87.42	49.88	88.13	54.62	88.37
MaxLogit†	60.88	88.03	44.83	91.16	55.54	87.45	48.72	88.63	52.49	88.82
Prompt learning based methods					1-shot					
CoOp_{MCM}	$41.14^{\pm 9.39}$	$91.47^{\pm 2.12}$	$39.06^{\pm 4.17}$	$91.68^{\pm 0.83}$	$45.38^{\pm 4.88}$	$89.16^{\pm 1.28}$	$51.37^{\pm 3.09}$	$87.82^{\pm 1.12}$	$44.24^{\pm 1.39}$	$90.03^{\pm 0.32}$
CoOp_{GL}	$23.30^{\pm 6.37}$	$94.59^{\pm 1.51}$	$32.08^{\pm 3.87}$	$92.21^{\pm 0.95}$	$39.22^{\pm 4.55}$	$89.59^{\pm 1.58}$	$55.78^{\pm 3.46}$	$83.26^{\pm 1.38}$	$37.59^{\pm 0.34}$	$89.91^{\pm 0.31}$
$\text{LoCoOp}_{\text{MCM}}$	$36.64^{\pm 4.87}$	$92.78^{\pm 0.93}$	$31.86^{\pm 3.79}$	$93.50^{\pm 0.92}$	$38.81^{\pm 3.37}$	$90.70^{\pm 0.97}$	$48.11^{\pm 2.31}$	$89.43^{\pm 0.68}$	$38.85^{\pm 2.67}$	$91.59^{\pm 0.58}$
$\text{LoCoOp}_{\text{GL}}$	$21.97^{\pm 2.91}$	$95.39^{\pm 0.59}$	$24.95^{\pm 2.37}$	$\mathbf{94.42}^{\pm 0.59}$	$34.14^{\pm 2.66}$	$91.14^{\pm 0.69}$	$49.04^{\pm 2.77}$	$87.73^{\pm 0.96}$	$32.53^{\pm 2.19}$	$92.17^{\pm 0.47}$
IDLike	$17.73^{\pm 1.91}$	$96.68^{\pm 0.31}$	$48.17^{\pm 1.39}$	$89.53^{\pm 0.58}$	$50.43^{\pm 6.58}$	$88.27^{\pm 2.27}$	$29.12^{\pm 7.64}$	$93.25^{\pm 2.16}$	$36.36^{\pm 3.86}$	$91.93^{\pm 1.18}$
LSN†	$59.28^{\pm 7.02}$	$87.20^{\pm 3.15}$	$40.15^{\pm 0.82}$	$91.47^{\pm 0.14}$	$46.11^{\pm 1.86}$	$88.74^{\pm 0.57}$	$60.34^{\pm 0.14}$	$83.92^{\pm 0.42}$	$51.47^{\pm 1.53}$	$87.84^{\pm 0.58}$
NegPrompt†	$65.03^{\pm 8.69}$	$84.56^{\pm 2.52}$	$44.39^{\pm 1.66}$	$89.63^{\pm 0.66}$	$51.31^{\pm 6.21}$	$86.55^{\pm 2.19}$	$87.60^{\pm 1.61}$	$63.76^{\pm 3.02}$	$62.08^{\pm 3.71}$	$81.13^{\pm 1.78}$
SCT_{MCM}	$41.93^{\pm 12.17}$	$91.77^{\pm 2.40}$	$30.39^{\pm 2.17}$	$93.76^{\pm 0.39}$	$38.73^{\pm 2.54}$	$90.78^{\pm 0.34}$	$46.78^{\pm 3.51}$	$88.96^{\pm 1.16}$	$39.46^{\pm 3.39}$	$91.32^{\pm 0.90}$
SCT_{GL}	$20.57^{\pm 10.2}$	$95.63^{\pm 1.89}$	$\mathbf{24.56}^{\pm 3.03}$	$94.39^{\pm 0.53}$	$33.27^{\pm 2.96}$	$91.27^{\pm 0.57}$	$48.12^{\pm 2.97}$	$86.76^{\pm 0.89}$	$31.62^{\pm 3.19}$	$92.01^{\pm 0.77}$
FA_{MCM}(Ours)	$25.50^{\pm 2.72}$	$94.72^{\pm 0.27}$	$36.24^{\pm 3.15}$	$92.40^{\pm 0.95}$	$35.38^{\pm 2.75}$	$\mathbf{91.99}^{\pm 0.68}$	$\mathbf{28.34}^{\pm 0.87}$	$\mathbf{93.95}^{\pm 0.19}$	$31.37^{\pm 1.07}$	$93.25^{\pm 0.32}$
FA_{GL}(Ours)	$\mathbf{14.12}^{\pm 1.32}$	$\mathbf{96.76}^{\pm 0.10}$	$29.99^{\pm 1.57}$	$92.95^{\pm 0.66}$	$\mathbf{32.48}^{\pm 1.48}$	$91.83^{\pm 0.49}$	$34.66^{\pm 1.21}$	$91.50^{\pm 0.36}$	$\mathbf{27.81}^{\pm 0.44}$	$\mathbf{93.26}^{\pm 0.27}$
Prompt learning based methods					16-shot					
CoOp_{MCM}	$30.26^{\pm 1.98}$	$93.43^{\pm 0.81}$	$34.69^{\pm 0.43}$	$92.59^{\pm 0.14}$	$41.91^{\pm 0.71}$	$90.11^{\pm 0.23}$	$44.68^{\pm 2.11}$	$89.95^{\pm 0.48}$	$37.89^{\pm 0.71}$	$91.52^{\pm 0.29}$
CoOp_{GL}	$15.96^{\pm 1.67}$	$96.11^{\pm 0.55}$	$27.26^{\pm 1.99}$	$93.29^{\pm 0.46}$	$35.36^{\pm 2.08}$	$90.58^{\pm 0.64}$	$48.63^{\pm 2.11}$	$86.11^{\pm 0.59}$	$31.81^{\pm 1.27}$	$91.51^{\pm 0.39}$
$\text{LoCoOp}_{\text{MCM}}$	$27.35^{\pm 3.19}$	$94.12^{\pm 0.80}$	$30.93^{\pm 1.19}$	$93.75^{\pm 0.26}$	$38.26^{\pm 1.53}$	$91.12^{\pm 0.27}$	$41.36^{\pm 2.56}$	$90.99^{\pm 0.53}$	$34.47^{\pm 0.73}$	$92.49^{\pm 0.14}$
$\text{LoCoOp}_{\text{GL}}$	$18.46^{\pm 1.38}$	$95.85^{\pm 0.63}$	$22.43^{\pm 0.96}$	$95.15^{\pm 0.22}$	$31.53^{\pm 1.52}$	$92.15^{\pm 0.22}$	$43.35^{\pm 3.12}$	$89.38^{\pm 0.78}$	$28.94^{\pm 1.29}$	$93.13^{\pm 0.17}$
IDLike	$19.23^{\pm 10.5}$	$96.70^{\pm 1.58}$	$54.15^{\pm 2.88}$	$87.64^{\pm 1.19}$	$56.63^{\pm 0.07}$	$85.86^{\pm 0.44}$	$34.69^{\pm 6.41}$	$91.90^{\pm 2.31}$	$41.18^{\pm 1.73}$	$90.53^{\pm 0.01}$
LSN†	$36.17^{\pm 4.81}$	$92.66^{\pm 1.16}$	$34.27^{\pm 0.44}$	$93.53^{\pm 0.20}$	$41.47^{\pm 0.85}$	$90.52^{\pm 0.37}$	$46.43^{\pm 0.60}$	$89.38^{\pm 0.24}$	$39.58^{\pm 0.73}$	$91.53^{\pm 0.09}$
NegPrompt†	$37.79^{\pm 0.11}$	$90.49^{\pm 0.01}$	$32.11^{\pm 3.77}$	$92.25^{\pm 1.00}$	$35.52^{\pm 0.41}$	$91.16^{\pm 0.03}$	$43.93^{\pm 9.09}$	$88.38^{\pm 3.31}$	$37.34^{\pm 1.41}$	$90.57^{\pm 0.59}$
SCT_{MCM}	$29.41^{\pm 2.19}$	$93.76^{\pm 0.55}$	$27.28^{\pm 2.80}$	$94.22^{\pm 0.40}$	$36.35^{\pm 2.13}$	$91.16^{\pm 0.33}$	$42.25^{\pm 1.89}$	$90.52^{\pm 0.48}$	$33.82^{\pm 1.78}$	$92.42^{\pm 0.35}$
SCT_{GL}	$15.19^{\pm 2.16}$	$\mathbf{96.71}^{\pm 0.44}$	$\mathbf{20.00}^{\pm 0.61}$	$\mathbf{95.57}^{\pm 0.11}$	$\mathbf{29.71}^{\pm 0.85}$	$92.37^{\pm 0.12}$	$44.17^{\pm 0.82}$	$88.59^{\pm 0.39}$	$27.27^{\pm 0.44}$	$93.31^{\pm 0.17}$
FA_{MCM}(Ours)	$25.79^{\pm 1.48}$	$94.29^{\pm 0.35}$	$33.54^{\pm 1.17}$	$93.02^{\pm 0.19}$	$33.77^{\pm 1.64}$	$\mathbf{92.64}^{\pm 0.42}$	$\mathbf{23.17}^{\pm 1.31}$	$\mathbf{95.14}^{\pm 0.26}$	$29.07^{\pm 1.11}$	$93.77^{\pm 0.19}$
FA_{GL}(Ours)	$\mathbf{14.49}^{\pm 1.27}$	$96.48^{\pm 0.29}$	$27.65^{\pm 1.08}$	$93.46^{\pm 0.18}$	$31.09^{\pm 1.38}$	$92.44^{\pm 0.34}$	$29.50^{\pm 0.89}$	$92.93^{\pm 0.18}$	$\mathbf{25.68}^{\pm 0.58}$	$\mathbf{93.82}^{\pm 0.11}$

Method	1-shot	4-shot	16-shot
CoOp	$67.44^{\pm 0.50}$	$69.71^{\pm 0.07}$	$70.99^{\pm 0.14}$
LoCoOp	$67.40^{\pm 0.64}$	$69.55^{\pm 0.10}$	$71.53^{\pm 0.17}$
IDLike	$68.17^{\pm 0.57}$	$68.91^{\pm 0.14}$	$69.46^{\pm 0.02}$
SCT	$68.63^{\pm 0.13}$	$69.93^{\pm 0.22}$	$\mathbf{71.78}^{\pm 0.05}$
FA(Ours)	$\mathbf{68.67}^{\pm 0.39}$	$\mathbf{69.96}^{\pm 0.04}$	$71.02^{\pm 0.09}$

Conventional OOD detection benchmark using ImageNet-1k as the ID dataset

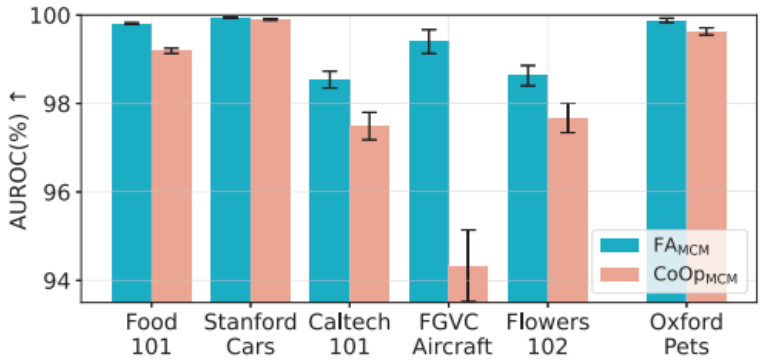
Experiments

Method	OpenImage-O		NINCO		ImageNet-O		Average	
	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑
CoOp_{MCM}	$37.95^{\pm 0.59}$	$91.99^{\pm 0.22}$	$78.00^{\pm 0.23}$	$73.00^{\pm 0.59}$	$70.14^{\pm 0.86}$	$\underline{81.34}^{\pm 0.83}$	$62.03^{\pm 0.38}$	$82.11^{\pm 0.48}$
CoOp_{GL}	$\underline{32.23}^{\pm 1.00}$	$92.22^{\pm 0.25}$	$72.04^{\pm 0.81}$	$75.35^{\pm 0.88}$	$67.35^{\pm 1.19}$	$78.44^{\pm 0.94}$	$\underline{57.21}^{\pm 0.96}$	$81.99^{\pm 0.58}$
$\text{LoCoOp}_{\text{MCM}}$	$37.03^{\pm 0.94}$	$92.24^{\pm 0.21}$	$77.96^{\pm 0.63}$	$72.78^{\pm 0.79}$	$70.96^{\pm 0.75}$	$81.25^{\pm 0.39}$	$61.98^{\pm 0.55}$	$82.09^{\pm 0.35}$
$\text{LoCoOp}_{\text{GL}}$	$33.87^{\pm 1.23}$	$\underline{92.53}^{\pm 0.07}$	$75.16^{\pm 0.92}$	$72.97^{\pm 0.39}$	$68.06^{\pm 1.92}$	$81.19^{\pm 0.25}$	$59.03^{\pm 1.27}$	$82.23^{\pm 0.17}$
IDLike	$54.43^{\pm 2.35}$	$87.89^{\pm 0.45}$	$78.93^{\pm 1.79}$	$69.32^{\pm 0.42}$	$84.08^{\pm 1.45}$	$67.45^{\pm 0.86}$	$72.48^{\pm 0.89}$	$74.89^{\pm 0.28}$
SCT_{MCM}	$37.28^{\pm 1.16}$	$92.04^{\pm 0.33}$	$78.51^{\pm 0.86}$	$71.09^{\pm 1.19}$	$71.45^{\pm 0.54}$	$81.29^{\pm 0.15}$	$62.41^{\pm 0.73}$	$81.48^{\pm 0.52}$
SCT_{GL}	$32.24^{\pm 0.34}$	$\mathbf{92.75}^{\pm 0.17}$	$74.15^{\pm 0.53}$	$73.14^{\pm 0.94}$	$68.35^{\pm 0.62}$	$80.84^{\pm 0.45}$	$58.25^{\pm 0.39}$	$82.24^{\pm 0.44}$
FA_{MCM}(Ours)	$37.93^{\pm 1.65}$	$91.84^{\pm 0.45}$	$\underline{70.44}^{\pm 1.24}$	$\underline{77.77}^{\pm 0.94}$	$\underline{63.84}^{\pm 1.36}$	$\mathbf{82.71}^{\pm 0.30}$	$57.40^{\pm 1.34}$	$\mathbf{84.08}^{\pm 0.47}$
FA_{GL}(Ours)	$\mathbf{32.10}^{\pm 0.89}$	$92.39^{\pm 0.28}$	$\mathbf{65.61}^{\pm 1.32}$	$\mathbf{79.01}^{\pm 0.71}$	$\mathbf{63.13}^{\pm 0.98}$	$80.39^{\pm 0.21}$	$\mathbf{53.61}^{\pm 0.89}$	$\underline{83.93}^{\pm 0.31}$

Challenging OOD detection benchmark using ImageNet-1k as the ID dataset

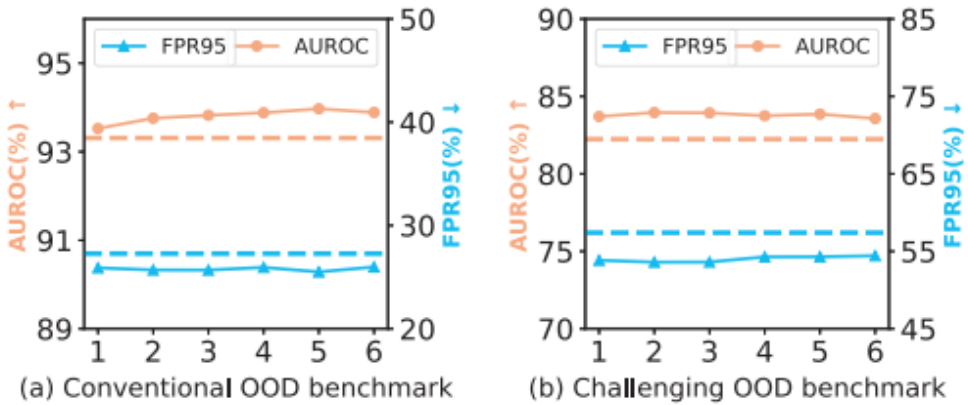
Experiments

ID Dataset	Method	iNaturalist		SUN		Places		Textures		Average	
		FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑
UCF101	CoOp_{MCM}	$14.40^{\pm 6.61}$	$97.41^{\pm 1.36}$	$37.09^{\pm 3.94}$	$93.14^{\pm 0.48}$	$40.07^{\pm 2.77}$	$91.89^{\pm 0.65}$	$33.24^{\pm 4.21}$	$93.19^{\pm 1.22}$	$31.21^{\pm 3.58}$	$93.91^{\pm 0.45}$
	CoOp_{GL}	$9.65^{\pm 6.35}$	$98.18^{\pm 0.98}$	$28.54^{\pm 5.07}$	$94.51^{\pm 0.92}$	$32.94^{\pm 1.74}$	$93.28^{\pm 0.32}$	$35.96^{\pm 2.65}$	$91.83^{\pm 1.42}$	$26.77^{\pm 3.22}$	$94.45^{\pm 0.42}$
	$\text{LoCoOp}_{\text{MCM}}$	$2.38^{\pm 0.98}$	$99.21^{\pm 0.22}$	$19.73^{\pm 4.24}$	$96.11^{\pm 0.78}$	$22.57^{\pm 3.01}$	$94.96^{\pm 0.77}$	$15.89^{\pm 3.12}$	$96.12^{\pm 0.67}$	$15.14^{\pm 1.91}$	$96.60^{\pm 0.33}$
	$\text{LoCoOp}_{\text{GL}}$	$1.13^{\pm 0.22}$	$99.65^{\pm 0.09}$	$17.93^{\pm 4.09}$	$96.51^{\pm 0.89}$	$22.34^{\pm 4.58}$	$95.03^{\pm 1.06}$	$18.17^{\pm 3.01}$	$95.25^{\pm 0.71}$	$14.89^{\pm 2.49}$	$96.61^{\pm 0.58}$
	IDLike	$66.63^{\pm 27.6}$	$89.75^{\pm 5.52}$	$87.88^{\pm 6.56}$	$78.04^{\pm 6.14}$	$84.47^{\pm 12.1}$	$76.04^{\pm 8.53}$	$49.07^{\pm 11.2}$	$91.19^{\pm 2.74}$	$72.01^{\pm 14.0}$	$83.76^{\pm 5.64}$
	SCT_{MCM}	$1.51^{\pm 0.51}$	$99.44^{\pm 0.10}$	$24.17^{\pm 5.54}$	$95.21^{\pm 0.88}$	$26.62^{\pm 4.41}$	$93.75^{\pm 0.81}$	$15.34^{\pm 3.84}$	$95.98^{\pm 0.64}$	$16.91^{\pm 3.20}$	$96.09^{\pm 0.48}$
	SCT_{GL}	$0.79^{\pm 0.27}$	$99.75^{\pm 0.04}$	$18.69^{\pm 2.41}$	$95.86^{\pm 0.56}$	$23.77^{\pm 2.47}$	$93.98^{\pm 0.47}$	$17.44^{\pm 2.62}$	$94.52^{\pm 0.54}$	$15.17^{\pm 1.56}$	$96.03^{\pm 0.25}$
	FA_{MCM}(Ours)	$\mathbf{0.12}^{\pm 0.11}$	$\mathbf{99.94}^{\pm 0.01}$	$\mathbf{3.06}^{\pm 1.33}$	$\mathbf{99.34}^{\pm 0.23}$	$\mathbf{4.68}^{\pm 1.00}$	$\mathbf{99.08}^{\pm 0.26}$	$\mathbf{2.68}^{\pm 0.51}$	$\mathbf{99.31}^{\pm 0.43}$	$\mathbf{2.63}^{\pm 0.58}$	$\mathbf{99.42}^{\pm 0.22}$
	FA_{GL}(Ours)	$0.13^{\pm 0.02}$	$99.94^{\pm 0.02}$	$4.98^{\pm 1.02}$	$99.01^{\pm 0.22}$	$7.46^{\pm 0.86}$	$98.44^{\pm 0.27}$	$5.67^{\pm 0.83}$	$98.75^{\pm 0.19}$	$4.56^{\pm 0.47}$	$99.03^{\pm 0.13}$
EuroSAT	CoOp_{MCM}	$99.14^{\pm 0.73}$	$31.95^{\pm 9.17}$	$98.29^{\pm 0.49}$	$44.36^{\pm 4.19}$	$97.28^{\pm 0.85}$	$47.51^{\pm 4.63}$	$86.69^{\pm 4.37}$	$69.17^{\pm 3.15}$	$95.35^{\pm 1.03}$	$48.25^{\pm 3.09}$
	CoOp_{GL}	$92.45^{\pm 4.26}$	$57.59^{\pm 10.4}$	$90.97^{\pm 3.49}$	$61.89^{\pm 4.83}$	$88.69^{\pm 3.85}$	$64.00^{\pm 4.92}$	$62.37^{\pm 1.93}$	$83.58^{\pm 2.31}$	$83.62^{\pm 1.18}$	$66.77^{\pm 2.83}$
	$\text{LoCoOp}_{\text{MCM}}$	$96.51^{\pm 3.33}$	$53.49^{\pm 8.06}$	$93.52^{\pm 5.68}$	$55.59^{\pm 9.19}$	$92.88^{\pm 4.05}$	$57.89^{\pm 5.19}$	$76.45^{\pm 7.48}$	$78.59^{\pm 2.66}$	$89.84^{\pm 4.29}$	$61.39^{\pm 3.90}$
	$\text{LoCoOp}_{\text{GL}}$	$86.13^{\pm 12.2}$	$72.68^{\pm 7.97}$	$87.81^{\pm 6.37}$	$68.58^{\pm 6.40}$	$85.93^{\pm 4.15}$	$69.90^{\pm 3.40}$	$54.89^{\pm 4.67}$	$87.00^{\pm 1.01}$	$78.69^{\pm 4.65}$	$74.54^{\pm 3.59}$
	IDLike	$98.02^{\pm 2.13}$	$73.32^{\pm 5.83}$	$95.74^{\pm 2.84}$	$65.19^{\pm 6.86}$	$94.79^{\pm 3.67}$	$69.44^{\pm 5.06}$	$63.34^{\pm 9.10}$	$86.68^{\pm 2.98}$	$87.97^{\pm 4.21}$	$73.66^{\pm 5.04}$
	SCT_{MCM}	$99.79^{\pm 0.22}$	$32.90^{\pm 6.73}$	$98.36^{\pm 0.84}$	$44.30^{\pm 3.41}$	$97.77^{\pm 0.86}$	$48.08^{\pm 1.64}$	$83.98^{\pm 6.92}$	$73.25^{\pm 3.58}$	$94.97^{\pm 2.19}$	$49.63^{\pm 2.39}$
	SCT_{GL}	$96.57^{\pm 5.41}$	$58.21^{\pm 5.31}$	$92.41^{\pm 4.55}$	$61.54^{\pm 3.47}$	$90.49^{\pm 5.02}$	$63.33^{\pm 2.71}$	$55.29^{\pm 14.7}$	$85.48^{\pm 3.68}$	$83.69^{\pm 7.02}$	$67.14^{\pm 2.20}$
	FA_{MCM}(Ours)	$85.24^{\pm 6.88}$	$78.62^{\pm 3.18}$	$\mathbf{36.24}^{\pm 10.8}$	$\mathbf{92.13}^{\pm 2.44}$	$\mathbf{28.17}^{\pm 10.2}$	$\mathbf{94.05}^{\pm 1.68}$	$10.36^{\pm 3.97}$	$97.71^{\pm 0.69}$	$\mathbf{39.99}^{\pm 7.79}$	$\mathbf{90.63}^{\pm 1.91}$
	FA_{GL}(Ours)	$\mathbf{77.82}^{\pm 7.28}$	$\mathbf{86.88}^{\pm 2.15}$	$56.10^{\pm 8.03}$	$88.43^{\pm 2.26}$	$49.87^{\pm 7.69}$	$89.07^{\pm 1.49}$	$\mathbf{8.99}^{\pm 2.51}$	$\mathbf{97.74}^{\pm 0.36}$	$48.19^{\pm 6.23}$	$90.53^{\pm 1.42}$



Using various ID datasets

Experiments



Sensitivity study of hyperparameter K

Forced-M-Init	Original-M-Init	FA_{MCM} FPR95↓	FA_{MCM} AUROC↑	FA_{GL} FPR95↓	FA_{GL} AUROC↑
-	-	$40.71^{\pm 4.00}$	$91.28^{\pm 0.82}$	$34.88^{\pm 2.92}$	$92.12^{\pm 0.76}$
✓	-	$40.81^{\pm 2.25}$	$91.25^{\pm 0.64}$	$36.03^{\pm 2.84}$	$91.62^{\pm 1.01}$
-	✓	$31.35^{\pm 1.34}$	$93.35^{\pm 0.34}$	$26.39^{\pm 1.02}$	$93.64^{\pm 0.19}$
✓	✓	$29.07^{\pm 1.11}$	$93.77^{\pm 0.19}$	$\mathbf{25.68}^{\pm 0.58}$	$\mathbf{93.82}^{\pm 0.11}$

Ablation study of different initialization scenarios for the forced and original prompts

Method	MCM FPR95↓	MCM AUROC↑	GL-MCM FPR95↓	GL-MCM AUROC↑
FA_{CE}	$41.43^{\pm 3.11}$	$91.01^{\pm 0.67}$	$34.29^{\pm 1.94}$	$90.99^{\pm 0.56}$
FA_{FCE-K}	$29.07^{\pm 1.11}$	$93.77^{\pm 0.19}$	$\mathbf{25.68}^{\pm 0.58}$	$\mathbf{93.82}^{\pm 0.11}$

Influence of the FCE loss

Shared-Vector	M-Init	FA_{MCM} FPR95↓	FA_{MCM} AUROC↑	FA_{GL} FPR95↓	FA_{GL} AUROC↑
-	-	$44.02^{\pm 2.24}$	$90.17^{\pm 0.42}$	$46.29^{\pm 1.33}$	$88.24^{\pm 0.41}$
✓	-	$40.71^{\pm 4.00}$	$91.28^{\pm 0.82}$	$34.88^{\pm 2.92}$	$92.12^{\pm 0.76}$
-	✓	$34.69^{\pm 0.62}$	$92.57^{\pm 0.12}$	$29.27^{\pm 0.55}$	$92.82^{\pm 0.80}$
✓	✓	$29.07^{\pm 1.11}$	$93.77^{\pm 0.19}$	$\mathbf{25.68}^{\pm 0.58}$	$\mathbf{93.82}^{\pm 0.11}$

Ablation study of the shared learnable vector

Thank You !