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

Atsuyuki Miyai[1], Qing Yu[1,2], Go Irie[3], Kiyoharu Aizawa[1]

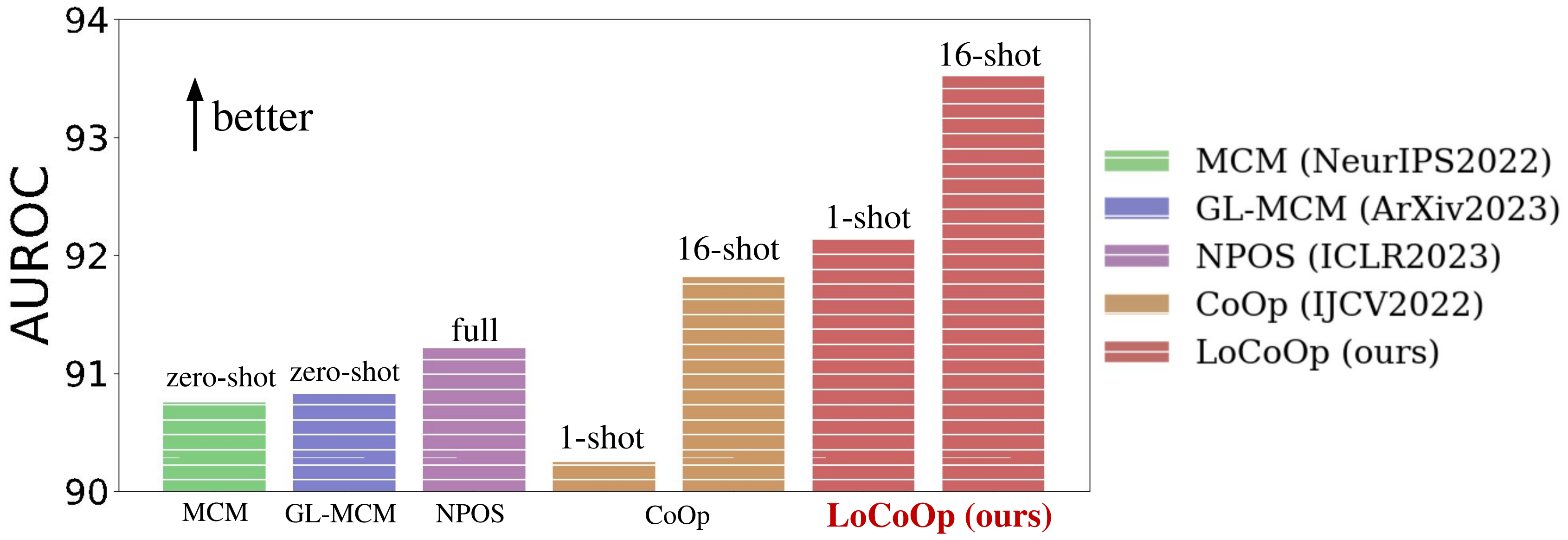
[1] The University of Tokyo [2] LY Corporation [3] Tokyo University of Science




Our findings

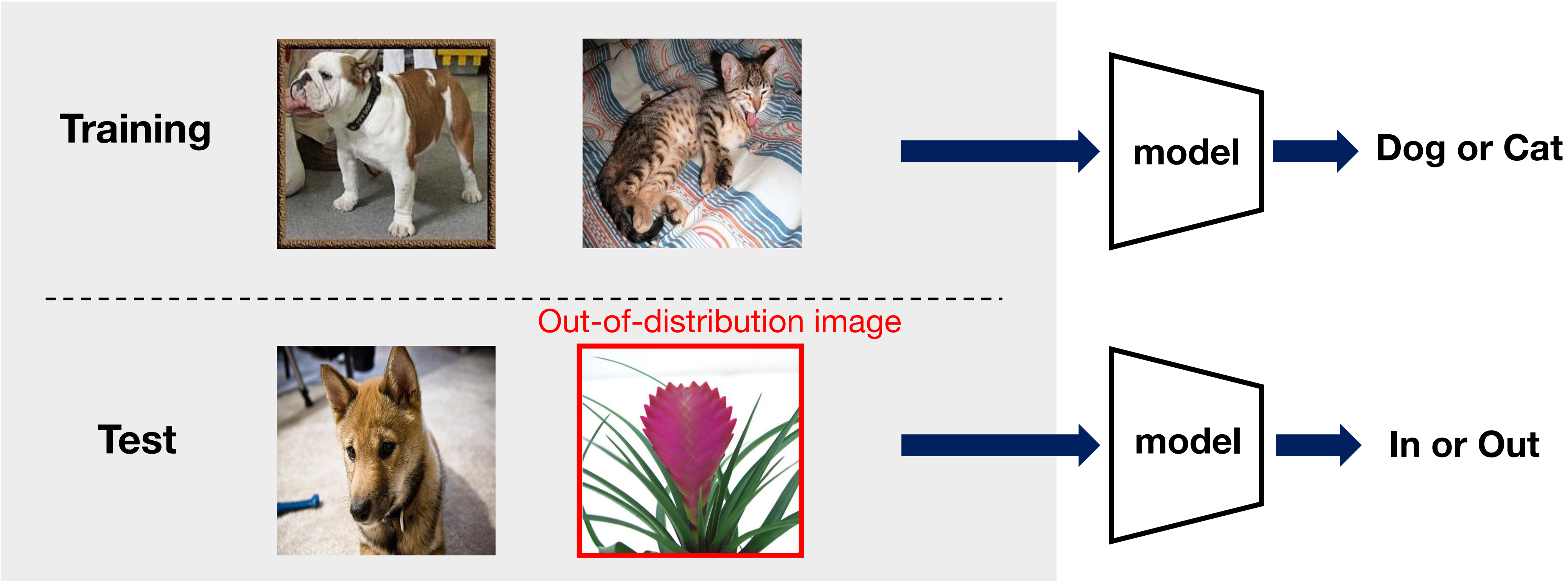
- We propose LoCoOp (**Lo**cal regularized **Co**ntext **Op**timization).
  - LoCoOp is the first prompt learning approach for OOD detection.

LoCoOp outperforms prior CLIP-based methods.
Even with 1-shot, LoCoOp outperforms them.



Out-of-distribution (OOD) detection

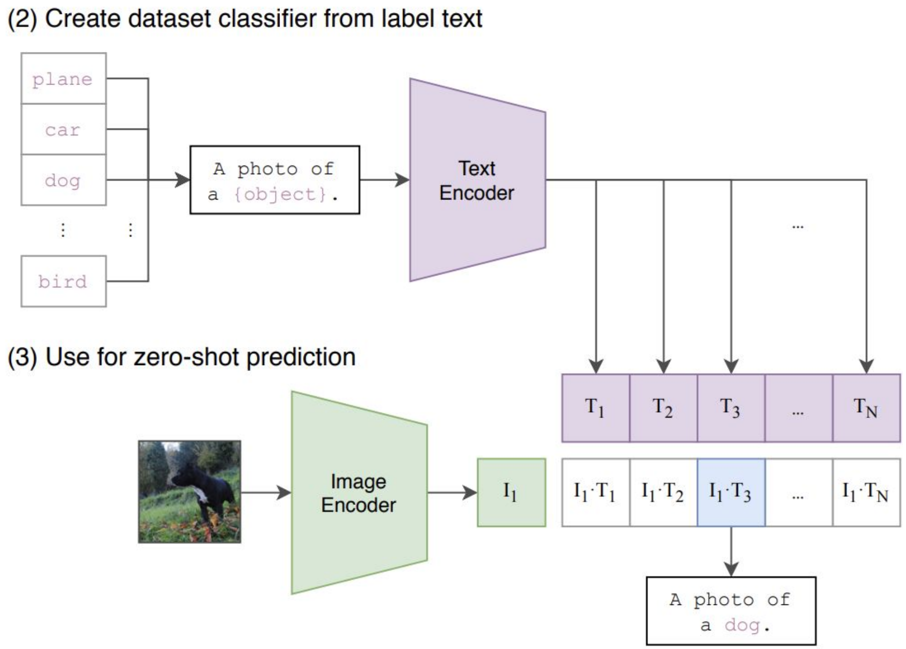
- Detect samples in **out-of-distribution** at test-time



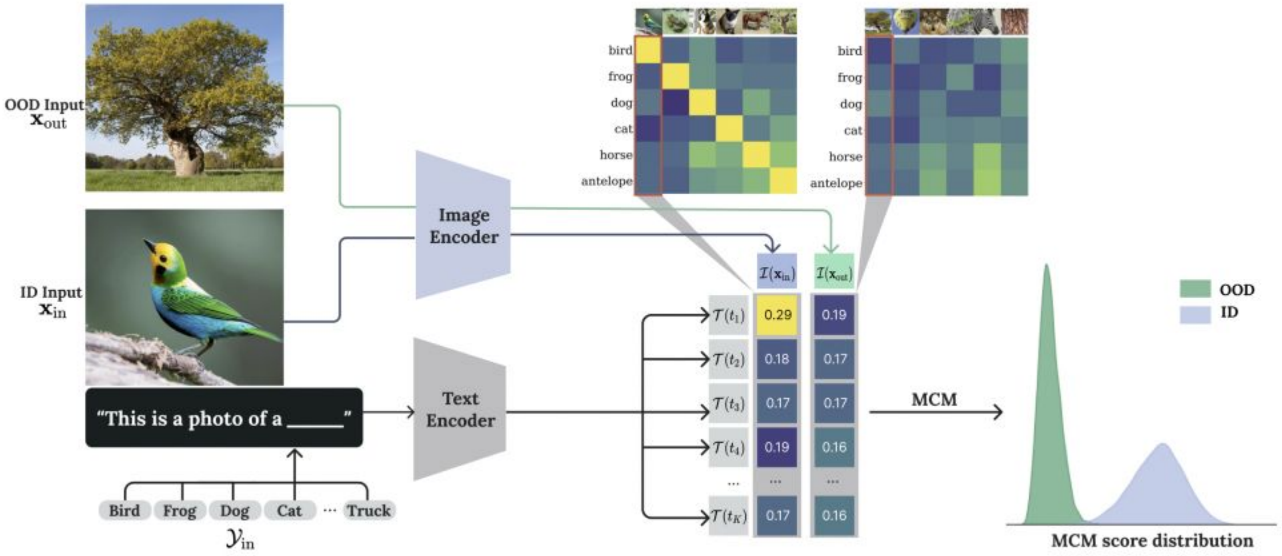
Background

- The application of CLIP [Radford+, ICML2021] to OOD detection attracts attention.

CLIP [Radford+, ICML2021],



MCM [Ming+, NeurIPS2022],



CLIP for OOD detection

Motivation

- Previous CLIP-based methods assumed the extreme settings.
  - Zero-shot OOD detection method (e.g., MCM [Ming+, NeurIPS2022])
    - Pros: No training cost
    - Cons: Domain Gap from ID data

    → **Suboptimal performance**
  - Fully supervised OOD detection method (e.g., NPOS [Tao+, ICLR2023])
    - Pros: No domain gap from ID data
    - Cons: Large training cost, Destroy the CLIP's representations

    → **Suboptimal performance**

We aim to develop an **efficient** and **effective** tuning approach !

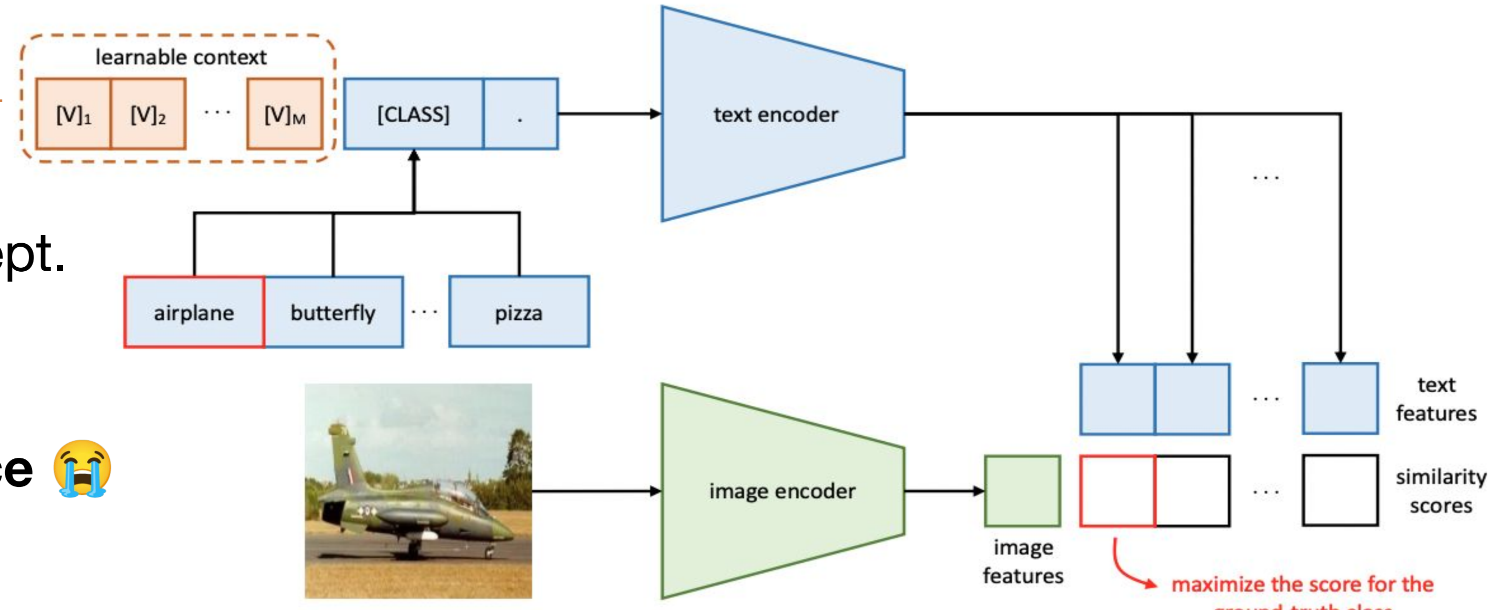
Baseline

- CoOp [Zhou+, IJCV2022]
  - Representative work for tuning CLIP via prompt learning.

These embeddings might include “**tree**” (OOD) concept.

⬇

It degrades the performance 😭

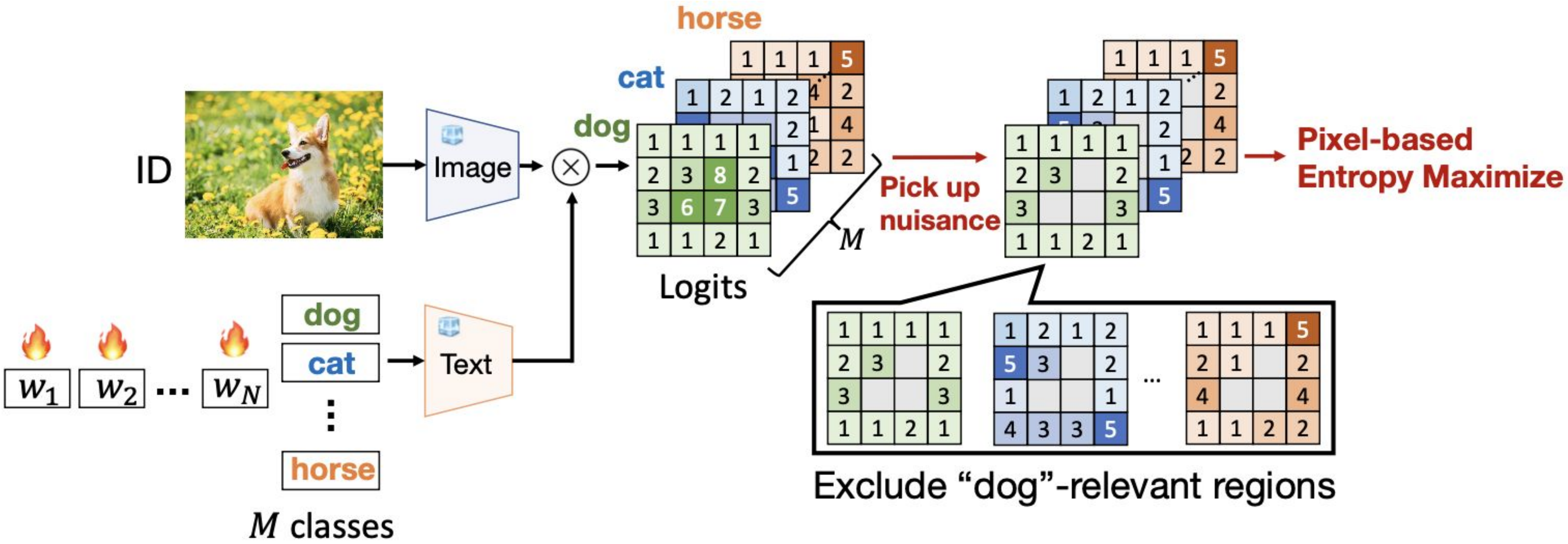


ID: Airplane

Limited performance in OOD detection due to **the potential presence of ID-irrelevant information** in text embeddings.

Proposed method

- We propose LoCoOp (**Lo**cal regularized **Co**ntext **Op**timization)
  - LoCoOp utilizes the portions of local features as OOD during training.
    - **It can remove the ID-irrelevant features in the ID class text embeddings and enhance the separation between ID and OOD.**



Main results on ImageNet OOD benchmarks

- LoCoOp outperforms other methods with very few training data.

	iNaturalist		SUN		Places		Texture		**Average**	
Method	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	AUROC↑	FPR95↓	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 [32]*	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 [49]†	32.19	93.16	54.01	87.19	60.67	83.75	53.94	87.18	50.20	87.82
KNN [44]†	29.17	94.52	35.62	92.67	39.61	91.02	64.35	85.67	42.19	90.97
NPOS_{MCM} [45]†	16.58	96.19	43.77	90.44	45.27	89.44	46.12	88.80	37.93	91.22
NPOS_{MCM} [45]*	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
$\text{LoCoOp}_{\text{MCM}}$ (ours)	38.49	92.49	33.27	93.67	39.23	91.07	49.25	89.13	40.17	91.53
$\text{LoCoOp}_{\text{GL}}$ (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
$\text{LoCoOp}_{\text{MCM}}$ (ours)	23.06	95.45	32.70	93.35	39.92	90.64	40.23	**91.32**	33.98	92.69
$\text{LoCoOp}_{\text{GL}}$ (ours)	16.05	**96.86**	**23.44**	**95.07**	**32.87**	91.98	42.28	90.19	**28.66**	**93.52**

Conclusion

- We first tackle CLIP-based few-shot OOD detection.
- We propose a novel prompt learning approach called LoCoOp.
  - LoCoOp leverages the portions of CLIP local features as OOD features for OOD regularization.
- LoCoOp brings substantial improvements over existing methods on the large-scale ImageNet OOD benchmarks.

LoCoOp-arXiv

code