DOI: 10.11992/tis.201706040
网络出版地址: http://kns.cnki.net/kcms/detail/23.1538.TP.20180411.0849.002.html

自动驾驶场景下小且密集的交通标志检测

葛园园[1], 许有疆[1], 赵帅[2], 韩亚洪[1]
（1. 天津大学 计算机科学与技术学院, 天津 300350; 2. 中国汽车技术研究中心 数据资源中心, 天津 300300）

摘　要: 在自动驾驶场景中, 交通标志的检测和识别对行车周围环境的理解至关重要。行车过程中拍摄的图片中存在许多较小的交通标志, 它们很难被现有的物体检测方法检测到。为了能够精确地检测到这部分小的交通标志, 我们提出了用浅层 VGG16 网络作为物体检测框架 R-FCN 的主体网络, 并改进 VGG16 网络, 主要有两个改进点: 1) 减小特征图缩放倍数, 去掉 VGG16 网络卷积 conv4_3 后面的特征图, 使用 RPN 网络在浅层卷积 conv4_3 上提取候选框; 2) 特征拼层, 将尺度相同的卷积 conv4_1、conv4_2、conv4_3 层的特征拼接起来形成组合特征 (aggregated feature)。改进后的物体检测框架能够检测到更多的小物体, 在驭势科技提供的交通标志数据集上取得了很好的性能, 检测的准确率 mAP 达到了 65%。

关键词: 交通标志; 目标检测; 深度学习; 组合特征; 卷积神经网络; 特征图; 候选框; 自动驾驶
中图分类号: TP183　　**文献标志码:** A　　**文章编号:** 1673−4785(2018)03−0366−07

中文引用格式: 葛园园, 许有疆, 赵帅, 等. 自动驾驶场景下小且密集的交通标志检测[J]. 智能系统学报, 2018, 13(3): 366–372.
英文引用格式: GE Yuanyuan, XU Youjiang, ZHAO Shuai, et al. Detection of small and dense traffic signs in self-driving scenarios[J]. CAAI transactions on intelligent systems, 2018, 13(3): 366–372.

Detection of small and dense traffic signs in self-driving scenarios

GE Yuanyuan[1], XU Youjiang[1], ZHAO Shuai[2], HAN Yahong[1]
(1. School of Computer Science and Technology, Tianjin University, Tianjin 300350, China; 2. Data Resource Center, China Automotive Technology and Research Center, Tianjin 300300, China)

Abstract: In self-driving scenarios, the detection and recognition of traffic signs is critical to understanding the driving environment. The plethora of small traffic signs are hard to detect by the existing object detection technology. To detect these small traffic signs accurately, we propose the use of the shallow network VGG16 as the R-FCN 's backbone and the modification of the VGG16 network. There are mainly two improvements in the VGG16 network. First, we reduce the multiple zooming of feature maps, remove the feature maps behind the VGG16 network convolution conv4_3, and use the RPN network to extract the region proposal in the shallow convolution conv4_3 layer. We then concatenate the feature maps. The features of the layers of the convolutions conv4_1, conv4_2, and conv4_3 are adjoined to form an aggregated feature. The improved object detection framework can detect more small objects. We use a dataset of traffic signs to test the performance and mAP accuracy.

Keywords: traffic sign; object detection; deep learning; aggregate feature; CNN; feature map; region proposal; self-driving

近年来, 随着深度学习技术[1-8]的发展, 自动驾驶引起了人们的广泛关注。在自动驾驶场景中, 交通标志的检测和识别起着非常重要的作用。精确的检测对后续的识别、辅助定位和导航起着决定性的作用。在真实的拍摄场景中, 交通标志种类众多, 大小不一, 存在着颜色差异, 且受到天气、光照、拍摄角度等因素的影响。这些复杂的因素使得交通标志的检测变得非常困难, 尤其是图片中存在大量小且密集的交通标志。为了能够更好地解决小物体的

收稿日期: 2017−06−10.　**网络出版日期:** 2018−04−11.
基金项目: 国家自然科学基金项目 (61472276).
通信作者: 韩亚洪. E-mail: yahong@tju.edu.cn.

检测, 本文首先对已有的基于深度学习的目标检测做了研究和总结, 然后提出自己改进的目标检测框架。

现有的基于深度学习的目标检测算法大致可以分为两大类: 一类是基于候选框提取的目标检测算法, 另一类是基于回归方法的目标检测算法。下面介绍这两大类目标检测算法的代表性检测框架。

R-CNN 是 Ross Girshick 等[9]最早提出的用卷积神经网络做目标检测的框架, 它是基于候选框提取的目标检测算法。首先通过 selective search 算法[10]提取候选框, 然后用这些候选框微调卷积神经网络训练 SVM(支持向量机) 分类器, 进行边框回归等。R-CNN 在 VOC2012 数据集上 mAP[11]达到了 53.3%, 和之前该数据集上最好的检测结果相比提高了 30%。但 R-CNN 也存在着很大的缺陷, 浪费磁盘空间, 训练时间长, 检测速度慢。

SPP-Net[12]在 R-CNN 的基础上作了改进, 提出了共享卷积层策略, 采用空间金字塔下采样 (spatial pyramid pooling), 将每个候选框映射后的特征归一化到同样尺度, 输入到后面的全连接层中。SPP-Net 的检测速度比 R-CNN 快了 24 ~ 102 倍, 检测精度也得到了一定的提高。但 SPP-Net 的训练过程包括多个阶段, 比较慢。

Fast R-CNN[13]将 R-CNN 与 SPP-Net 的空间金字塔采样融合到一起, 提出了候选框下采样层 (ROI pooling layer), 直接将映射后的候选框下采样到 7×7 大小的特征图输入到后面的全连接层中。Fast R-CNN 还提出了多任务损失函数, 直接对候选框进行分类和边框回归, 检测速度得到了很大提升, 使得候选框提取成为限制目标检测速度的一个计算瓶颈。Faster R-CNN[14]为了解决候选框提取问题, 提出了 RPN(region proposal network) 网络来提取候选框, 将 RPN 网络和 Fast R-CNN 网络结合到了一起。Faster R-CNN 的检测精度和速度得到了很大的提高。

R-FCN[15]在 Faster R-CNN 的基础上作了进一步的改进, 提出位置敏感权重图 (position-sensitive score maps) 来解决图像分类时的旋转不变性和物体检测上位置的旋转可变性之间的矛盾。R-FCN 的位置敏感权重图策略使得网络变成了一个全卷积网络, 不再有全连接层, 进一步加快了目标检测的速度和准确率。

以上目标检测算法都是深度学习方面基于候选框的目标检测框架。另一类基于回归方法的目标检测框架的典型代表是 YOLO[16]和 SSD[17], 它们不必先提取候选框, 再进行候选框的分类和位置调整, 而是直接对图像进行划分网格, 在每个网格对应位置回归出目标位置和类别信息, 它们网络训练的整个过程都是端到端的, 检测速度非常快, 可以达到实时要求。

对于自动驾驶场景下小且密集的交通标志检测来说, 上述两大类物体检测框架都存在着一定的缺陷: 基于候选框的检测框架相比基于回归方法的检测框架来说, 检测精度会高很多, 但速度比较慢; 基于回归方法的检测框架虽然检测速度比较快, 但由于是对图像进行暴力网格划分, 对于物体的检测精度比较差, 尤其是对于小且密集物体的检测, 效果非常差。因此, 为了保证自动驾驶场景中小且密集的交通标志的检测精度, 本文提出了对基于候选框目标检测框架中速度最快的 R-FCN[15]框架进行改进, 进一步保证了小且密集的交通标志的检测精度。

1　检测框架

R-FCN[15]目标检测框架分别以 ResNet-50、ResNet-101[18]作为提取特征的主体网络, 在 PASCAL VOC[11]数据集上取得了不错的检测效果, 但是交通标志检测数据集中存在很多较小的物体, 卷积网络层数越深, 最后一层卷积特征图上的特征点对应于原图的感受野越大, 对小物体的位置定位比较困难, 检测效果比较差。因此, 本文采用层数不是很深的 VGG16[19]网络作为 R-FCN 的主体网络, 并在此基础上针对小物体的检测作进一步的改进。本文提出的物体检测框架主要是对 R-FCN 检测框架的主体网络进行改进, 网络的训练过程和 R-FCN 中的训练过程保持一致。图 1 是本文提出的基于改进版 VGG16 的 R-FCN 小物体检测框架。该检测框架主要是对 VGG16 作了两个改进, 第一个改进是将 VGG16 的卷积层 conv5_1、conv5_2、conv5_3 去掉, 同时去掉 conv4_3 后面的 pool4 层; 第二个改进是特征拼层, 将 conv4_1、conv4_2、conv4_3 通过 L2 normalization[20] 标准化, 然后再拼接到一起, 称之为组合特征, 然后输入到后续网络中。RPN 网络在标准化后的卷积 conv4_3 上提取候选框, 将候选框和网络最后一组卷积层特征输入到 PSROIPooling 层来做后续的物体分类和边框回归。

2　改进方法

本文提出的物体检测框架主要针对交通标志数据集中的小物体的检测。网络层数越深, 最后一层卷积层的感受野越大, 会破坏小物体在特征图上的位置信息。R-FCN[15]物体检测框架使用 ResNet-50、ResNet-101 作为主体网络时, 可以很好地检测到较大的物体, 但是对于存在大量密集的小物体的数据

集来说，检测结果没有浅层的 VGG16 效果好。因此，本文采用 VGG16[19]作为 R-FCN 的主体网络，VGG16 从 conv1_1 到 conv5_3 总共有 13 个卷积层，每一层的卷积核大小、移动步长都是相同的，这 13 个卷积层可以分为 5 组，第一、二组分别有 2 个卷积层，第三、四、五组分别有 3 个卷积层，每一组的特征图大小相同，前 4 组特征每组后面都接了 1 个步长为 2 的下采样层。本文提出的两个关键改进点都是对 VGG16 网络进行改进，下面对这两个关键改进点作具体说明。

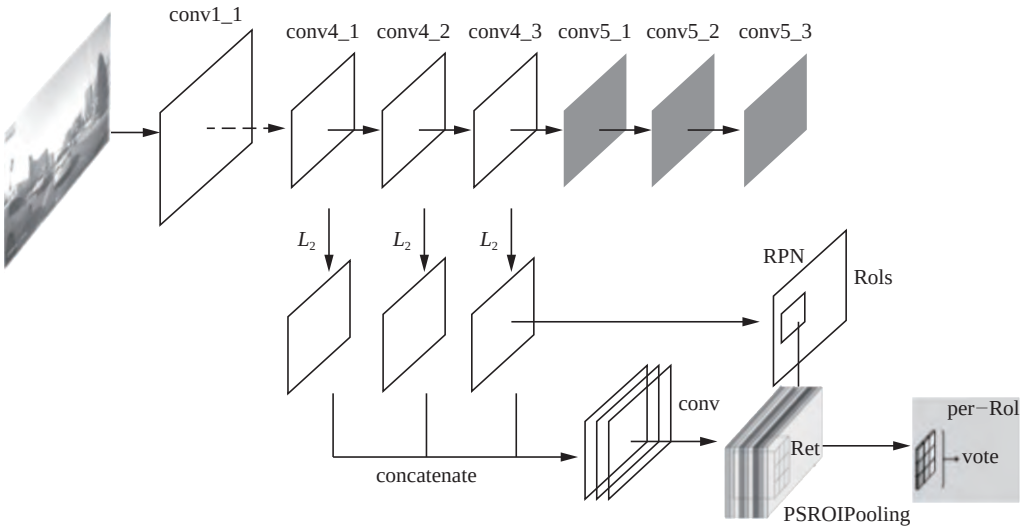


图 1　基于改进版 VGG16 的 R-FCN 小物体检测框架

Fig. 1　Small-size object detection architecture based on modified VGG16's R-FCN

2.1　减小特征图缩放倍数

VGG16 作为 R-FCN 的主体网络，在卷积 conv5_3 上通过 RPN 网络提取候选框，从 conv1_1 到 conv5_3，中间经过了 4 次步长为 2 的下采样层，原始图像到达 conv5_3 层边长缩小了 16 倍，conv5_3 特征图上的一个特征点相当于对应原图的 16×16 个像素点。当 RPN 网络从 conv5_3 上产生候选框时，有边长小于 16 个像素的候选框被过滤掉了，因为边长小于 16 个像素的候选框无法映射到从 conv5_3 上卷积得到的后续特征图上，这就造成许多边长小于 16 个像素的小物体无法被检测到，或者检测不准确。针对这一问题，本文提出减小特征图缩放倍数的策略，也就是说，去掉卷积 conv5_1、conv5_2、conv5_3 和 pool4 层，在 conv4_3 层上通过 RPN 产生候选框，再映射到从 conv4_3 上得到的后续特征图上。这样，原图到 conv4_3 边长缩小了 8 倍，边长在 8 ~ 16 范围内的小物体可以得到更好的检测效果。

2.2　特征拼层

RPN 网络产生的边长较小的候选框和最后一层卷积特征作映射后得到的特征尺度特别小，只用 conv4_3 层输入到后续的网络中进行分类和边框回归特征不充足，对于小物体位置定位来说检测效果不是很好，还有待提升。因此，为了能够丰富小物体的特征信息，使得对于小物体的位置定位更加精确，提出了特征拼层的策略。首先对卷积 conv4_1、conv4_2、conv4_3 进行 L2 normalization[12]。假设用向量$\boldsymbol{X}=(x_1,x_2,\cdots,x_d)$来表示一个输入特征层的所有元素，对该特征图进行如下标准化：

$$\widehat{\boldsymbol{X}}=\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|_2} \tag{1}$$

$$\|\boldsymbol{X}\|_2=\left(\sum_{i=1}^{d}|x_i|^2\right)^{1/2} \tag{2}$$

式中$\widehat{\boldsymbol{X}}$表示标准化后的特征图向量。标准化后的特征向量元素值比较小，使得网络训练变得比较困难，为了更好地训练网络，通常会对标准化后的元素通过一个比例因子γ_i进行缩放，假设变化后的元素为y_i，则

$$y_i=\gamma_i\widehat{x_i} \tag{3}$$

最后得到变化后的特征图向量$\boldsymbol{Y}=(y_1,y_2,\cdots,y_d)$。在网络训练过程中，假设 l 是需要优化的损失函数，则关于缩放因子γ_i和输入特征图的更新规则如下：

$$\frac{\partial l}{\partial\widehat{\boldsymbol{X}}}=\frac{\partial l}{\partial\boldsymbol{\gamma}}\cdot\gamma \tag{4}$$

$$\frac{\partial l}{\partial\boldsymbol{X}}=\frac{\partial l}{\partial\widehat{\boldsymbol{X}}}\left(\frac{\boldsymbol{I}}{\|\boldsymbol{X}\|_2}-\frac{\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}}{\|\boldsymbol{X}\|_2^3}\right) \tag{5}$$

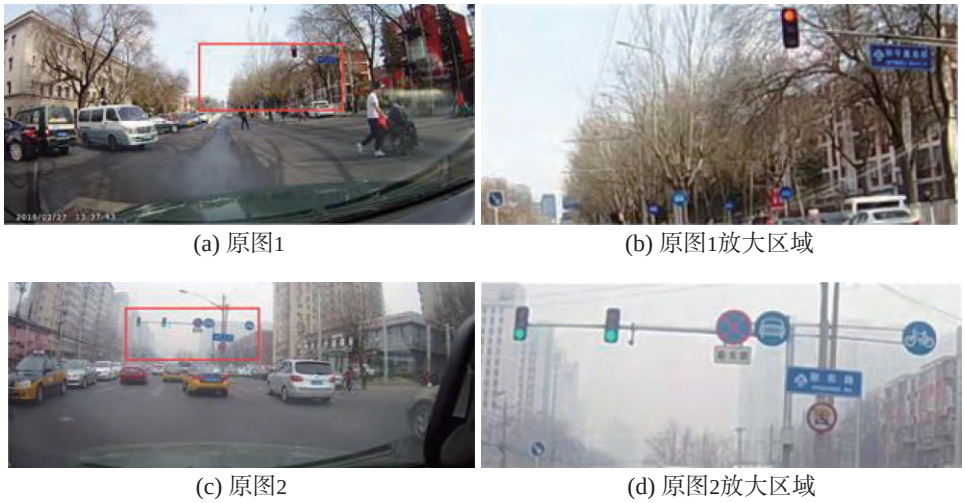
$$\frac{\partial l}{\partial\gamma_i}=\sum_{y_i}\frac{\partial l}{\partial y_i}\widehat{x_i} \tag{6}$$

使用如上标准化方式对 VGG16 同组的卷积层 conv4_1、conv4_2、conv4_3 进行标准化，然后将它们拼接起来，称之为组合特征，再输入到后续的网络中训练模型，然后进行分类和边框回归，这一改进能够检测到更多的小物体，小物体定位精确度也得到了一定的提升。

3　实验结果及分析

3.1　数据集

本文实验采用的交通标志检测数据集是由驭势科技提供的。数据集中的图片都是在真实行车环境中拍摄的, 总共有 10 000 张分辨率为 1 280×720 的图片。将该数据集划分为 3 部分, 4 000 张用来作训练集, 1 000 张用来作验证集, 5 000 张用来作测试集, 本文所有实验都是以 VGG16 在 ImageNet 上训练好的模型作为基础, 在此基础上进行微调, 每组实验迭代 70 000 次, 每 10 000 次保存一个模型, 挑选验证集上最好的模型在测试集上进行测试。图 2 是交通标志数据集中的一些样本图, 图 (b) 是图 (a) 中矩形框区域的放大图, 图 (d) 是图 (c) 中矩形框区域的放大图, 可以看到图片中存在大量较小的交通标志。图 3 是训练集中交通标志尺寸分布图。横轴是训练集中交通标志短边的范围, 纵轴是某个范围内的交通标志数量占训练集中所有交通标志的比例。从图 3 中可以看出有很大一部分交通标志短边范围在 8 ~ 16 个像素之间, 如果在 VGG16 网络的卷积 conv5_3 上产生候选框, 从原图到 conv5_3 边长缩小了 16 倍, 这部分候选框将很难被检测到。因此, 本文通过降低特征图缩小倍数, 采用 conv4_3 层的特征来产生候选框, 这样短边范围在 8 ~ 16 这一部分的交通标志将能够被更好地检测出来。



(a) 原图1　　(b) 原图1放大区域

(c) 原图2　　(d) 原图2放大区域

图 2　交通标志数据集样本图

Fig. 2　Sample images from traffic sign dataset

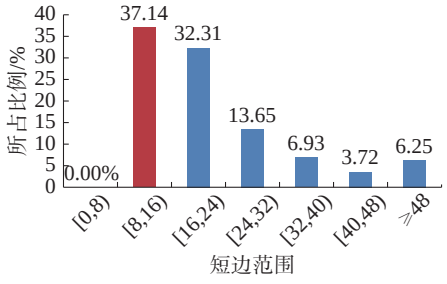


图 3　交通标志尺寸分布图

Fig. 3　Traffic sign's size distribution

3.2　实验结果

本文实验采用 mAP(mean average precision)[11] 来衡量最终的测试结果, mAP 值越大, 表明检测出来的结果越准确。表 1 是以不同深度的卷积神经网络作为 R-FCN 的主体网络的测试结果。从 mAP 值中可以看出, 主体网络的层数越深, 小物体的检测效果越不好。这是因为神经网络层数越深, 最后一层的特征图感受野越大, 候选框在映射到最后一层特征图时对于小物体的定位不是很准确。这也是采用 VGG16 作为 R-FCN 的主体网络来解决小物体检测的原因。

表 1　基于不同神经网络 R-FCN 交通标志检测结果

Table 1　Results based on different CNN's R-FCN

模型	mAP(验证集)	mAP(测试集)
ReNet-101+R-FCN	0.497	0.499
ResNet-50+R-FCN	0.514	0.519
VGG16+R-FCN	0.542	0.537

表 2 是采用不同改进版的 VGG16 作为 R-FCN 的主体网络, 分别将 conv3_3、conv4_3、conv5_3 输入到 RPN 网络来提取候选框, 并将该层的特征输入到后续的 R-FCN 网络中, RPN 网络提取的候选框映射到最后一个卷积层, 然后再进行分类和边框回归。从表 2 中可以看出, 在 R-FCN 物体检测框架上, 当降低特征图缩放倍数, 用 VGG16 的卷积 conv4_3

层特征提取候选框时，测试集上检测结果从 0.537 提升到 0.637，提高了 10 个百分点；用 VGG16 的卷积 conv3_3 层特征提取候选框时，检测结果从 0.537 提升到 0.596，提高了约 6%，没有在卷积 conv4_3 层上提取候选框的检测效果好。这是因为卷积 conv3_3 层的特征在网络中处于较浅层特征，抽象程度不够，不利于候选框分类。

表 2　基于 VGG16 网络不同层的 R-FCN 交通标志检测结果

Table 2　Results based on VGG16's different-layer R-FCN

模型	mAP(验证集)	mAP(测试集)
VGG16(conv3_3)+R-FCN	0.598	0.596
VGG16(conv4_3)+R-FCN	0.639	0.637
VGG16(conv5_3)+R-FCN	0.542	0.537

表 3 是采用 VGG16 作为 R-FCN 的主体网络，并对主体网络 VGG16 进行修改，对 conv4_1、conv4_2、conv4_3 分别进行不同的特征组合，并将组合特征输入到 R-FCN 的后续网络中。可以看到，基本上任意两层特征拼层输入到 R-FCN 的后续网络中效果都有提升，特别是将 3 层特征拼接到一起输入到 R-FCN 的后续网络中效果是最好的，达到了 65%。

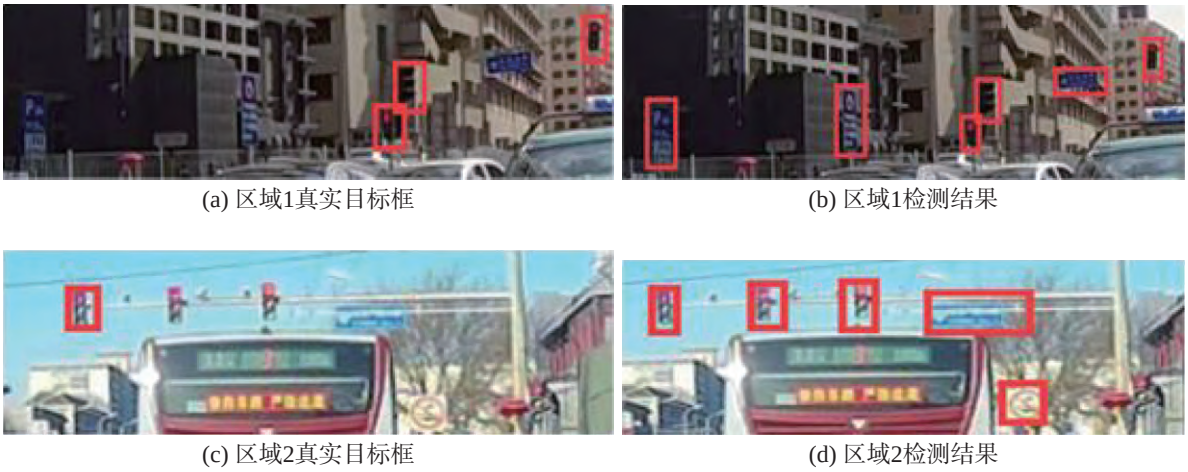
表 3　基于卷积 conv4_1、conv4_2、conv4_3 不同组合特征的检测结果

Table 3　Results based on conv4_1, conv4_2, and conv4_3 with different aggregated features

模型	mAP (验证集)	mAP (测试集)
VGG16(conv4_3)+R-FCN	0.639	0.637
VGG16(conv4_1+conv4_2)+R-FCN	0.648	0.642
VGG16(conv4_1+conv4_2)+R-FCN	0.648	0.642
VGG16(conv4_2+conv4_3)+R-FCN	0.640	0.636
VGG16(conv4__1+conv4_2+conv4_3) +R-FCN	0.650	0.650

3.3　结果展示

图 4 展示了本文提出的检测框架检测出来的交通标志与数据集提供的真实目标框作比较，图 (a)、(c) 中的边框是数据集给出的真实目标框，图 (b)、(d) 中的边框是使用本文的检测框架检测出来的结果。可以看出我们的检测框架检测出来的结果更好，可以检测到交通标志数据集中真实目标框漏标的交通标志，数据集中交通标志的漏标情况降低了本文实验的结果 mAP[11]。



(a) 区域1真实目标框　(b) 区域1检测结果

(c) 区域2真实目标框　(d) 区域2检测结果

图 4　检测结果与 ground truth 对比

Fig. 4　Comparison between the detection results and the ground truth

图 5 是从两张原始图中裁剪出来的部分区域，图 (a)、(b)、(c) 分别是使用不同模型的检测结果，图 (a) 是使用卷积 conv5_3 得到的检测结果，图 (b) 是使用卷积 conv4_3 得到的检测结果，图 (c) 是将卷积 conv4_1、conv4_2、conv4_3 组合层特征输入到后续网络得到的检测结果，可以看出两个关键改进点结合起来对于小物体的检测效果是最好的。

4　结束语

本文提出了针对小物体检测的物体检测框架。用 VGG16 作为 R-FCN 的主体网络，并对 VGG16 网络进行改进。改进后的检测框架对小物体的检测有了很大的提升，在驭势科技提供的存在大量小交通标志的数据集上取得了很好的效果。交通标志的精确检测对自动驾驶起着非常重要的作用，未来自动驾驶的真正到来时代离不开交通标志的精确检测和分类。本文只是将图片中的所有交通标志检测出来，并没有对这些交通标志进行分类，未来将会继续研究如何更精确地检测到较小的交通标志，并对检测出来的交通标志进行分类。

(a) conv5_3

(b) conv4_3

(c) conv4_1、conv4_2、conv4_3

图 5　使用不同层及组合层检测结果

Fig. 5　Using different layers and aggregated-layer-detection results

参考文献:

[1] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Proceedings of Advances in Neural Information Processing Systems. Stateline, NV, USA, 2012: 1097-1105.

[2] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA, 2015: 1–9.

[3] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[J]. arXiv:1412.7062, 2015.

[4] YU Gang, YUAN Junsong. Fast action proposals for human action detection and search[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA, 2015: 1302–1311.

[5] HONG S, YOU T, KWAK S, et al. Online tracking by learning discriminative saliency map with convolutional neural network[C]//Proceedings of the 32nd International Conference on Machine Learning. Lille, France, 2015: 597–606.

[6] WANG Naiyan, YEUNG D Y. Learning a deep compact image representation for visual tracking[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, USA, 2013: 809–817.

[7] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, 2016: 2818–2826.

[8] SAHA S, SINGH G, SAPIENZA M, et al. Deep learning for detecting multiple space-time action tubes in videos[J]. arXiv: 1608.01529, 2016.

[9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA, 2014: 580–587.

[10] 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(2): 154–171.

[11] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The Pascal visual object classes (VOC) challenge[J]. International journal of computer vision, 2010, 88(2): 303–338.

[12] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland, 2014: 346–361.

[13] GIRSHICK R. Fast R-CNN[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 1440–1448.

[14] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Proceedings of 2015 Advances in Neural Information Processing Systems. Montréal,Canada, 2015: 91-99.

[15] Li Y, He K, Sun J. R-fcn: Object detection via region-

based fully convolutional networks[C]//Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 379-387.

[16] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, 2016: 779–788.

[17] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of 14th European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 21–37.

[18] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, 2016: 770–778.

[19] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2014.

[20] LIU Wei, RABINOVICH A, BERG A C. ParseNet: looking wider to see better[J]. arXiv: 1506.04579, 2015.

作者简介：

葛园园，女，1991 年生，硕士研究生，主要研究方向为物体检测。

许有疆，男，1992 年生，硕士研究生，主要研究方向为视频动作识别。

赵帅，男，1988 年生，硕士研究生，主要研究方向为深度学习与机器学习、车辆动力学、自动驾驶技术、驾驶行为分析。