

INTRODUCTION

- **Goal:** Boosting the performance of weakly-supervised object detectors (WSODs) with a *few carefully selected fully-annotated images*.
- **Motivations:**
 - ▷ WSODs require **only image tags** annotation for training.
 - ▷ But achieve lower performances than fully-supervised object detectors.
 - ▷ We want to **narrow the gap** between weakly- and fully-supervised object detectors.
 - ▷ WSODs suffer some **well-known confusions**. Addressing them will make the detectors more effective.

CONTRIBUTIONS

- We introduce a new approach to object detection that **combines weakly-supervised and active learning**.
- We introduce **BiB**, an active selection strategy that is **tailored** to address the limitations of weakly-supervised object detectors.
- **BiB** demonstrates a better **detection performance/annotation cost trade-off** than both weakly- and fully-supervised object detection.

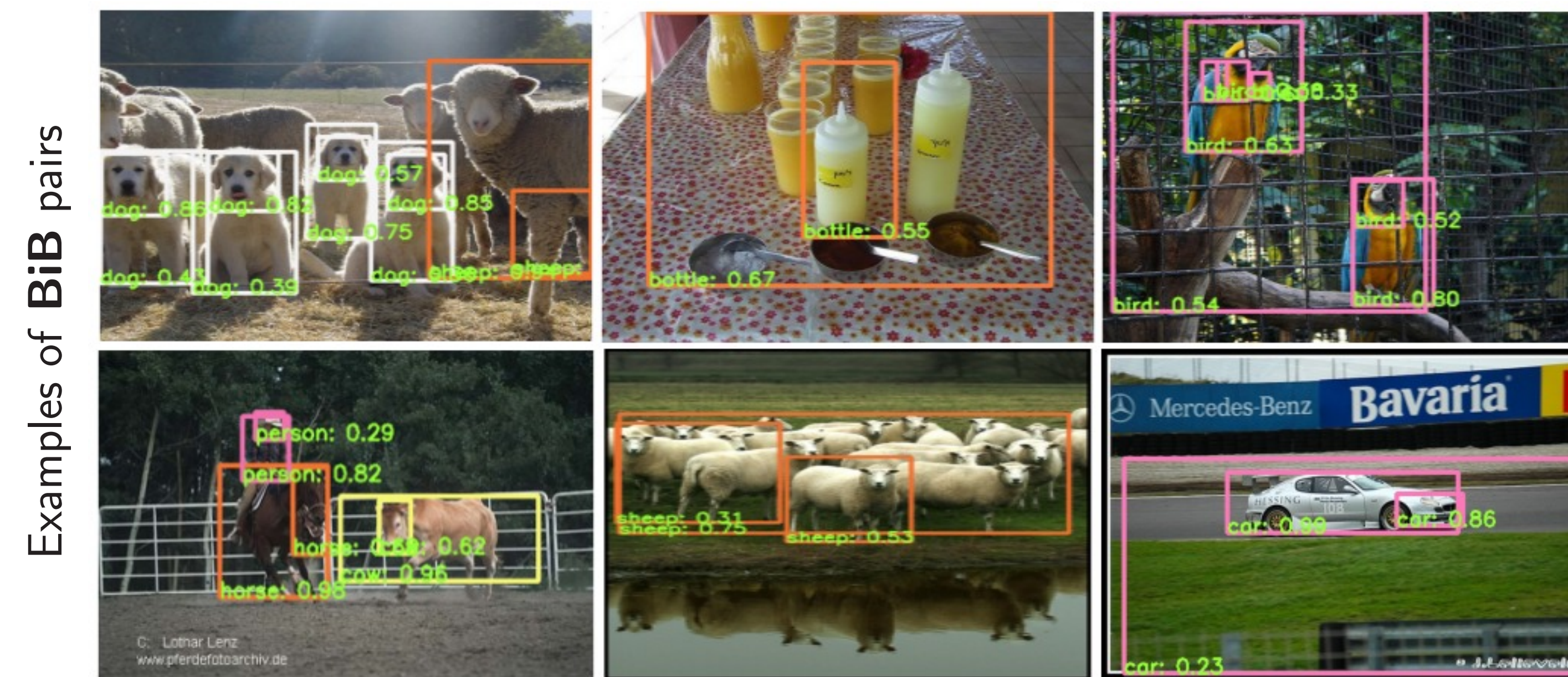
DISCOVER PROBABLE WSODs MISTAKES

Typical confusions of WSODs:

predictions focusing only on *discriminative object parts* or *grouping instances* of objects.

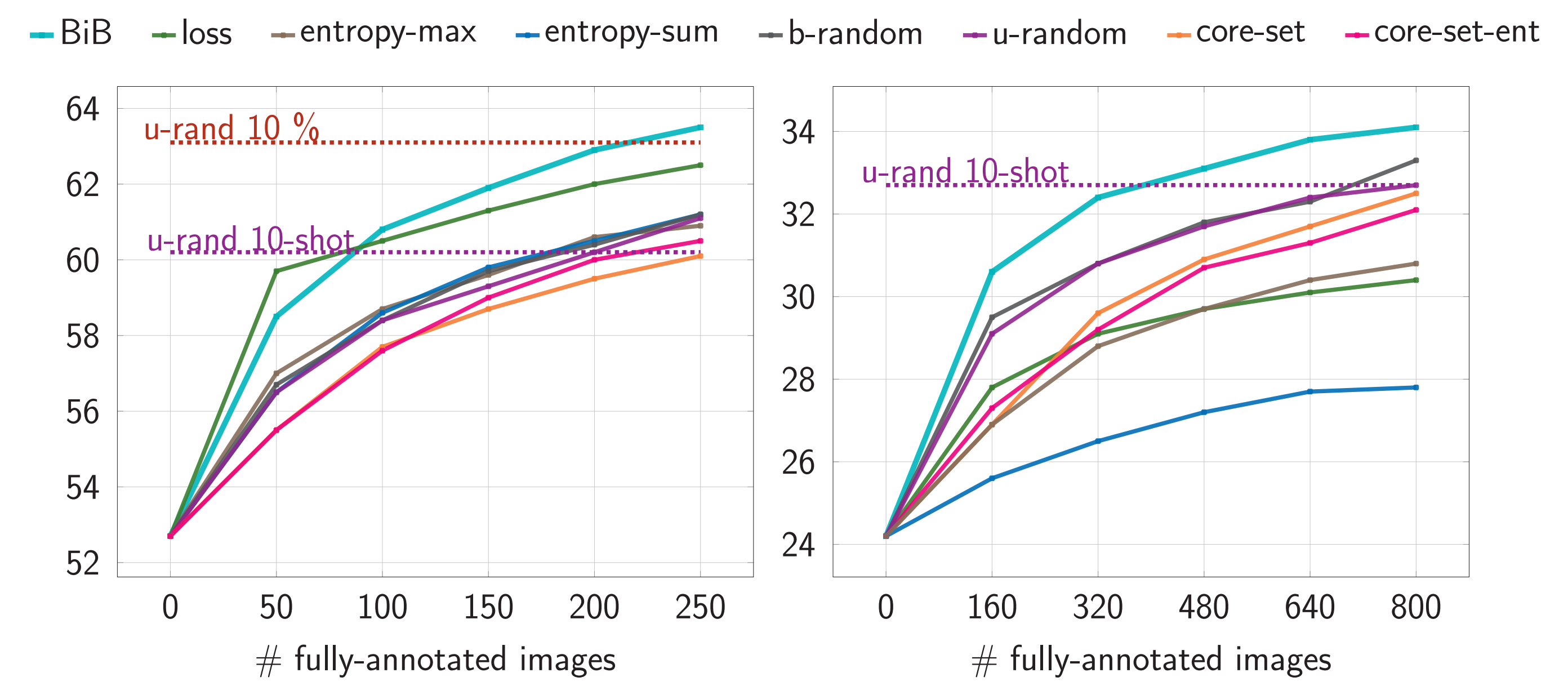
Box-in-box (BiB) pairs of regions:

pairs of predictions of the *same class* s.t. one is “contained” in the other.



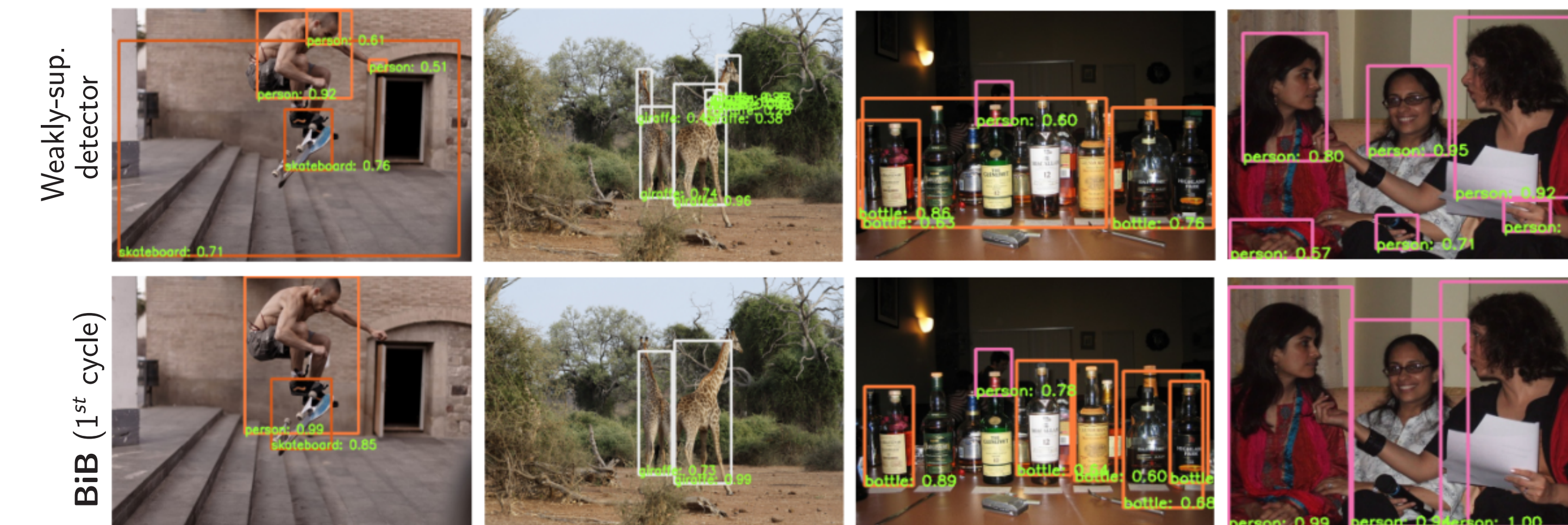
EXPERIMENTAL RESULTS

- **Datasets:** COCO2014 [47], VOC07 [24],
- **Evaluation Metrics:** Average precision (AP50 and AP).
- **Comparison of active learning strategies**



Results (AP50) on VOC07 (left) and COCO (right) dataset

- **Examples of improved detections:**



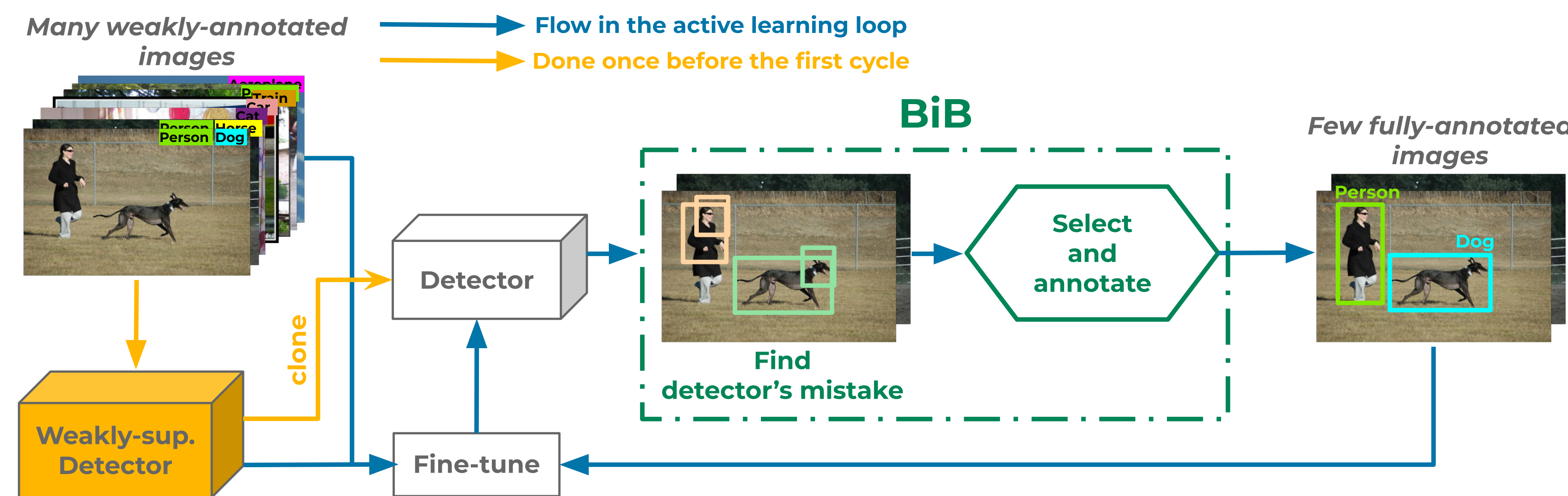
- **Comparison to the state of the art:**

Setting	Method	VOC07		COCO	
		AP50	AP50	AP	AP
Fully supervised	Fast RCNN [32]	66.9	38.6	18.9	
	Faster RCNN [54]	69.9	41.5	21.2	
WSOD	WSDN [7]	34.8	-	-	
	OICR [69]	41.2	-	-	
	C-MIDN [29]	52.6	21.4	9.6	
	WSOD2 [80]	53.6	22.7	10.8	
	MIST-CDB [55]	54.9	24.3	11.4	
	CASD [38]	56.8	26.4	12.8	
Weak & few strong (10-shot)	BCNet [49]	57.1	-	-	
	OAM [6]	59.7	31.2	14.9	
	Ours (u-rand)	60.2	32.7	16.4	
	Ours (BiB)	62.9	34.1	17.2	

- **Ablation study on VOC07:**

DifS	K selection im. reg. BiB	Number of images annotated				
		50	100	150	200	250
✓		56.3 ± 0.4	58.0 ± 0.5	58.9 ± 0.4	60.0 ± 0.3	60.5 ± 0.4
		56.5 ± 0.4	58.4 ± 0.4	59.3 ± 0.7	60.2 ± 0.4	61.1 ± 0.5
✓	✓	57.1 ± 0.4	58.3 ± 0.5	59.3 ± 0.6	59.8 ± 0.4	60.3 ± 0.4
✓	✓	58.4 ± 0.4	60.2 ± 0.4	61.5 ± 0.6	62.6 ± 0.4	63.4 ± 0.3
✓	✓	57.9 ± 0.7	60.1 ± 0.4	61.2 ± 0.5	62.1 ± 0.5	62.6 ± 0.4
		58.5 ± 0.8	60.8 ± 0.5	61.9 ± 0.4	62.9 ± 0.5	63.5 ± 0.4

OVERALL APPROACH



Active learning pipeline:

- Train a weakly-supervised object detector.
- // *Active learning loop*
- Repeat
 - ▷ Select images to fully label.
 - ▷ Ask human annotators to draw bounding boxes around objects in them.
 - ▷ Fine-tune the weakly-supervised object detector with all annotations.

BiB selection:

- Find BiB pairs in all images.
- // *kmeans++ initialization*
- Repeat until enough images are selected:
 - ▷ Compute the distance between BiB pairs in selected images and those in other images.
 - ▷ Pick a BiB pair with probability proportional to its distance to the pairs in selected images.
 - ▷ Select the image containing the chosen pair.

