

Learning Transferable Reward for Query Object Localization with Policy Adaptation



Tingfeng Li



Shaobo Han



Martin Renqiang Min



Dimitris N. Metaxas

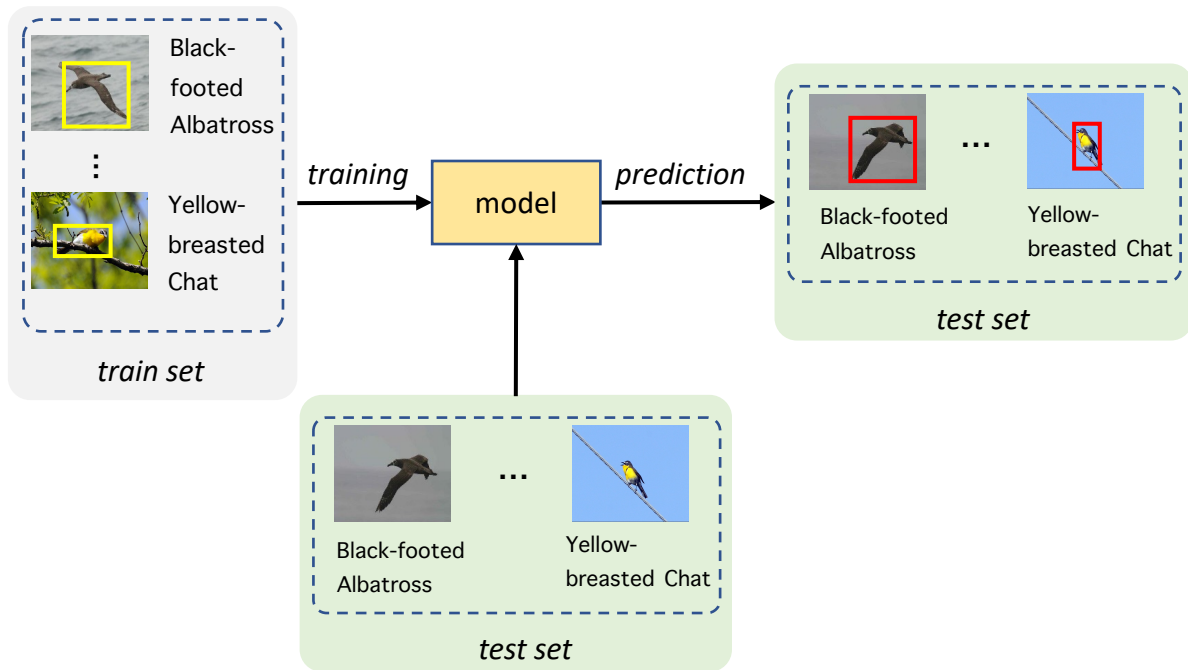


RUTGERS

NEC

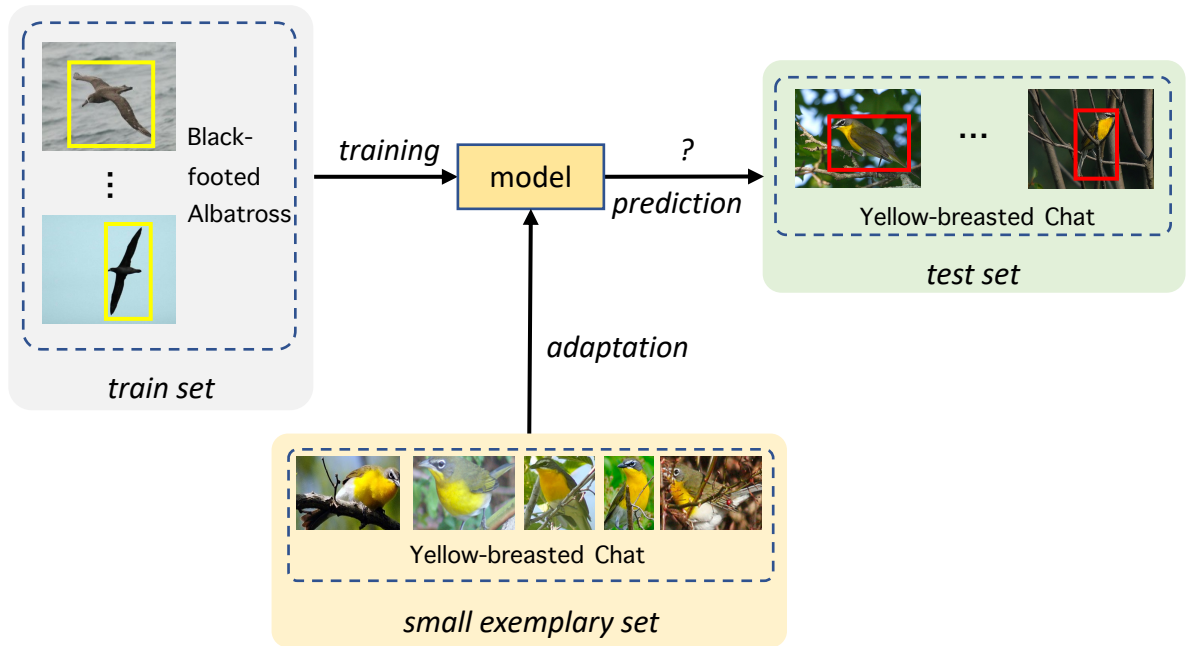
NEC Laboratories America

Fully Supervised Object Localization



- class-agnostic
- many classes of labeled data with box annotations for training
- Not transferable, i.e., not easy to add a new class in test time

Query Object Localization

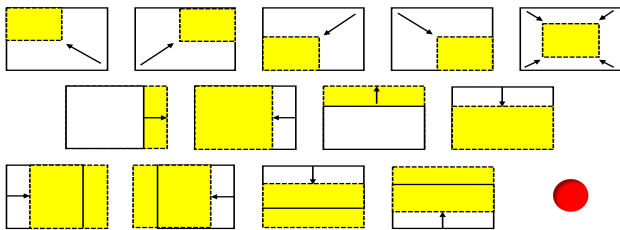


- class-specific, suitable for aerial imagery, robotic manipulation,...
- one class for training
- explicit transferable

Localization by Reinforcement Learning

- States: RoI feature + internal state within RNN

- Actions



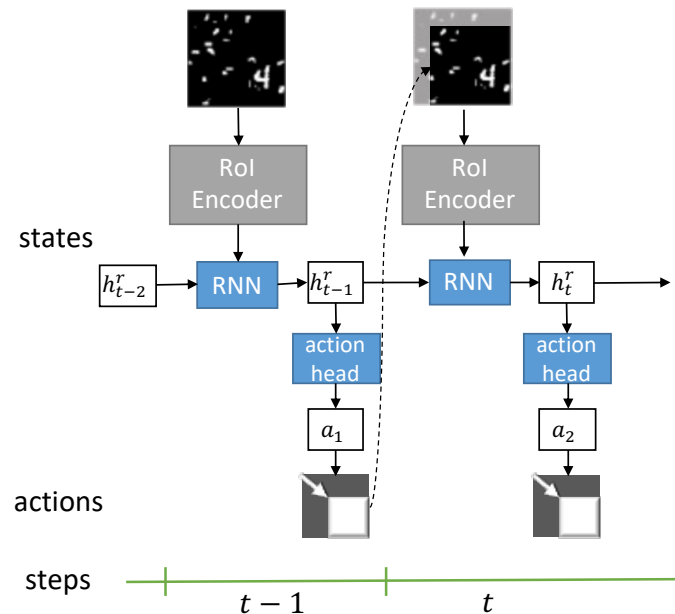
- Optimization

- Policy gradient: $Loss_{policy} = -\log p * reward$

- Reward

$$R = ||\mathbf{b}_{t-1} - \mathbf{c}|| - ||\mathbf{b}_t - \mathbf{c}||$$

Prototype embedding

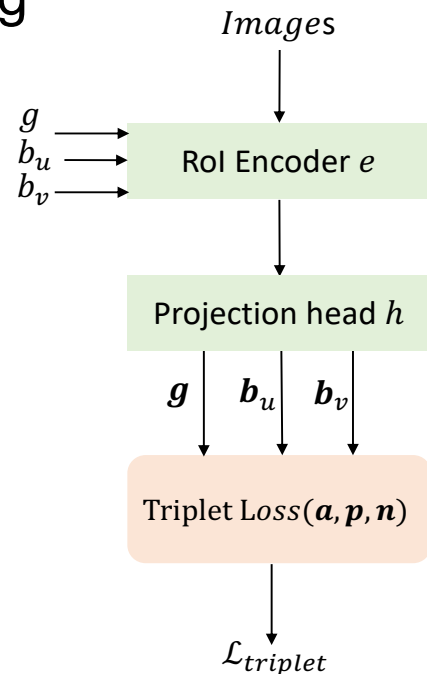


Learn Ordinal Reward

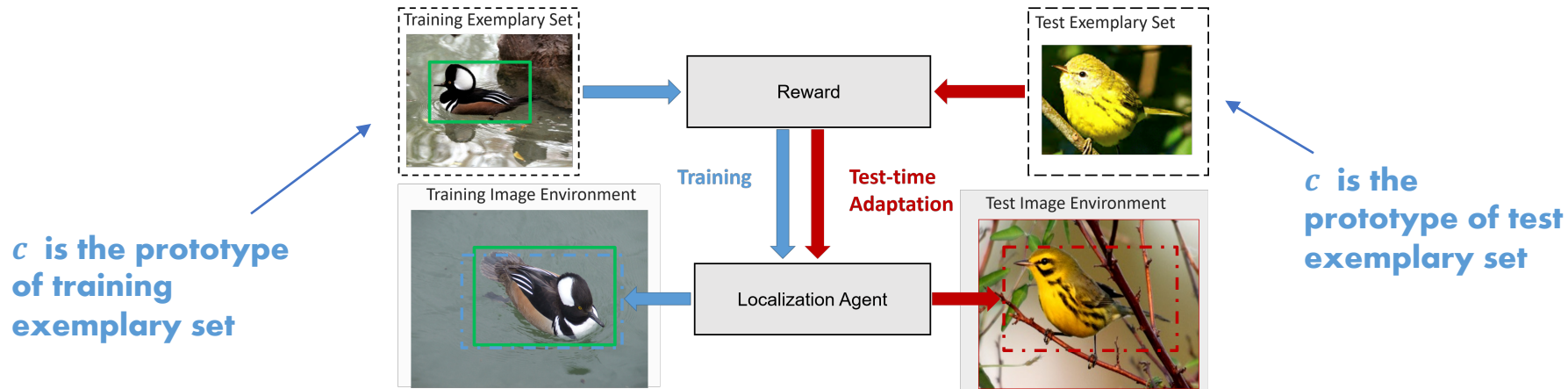
- Ordinal property is not naturally existing in off-the-shelf backbones

- ✗ CLIP pre-trained ViT
- ✗ ImageNet pre-trained VGG16
- ✗ Faster RCNN pre-trained VGG16
- ✓ Triplet loss trained backbones

- Training



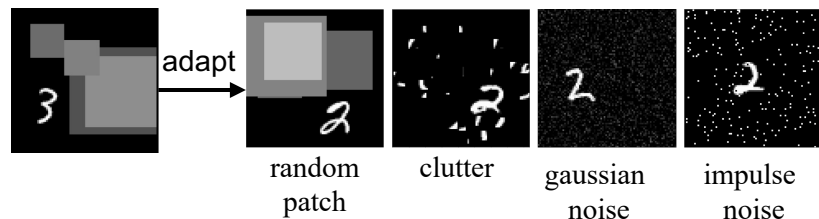
Test-time Policy Adaptation



$$R = ||\mathbf{b}_{t-1} - \mathbf{c}|| - ||\mathbf{b}_t - \mathbf{c}||$$

Experiments – New Classes/Background

adapt	random patch	clutter	impulse noise	gaussian noise	mean
	97.6 $\pm$ 0.4	39.6 $\pm$ 0.5	22.1 $\pm$ 0.7	66.2 $\pm$ 2.0	56.4
✓	100.0$\pm$0.0	97.4$\pm$0.3	99.9$\pm$0.1	100.0$\pm$0.0	99.3



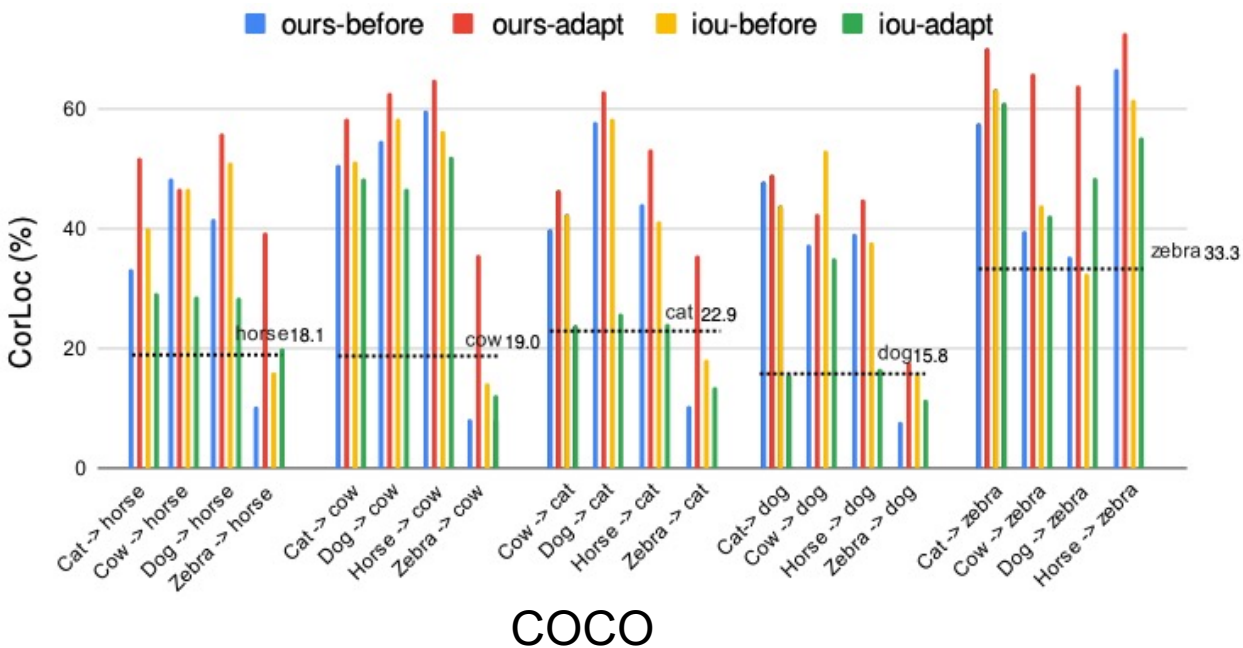
MNIST



CUB

	adapt	warbler (new)	wren	sparrow	oriole	kingfisher	vireo	gull	mean
DDT		73.8	78.6	71.2	74.5	78.0	69.2	93.3	76.9
ours		85.5 $\pm$ 1.1	82.9 $\pm$ 2.6	81.3 $\pm$ 3.7	77.9 $\pm$ 0.7	78.9 $\pm$ 0.6	82.2 $\pm$ 4.6	86.3 $\pm$ 3.5	82.1
	✓	89.7$\pm$1.1	91.0$\pm$1.1	89.3$\pm$1.5	85.0$\pm$0.8	85.9$\pm$4.4	90.0$\pm$0.9	93.9$\pm$0.5	89.3

Experiments – Ordinal vs. IoU Reward



	training reward	adaptation reward
our method	ordinal embedding	ordinal embedding
iou-based method	IoU	ordinal embedding

- state representation learning
- continuous vs. discrete reward
- adaptable vs. not adaptable reward

Our agent trained with ordinal reward can adapt to new classes.

Experiments – Adaptation vs. Fine-tuning

	before finetune	before adapt	after finetune	after adapt
	Faster rcnn	ours	Faster rcnn	ours
Cat -> horse	20.93	33.32	37.73	51.89
Cow -> horse	54.79	48.41	68.04	46.80
Dog -> horse	38.52	41.50	58.01	55.89
Zebra -> horse	1.12	10.29	6.04	39.22
Cat -> cow	40.52	50.85	58.55	58.58
...	...	...	...	...
Cat -> zebra	10.64	57.58	37.97	70.28
Cow -> zebra	4.42	39.64	19.64	65.80
Dog -> zebra	2.29	35.27	15.88	63.91
Horse -> zebra	7.86	66.82	29.30	72.83
mean	30.32 $\pm$ 25.0	39.51$\pm$17.9	41.17 $\pm$ 25.3	52.04$\pm$13.9

Test-time adaptation beats fine-tuning on target domains

Experiments – Compare to Few-shot Detectors

		target	cat	dog	cow	horse	mean
multi-way few-shot	{	FRCN+fc-full	13.1	3.1	3.7	7.4	6.8
		TFA w/ fc	29.1	13.0	5.0	10.7	14.4
		TFA w/cos	28.0	10.3	4.5	8.9	12.9
one-way multi-shot <i>explicit</i> adaptation	→	ours-before	23.0	20.6	24.5	21.2	22.3
		ours-adapt	40.3	33.5	43.1	40.2	39.3

COCO Dataset

One-way multi-shot + explicit adaptation performs better

Thanks for watching!



[paper](#)



[code](#)