CenterNet: Keypoint Triplets for Object Detection

2019.12.4

University of Chinese Academy of Sciences University of Oxford Huawei Noah's Ark Lab

Content

Baseline

what is the baseline for this paper?

The Baseline for this paper is cornernet, which is a key-point-based object detection algorithm that removes the anchor box mechanism compared to the previous object detection.

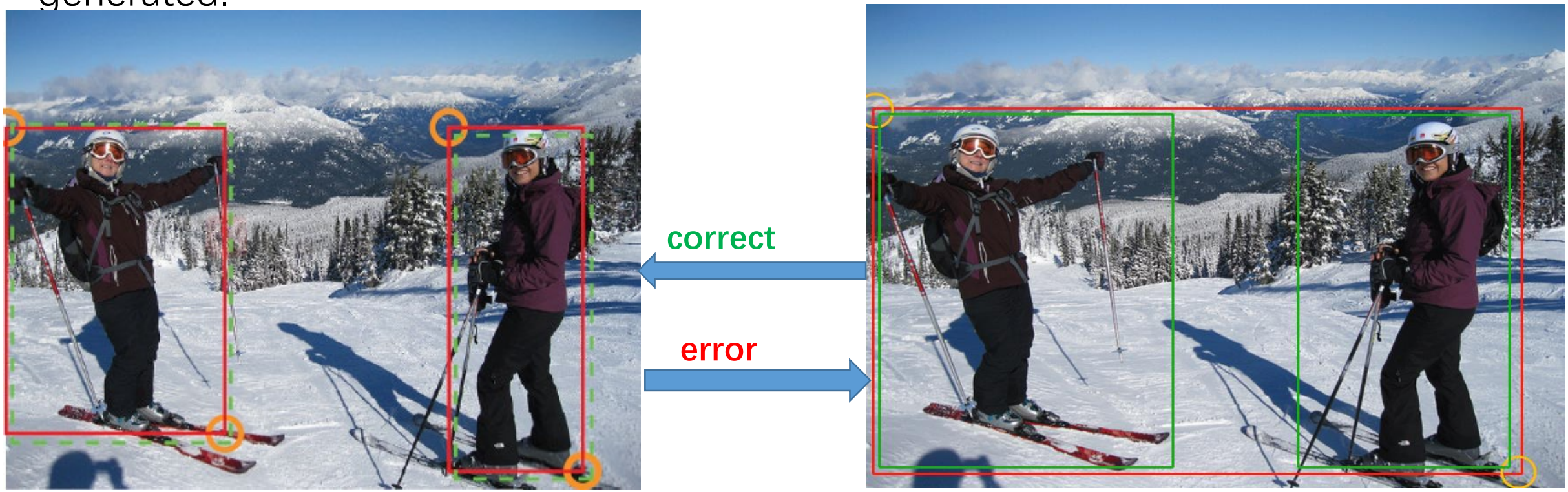
what are the advantages of the baseline?

The advantages of baseline is that there is no need to anchor box a lot and manually set a lot of hyper parameters, which improves the speed of the whole network.

Baseline

what are the disadvantages of the baseline?

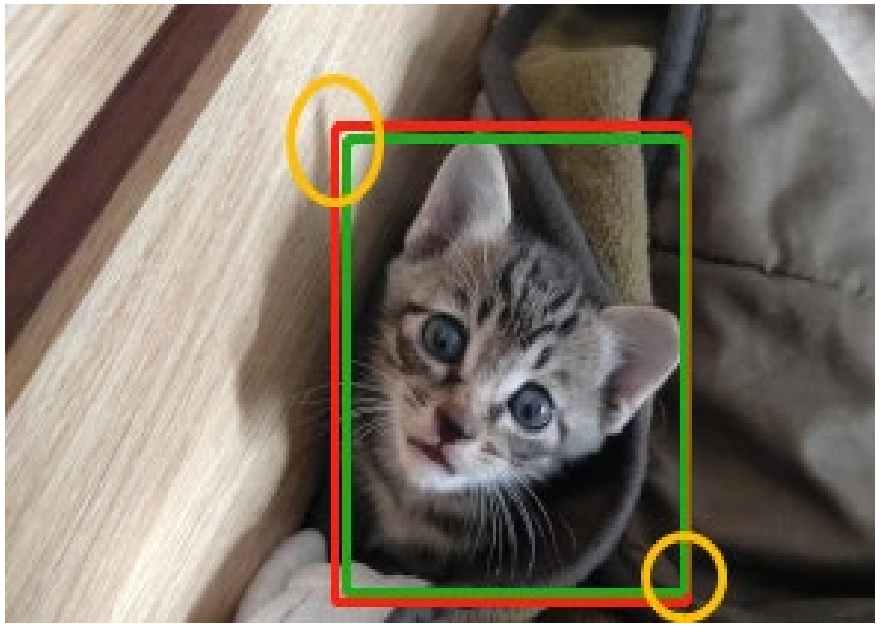
Firstly, without the help of anchor box and global information, any two corner points can form a predicted bounding box, many error bouding boxes will be generated.



Baseline

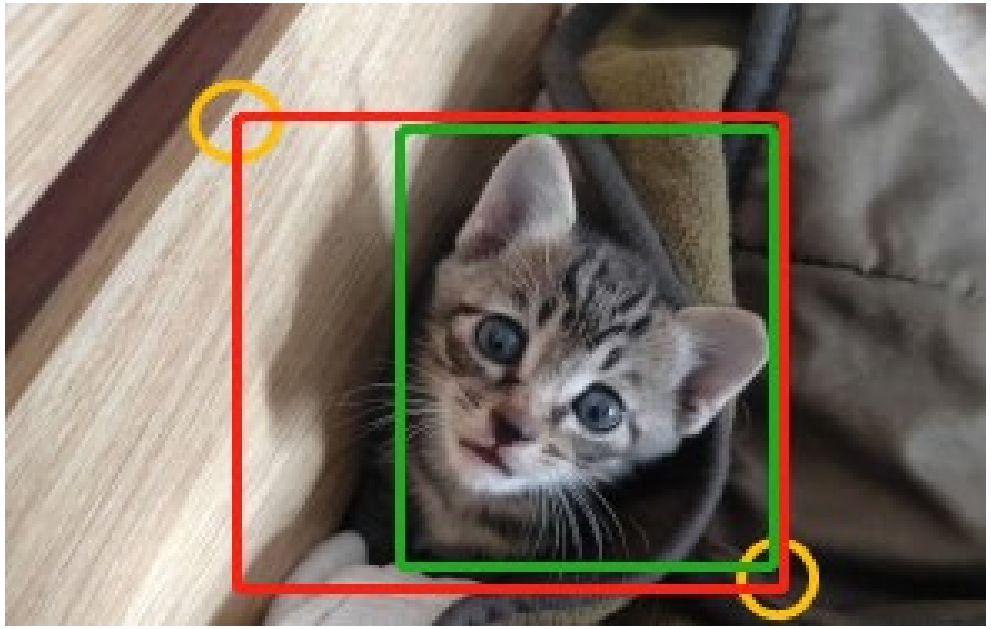
what are the disadvantages of the baseline?

Secondly, baseline is sensitive to edges, which leads to many corner points being sensitive to the edges of the background as well, so the wrong corner points are also detected in the background.



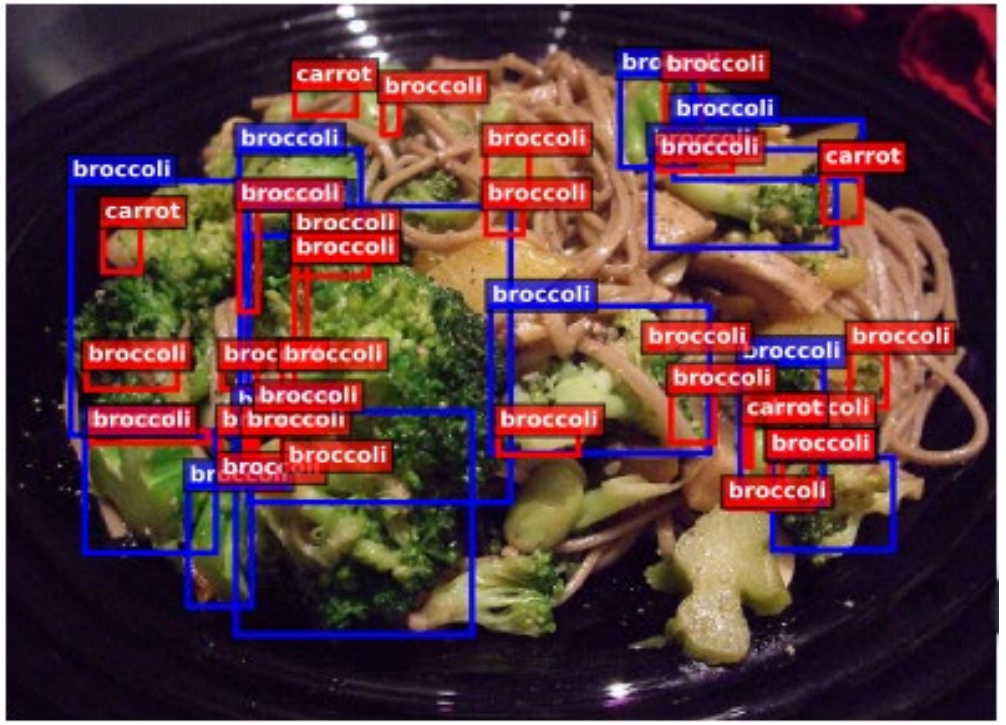
correct

error



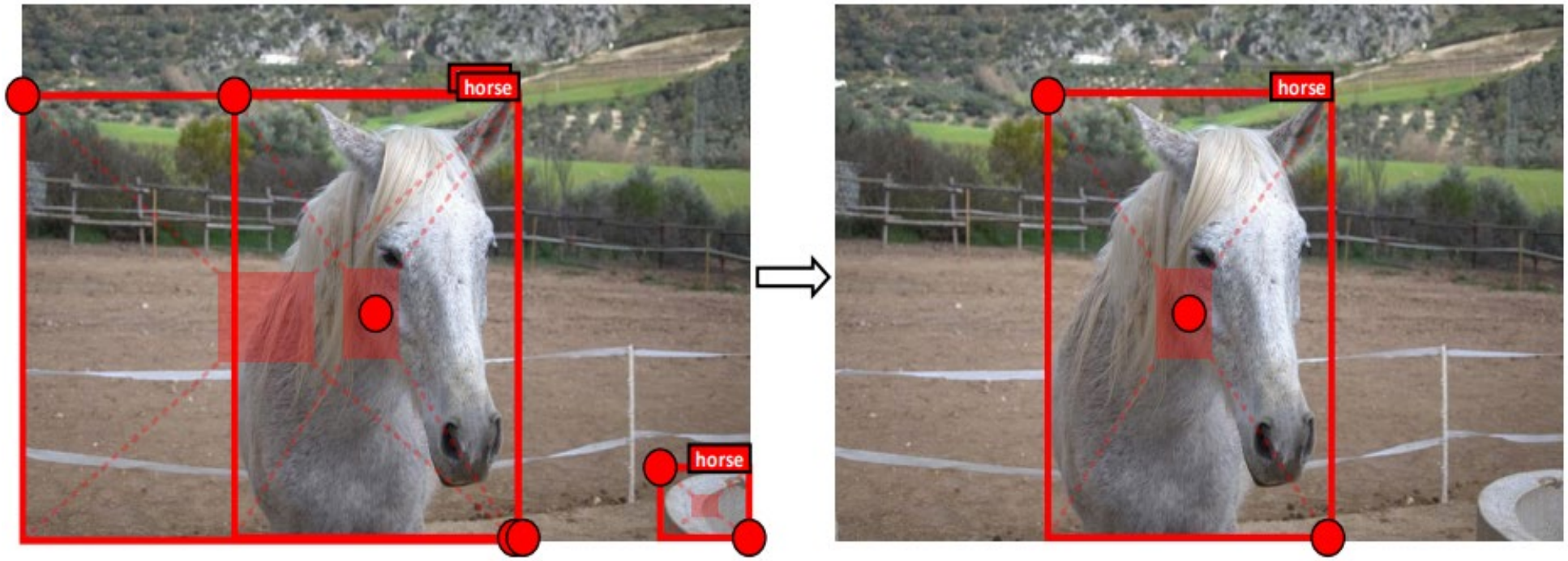
motivation

In object detection, keypoint-based approaches often suffer a large number of incorrect predicted bounding boxes, due to the lack of the internal information of the detect object.



Corresponding solution

intuition : Make good use of the internal information of the detect object



If the predicted bounding box is accurate, the probability of detecting the object center point in its central region of the predicted bounding box is high, and vice versa.

Corresponding solution

specific steps

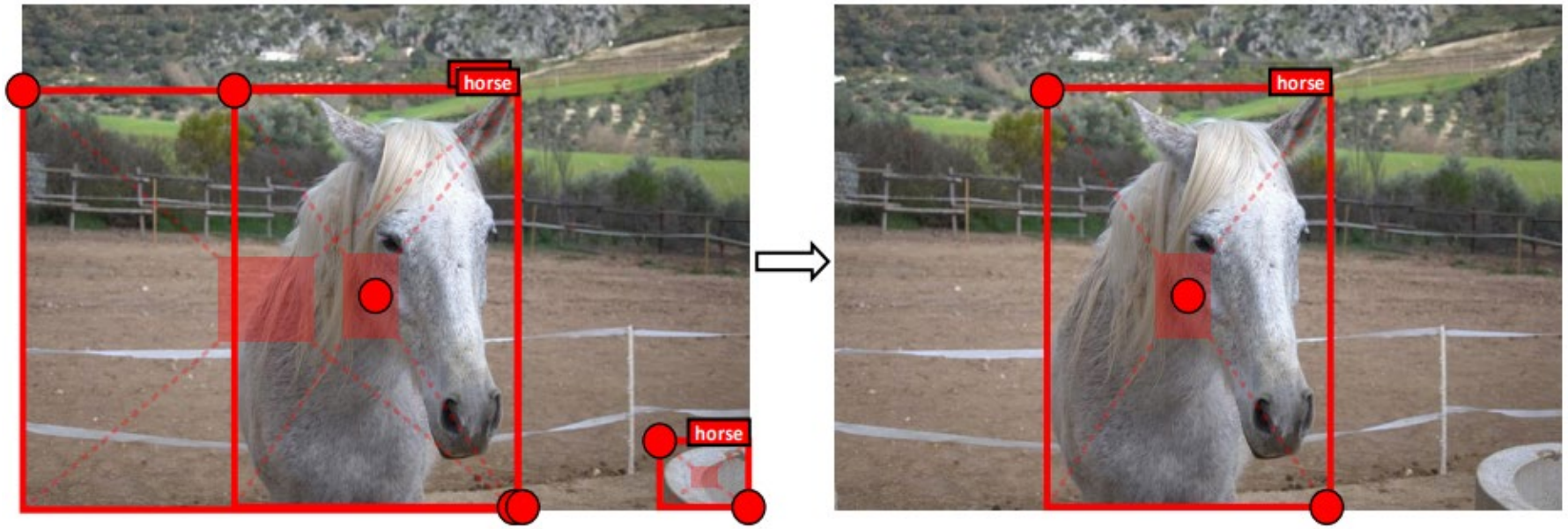
Step1:for each object generate tow corner key points and one center key point

Step2:the tow corner key points are used to generate the predicted bounding box

Step3:define a central region for each predicted bounding box

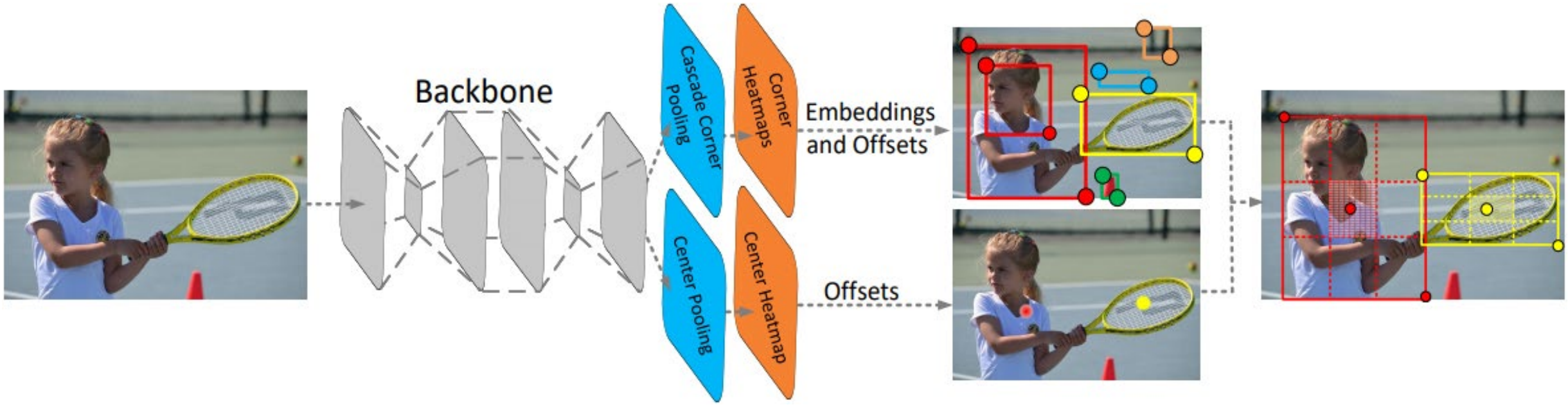
Step4:determine if the central region of each predicted bounding box contains a center key point

Step5:if there is a center key point in the central region of the prediction bounding box, preserve the bounding box ,otherwise remove it.



Our Approach

Centernet.

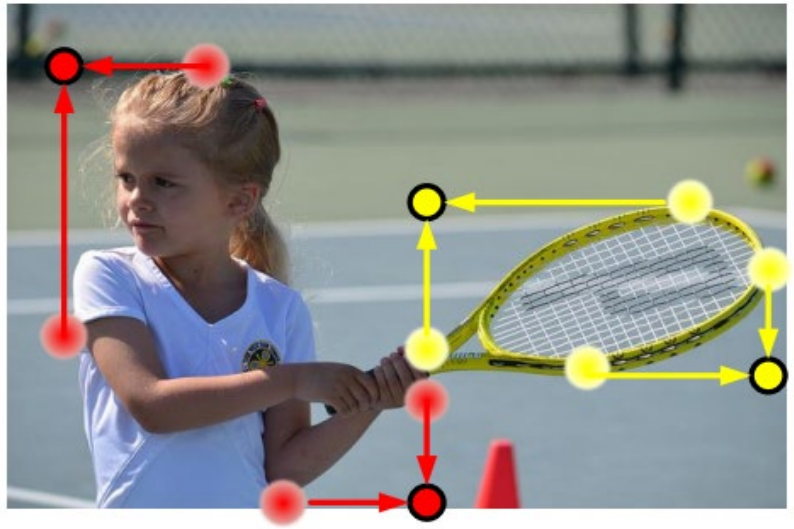


Centernet use cascade corner pooling and center pooling to output two corner keypoints heatmaps and a center keypoint heatmap.

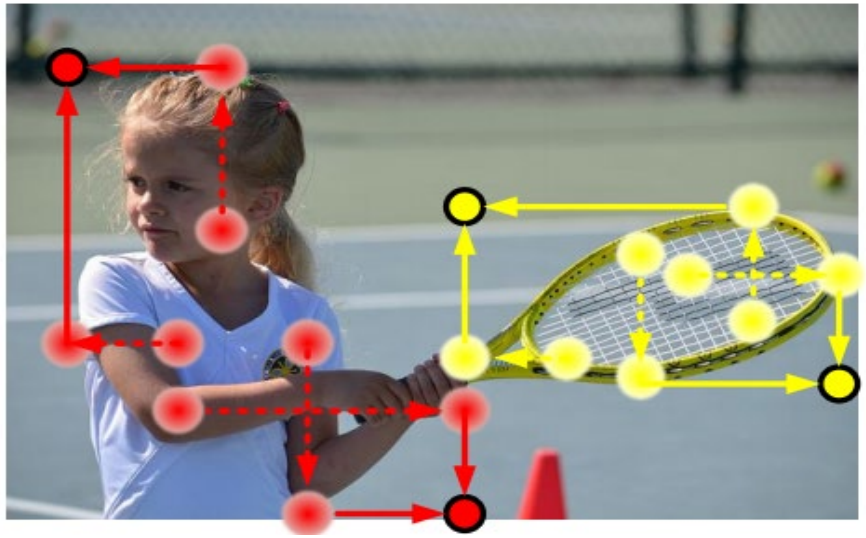
Then the center keypoint are used to determine the final bounding box.

Our Approach

Corner pooling vs Cascade corner pooling



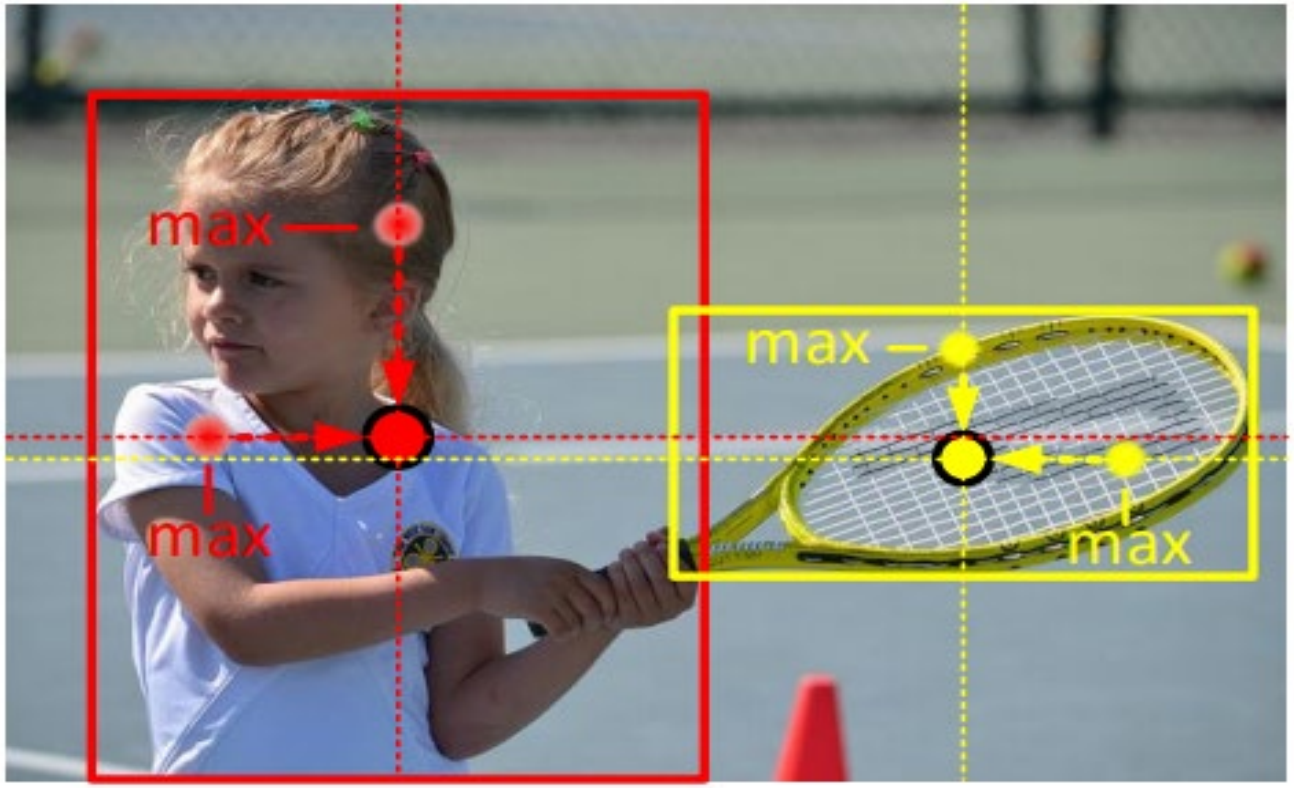
Corner pooling only takes the maximum values in Horizontal and vertical directions



Cascade corner pooling takes the maximum values in both boundary directions and interna directions(along the dotted line in the figure) o objects.

Our Approach

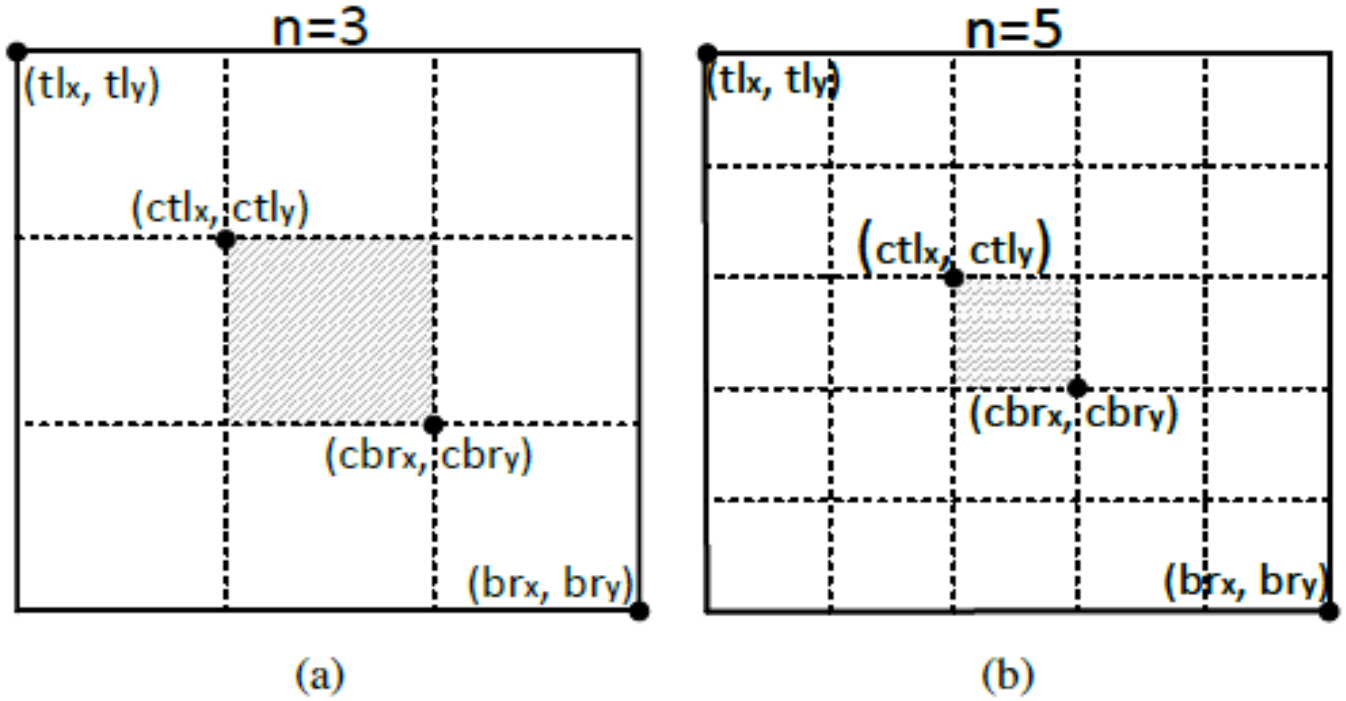
Center pooling



Center pooling takes the maximum values in both horizontal and vertical directions.
provide information other than the location of the center point.

Our Approach

Central region



$$\begin{cases} \mathrm{ctl_x} = \dfrac{(n+1)\mathrm{tl_x} + (n-1)\mathrm{br_x}}{2n} \\ \mathrm{ctl_y} = \dfrac{(n+1)\mathrm{tl_y} + (n-1)\mathrm{br_y}}{2n} \\ \mathrm{cbr_x} = \dfrac{(n-1)\mathrm{tl_x} + (n+1)\mathrm{br_x}}{2n} \\ \mathrm{cbr_y} = \dfrac{(n-1)\mathrm{tl_y} + (n+1)\mathrm{br_y}}{2n} \end{cases}$$

Divide the prediction bounding box into n smaller regions, and take the middle one as the central region

Our Approach

Loss function

$$\mathrm{L} = \mathrm{L}_{\mathrm{det}}^{\mathrm{co}} + \mathrm{L}_{\mathrm{det}}^{\mathrm{ce}} + \alpha \mathrm{L}_{\mathrm{pull}}^{\mathrm{co}} + \beta \mathrm{L}_{\mathrm{push}}^{\mathrm{co}} + \gamma \left(\mathrm{L}_{\mathrm{off}}^{\mathrm{co}} + \mathrm{L}_{\mathrm{off}}^{\mathrm{ce}} \right)$$

$$L_{det} = \frac{-1}{N} \sum_{c=1}^{C} \sum_{i=1}^{H} \sum_{j=1}^{W} \begin{cases} (1 - p_{cij})^{\alpha} \log \left(p_{cij} \right) & \text{if } y_{cij} = 1 \\ (1 - y_{cij})^{\beta} \left(p_{cij} \right)^{\alpha} \log \left(1 - p_{cij} \right) & \text{otherwise} \end{cases}$$

$$L_{pull} = \frac{1}{N} \sum_{k=1}^{N} \left[\left(e_{t_k} - e_k \right)^2 + \left(e_{b_k} - e_k \right)^2 \right],$$

$$L_{push} = \frac{1}{N(N-1)} \sum_{k=1}^{N} \sum_{\substack{j=1 \\ j \neq k}}^{N} \max \left(0, \Delta - |e_k - e_j| \right),$$

$$L_{off} = \frac{1}{N} \sum_{k=1}^{N} \mathrm{SmoothL1Loss} \left(\boldsymbol{o}_k, \hat{\boldsymbol{o}}_k \right)$$

Experiment

In terms of accuracy, the experimental results showed that CenterNet obtained 47% of the AP, surpassing all known one-stage detection methods and leading by a wide margin of at least 4.9%.

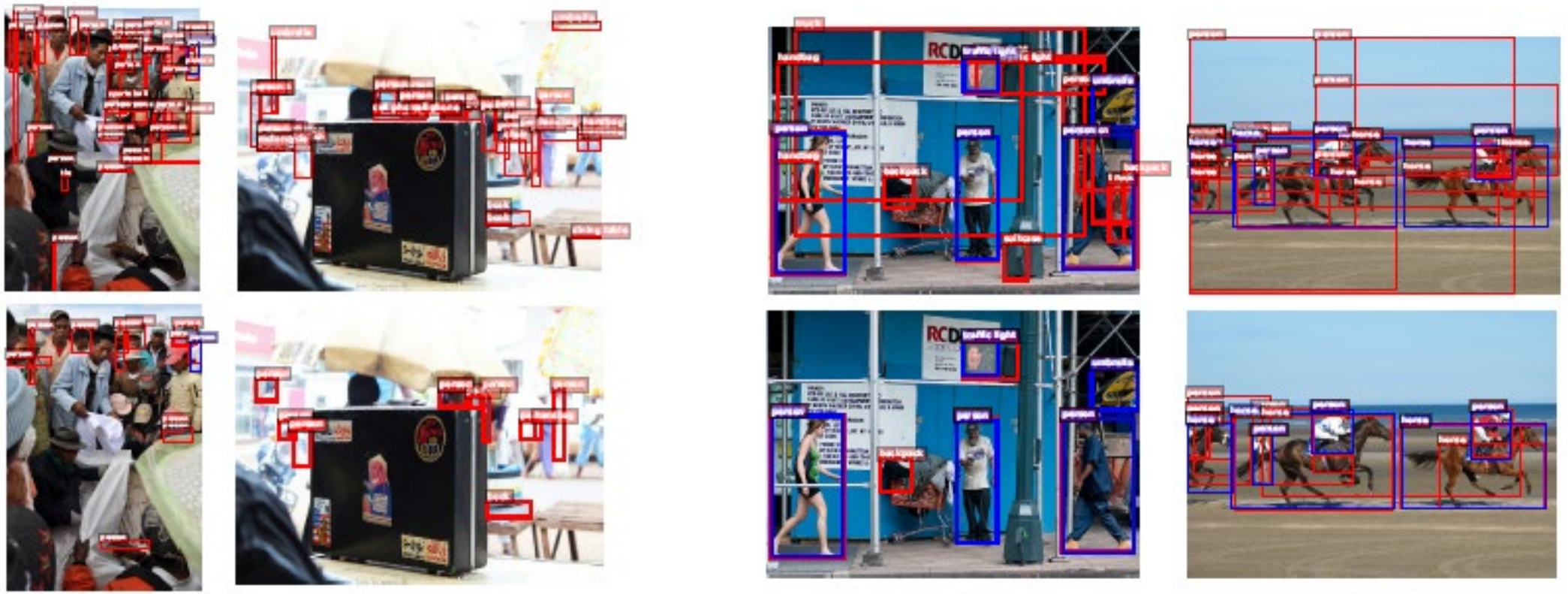
Method	Backbone	Train input	Test input	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_{100}	AR_S	AR_M	AR_L
One-stage:															
YOLOv2 [32]	DarkNet-19	544×544	544×544	21.6	44.0	19.2	5.0	22.4	35.5	20.7	31.6	33.3	9.8	36.5	54.4
DSOD300 [34]	DS/64-192-48-1	300×300	300×300	29.3	47.3	30.6	9.4	31.5	47.0	27.3	40.7	43.0	16.7	47.1	65.0
GRP-DSOD320 [35]	DS/64-192-48-1	320×320	320×320	30.0	47.9	31.8	10.9	33.6	46.3	28.0	42.1	44.5	18.8	49.1	65.0
SSD513 [27]	ResNet-101	513×513	513×513	31.2	50.4	33.3	10.2	34.5	49.8	28.3	42.1	44.4	17.6	49.2	65.8
DSSD513 [8]	ResNet-101	513×513	513×513	33.2	53.3	35.2	13.0	35.4	51.1	28.9	43.5	46.2	21.8	49.1	66.4
RefineDet512 (single-scale) [45]	ResNet-101	512×512	512×512	36.4	57.5	39.5	16.6	39.9	51.4	-	-	-	-	-	-
CornerNet511 (single-scale) [20]	Hourglass-52	511×511	ori.	37.8	53.7	40.1	17.0	39.0	50.5	33.9	52.3	57.0	35.0	59.3	74.7
RetinaNet800 [24]	ResNet-101	800×800	800×800	39.1	59.1	42.3	21.8	42.7	50.2	-	-	-	-	-	-
CornerNet511 (multi-scale) [20]	Hourglass-52	511×511	≤1.5×	39.4	54.9	42.3	18.9	41.2	52.7	35.0	53.5	57.7	36.1	60.1	75.1
CornerNet511 (single-scale) [20]	Hourglass-104	511×511	ori.	40.5	56.5	43.1	19.4	42.7	53.9	35.3	54.3	59.1	37.4	61.9	76.9
RefineDet512 (multi-scale) [45]	ResNet-101	512×512	≤2.25×	41.8	62.9	45.7	25.6	45.1	54.1						
CornerNet511 (multi-scale) [20]	Hourglass-104	511×511	≤1.5×	42.1	57.8	45.3	20.8	44.8	56.7	36.4	55.7	60.0	38.5	62.7	77.4
CenterNet511 (single-scale)	Hourglass-52	511×511	ori.	41.6	59.4	44.2	22.5	43.1	54.1	34.8	55.7	60.1	38.6	63.3	76.9
CenterNet511 (single-scale)	Hourglass-104	511×511	ori.	44.9	62.4	48.1	25.6	47.4	57.4	36.1	58.4	63.3	41.3	67.1	80.2
CenterNet511 (multi-scale)	Hourglass-52	511×511	≤1.8×	43.5	61.3	46.7	25.3	45.3	55.0	36.0	57.2	61.3	41.4	64.0	76.3
CenterNet511 (multi-scale)	Hourglass-104	511×511	≤1.8×	**47.0**	**64.5**	**50.7**	**28.9**	**49.9**	**58.9**	**37.5**	**60.3**	**64.8**	**45.1**	**68.3**	**79.7**

Experiment

Centernet ranks among the top of state-of-the-art two-stage detectors.

Method	Backbone	Train input	Test input	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	AR_1	AR_{10}	AR_{100}	AR_S	AR_M	AR_L
Two-stage:															
DeNet [40]	ResNet-101 [14]	512×512	512×512	33.8	53.4	36.1	12.3	36.1	50.8	29.6	42.6	43.5	19.2	46.9	64.3
CoupleNet [47]	ResNet-101	ori.	ori.	34.4	54.8	37.2	13.4	38.1	50.8	30.0	45.0	46.4	20.7	53.1	68.5
Faster R-CNN by G-RMI [16]	Inception-ResNet-v2 [39]	~ 1000×600	~ 1000×600	34.7	55.5	36.7	13.5	38.1	52.0	-	-	-	-	-	-
Faster R-CNN +++ [14]	ResNet-101	~ 1000×600	~ 1000×600	34.9	55.7	37.4	15.6	38.7	50.9	-	-	-	-	-	-
Faster R-CNN w/ FPN [23]	ResNet-101	~ 1000×600	~ 1000×600	36.2	59.1	39.0	18.2	39.0	48.2	-	-	-	-	-	-
Faster R-CNN w/ TDM [37]	Inception-ResNet-v2	-	-	36.8	57.7	39.2	16.2	39.8	52.1	**31.6**	**49.3**	**51.9**	**28.1**	**56.6**	**71.1**
D-FCN [7]	Aligned-Inception-ResNet	~ 1000×600	~ 1000×600	37.5	58.0	-	19.4	40.1	52.5	-	-	-	-	-	-
Regionlets [43]	ResNet-101	~ 1000×600	~ 1000×600	39.3	59.8	-	21.7	43.7	50.9	-	-	-	-	-	-
Mask R-CNN [12]	ResNeXt-101	~ 1300×800	~ 1300×800	39.8	62.3	43.4	22.1	43.2	51.2	-	-	-	-	-	-
Soft-NMS [2]	Aligned-Inception-ResNet	~ 1300×800	~ 1300×800	40.9	62.8	-	23.3	43.6	53.3	-	-	-	-	-	-
Fitness R-CNN [41]	ResNet-101	512×512	1024×1024	41.8	60.9	44.9	21.5	45.0	57.5	-	-	-	-	-	-
Cascade R-CNN [4]	ResNet-101	-	-	42.8	62.1	46.3	23.7	45.5	55.2	-	-	-	-	-	-
Grid R-CNN w/ FPN [28]	ResNeXt-101	~ 1300×800	~ 1300×800	43.2	63.0	46.6	25.1	46.5	55.2	-	-	-	-	-	-
D-RFCN + SNIP (multi-scale) [38]	DPN-98 [5]	~ 2000×1200	~ 2000×1200	45.7	**67.3**	51.1	29.3	48.8	57.1	-	-	-	-	-	-
PANet (multi-scale) [26]	ResNeXt-101	~ 1400×840	~ 1400×840	**47.4**	67.2	**51.8**	**30.1**	**51.7**	**60.0**	-	-	-	-	-	-
CenterNet511 (single-scale)	Hourglass-52	511×511	ori.	41.6	59.4	44.2	22.5	43.1	54.1	34.8	55.7	60.1	38.6	63.3	76.9
CenterNet511 (single-scale)	Hourglass-104	511×511	ori.	44.9	62.4	48.1	25.6	47.4	57.4	36.1	58.4	63.3	41.3	67.1	80.2
CenterNet511 (multi-scale)	Hourglass-52	511×511	≤1.8×	43.5	61.3	46.7	25.3	45.3	55.0	36.0	57.2	61.3	41.4	64.0	76.3
CenterNet511 (multi-scale)	Hourglass-104	511×511	≤1.8×	**47.0**	**64.5**	**50.7**	**28.9**	**49.9**	**58.9**	**37.5**	**60.3**	**64.8**	**45.1**	**68.3**	**79.7**

Experiment



(a) (b) (c) (d)

(a)and (b) show that CenterNet can effectively remove the small error predicted bbox

(c) and (d) show that CenterNet can effectively remove the middle error predicted bbox

Thanks