

# Anchor-free第一篇-CornerNet 论文总结

---

作者：小哲

微信公众号：小哲AI

github: <https://github.com/lxztju/notes>

## Anchor-free第一篇-CornerNet 论文总结

1. 论文摘要
2. 算法实现效果
3. 论文主要思想及创新点
4. 论文架构
5. 论文的细节
  1. 检测角点
  2. 分组匹配角点
  3. corner pooling
  4. hourglass Network

## 1. 论文摘要

---

文章提出一种新型的目标检测算法，CornerNet，利用一个单一的卷积神经网络检测一对关键点（左上角与右下角）作为一个对象的bbox。CornerNet利用一组关键点检测目标物体，消除了通常在的one-stage算法中anchor的设计。论文中提出了一种新型的pooling层帮助更好的定位corner——corner pooling，CornerNet在COCO上实现了42.1AP，超出所有显存的one-stage感知器。

## 2. 算法实现效果

---

在COCO数据集上的实验结果对比，在各种one-stage方法中得到了SOTA的结果。

Method	Backbone	AP	AP <sup>50</sup>	AP <sup>75</sup>	AP <sup>s</sup>	AP <sup>m</sup>	AP <sup>l</sup>	AR <sup>l</sup>	AR <sup>10</sup>	AR <sup>100</sup>	AR <sup>s</sup>	AR <sup>m</sup>	AR <sup>l</sup>
<b>Two-stage detectors</b>													
DeNet [39]	ResNet-101	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 [46]	ResNet-101	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 [38]	34.7	55.5	36.7	13.5	38.1	52.0	-	-	-	-	-	-
Faster R-CNN+++ [15]	ResNet-101	34.9	55.7	37.4	15.6	38.7	50.9	-	-	-	-	-	-
Faster R-CNN w/ FPN [22]	ResNet-101	36.2	59.1	39.0	18.2	39.0	48.2	-	-	-	-	-	-
Faster R-CNN w/ TDM [35]	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	37.5	58.0	-	19.4	40.1	52.5	-	-	-	-	-	-
Regionlets [43]	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	-	-	-	-	-	-
Soft-NMS [2]	Aligned-Inception-ResNet	40.9	62.8	-	23.3	43.6	53.3	-	-	-	-	-	-
LH R-CNN [21]	ResNet-101	41.5	-	-	25.2	45.3	53.1	-	-	-	-	-	-
Fitness-NMS [40]	ResNet-101	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	-	-	-	-	-	-
D-RFCN + SNIP [37]	DPN-98 [5]	45.7	67.3	51.1	29.3	48.8	57.1	-	-	-	-	-	-
<b>One-stage detectors</b>													
YOLOv2 [31]	DarkNet-19	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 [33]	DS/64-192-48-1	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 [34]	DS/64-192-48-1	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 [25]	ResNet-101	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 [10]	ResNet-101	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	36.4	57.5	39.5	16.6	39.9	51.4	-	-	-	-	-	-
RetinaNet800 [23]	ResNet-101	39.1	59.1	42.3	21.8	42.7	50.2	-	-	-	-	-	-
RefineDet512 (multi scale) [45]	ResNet-101	41.8	62.9	45.7	25.6	45.1	54.1	-	-	-	-	-	-
CornerNet511 (single scale)	Hourglass-104	40.5	56.5	43.1	19.4	42.7	53.9	35.3	54.3	59.1	37.4	61.9	76.9
CornerNet511 (multi scale)	Hourglass-104	42.1	57.8	45.3	20.8	44.8	56.7	36.4	55.7	60.0	38.5	62.7	77.4

### 3. 论文主要思想及创新点

主要思想：将目标检测的问题看作一个关键点检测的问题， 利用物体目标bbox的左上与右下角作为得到预测的bbox，因此在cornetnet中就不存在anchor的概念。因此避免了anchor使用的两个弊端：

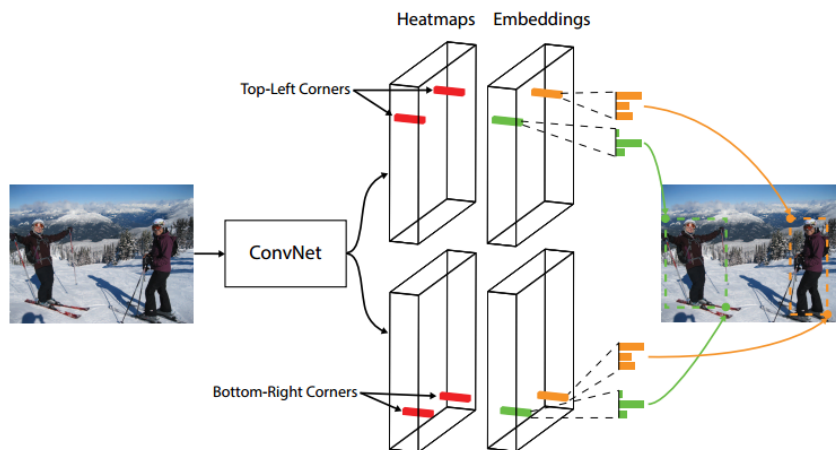
1. 正负样本的极度不均衡， anchor的数目往往很多，造成负样本过多
2. 引入大量的设计超参数，例如anchor的数目，大小，长宽比。

因此这种算法是一种anchor free的算法。

总结这种算法：主要就是两个关键点(top left and bottom right)的检测与匹配问题

### 4. 论文架构

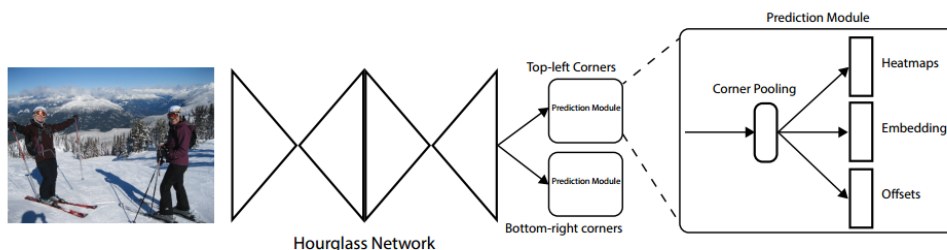
算法的主要架构：



**Fig. 1.** We detect an object as a pair of bounding box corners grouped together. A convolutional network outputs a heatmap for all top-left corners, a heatmap for all bottom-right corners, and an embedding vector for each detected corner. The network is trained to predict similar embeddings for corners that belong to the same object.

主要结构：

1. 利用hourglasses作为backbone提取特征
2. 生成两组feature map（heat map）分别表示top-left与bottom-right
3. 然后根据一定的方法匹配两个关键点。



**Fig. 4.** Overview of CornerNet. The backbone network is followed by two prediction modules, one for the top-left corners and the other for the bottom-right corners. Using the predictions from both modules, we locate and group the corners.

论文的整体结构图，如上图所示，经过backbone之后的特征图经过corner pooling之后，生成heatmap，embedding与offsets，

embedding vector用来评估两个关键点之间的距离，offsets是为了对预测bbox进行微调，因为在将下采样后的feature map映射回原图时进行取整会产生量化误差。

主要的结构部分：

- hourglass backbone
- corner pooling

## 5. 论文的细节

### 1. 检测角点

预测两个feature map，一个代表了左上角，一个代表右下角，每一组角点都预测 $C \times H \times W$ ，其中HW为heatmap的长宽，C为类别的数目（不包括背景类别）。如下图对于每一个corner都有一个positive的ground-truth，其他的位置都是负值。对于负样本的惩罚随着距离真值的距离不断递减的（依据2维高斯函数递减）。



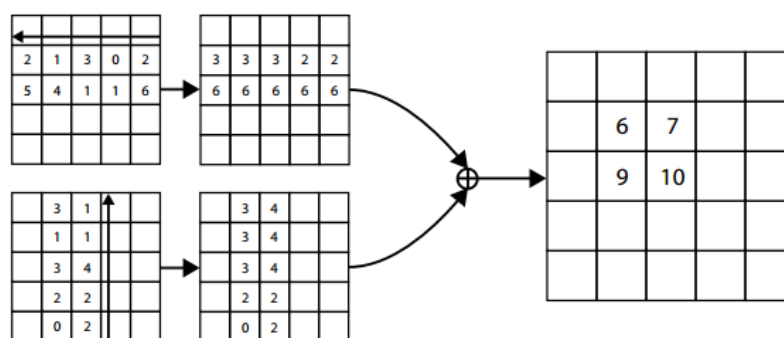
**Fig. 5.** “Ground-truth” heatmaps for training. Boxes (*green dotted rectangles*) whose corners are within the radii of the positive locations (*orange circles*) still have large overlaps with the ground-truth annotations (*red solid rectangles*).

## 2. 分组匹配角点

一张图像上可能有多个物体，因此会检测出很多的角点，采用embedding的方法进行匹配，网络对于每一个角点预测一个embedding向量，对于属于同一个object的top-left与bottom-right的点，他们的embedding向量之间的距离非常小。（利用二者之间的距离来匹配对应的top-left与bottom-down角点）。

## 3. corner pooling

这是论文中的一种新提出的方法，直接上图看看怎么计算（以top-left为例）：



判断某个点是否为top-left角点，上下两组特征图分别表示left与top的特征图，对于一个给定的特征图对于left，分别从当前位置到最右侧的vector位置取maxpooling，同理top是从当前位置到最下端的vector取maxpooling，然后将二者相加。

这个层在预测module中产生heatmap，embedding与offsets。

## 4. hourglass Network

这个网络的hourglass模块先下采样，然后进行上采样。由于下采样的过程中会丢失大量的信息，加入了skip layers，这种架构里采用了一个整体的架构能够轻松的实现全局与局部信息的捕捉。

hourglass 论文: <https://arxiv.org/pdf/1611.05424>

---

CornerNet的论文地址: <https://arxiv.org/abs/1808.01244>