文章编号：2095-2163(2019)06-0006-07 中图分类号：TP391 文献标志码：A

语境信息约束下的多目标检测网络

邬亚男，李君君，张彬彬
（合肥工业大学 计算机与信息学院，合肥 230601）

摘 要：目标检测问题一直是计算机视觉以及机器学习领域非常重要的研究课题，并且在交通监控、医学影像、辅助驾驶等方面有着广泛的应用。由于现实任务对于检测速度和精度的要求，目标检测一直是计算机视觉领域具有挑战性的任务。语境信息可以作为推理的关键证据应用于多目标识别领域。由此，提出语境信息约束下的直接预测目标类别和目标位置的多目标检测网络。该网络采取端对端的训练方式，分层提取特征，并利用语境信息微调网络的输出结果以更好地进行实时预测。在 PASCAL VOC 2007 数据集上的定性及定量实验结果，证明了深度语境网络下的目标检测模型具有显著的目标检测性能，优于当前先进的方法。实验证明，利用语境信息可以为目标检测提供有效的判定依据，提高检测的准确率。

关键词：目标检测；语境信息；实时检测；卷积神经网络

Multi-object detection network constrained by context information

WU Yanan, LI Junjun, ZHANG Binbin
(School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China)

【Abstract】 Object detection has always been a very important research topic in the field of computer vision and machine learning, and has a wide range of applications in traffic monitoring, medical imaging, and assisted driving. Object detection has always been a challenging task in the field of computer vision due to the requirements of real-world tasks for detection speed and accuracy. Context information can be applied to the field of multi-object detection as the key evidence of reasoning. Therefore, we propose a multi-object detection network that directly predicts object categories and object locations under the constraint of context information. The network adopts an end-to-end training method, hierarchically extracts features and uses context information to fine tune the output of the network to better predict in real time. The qualitative and quantitative experimental results on the PASCAL VOC 2007 dataset demonstrate that the object detection model under the deep context network has significant object detection performance, which is superior to the current advanced methods. Experiments show that the use of context information can provide effective evidence for object detection and improve the accuracy of detection.

【Key words】 Object detection; context information; real-time detection; convolutional neural network

0 引 言

目标检测的任务是集目标分类、目标定位两者之所长，检测输入图像中所有感兴趣的目标的类别属性和位置属性，输出相对应的概率标签，表明将目标分类为该类别的概率，明确输入图像中感兴趣物体的位置和范围，以矩形框表示物体的所在。目标检测问题一直是计算机视觉以及机器学习领域非常重要的研究课题，并且在视频监控[1]、行人检测[2]、行为识别[3]、场景理解[4]等方面有着广泛的应用。

传统目标检测模型主要由人工设计特征以及分类决策构成。通过人工设计特征表达，然后设计相应的分类器对目标进行检测。虽然这些手工制作的方法取得了令人瞩目的成功，但其在实践中不能灵活捕获图片信息，这可能会阻碍性能进一步提高。随着机器学习理论逐步完善以及深度学习技术的日益发展，深度网络模型不断发展壮大，对于特征的表达能力日益增强，检测精度也得以提升。目标检测任务从传统模型逐渐向基于深度学习的模型研究，涌现了一大批深度网络下的目标检测模型。尽管如此，由于现实任务高精度、高速度的目标检测需求，当前的目标检测结果仍然差强人意。因此，深度检测模型设计仍然面临着巨大的压力，仍然是亟待优化和解决的具有挑战性的研究课题。

语境线索在搜索和检测物体中有着重要作用，并且在计算机视觉和认知神经科学等方面有着重要的应用。语境有助于图像理解，符合现实世界的客观规律，语境信息对于人类识别物体也至关重要，计

作者简介：邬亚男(1993-)，女，硕士研究生，主要研究方向：图像与智能信息处理；李君君(1995-)，男，硕士研究生，主要研究方向：计算机视觉、人工智能、模式识别；张彬彬(1993-)，男，硕士研究生，主要研究方向：计算机视觉、图像分析与理解、模式识别。
通讯作者：邬亚男 Email：wyn19931106@163.com

收稿日期：2019-04-27

算机视觉的许多研究证明，通过适当的语境建模能够有效改进识别算法。由于视觉对象在其外观、动作等方面变化很大，通常难以仅使用局部线索来学习鲁棒模型。同时，由于物体几乎不是孤立地发生的，其语境信息，可以用来评估目标检测模型的输出并提高检测性能。本文的主要贡献如下：

(1)在 SSD 模型基础上，提出语境信息约束下的直接预测目标类别和目标位置的多目标检测网络，该网络采取端对端的训练方式，分层提取特征并进行实时的目标检测。

(2)采用语境信息作为约束条件，预测目标类别和目标位置，利用语境信息微调网络的输出结果，以更好地进行实时预测。

(3)在 PASCAL VOC 2007 数据集上的实验结果，证明了本文方法在公开数据集测试中具有显著的目标检测性能，优于当前先进的方法。

1　相关工作

针对图像目标检测问题，通常有两种常见的目标检测模型，一种为基于滑动窗口的目标检测模型，另一种为基于区域提议的目标检测模型。在卷积神经网络出现之前，DPM[5] 和选择性搜索[6] 受到了许多的关注。在 R-CNN[7] 结合选择性搜索、区域提议以及卷积神经网络带来显著改进后，基于区域提议的目标检测方法变得流行。

SPPnet[8] 显著加快了原有的 R-CNN 方法，其引入了一个空间金字塔池化层，对区域大小和尺度更加鲁棒，并允许分类层重用多个图像分辨率下生成的特征映射上计算的特征。Fast R-CNN[9] 扩展了 SPPnet，使得其可以通过最小化置信度和边界框回归的损失，来对所有层进行端到端的微调，并初次利用 MultiBox[10] 学习目标信息。然而，Fast R-CNN 仍然选择使用选择性搜索进行区域提议，浪费了太多的检测时间。据此，Faster R-CNN[11] 提出区域提议网络进行区域提议，并引入了一种方法，通过微调共享卷积层和预测层将区域提议网络和 Fast R-CNN 结合在一起，使用区域提议网络池化中级特征，提升了检测速度。

基于滑动窗口的目标检测模型完全跳过提出步骤，直接预测多个类别的边界框和置信度。OverFeat[12] 是首先利用滑动窗口进行目标检测的方法，在知道了底层目标类别的置信度之后，直接从最顶层的特征映射的每个位置预测边界框。之前常见的检测方法都将检测转换为分类问题，而 YOLO[13] 另辟蹊径，仅仅经过一个神经网络，将检测转换为回归问题，从而实现端对端优化。YOLO 使用整个最顶层的特征映射来预测多个类别和边界框(这些类别共享)的置信度。YOLO 经过不断更新多次优化升级得到目前的 YOLOv3[14]，并在设计上进行了一定的改进：首先融合先前的 Darknet-19 网络以及残差网络，设计出 DarkNet-53 网络进行特征提取，功能更加强大。此外，YOLOv3 还能够进行跨尺度预测，利用金字塔网络的概念预测出三个不同的尺度上边界框。

语境信息可以作为推理的关键证据应用于多目标识别领域。然而，上述研究忽略了语境信息的重要作用，仅仅利用设计的目标检测器检测对象类别及位置，结果可能会违反现实世界中的规律。在考虑语境信息之后，准确性得到了很大改善。在传统模型中，检测算法由人工设计特征及浅层分类器构成。语境信息可作为正则化约束条件[15-16]，调整检测结果以提高性能，也可以约束深度学习模型，利用这种丰富的且有区别的语境信息有助于机器获取行为发生时相应的场景信息，获得图像内容的理解，提高检测的准确度。例如，Bell[17] 等人也分别对语境和外部场景进行了建模。利用空间循环神经网络分别对感兴趣区域的外部环境整合了多尺度语境，有助于特定的小目标检测。Zhe [18] 等人通过可学习的直方图层在端到端训练中学习深度神经网络中的统计语境特征，将可学习的直方图层集成到深层网络中，探索了语义分割和目标检测两个视觉问题。Heilbron[19] 等人提出语境级联模型，通过采用与人类活动相关的语义先验，语境级联模型产生高质量的特定类别的行动提议，并通过级联的方式抑制无关的活动提议。

2　语境信息约束下的多目标检测网络

利用语境信息作为约束条件，能够准确且有效地捕捉图片中除了目标物本身之外的所有信息(包括其它目标信息和背景信息)。语境信息作为目标检测推理过程的关键证据，具有重要的作用及意义。以此作为切入点，构建语境信息约束下实时的多目标检测网络，如图 1 所示。该网络分层提取特征并依次进行边框回归和分类，从而得到图像中所有感兴趣的目标类别属性和位置属性。

语境信息约束下的多目标检测网络的主要贡献如下：

(1)在 SSD 模型基础上进行改进，提出语境信

息约束下能够端对端训练的多目标检测网络，并依次进行边框回归和分类。

(2)采用语境信息约束网络输出结果，微调网络的输出结果以更好地进行实时预测。

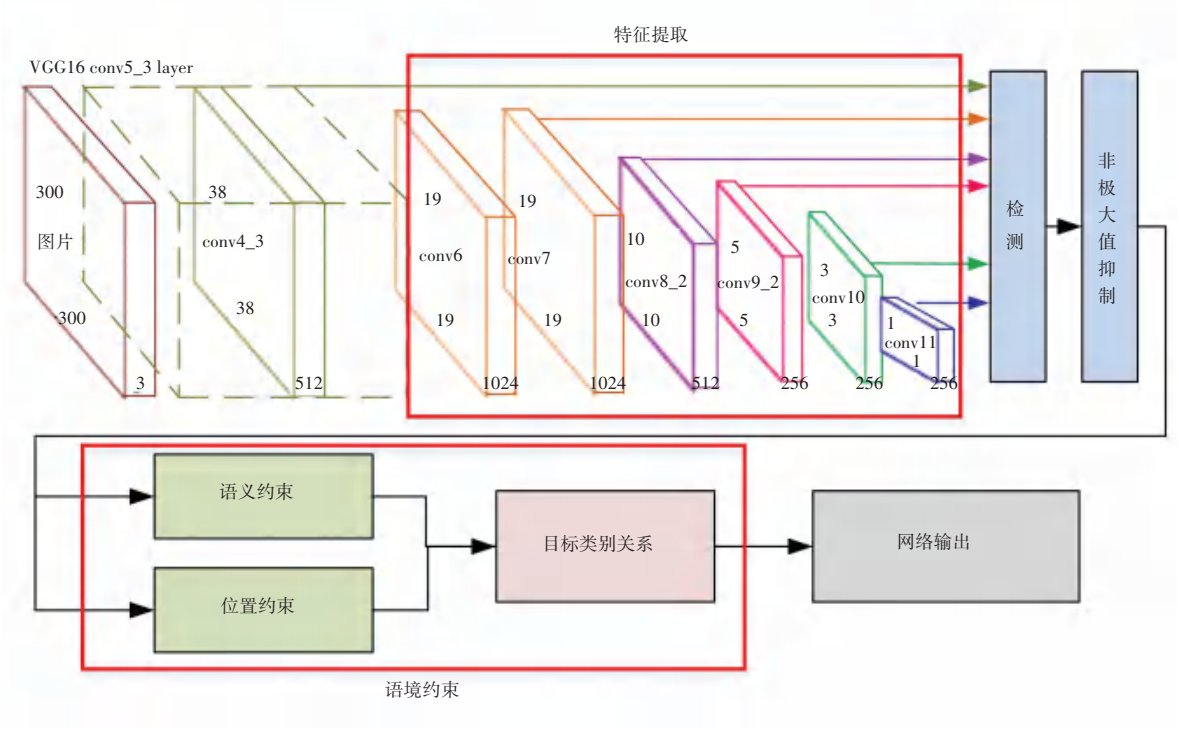

图 1 语境信息约束下的多目标检测网络

Fig. 1 Multi-object detection network constrained by context information

2.1 语境约束

多目标检测网络经过初步训练，将得到训练集图像所有候选框中目标的语义类别、标签以及所有候选框的位置坐标。由此，可以计算出候选框中心位置点的坐标。已知语义类别集合，可得候选目标 o 的位置属性 v，以及候选目标的语义属性 a。定义语义存在矩阵，统计每一幅图像中出现的类别，语义存在矩阵，统计所有训练集图像中同时出现的类别，对即可得语义类别共现频率矩阵，统计训练集图像中同时出现的类别频率。由候选目标的语义属性以及语义类别共现频率矩阵作为语义信息获取目标类别关系。目标类别之间语义约束置信度获取方法如下：

$$c_{semantic}(a_m,a_n)=\frac{\sum_{I^{(n)}\in I_{\text{train}}}\left(\sum_{v_i}\mathbb{I}(o_{v_i}=a_m)\times\sum_{v_j}\mathbb{I}(o_{v_j}=a_n)\right)}{\sum_{I^{(n)}\in I_{\text{train}}}\sum_{v_j}\mathbb{I}(o_{v_j}=a_n)},\tag{1}$$

其中，I 表示训练集图像 $I^{(n)}\in I_{\text{train}}$；$a$ 表示语义类别集合 $a_m,a_n\in a$；v 表示候选框 $v_i,v_j\in v$；o 候选框中目标的语义类别标签；I 表示计数函数。

由目标对的位置信息可计算类别间的相对位置信息。该信息是一个向量，包含两个类别间的距离和角度信息，由目标对的相对位置信息可计算类别间相对位置，分别作为位置信息获取目标类别关系。目标类别之间位置约束置信度获取方法如下：

$$\vec{R}_{i,j,m,n}=[\Delta x_{ij},\Delta y_{ij}],o_{v_i}=a_m,o_{v_j}=a_n.\tag{2}$$

$$c_{location}(a_m,a_n)=c_{semantic}(a_m,a_n)\cdot\sum_{I^{(n)}\in I_{\text{train}}}f(\vec{R}_{i,j,m,n},\mu,\sigma^2).\tag{3}$$

其中，$[x_i,y_i]$ 表示候选框的中心位置坐标；$[\Delta x_{ij},\Delta y_{ij}]$ 表示属于两个类别 a_m,a_n 之间的候选框 v_i,v_j 相对位置；u 为目标对的相对位置均值；σ^2 为目标对的相对位置方差；f 为标准正态分布函数。

根据捕获的目标类别关系，微调候选目标框的类别得分。通过语境约束 $c_{semantic}$ 以及 $c_{location}$ 判断后，对于每张图片的每个目标，考虑所有与之相关的候选框类别，得到最终类别置信度 c。

2.2 网络模型

语境信息约束下的多目标检测网络与 Faster R-CNN 中的区域提议网络非常相似，也使用了一组固定的边界框进行预测，类似于 RPN 中的锚边界

框。但并不是使用这些来池化特征并评估另一个分类器，而是为每个目标类别在每个边界框中同时生成一个分数。因此，本文的方法避免了将区域提议网络与 Fast R-CNN 合并的复杂性，并且更容易训练，更快且更直接地集成到其它任务中。语境信息约束下的多目标检测网络使用默认边界框，方法比现有方法更灵活，可以在不同尺度的多个特征映射的每个特征位置上使用不同长宽比的默认边界框。如果只从最顶层的特征映射的每个位置使用一个默认框，语境信息约束下的多目标检测网络将具有与 OverFeat 相似的架构，若使用整个最顶层的特征映射，并添加一个全连接层进行预测来代替卷积预测器，并且没有明确地考虑多个长宽比，将近似地再现 YOLO。

利用在 ILSVRC CLS-LOC 数据集上已经预先训练好的 VGG16 网络，作为基础卷积神经网络架构。通过多层卷积分层提取特征，直接检测图像中的目标，最终获得固定尺度的边界框以及对象的类别。将多个卷积层添加到 VGG16 第五个卷积层之后，VGG 第六、七层原本为全连接层，将其替换为卷积层，从其中重采样参数。然后继续增加 4 个卷积层，所以总共含有 6 种尺度的卷积层，尺寸逐渐减小，每种尺度的卷积层都能学习到一个特征图，将 6 种尺度卷积层学习的多个特征图预测组合在一起，以便于针对不同尺度的目标检测问题。SSD 模型训练一开始，需要将真实信息分配给固定的检测器，输出集合中的特定输出，一旦确定了这个分配，损失函数和反向传播就可以应用端到端了。通过计算交叠面积，保证唯一的默认框与真实标签一一对应，之后只要交叠面积大于设定的阈值就将默认框与真实标签进行随机匹配，使网络可以预测同个位置的多个框。测试时，根据训练好的模型得到每个候选框中的目标类别的预测分数，同时对候选框进行调整以便于更好地匹配目标形状。使用随机梯度下降，对得到的语境信息约束下的实时多目标检测网络模型进行微调，初始学习率为 10^{-3}，动量为 0.9，权重衰减为 0.0005，批处理数据大小为 32。针对每个数据集的学习速率衰减策略略有不同，本文使用 Caffe 框架[20]，实验环境硬件平台配置为 CoreX i7-6800k 6 核、3.4 GHz CPU、2 块 NIVDIA GTX1080 8 GB 显卡。整个语境信息约束下的实时多目标检测网络训练过程的损失函数为：

$$L(x,c,l,g)=\frac{1}{N}(L_{conf}(x,c)+\alpha L_{loc}(x,l,g)). \tag{4}$$

其中，x 为网络输入；c 为类别置信度；l 和 g 分别为预测框和真实标签；L_{conf} 代表置信损失，置信度损失是在多类别置信度上的 softmax 损失；L_{loc} 代表定位损失，定位损失是预测框与真实框参数之间的 Smooth L1 损失。

3　实验结果与分析

3.1　数据集与评价标准

由于 PASCAL VOC 2007 数据集是一个有着明显层级结构的多类别、多目标类型的数据集，类别、目标之间具有明显的共生关系。所以本文选择在 PASCAL VOC 2007 数据集上评估 DCN 方法。该数据集包括 20 个对象类别，9 963 张高质量图片。其中 5 011 张训练和验证集图像，4 952 张测试图像，包含 24 640 个已标注的目标，该数据集为图像分割、目标识别和目标检测提供标准。为了证明 DCN 方法的检测精度，本文采用类别平均精准度（AP，Average Precision）值和平均 AP（mean AP）值进行评测。

$$\begin{cases} \mathrm{Precision}_{c_i}=\dfrac{N(TruePositives)_{c_i}}{N(TotalObjects)_{c_i}}, \\ AP_{c_i}=\dfrac{\sum \mathrm{Precision}_{c_i}}{N(TotalImages)_{c_i}}, \\ mAP=\dfrac{\sum AP_{c_i}}{N(Classes)}. \end{cases} \tag{5}$$

其中，$N(TruePositives)_{c_i}$ 为正确预测的数量；$N(TotalObjects)_{c_i}$ 该图像中给定类别 c_i 的实际目标的数量；$\mathrm{Precision}_{c_i}$ 为类 c_i 的精度。平均 AP 值即为所有类别的平均精度求和除以所有类别。

实验方法中，如果预测窗口与数据集标定的 ground-truth 窗口面积 IoU 重叠大于 50%（mAP@IoU=0.5），则认为此提议窗口的检测是正确的，否则是错误的。

3.2　实验结果分析

本文在 PASCAL VOC 2007 数据集上进行实验验证，并与当前先进检测方法 R-CNN[7]、Fast R-CNN[9]、Faster R-CNN[11]、G-CNN[21]、OHEM[22]、SSD[23] 进行对比，以验证语境信息约束下的多目标检测网络方法的目标检测效果。

表 1 给出了不同线索下语境信息约束下，多目标检测网络在 PASCAL VOC 2007 数据集上得到平均准确率值，以证明不同线索对检测结果的影响。其中 semantic 代表语义约束，location 代表位置约

束。由表 1 可得,语境信息约束分别作用以及共同作用下的多目标检测网络对于检测精度的提升效果。

表 1 不同线索在 PASCAL VOC 2007 数据集下的平均检测精度 mAP(%)

Tab. 1 Detection Results on VOC 2007 test under different clues

clue	mAP
SSD[23]	68.0
SSD+semantic	70.2
SSD+location	69.9
ours	72.1

表 2 给出了语境信息约束下的多目标检测网络与对比方法,分别在 PASCAL VOC 2007 数据集上的平均 AP 值以及 20 个类别条件下得到平均准确率值。由表 2 可得,在 PASCAL VOC 2007 数据集下以及 20 个类别条件下得到的平均准确率,总体优于当前先进方法。

语境信息约束下的多目标检测网络能够在一定程度上处理漏检(chair、bird)、误检(potted plant、sofa)等问题,针对检测错误以及不准确等问题进行修正,提升检测的精准度,具有更好的检测效果,如图 2 所示。

表 2 不同方法在 PASCAL VOC 2007 数据集下的平均检测精度 mAP(%)

Tab. 2 Detection Results on VOC 2007 test under different methods

method	R-CNN[7]	Fast R-CNN[9]	Faster R-CNN[11]	G-CNN[21]	OHEM[22]	SSD300[23]	ours
aero	68.1	**74.5**	70.0	68.3	71.2	73.4	73.6
bicycle	72.8	78.3	80.6	77.3	78.3	77.5	**82.7**
bird	56.8	69.2	**70.1**	68.5	69.2	64.1	68.9
boat	43.0	53.2	57.3	52.4	57.9	59.0	**59.1**
bottle	36.8	36.6	49.9	38.6	46.5	38.9	**51.9**
bus	66.3	77.3	78.2	78.5	81.8	75.2	**83.1**
car	74.2	78.2	79.4	79.5	79.1	**80.8**	80.4
cat	67.6	82.0	83.0	81.0	83.2	78.5	**92.6**
chair	34.4	40.7	52.2	47.1	47.9	46.0	**53.0**
cow	63.5	72.7	75.3	73.6	**76.2**	67.8	75.8
table	54.5	67.9	67.2	64.5	68.9	**69.2**	68.7
dog	61.2	79.6	80.3	77.2	**83.2**	76.6	82.7
horse	69.1	79.2	79.8	80.5	80.8	82.1	**87.1**
mbike	68.6	73.0	75.0	75.8	75.8	**77.0**	70.9
person	58.7	69.0	76.3	66.6	72.7	72.5	**81.0**
plant	33.4	30.1	39.1	34.3	39.9	**41.2**	36.7
sheep	62.9	65.4	68.3	65.2	67.5	64.2	**69.2**
sofa	51.1	**70.2**	67.3	64.4	66.2	69.1	66.9
train	62.5	75.8	**81.1**	75.6	75.6	78.0	77.1
tv	64.8	65.8	67.6	66.4	75.9	68.5	**80.0**
mAP	58.5	66.9	69.9	66.8	69.9	68.0	**72.1**

图 3 给出了 PASCAL VOC 2007 数据集上 6 个示例类别的平均精准度柱状图结果,验证了利用语境信息约束能够提升多目标检测网络的检测效果,语义约束和位置约束对于目标检测有着重要的作

用。

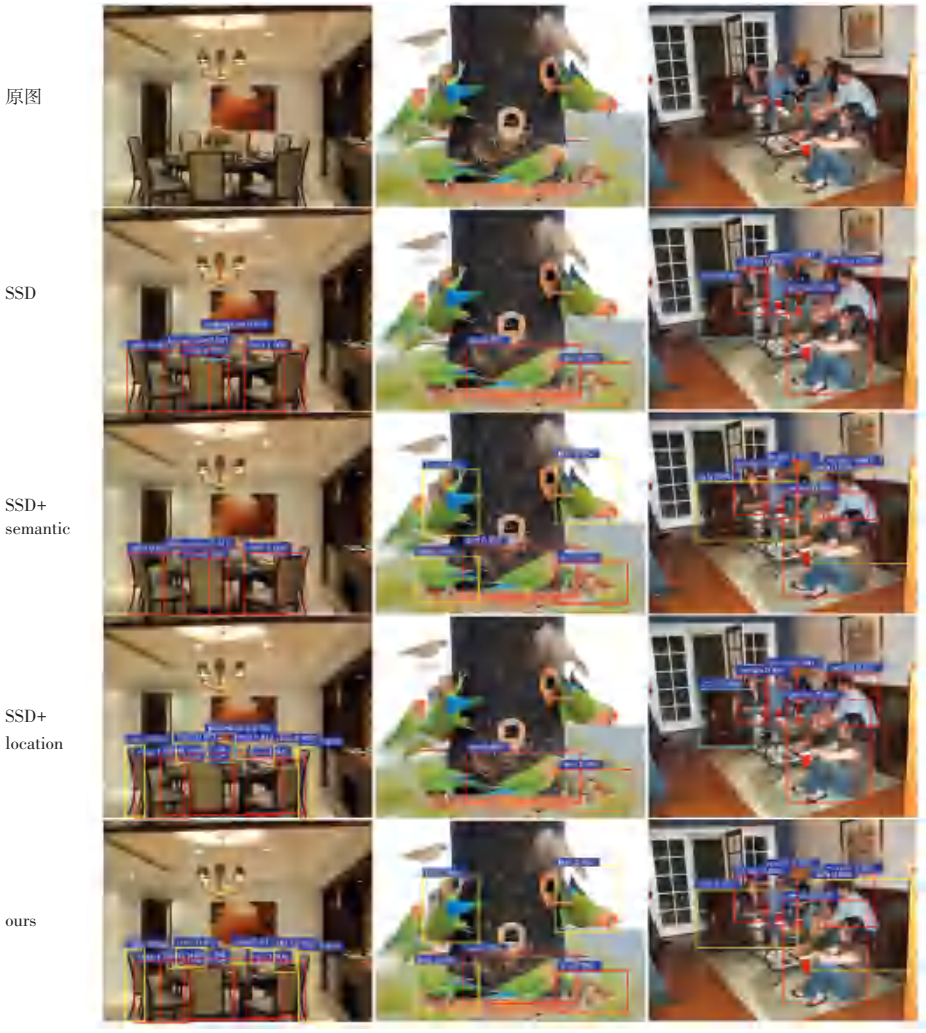

图 2　PASCAL VOC 2007 数据集目标检测实例

Fig. 2　Object detection instance in PASCAL VOC 2007 dataset

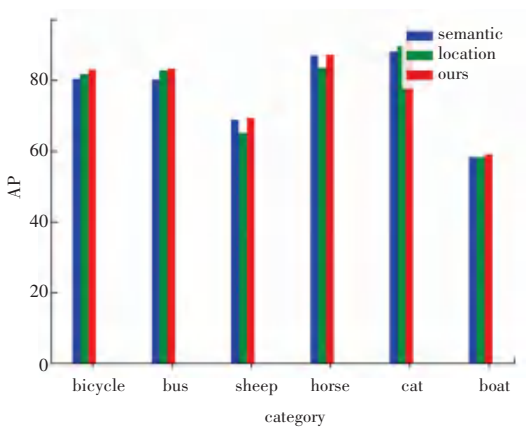

图 3　PASCAL VOC 2007 数据集类别平均精准度

Fig. 3　PASCAL VOC 2007 dataset category average precision

4　结束语

本文介绍了语境信息约束下的多目标检测网络，是一种快速的单次多类别目标检测器，模型的关键特性是使用网络顶部多个特征映射的多尺度卷积边界框输出。这种表示能够高效地建模可能的边界框形状空间。语境信息约束下的多目标检测网络在准确性和速度方面与其对应的最先进的目标检测器相比毫不逊色。在 PASCAL VOC 2007 数据集上的实验结果证明了本文方法在公开数据集测试中具有显著的目标检测性能，提高了检测精度，优于当前先进的方法。在此基础上仍然存在许多可以深入研究的方向，其中有前景的未来方向是探索其作为系统的一部分，使该模型作为目标检测组件的大型系统有用的构建模块，同时检测和跟踪视频中的目标。

参考文献

[1] 傅赟，王桂丽，周旭廷，等. 交通监控系统中视频运动目标检测算法研究[J]. 计算机技术与发展，2017，27(7)：156-158.

[2] 王国华，刘琼，庄家俊. 基于局部特征的车载红外行人检测方法研究[J]. 电子学报，2015，43(7)：1444-1448.

[3] 马钰锡,谭励,董旭,于重重. 面向智能监控的行为识别[J]. 中国图像图形学报, 2019, 24(2): 132-144.
[4] 庄严, 陈东, 王伟, 等. 移动机器人基于视觉室外自然场景理解的研究与进展[J]. 自动化学报, 2017, 36(1): 1-11.
[5] FELZENSZWALB P F, MCALLESTER D A, RAMANAN D. A Discriminatively Trained, Multiscale, Deformable Part Model [C]// IEEE Computer Conference on Computer Vision and Pattern Recognition, 2008: 24-26.
[6] Uijlings J R R , K. E. A. van de Sande. Selective Search for Object Recognition[J]. International Journal of Computer Vision, 2013, 104(2):154-171.
[7] GIRSHICK R, DONAHUE J, DARRELLAND T, et al. Rich feature hierarchies for object detection and semantic segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
[8] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9):1904-16.
[9] Ross B. Girshick . Fast R-CNN [C]// IEEE International Conference onComputer Vision, 2015:1440-1448.
[10] ERHAN D, SZEGEDY C, TOSHEV A, et al. Scalable Object Detection Using Deep Neural Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2014: 2155-2162.
[11] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016: 39(6):1137-1149.
[12] SERMANET P, EIGEN D, ZHANG X, et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2014: 21-32.
[13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//IEEE Conference on Computer Vision & Pattern Recognition, 2016: 779-788.
[14] REDMON J, FARHADI A. YOLOv3: an incremental improvement [C]//IEEE Conference on Computer Vision & Pattern Recognition. 2018: 47-67.
[15] CHOI W, CHAO Y W, PANTOFARU C, et al. Indoor Scene Understanding with Geometric and Semantic Contexts [J]. International Journal of Computer Vision, 2015, 112(2):204-220.
[16] SUNGBAEK Y, HYUNJIN P, YI J. An Efficient Bayesian Approach to Exploit the Context of Object-Action Interaction for Object Recognition:[J]. Sensors, 2016, 16(7):981.
[17] BELL S, ZITNICK C L, BALA K, et al. Inside-Outside Net: Detecting Objects in Context with Skip Pooling and Recurrent Neural Networks[J]. 2015.
[18] ZHE W, LI H, OUYANG W, et al. Learnable Histogram: Statistical Context Features for Deep Neural Networks [C]// European Conference on Computer Vision,2016: 246-262.
[19] HEILBRON F C, BARRIOS W, ESCORCIA V, et al. SCC: Semantic Context Cascade for Efficient Action Detection [C]// IEEE Conference on Computer Vision & Pattern Recognition, 2017:3175-3184.
[20] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[J]. ACM Multimedia 2014: 675-678.
[21] NAJIBI M, RASTEGARI M, DAVIS L S. G-CNN: An Iterative Grid Based Object Detector [C]// Proceedings of IEEE Conference on Computer Vision & Pattern Recognition, 2016:2369-2377.
[22] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training Region-Based Object Detectors with Online Hard Example Mining[C]// Proceedings of IEEE Conference on Computer Vision& Pattern Recognition, 2016:761-769.
[23] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// Proceedings of European Conference on Computer Vision, 2016: 21-37.

(上接第5页)

[3] Kruskal J B. On the shortest spanning subtree of a graph and the traveling salesman problem [J]. Proceedings of the American Mathematical society, 1956, 7(1): 48-50.
[4] Zou Z, Gao H, Li J. Discovering frequent subgraphs over uncertain graph databases under probabilistic semantics [C]// Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010: 633-642.
[5] Provan J S, Ball M O. The complexity of counting cuts and of computing the probability that a graph is connected[J]. SIAM Journal on Computing, 1983, 12(4): 777-788.
[6] Sevon P, Eronen L, Hintsanen P, et al. Link discovery in graphs derived from biological databases[C]//Data Integration in the Life Sciences. Springer Berlin Heidelberg, 2006: 35-49.
[7] Kamousi P, Suri S. Stochastic Minimum Spanning Trees and Related Problems[C]//ANALCO. 2011: 107-116.
[8] Huang L, Li J. Minimum spanning trees, perfect matchings and cycle covers over stochastic points in metric spaces[J]. arXiv preprint arXiv:1209.5828, 2012.
[9] Frieze A M. On the value of a random minimum spanning tree problem[J]. Discrete Applied Mathematics, 1985, 10(1): 47-56.
[10] Steele J M. On Frieze´s χ (3) limit for lengths of minimal spanning trees[J]. Discrete Applied Mathematics, 1987, 18(1): 99-103.
[11] Ball M O. Computational complexity of network reliability analysis: An overview[J]. Reliability, IEEE Transactions on, 1986, 35(3): 230-239.
[12] Provan J S, Ball M O. The complexity of counting cuts and of computing the probability that a graph is connected[J]. SIAM Journal on Computing, 1983, 12(4): 777-788.
[13] Soliman M A, Ilyas I F, Chen-Chuan Chang K. Top-k query processing in uncertain databases[C]//Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on. IEEE, 2007: 896-905.
[14] Pei J, Jiang B, Lin X, et al. Probabilistic skylines on uncertain data[C]//Proceedings of the 33rd international conference on Very large data bases. VLDB Endowment, 2007: 15-26.
[15] Jin R, Liu L, Aggarwal C C. Discovering highly reliable subgraphs in uncertain graphs[C]//Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2011: 992-1000.