

## 基于改进的R-FCN带纹理透明塑料裂痕检测

关日钊, 陈新度, 吴磊, 徐焯基

广东工业大学 机电工程学院, 广州 510006

**摘要:**为了解决利用传统的机器学习方法来检测带纹理透明塑料裂痕的检测精度和识别率不高的问题,提出一种改进的基于区域的全卷积网络(Region-based Fully Convolutional Networks, R-FCN)检测方法,通过对R-FCN中的残差网络(Residual Network, ResNet)特征提取网络进行混合尺度感受野融合处理,弥补了原网络对微小裂痕敏感度不高的缺点。实验表明,改进后的R-FCN检测方法的裂痕检测精度比基于传统机器学习支持向量机(Support Vector Machine, SVM)检测方法的裂痕检测准确率高20%左右,比未改进的R-FCN检测方法的检测准确率高8%,证明了该方法的有效性。

**关键词:**裂痕检测;支持向量机(SVM);基于区域的全卷积网络(R-FCN);残差网络(ResNet);感受野

**文献标志码:**A **中图分类号:**TP24 **doi:**10.3778/j.issn.1002-8331.1712-0040

关日钊, 陈新度, 吴磊, 等. 基于改进的R-FCN带纹理透明塑料裂痕检测. 计算机工程与应用, 2019, 55(6): 168-172.

GUAN Rizhao, CHEN Xindu, WU Lei, et al. Textured transparent plastics crack detection based on improved R-FCN. Computer Engineering and Applications, 2019, 55(6): 168-172.

### Textured Transparent Plastics Crack Detection Based on Improved R-FCN

GUAN Rizhao, CHEN Xindu, WU Lei, XU Zhuoji

College of Mechanical and Electrical Engineering, Guangdong University of Technology, Guangzhou 510006, China

**Abstract:** To solve the problem of the detection accuracy and recognition using traditional machine learning method to detect texture with transparent plastic crack rate is not high, an improved detection method based on Region-based Fully Convolutional Networks(R-FCN) is proposed. It makes up for the shortcomings of the original network's low sensitivity to tiny cracks by using mixed-scale receptive field fusion procession in Residual Network(ResNet). Experimental results show that the crack detection accuracy based on improved R-FCN is about 20% higher than that based on Support Vector Machine(SVM), and is about 8% higher than that based on R-FCN without improvement. The validity of the method is proved.

**Key words:** crack detection; Support Vector Machine(SVM); Region-based Fully Convolutional Networks(R-FCN); Residual Network(ResNet); receptive field

### 1 引言

随着图像处理技术的不断发展,基于机器视觉的物体表面缺陷检测方法成为了快速而且有效的检测方式之一,广泛应用于印刷、电子、材料加工等行业,逐渐替代人工检测。一般的机器视觉检测缺陷的过程由以下几个步骤组成:图像获取、图像增强、图像分割、缺陷特征提取和缺陷特征识别与判断。其中缺陷特征提取和缺陷特征识别与判断这两个步骤成为了整个检测过程

的重点和难点。

针对上述两个难点,国内外学者对此进行了大量的研究。从研究方法来看主要分为三种:基于传统图像处理方法、基于机器学习方法和基于深度学习方法。国内周江<sup>[1]</sup>等通过二值化和闭操作对图像进行预处理,然后将目标区域面积和模版区域面积进行了差分运算,选出差值大于规定值的区域作为缺陷区域,实现了磁钢片裂纹缺陷检测;周显恩<sup>[2]</sup>等针对啤酒瓶口图像中缺陷灰度

**基金项目:**广东省省级科技计划项目(No.2015B010919001, No.2016B090912002, No.2013JDXM30)。

**作者简介:**关日钊(1992—),男,硕士生,研究领域为机器视觉,深度学习, E-mail: trustguan@foxmail.com;陈新度(1967—),男,博士,教授,研究领域为基于机器人的智能制造单元,机器视觉,深度学习;吴磊(1978—),男,讲师,研究领域为计算机集成与制造,机器视觉;徐焯基(1993—),男,硕士生,研究领域为机器视觉,深度学习。

**收稿日期:**2017-12-04 **修回日期:**2018-01-26 **文章编号:**1002-8331(2019)06-0168-05

**CNKI网络出版:**2018-04-28, <http://kns.cnki.net/kcms/detail/11.2127.TP.20180428.1024.003.html>

变化范围大的问题,提出了残差分析动态阈值分割与全局阈值分割结合的方法,对啤酒瓶口缺陷进行检测。传统的机器学习方法是通过构造大量的目标特征,然后对特征进行训练和分类,得出一个训练后的模型。最后通过这个模型可以对待检测的样品进行识别与分类。国内胡方尚<sup>[3]</sup>等针对印刷品缺陷检测的问题,提出了一种基于改进多类支持向量机(SVM)的印刷缺陷检测方法,与传统的一对一型向量机方法相比,其检测速度更快和识别准确率更高。国外Krummenacher<sup>[4]</sup>等利用SVM检测方法对火车车轮进行缺陷分类检测,准确率达到88%以上。随着GPU技术的发展,同属人工智能范畴的深度学习也在近几年内得到迅速的发展,特别在图像领域方面。在2012年,Hinton课题组通过构建CNN网络AlexNet<sup>[5]</sup>在ImageNet图像识别比赛中一举得到冠军,自那年后,深度学习正式进入一个爆发式的发展阶段。与传统的机器学习不同的是,深度学习不需要人为构造目标特征,只需要通过卷积神经网络对图像进行特征提取,这种通过自动学习提取出来的特征更加具有代表性。在缺陷检测方面,王宪保<sup>[6]</sup>等通过建立深度置信网络对太阳能电池片进行准确快速的缺陷检测;余永维<sup>[7]</sup>等提出了一种基于深度学习特征匹配的铸件缺陷三维定位方法,定位误差小于5.52%。

本文研究的对象是一种带纹理的透明塑料制品,其裂痕检测难点主要有两个,一个是塑料表面具有竖直方向和与水平方向夹角10°条纹,容易与裂痕混淆;另外一个是在相同强度背光光照下,每一个塑料件由于材料不均匀导致透光程度不一样,利用传统的模版差分方法显然存在问题。针对上述两个难点,提出一种改进的R-FCN方法对这种塑料制品进行检测与定位,通过对R-FCN中的ResNet特征提取网络进行混合尺度感受野融合处理来提高网络对微小裂痕的敏感度。实验结果表明,与传统的SVM方法和未改进的R-FCN方法相比,经过改进后的R-FCN方法在准确率上比前两者高。

2 基于区域的全卷积网络(R-FCN)

目标检测技术就是将原始图像中所感兴趣的区域(Region Of Interest, ROI)从原始图像的背景中加以区分出来,并准确定位ROI区域的位置<sup>[8]</sup>。在2012年的ImageNet竞赛中,Alex等人组成的团队采用的卷积神经网络CNN在图像分类和目标定位中分别达到84.7%和39.1%的正确率。相对传统的机器学习方法无论在速度上还是正确率上都有明显的提高。随后Girshick等人在此基础上提出了基于CNN特征图的区域定位网络结果模型R-CNN<sup>[9]</sup>以及R-CNN网络结构模型的改进模型Fast R-CNN<sup>[10]</sup>和Faster R-CNN<sup>[11]</sup>,进一步地提高了目标检测定位的正确度和速度。

R-FCN是Dai<sup>[12]</sup>等人在Faster R-CNN模型基础上

提出的。整个网络不存在全连接层,都是可学习的卷积层,并且几乎一张图像上的所有计算都是共享的,相比Faster R-CNN来说大大地节省了计算的开销<sup>[13]</sup>。其网络流程如图1所示,简单概括为以下几个步骤:

- (1)利用RestNet进行特征提取,获得一系列特征图。
- (2)与Faster R-CNN一样,利用区域提名网络(Region Proposal Network, RPN)生成一系列感兴趣区域(ROI)。
- (3)在RPN网络最后添加一层特殊卷积层对每一个ROI构建一组位置敏感得分图(position-sensitive score maps),每个得分图保存目标的空间位置信息。
- (4)对每一组位置敏感得分图进行均值池化处理,经过投票后得到一组向量,然后送入softmax层计算并输出每一类的概率。

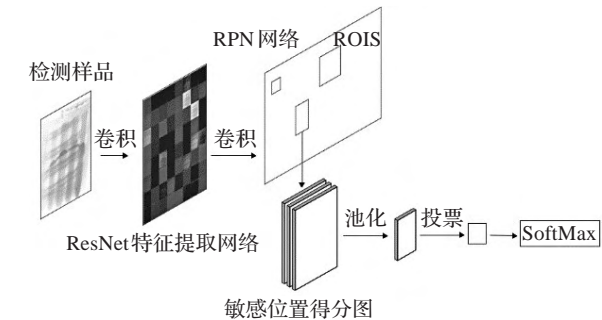


图1 R-FCN网络图

从上述过程中可以看出,在网络的池化层后面并没有再连接其他卷积层或全连接层,这样整个网络不仅可以进行端对端(end to end)的训练,而且所有层的计算都是在整个图像上共享,这也成为了在速度上比Faster R-CNN快的一个重要原因。

此外,在过程中提到的位置敏感得分图是整个网络的核心地方。如图2所示,对一个大小为的ROI,一个箱子(bin)的大小为 $\frac{w}{k} \times \frac{h}{k}$ , $k$ 是预定义的超参数,在论文的实现中,取3。

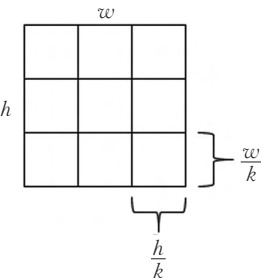


图2 ROI示意图

在最后的卷积层中为每类产生 $k^2$ 个得分图。对第 $(i,j)$ 个箱子 $(0 \leq i,j \leq k-1)$ ,可以根据式(1)对一个敏感ROI进行池化操作。

$$r_c(i,j|\Theta) = \frac{1}{n} \sum_{(x,y) \in bin(i,j)} z_{i,j,c}(x+x_0,y+y_0|\Theta) \tag{1}$$

其中  $\Theta$  为网络参数,  $x_0$  和  $y_0$  为 ROI 左上角的坐标,  $bin(i, j)$  为 ROI 第  $(i, j)$  个空间位置,  $(x, y)$  为该空间位置的点的坐标,  $n$  为该空间位置内的点数<sup>[14]</sup>。

### 3 特征提取网络

#### 3.1 残差网络(ResNet)

ResNet是何凯明<sup>[15]</sup>团队在2015年提出的,并在当年的ImageNet分类比赛中获得了冠军。一般情况下,深度学习网络的深度对最后的分类和识别的效果存在着较大的影响,网络深度越大,分类和识别的效果越好。但实际上当仅仅利用简单的网络堆叠方法去增加网络深度时,梯度消失的现象会越来越明显,导致网络最后出现“退化”问题,所谓“退化”是指当网络模型的层数加深时,错误率却变高了<sup>[16]</sup>。而这个“退化”问题的产生原因可以归结于优化问题,即当网络模型变复杂后,随机梯度下降法(SGD)的优化变得更加困难了,导致最终训练出来的模型达不到好的学习效果。

ResNet的出现正是要解决当网络深度加深时候出现的退化问题。提出了一个残差结构,如图3所示。

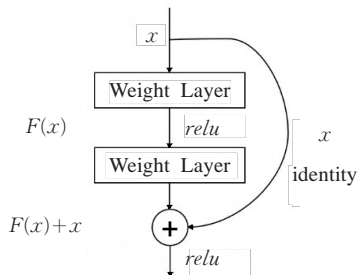


图3 残差结构图

从图3中可以看出,网络添加了一个恒等映射(identity mapping),将原始所需要学习的函数  $H(x)$  转换为  $F(x)+x$ ,这两种表达的效果相同,但是优化的难度并不相同,  $F(x)$  的优化会比  $H(x)$  简单,这是因为在图像处理中的残差向量编码可以通过一个转换,将一个问题分解成多个尺度的残差问题,能够起到优化训练的效果。此外,在通过第二个激活函数输出之前在主输出中叠加了原始输入,即增加了一个捷径连接(short-cut connection)。捷径连接使数据流通过跨层的方式来进行非线性叠加运算并且不会给网络增加额外的参数和计算量,同时可提高网络训练的速度以及训练效果<sup>[17-18]</sup>。当模型的层数加深时,这个简单的结构能够很好地解决退化问题。

#### 3.2 感受野

感受野是卷积神经网络中最重要的概念之一,它表示卷积神经网络每一层输出的特征图上的像素点在原始图像上映射的区域大小,如图4所示。

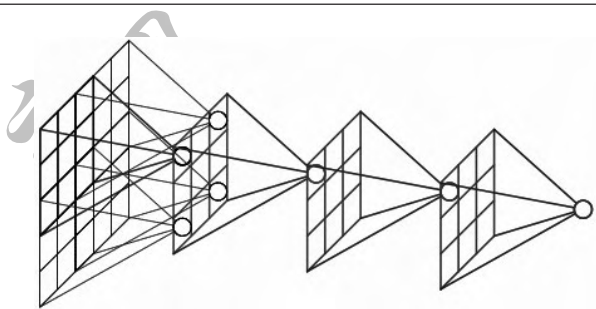


图4 感受野示意图

与全连接网络不同的是,在卷积神经网络中每一层的每一个单元格的值只依赖于输入的一个区域<sup>[19]</sup>。简单来说,感受野的大小决定了网络最终输出的特征向量所包含原图特征信息量的大小。而深层卷积层的感受野大小和它之前所有层的滤波器大小和步长有关系。通常在计算最深层感受野的时候采用“从上到下”的方式计算,即先计算最深层在上一层上的感受野,然后逐渐传递到第一层,具体如式(2)所示:

$$RF=(RF'-1)\times stride+fsize \tag{2}$$

其中  $RF$  为最深层感受野,  $RF'$  为当前最深层相对于上一层的感受野,  $stride$  为当前层的步长,  $fsize$  为当前层的卷积核大小。表1为ResNet-50前五层结构感受野的计算情况。

表1 ResNet-50前五层感受野数据

layer	fsize	stride	RF
Conv_1	7	2	7
Pool_1	3	2	11
Conv_2	1	1	11
Conv_3	3	1	19
Conv_4	1	1	19

#### 3.3 混合尺度感受野融合

根据前面讨论的ResNet原理可以构造出最后的实际网络应该由图5(a)中的残差模块叠加组合而成。但是为了节省计算的成本,对此残差块做了优化,如图5(b)所示,即将两个  $3\times3$  的卷积层替换为  $1\times1$ 、 $3\times3$ 、 $1\times1$  三个卷积层,这样一来,既保持了精度又减少了计算量。

ResNet-50是由图5(b)中的残差块构成的50层网络,共有16个残差块。在3.2节中提到神经网络的感受

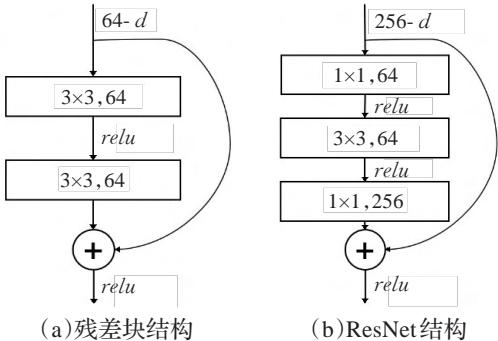


图5 残差块结构图



野决定了每一层输出的特征图在原始图像上映射的范围,而决定感受野大小与卷积核的大小和卷积步长有关。一般情况下是通过改变卷积核的大小来改变感受野的大小,卷积核越大,最终获取的特征信息量就越多。本文的实验对象是一种带纹理的透明塑料件,需要对其进行裂痕检测和定位,而由于有些裂痕比较小,如果直接用ResNet-50去训练和检测的话,会出现漏检的情况,为了解决这种问题,本文提出一种混合尺度感受野模型,在ResNet-50原生网络的基础下,把网络中一部分带有3×3大小的卷积核替换为5×5大小的卷积核。如图6(a)所示。

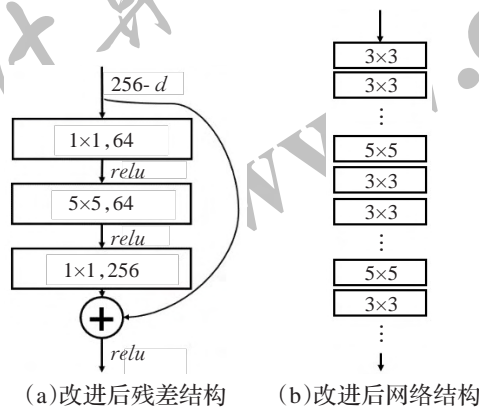


图6 混合尺度感受野模型

ResNet-50网络最终输出的特征图大小可以根据式(3)<sup>[20]</sup>计算为7×7,其中F表示卷积核大小,P表示边界填充间隔,S表示步长。而根据3.2节计算感受野的方法来计算,可以算出网络修改前后的最深感受野分别为164和212,这就意味着在这7×7单元大小的特征图中,修改后的每一个单位在原始图像上映射的区域比修改前的要大,即包含原图的信息要更多,提高了网络对小裂痕的敏感度。

$$\begin{cases} W_{dst} = (W_{src} - F + 2P) / S + 1 \\ H_{dst} = (H_{src} - F + 2P) / S + 1 \end{cases} \quad (3)$$

此外,考虑到增加卷积核的大小会增加训练参数,所以两种卷积核的残差块的数目比不能设为1:1,本文实验将它们数目比设为3:1,即具有3×3和5×5大小卷积核的残差块个分别为12和4,并且后者要等距穿插在前者之中,如图6(b)所示,具体的训练效果和检测效果将在第4章给出。

4 实验步骤与结果分析

4.1 实验步骤

本实验的实验环境是Caffe+GTX1070,利用ImageNet预训练模型进行微调(fine tune),微调是指利用已经训练好的模型的参数进行网络参数的初始化,这样可以节省训练时间和提高训练的精度。为了防止梯度

过大,网络初始学习率为0.000 1,权重衰减率为0.000 5,动量设置为0.9。此外训练样品共850张图片,训练迭代次数为3万次,类别数目为2,分别是裂痕和背景。总体流程如图7所示。

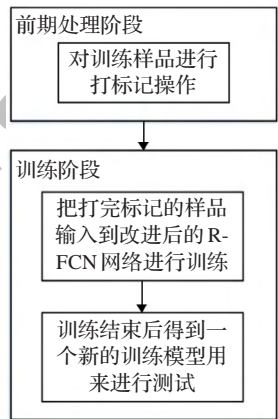


图7 训练流程图

在训练阶段中,每个ROI的损失函数为交叉熵损失与边界框回归损失的总和,可表示为式(4):

$$L(s, t_x, y, w, h) = L_{cls}(s_c^*) + \lambda [c^* > 0] L_{reg}(t, t^*) \quad (4)$$

其中 $c^* > 0$ 说明ROI的真实标签为裂痕, $t_i$ 与 $t_i^*$ 分别表示预测框与真实框的4个参数化信息。 $L_{reg}$ 与Faster R-CNN的边界框损失回归相同,如式(5)所示,其中R函数为Fast R-CNN中定义的Smooth L1,如式(6)所示。

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) \quad (5)$$

$$R(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & |x| \geq 1 \end{cases} \quad (6)$$

图8为训练阶段中总的损失值随迭代次数的变化曲线图,从图中可以看出损失函数在迭代到15 000次的时候已经趋于收敛。

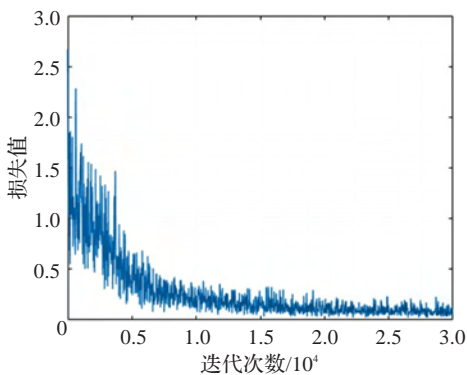


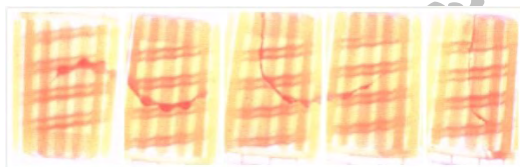
图8 训练损失值图

4.2 结果分析

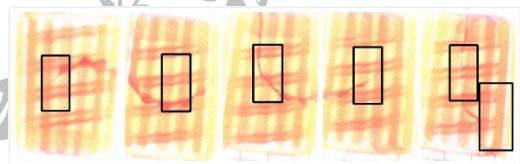
对比实验中采用正负样品图片各100张,对比对象有三个,分别是基于HOG+SVM机器学习模型、基于R-FCN网络模型和样品经过改进后的R-FCN网络模型。具体实验数据表2所示,图9为部分检测实际效果,黑色

表2 三种方法实验数据

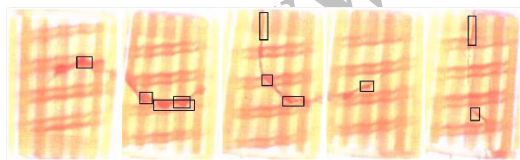
方法	准确率	召回率	单张检测时间/ms
SVM+HOG	0.658	0.79	942
R-FCN	0.833	0.93	86
本文	0.914	0.97	98



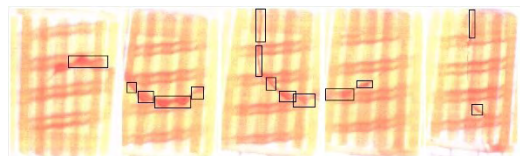
(a)待测样品



(b)SVM+HOG 检测结果



(c)R-FCN 检测效果



(d)改进R-FCN 检测效果

图9 实际测试效果

方框区域为裂痕区域。

从表2中可以看出,传统的SVM机器学习方法在三种方法中准确率最低而且检测时间最长;而基于R-FCN方法中,经过改进后的方法比原方法在准确率和召回率上分别高8%和4%。由于改进后的网络增加了 $5\times 5$ 的卷积核,导致计算量提高了一点,所以在时间上,改进后的方法在时间上会长一点,但并不影响在线检测的要求。

根据图9中的检测结果来看,在检测到裂痕的情况下,基于SVM检测方法框选的区域过大,而基于R-FCN方法在定位上更加的准确,框选的部分几乎为裂痕的最小包围盒。此外,从图9(c)和图9(d)可以看出,经过改进后的网络检测到的裂痕区域会比未经过改进的网络要多,这主要得益于改进后网络的感受野增大了,提高了网络对微小裂痕的敏感度。

## 5 结束语

本文研究的内容为透明带纹理的塑料制品的裂痕检测问题。提出一种改进的R-FCN方法对这种塑料制

品进行检测与定位,通过对R-FCN中的ResNet特征提取网络进行混合尺度感受野融合处理来提高网络对微小裂痕的敏感度。从实验结果来看,基于改进后的R-FCN的网络性能比原生网络性能要好,证明了本文提出的方法具有有效性。此外,由于神经网络的训练和测试对计算机硬件的要求较高,导致硬件成本提高,从而成为了把深度学习技术推广到工业检测领域的一个限制性因素,因此如何通过优化网络,在保证检测率的情况下降低硬件成本的问题将成为本文下一步的研究方向。

## 参考文献:

- [1] 周江,任锟,帅英琦,等.基于机器视觉的磁钢片缺陷检测研究[J].机电工程,2014,31(12):1541-1546.
- [2] 周显恩,王耀南,朱青,等.基于机器视觉的瓶口缺陷检测方法研究[J].电子测量与仪器学报,2016,30(5):702-713.
- [3] 胡方尚,郭慧.基于改进多类支持向量机的印刷缺陷检测[J].华东理工大学学报(自然科学版),2017,43(1):143-148.
- [4] Krummenacher G, Ong C S, Koller S, et al. Wheel defect detection with machine learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(4): 1176-1187.
- [5] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems, 2012: 1097-1105.
- [6] 王宪保,李洁,姚明海,等.基于深度学习的太阳能电池片表面缺陷检测方法[J].模式识别与人工智能,2014,27(6):517-523.
- [7] 余永维,杜柳青,曾翠兰,等.基于深度学习特征匹配的铸件微小缺陷自动定位方法[J].仪器仪表学报,2016,37(6):1364-1370.
- [8] 颜伟鑫.深度学习及其在工件缺陷自动检测中的应用研究[D].广州:华南理工大学,2016.
- [9] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [10] Girshick R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [11] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems, 2015: 91-99.
- [12] Dai J, Li Y, He K, et al. R-FCN: object detection via region-based fully convolutional networks[C]//Advances in Neural Information Processing Systems, 2016: 379-387.
- [13] Zhang H, Kyaw Z, Yu J, et al. PPR-FCN: weakly supervised visual relation detection via parallel pairwise R-FCN[J]. arXiv preprint arXiv:1708.01956, 2017.

(下转第264页)