

# HoughNet: Integrating near and long-range evidence for bottom-up object detection

---

Nermin Samet, Samet Hicsonmez, Emre Akbas

# OUTLINE

- Step-by-step animation of the voting process
- Details about the “Mini COCO” (COCO `minitrain`) dataset
- More visual results

# Vote Aggregation

In the following slides, we illustrate the vote aggregation process for two steps, for a specific class.

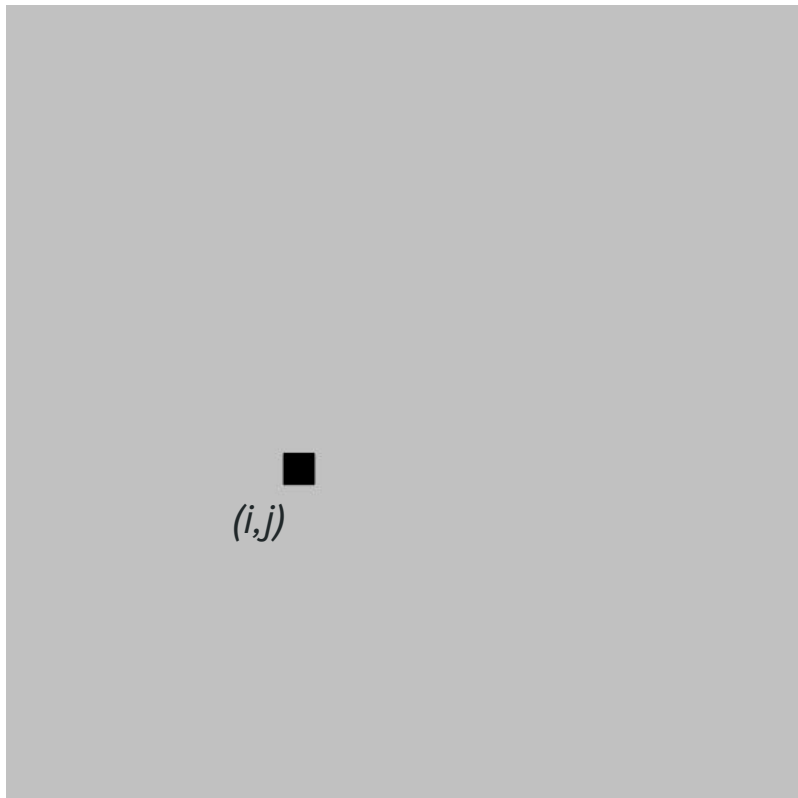
**E** is the visual evidence tensor for a specific class. Its size is  $H \times W \times R$ .

**O** is the corresponding object presence map with size  $H \times W$ .

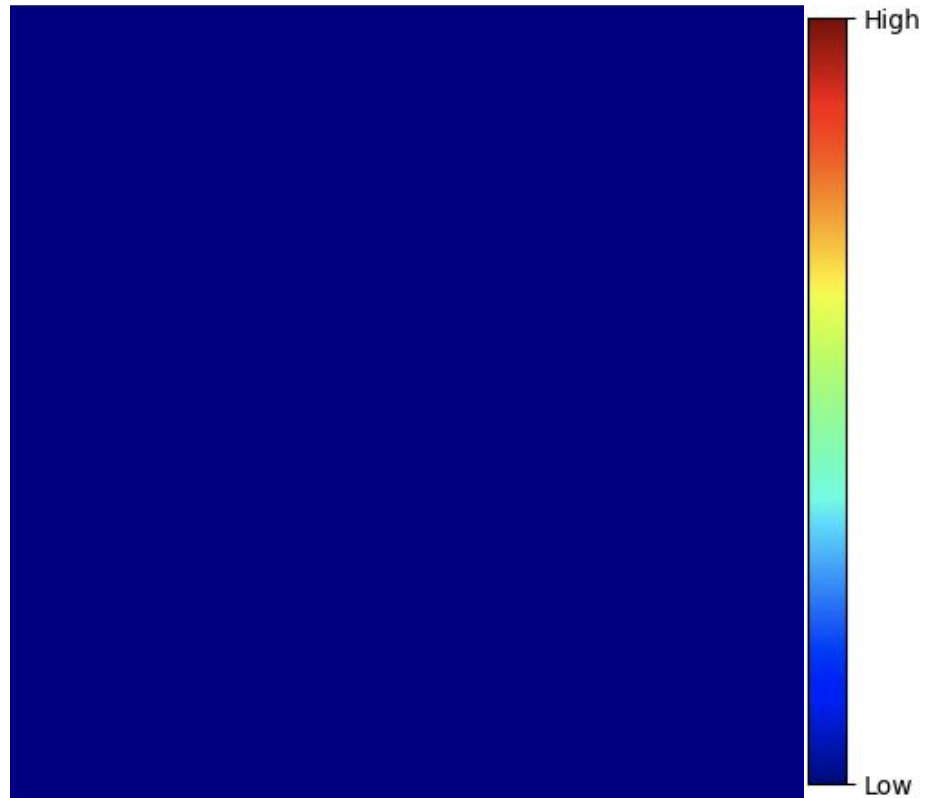
The colors in **O** indicate accumulated vote strength.

Let us pick an arbitrary location  $(i,j)$  and an arbitrary channel in  $\mathbf{E}$  (e.g., the 3<sup>rd</sup> channel is illustrated below).

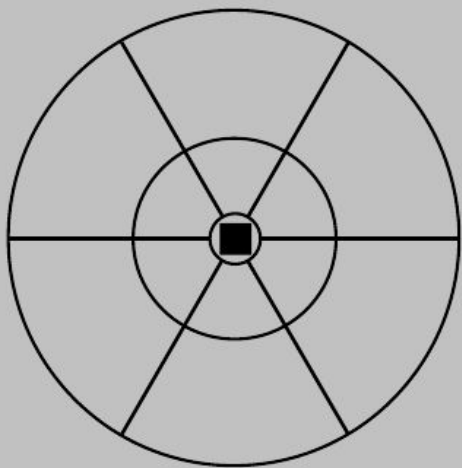
The 3<sup>rd</sup> channel of  $\mathbf{E}$



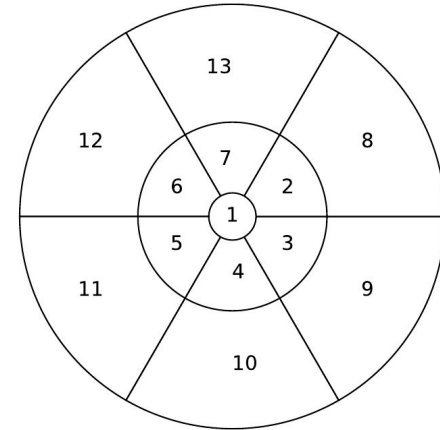
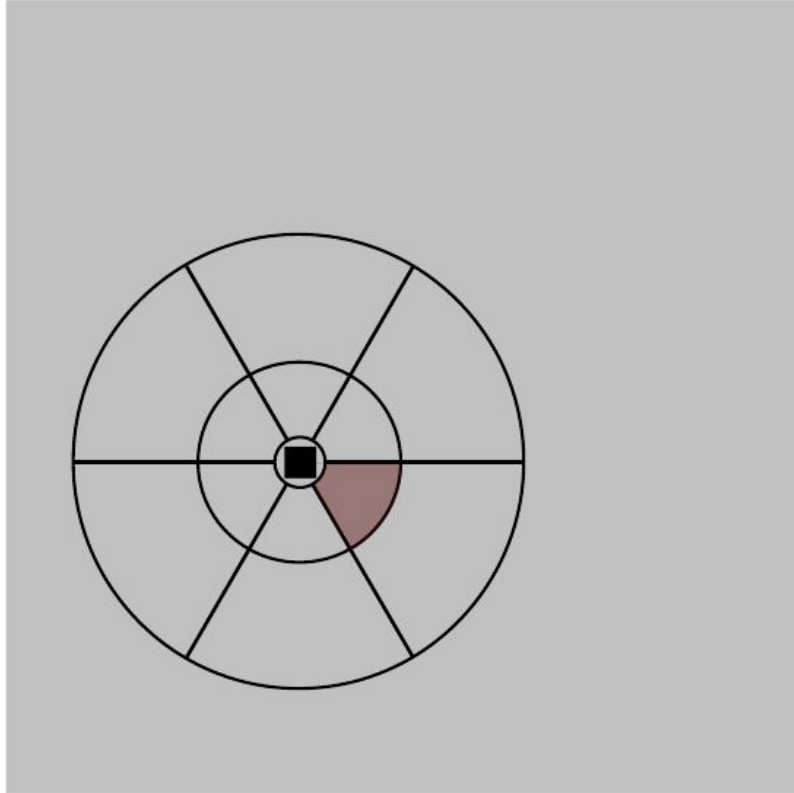
$\mathbf{O}$  is initialized with zeros



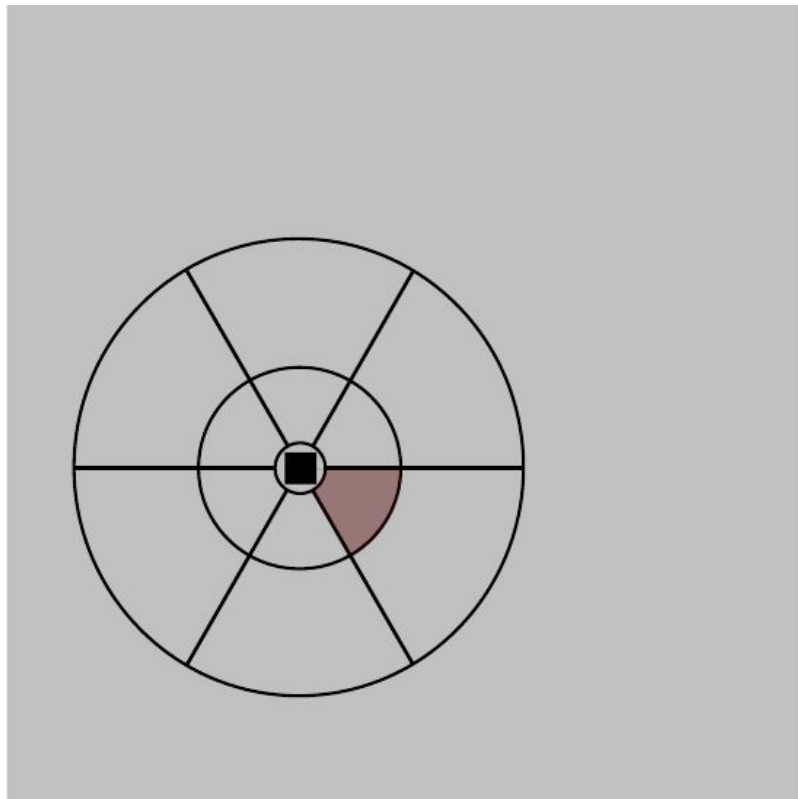
We place the vote field centered at the location  $(i,j)$ .



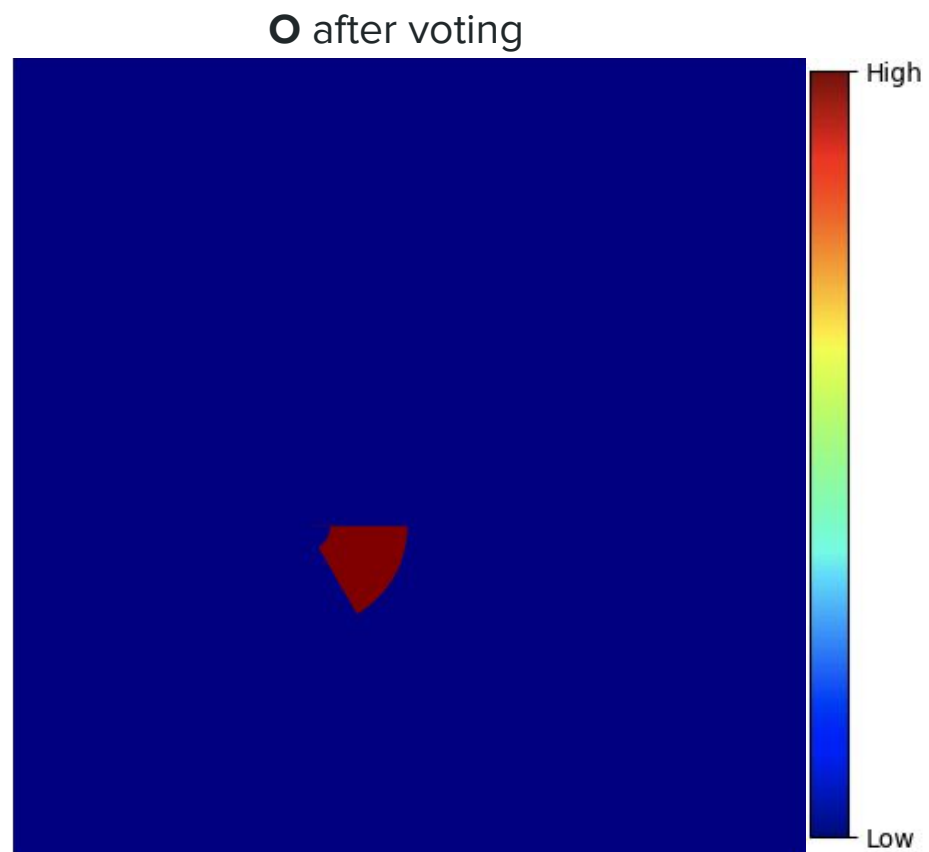
Since, we are using the 3<sup>rd</sup> channel, the relevant region of the vote field is 3 (shown with pink color on **E**).



the vote field

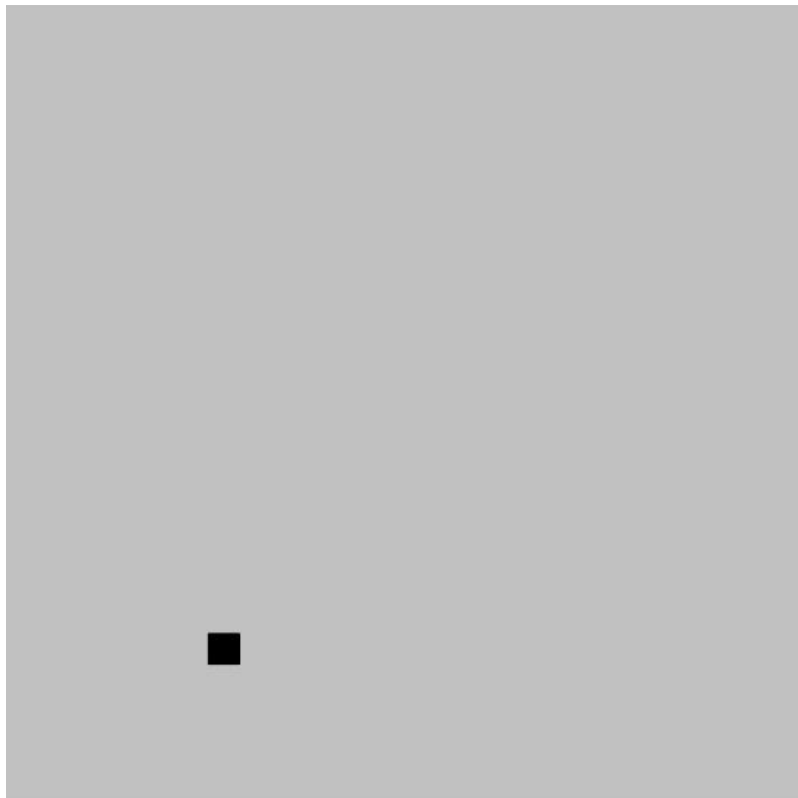


voting  
→

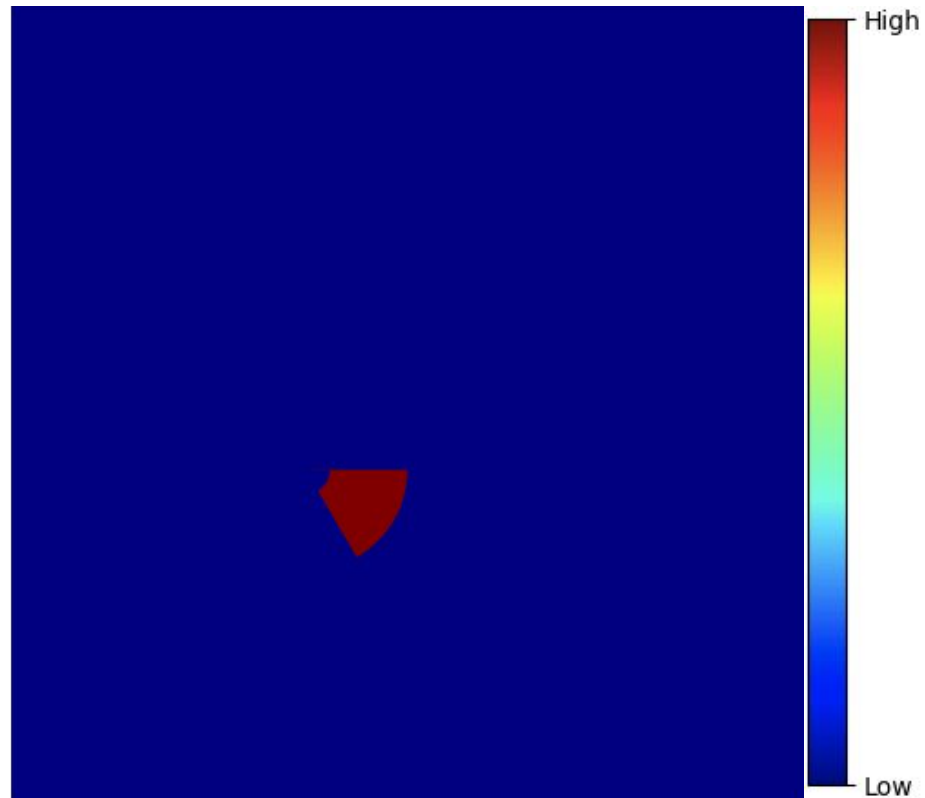


To further illustrate the voting process, let us pick another location (shown below on **E**) and another (e.g. the 8th) channel in **E**.

The 8<sup>rd</sup> channel of **E**

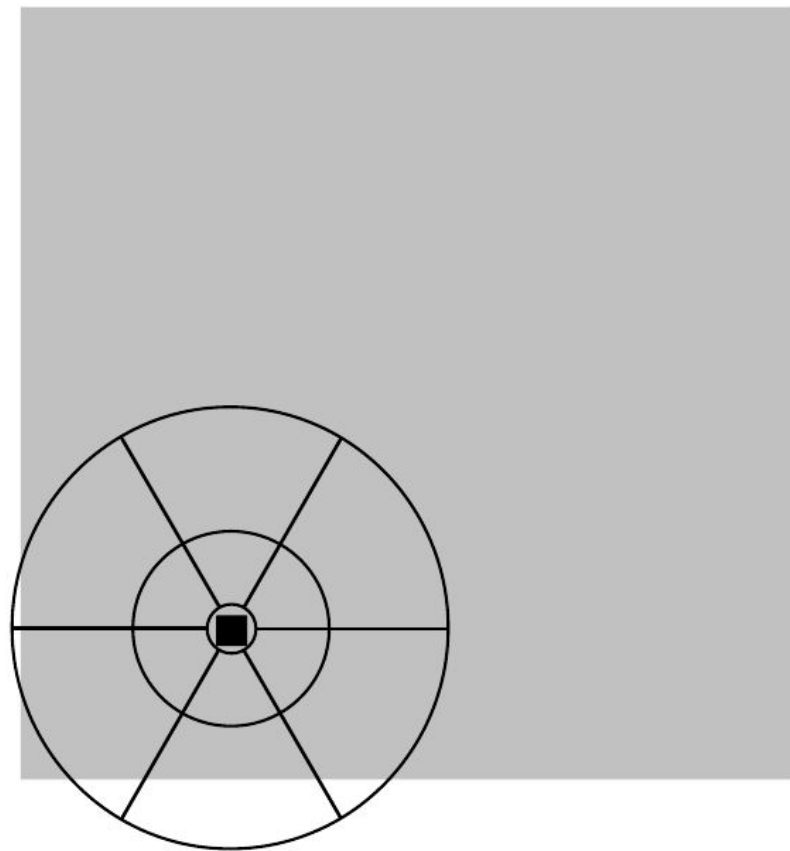


**O**

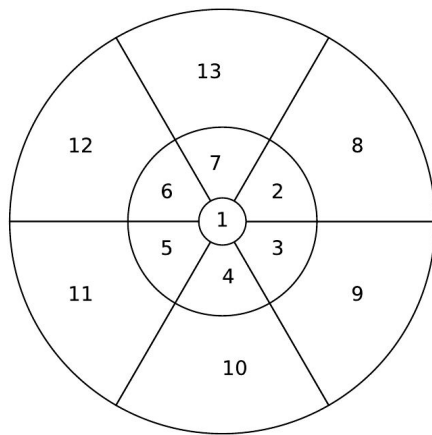
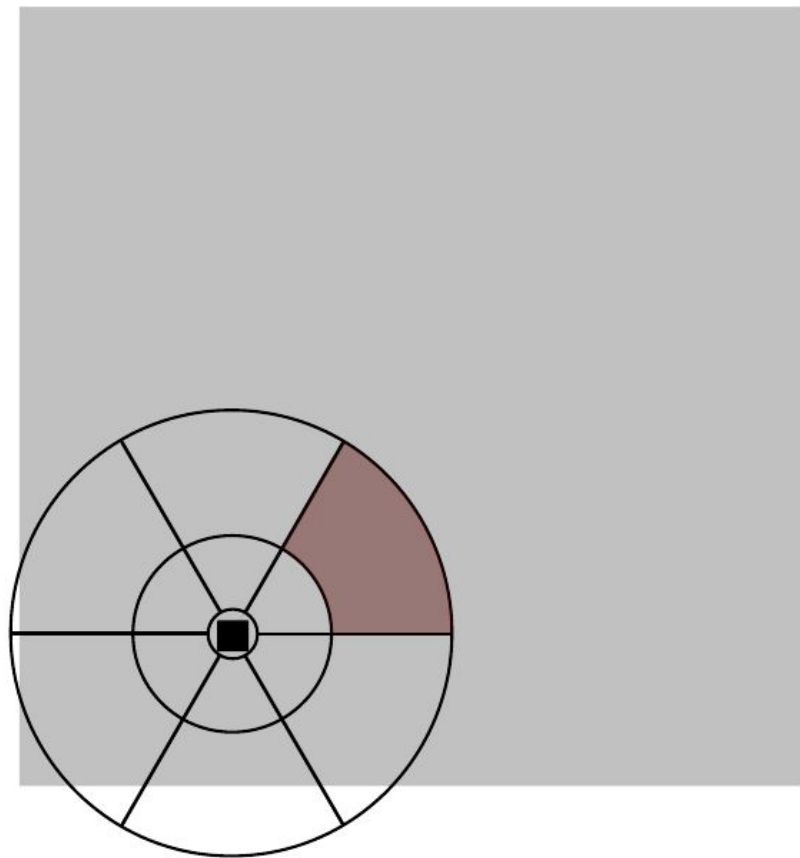




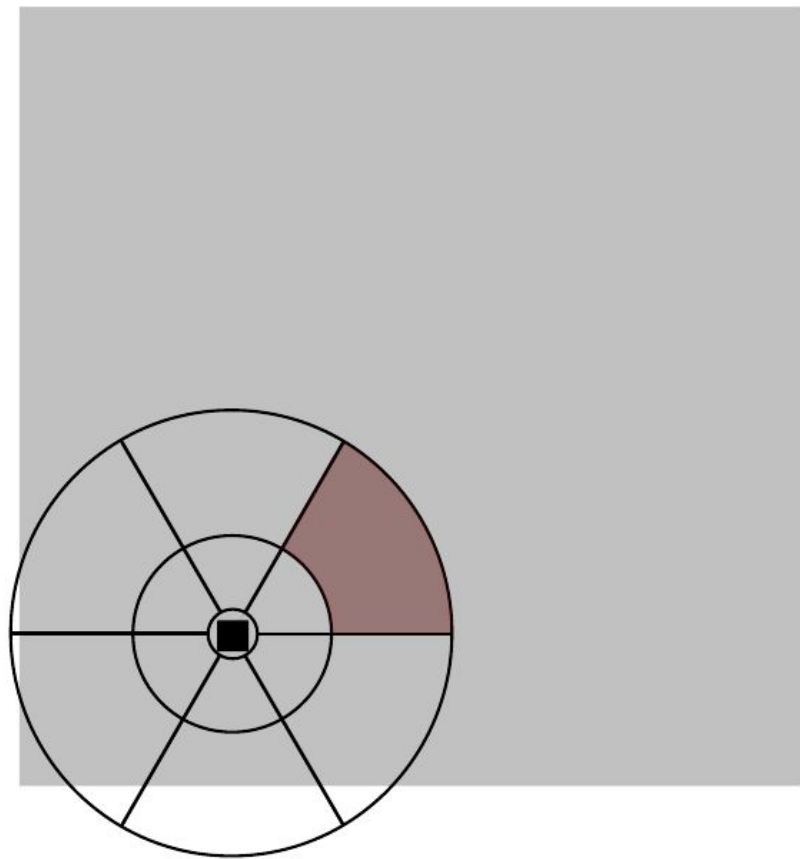
Place the vote field centered at the location.



The region **8** marks the target area to be voted on.

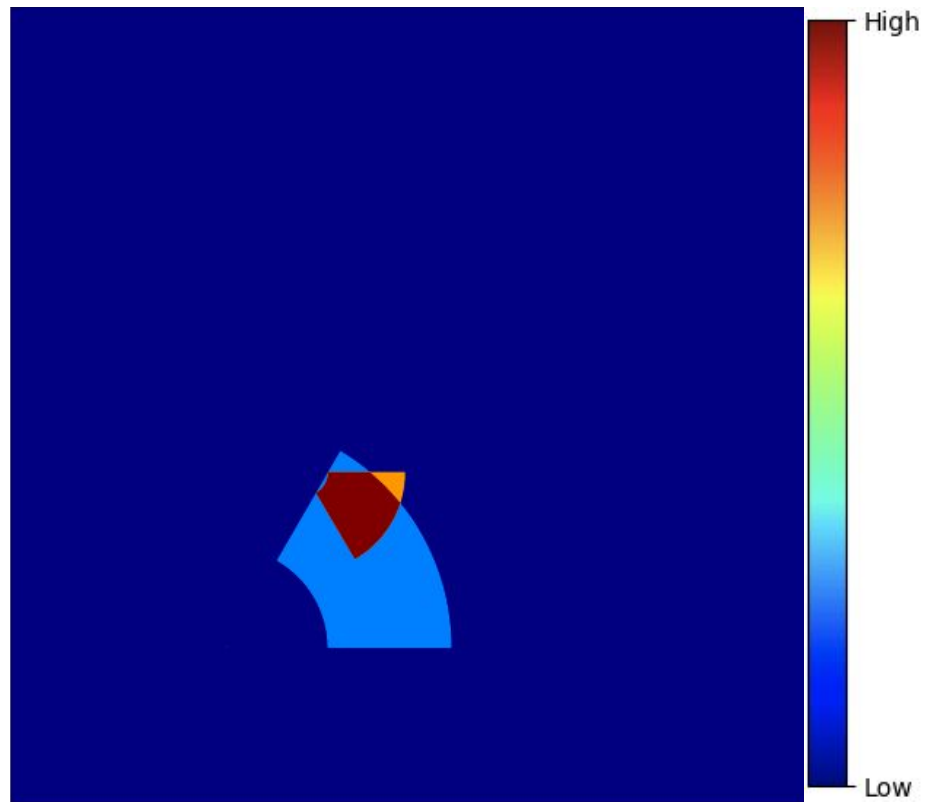


the vote field



voting  
→

○ after voting



# Mini COCO (i.e. COCO minitrain)

In the following slides,

- We present the full set of results trained on the COCO minitrain benchmark.
- We present the correlation between COCO train2017 and COCO minitrain.
- We present several statistics about COCO minitrain.
  - We show *Person* class separately, because it is the most dominant class in the dataset. When we add person class to the figures, it becomes harder to see the details of other classes.

# COCO minitrain Benchmark Results

Method	BackBone	Scale	$AP$	$AP_{50}$	$AP_{75}$	$AP_S$	$AP_M$	$AP_L$
<i>Two-stage detectors:</i>								
Faster R-CNN	ResNet-50 w FPN	800	27.7	48.8	28.4	14.7	29.8	36.4
Mask R-CNN	ResNet-50 w FPN	800	28.5	49.5	29.4	14.7	30.7	37.6
<i>One-stage detectors:</i>								
RetinaNet	ResNet-50 w FPN	800	25.7	43.1	26.8	12.1	28.6	34.2
CornerNet	Hourglass-104	511	28.4	41.8	29.5	11.3	29.6	39.2
ExtremeNet	Hourglass-104	511	27.3	39.4	28.9	12.5	29.6	38.0

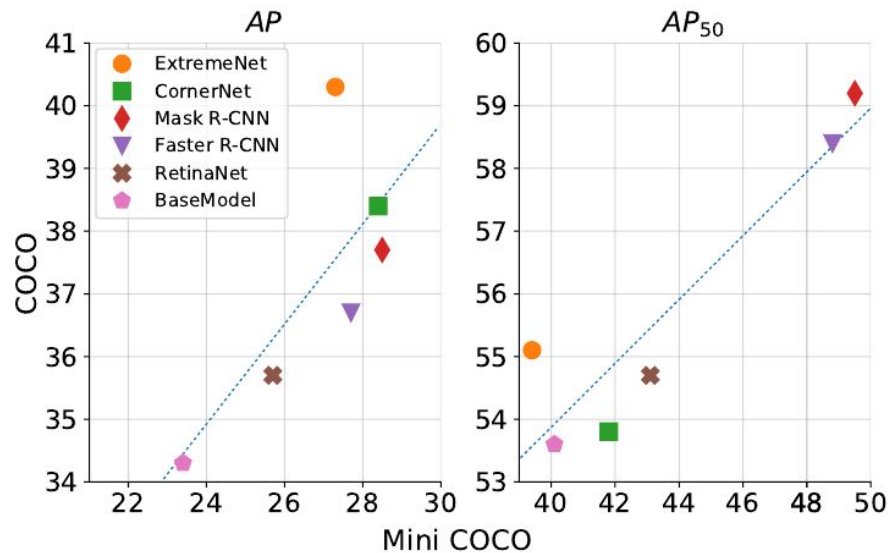
COCO val2017 performances. Models were trained on COCO minitrain.

# COCO minitrain Benchmark Results

Method	Backbone	Scale	MiniCOCO			COCO		
			$AP$	$AP_{50}$	$AP_{75}$	$AP$	$AP_{50}$	$AP_{75}$
<i>Two-stage detectors:</i>								
Faster R-CNN	ResNet-50 w FPN	800	27.7	48.8	28.4	36.7	58.4	39.6
Mask R-CNN	ResNet-50 w FPN	800	28.5	49.5	29.4	37.7	59.2	40.9
<i>One-stage detectors:</i>								
RetinaNet	ResNet-50 w FPN	800	25.7	43.1	26.8	35.7	54.7	38.5
CornerNet	Hourglass-104	511	28.4	41.8	29.5	38.4	53.8	40.9
ExtremeNet	Hourglass-104	511	27.3	39.4	28.9	40.3	55.1	43.7
HoughNet (ours)	ResNet-101	512	23.4	40.1	23.6	34.3	53.6	36.6

COCO minitrain vs COCO train2017. Results are given on the COCO val2017.

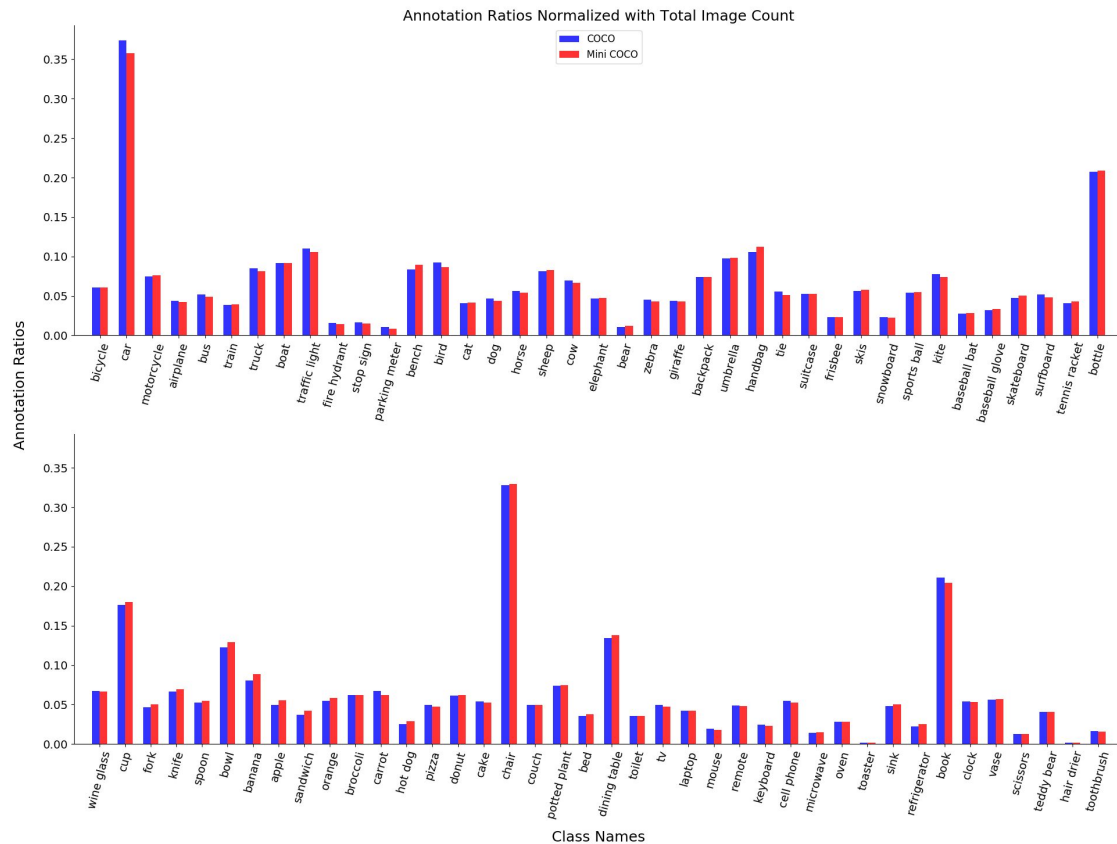
## Correlation between COCO train2017 and COCO minitrain



Positive correlation between COCO train2017 and COCO minitrain results. The Pearson correlation coefficients are 0.74 and 0.92 for COCO evaluation metrics AP and AP50 respectively. This figure is based on the table from the previous slide.

# COCO minitrain Statistics

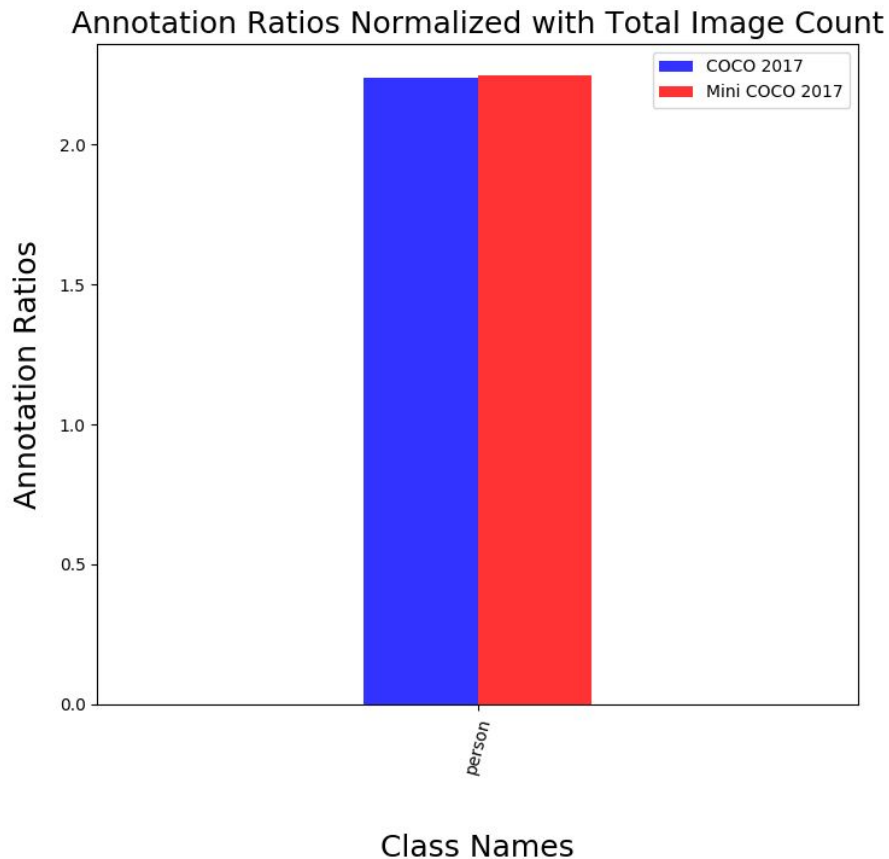
- *Total annotations (i.e. object instances) normalized with total image counts in the dataset.*





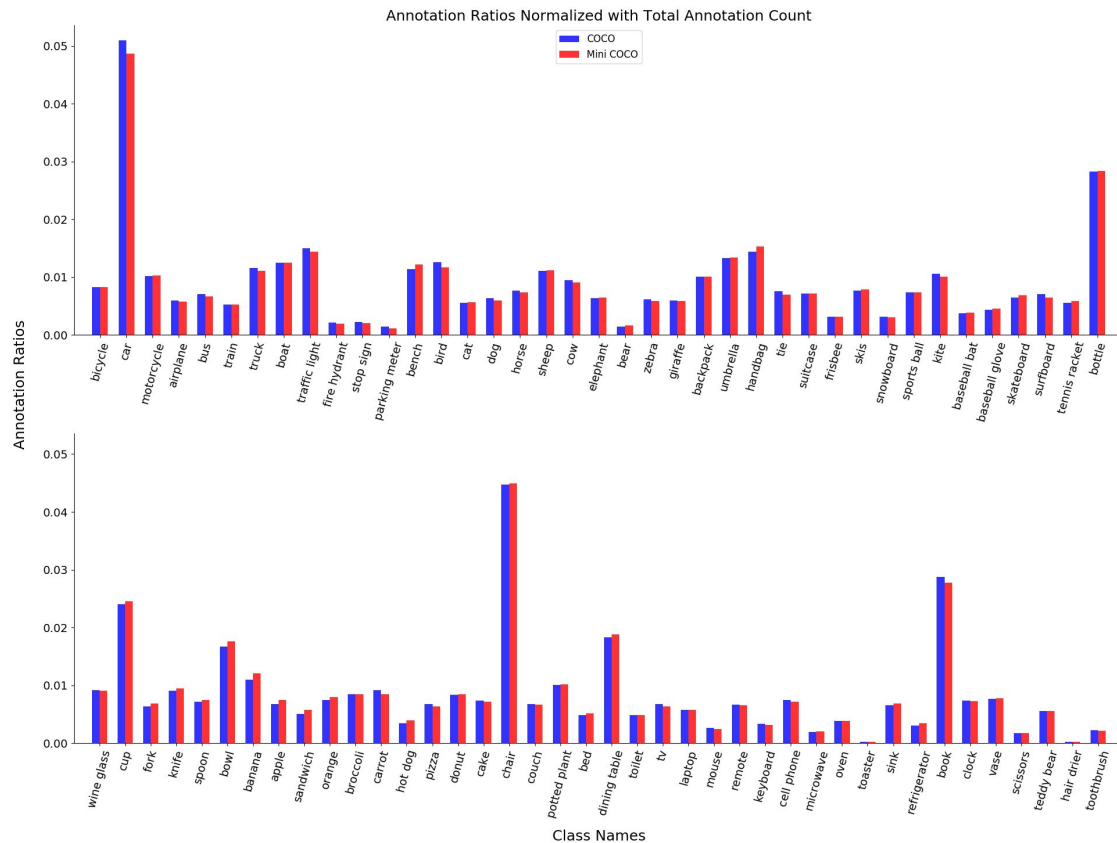
# COCO minitrain Statistics

- *Person* annotations normalized with total image counts in the dataset.



# COCO minitrain Statistics

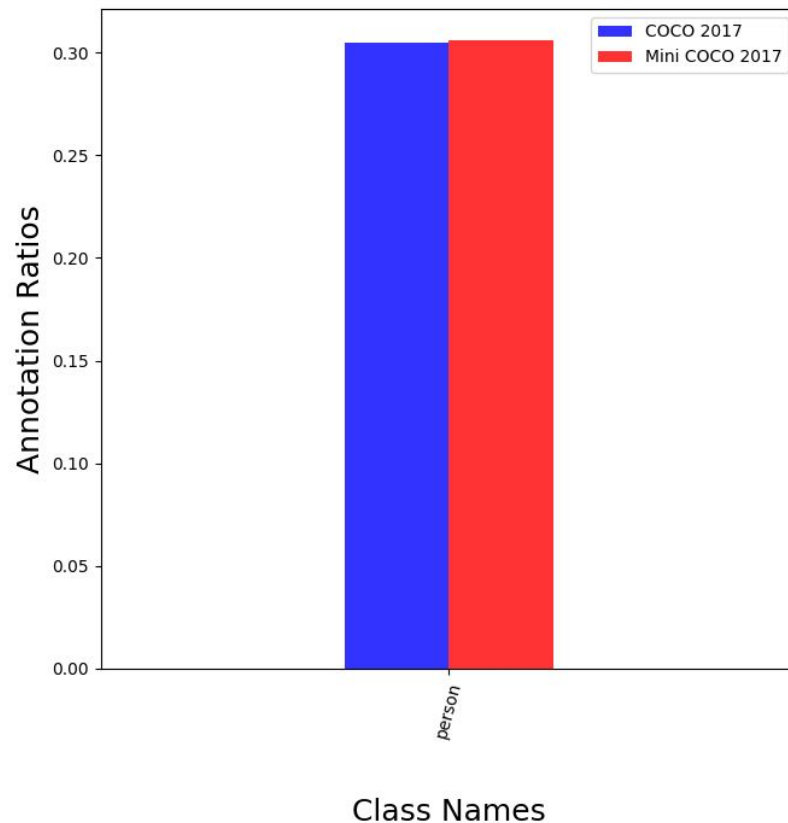
- Total annotations normalized with total annotation counts in the dataset.



# COCO minitrain Statistics

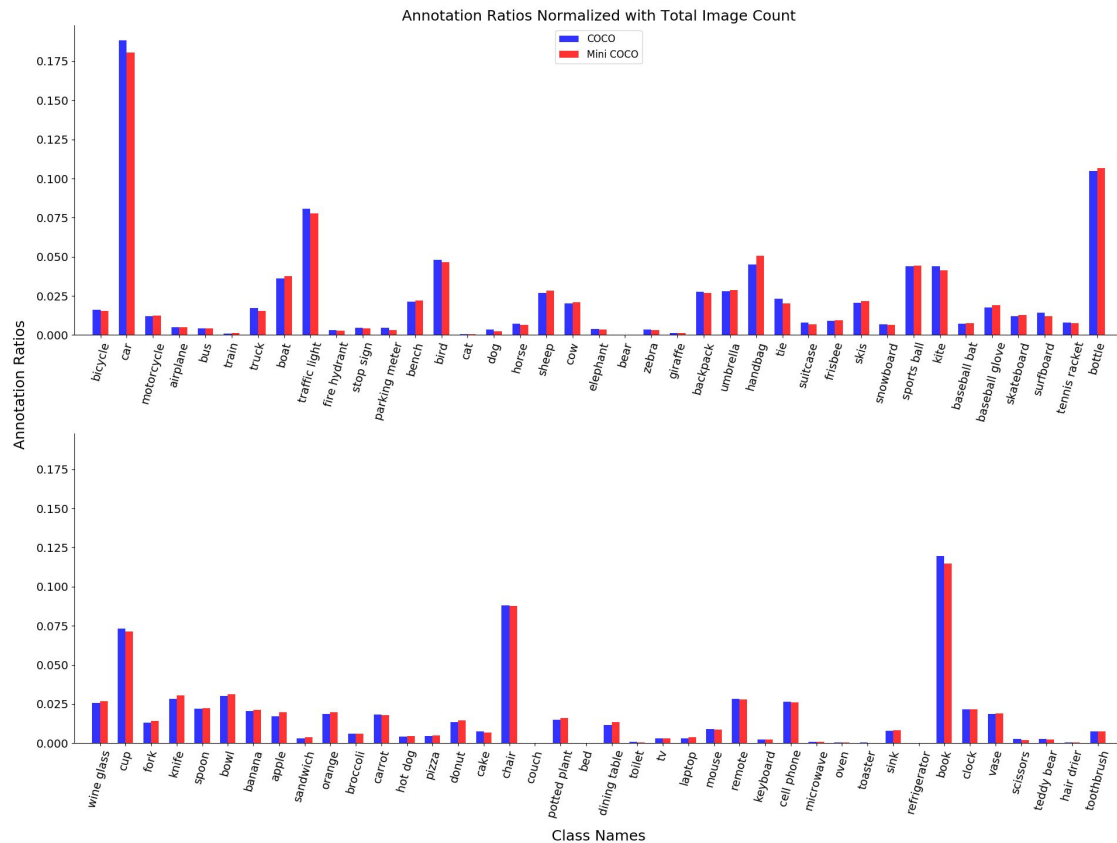
- *Person* annotations normalized with total annotation counts in the dataset.

Annotation Ratios Normalized with Total Annotation Count



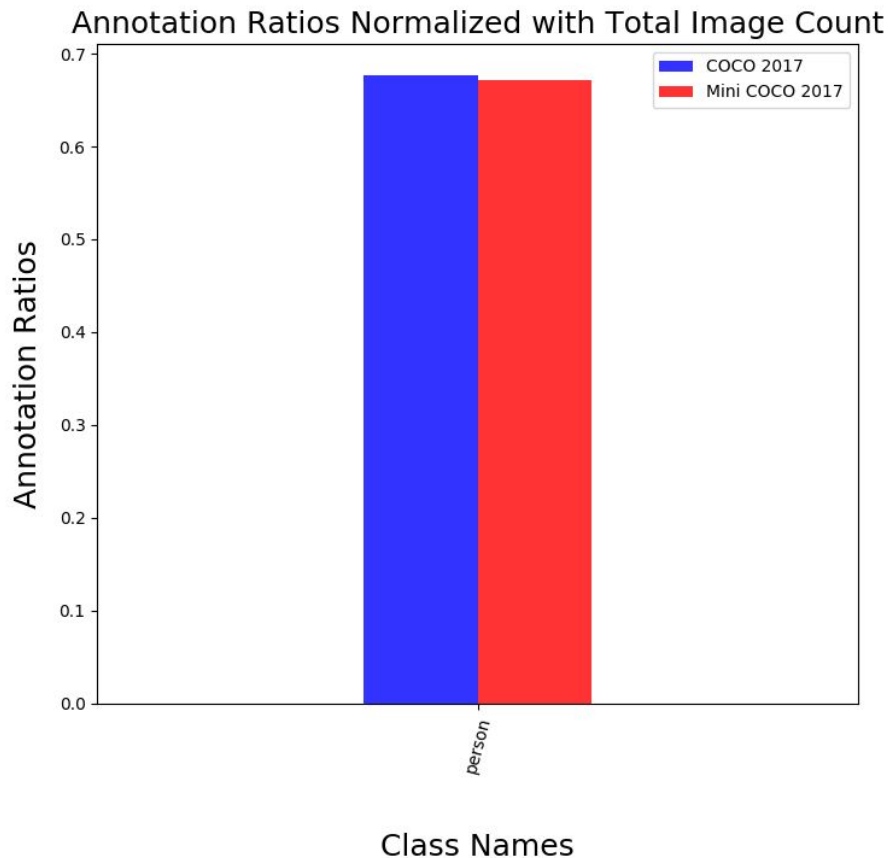
# COCO minitrain Statistics

- *Small* annotations normalized with total image counts in the dataset.



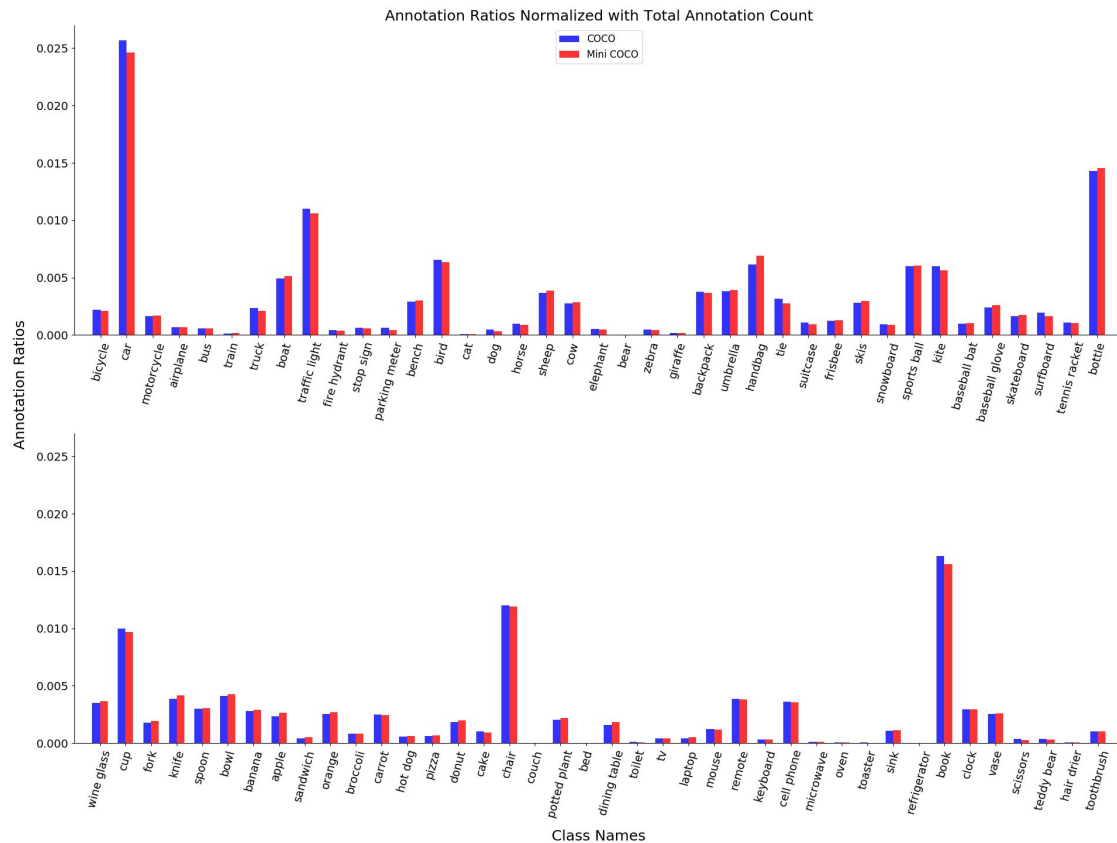
# COCO minitrain Statistics

- *Small Person* annotations normalized with total image counts in the dataset.



# COCO minitrain Statistics

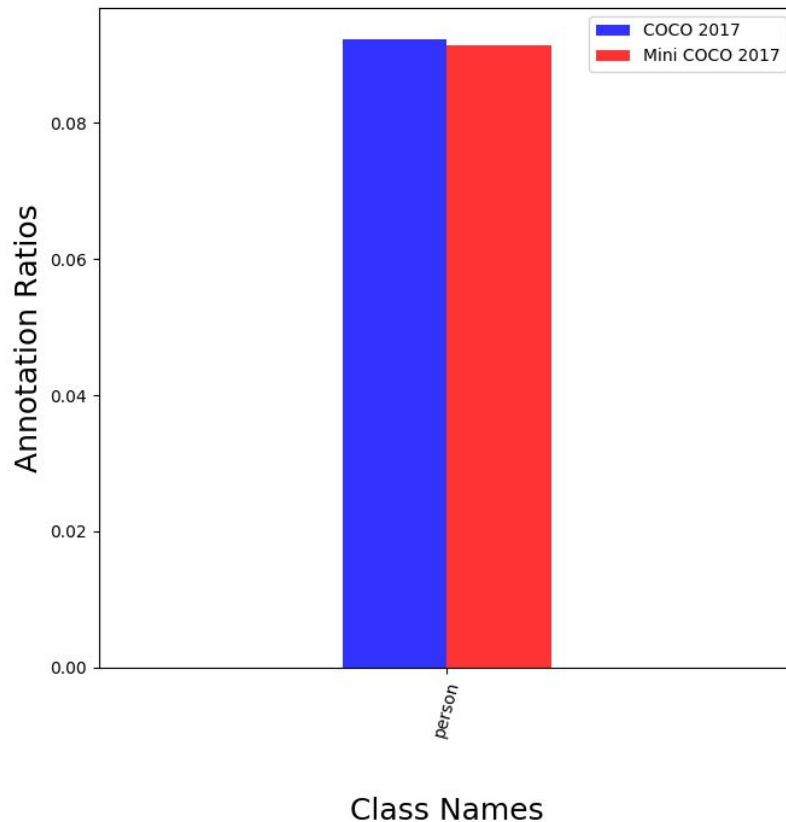
- *Small* annotations normalized with total annotation counts in the dataset.



# COCO minitrain Statistics

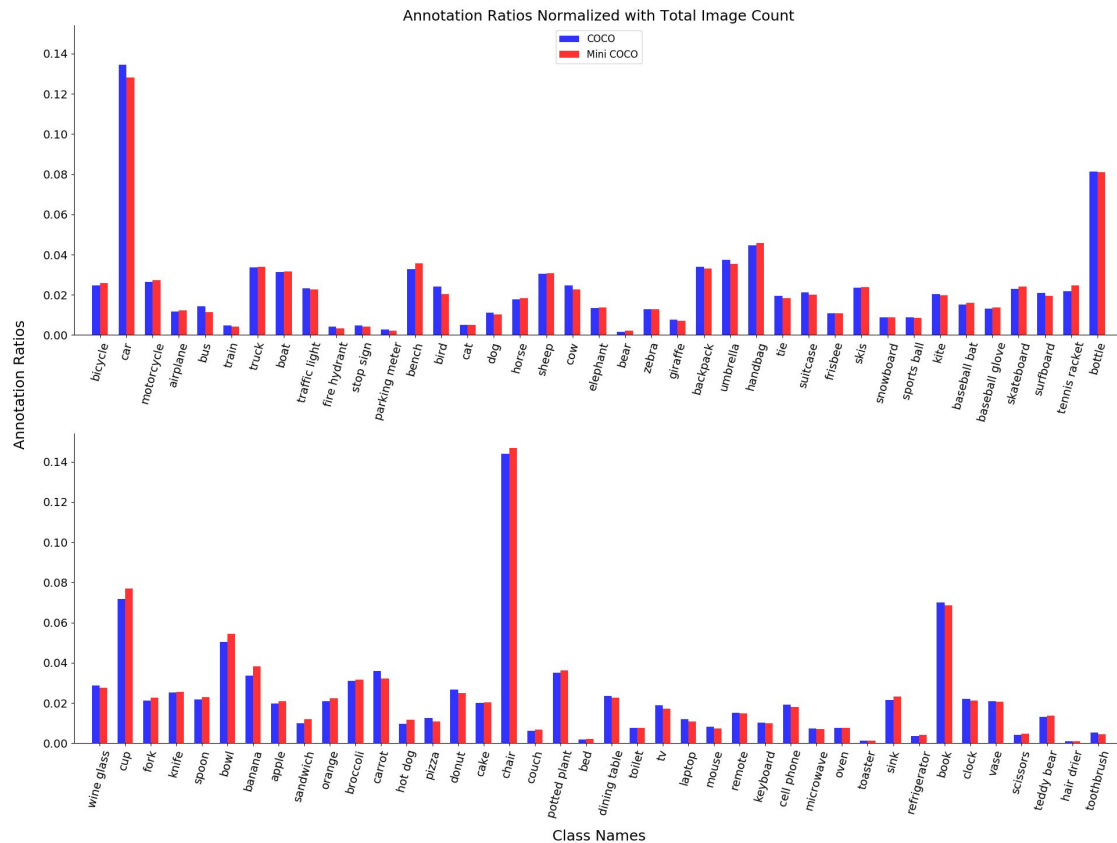
- *Small Person* annotations normalized with total annotation counts in the dataset.

Annotation Ratios Normalized with Total Annotation Count



# COCO minitrain Statistics

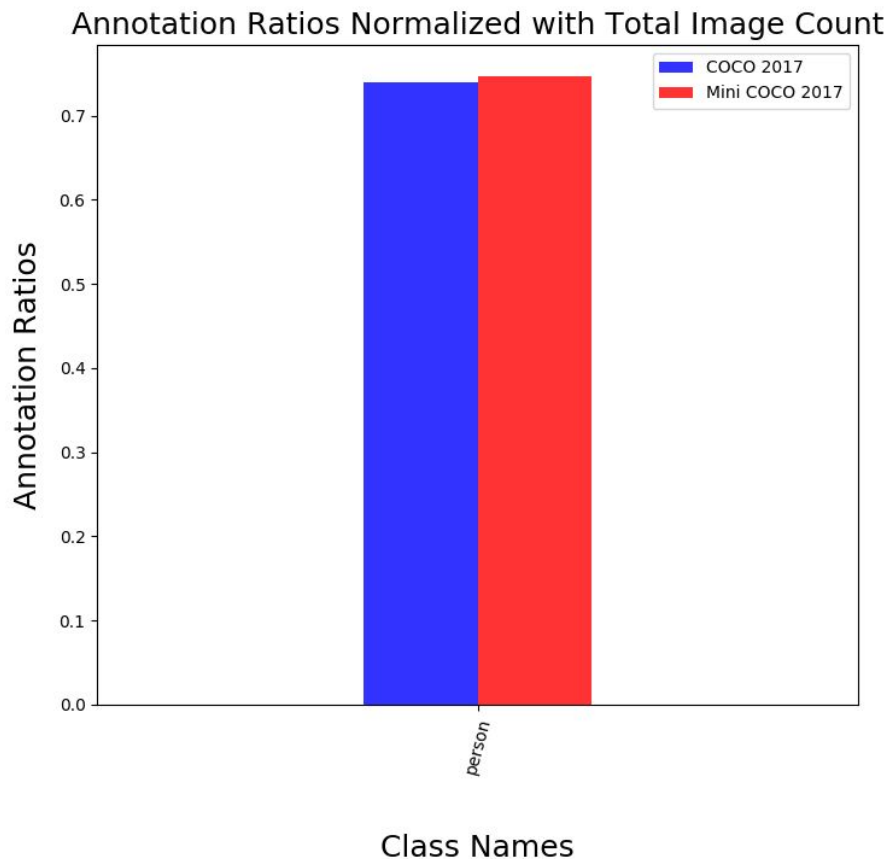
- *Medium* annotations normalized with total image counts in the dataset.





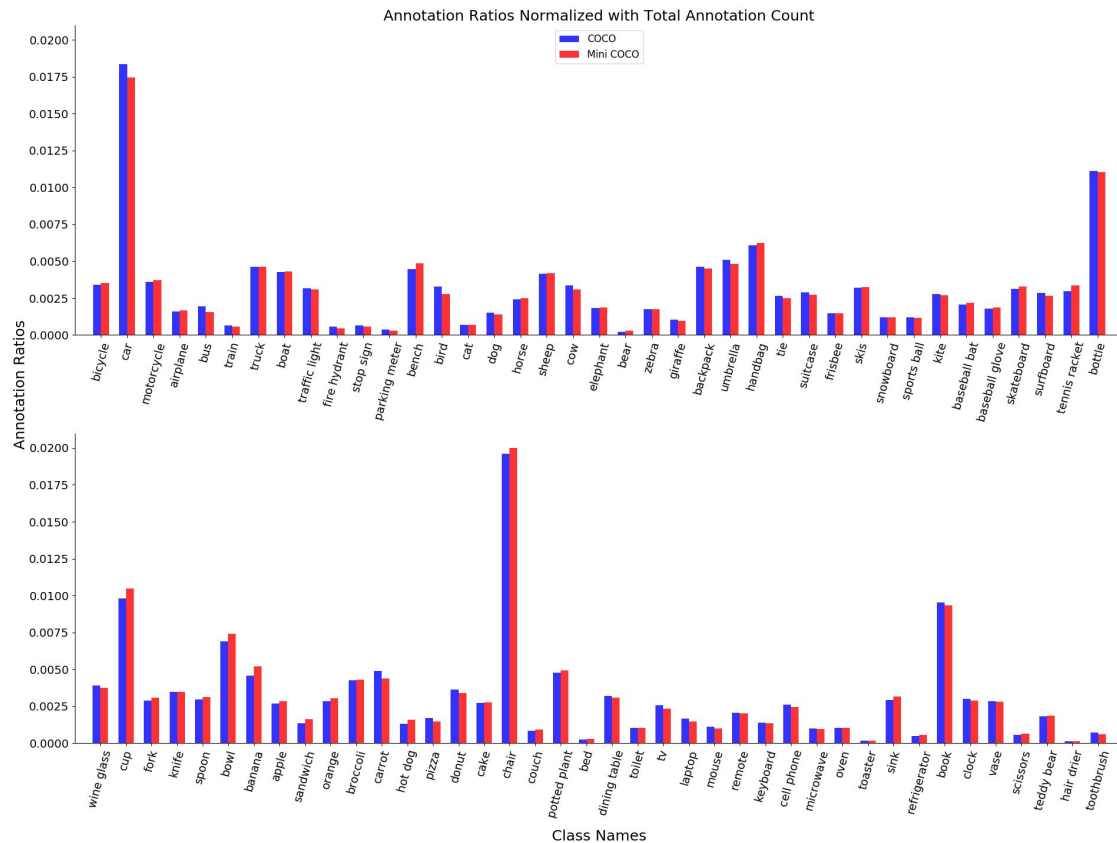
# COCO minitrain Statistics

- Medium *Person* annotations normalized with total image counts in the dataset.



# COCO minitrain Statistics

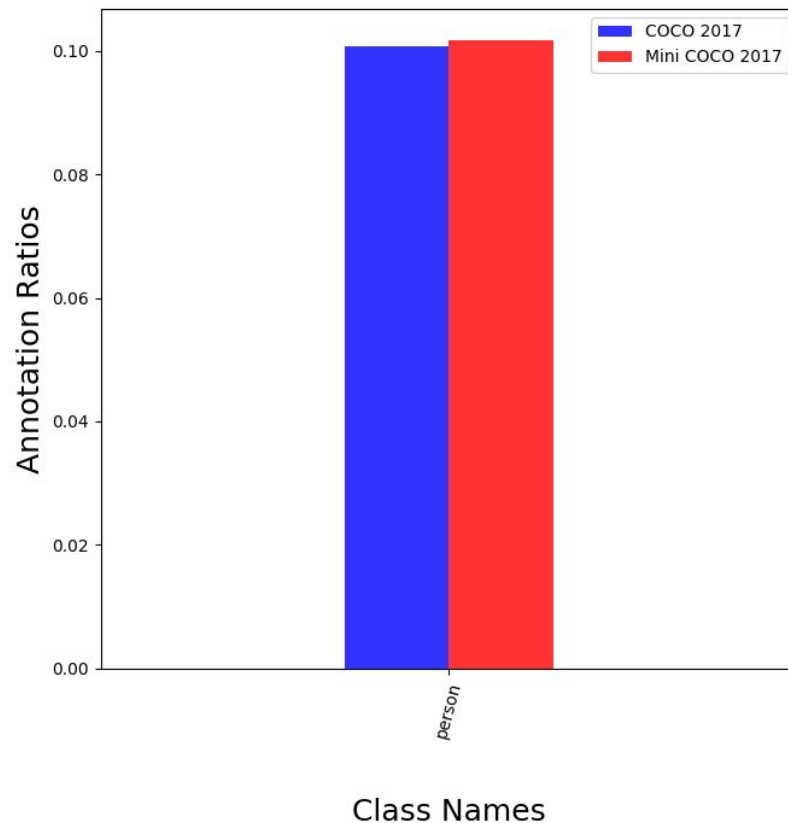
- *Medium* annotations normalized with total annotation counts in the dataset.



# COCO minitrain Statistics

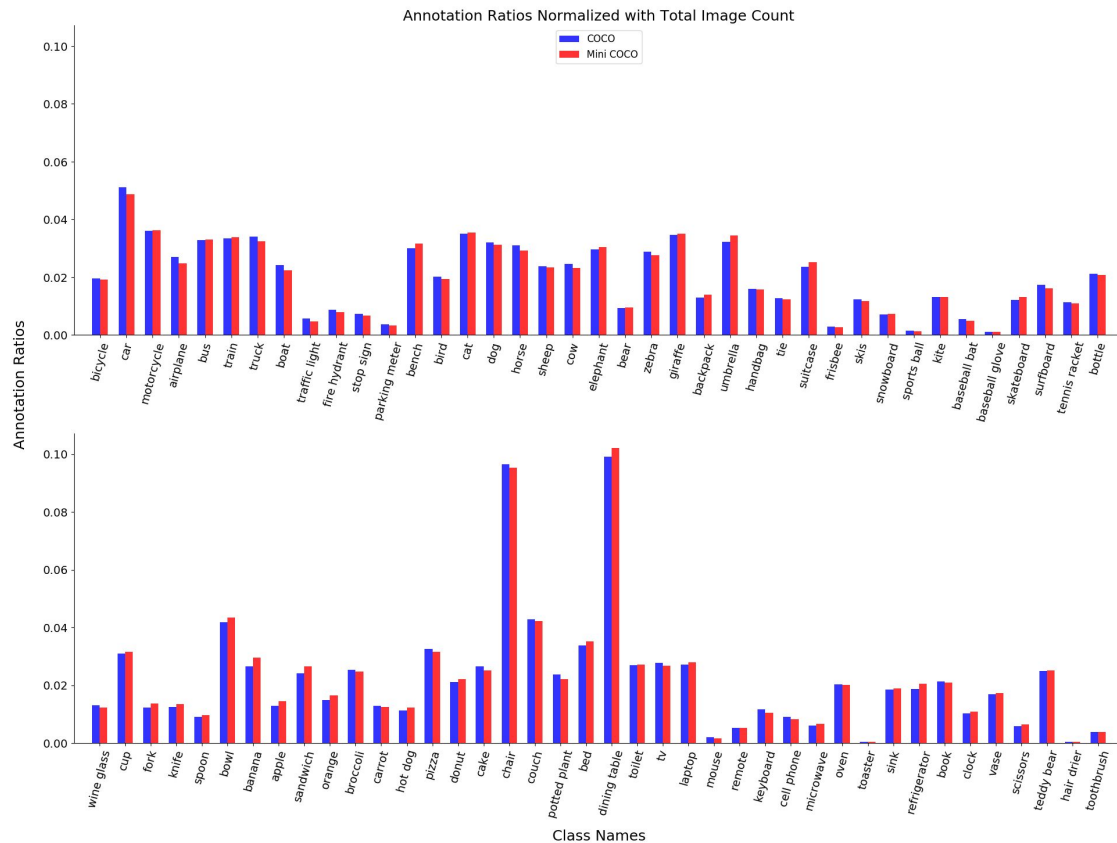
- Medium *Person* annotations normalized with total annotation counts in the dataset.

Annotation Ratios Normalized with Total Annotation Count



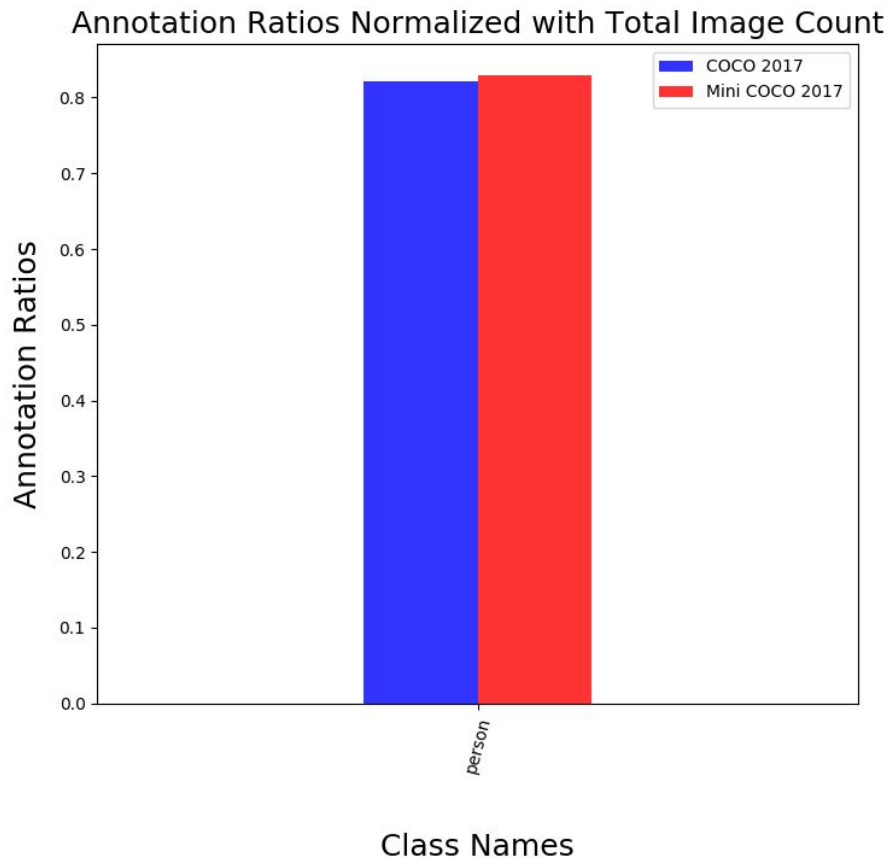
# COCO minitrain Statistics

- *Large* annotations normalized with total image counts in the dataset.



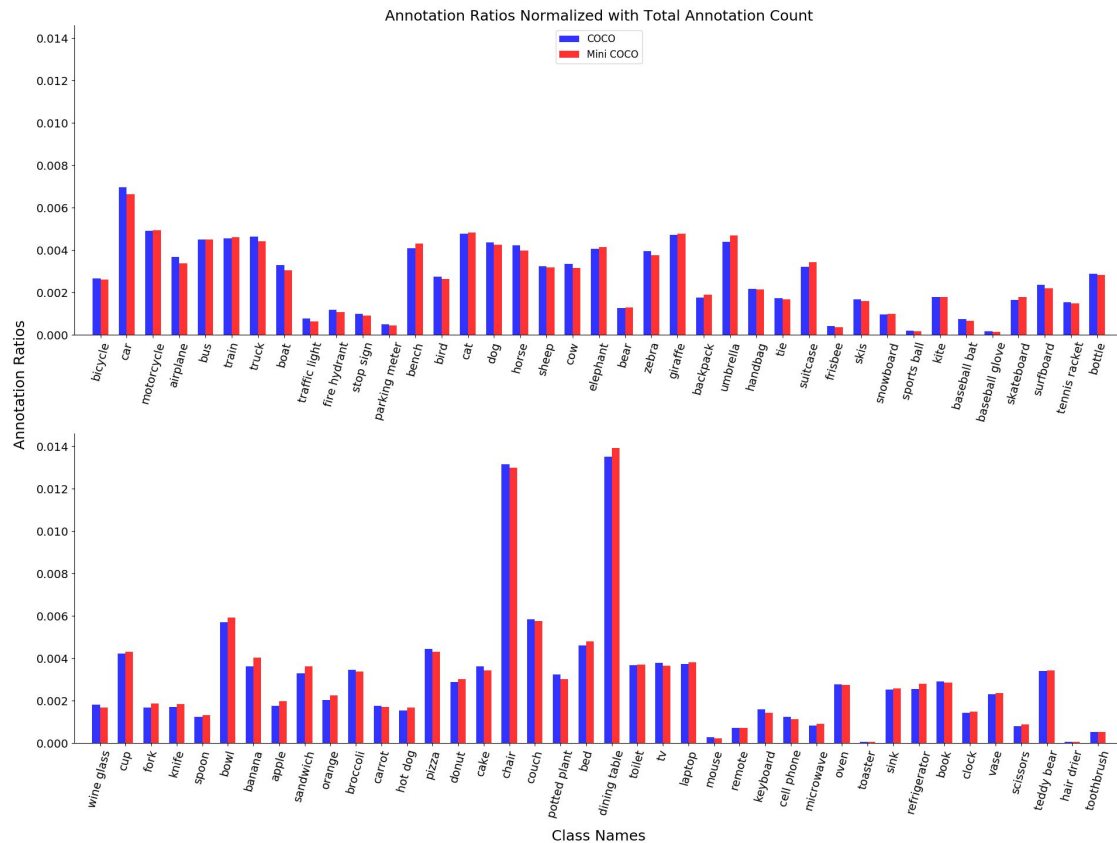
# COCO minitrain Statistics

- Large *Person* annotations normalized with total image counts in the dataset.



# COCO minitrain Statistics

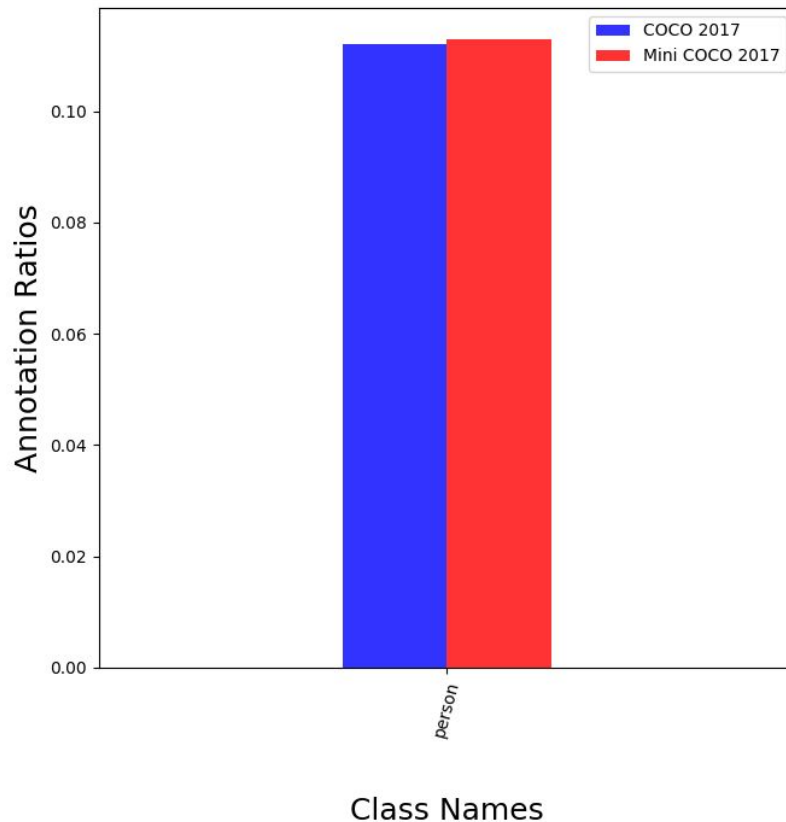
- *Large* annotations normalized with total annotation counts in the dataset.



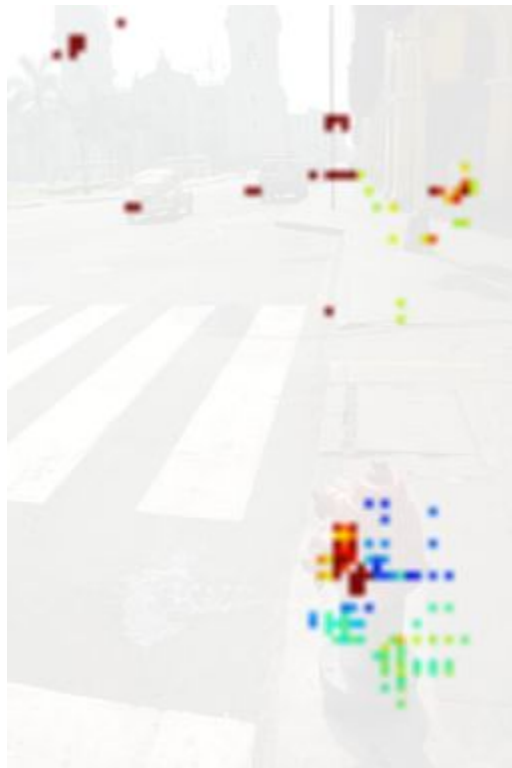
# COCO minitrain Statistics

- Large *Person* annotations normalized with total annotation counts in the dataset.

Annotation Ratios Normalized with Total Annotation Count



## More Visual Results



*Fire hydrant detection gets strong votes from cars, person, buildings and road.*



## More Visual Results



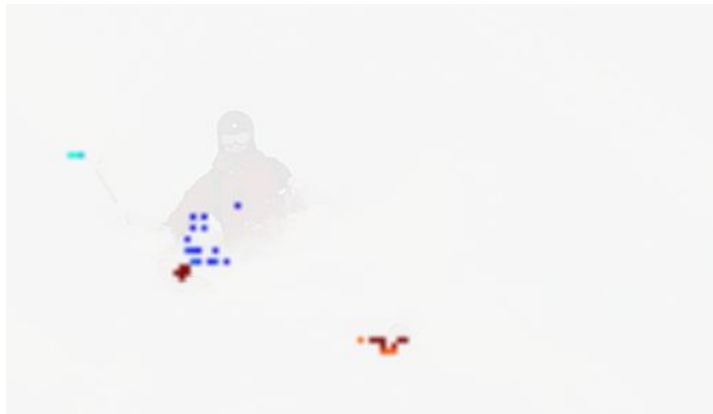
*Tennis racket  
detection gets  
strong votes  
from person.*

# More Visual Results



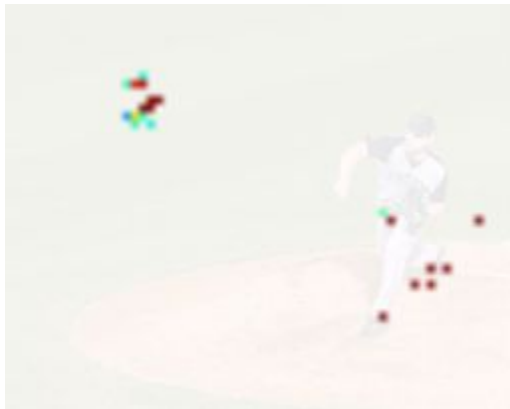
*Kite detection gets strong votes from person and sky.*

# More Visual Results



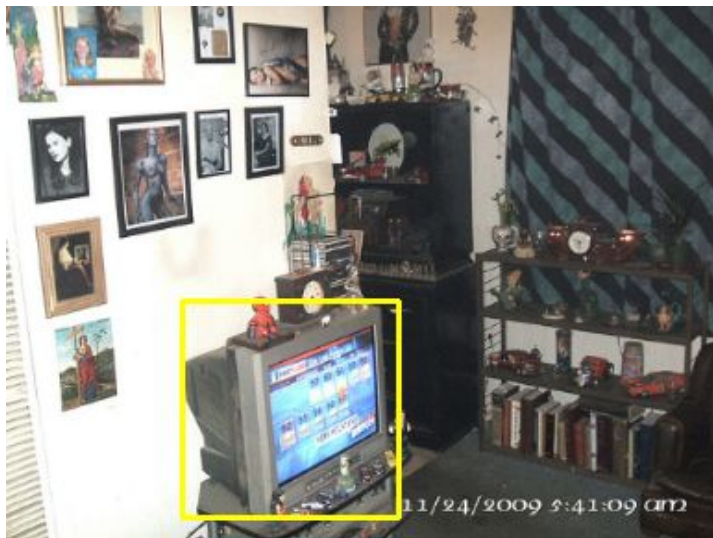
*Ski detection gets strong votes from other ski, ski baton and person.*

## More Visual Results



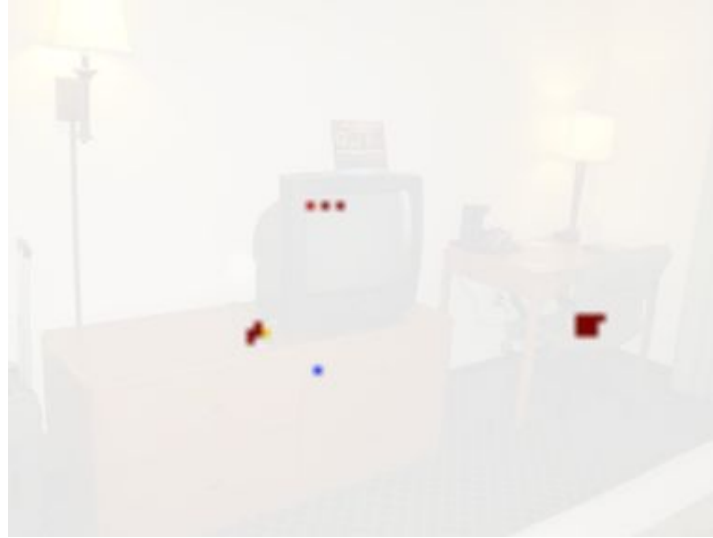
*Sports ball* detection gets strong votes from *person*.

# More Visual Results



*Television* detection gets strong votes from common things in a living room such as *paintings at the wall* and *books in the shelf*.

## More Visual Results



*Remote* detection gets strong votes from *television* and *chair* objects.

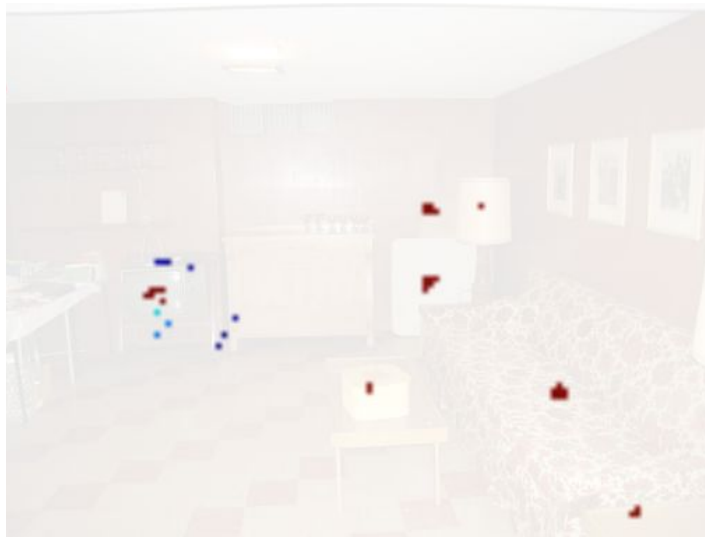
# More Visual Results



*Television* detection gets strong votes from things in a living room such as *lamp* (is not among 80 classes of COCO dataset), *chair* and *couch*.



# More Visual Results



*Television detection gets strong votes from things in a kitchen such as lamp (is not among 80 classes of COCO dataset), and couch.*