特征融合与分割引导的弱监督目标检测

柴文光,蔡春波

(广东工业大学 计算机学院,广东 广州 510006)

摘　要:基于卷积神经网络(CNNs)的区域建议生成方法(PRN)是通过实例级注释进行训练所得,也是当前全监督目标检测(FSOD)的重要组成部分。由于实例级注释耗时耗力,而图像级注释相比之下更容易收集,因此仅使用图像级注释的弱监督目标检测(WSOD)引起了众多研究者的关注。当前,WSOD依赖于诸如选择性搜索之类标准的区域建议生成方法,这些方法容易生成大量有噪的建议框,导致其存在无法拟合真实的目标对象。鉴于此,基于卷积特征多层融合以及分割引导策略获取高质量建议框,具体而言,利用卷积网络深层信息进行多层融合,以及边缘信息获取初始的候选建议框,然后通过弱监督语义分割的一致性准则,将分割映射分为水平和垂直两个变量得到目标一致性表示,从而提取高质量的建议框。在PASCAI VOC2007数据集上的实验结果表明,该方法在分类和定位检测中展现了优秀的性能,平均精度(mAP)和定位精度(CorLoc)准确率分别达51.0%、71.2%。

关键词:特征融合;分割引导;目标一致性;弱监督目标检测

DOI:10.11907/rjdk.211291　　开放科学(资源服务)标识码(OSID):

中图分类号:TP301　　**文献标识码**:A　　**文章编号**:1672-7800(2022)001-0114-06

Feature Fusion and Segmentation Guided Weakly Supervised Object Detection

CHAI Wen-guang, CAI Chun-bo

(*School of Computer Science*, *Guangdong University of Technology*, *Guangzhou* 510006, *China*)

Abstract: The region proposal generation method (ie, PRN) based on convolutional neural networks (CNNs) is trained through instance-level annotations, and is also an important part of the current fully supervised target detection (FSOD). Because instance-level annotations are time-consuming and labor-intensive, while image-level annotations are easier to collect, so weakly supervised object detection (WSOD) that only uses image-level annotations has attracted the attention of many researchers. The current WSOD relies on standard region proposal generation methods such as selective search. These methods are prone to generate a large number of noisy proposal boxes, resulting in their existence that cannot fit the real target object. This paper is based on the multi-layer fusion of convolutional features and the segmentation guidance strategy to obtain high-quality proposal boxes. Specifically, the deep information of the convolutional network is used for multi-layer fusion, and the edge information is used to obtain the initial candidate proposal boxes, and then through weakly supervised semantic segmentation The consistency criterion divides the segmentation map into two variables, horizontal and vertical, to obtain the target consistency representation, thereby extracting high-quality proposal boxes. The experimental results on the PASCAI VOC2007 dataset show that the method in this paper exhibits excellent performance in classification and localization detection, with mean of average precision (mAP) and localization (CorLoc) reaching 51.0% and 71.2% accuracy rates, respectively.

Key Words: feature fusion; segmentation guidance; object consistency; weakly supervised object detection

0　引言

近十年来,卷积神经网络(Convolutional Neural Networks,CNN)的发展与大规模、带精确实例级注释的训练数据集(PASCAI VOC[1])的应用,极大地推动了计算机视觉领域的进步。其中,目标检测方向更是在获得飞速发展的同时被广泛应用于AR、人脸识别等具体场景中[2-3]。然而,获

收稿日期:2021-03-04

基金项目:国家自然科学基金项目(61907009)

作者简介:柴文光(1969-),男,博士,广东工业大学计算机学院副教授、硕士生导师,研究方向为数字水务、IT治理、物联网;蔡春波(1994-),男,广东工业大学计算机学院硕士研究生,研究方向为图像处理、目标检测。

取此类数据集往往费时费力，这就阻碍了目标检测技术的进一步发展。故相较于实例级注释，更容易被收集的图像级注释被提出并被进一步运用于目标检测中。因此，利用图像级注释训练检测器的弱监督目标检测（Weakly Supervised Object Detection，WSOD）[4-5]方法成为近年来目标检测领域的研究热点。众多研究者的加入使得WSOD方法不断地被改进、优化，令其在识别速度或是识别精度上都得到了大幅提升。

最初的WSOD方法大多基于多实例学习（Multiple-Instance Learning，MIL）[6]，其主要是将图像作为一个包（其中一个正包中至少包含一个正实例，而一个负包中的所有实例皆为负实例）、对象建议作为实例，利用这些包训练实现弱监督目标检测器。但是在MIL中，对于来自同一类的多个对象实例中特定分数较低的对象实例极有可能被判为负实例。此类情况下，所选择的对象实例外观变化与大小差异较小，不足以训练出具有较强辨别能力的检测分类器。此外，在训练过程中，缺失的实例同样可能会被选择为负实例，这将进一步降低检测分类器的识别能力。针对这一问题，近期研究者依据CNN强大的表征学习能力，提出端到端MIL网络，如OICR（Online Instance Classifier Refinement）、PCL（Proposal Cluster Learning）等。在端到端MIL网络中，其将实例分类检测问题视为潜在的包内模型学习（图像分类）问题，通过实例级标签训练分类器以区分正负实例，提取得分最高的正实例。但由于在WSOD中图像缺乏对象实例级标签，使得WSOD与全监督目标检测（Fully Supervised Object Detection，FSOD）方法之间存在巨大的性能差异。

当前，主流的WSOD方法是利用区域建议生成包含对象的建议，而后根据该建议的特征进行实例学习，实现图像分类与定位。因此，如何获取高质量的区域建议成为目前WSOD的一大挑战。针对这一挑战，本文提出基于卷积特征多层融合以及分割引导策略的候选建议框生成方法用于获取高质量的区域建议，使得这些建议尽可能多地包含目标。该方法具体过程如图1所示：①结合EdgeBoxes选取初始候选框，同时对VGG16网络各层的特征映射进行融合，利用其边缘结构特征过滤掉无关的建议框；②结合2阶段获取的建议框经过弱监督语义分割网络，计算分割映射每个建议框得分。此处考虑到一致性准则（只要建议框在目标边界内，则所有建议框的得分应一致），本文给出一种新的目标一致性表示方式。具体地，将分割置信映射投影到水平和垂直两个方向，通过计算得到目标一致性表示。在新的表示中，只要建议框不超过目标边界，得分就始终保持一致。最后，进行MIL以完成图像定位及分类。相比当前的弱监督目标检测方法，本文方法在具备优秀性能的同时能够轻易地嵌入到其他弱监督目标检测中。

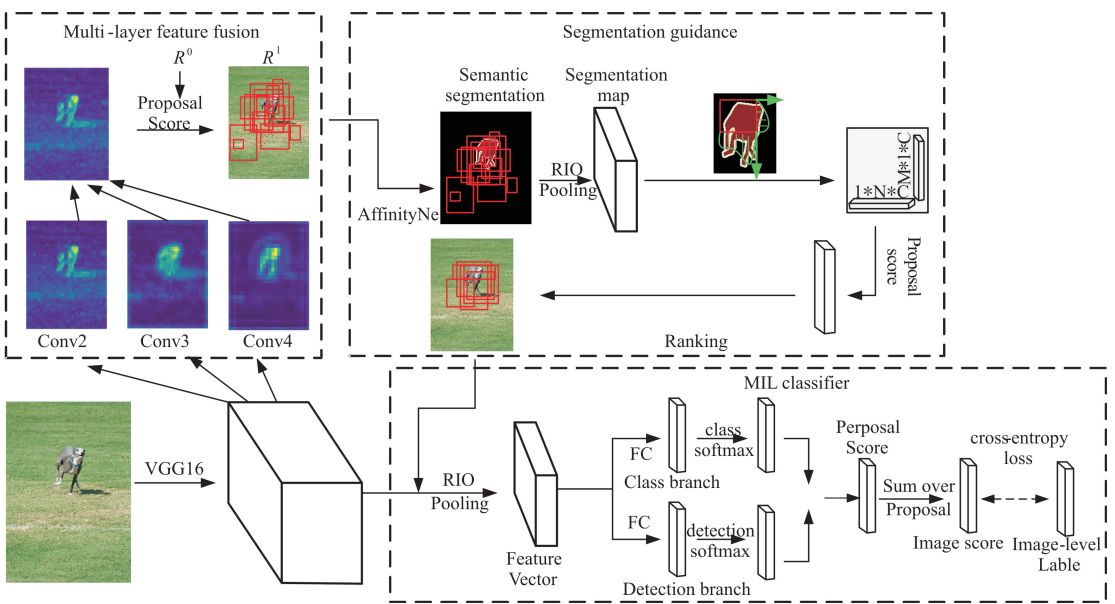

Fig. 1　Specific process of the proposed method

图1　本文方法具体过程

1　相关工作

1.1　弱监督目标检测

通过在卷积神经网络（CNN）中加入多实例学习（MIL），近年来WSOD性能得到了显著提高。Bilen等[4]率先在MIL中引入端到端弱监督深度检测网络（Weakly Supervised Deep Detection Networks，WSDDN），这激发了许多研究者的兴趣；Tang等[7]通过引入在线实例分类器优化（OICR），进一步改善了WSOD的工作，但是其在优化过程中容易陷入局部优化；Tang等[8]基于图中心聚类和平均MIL损失函数的方法进一步改进了OICR，以提高WSOD的检测性能；Wan等[9]提出一种最小熵潜模型，通过最小化全局和局部熵实现对目标进行分类和定位；Wan等[10]将连续

优化方法引入MIL中，利用一组平滑的损失函数松弛原始损失函数，成功解决了MIL中的非凸优化问题。相比之下，本文采用特征融合及目标一致性表示的方法，仅获取少数高质量的建议框就能完成目标检测器训练，并且该目标检测器同样具备优秀的检测性能。

1.2 区域建议生成

关于区域建议生成常用的两种方法是选择性搜索(SS)[11]和边缘盒(EB)[12]。其中，SS基于超像素合并方法生成建议，但其生成的建议对物体真实边界框的拟合效果普遍不佳。而EB则通过提取图像边缘，然后评估滑动窗口框的客观得分并生成建议。目前，已有的全监督目标检测都是基于CNN的区域建议生成方法[13]，使用边界框注释作为监督以训练建议网络，利用其精确的注释将边界框回归到真实的目标位置。但是为了保证其高性能，这些方法需要通过有效的边框注释甚至像素级标注[14]以训练网络模型，这与WSOD在训练时只使用图像级注释的要求有所差异。因此，弱监督训练下生成高质量的建议框更具有挑战性。针对这一挑战，本文基于EB方法，利用CNN中的低层信息生成的边缘反馈映射进行多层融合[15]，并结合语义分割[16]引导获取高质量的区域建议。实验结果表明，该方法取得了良好的WSOD性能。

1.3 弱监督语义分割

鉴于弱监督语义分割的优秀特性，Diba等[17]首次将语义分割嵌入WSOD中，提出利用分段知识对目标分数较低的建议框进行过滤的级联卷积网络，但其无法有效地对仅包含部分对象的高置信框进行过滤；Wei等[18]则通过引入两种分割度量，高纯度和完整地完成对建议框的挖掘；Gao等[19]利用分割映射生成具有丰富上下文信息的实例建议；Li等[20]将弱监督目标检测和分割任务整合到检测—分割循环协同这一多任务学习框架中。本文采用语义分割引导，对分割置信映射进行了修正，以减少外部无关信息的影响，从而获取高质量的候选建议框。具体而言，通过特征融合得到候选建议框，而后采用Ahn等[21]提出的AffinityNet提前生成弱监督语义分割结果，结合目标一致性计算每个建议框的客观分数，再将计算结果进行排序，选取得分高的建议作为输入，并将其输入到弱监督目标检测网络训练分类器。

2 高质量候选区域生成

2.1 特征融合

多层特征融合：给定图像，一方面通过EdgeBoxes(EB)获得初始的候选建议框$R^0=\{R_1^0,R_2^0,\cdots,R_n^0\}$，另一方面，利用VGG16[22]网络结构前向传播计算不同卷积层在通道维度上的平均值进而得到卷积的响应映射，并将其调整到原始图像大小进行多层信息融合，得到图像的特征映射；而后将R^0映射到图像特征，评估初始建议的客观分数，剔除大部分和背景相应的框。如图2所示，后面的层往往会响应更多的语义特征，提供局部对象的有用信息，这些层的响应映射与显著图相似。利用第二层到第四层特征为建议框生成类似边缘的响应映射。基于卷积层输出的特征映射$F\in\mathbb{R}_{C\times W\times H}$，其中C、W、H分别代表通道、宽、高，响应映射$M\in R_{W\times H}$由式(1)计算所得，$f_{cwh}$、$m_{wh}$分别代表$F$、$M$，计算每个通道的平均值，然后归一化。

$$r_{wh}=\frac{1}{C}\sum_{1}^{C}f_{cwh},r_{wh}\leftarrow\frac{r_{wh}}{max_{w',h'}r_{w'h'}} \tag{1}$$

首先，上述每层卷积特征调整到原始图像大小并将它们相加，得到类似边缘的特征映射；然后，计算每个初始建议框中存在的边的数量以评估初始建议框R^0的客观得分，并执行非极大值抑制(Non-Maximum Suppression，NMS)；最后，根据客观得分对候选建议框进行排序，选取得分高的框，得到新的集合$R^1=\{R_1^1,R_2^1,...,R_n^1\}$。

2.2 目标一致性表示

通过多层融合生成的候选框仍具有一定的非相关域，因为类边缘的影响映射在背景区域上也有较高响应。为了解决该问题，本文结合分割图像映射选取高质量目标建议，将第一阶段生成的候选框作为输入，经过AffinityNet生成语义分割结果，从而获得目标建议的集合$R^2=\{R_1^2,R_2^2,\cdots,R_n^2\}$，$R_i\in\mathbb{R}_{W_i\times H_i\times(C+1)}$，$W_i$、$H_i$表示第$i$个目标建议的宽和高，$C$表示目标类别。计算分割映射目标建议的客观分数$OR_i$如式(2)所示。

$$OR_i=f(R_i) \tag{2}$$

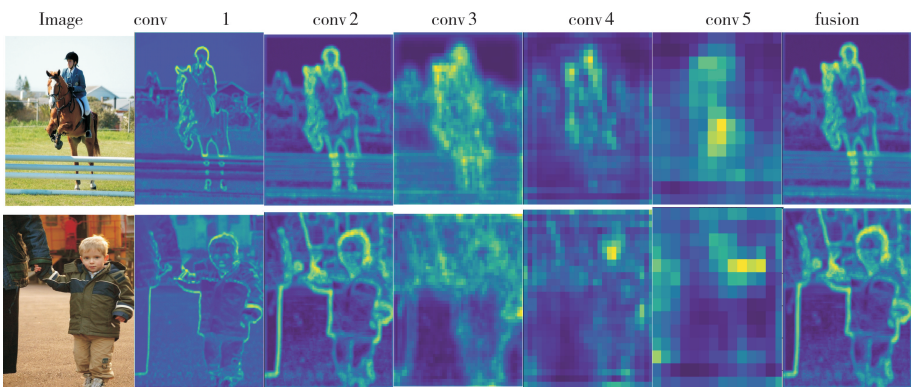

Fig. 2 Response and fusion of different convolutional layers of VGG16 network

图2 VGG16网络不同卷积层的响应及融合

$f(\cdot)$是计算目标分数的关键因素，f函数的不同表示方法主要有两种：一种是基于分割的上下文方法，一种是目标一致性表示方法。为了获得更多完整目标，文献[17]选择边界框内具有较高建议分数和周围上下文区域较低分数的框。

$$OR_i = avg(R_i) - avg(\hat{R}_i) \tag{3}$$

$\hat{R}_i$是R_i周围上下文区域，当框靠近目标的边界时，尤其是对于不规则的目标，上下文客观分数可以忽略不计，即使框在目标边界内，OR_i依然会开始减少。

如图3所示，当$R1$、$R2$、$R3$同时包围目标区域时，它们的客观得分会随着框的增大而缩小，即$OR_1 < OR_2 < OR_3$。为了完善此问题，只要候选框在目标边界内，就需要上下文一致的客观得分，即$OR_1 = OR_2 = OR_3$。因此，本文提出了目标一致性表示OCR，将R_i投影到与边界框水平垂直的方向，以获得目标一致的分数$OCRS_i$：

$OCRS_i = \{R_i^x, R_i^y\}$，其中$R_i^x \in \mathbb{R}_{M_i \times 1 \times (C+1)}$、$R_i^y \in \mathbb{R}_{1 \times N_i (C+1)}$分别表示$S_i$水平垂直方向的投影。$R_i^x = Max_x(R_i)$、$R_i^y = Max_y(R_i)$分别表示获取水平、垂直方向的最大值。

$$OR_i = (avg(R_i^x) + avg(R_i^y)) - (avg(\hat{R}_i^x) + avg(\hat{R}_i^y)) \tag{4}$$

最后，所有目标候选框$R^2 = \{R_1^2, R_2^2, \cdots, R_n^2\}$按客观分数$OR_i$进行排序，获取前300个目标建议，将其送入弱监督目标检测网络，从这些更完整的候选框中挖掘目标信息。

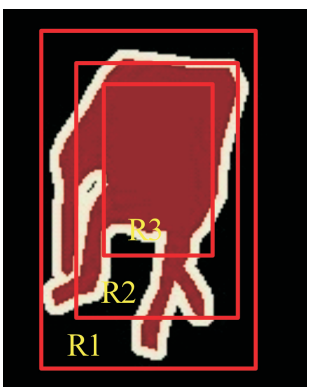

Fig. 3　Object consistency representation

图3　目标一致性表示

2.3　弱监督目标检测网络

本文使用弱监督目标检测网络（WSDDN）获取建议分数，用于后续的建议框选择和目标检测器优化。主要工作如下：给一张图像I及图像注释$y = [y_1, y_2, \cdots, y_C]$，通过区域生成方法获取目标候选框$R^2 = \{R_1^2, R_2^2, \cdots, R_n^2\}$，$y_c = 1$或0表示图像I中有或没有目标类别$c$，$C$表示类别数量。然后目标候选框经过RIO pooling得到特征建议向量，经过两个全连接层分别摄入到分类和检测两个平行分支，得到两个矩阵向量，X_{cls}、X_{det}，$X \in \mathbb{R}_{C \times N}$，然后通过Softmax层$\sigma(\cdot)$沿着纵向类别和横向目标对两个矩阵进行归一化，获取每个目标建议的分类和检测分数。

$$\sigma(X_{cls}) = \frac{e^{X_{cls}}}{\sum_{i=1}^{C} e^{X_{cls}}} \tag{5}$$

$$\sigma(X_{cls}) = \frac{e^{X_{det}}}{\sum_{i=1}^{N} e^{X_{det}}} \tag{6}$$

目标建议的实例级分数按元素乘积计算得到$X = \sigma(X_{cls}) \odot \sigma(X_{det})$，其用于建议选择和目标检测器优化。最后，计算第$c$个类的图像级预测分数。

$$p_c = \sum_{i=1}^{N} X_i, c \tag{7}$$

对于WSDDN模型训练，将式(7)中的p_c作为第c类图像的预测概率，采用随机梯度下降方法优化损失函数，分类损失函数为：

$$L_{mlc} = -\sum_{c=1}^{C} \{y_c \log p_c + (1 - y_c) \log(1 - p_c)\} \tag{8}$$

3　实验

本文主要在PASCAL VOC2007数据集进行实验，共9 963张图像包含20个目标类别。将数据集切分为包含5 009张图像的训练集（验证集）与包含4 954张图像的测试集。实验过程中只使用图像级标签进行训练。以两种指标评估性能：①测试集的平均精度（Average Precision，AP）和所有类别的平均AP（mean of AP，mAP）；②CorLoc对trainval集进行定位精度评估，即预测框与真实框的重叠率超过50%。本文采用EdgeBoxes生成初始的目标候选框，将ImageNet上预训练的VGG16网络作为主干网络获取图像信息，通过特征融合和分割目标一致的表示方法，在获取高质量候选建议框实验中，证明了其方法的有效性，同时采用WSDDN进行模型训练。

采用本文方法可以更好地处理对象边界处的情况，产生更完整的检测区域。为验证其有效性，本文将（Intersection over Union）IoU阈值（预测边界框与真实边界框的IoU）设置不同的参数，观察其性能差距。当IoU阈值从0.5提高到0.7时，结果变化如表1所示，由这些结果可以得出结论，本文方法能够选择更精确边界的高质量区域框。同时选择近年来优秀的检测方法进行比较，WSDDN[4]、OCIR[6]、PCL[7]、SLV[8]、C-MIL[10]、WCCN[16]、TS2C[17]，如表2、表3、表4所示。与当前优秀的WSOD方法比较，本文方法在VOC2007测试集的AP（%）和验证集的CorLoc（%）上展示了其良好结果。同时，如表4所示，本文方法在定位准确率上展现出非常优秀的性能，能定位更完整的目标对象。图4也展示了本文方法在图像中的定位效果，图中绿色的框为真实注释框，红色的框为本文检测的定位结果（即IOU>0.5）（彩图扫OSID码可见）。

Table 1　mAP of different thresholds on the VOC2007 test set

表1　VOC2007测试集上不同阈值的mAP

IoU阈值	0.5	0.7	ratio(%)
WCCN[16]	39.4	18.3	46.2
TS2C[17]	41.7	20.9	49.7
Our	42.2	21.8	53.1

Table 2 Category detection precision of VOC2007 test set

表 2 VOC2007 测试集类别检测精度

(%)

方法	WS-DDN[4]	OICR[6]	PCL[7]	SLV[8]	C-MIL[10]	WCCN[16]	TS2C[17]	Our
aero	46.4	58.5	57.1	65.6	62.5	49.5	59.3	63.3
bike	58.3	62.0	67.1	71.4	58.4	60.6	57.5	65.7
bird	35.5	35.1	40.9	49.0	49.5	38.6	43.7	50.6
boat	25.9	16.9	16.9	37.1	32.1	29.2	27.3	30.4
bottle	14.0	17.4	18.8	24.7	19.8	16.2	13.5	26.3
bus	66.7	63.2	65.1	69.6	70.5	70.8	63.9	66.8
car	53.0	60.8	63.7	70.3	66.1	56.9	61.7	59.3
cat	39.2	34.4	45.3	70.6	63.4	42.5	59.9	49.9
chair	8.9	8.2	17.0	30.8	20.0	10.9	24.1	29.2
cow	41.8	49.7	56.7	63.1	60.5	44.1	46.9	65.4
table	26.6	41.0	48.9	36.0	52.9	29.9	36.7	49.1
dog	39.6	31.3	33.2	61.4	53.5	42.2	45.6	64.4
horse	44.7	51.9	54.4	65.3	68.9	47.9	39.9	66.7
mbike	59.0	64.8	68.3	68.4	8.4	64.1	62.6	66.9
person	10.8	13.6	16.8	12.4	24.6	13.8	10.3	20.8
plant	17.3	23.1	25.7	29.9	51.8	23.5	23.6	18.3
sheep	40.7	41.6	45.8	52.4	58.7	45.9	41.7	48.1
sofa	49.6	48.4	52.2	60.0	66.7	54.1	52.4	50.7
train	56.9	58.9	59.1	67.6	63.5	60.8	58.7	65.6
tv	50.8	58.7	62.0	64.5	50.5	54.5	56.6	60.7

Table 3 Localization detection precision of VOC2007 validation set

表 3 VOC2007 验证集定位检测精度

(%)

方法	WS-DDN[4]	OICR[6]	PCL[7]	SLV[8]	C-MIL[10]	WCCN[16]	TS2C[17]	Our
aero	65.1	81.7	79.6	–	–	83.9	84.2	87.3
bike	58.8	80.4	85.5	–	–	72.8	74.1	85.0
bird	58.5	48.7	62.2	–	–	64.5	61.3	75.6
boat	33.1	49.5	47.9	–	–	44.1	52.1	56.8
bottle	39.8	32.8	37.0	–	–	40.1	32.1	50.1
bus	68.3	81.7	83.8	–	–	65.7	76.7	83.5
car	60.2	85.4	83.4	–	–	82.5	82.9	89.3
cat	59.6	40.1	43.0	–	–	58.9	66.6	74.9
chair	34.8	40.6	38.3	–	–	33.7	42.3	41.2
cow	64.5	79.5	80.1	–	–	72.5	70.6	88.4
table	30.5	35.7	50.6	–	–	25.6	39.5	50.3
dog	43.0	33.7	30.9	–	–	53.7	57.0	68.4
horse	56.8	60.5	57.8	–	–	67.4	61.2	78.7
mbike	82.4	88.8	90.8	–	–	77.4	88.4	90.6
person	25.5	21.8	27.0	–	–	26.8	9.3	26.1
plant	41.6	57.9	58.2	–	–	49.1	54.6	59.7
sheep	61.5	76.3	75.3	–	–	68.1	72.2	82.1
sofa	55.9	59.9	68.5	–	–	27.9	60.0	67.8
train	65.9	75.3	75.7	–	–	64.5	65.0	82.8
tv	63.7	81.4	78.9	–	–	55.7	70.3	85.7

Table 4 mAP of the VOC2007 test set and the localization precision CorLoc of validation set

表 4 VOC2007 测试集的 mAP 和验证集的定位精度 CorLoc

方法	mAP(%)	CorLoc(%)
WSDDN[4]	39.3	53.5
OICR[6]	42.0	60.2
PCL[7]	43.5	62.7
SLV[8]	49.2	69.2
C-MIL[10]	50.5	65.0
WCCN[16]	46.6	56.7
TS2C[17]	50.8	61.0
Our	51.0	71.2

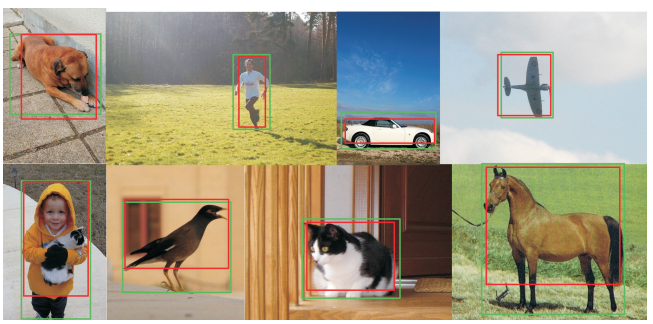

Fig. 4 Example of object detection

图 4 目标检测示例

4 结语

本文提出了一种基于特征融合和分割引导的弱监督目标检测，主要通过 CNN 网络的底层信息进行特征融合和分割引导的一致性表示方法获取高质量的候选框。同时，提出一种新的评估建议标准，即目标一致性表示，不仅能获取更完整的目标建议框，而且可以很容易地嵌入到流行的弱监督目标检测框架中。与当前优秀的方法相比，在 VOC2007 数据集上的定位准确率达到了非常优秀的性能，类别检测精度也达到了同等性能，甚至在单个类别的精度上有显著提升。尽管当前弱监督学习已经取得了不错的性能，但与完全监督学习相比还存在很大差距，但完全监督学习又依赖于精确的注释数据，因此弱监督目标检测研究在生活场景中更具有实用性。下一步研究中，将考虑轻量级的网络结构，增加模型训练速率。

参考文献：

[1] EVERINGHAM M, ESLAMI S M A, VAN GOOL L, et al. The pascal visual object classes challenge: a retrospective[J]. International Journal of Computer Vision, 2015, 111: 98-136.

[2] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[DB/OL]. https://arxiv.org/abs/1506.01497, 2015.

[3] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[4] BILEN H, VEDALDI A. Weakly supervised deep detection networks [C]//Proceedings of the IEEE Conference on Computer Vision and

Pattern Recognition,2016: 2846-2854.

[5] ZHOU X L, CHEN X J, CHEN S Y, et al. Weakly supervised learning-based object detection: a survey[J]. Computer Science, 2019,46(11):50-57.
周小龙,陈小佳,陈胜勇,等. 弱监督学习下的目标检测算法综述[J]. 计算机科学,2019,46(11):50-57.

[6] MARON O, LOZANO-PÉREZ T. A framework for multiple-instance learning[C]//Advances in Neural Information Processing Systems Conference, 1998: 570-576.

[7] TANG P, WANG X, BAI X, et al. Multiple instance detection network with online instance classifier refinement[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017: 2843-2851.

[8] TANG P, WANG X, BAI S, et al. PCL: proposal cluster learning for weakly supervised object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(1): 176-191.

[9] WAN F, WEI P, JIAO J, et al. Min-entropy latent model for weakly supervised object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018: 1297-1306.

[10] WAN F, LIU C, KE W, et al. C-MIL: Continuation multiple instance learning for weakly supervised object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019: 2199-2208.

[11] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104: 154-171.

[12] ZITNICK C L, DOLLÁR P. Edge boxes: locating object proposals from edges[C]//European Conference on Computer Vision, 2014: 391-405.

[13] KUO W, HARIHARAN B, MALIK J. Deepbox: learning objectness with convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision,2015: 2479-2487.

[14] PINHEIRO P O, LIN T Y, COLLOBERT R, et al. Learning to refine object segments[C]//European Conference on Computer Vision, 2016: 75-91.

[15] ZHOU Z B. Research on ECG signal recognition method based on feature fusion CNN model[J]. Software Guide, 2020,19(11):47-49.
周志波. 基于特征融合 CNN 模型的 EGC 信号识别方法研究[J]. 软件导刊,2020,19(11):47-49.

[16] LU X, LIU Z. A review of image semantic segmentation based on deep learning[J]. Software Guide, 2021,20(1):243-244.
卢旭,刘钊. 基于深度学习的图像语义分割技术综述[J]. 软件导刊,2021,20(1):243-244.

[17] DIBA A, SHARMA V, PAZANDEH A, et al. Weakly supervised cascaded convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 914-922.

[18] WEI Y, SHEN Z, CHENG B, et al. Ts2c: tight box mining with surrounding segmentation context for weakly supervised object detection[C]//Proceedings of the European Conference on Computer Vision (ECCV),2018: 434-450.

[19] GAO Y, LIU B, GUO N, et al. C-MIDN: Coupled multiple instance detection network with segmentation guidance for weakly supervised object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision,2019: 9834-9843.

[20] LI X, KAN M, SHAN S, et al. Weakly supervised object detection with segmentation collaboration[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision,2019: 9735-9744.

[21] AHN J, KWAK S. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018: 4981-4990.

[22] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [DB/OL]. https://arxiv.org/abs/1409.1556,2014.

(责任编辑:孙　娟)