

---

## MULTIBOX SAMPLE SELECTION FOR ACTIVE OBJECT DETECTION

---

*Jiaxiang Dong, Li Zhang\**



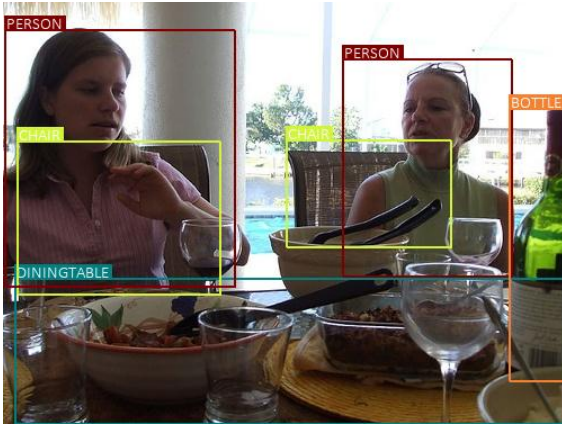
*Jiaxiang Dong*



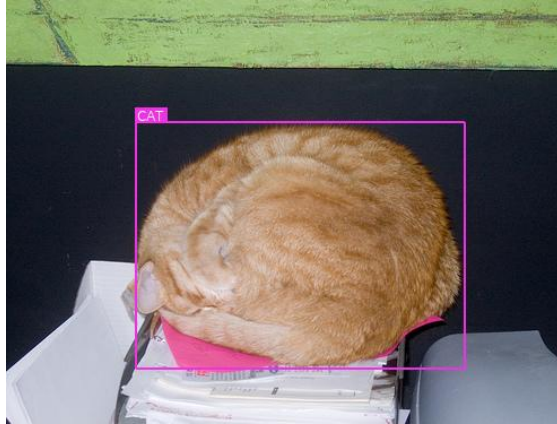
*Li Zhang*

# Object Detection

*Need to*  
- localization  
- classification



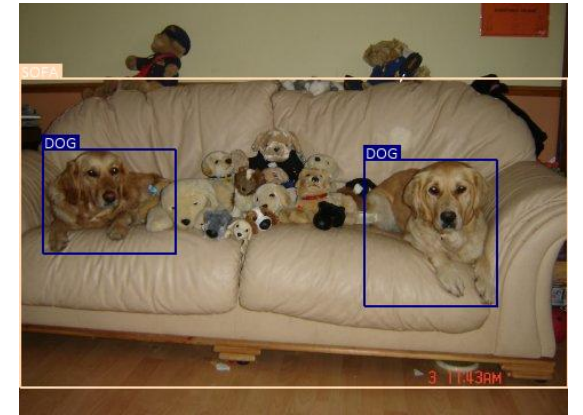
chair: 0.6



cat: 0.55



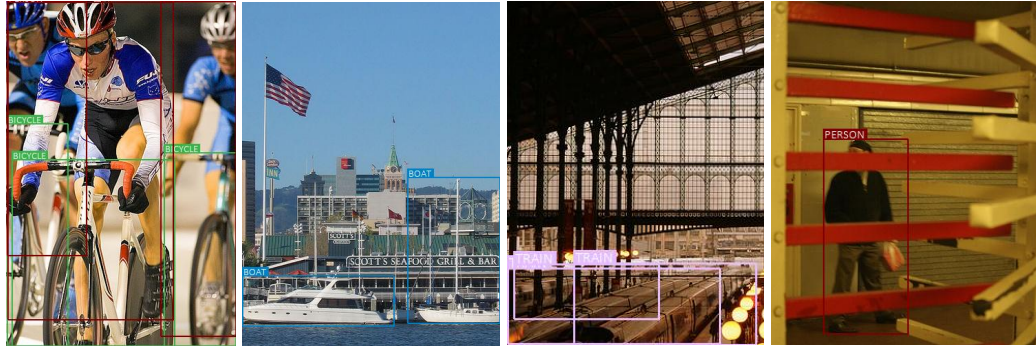
rooster: 0.6



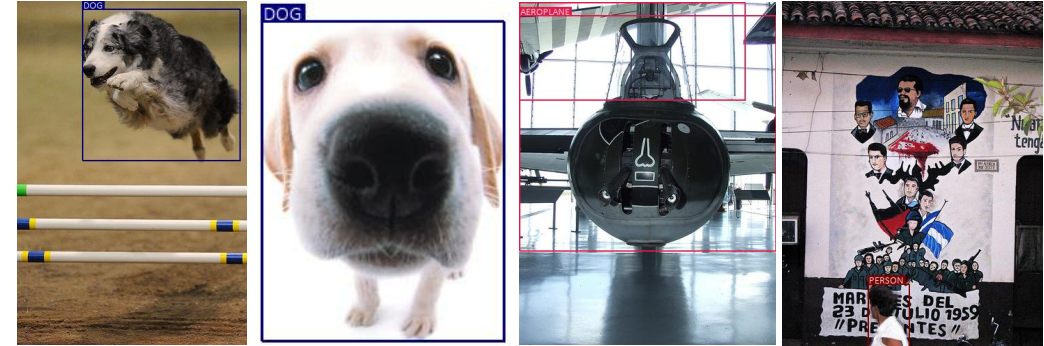
dog: 0.7

**Object detection aims to locate objects in images and classify them into correct categories.**

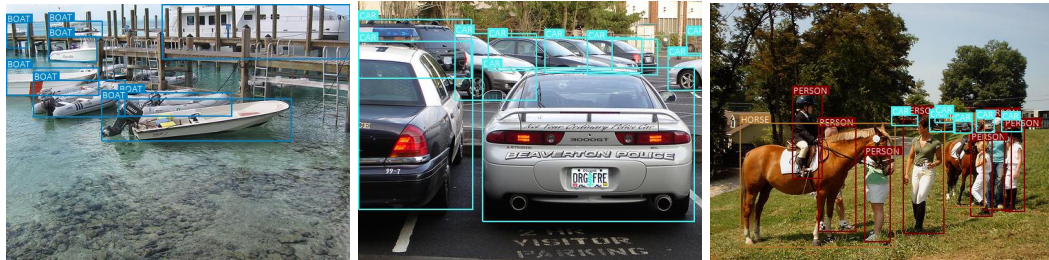
# Deep Object Detection



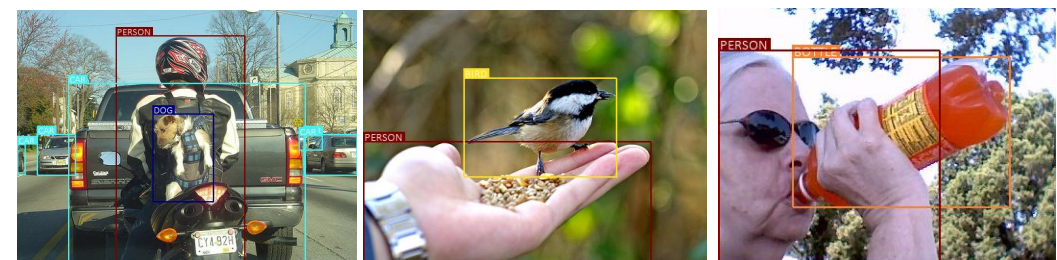
complex scenarios



uncertain objects



a large amount of objects

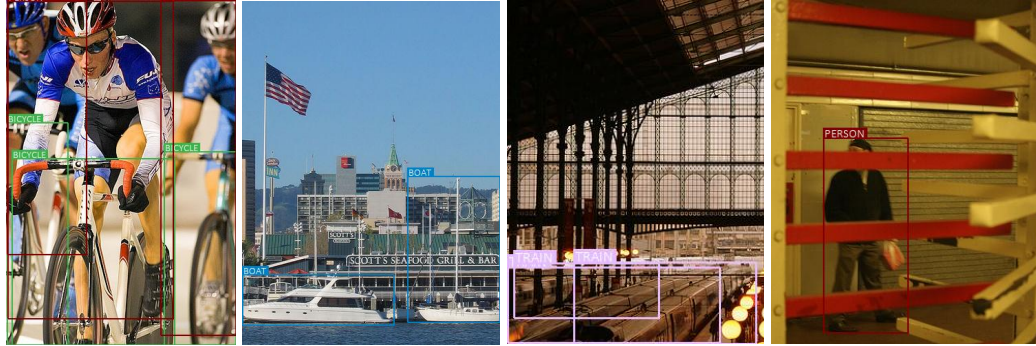


complex objects

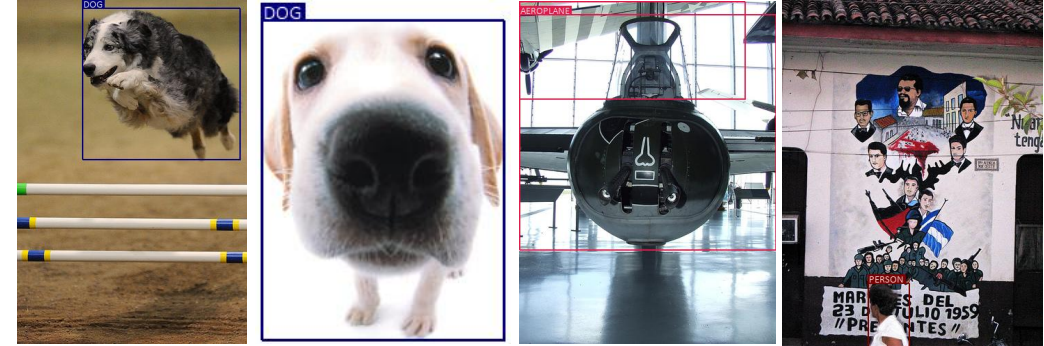
- The advanced performance relies on the large-scale annotated datasets.
- The strong annotation usually costs enormous labor work.



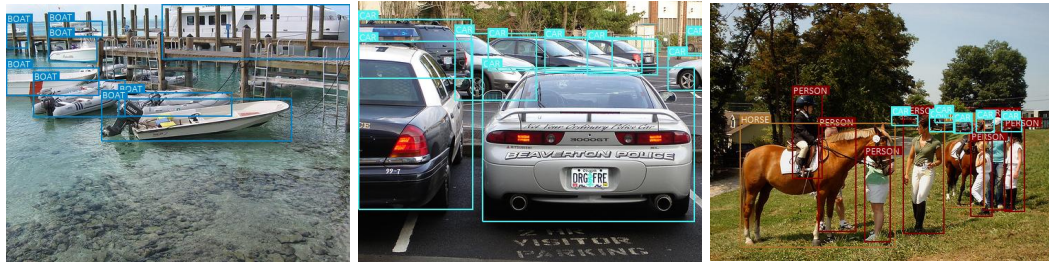
# Deep Object Detection



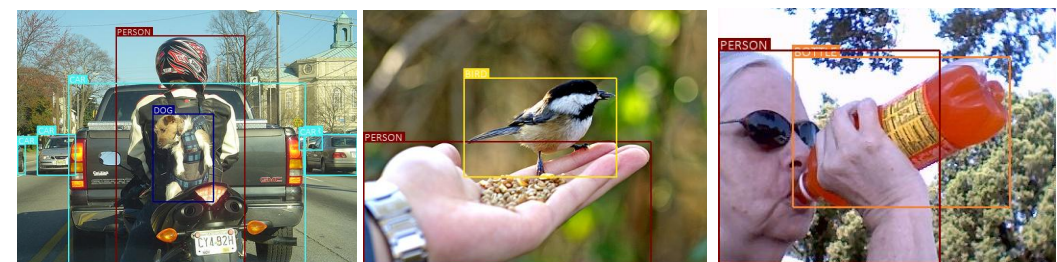
complex scenarios



uncertain objects



a large amount of objects



complex objects

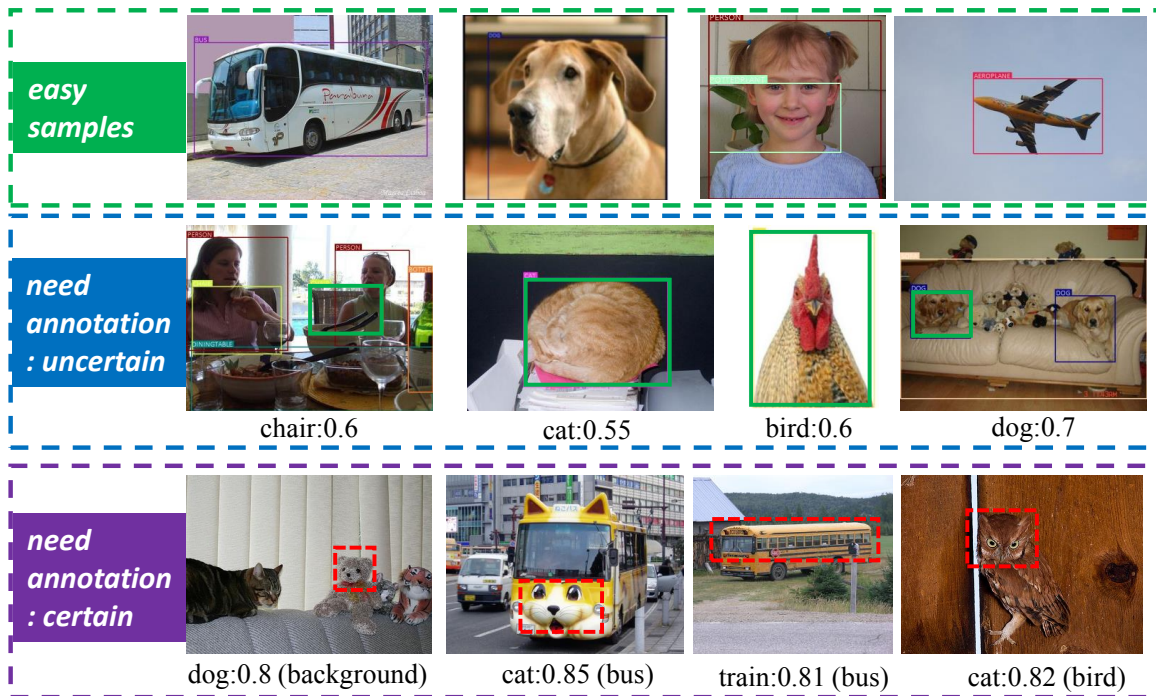
**DOD** ➡

**How to achieve the state-of-the-art performance with less supervision?**



# Active Object Detection

Active object detection (AOD) employs *a query selection criterion* to select the most informative samples to annotate, which maximally *boosts the performance with limited labeled data*.



→ *The main object in the image and occupy a large portion of the image.*

→ *Mostly non-primary objects or hard to recognize.*

→ *Outliers, which can create noise to the model. They are mostly extremely small objects.*

# Challenges in AOD

Different from well-studied active learning for image classification:

- **Firstly** it requires multiple box predictions on each image instead of a single prediction.
- **Secondly** outliers exist in the detection datasets such as extremely small objects or not clear images. Outliers not only mislead the model training but also waste the labeling budget.



# Motivation

Current AOD works typically regard each *final box prediction as a single prediction* and apply active learning criterion to it.

## **Shortcoming**

Fail to consider the multiple boxes in an entirety.

Suffer from overwhelming uninformative boxes

Miss some informative boxes.

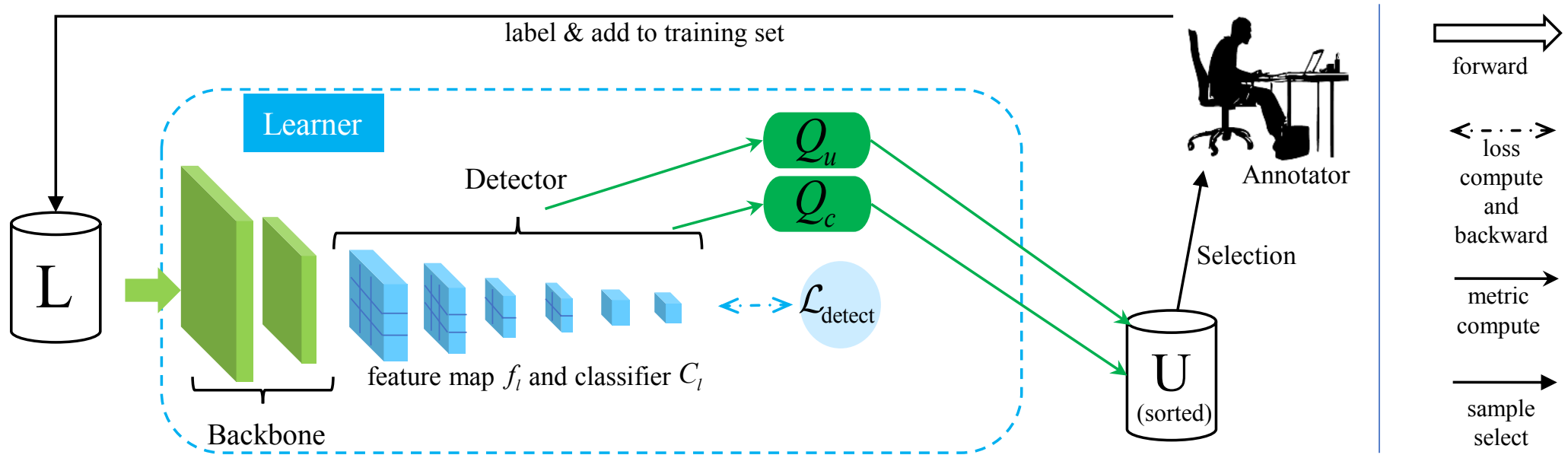
# Contribution

We divide sub-optimally predicted images into uncertainly and certainly predicted images and propose a Multi-Box Sample Selection (MSS) criterion.

- *The MultiBox Sample Selection criterion (MSS)* is composed of *Multi-Box Uncertainty (MBU)* and *Multi-Box Committee (MBC)* to tackle sub-optimally predicted images.
- For MBU, we analyze the optimal MultiBox prediction and assess *the uncertainty* based on the box that best reflects the uncertainty of the whole image.
- For MBC, we find *that certain but incorrect predictions* are usually inconsistent among nearby boxes, so we leverage the anchor architecture of the detection network to form the committee.



# Architecture



## Feature map

## Convolution

## Bounding boxes predictions

## Aggregating

<p>Image <math>x</math> is input into a truncated base network, which can be any convolutional backbone network.</p> <p><b>1</b></p>	<p>Then <math>F(x)</math> is input to a series of convolution layers <math>G_1, G_2, \dots, G_N</math> to generate multi-scale feature maps <math>f_1; \dots; f_N</math>.</p> <p><b>2</b></p>	<p>Each cell as an anchor and design <math>M</math> bounding boxes of varied size around it. Each box prediction corresponds to a tuple <math>(l_i^{cx}, l_i^{cy}, l_i^w, l_i^h, c_i)</math>.</p> <p><b>3</b></p>	<p>Use <math>C_1, C_2, \dots, C_M</math> to denote the classifier for each predicted box.</p> <p><b>4</b></p>
--	---	---	---

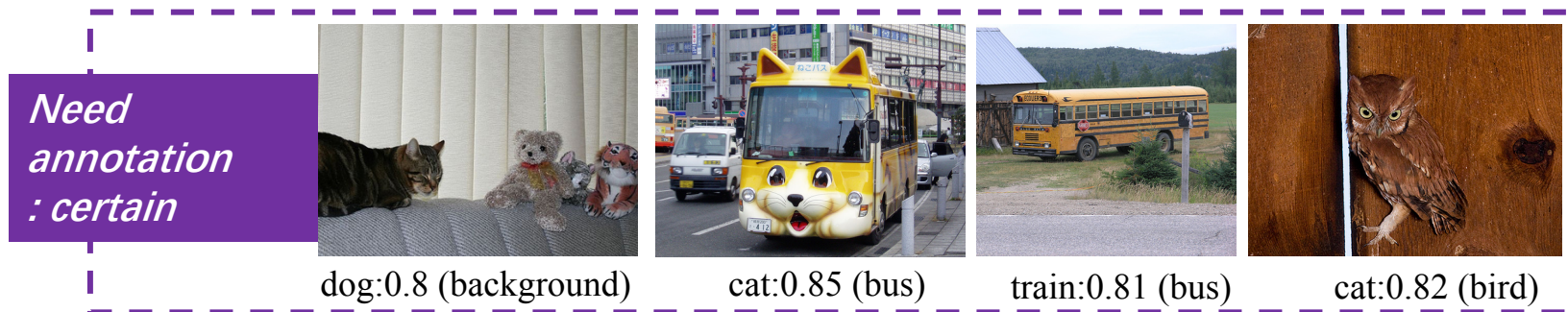
# Multi-Box Sample Selection

We propose a **Multi-Box Sample Selection (MSS)** criterion. We prioritize images that are not predicted well by the detector since such images can improve the model more. We divide these sub-optimally predicted images into two categories: *Uncertain Images* and *Certain Images*.



- Uncertain sub-optimally predicted images

***Low prediction confidence!***



- Certain sub-optimally predicted images

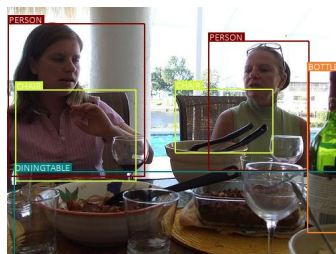
***High prediction confidence, but the prediction is incorrect!***



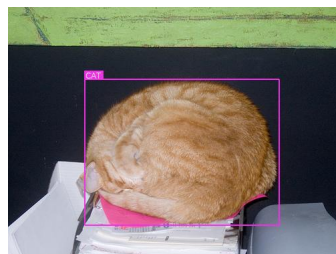
# Multi-Box Sample Selection (MSS)

*: Uncertain*

*Need  
Annotation*



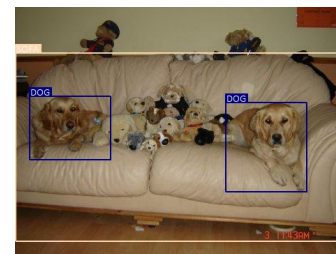
chair:0.6



cat:0.55



bird:0.6



dog:0.7

**Multi-Box Uncertainty (MBU)**

*: Certain, but incorrect*

*Need  
Annotation*



dog:0.8 (background)



cat:0.85 (bus)



train:0.81 (bus)



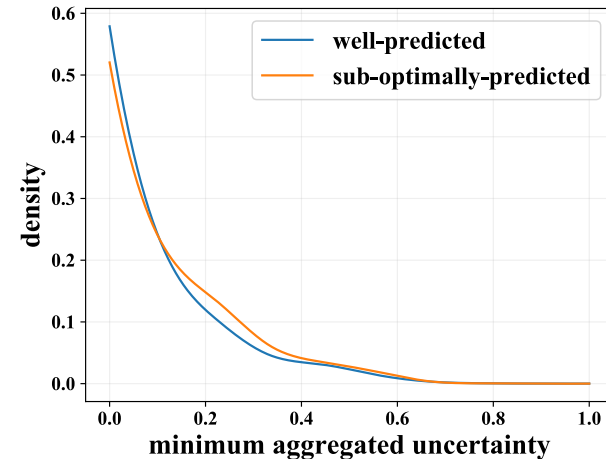
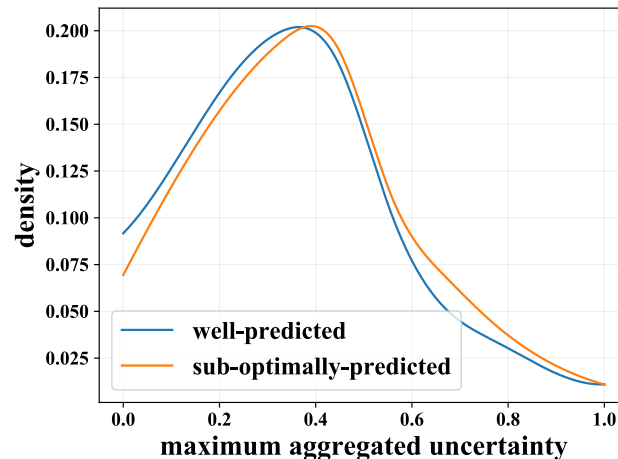
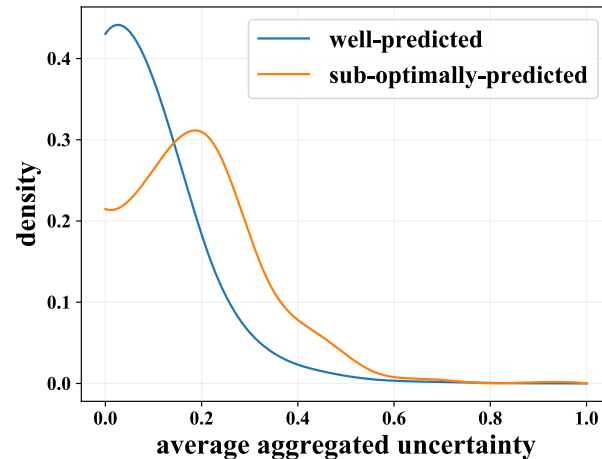
cat:0.82 (bird)

**Multi-Box Sample  
Selection (MSS)**

**Multi-Box Committee (MBC)**

# Multi-Box Uncertainty (MBU)

Prior active learning methods usually focus on *uncertainty based on single image prediction, which can not solve the MultiBox prediction*. An easy extension is to aggregate the uncertainty of each box by some aggregation function such as 'average', 'maximum' or 'minimum'.



**Simple aggregation of single prediction uncertainty can not select sub-optimally predicted images well.**



# Multi-Box Uncertainty (MBU)

- A novel uncertainty measurement for MultiBox predictions

- 1、** If one foreground class  $k$  appears in the image, for a certain and correct MultiBox prediction, there at least exists one box  $l_i$  having high probability  $c_{ik}$ , which means that **the highest  $k$ -th class.** probability among all the predicted boxes should be extremely high, **i.e.  $\max_{i \in [1, M]} c_{ik}$  is close to 1.**
- 2、** If the class does not exist in the image, all boxes should have low probability on this foreground class, **i.e.  $\max_{i \in [1, M]} c_{ik}$  is close to 0.**

**In both cases, if  $\max_{i \in [1, M]} c_{ik}$  is far from both 0 and 1, the Multi-Box prediction for class  $k$  is definitely uncertain.**

$$Q_u = \sum_{k=1}^K \min \left( \left| \max_{i \in [1, M]} c_{ik} - 0 \right|, \left| \max_{i \in [1, M]} c_{ik} - 1 \right| \right). \quad (2)$$

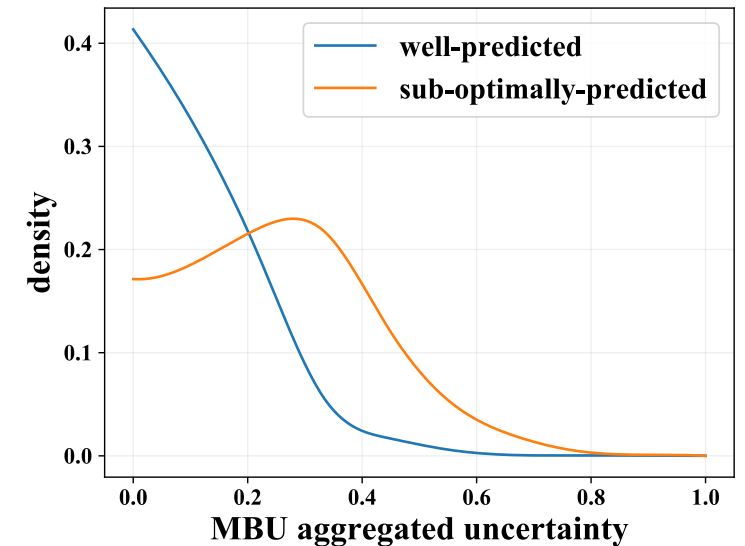
# Multi-Box Uncertainty (MBU)

A novel uncertainty measurement for MultiBox predictions

**Case 1** If one foreground class  $k$  appears in the image, for a certain and correct MultiBox prediction, there at least exists one box  $l_i$  having high probability  $c_{ik}$ , which means that **the highest  $k$ -th class** probability among all the predicted boxes should be extremely high, **i.e.  $\text{Max}_{i \in [1, M]} c_{ik}$  is close to 1.**

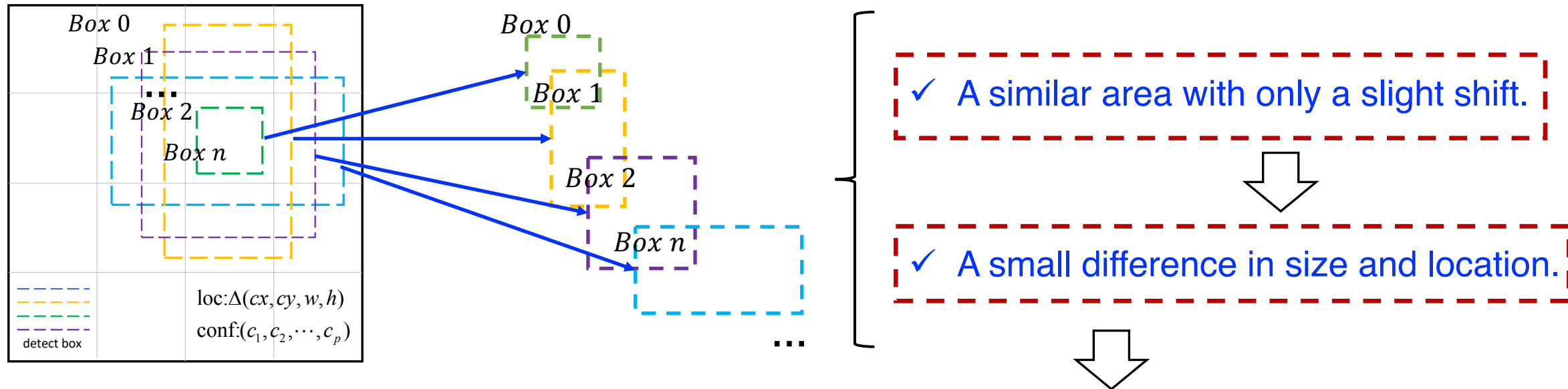
**Case 2** If the class does not exist in the image, all boxes should have low probability on this foreground class, **i.e.  $\text{max}_{i \in [1, M]} c_{ik}$  is close to 0.**

**MBU avoids the shortcoming of simple aggregation of uncertainty and can largely discriminate well- and sub-optimally-predicted samples.**



# Multi-Box Committee (MBC)

*For certain but sub-optimally predicted images*, the prediction itself is incorrect. We design **Multi-Box Committee (MBC)** to tackle these images, which measures the disagreement of multiple detectors.

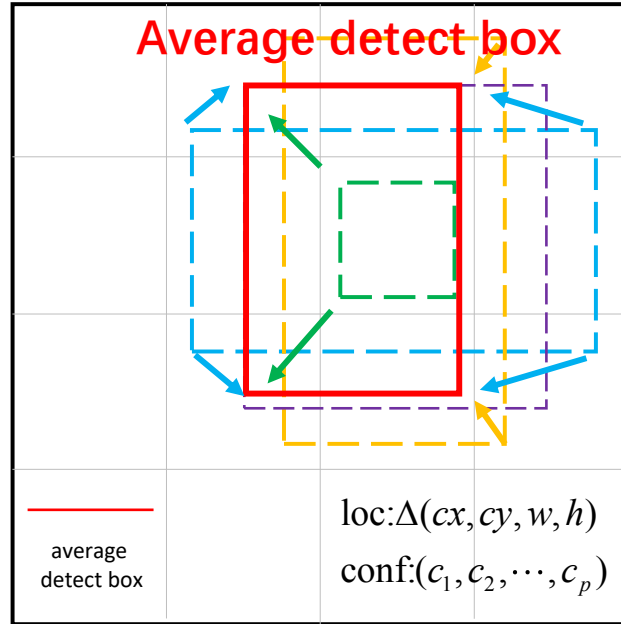


Bounding boxes around  
an anchor forms a committee

**They should give consistent predictions!**



# Multi-Box Committee (MBC)



- *The disagreement of bounding boxes at the same cell* reflects the sub-optimal prediction and indicates that *the sample is valuable to annotate*.
- MBC *constructs committee* as bounding boxes at each anchor and *compute the variance*.

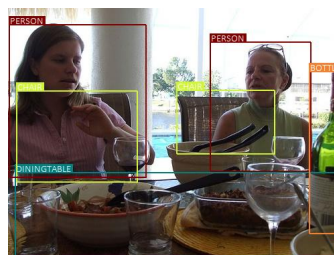
$$Q_c = \left( \max_{\substack{c_i \in l_i \\ l_i \in \mathbf{l}_a}} \left( \sum_{k=1}^{|\mathcal{C}|} c_{ik} \right) \right) \left( \frac{1}{M_a} \sum_{l_i \in \mathbf{l}_a} \left\| l_i - \frac{1}{M_a} \sum_{l_i \in \mathbf{l}_a} l_i \right\|_2 \right). \quad (3)$$

# Multi-Box Sample Selection (MSS)

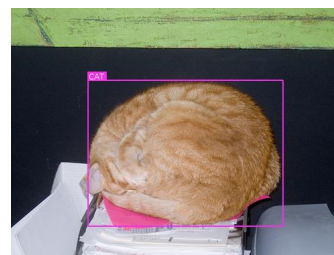
**Sub-optimally  
Prediction**

*: Uncertain*

*Need  
Annotation*



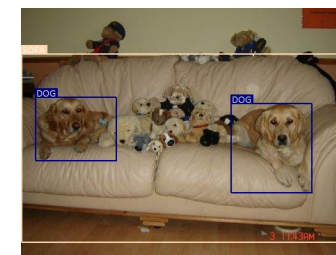
chair:0.6



cat:0.55



bird:0.6

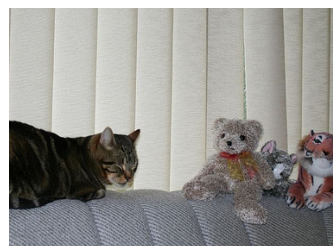


dog:0.7

**Multi-Box Uncertainty (MBU)**

*: Certain, but incorrect*

*Need  
Annotation*



dog:0.8 (background)



cat:0.85 (bus)



train:0.81 (bus)



cat:0.82 (bird)

**Complementary**

**Multi-Box Committee (MBC)**

# Experiment

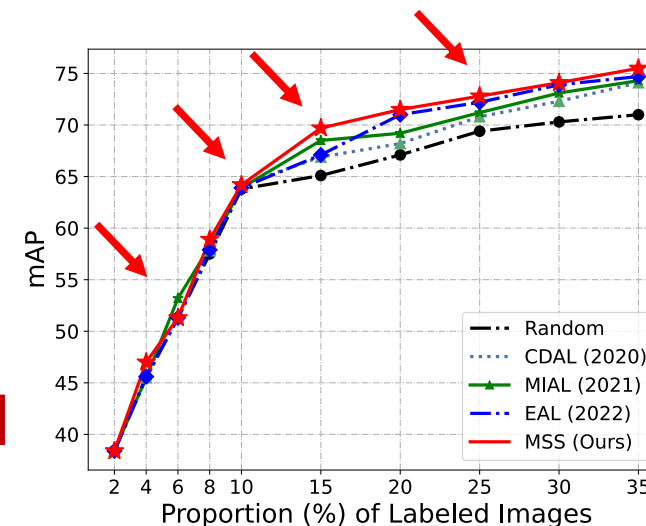
In-domain datasets *Pascal VOC* and *COCO*.

Table 1. Pascal VOC for mAP (Iou 0.5) on SSD300 results

Method	15%	20%	25%	30%	35%	100% (Oracle)
Random	65.1±0.5	67.1±0.4	69.4±0.4	70.3±0.3	71.0±0.2	77.6±0.2
Det-Ent [10]	66.1±0.4	67.7±0.4	69.8±0.3	70.8±0.2	72.1±0.2	
SEAS [2]	66.1±0.4	67.9±0.4	70.1±0.3	71.0±0.2	72.4±0.2	
WBPM [31]	67.1±0.4	68.2±0.3	70.5±0.3	72.0±0.2	73.7±0.2	
LAAL [18]	66.7±0.4	68.1±0.3	70.3±0.2	71.8±0.2	73.4±0.2	
LPM [42]	66.8±0.4	68.2±0.2	70.7±0.2	72.1±0.2	73.9±0.2	
MSS	70.2±0.3	71.5±0.3	72.8±0.3	74.1±0.2	75.3±0.2	

Table 2. COCO 20% trainval on SSD300, test-dev2015 detection results

Network	Method	Avg. Precision, IoU			Avg. Precision, Area			Avg. Recall, #Dets			Avg. Recall, Area		
		0.5:0.95	0.5	0.75	S	M	L	1	10	100	S	M	L
SSD	Random	19.9	36.3	19.5	4.9	19.6	31.1	20.5	30.3	32.3	9.8	35.3	49.5
	WBPM [31]	21.3	37.4	21.0	5.2	21.0	32.2	20.9	31.3	33.0	9.9	36.3	51.0
	LPM [42]	21.4	37.6	21.2	5.2	21.1	32.2	20.9	31.4	33.1	9.9	36.4	51.2
	MSS	22.8	39.1	22.7	5.6	22.4	33.5	21.6	32.7	33.8	10.2	37.1	52.6
	Full (Oracle)	25.1	43.1	25.8	6.1	26.4	40.5	23.6	35.2	37.3	11.6	40.5	56.0
RetinaNet	Random	26.5	41.4	24.5	9.6	26.8	37.1	26.5	36.3	37.8	15.4	40.2	54.3
	WBPM [31]	28.2	44.0	26.5	11.0	28.3	39.5	27.6	38.2	39.3	16.3	41.5	56.4
	LPM [42]	28.3	44.2	26.7	11.0	28.4	39.7	27.6	38.3	39.5	16.3	41.6	56.6
	MSS	30.0	46.7	28.7	12.4	31.2	41.5	28.8	40.1	41.0	17.0	43.1	58.5
	Full (Oracle)	34.3	53.2	36.9	16.2	37.4	47.4	32.6	44.2	46.3	20.8	48.7	65.4



**MSS achieves consistent SOTA and surpasses previous baselines !**

# Experiment

Cross-domain datasets *Pascal VOC to Clipart1k ( $P \rightarrow C$ )* and *Pascal VOC to Watercolor ( $P \rightarrow W$ )*

Table 3. Clipart1k and Watercolor (mAP) results. 5% and 10% are target data select ratio.

Method	Clipart1k		Watercolor	
	5%	10%	5%	10%
STABR	35.7		49.9	
SWDA	38.1		53.3	
Random	31.4±0.4	39.0±0.3	47.2±0.3	54.2±0.2
MSS	35.2±0.3	43.3±0.2	51.3±0.2	57.1±0.2

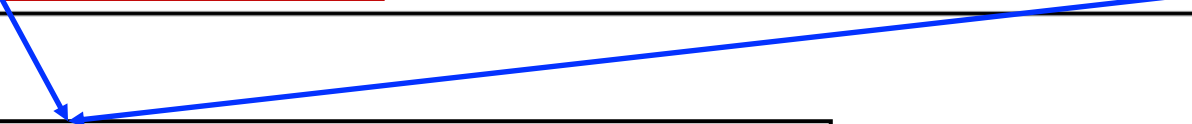
**MSS shows good transferability among different conditions !**



# Ablation Study

Table 4. Ablation Study of Pascal VOC and COCO on SSD300.

Dataset	Metric	random	w/ MBU	w/ MBC	w/ MBO	w/ MBU+MBC	MSS
Pascal VOC 35%	mAP (0.5)	71.0 $\pm$ 0.2	73.7 $\pm$ 0.2	72.5 $\pm$ 0.3	72.1 $\pm$ 0.2	74.8 $\pm$ 0.2	75.3 $\pm$ 0.2
COCO 20%	AP (0.5:0.95)	26.5 $\pm$ 0.1	28.8 $\pm$ 0.2	27.9 $\pm$ 0.3	27.3 $\pm$ 0.2	29.5 $\pm$ 0.2	30.0 $\pm$ 0.2

- 
- 1、 Both w/ MBU and w/ MBC outperform random selection.
  - 2、 MSS outperforms both w/ MBU and w/ MBC.

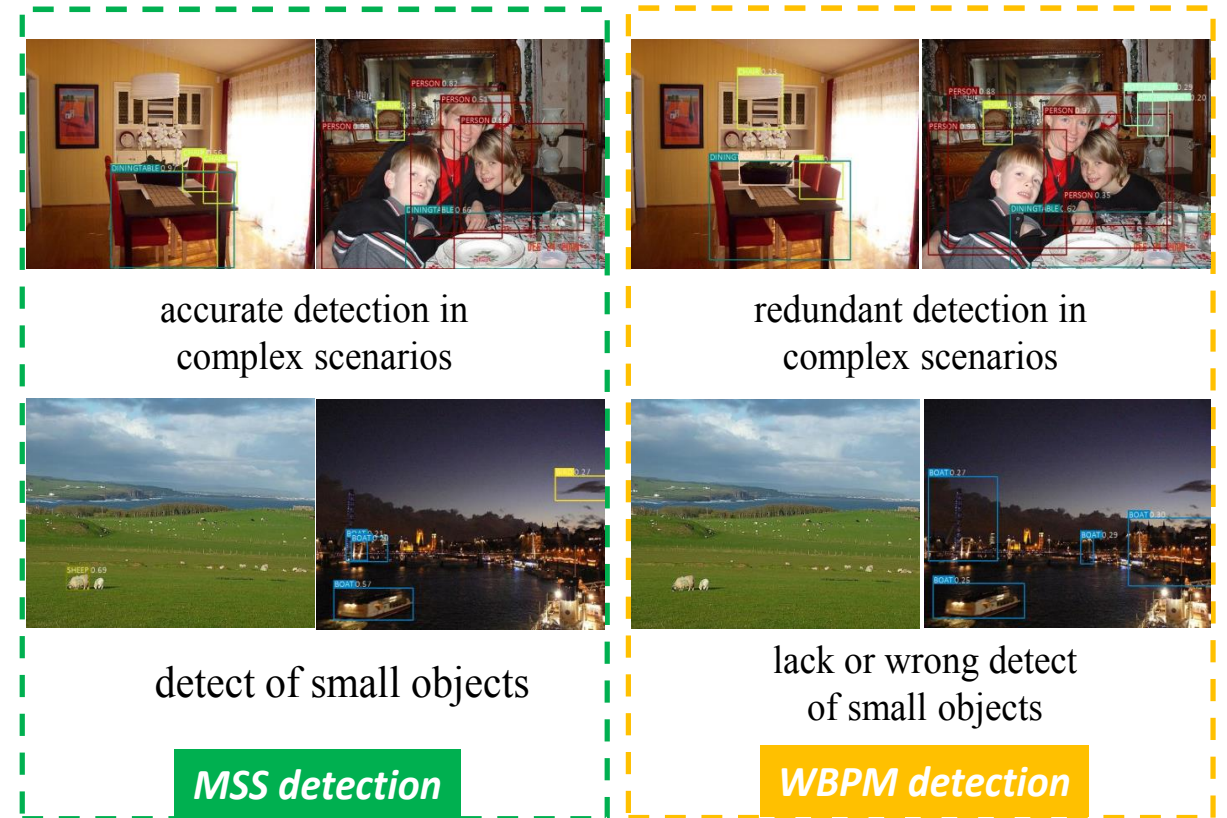
**MBU and MBC are complementary to each other and can get better results when used together !**

# Show Cases

Compared to *WBPM*, *MBU alone*, *MBC alone* and *MSS*.



sample selected



difficult cases



**Thank You!**  
djx20@mails.tsinghua.edu.cn