RepPoints: Point Set Representation for Object Detection

Ze Yang* yangze@pku.edu.cn
Shaohui Liu* b1ueber2y@gmail.com
Han Hu hanhu@microsoft.com
Liwei Wang wanglw@cis.pku.edu.cn
Stephen Lin stevelin@microsoft.com

Code link: https://github.com/microsoft/RepPoints

ICCV 2019 Seoul, Korea

Motivation

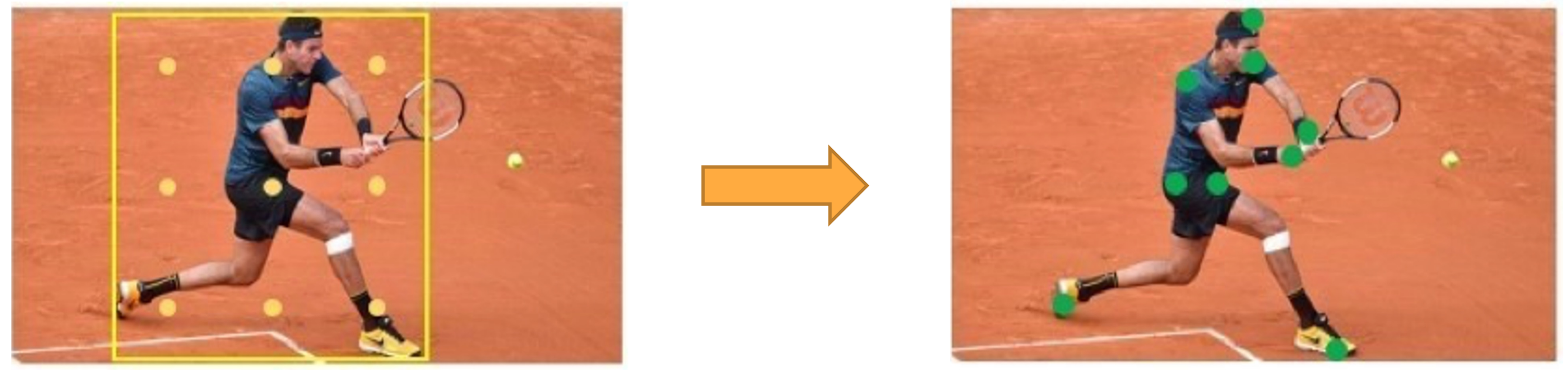
Why bounding box?

- Bounding box is convenient to annotate with little ambiguity.
- It is convenient to use the bounding box representation to facilitate feature extraction.

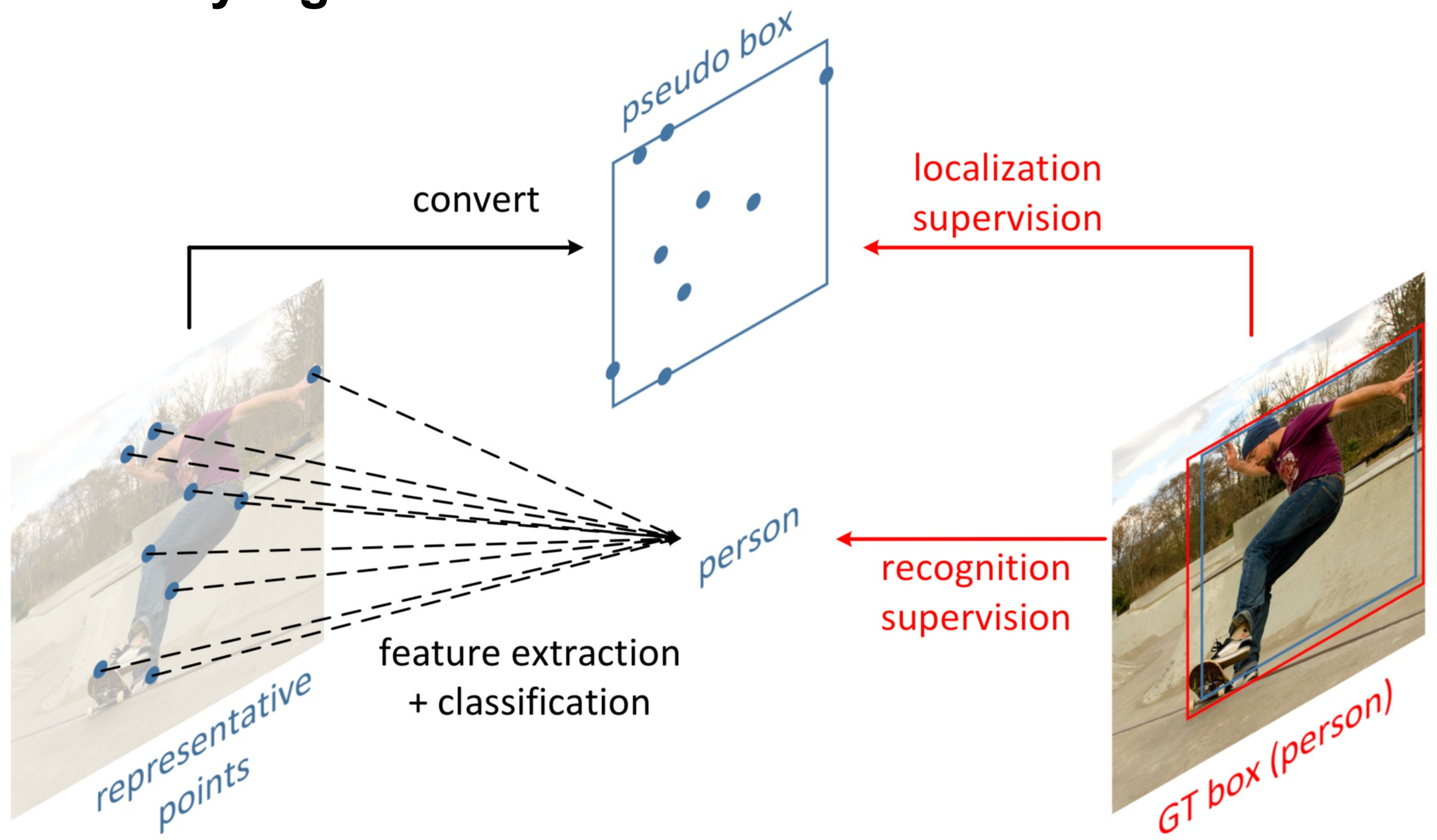
Limitations.

- Coarse object feature extraction.
- Unable to tackle irregular object (like road).

semantic + geometric representation



RepPoints is a new representation for object detection that consists of a set of points which indicate the **spatial extent** of an object and **semantically significant** local areas.



Methodology

How to transform point set to bounding box?

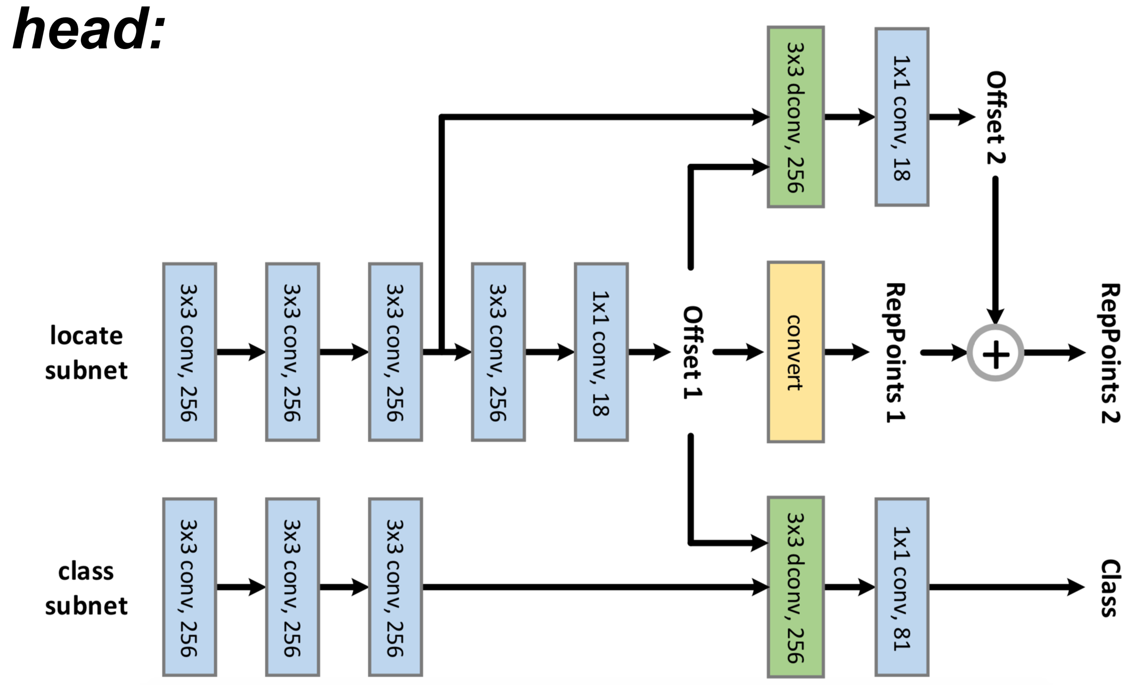
- *Min-max function*: Min-max operation over both axes are performed to acquire rectangular box.
- *Moment-based function*. The first-order moment statistics are used to estimate the center points. The second-order moment statistics are used to estimate the scale of rectangular box, where the scale is multiplied by globally shared learnable multipliers.

Refinement: RepPoints vs Bounding Box.

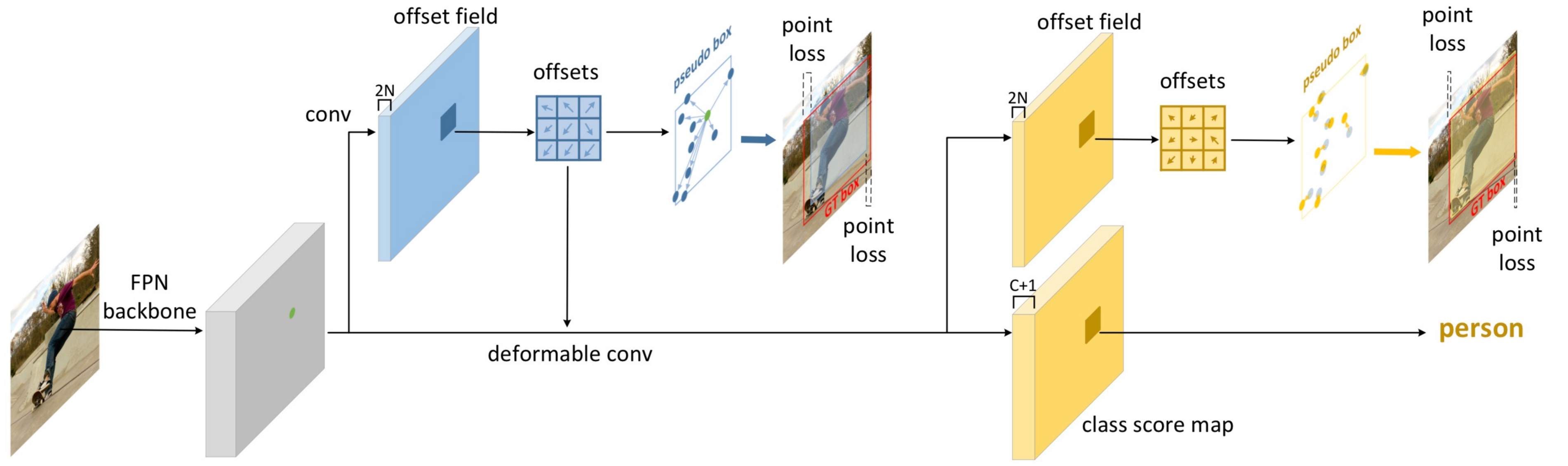
$$\mathcal{R}_r = \{(x_k + \Delta x_k, y_k + \Delta y_k)\}_{k=1}^{n} \quad \textbf{vs} \quad \mathcal{B}_r = (x_p + w_p \Delta x_p, y_p + h_p \Delta y_p, w_p e^{\Delta w_p}, h_p e^{\Delta h_p})$$

2N degree of freedom — 4 degree of freedom

RepPoints head:



Pipeline:



Experiments

Quantitative.

	Backbone	Anchor-Free	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Faster R-CNN w. FPN [24]	ResNet-101		36.2	59.1	39.0	18.2	39.0	48.2
RefineDet [46]	ResNet-101		36.4	57.5	39.5	16.6	39.9	51.4
RetinaNet [25]	ResNet-101		39.1	59.1	42.3	21.8	42.7	50.2
Deep Regionlets [44]	ResNet-101		39.3	59.8	-	21.7	43.7	50.9
Mask R-CNN [13]	ResNeXt-101		39.8	62.3	43.4	22.1	43.2	51.2
FSAF [50]	ResNet-101		40.9	61.5	44.0	24.0	44.2	51.3
Cascade R-CNN [2]	ResNet-101		42.8	62.1	46.3	23.7	45.5	55.2
CornerNet [21]	Hourglass-104	✓	40.5	56.5	43.1	19.4	42.7	53.9
ExtremeNet [48]	Hourglass-104	✓	40.1	55.3	43.2	20.3	43.2	53.1
RPDet	ResNet-101	✓	41.0	62.9	44.3	23.6	44.1	51.7
RPDet	ResNet-101-DCN	✓	**42.8**	**65.0**	**46.3**	**24.9**	**46.2**	**54.7**
RPDet (ms train)	ResNet-101-DCN	✓	**45.0**	**66.1**	**49.0**	**26.6**	**48.6**	**57.5**
RPDet (ms train & ms test)	ResNet-101-DCN	✓	**46.5**	**67.4**	**50.9**	**30.3**	**49.7**	**57.1**

Comparison with state-of-the-art

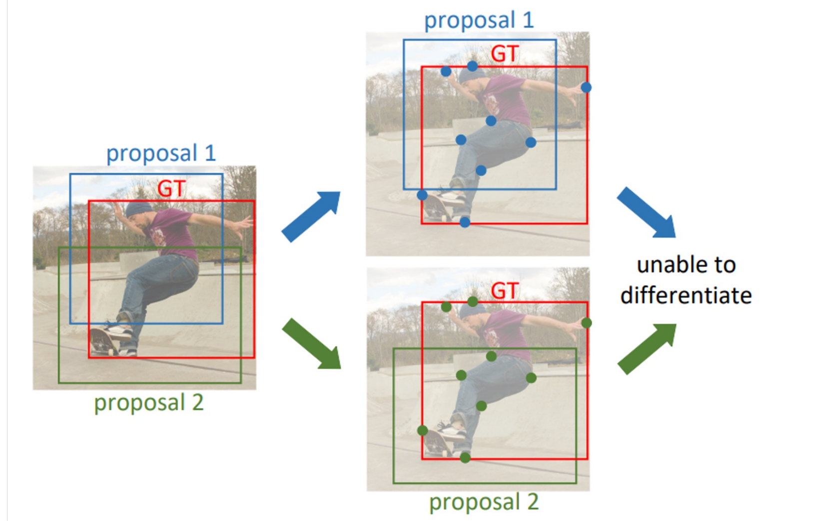
method	backbone	ms train	ms test	AP
RPDet	R-50			38.6
	R-50	✓		40.8
	R-50	✓	✓	42.2
	R-101			40.3
	R-101	✓		42.3
	R-101	✓	✓	44.1
	R-101-DCN			43.0
	R-101-DCN	✓		44.8
	R-101-DCN	✓	✓	46.4
	X-101-DCN			44.5
	X-101-DCN	✓		45.6
	X-101-DCN	✓	✓	46.8

Comparison with different backbone

Qualitative.

Visualization of the learnt RepPoints and the induced bounding boxes on several examples from the COCO [26] minival set (using the pseudo box converting function T_1). In general, the learned RepPoints are located on extreme or semantic keypoints of objects.

Relation to deformable RoI pooling



Future direction: denser and finer