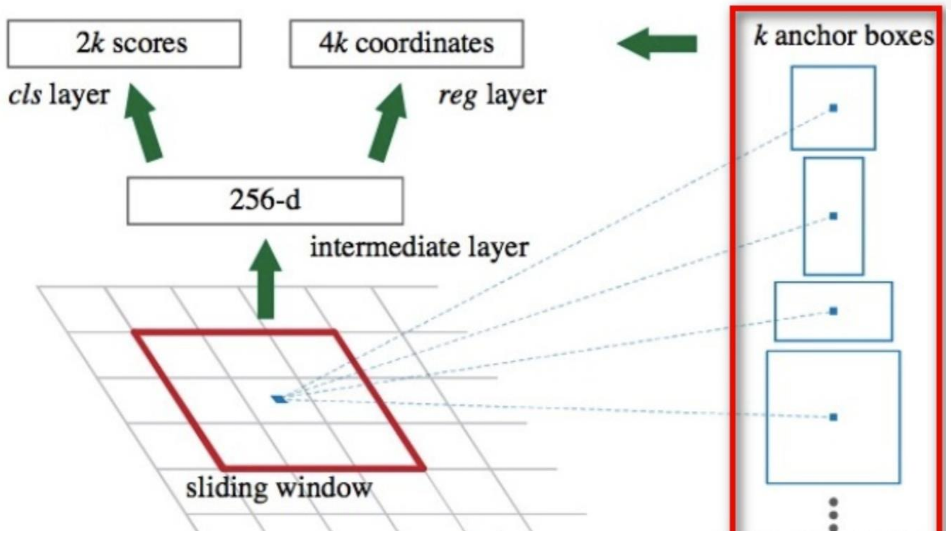
Paper reading

段永杰

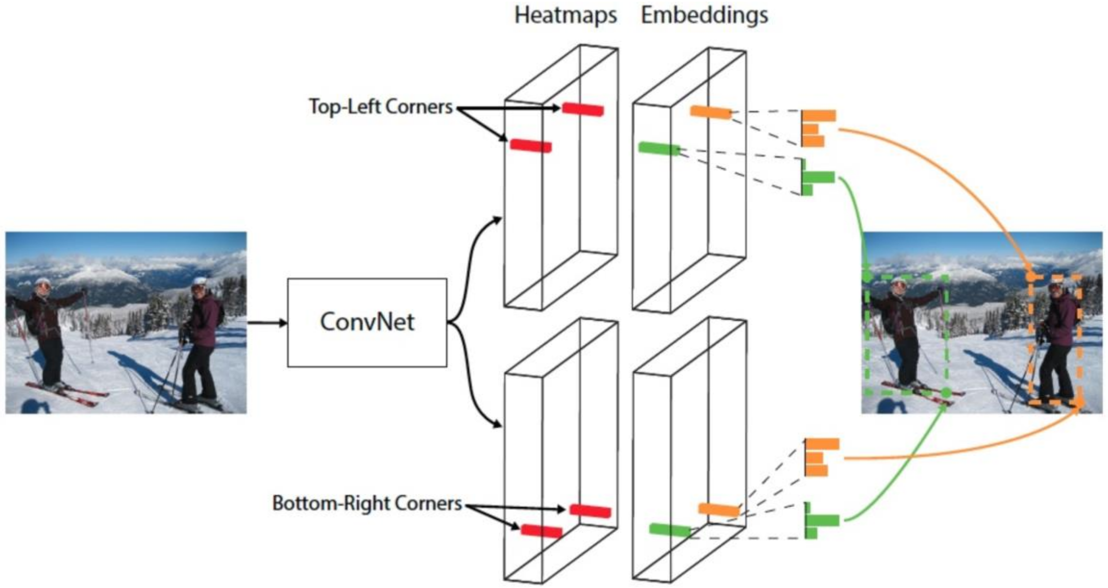
2020/6/11

Anchor related

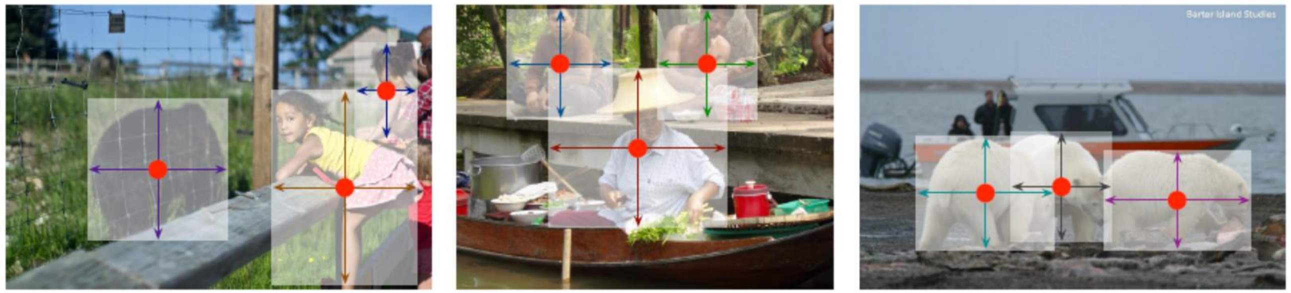
- Anchor-based
  - Yolo V2/V3、Faster-RCNN、SSD
- Anchor-free
  - CenterNet、CornerNet、ExtremeNet、FCOS



Faster-RCNN



CornerNet



CenterNet

Region Proposal by Guided Anchoring

Jiaqi Wang, Kai Chen, Shuo Yang, Chen Change Lo, Dahua Lin

1. CUHK - SenseTime Joint Lab, The Chinese University of Hong Kong
2. Amazon Recognition
3. Nanyang Technological University

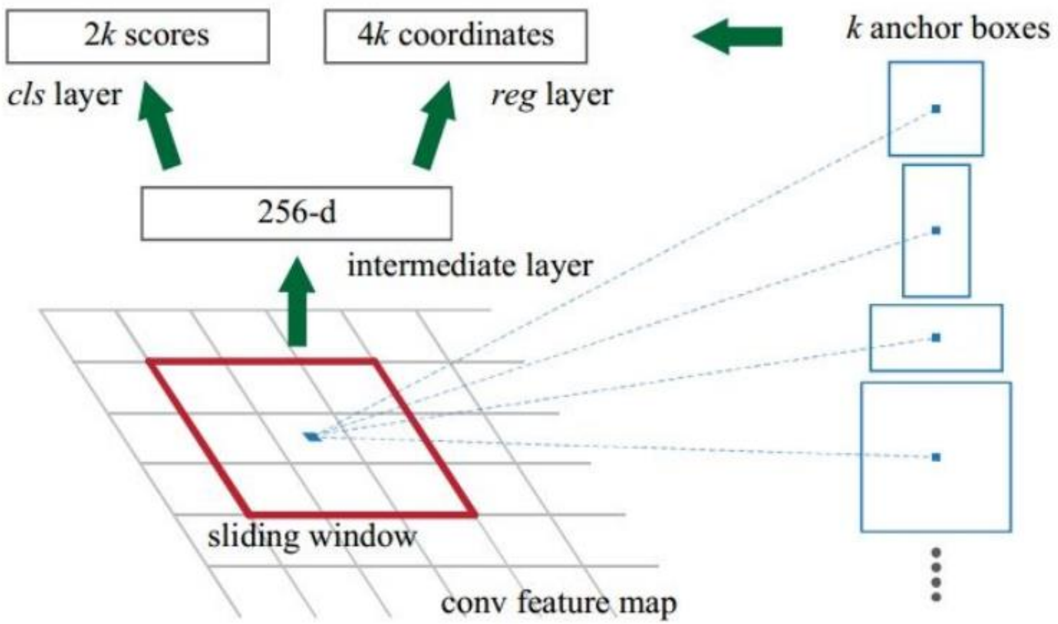
CVPR 2019

Abstract

- Motivation
  - Anchor：物体检测中人为设计的一组基准框
  - 现有anchor-based方法的问题
    - 超参
    - 特殊比例物体
    - Anchor 数目过多
- Contribution
  - Guided anchoring：根据图像特征指导anchor的生成
  - Feature adaption：修正特征图与anchor之间的匹配关系

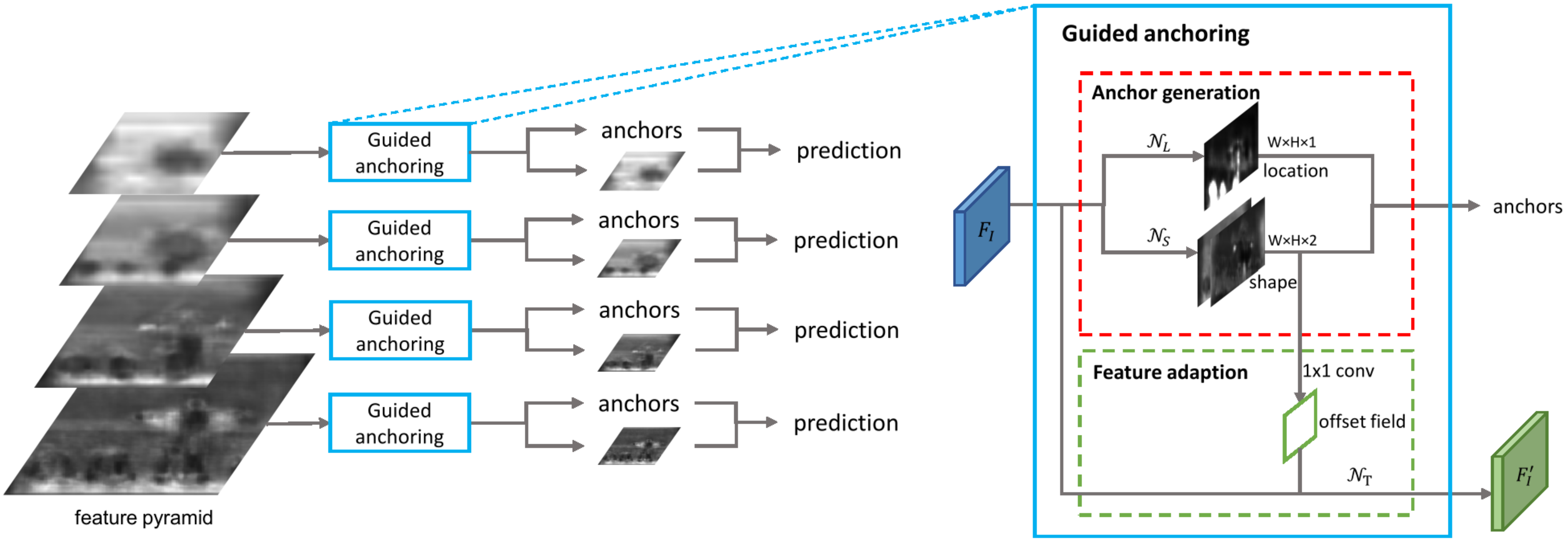
Motivation

- 常见的anchor生成方法
  - 大量数据聚类，存在泛化问题
  - 一系列特定尺度和长宽比，anchor数目对物体比例敏感



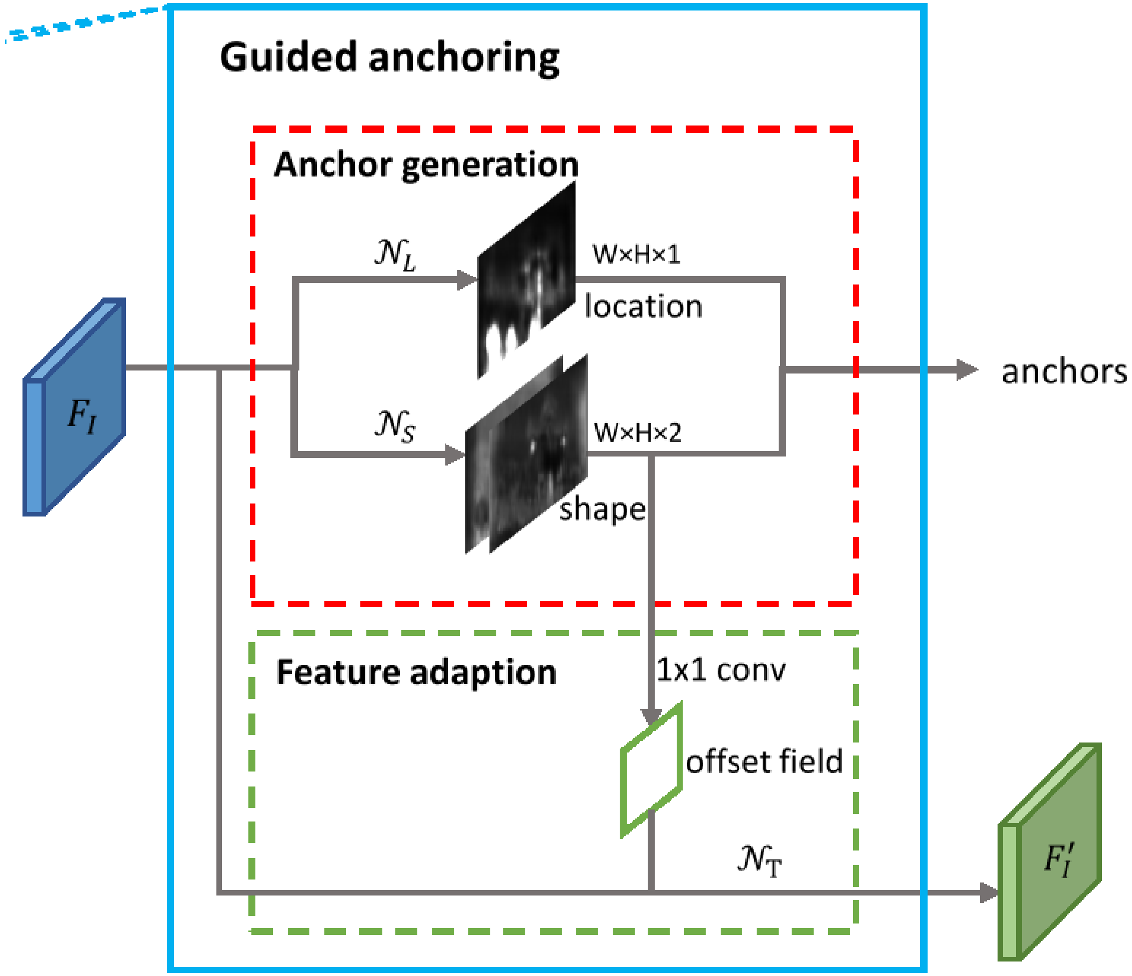
Methods

- 根据图像特征生成稀疏的、形状根据位置可变的anchor



Methods

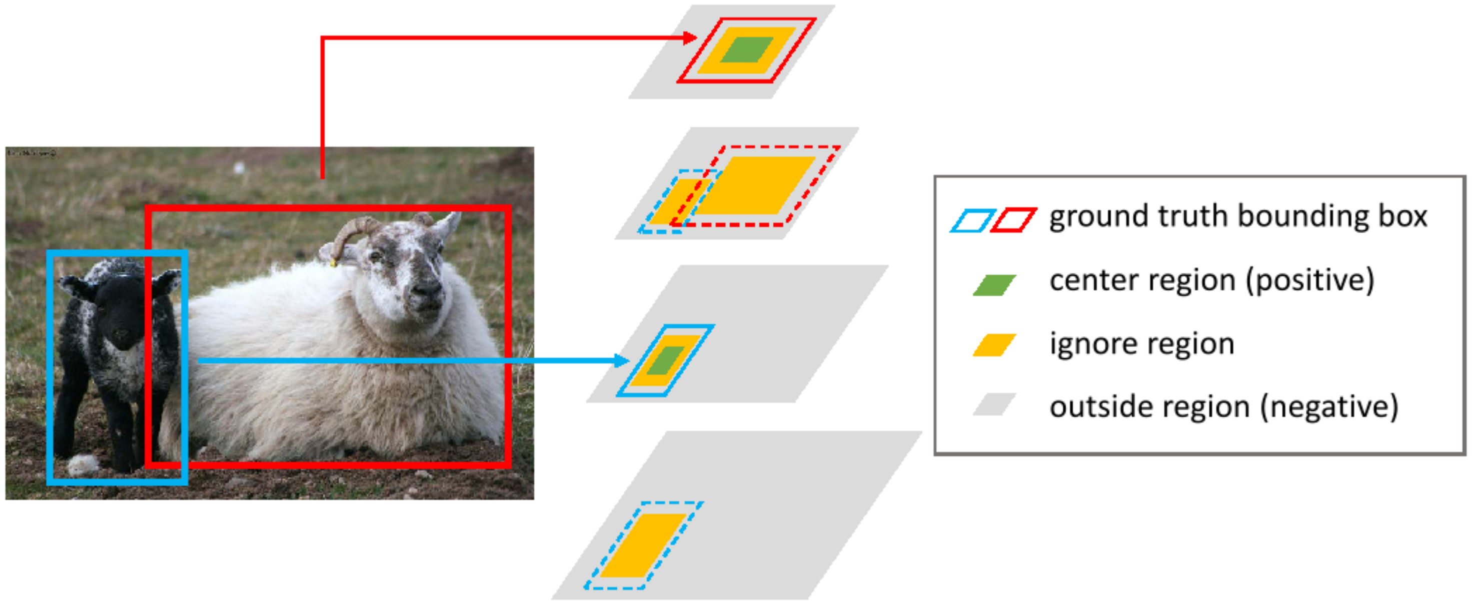
- Anchor位置预测
  - 1x1 conv + sigmoid
  - 滤除大量背景区域
- 训练样本
  - 三种区域采样
- Loss
  - Focal loss
  - 针对小区域、困难样本



Methods

- 训练样本

FPN，多尺度



- 正样本：中心区域，σ_1*bbox
- 忽略：σ_2*bbox 除去中心区域
- 负样本：bbox外部

仅在对应尺度的特征层！
相邻尺度的对应位置忽略

Methods

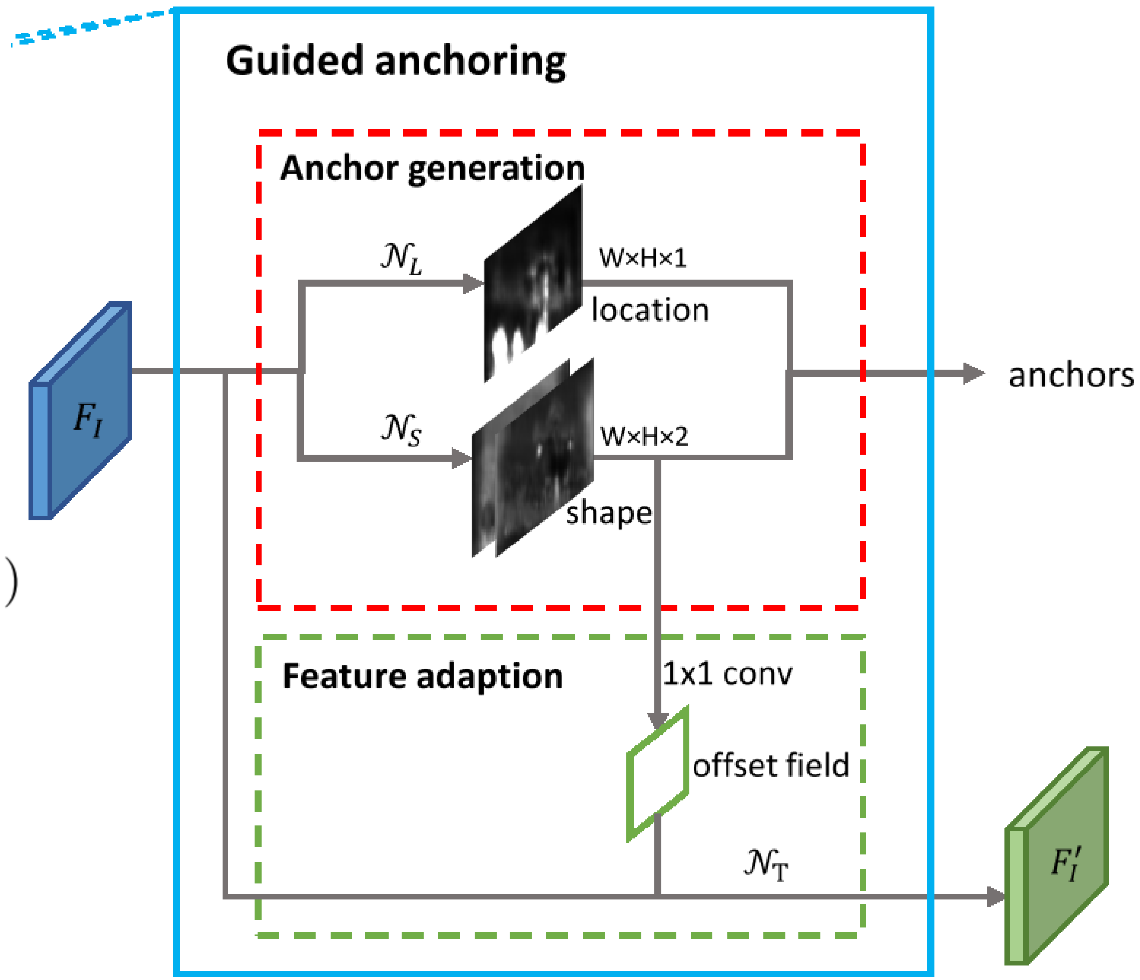
- Anchor形状预测
  - 1x1 conv + transform layer

$$w = \sigma \cdot s \cdot e^{dw}, \quad h = \sigma \cdot s \cdot e^{dh}$$

- Loss函数
  - bounded IoU loss

$$\mathcal{L}_{shape} = \mathcal{L}_1(1 - \min(\frac{w}{w_g}, \frac{w_g}{w})) + \mathcal{L}_1(1 - \min(\frac{h}{h_g}, \frac{h_g}{h}))$$

  - 本质和IoU loss类似



Methods

- Anchor assignment
  - 传统方法：选取与GT的IoU最大的anchor来计算loss反传
  - Anchor的长宽是预测得到的，非固定

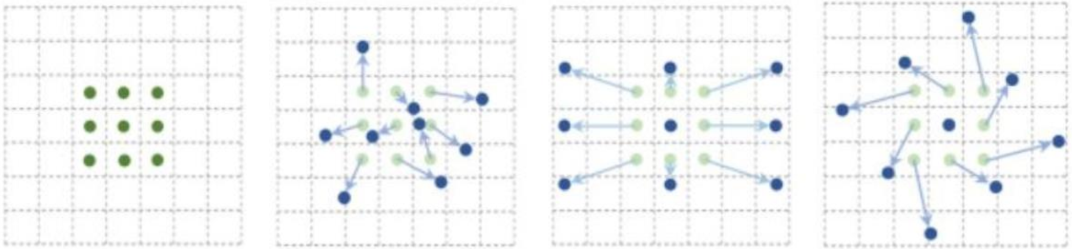
$$\mathrm{vIoU}(a_{\mathbf{wh}}, \mathrm{gt}) = \max_{w>0,h>0} \mathrm{IoU}_{normal}(a_{wh}, \mathrm{gt})$$

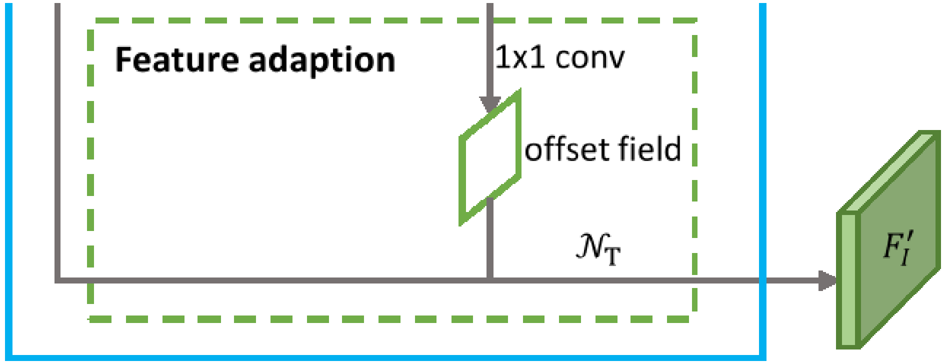
  - 采样方式
    - 采样w和h，文章中设置为9组
    - 计算IoU最大的bbox作为当前anchor的GT
    - 计算loss优化w和h

Methods

- Feature adaption
  - 同一层特征图不同位置，感受野大小/形状相同
  - 使用同一层特征图/卷积核，却代表不同大小/形状的anchor

  - 解决办法
    - 把 anchor 的形状信息直接融入到特征图中
    - 1x1 conv 得到 offset
    - 3x3 deformable conv



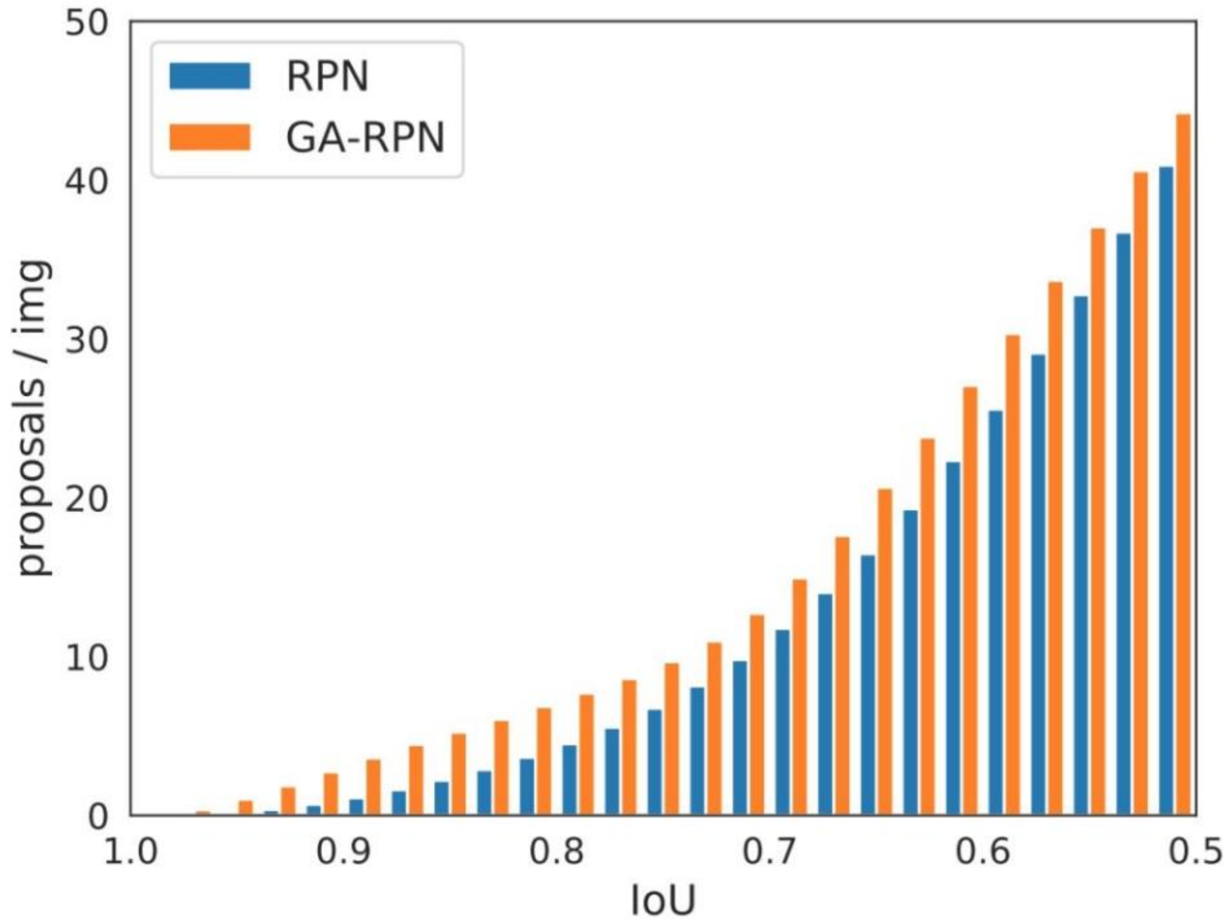


Methods

- 高质量proposal
  - Proposal的整体质量好了很多，但是在检测性能上却没有太大提升
  - 解决办法：
    - 减少proposal数目
    - 提高正样本的IoU阈值（!）

Table 7: Exploration of utilizing high-quality proposals.

proposal	num	IoU thr	AP	AP_{50}	AP_{75}
RPN	1000	0.5	36.7	58.8	39.3
	1000	0.6	37.2	57.1	40.5
	300	0.5	36.1	57.6	39.0
	300	0.6	37.0	56.3	39.5
GA-RPN	1000	0.5	37.4	**59.9**	40.0
	1000	0.6	38.9	59.0	42.4
	300	0.5	37.5	59.6	40.4
	300	0.6	**39.4**	59.3	**43.2**



Results

- 数据集：MS COCO 2017
- Bacbone：resnet-50，FPN
- AR：average recall，AP：average precision

- Detection 结果

Method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Fast R-CNN	37.1	59.6	39.7	20.7	39.5	47.1
GA-Fast-RCNN	**39.4**	59.4	42.8	21.6	41.9	50.4
Faster R-CNN	37.1	59.1	40.1	21.3	39.8	46.5
GA-Faster-RCNN	**39.8**	59.2	43.5	21.8	42.6	50.7
RetinaNet	35.9	55.4	38.8	19.4	38.9	46.5
GA-RetinaNet	**37.1**	56.9	40.0	20.1	40.1	48.0

Results

- Region proposal 结果

Method	Backbone	AR_{100}	AR_{300}	AR_{1000}	AR_S	AR_M	AR_L	runtime (s/img)
SharpMask [24]	ResNet-50	36.4	-	48.2	6.0	51.0	66.5	0.76 (unfair)
GCN-NS [22]	VGG-16 (SyncBN)	31.6	-	60.7	-	-	-	0.10
AttractioNet [10]	VGG-16	53.3	-	66.2	31.5	62.2	77.7	4.00
ZIP [16]	BN-inception	53.9	-	67.0	31.9	63.0	78.5	1.13
RPN	ResNet-50-FPN	47.5	54.7	59.4	31.7	55.1	64.6	**0.09**
	ResNet-152-FPN	51.9	58.0	62.0	36.3	59.8	68.1	0.16
	ResNeXt-101-FPN	52.8	58.7	62.6	37.3	60.8	68.6	0.26
RPN+9 anchors	ResNet-50-FPN	46.8	54.6	60.3	29.5	54.9	65.6	0.09
RPN+Focal Loss [19]	ResNet-50-FPN	50.2	56.6	60.9	33.9	58.2	67.5	0.09
RPN+Bounded IoU Loss [29]	ResNet-50-FPN	48.3	55.1	59.6	33.0	56.0	64.3	0.09
RPN+Iterative	ResNet-50-FPN	49.7	56.0	60.0	34.7	58.2	64.0	0.10
RefineRPN	ResNet-50-FPN	50.2	56.3	60.6	33.5	59.1	66.9	0.11
GA-RPN	ResNet-50-FPN	**59.2**	**65.2**	**68.5**	**40.9**	**67.8**	**79.0**	0.13

Results

- Region proposal 结果

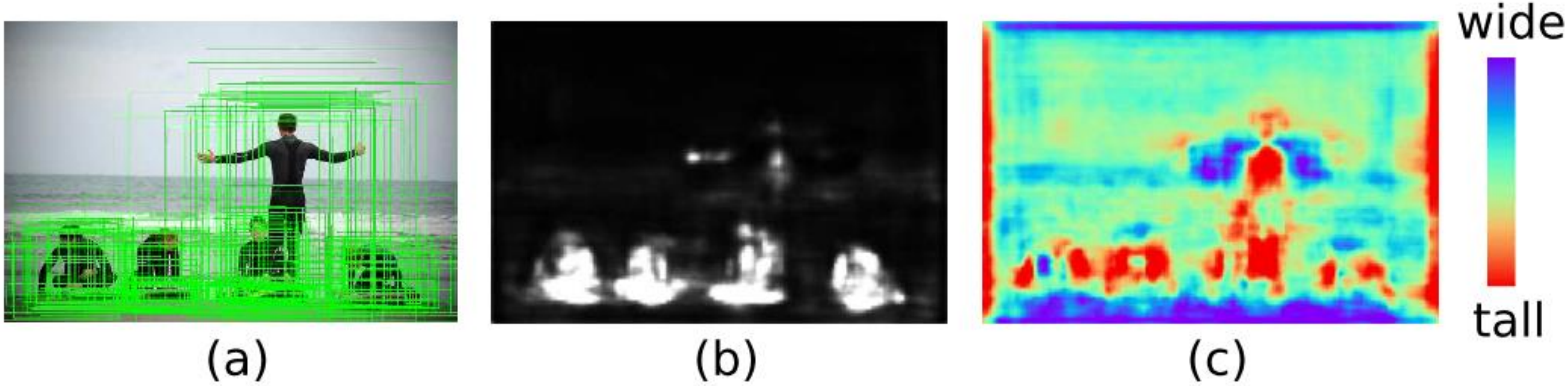
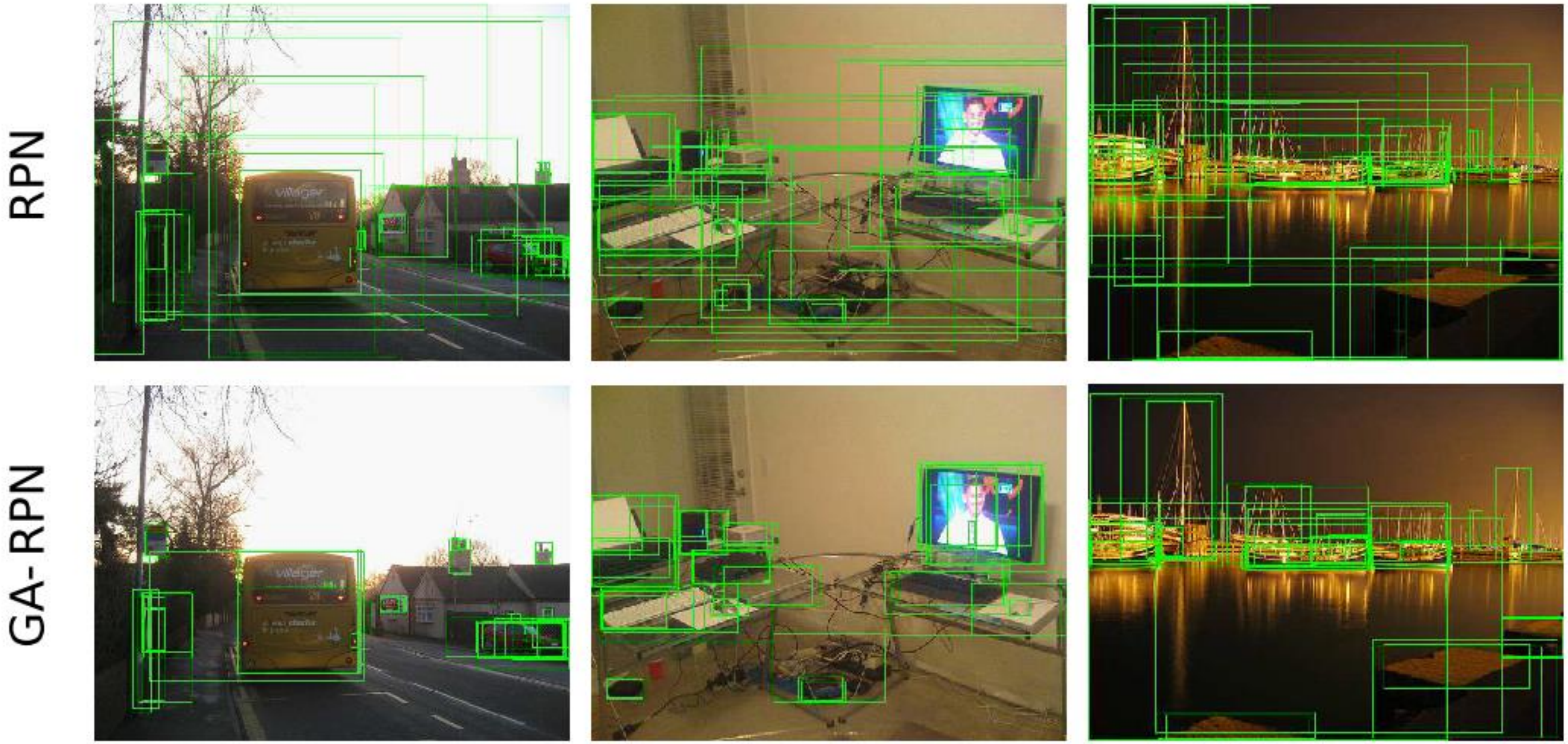


Figure 4: Anchor prediction results. (a) input image and predict anchors; (b) predicted anchor location probability map; (c) predicted anchor aspect ratio.

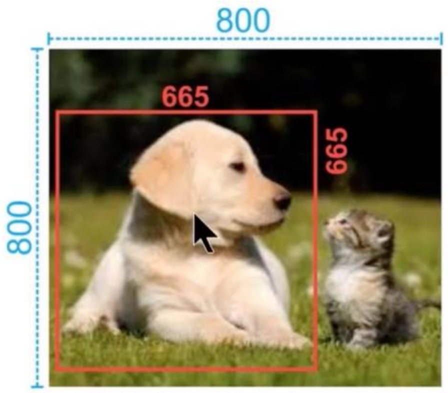
Results

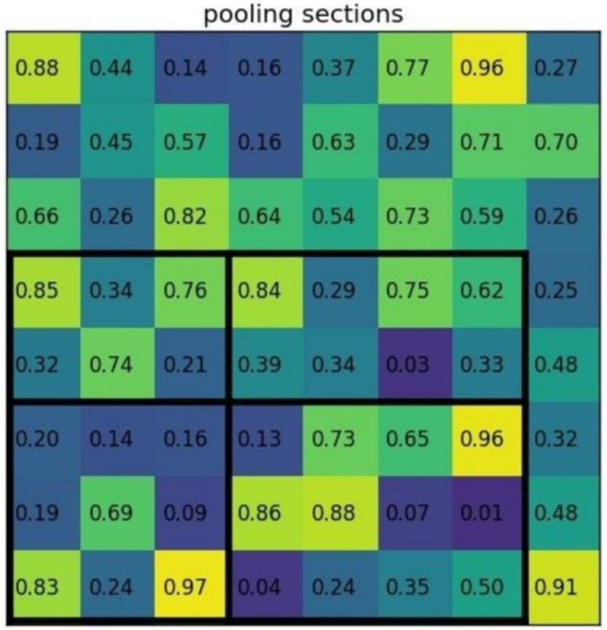
- Region proposal 结果



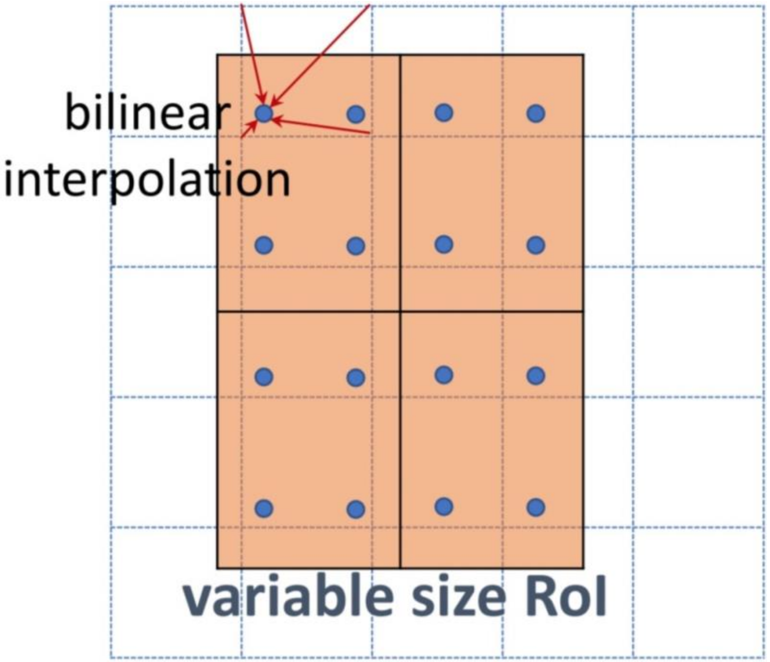
Summary

- Anchor设计准则
  - Alignment
    - 每个anchor的中心就是当前像素点，即不做anchor中心的回归预测（仅预测形状）
  - Consistency
    - 对应两阶段检测器中的ROI pooling或者 ROI align 操作





ROI pooling



ROI align

Summary

- 方法
  - Anchor设计中的alignment和consistency准则，分别对应两个分支及FA
  - 高质量proposal，减少数量并提高IoU阈值
  - Location预测不需要很精确
  - 类似coarse-to-fine的思路
- 局限
  - 相邻很近的物体（FPN可缓解）
  - 极端形状物体（anchor-based方法的问题）

RepPoints: Point Set Representation for Object Detection

Ze Yang[1†*], Shaohui Liu[2,3†*], Han Hu[3], Liwei Wang[1], Stephen Lin[3]

1. Peking University
2. Tsinghua University
3. Microsoft Research Asia

ICCV 2019

Abstract

- Motivation
  - Bounding box：规则且相对固定的框，定位粗糙
  - 定义一组有代表性的点（representive points），对目标更精确表示

- Contribution
  - 通过一组点集提供更细粒度的位置表示和便于分类的信息
  - 摆脱了bounding box 以及 anchor 的限制

Motivation

- 传统的bbox
  - 表示为四维向量：***B* = (*x*, *y*, *w*, *h*)**
  - Feature map ->
  - anchor ->
  - proposal ->
  - ROI pooling/align ->
  - 优化**B**

$$\left(\frac{x_t - x_p}{w_p}, \frac{y_t - y_p}{h_p}, \log\frac{w_t}{w_p}, \log\frac{h_t}{h_p}\right).$$

$$\text{bbox anchors} \xrightarrow{\text{bbox reg.}} \text{bbox proposals (S1)}$$
$$\xrightarrow{\text{bbox reg.}} \text{bbox proposals (S2)}$$
$$\ldots$$
$$\xrightarrow{\text{bbox reg.}} \text{bbox object targets}$$

Methods

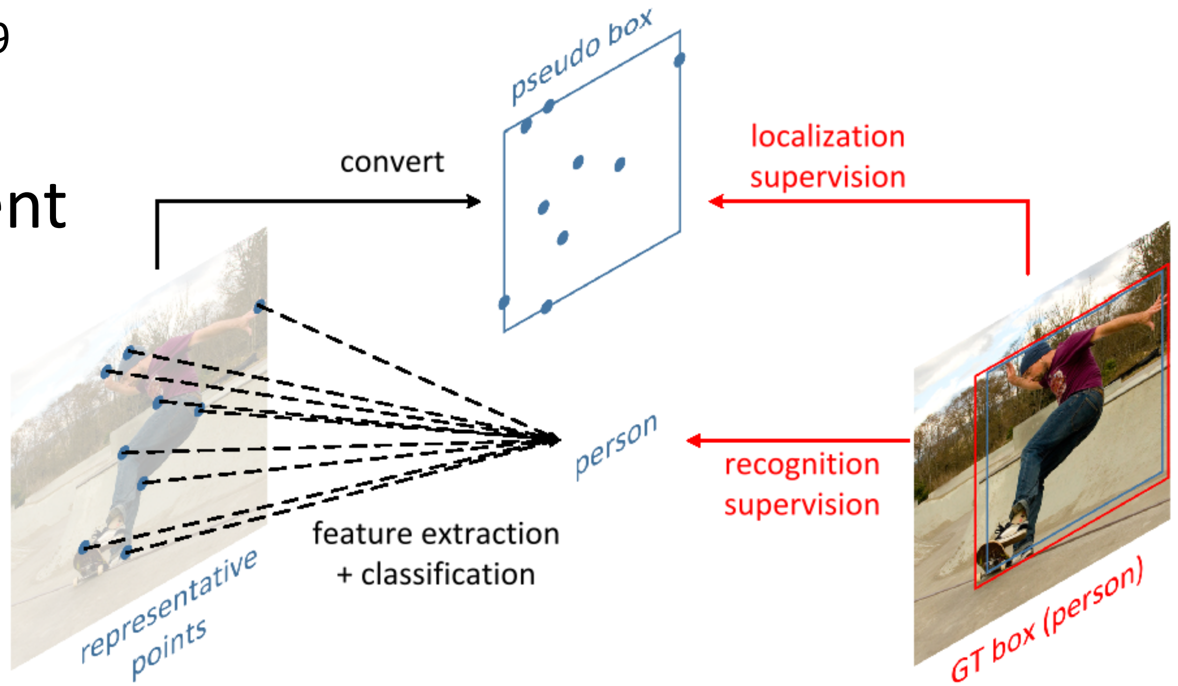
- BBox只考虑目标的矩形空间范围，不考虑形状、姿态和语义上重要的局部区域的位置
- 因此 RepPoints 建立一组自适应的特征点集代替B

$$\mathcal{R} = \{(x_k, y_k)\}_{k=1}^{n}$$

文章设定n=9

- BBox refinement -> RepPoints refinement

$$\mathcal{R}_r = \{(x_k + \Delta x_k, y_k + \Delta y_k)\}_{k=1}^{n}$$



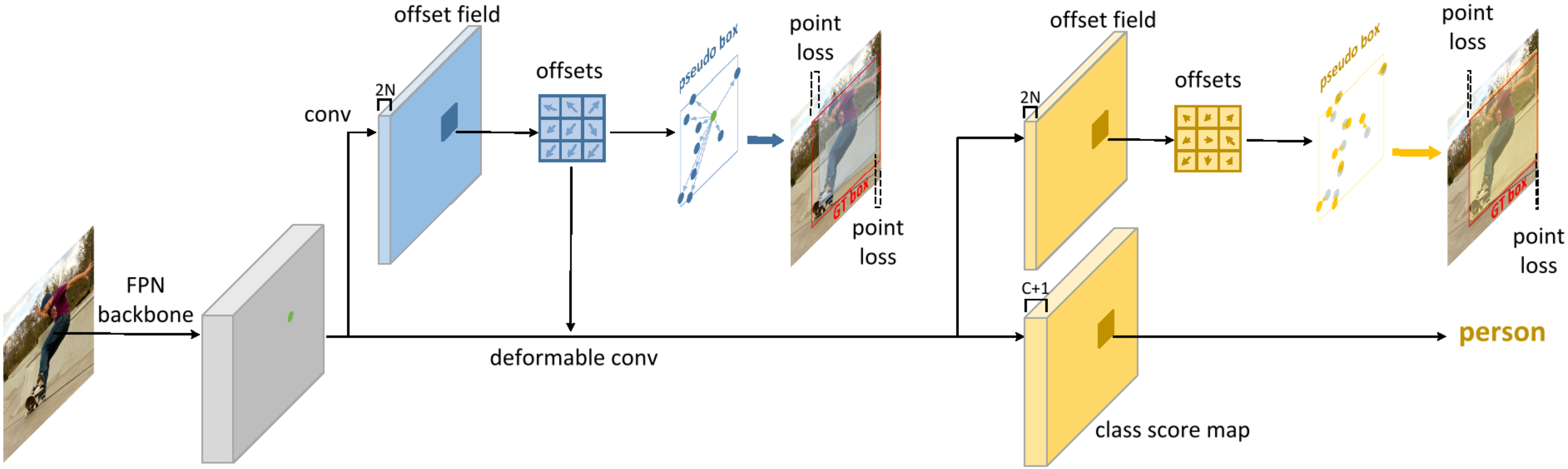
Methods

- RepPoints 转换 pseudo-BBox
  - Min-max function
    - 所有点在XY方向上的极小值与极大值
  - Partial min-max function
    - 部分点在XY方向上的极小值与极大值
  - Moment-based function
    - 所有点均值作为中心点，二阶矩乘上系数作为长宽（类似方差概念）

- Learning RepPoints
  - 定位损失：pseudo-BBox的左上角和右下角，smooth L1 loss
  - 分类损失

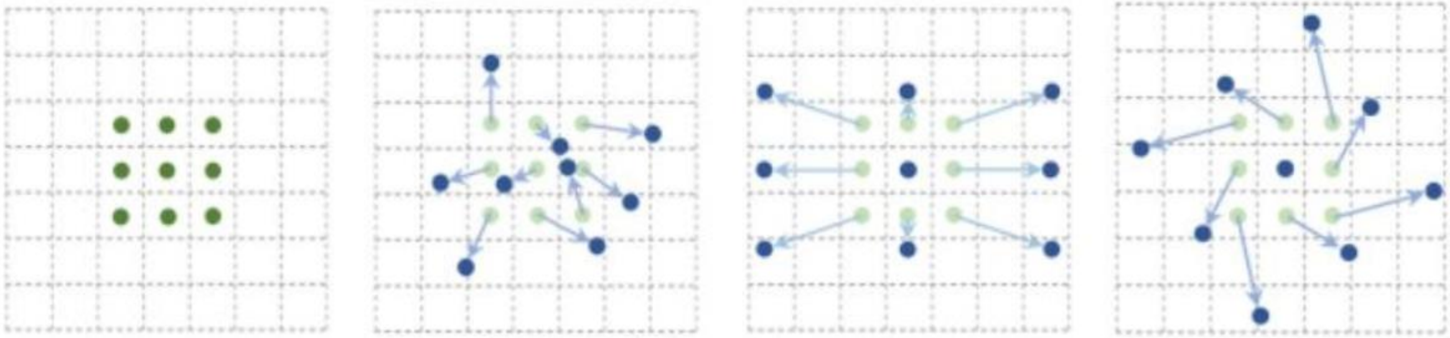
Methods

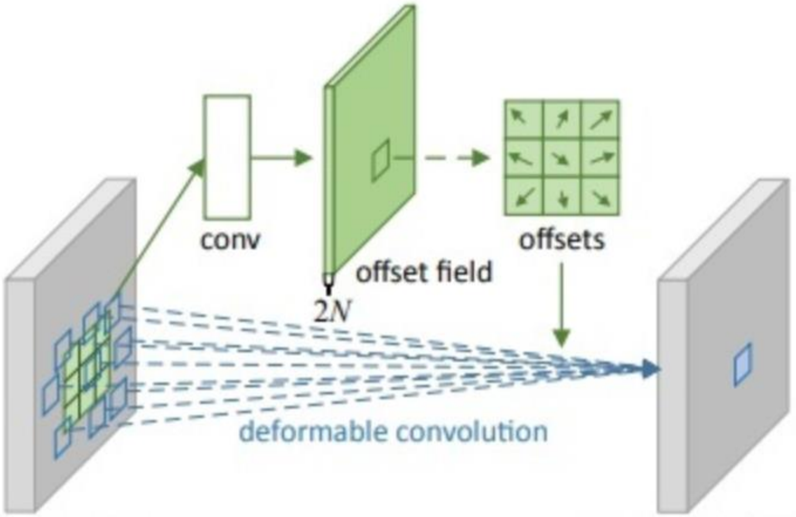
- （类）两阶段方法
- 结合 deformable convolution



Methods

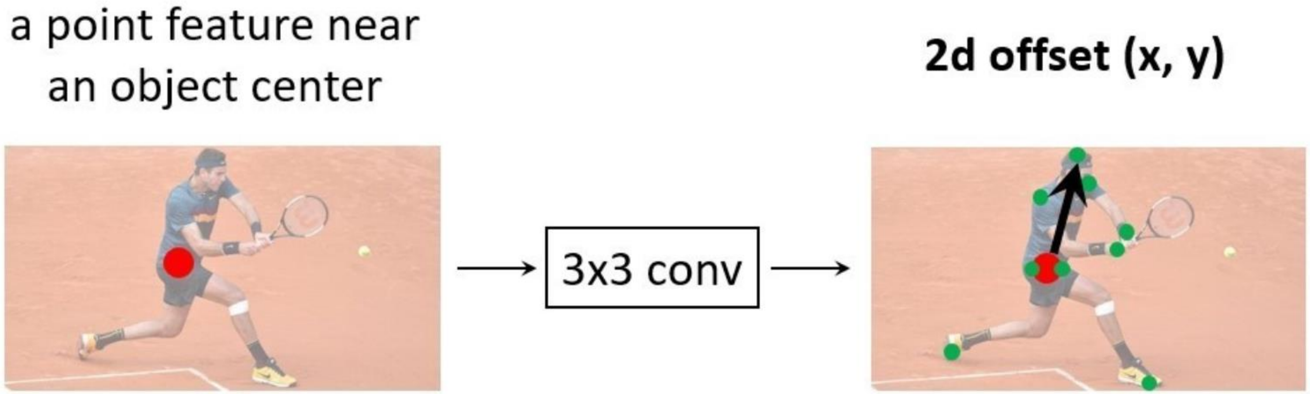
- Deformable convolution
  - 学习偏移量，集中至感兴趣区域

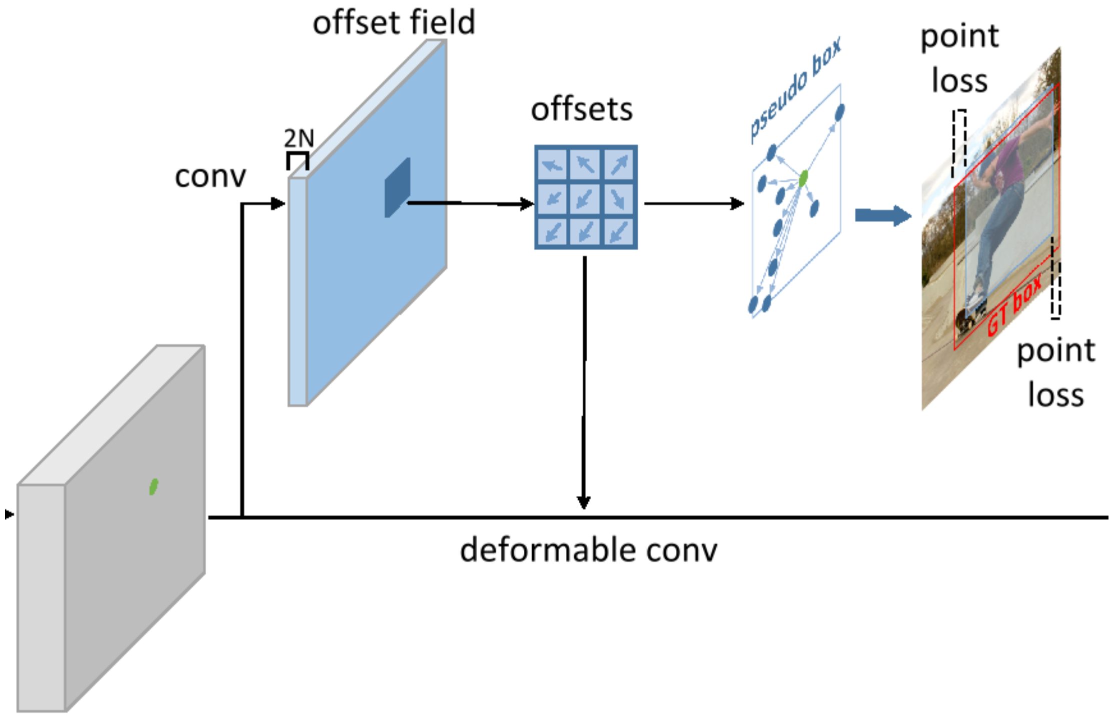




Methods

- RPDet： an Anchor Free Detector
  - 初始化：预测物体中心点

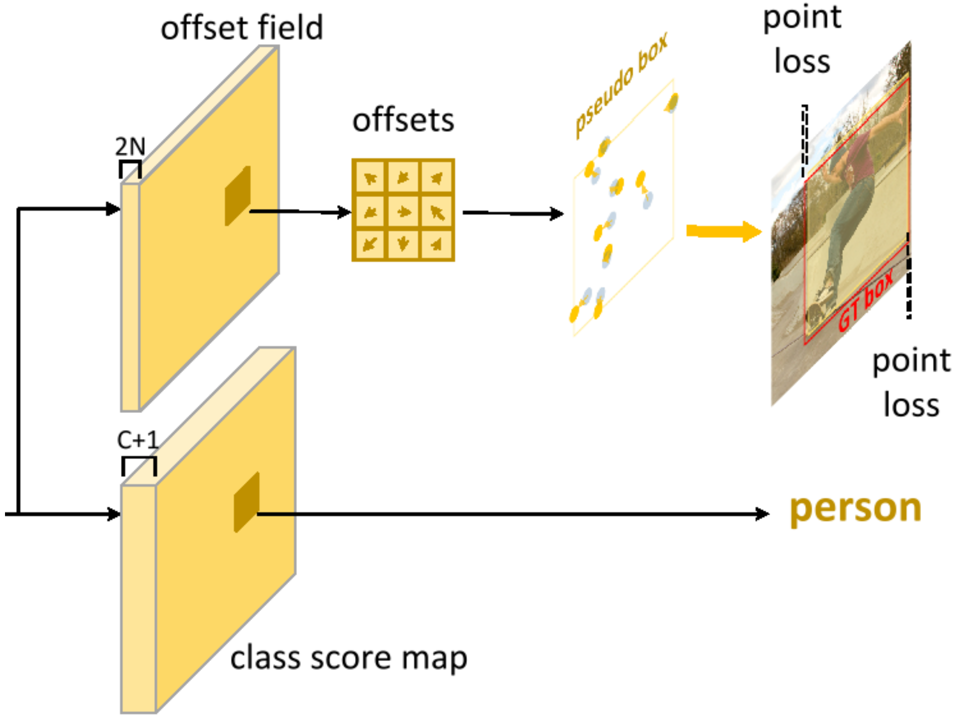




  - 第一组RepPoints点，距中心点的偏移
  - 定位loss + 分类loss

Methods

- RPDet： an Anchor Free Detector
  - 第二组RepPoints点，距第一组点的偏移
  - Refinement
  - 仅定位loss
  - 目的在于预测更精确的位置



Results

- 数据集： MS COCO 2017
- Bacbone： resnet-50， FPN
- AR： average recall
- AP： average precision
- Detection 结果

Representation	Backbone	AP	AP_{50}	AP_{75}
Bounding box	ResNet-50	36.2	57.3	39.8
RepPoints (ours)	ResNet-50	**38.3**	**60.0**	**41.1**
Bounding box	ResNet-101	38.4	59.9	42.4
RepPoints (ours)	ResNet-101	**40.4**	**62.0**	**43.6**

method	backbone	# anchors per scale	AP
RetinaNet [28]	ResNet-50	3×3	35.7
FPN-RoIAlign [27]	ResNet-50	3×1	36.7
YOLO-like	ResNet-50	-	33.9
RPDet (ours)	ResNet-50	-	**38.3**
RetinaNet [28]	ResNet-101	3×3	37.8
FPN-RoIAlign [27]	ResNet-101	3×1	39.4
YOLO-like	ResNet-101	-	36.3
RPDet (ours)	ResNet-101	-	**40.4**

Table 4. Comparison of the proposed method (RPDet) with an anchor-based method (RetinaNet, FPN-RoIAlign) and an anchor-free method (YOLO-like). The YOLO-like method is adapted from the YOLOv1 method [35] by additionally introducing FPN [27], GN [48] and focal loss [28] into the method for better accuracy.

Results

- MS-COCO

	Backbone	Anchor-Free	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
YOLOv2 [36]	DarkNet-19		21.6	44.0	19.2	5.0	22.4	35.5
SSD [31]	ResNet-101		31.2	50.4	33.3	10.2	34.5	49.8
YOLOv3 [37]	DarkNet-53		33.0	57.9	34.4	18.3	35.4	41.9
DSSD [10]	ResNet-101		33.2	53.3	35.2	13.0	35.4	51.1
Faster R-CNN w. FPN [27]	ResNet-101		36.2	59.1	39.0	18.2	39.0	48.2
RefineDet [52]	ResNet-101		36.4	57.5	39.5	16.6	39.9	51.4
RetinaNet [28]	ResNet-101		39.1	59.1	42.3	21.8	42.7	50.2
Deep Regionlets [49]	ResNet-101		39.3	59.8	-	21.7	43.7	50.9
Mask R-CNN [14]	ResNeXt-101		39.8	62.3	43.4	22.1	43.2	51.2
FSAF [56]	ResNet-101		40.9	61.5	44.0	24.0	44.2	51.3
LH R-CNN [26]	ResNet-101		41.5	-	-	25.2	45.3	53.1
Cascade R-CNN [2]	ResNet-101		42.8	62.1	46.3	23.7	45.5	55.2
CornerNet [24]	Hourglass-104	✓	40.5	56.5	43.1	19.4	42.7	53.9
ExtremeNet [54]	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**

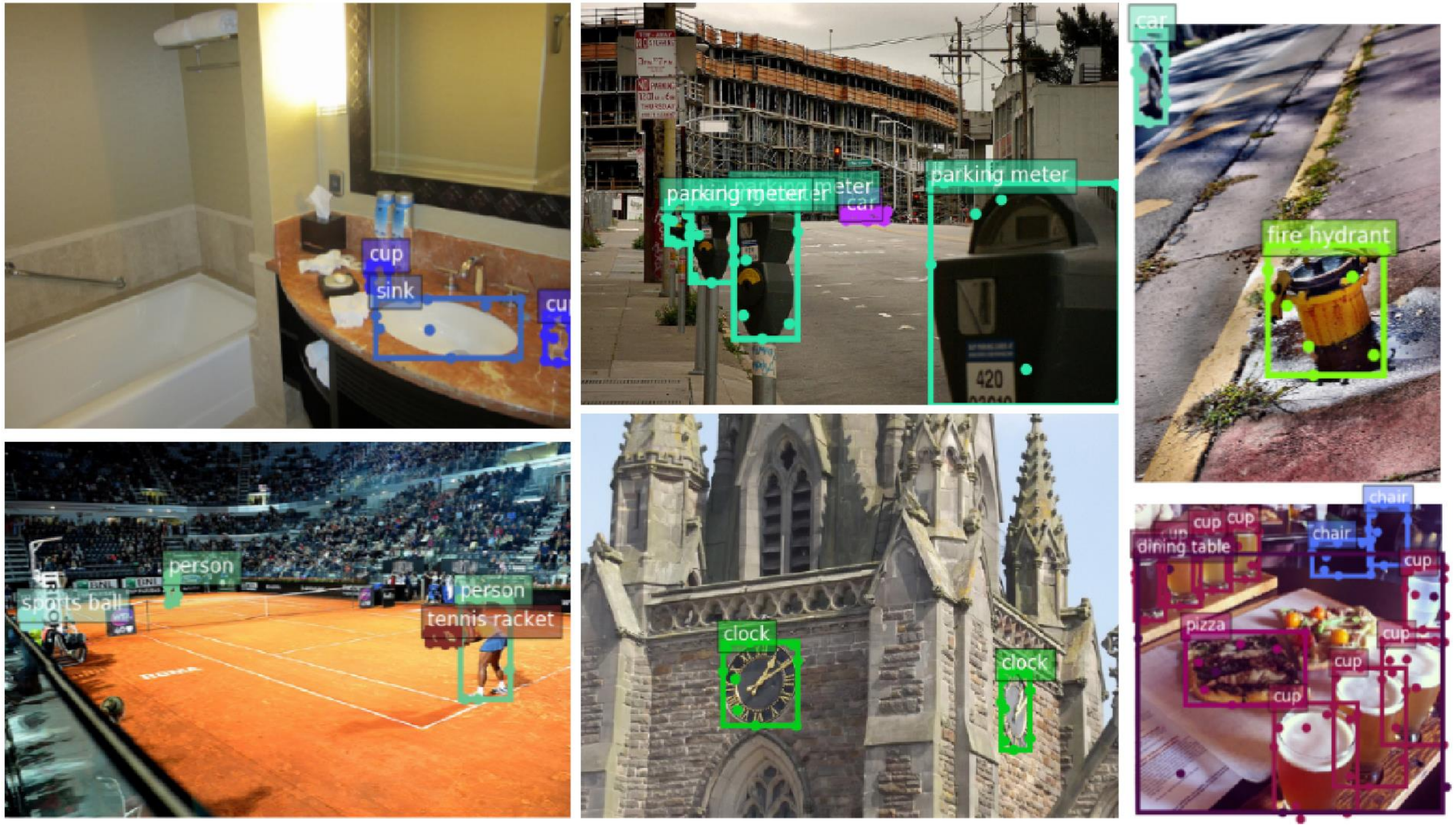
Results

- RepPoints 转换 pseudo-Bbox 方式

pseudo box converting function	AP	AP_{50}	AP_{75}
$\mathcal{T} = \mathcal{T}_1$: min-max	38.2	59.7	40.7
$\mathcal{T} = \mathcal{T}_2$: partial min-max	38.1	59.6	40.5
$\mathcal{T} = \mathcal{T}_3$: moment-based	38.3	60.0	41.1

Results

- RepPoints 可视化



Summary

- 方法
  - 新的表示方式，能够对物体的形状、姿态等更精细地表达
  - RepPoints 与 deformable convolution 结合
  - 某种意义上的 anchor，但是点集是任意的
- 局限
  - 点集的语义性不足