

# 偏移垃圾箱分类网络，用于精确对象检测

Heqian Qiu, Hongliang Li<sup>\*</sup>, Qingbo Wu, Hengcan Shi University  
of Electronic Science and Technology of China Chengdu, China

hqqiu@std.uestc.edu.cn, hlli@uestc.edu.cn, qbwu@uestc.edu.cn, shihc@std.uestc.edu.cn

## Abstract

对象检测结合了对象分类和OB-JECT本地化问题。大多数现有物体检测方法通常通过利用具有平滑L1损耗函数的回归网络训练来定位对象，以预测候选框和对象之间的开关。但是，此损失函数在具有大错误的不同样本上应用相同的惩罚，从而导致次优回归网络和不准确的偏移。在本文中，我们提出了一种偏移箱分类网络，其优化了跨熵损耗来预测更准确的偏移。它不仅为不同的样本提供了不同的惩罚，还避免了由大错误引起的渐变爆炸问题。具体地，我们将持续的偏移离散到多个箱中，并预测每个偏移箱的概率。此外，我们提出了基于期望的偏移预测和分层聚焦方法来提高预测精度。在Pascal VOC和MS-Coco数据集上的广泛实验证明了我们所提出的方法的有效性。我们的方法通过大的边距来实现基线方法

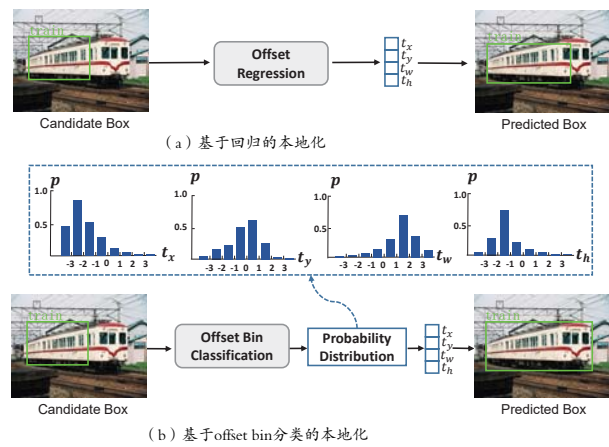


图1.基于典型的偏移回归的ob-ject检测方法和我们提出的偏移箱分类方法的比较。(a)典型的对象检测方法基于偏移回归定位对象。(b)所提出的方法根据不同的偏移箱上的输出概率分布定位ob-jects。典型的偏移回归方法从候选盒朝向物体中的有限偏移，而偏移箱分类方法避免了该问题。L.

## 1. Introduction

对象检测是一个基本且挑战的愿景任务，包括对象分类和对象本地化问题。一组广泛的计算机视觉应用，如自动驾驶[7,17,39–41]，视频监控[6,24]和机器人[38,42,45]将受益于准确的对象本地化。大多数最先进的对象检测方法[1,8,11,12,20,21,26,30,31,35,44]首先生成一系列候选盒，然后预测这些盒子的偏移找到对象，如图1(a)所示。由于偏移是连续值，因此这些方法通过使用L2或平滑的L1损耗优化的杠杆作用回归网络来预测它们。但是，如[9]所研究，

2丢失[10]可能导致偏移误差时造成梯度爆炸。为了避免这个问题，平滑的L1损失[9]通过剪切它们的梯度，削弱了样本与大误差的影响。虽然平滑的L1损失解决了梯度爆炸问题，但它不能惩罚具有大错误的样本，这导致候选盒和对象之间的次优回归网络和不准确的偏移。例如，在图1(a)中，列车物体不能被边界框紧密包围。为了解决这个问题，我们提出了一个偏移的垃圾箱分类网络来预测更准确的偏移，如图1(b)所示。所提出的方法采用具有跨熵损耗的分类网络，而不是平滑的L1或L2损耗。一方面，它给出了不同的偏移错误足够的惩罚。在这一点

<sup>\*</sup>Corresponding author.

另一方面，它避免了梯度爆炸问题。从来没有一个分类网络只能预测离散的偏移值。因此，我们提出了基于期望的偏移预测和分层聚焦偏移偏移，以进一步提高预测精度。具体地，我们使用均匀的离散化量化连续偏移到多个箱中，然后用交叉熵丢失训练偏移箱分类网络以预测偏移箱的概率分布。灵感来自[37]，通过计算离散化偏移箱的Softmax预期值，将分类结果转换为对象位置。同时，我们提出了一个分层聚焦偏移预测网络，以逐渐改进偏移箱，以获得更精确的对象本地化。我们验证了我们对于两个公共对象检测数据集的方法的有效性，包括Pascal VOC和MS-Coco数据集。结果表明，我们提出的方法有利于准确的对象检测。我们的贡献可以归纳如下：我们提出了一个偏移垃圾箱分类网络，以预先描述更准确的偏移而不是通过平滑L1或L2损耗优化的回归网络。

为了进一步产生更精确的对象本地化，我们提出了基于期望的偏移预测和Hierarchical聚焦偏移预测。

两个常见数据集上的广泛实验证明了所提出的方法的有效性。

## 2. Related Work

对象探测器：现代物体检测框架通常可以被归类为两级和单级的解决方案。在两阶段检测器中[1,8,11,12,20,21,26,30,31,35,44]，首先产生可能包含Object-Keypoints的稀疏区域提案集，然后它们的特征是提取以下分类和本地化。代表性方法，包括更快的R-CNN [35]，FPN [20]和掩模R-CNN [12]，已经在各种基准上实现了主导的性能。与两级探测器相比，单级探测器[18,19,21,23,32-34]达到高推理速度，例如yolo [32-34]，SSD [23]，视网膜[21]。它们通常跳过区域提案基因，并直接在锚箱方案之后预测边界框。虽然这些方法已成功进行了对象，但实现了准确对象本地化仍然是一个具有挑战性的问题。绑定框回归：为了解决对象本地化的问题，大多数物体检测方法[1,8-11,15,26,28,44]利用边界框重新定位网络预测四个将候选框转换为对象的坐标。R-CNN [10]通过培训具有L2损耗的线性回归模型来预测这些偏移。但是，当有一些具有大错误的样本时，很容易引起梯度利用。换 -

L2损失，FAST R-CNN

[9]提出平滑L1损失，以减少样品的效果大的误差，这已被广泛接受对象检测中的回归。平衡L1丢失[28]进一步提高了样本的渐变贡献，以小错误来重新平衡所涉及分类和本地化任务以及具有不同属性的样本。不同的方法KL丢失[14]将地面真理界限框架的含糊存在，并学习了更多准确对象本地化的界限框回归和本地化方差。此外，Unitbox [46]和Giou

[36]直接使用评估度量作为对象函数来解决优化常用距离损耗和最大化度量值之间的间隙。但是，很难用同样的IOU优化不同的边界框。此外，一系列对象检测器[1,8,11,26,44]试图通过迭代地重复绑定框来改善对象定位。它们都级联多个重新格雷斯，并在每次迭代到下一个边界框回归线后喂养检测结果。Cascade R-CNN

[1]考虑了在每次迭代时检测输出的分布和再采样的边界框，以保证检测器质量与测试之间的匹配。然而，随着迭代的数量增加，它是非单调的，以提高基因准确性。iou-net [15]建议预测IOU与匹配的基础是指导边界框的回归的本地化信心。我们提出了一个偏移垃圾箱分类网络，跨熵丢失来实现更准确的对象本地化，这也是有效地转动其他计算机视觉区域。对于考试，[27]预测检测热量和用于人类姿势估计的关联嵌入标签。

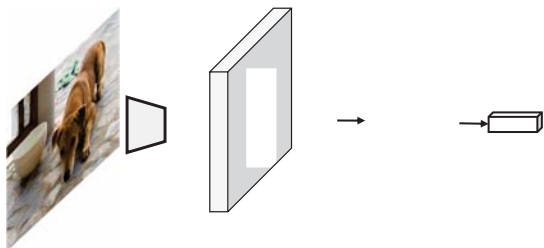
[5]通过使用序数回归损耗而不是L2损耗训练深度估计网络。最近，一些无锚方法[16,43,47]直接预测边界框的关键点的热量，并引入不同种类的损耗函数来细化和组用于最终检测到的边界框的这些关键点。Cornersnet

[16]使用ShimpleL1丢失来分配本地偏移，并拉出损耗并推动丢失，以限制关键点之间的差异。CenterNet

[47]使用两个L1损耗功能回归局部偏移和对象大小。FCOS [43]使用IOU丢失来分配边界框的区域。与所提出的方法不同，它们通常需要仔细组用于最终对象的关键点。

## 3. Approach

在本节中，我们首先审查和分析传统边界框回归的问题。然后，我们引发了我们建议的偏移垃圾箱分类网络以遵守这个问题，这是基于流行的FPN [20]实现的。



连续偏移范围

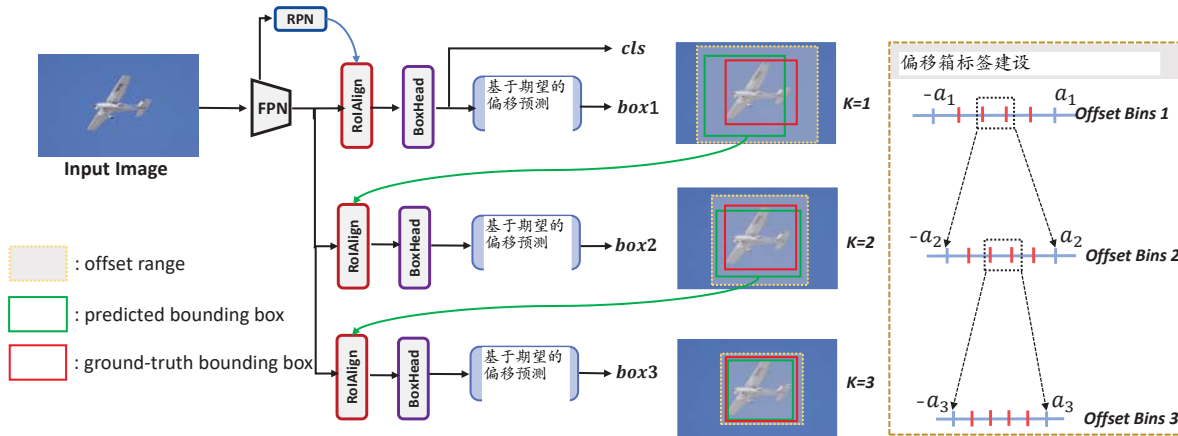


图3.所提出的分层聚焦偏移预测的架构。在这里，我们在分层聚焦偏移预测中显示三个阶段。填充有灰色的黄色虚线盒子表示每个阶段的偏移量程。绿色盒子和红色框代表每个阶段的预测盒子和地面真相盒。每个阶段的偏移量级在前一级的偏移区内定义。

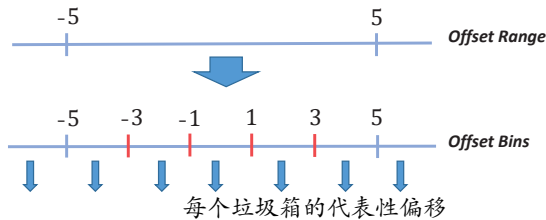


图4.偏移箱施工的插图。偏移范围 $(-5, 5)$ 被均匀地离散化为五个垃圾箱，每个箱子的中位数为他们的代表性偏移。另外，范围 $(-\infty, -5]$ 和 $[5, +\infty)$ 分别使用端点-5和5分别作为其代表性偏移。

每个垃圾箱的Cive偏移可以表示如下：

$$t_{i,l}^r = \begin{cases} -a + (l + \frac{1}{2}) * w & l \in [0, m] \\ -a & l = m + 1 \\ a & l = m + 2 \end{cases} \quad (5)$$

如果 $TRI, L$ 是对应于绑定框的坐标 $I$ 的BIN标签 $L$ 对应的代表性偏移。从0到 $m$ 的标签的重复偏移被表示为每个箱的中值，另一个标签表示为端点的偏移量。

### 3.2.2 Network Learning

基于离散化偏移箱标签，将对象本地化作为多级分类问题而不是直接回归是直接的。如图2所示，候选盒被馈送到后骨网络FPN [20]的盒头中，以产生其偏移栅格评分Vec-tor  $S \in R^{4n}$ ，其中4是界定框的四个坐标， $n$ 是偏移垃圾箱的数量。然后我们重塑

分数矢量到 $R^{4 \times n}$ 并分别通过SoftMax函数分别对每个坐标的分数矢量分别为概率的形式，如下所示：

$$p_{i,l} = \frac{\exp(s_{i,l})}{\sum_{l=0}^{n-1} \exp(s_{i,l})} \quad (6)$$

其中 $p_{i,l}$ 表示第 $i$ 个坐标偏移的概率属于 $l$ -th bin。偏移箱分类器的损耗功能LBIN作为交叉熵丢失，使：

$$L_{bin}(p_{i,l}, l) = - \sum_{i \in \{x,y,w,h\}} \sum_{l=0}^{n-1} y_l * \log p_{i,l} \quad (7)$$

其中计算丢失时，当地面真理类被标记为 $l$ ，其中 $y_l \in \{0,1\}$ 。关于输出得分SB的梯度

可以导出分类器层的IL，如下所示：

$$\frac{\partial L_{bin}}{\partial s_i} = \begin{cases} - \sum_{i \in \{x,y,w,h\}} (p_{i,l} - 1), & y_l = 1 \\ - \sum_{i \in \{x,y,w,h\}} (p_{i,l}), & y_l = 0 \end{cases} \quad (8)$$

基于上述公式，偏移是有界的，其规范限制为 $[0,1]$ ，与L2损耗功能相比，所有样品更稳定。同时，它有效考虑了使用基于预测概率 $P_{i,l}$ 的不同梯度概要的样本与平滑L1损耗相比。最后，我们使用损失函数1到端到端的列车我们的网络准确对象检测：

$$L = L_{cls} + \lambda_{bin} L_{bin} \quad (9)$$

其中LCL表示对象分类的损失，偏移箱分类丢失Lbin用于对象的本地化。 $\lambda_{bin}$ 是控制这些损失之间平衡的重量。在本文中，我们将 $\lambda_{bin}$ 设置为1。

方法期望分层AP50 AP60 AP70 AP80 AP90							
Bounding Box Regression [20]	45.0	<b>74.5</b>	<b>69.5</b>	57.6	36.0	6.6	
Bin Classification	45.8	73.3	67.9	57.2	39.6	9.8	
Bin Classification	✓	47.5	74.0	69.0	58.8	41.5	13.6
Bin Classification	✓	47.5	72.8	67.9	58.1	42.0	16.0
Bin Classification	✓	✓	<b>49.0</b>	73.2	68.4	<b>59.0</b>	<b>44.3</b> <b>19.6</b>

表1.每个组分在所提出的方法中的影响。结果是在VOC2007测试集[4]上报告。具有Reset-50-FPN [20]的基线方法通过边界框回归方法定位对象。期望和分层代表基于期望的偏移预测和分层聚焦偏移预测。

### 3.2.3 Expectation-based Offset Prediction

由于偏移是具有高精度的连续值，因此分类网络仅预测离散偏移值。因此，我们提出了两种不同的方法来提高检测结果的预测：基于期望的偏移预测和分层聚焦偏移预测。对于图2中的基于期望的偏移预测方法，我们利用不同偏移偏移频率的概率分布来估计预测的偏移 $T_i$ ，其由SoftMax预期值而不是最大值计算，如下所示：

$$t_i = \mathbb{E}(T_i^r) = \sum_{l=0}^{n-1} (p_{i,l} * t_{i,l}^r) \quad (10)$$

其中 $\text{tri} = \{\text{tri}, 0, \text{tri}, 1, \dots, \text{tri}, n-1\}$ 表示 $n$ 个垃圾箱的一组代理离散偏移。符号 $\mathbb{E}$ 表示离散偏移的期望。

#### 3.2.4 分层聚焦偏移预测

此外，我们提出了一种分层聚焦偏移预测，具有粗细的策略，逐渐改进箱间隔，如图3所示。当BIN间隔非常小时，离散值将更接近目标值。假设K-TH阶段有K阶段和NK箱。在每个阶段，偏移范围 $(-AK, AK)$ 定义在前一级的偏移区内。所以，距离的宽度 $W_k$ 可以表示为 $W_{k-1}$

NK。然后，我们预测每个阶段的偏移 $T_{ki}$ 类似于第3.2.3节。最终预测的偏移可以计算为：

$$t_i = \sum_{k=1}^K t_i^k \quad (11)$$

如图3所示，在第一阶段，我们预测RPN生成的候选框之间的摘录和偏移范围内的OBJECTS $(-A_1, A_1)$ 。随后，在每个阶段，我们预测先前的禁区内的更精细的偏移。通过逐步分类偏移，我们可以获得更精确的边界框。

## 4. Experiments

为了评估所提出的偏移垃圾箱分类网络的有效性，我们对两个标准对象检测数据集进行了广泛的实验，包括Pascal VOC数据集[4]和MS-Coco DataSet [22]。数据集。Pascal VOC数据集[4]包含20个OB-JECT类别，其中包括Pascal VOC2007数据集和Pascal VOC2012数据集。以下[35]，我们在VOC 2007 TrainVal和VOC2012 TrainVal集合上培训我们的网络，其中包括5011和11540张图像，并评估包含4952图像的VOC2007测试集。MS-Coco DataSet [22]涉及80个对象类别，其具有比Pascal VOC数据集更大的刻度。在常见的做法之后评估指标。我们采用标准的Coco风格的平均精度（AP）来测量各种质量的检测性能，其平均地图跨越0.5至0.95的阈值，间隔为0.05。它还包括小规模AP，中等尺度APM和大型APL跨的AP。实施细节。为了进行公平比较，我们实现了基于Pytorch [29]和MMDETECTION的所有实验[2]。基于Reset-50和Resnet-101 [13]作为基线网络，我们使用FPN [20]。在典型的约定之后，我们在Pascal VOC数据集[4]上采用 $1000 \times 600$ 的输入图像比例，以及MS-Coco DataSet上的 $1333 \times 800$ 的比例[22]。我们培训结束于结束的结果，以2个GPU（每GPU 2张图像）为12时代。初始学习速率设定为0.005，并在8月和11个时期后减少0.1因素0.1。除非另有指定，否则所有其他超参数均遵循MMDetection中的默认设置[2]。损耗重量 $\lambda$  Bin设置为1。偏移范围A和箱数N分别设定为3和20。在分层聚焦偏移预测中，阶段K的数量被设置为2。

### 4.1. Ablation Study

在本节中，我们验证了基线Reset-50-FPN [20]的有效性。没有损失普遍性，我们



Method	$AP$	$AP_{50}$	$AP_{60}$	$AP_{70}$	$AP_{80}$	$AP_{90}$
$L_2$ Loss [10]	44.7	72.6	67.6	56.8	37.4	7.8
$Smooth L_1$ Loss [20]						
$\beta = 1.0$	45.0	<b>74.5</b>	<b>69.5</b>	57.6	36.0	6.6
$\beta = 1.5$	44.3	73.9	68.6	56.5	34.9	6.4
$\beta = 2.0$	44.2	74.3	68.9	56.1	33.9	6.2
Bin Classification	<b>47.5</b>	74.0	69.0	<b>58.8</b>	<b>41.5</b>	<b>13.6</b>

表2.不同损失功能的有效性。 $\beta$ 表示平滑L1损耗功能中的分割点。结果是在VOC2007测试集[4]上报告。

Stage	$AP$	$AP_{50}$	$AP_{60}$	$AP_{70}$	$AP_{80}$	$AP_{90}$
$K = 1$	47.5	<b>74.0</b>	<b>69.0</b>	58.8	41.5	13.6
$K = 2$	<b>49.0</b>	73.2	68.4	<b>59.0</b>	<b>44.3</b>	<b>19.6</b>
$K = 3$	48.8	73.3	68.3	58.5	43.6	19.1

表3.所提出的Hi-批准聚焦偏移预测方法中阶段数量的有效性。结果是在VOC2007测试集[4]上报告。

形成消融研究，以揭示每个组分在Pascal VOC数据集中提出的方法[4]。主要成分分析。我们分析了表1中每个提出的组分的效果。简单地通过所提出的偏移箱分类方法估算OB-JECT定位，与基线边界框回归方法相比将AP改善0.8% [20]。介绍基于期望的偏移预测和分层聚焦偏移预测与基线相比，达到2.5%的增益，这进一步提高了预测的预测。基于期望的偏移预测进入AC-计算其他偏移频体中的样本的概率，以便与ES-时分偏移，并一致地改善了不同的IOU度量标准。分层聚焦偏移预测在高IOU度量中执行更好。原因是它在更精细的偏移箱内预测更精确的偏移。

Ulti-Menty，我们的完整方法优于基线边界框回归方法4.0%。结果表明，所提出的方法在更准确的对象检测方面的有效性，尤其是在高IOU度量中执行比赛。不同损失函数预测偏移的有效性。表2中所示的不同损失功能的有效性如图2所示。基于相同的骨干网Reset-50-FPN [20]，我们调整回归损耗平滑L1的Di-Vision点 $\beta$ ，以便基于处理更多的样品足够的渐变贡献。然而，当我们把 $\beta$ 设置为更大的值时，检测性能AP减少。一种可能的原因是网络学习由具有大距离误差的一些样本主导。与光滑的L1损耗和L2损耗相比，我们的方法如图所示执行更好的性能

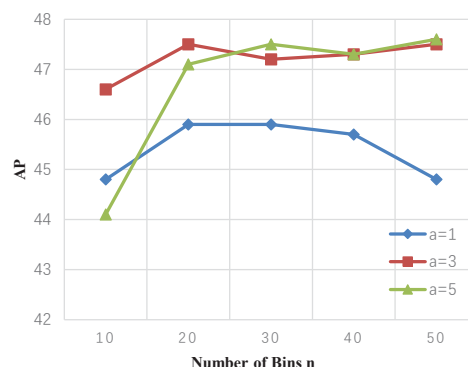


图5.箱体分类对具有不同超参数的偏移箱标签的有效性。横轴表示箱数N，垂直轴代表检测性能AP。蓝线，红线和绿线分别表示偏移量A = 1,3,5。

在表2中，通过偏移箱分类来缓解问题。

偏移箱标签的设置。图5显示了具有不同超参数的偏移箱标签的垃圾箱分类的EF-效力。A和N分别表示分割偏移范围和箱数的终点。当箱数n是固定的时，可以看出，对于a = 1，检测性能降低，而性能类似于a = 3和a = 5。这是因为忽略了大于1的偏移量的许多样本在训练期间，如果a = 1.当端点a = 3或5时，可以观察到当箱n的数量设置为20到50时，检测性能非常靠近，从而坚固到长距离偏移箱数。另外，当n小时，检测性能相对较差（即n = 10）。要使用箱号码平衡性能，我们在实验中选择A = 3和N = 20。

分层聚焦偏移预测中的阶段数。阶段数量的有效性

基线方法我们的方法基线方法我们的方法



图6. VOC2007测试集基准方法与建议的偏移箱分类方法之间的可视化比较[4]。第一列和第三列显示基线方法的检测结果。第二个和第四列表明我们方法的检测结果。

Method	Backbone	$AP$	$AP_{50}$	$AP_{60}$	$AP_{70}$	$AP_{80}$	$AP_{90}$
Faster R-CNN* [35]	ResNet-50-FPN	45.0	<b>74.5</b>	<b>69.5</b>	57.6	36.0	6.6
Our +Faster R-CNN [35]	ResNet-50-FPN	<b>49.0</b>	73.2	68.4	<b>59.0</b>	<b>44.3</b>	<b>19.6</b>
Faster R-CNN* [35]	ResNet-101-FPN	47.8	<b>75.5</b>	<b>70.6</b>	60.3	41.3	10.5
Our+Faster R-CNN [35]	ResNet-101-FPN	<b>50.8</b>	74.0	69.5	<b>60.8</b>	<b>47.2</b>	<b>22.5</b>
Cascade R-CNN* [1]	ResNet-50-FPN	49.5	73.1	<b>69.0</b>	<b>61.0</b>	45.9	18.1
Our+Cascade R-CNN [1]	ResNet-50-FPN	<b>50.4</b>	<b>73.3</b>	68.9	60.4	<b>46.5</b>	<b>22.2</b>
Cascade R-CNN* [1]	ResNet-101-FPN	51.0	73.6	69.6	61.9	48.3	21.1
Our+Cascade R-CNN [1]	ResNet-101-FPN	<b>51.9</b>	<b>73.9</b>	<b>69.8</b>	<b>62.1</b>	<b>48.7</b>	<b>25.0</b>

表4.与VOC2007测试集上的最先进方法进行比较[4]。符号\*代表基于MMDetection [2]的重新实现结果。

分层聚焦偏移预测显示在表3中。根据图5中的分析，我们将每个阶段的箱数NK设置为相同（ $NK = 20$ ， $k = 1, 2, 3$ ）和端点 $A1 = 3$ 在第一阶段。因此，第三阶段中的第二级和 $A3$ 中的偏移范围 $A2$ 的终点分别设定为0.15和0.015。可以看出，当阶段 $k = 2$ 的数量时，检测结果 $AP$ 与仅一个阶段相比提高了1.6%。在第二阶段，箱的宽度已经在非常小的范围内。添加第三阶段，检测性能接近第二级。可以看出，具有两个阶段的BIN分类可以实现更好的检测性能。

可视化比较。图6显示了基线方法[20]与所提出的偏移箱分类方法之间的求助性比较。可以观察到基线方法[20]分配一些不紧密围绕物体的边界框

图6的第一行图像，而我们的方法可以更准确地检测对象。图6的第二行图像显示，由于低质量边界框，在基线方法[20]中错过了汽车对象和人物对象。

#### 4.2. 与最先进的方法进行比较

结果Pascal VOC数据集。我们在表4中我们将我们使用两个基线[1,20]上的两个基线[1,20]的方法进行比较。进行公平比较，我们采用了我们的方法和相应基线的相同参数设置。我们通过所提出的方法将边界框回归网络放置验证其有效性。因为Cascade R-CNN [1]是一个多级对象检测器，所以我们用图2中的胶印箱级分支替换Cascade R-CNN中每个阶段的再渗起分支。为了减少参数的数量，偏移量宾宾分类分支这里不包括图3中的分层专注。我们设置了

Method	Backbone	$AP$	$AP_{50}$	$AP_{75}$	$AP_S$	$AP_M$	$AP_L$
YOLOv2 [33]	DarkNet-19	21.6	44.0	19.2	5.0	22.4	35.5
SSD512 [23]	ResNet-101	31.2	50.4	33.3	10.2	34.5	49.8
RetinaNet [21]	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
Faster R-CNN [20]	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Deformable R-FCN [3]	Inception-ResNet-v2	37.5	58.0	40.8	19.4	40.1	52.5
Mask R-CNN [12]	ResNet-101-FPN	38.2	60.3	41.7	20.1	41.1	50.2
Libra R-CNN [28]	ResNet-101-FPN	40.3	61.3	43.9	22.9	43.1	51.0
KL Loss [14]	ResNet-50-FPN	39.2	57.6	42.5	21.2	41.8	52.5
Grid R-CNN [25]	ResNet-101-FPN	41.5	60.9	44.5	23.3	44.9	53.1
IoU-Net [15]	ResNet-101-FPN	40.6	59.0	-	-	-	-
Cascade R-CNN [1]	ResNet-101-FPN	<b>42.8</b>	<b>62.1</b>	<b>46.3</b>	<b>23.7</b>	<b>45.5</b>	<b>55.2</b>
Faster R-CNN* [20]	ResNet-50-FPN	36.6	58.8	39.6	21.6	39.8	45.0
Our+Faster R-CNN	ResNet-50-FPN	<b>40.5</b>	<b>59.6</b>	<b>43.1</b>	<b>22.6</b>	<b>43.1</b>	<b>51.0</b>
Faster R-CNN* [20]	ResNet-101-FPN	38.8	60.9	42.1	22.6	42.4	48.5
Our+Faster R-CNN	ResNet-101-FPN	<b>42.5</b>	<b>61.7</b>	<b>45.4</b>	<b>23.9</b>	<b>45.6</b>	<b>53.8</b>
Faster R-CNN* [20]	ResNeXt-101-FPN	41.9	<b>63.9</b>	45.9	<b>25.0</b>	45.3	52.3
Our+Faster R-CNN	ResNeXt-101-FPN	<b>43.2</b>	62.7	<b>46.3</b>	24.7	<b>46.4</b>	<b>54.8</b>
Cascade R-CNN* [1]	ResNet-50-FPN	40.7	59.3	44.1	23.1	43.6	51.4
Our+Cascade R-CNN	ResNet-50-FPN	<b>42.3</b>	<b>60.4</b>	<b>45.8</b>	<b>23.9</b>	<b>44.8</b>	<b>53.6</b>
Cascade R-CNN* [1]	ResNet-101-FPN	42.4	61.1	46.1	23.6	45.0	54.4
Our+Cascade R-CNN	ResNet-101-FPN	<b>44.4</b>	<b>62.6</b>	<b>48.3</b>	<b>24.7</b>	<b>47.5</b>	<b>56.7</b>
Cascade R-CNN* [1]	ResNeXt-101-FPN	43.7	62.6	47.5	25.3	46.7	55.5
Our+Cascade R-CNN	ResNeXt-101-FPN	<b>44.7</b>	<b>63.1</b>	<b>48.5</b>	<b>25.3</b>	<b>47.8</b>	<b>57.1</b>

表5.与MS-COCO测试开发套装上的最先进方法的比较[22]。符号\*代表基于MMDetection [2]的重新实现结果。

Cascade

R-CNN至2的级数分别在第一和第二阶段分别设定为0.5和0.7。这些基线通过我们的方法始终如一地改善，这证明了所提出的方法的优势和赋予。MS-Coco DataSet上的结果。此外，我们还在表5中的大规模MS-Coco Test-Dev集合[22]上的一些最先进的对象检测方法进行了比较了所提出的方法[22]。可以观察到所提出的方法显著优于这些最先进的方法。所提出的偏移箱分类方法可以将Reset-50-FPN，Reset-101-FPN和Resnext-101-FPN的更快R-CNN [20,35]的AP改善为3.9%，3.7%和1.3%。当我们将Cascade R-CNN [1]介绍给我们的方法时，结果AP可以达到相当大的精度42.3%，44.4%和44.7%。优越的性能证明了所提出的偏移箱分类方法的有效性。

首先使用偏移箱标签结构来将连续偏移分为几个箱。然后，偏移箱分类网络预测偏移箱的概率分布。此外，引入基于期望的偏移预测和分层聚焦偏移预测方法以将离散的分类结果转换为更精确的偏移。我们的方法在Pascal VOC和MS-Coco对象检测数据集中实现了卓越的性能。结果表明了我们所提出的方法的效果。确认。这项工作是由中国国家自然科学基金（61525102,61831005,61971095和61871078号）的支持。

## References

- [1] Zhaowei Cai和Nuno Vascomcelos. Cascade R-CNN: 渗透到高质量的物体检测。在计算机愿景和模式识别的IEEE会议上的诉讼程序中, pages 6154–6162, 2018. [2] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tian-

## 5. Conclusion

在本文中，我们提出了一个偏移的垃圾箱分类网络来实现更准确的对象检测。



heng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155, 2019. [3] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. In Advances in neural information processing systems, pages 379–387, 2016. [4] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. International journal of computer vision, 88(2):303–338, 2010. [5] Huan Fu, Mingming Gong, Chaohui Wang, Kayhan Batmanghelich, and Dacheng Tao. Deep ordinal regression net—work for monocular depth estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2002–2011, 2018. [6] Zhihang Fu, Yaowu Chen, Hongwei Yong, Rongxin Jiang, Lei Zhang, and Xian-Sheng Hua. Foreground gating and background refining network for surveillance object detection. IEEE Transactions on Image Processing, 28(12):6077–6090, 2019. [7] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 3354–3361. IEEE, 2012. [8] Spyros Gidaris and Nikos Komodakis. Object detection via a multi-region and semantic segmentation-aware cnn model. In Proceedings of the IEEE international conference on computer vision, pages 1134–1142, 2015. [9] Ross Girshick. Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 1440–1448, 2015. [10] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580–587, 2014. [11] Jicheng Gong, Zhao Zhao, and Nic Li. Improving multi-stage object detection via iterative proposal refinement. [12] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 2961–2969, 2017. [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016. [14] Yihui He, Chenchen Zhu, Jianren Wang, Marios Savvides, and Xiangyu Zhang. Bounding box regression with uncertainty for accurate object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2888–2897, 2019. [15] Borui Jiang, Ruixuan Luo, Jiayuan Mao, Tete Xiao, and Yunqiang Jiang. Acquisition of localization confidence for accurate object detection. In Proceedings of the European Conference on Computer Vision (ECCV), pages 784–799, 2018. [16] 嘿法律和贾邓。Cornernet: 将对象检测为配对键点。在欧洲经验 (ECCV) 上的欧洲会议上, 第734–750, 2018页。[17] Buyu Li, Wanli Ouyang, 陆胜, 兴宇曾和小刚王。GS3D: 高效的3D对象检测帧—用于自动驾驶。在IEEE电脑愿景和模式识别会议上的诉讼程序中, 第1019–1028, 2019页。[18] Wei Li, Hongliang Li, 青佛吴, 孝感和国王Ng。同时从寄生虫图像中检测和计数密集的车辆。IEEE工业电子产品交易, 66(12): 2019年9651–9662, 2019年。[19] 魏丽, 宏良李, 青湾吴, 凡人猛, 林峰徐和王朝王。Headnet: 用于头部检测的端到端自适应关系网络。视频技术电路和系统的IEEE交易, 2019年。特征用于对象检测的PYRA—中途网络。在IEEE计算机愿景和模式识别上的IEEE会议上, 第2117–2125, 2017页。[21] Tsung-yi Lin, Priya Goyal, Ross Girshick, Kaiming He和Piotr Dollár。致密物体检测的焦损。在IEEE国际计算机愿景会议上, 2017年第2980–2988, 2017页。[22] Tsung-yi Lin, Michael Maire, Serge Ippolito, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár和C劳伦斯Zitnick。Microsoft Coco: 上下文中的常见对象。在欧洲计算机愿景会议上, 第740–755页。斯普林斯, 2014年。[23] 魏刘, 龙玉米武奴罗维尔州, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu和Alexander C Berg。SSD: 单次拍摄多杆探测器。在欧洲在电脑愿景中的影响, 第21–37页。Springer, 2016年。[24] 魏刘, 胜凯辽, 和威勇湖。感知动态存储器的车辆检测中的动态记忆。电路技术的IEEE交易, 用于视频技术, 2019年。[25] 鑫璐, 李丽, 玉溪悦, 泉泉李, 俊杰燕。网格R-CNN。在IEEE愿景和模式识别的IEEE会议上, 2019年7363–7372, 2019。[26] Mahyar Najibi, Mohammad Rastegari和Larry S Davis。G-CNN: 基于迭代网格的物体检测器。在电脑愿景和模式识别上的IEEE会议上的记录中, 第2369–2377, 2016页。[27] Alejandro Newell和Jia Deng。Associata嵌入到图形的像素。在神经信息处理系统的进步中, 2017年第2171–2180, 2017页。[28] 江米鹿, 开辰, 建平史, 华军冯, 万利欧阳和大华林。天秤座R-CNN: 朝着对象检测的禁止学习。在计算机愿景和模式识别上的IEEE会议上, 第821–830, 2019页。[29] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary Devito, Zeming Lin, Al-Ban Desmaison, luca antiga和adam lerer。Pytorch中的自动差异化。2017年。

[30]鹤倾秋, 弘良李, 青江吴, 樊曼·孟, 国王Nggn, 恒中史。  
A2RMNET: 遥感图像中的对象检测的自适应方式多尺度网络。遥感, 11 (13): 1594, 2019。[31]贺倩邱, 弘良李, 青武, 樊曼梦, 林峰徐, 国王, 恒河史。分层配置文件嵌入用于对象检测。Muremedia, 2020年的IEEE经交。[32] Joseph Redmon, Santosh Divvala, Ross Girshick和Ali Farhadi。你只看一次: 统一, 实时对象模式。在电脑愿景和模式识别的IEEE会议上的讨论中, 2016年779-788, [33] Joseph Redmon和Ali Farhadi。  
YOLO9000: 更好, 更快, 更强。在计算机愿景和模式识别的IEEE会议上的讨论中, 2017年7263-7271, 2017。[34] Joseph Redmon和Ali Farhadi。YOLOV3: 增量改进。Arxiv预印迹arxiv: 1804.02767, 2018。[35]邵庆仁, 凯明, 罗斯吉伦克和剑孙。  
R-CNN更快: 与区域提案网络的实时对象检测。在神经信息的进展中, 2015年第91-99页的第91-99页。联盟的广义概括: 界限框回归的度量和损失。在IEEE愿景和模式识别的IEEE会议上, 2019年658-666, 2019。[37] Rasmus Rothe, Radu Timofte和Luc Van Gool。德克斯: 从单一形象的明显年龄的深入期望。在IEEE电脑视觉研讨会上的验收中, 2015年10-15, 第10-15页的第10-15页。[38]恒珊史, 河江李, 樊曼孟, 青武。用于引用表达式图像段的密钥-字感知网络。在欧洲电脑愿景 (ECCV) 会议上, 第38-54页, 2018年第38-54页。[39]恒珊史, 洪良李, 樊曼孟, 青博吴, 林峰徐和王尼吉。分层解析网: 从全局场景解析对象的语义场景。多媒体的IEEE交易, 20 (10): 2670-2682, 2018。[40]恒珊施, 洪良李, 青博武, 樊曼蒙和国王Ng。通过可释放的预测提高场景解析性能。在第26届ACM国际多媒体会议上, 第492-500页, 2018年第492-500页。[41]恒珊史, 弘良李, 青武和溜博宋。通过集成分类模型和基于方差的正规解析场景解析。在IEEE计算机愿景和模式识别会议上的诉讼程序中, 2019年第5307-5316页, 2019年。[42] Bugra Tekin, Sudipta Ninha和Pascal Fua。实时无缝单射6D对象姿态预测。在计算机愿景和模式识别上的IEEE会议上的报告中, 第292-301, 2018页。[43]志田, 春华沉, 郝辰和童。  
FCOS: 完全卷积的单阶段对象检测。在IEEE计算机舞台上的IEEE国际会议上的报告中, 2019年9627-9636, [44]斌杨, 朱杰燕, 镇磊和斯坦Z李。从图像中的曲目。在IEEE会议的诉讼中

在电脑视觉和模式识别上, 第6043-6051, 2016, 2016。[45] Tsun-yi Yang, Yi-Ting Chen, Yen-Yu Lin和Yung-yu Chuang。FSA-NET: 从单个图像中学习头部姿势估计的细粒度结构聚合。在计算机愿景和模式识别上的IEEE会议上, 2019年1087-1096, 2019年第1087-1096页。[46]嘉徽宇, 云宁江, 张阳王, 志宁曹, 托马斯黄。UnitBox: 高级对象检测网络。在第24届ACM国际会议上关于多媒体, 第516-520页的法律程序。ACM, 2016。[47]兴义周, Dequan Wang和Philipp Kr " ahenb " uhl。ob-jects作为积分。Arxiv预印迹Arxiv: 1904.07850, 2019。