文章编号:1674-2974(2019)10-0133-08 DOI:10.16339/j.cnki.hdxbzkb.2019.10.016

用于交通标志检测的窗口大小聚类残差SSD模型

宋青松,王兴莉,张超†,陈禹,宋焕生,KHATTAK Asad Jan
(长安大学 信息工程学院,陕西 西安 710064)

摘 要:SSD通常被认为适合于求解小目标图像检测问题,但在特征表征和检测效率两方面还存在改进空间.提出一种聚类残差SSD模型,一方面将原始SSD模型中的VGG16基础网络替换为更深的ResNet50残差网络,以改善特征表征能力.另一方面采用K-均值聚类算法取代盲目搜索机制,确定SSD中默认窗口的大小,以改善检测效率.针对德国交通标志检测数据集,模型获得了97.1% mAP和每幅图像0.07 s的检测速度.针对中国交通标志数据集,模型获得89.7% mAP和每幅图像0.08 s的检测速度.与原始SSD模型比较,本文所提模型的检测性能得到改善.

关键词:交通标志检测;深度学习;单拍多盒探测器(SSD);K-均值;聚类
中图分类号:TP391.4 **文献标志码**:A

A Residual SSD Model Based on Window Size Clustering for Traffic Sign Detection

SONG Qingsong, WANG Xingli, ZHANG Chao†,
CHEN Yu, SONG Huansheng, KHATTAK Asad Jan
(School of Information Engineering, Chang'an University, Xi'an 710064, China)

Abstract: Single Shot MultiBox Detector (SSD) is generally considered to be suitable for solving small target detection in images. However, its performance on feature extraction and detection efficiency is still required to be improved. A clustering residual SSD model is proposed in this paper. On one hand, in order to improve the feature extraction quality, the basic network VGG16 which consists of the original SSD model is replaced with a deeper residual network ResNet50. On the other hand, in order to improve the detection efficiency, K-means algorithm other than the blind search mechanism used in the original SSD model is exploited to find and determine the assignments of the sizes of default windows. For German traffic sign detection dataset, it obtains 97.1% mAP in detection accuracy and 0.07 s per image in detection efficiency. For Chinese traffic sign dataset, it obtains 89.7% mAP in detection accuracy and 0.08 s per image in detection efficiency. Compared with the original SSD model, the proposed model obtains the improved detection performance.

Key words: traffic sign detection; deep learning; Single Shot MultiBox Detector (SSD); K-mean; clustering

---

* **收稿日期**:2018-11-21
基金项目:国家自然科学基金资助项目(61201406,61572083),National Natural Science Foundation of China(61201406,61572083);中国博士后科学基金资助项目(2019M653516),China Postdoctoral Science Foundation(2019M653516)
作者简介:宋青松(1980—),男,河南周口人,长安大学副教授,工学博士
† 通讯联系人,E-mail:2016224007@chd.edu.cn

交通标志的检测与分类是智能驾驶领域重要研究课题之一. 传统的方法多为基于候选区域和分类器的两段式分类检测方法.首先,使用滑动窗口选定图像的某一区域作为候选区域;针对选定的候选区域提取诸如 HOG (Histogram of Orientated Gradient,HOG)、Haar、SIFT (Scale-invariant feature transform,SIFT)、LBP(Local Binary Pattern,LBP)等[1-4]特征. 然后,使用随机森林(Random Forest,RF)、支持向量机 (Support Vector Machine,SVM)、Adaboost 等[5-9]分类算法对提取的特征进行分类,得出该候选区域的检测结果. 文献[9]以图像中交通标志的颜色、形状、空间位置等作为特征,使用 Adaboost 算法训练决策树模型,以此生成候选区域;之后使用 SVM 给出候选区域的类别,这是一种典型的两段式分类检测方法. 该类方法在候选区域的选取上往往存在盲目性, 同一图像会生成数以千计的候选区域,难以满足实时性要求;同时人工提取的特征通常难以很好地表征图像,影响检测准确率.

文献[10]提出区域卷积神经网络(Region CNN, R-CNN), 使用特征表征能力强的卷积神经网络(Convolutional Neural Networks,CNN)[11]提取候选区域的特征,检测准确率得到了很大的提升. 文献[12]首先使用全卷积网络 (Fully Convolutional Network, FCN)[13] 分割出交通标志候选区域, 然后使用 CNN 对候选区域进行分类. 该类方法本质上仍然是两段式的分类检测方法,在检测实时性方面改善有限.

YOLO 算法[14]将图像网格化,生成一组默认窗口,进而在该组默认窗口中心区域检测目标,由于不依赖候选区域,YOLO 算法在检测实时性方面取得了巨大的突破; 但分类层使用的特征尺度单一,在检测准确率方面没有获得显著提升.图像金字塔[9,15]和多尺度特征往往有利于模型性能的提升.文献[15]使用图像金字塔进行交通标志检测并取得一定效果. SPP-Net[16]、FPN[17]等算法将多尺度特征加入到分类决策层,有效提升了检测准确率.文献[18]提出的 SSD 在一定程度上综合了 YOLO 算法的默认窗口生成机制和 FPN 使用的多尺度特征融合思想,在检测速度和准确率方面都取得了良好的提升,但对于交通标志小目标的检测问题,在特征表征和检测效率两方面还存在改进空间.

针对交通标志小目标检测问题,本文提出一种聚类残差 SSD 模型,一方面将原始 SSD 模型中的基础网络替换为 ResNet50[19],提升特征表征能力;另一方面,引入 K-均值聚类算法取代默认窗口的随机生成机制,改善检测效率.实验结果表明,所提模型能改善交通标志小目标的检测性能.

1 模型结构

本文所提算法整体流程如图 1 所示. 首先对数据进行预处理,扩充训练样本;然后对模型依次进行预训练和微调,微调过程中引入 K-均值聚类算法实现窗口大小的启发式生成,克服原始 SSD 模型窗口生成机制具有内在盲目性这一缺陷;最后对模型检测性能进行评价.

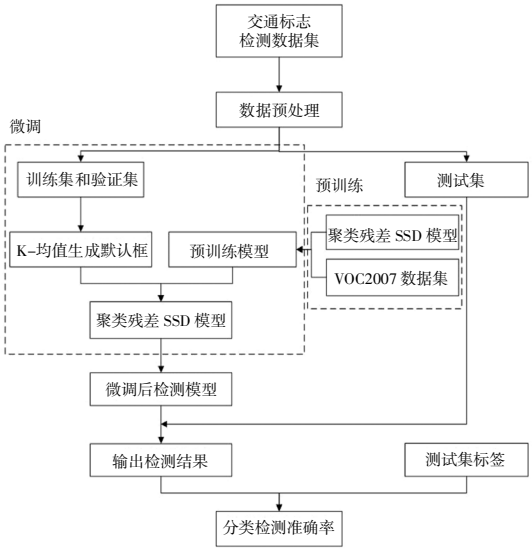


图 1 算法整体流程图

Fig.1 Flow chart of algorithm

1.1 原始 SSD 模型

原始 SSD 是以 VGG16[20]为基础网络,额外再依次堆叠 5 个卷积模块构成的一种深度 CNN 模型,共有 25 个卷积层和 5 个最大池化层. 输入为 512×512 像素图像,其中 5 个最大池化层将基础网络 VGG16 分隔为 6 个部分,前 5 个部分每部分包含卷积层的个数分别为 2、2、3、3 和 3,即卷积层 1~13,包含的卷积通道数分别为 64、128、256、512 和 512. 前 13 个卷积层卷积核大小均为 3×3,卷积跨度均为 1. 最大池化层的池化窗口为 2×2,跨度为 2. 第 6 个部分有两层卷积(卷积层 14、15),卷积层 14 的卷积核大小为 3×3,卷积通道为 1 024. 卷积层 15 的卷积核大小为 1×1,卷积通道为 1 024.

5 个额外堆叠的卷积模块中每个模块都有 2 层卷积(卷积层 16~25),第 1 个模块第 1 层卷积(卷积层 16) 的卷积核大小为 1×1, 第2 层卷积 (卷积层 17)的卷积核大小为 3×3. 卷积层 16 和 17 的跨度分

别为 1 和 2,通道数分别为 256 和 512. 第 2、3、4 个模块的第 1 层卷积(卷积层 18、20 和 22)的卷积核大小均为 1×1,跨度均为 1,通道数均为 128;第 2 层卷积(卷积层 19、21 和 23)的卷积核大小均为 3×3,跨度均为 2,通道数均为 256. 第 5 个模块第 1 层卷积(卷积层 24)的卷积核大小为 1×1,第 2 层卷积(卷积层 25)的卷积核大小为 4×4. 卷积层 24 和 25 的跨度均为 1,但通道数分别为 128 和 256. 原始 SSD 模型综合 10、15、17、19、21、23、25 这 7 层卷积层,将 7 层不同尺度的特征层用于目标的分类检测与位置回归[18].

1.2 ResNet50 残差网络

通常通过增加网络层数可以改善特征表征质量,但是由于受梯度消失或爆炸等梯度不稳定问题制约,以 VGG16 为基础的 SSD 模型难以通过进一步扩充网络深度以改善特征表征质量. 深度残差网络(ResNet)引入残差连接在一定程度上规避了梯度问题,这为改善 SSD 检测性能提供了可能.

常规深度神经网络的特征以连乘方式在层间传播,$H_L(x_i)$为特征分布函数,见式(1),x_i 为输入数据($i=1$)或第 i 层网络特征($i>1$),L 表示网络总层数,w_i 为网络第 i 层权重参数. 网络权重 w_i 与输入数据 x_i 经过卷积和激活函数 σ,最后以连乘方式输出特征,可能导致梯度不稳定发生[21].

$$H_L(x_i)=\prod_{l=i}^{L-i} w_l x_l$$
$$x_l=\sigma(w_{l-1}x_{l-1}) \tag{1}$$

ResNet 定义一个残差函数 $F(x_i)=H_L(x_i)-x_i$,将特征层间乘性传播改善为特征残差的层间加性传播,如公式(2)所示,从而避免了梯度不稳定问题发生. ResNet 的卷积层数可以达到数十甚至上百层,ResNet50 是一个典型模型,其深度可以达到 50 层[19]. 本文将 ResNet50 取代原始 SSD 模型中的 VGG16,以改善特征表征质量.

$$H_L(x_i)=x_i+\sum_{i=1}^{L} F(x_i,w_i) \tag{2}$$

1.3 K-均值聚类确定默认窗口大小

原始 SSD 模型采用基于默认窗口的目标预测检测. 默认窗口有两个自由参数:大小和长宽比.原始 SSD 模型中默认窗口的大小以随机方式指定,具体地,事先指定使用 7 个不同尺度的特征层作为预测层,通过

$$s_k=s_{\min}+\frac{s_{\max}-s_{\min}}{m-1}(k-1),k\in[1,m]$$

确定每个预测层中默认窗的基准大小[18]. 其中 m 表示预测层的数量,$m=7$;$s_{\min}$、$s_{\max}$ 分别表示第 1 和第 7 个预测层默认窗的基准大小;s_k 表示其他预测层默认窗的基准大小. 实践中 $s_{\min}$ 和 $s_{\max}$ 的值往往需要多次尝试,本质上是盲目搜索的,不具备任何启发性,导致检测效率受限.

本文采用 K-均值聚类算法取代原始 SSD 的盲目搜索方法生成默认窗的基准大小 s_k. 采用 K-均值算法对训练集中交通标志的大小聚类,鉴于选用 7 个不同尺度的特征层作为预测层,聚类中心个数指定取值 7,即聚成 7 个簇. 聚类中心点对应的窗口大小即为默认窗口的基准大小 s_k. 从而,训练集中检测目标的大小作为一种先验知识被聚类发现,指导默认窗口基准大小的选取.

1.4 聚类残差 SSD 模型

如图 2 所示,所提聚类残差 SSD 模型是以 ResNet50 作为基础网络,再额外叠加 5 个卷积模块构成的一个深度残差网络. 模型的参数设置如表 1 所示,网络共由 65 层,包含 59 个卷积层,5 个最大池化层. 输入为 512×512 像素图像,每经过池化层后输出的特征层大小都会减小为上层输入大小的 1/2,最后一层卷积层的大小为 1×1.

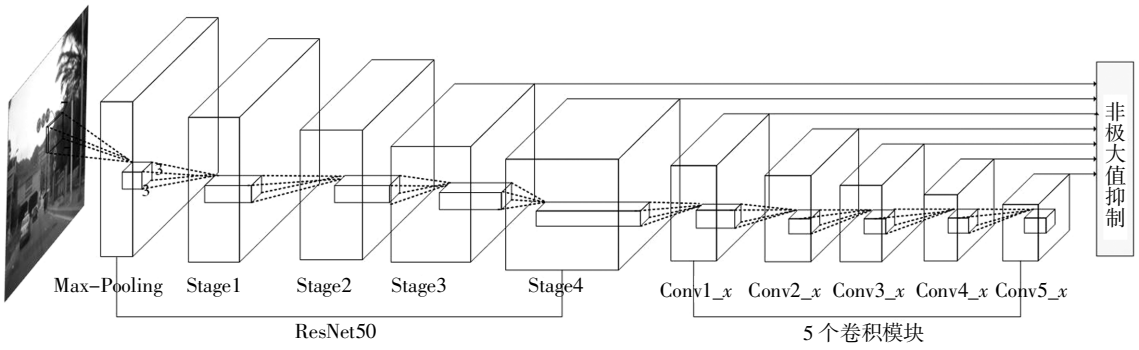


图 2　聚类残差 SSD 模型

Fig.2　The clustering residual SSD mode

表 1 聚类残差 SSD 模型的参数设置

Tab.1 Parameter settings of the cluster residual SSD model

结构	网络层	核大小	通道数	跨度	瓶颈结构	激活函数	输出
ResNet50	Conv	7×7	64	1	—	Relu	512×512
	Max-Pooling	3×3	64	2	—	Relu	256×256
	Stage1	1×1	64	1	×3	Relu	256×256
		3×3	64	1			
		1×1	256	1			
	Stage2	1×1	128	2	×4	Relu	128×128
		3×3	128	1			
		1×1	512	1			
	Stage3	1×1	256	2	×6	Relu	64×64
		3×3	256	1			
		1×1	1 024	1			
	Stage4	1×1	512	2	×3	Relu	32×32
		3×3	512	1			
		1×1	2 048	1			
5 个卷积模块	Conv1_1	1×1	256	1	—	Relu	32×32
	Conv1_2	3×3	512	1	—	Relu	32×32
	Max-Pooling	3×3	512	2	—	Relu	16×16
	Conv2_1	1×1	128	1	—	Relu	16×16
	Conv2_2	3×3	256	1	—	Relu	16×16
	Max-Pooling	3×3	256	2	—	Relu	8×8
	Conv3_1	1×1	128	1	—	Relu	8×8
	Conv3_2	3×3	256	1	—	Relu	8×8
	Max-Pooling	3×3	256	2	—	Relu	4×4
	Conv4_1	1×1	128	1	—	Relu	4×4
	Conv4_2	3×3	256	1	—	Relu	4×4
	Max-Pooling	3×3	512	2	—	Relu	2×2
	Conv5_1	1×1	128	1	—	Relu	2×2
	Conv5_2	3×3	256	2	—	Relu	1×1

已有研究表明,低层特征图含有更丰富的细节信息,对小目标的检测十分有用,而高层特征图具有较强的语义信息,适用于大目标的检测,结合高低多层特征图有利于不同尺度目标的检测. 首先,将模型中的 Stage3_5、Stage4_3、Conv1_2、Conv2_2、Conv3_2、Conv4_2 和 Conv5_2,7 种不同尺度的特征层用于预测目标,实现多尺度检测与识别. 接着,针对 7 种不同尺度的预测层,使用卷积核进行目标预测,同时输出目标分类置信度和目标框与预测框的相对位置偏移量. 记一个特征图的分辨率是 $m \times n$,每个像素单元指定 s_k 为基准大小的 b 个不同宽高比的默认框,每个默认框需要预测 c 个类别和 4 个相对偏移量 $\Delta(cx,cy,w,h)$,那么当前特征图有 $(c+4)\times b \times m \times n$ 个自由参数. 不同宽高比默认框的使用可以有效地离散输出框的形状,提高匹配精度和速度. 之后,当预测到该层有目标时,使用默认框与目标框进行匹配,匹配结果即为预测框. 本文使用 Jaccard Overlap 策略来匹配目标框和默认框[18],文中 Jaccard Overlap 的阈值设置为 0.5. 最后使用非极大值抑制去除冗余的预测框,本文非极大值抑制的阈值设为 0.6.

模型中总损失函数为定位损失和分类损失的加权和,其定义如式(3)所示,其中 L_{conf} 和 L_{loc} 分别表示分类损失和定位损失,N 是匹配的默认框个数,如果 $N=0$,则总的损失为 0,f 是每个预测框与目标框的匹配标志($f=1$ 表示匹配,$f=0$ 表示不匹配),例如 $f_{ij}^{p}=1$ 表示类别为 p 的第 i 个默认框与第 j 个目标框相匹配. 分类损失如式(4)所示,其中 c 表示类的置信度. 定位损失如式(5)所示,其中 l 表示预测框,g 表示目标框,d 为默认框,(cx,cy) 是相对中心点的偏移量[18].

$$L(f,c,l,g)=\frac{1}{N}(L_{\text{conf}}(f,c)+\alpha L_{\text{loc}}(f,l,g)) \tag{3}$$

$$L_{\text{conf}}(f,c)=-\sum_{i\in \text{Pos}}^{N} f_{ij}^{p}\log(\hat{c}_i^{p})-\sum_{i\in \text{Neg}}\log(\hat{c}_i^{0})$$

$$\hat{c}_i^{p}=\frac{\exp(c_i^{p})}{\sum_{p}\exp(c_i^{p})} \tag{4}$$

$$L_{\text{loc}}(f,l,g)=\sum_{i\in \text{Pos}}^{N}\sum_{m\in(cx,cy,w,h)} f_{ij}^{p}\,\text{smooth}_{L1}(l_i^{m}-\hat{g}_j^{m})$$

$$\hat{g}_j^{cx}=(g_j^{cx}-d_i^{cx})/d_i^{w},\ \hat{g}_j^{cy}=(g_j^{cy}-d_i^{cy})/d_i^{h}$$

$$\hat{g}_j^{w}=\log\left(\frac{g_j^{w}}{d_i^{w}}\right),\ \hat{g}_j^{h}=\log\left(\frac{g_j^{h}}{d_i^{h}}\right) \tag{5}$$

2 实验数据和性能评价指标

2.1 数据集

2.1.1 德国交通标志检测数据集(GTSDB)

GTSDB[22]数据库中的交通标志图像全部从自然场景中采集得到,如图 3 所示,有不同道路(高速公路、城市道路、乡村道路),不同光线(光线强和光线弱),不同天气(雨天、雾天、雪天)下的图像,合计 900 幅图像,每幅图像大小为 1 360×800 像素. 每一

幅图像有 1~4 个交通标志或者没有交通标志. 交通标志大小在 16×16~128×128 像素之间. 将所有交通标志按照如图 4 方式分为 3 类: 禁止标志 (Prohibitory)、指示标志(Mandatory)、危险标志(Danger). 900 幅图像被分为训练集和测试集两部分, 其中训练集为 600 幅图像,测试集为 300 幅图像.

图 3　GTSDB 交通场景图像

Fig.3　Traffic scene image of GTSDB

(a)禁止标志(Prohibitory)

(b)指示标志(Mandatory)

(c)危险标志(Danger)

图 4　GTSDB 交通标志类别图

Fig.4　Traffic sign class image of GTSDB

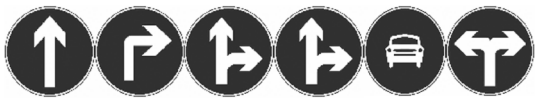
2.1.2　中国交通标志数据集 (CTSD)

CTSD[23]数据库中的图像是通过采集北京和厦门不同天气(晴天、雨天、大风)、不同道路(高速公路、城市道路、乡村道路)下的自然场景图像,部分如图 5 所示.图像为 1 024×768 和 1 270×800 像素两类. 一共有 1 100 幅图像, 训练集 700 幅, 测试集 400 幅.训练集中交通标志的大小在 20×20~380×378 像素之间,测试集中交通标志的大小在 26×26~573×557 像素之间. 将所有交通标志按照如图 6 所示方式分为 3 类: 禁止标志 (Prohibitory)、指示标志(Mandatory)、危险标志(Danger).

图 5　CTSD 交通场景图像

Fig.5　Traffic scene image of CTSD

(a)禁止标志(Prohibitory)



(b)指示标志(Mandatory)

(c)危险标志(Danger)

图 6　CTSD 交通标志类别图

Fig.6　Traffic sign class image of CTSD

2.2　训练样本扩充

为了改善模型鲁棒性, 扩充训练数据集. 扩充后的训练集包括①原始图像;②对原始图像再采样得到的图像块,与原图像目标的 Jaccard Overlap[18]分别为 0.1、0.3、0.5、0.7、0.9; ③将原始图像随机采样一部分. 采样后图像的大小为原始图像的 0.1~1 倍, 宽高比在 1/2~2 之间, 当目标框的中心在采样后的图像中时,裁去目标框落在图像外面的部分,保留重叠部分. 经过上述采样之后,将每个采样的小块调整到 512×512 像素,并以 0.5 的概率对其水平翻转.

2.3　性能评价

采用精确率(Precision)和召回率(Recall)的曲线 PR 所包围的面积 AP 来评价模型测试准确率. AP 取值越大,表明检测准确率越高. mAP 是所有类别 AP 的平均值. 其中 Precision 与假阳性样本个数(FP)和正阳性样本个数(TP)的关系,Recall 与假阴性样本个数(FN)和正阳性样本个数(TP)的关系如式(6)所示:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}} \tag{6}$$

3　实验结果及分析

针对原始 SSD 模型和所提 SSD 模型,开展对比实验.

3.1　预训练与微调

本文所提聚类残差 SSD 模型仅在预先训练好

的 ResNet50 网络的基础上进行微调. 预训练使用 VOC 2007 数据集[24]. 微调过程中,通过损失函数最小化达到模型最优.优化器选用 Adam;训练轮次为 500;学习率采用动态的方式,当轮次小于 300 时,学习率为 10^{-4};当轮次大于 300 时,学习率为 10^{-5}.

3.2 实验过程及结果

本文所用实验环境,硬件配置为 I7 7700K 处理器、16 G 内存和 Titan XP 显卡, 软件配置为 Ubuntu16.04、Python3.5 和 Keras.

3.2.1 原始 SSD 模型的训练

使用 2.1 中所介绍的 GTSDB 数据集, 对原始 SSD 模型进行训练,根据文献[18]中默认框的生成方式将 s_k 设置为[0.05,0.13,0.21,0.29,0.37,0.45,0.53], 训练和验证过程中损失函数如图 7 所示,训练集和验证集的损失不断减小,当轮次到 300 次时训练基本达到收敛. 将测试集在该模型上进行实验,测试结果如表 2 第 4 行所示,检测的 mAP 可以达到 94.5%, 每幅图像的检测速度是 0.05 s. 将原始 SSD 模型的基础网络替换为 ResNet50,检测结果如表 2 第 5 行所示,检测的 mAP 可以达到 96%,每幅图像的检测速度是 0.07 s. 以 ResNet50 为基础网络的 SSD 模型比原始 SSD 模型的 mAP 高 1.5%,每幅图像的检测速度慢 0.02 s,检测效率略降,准确率得到改善.

3.2.2 聚类残差 SSD 模型训练

使用 GTSDB 对所提模型进行训练. 首先根据 K-均值聚类算法将 GTSDB 训练集中交通标志的长和宽聚成 7 个簇,如图 8 所示,聚类中心点对应默认窗口的基准大小. 得到默认窗口的基准大小分别是 [8.65,14.69], [12.12,20.38], [15.69,26.48], [19.95,33.44], [25.25,41.88], [31.81,52.80] 和 [42.57,70.56]. 训练和验证过程中损失函数如图 7 所示, 当轮次到 300 次时认为模型已收敛. 从图中可以发现,所提模型最后收敛损失小于原始 SSD 模型的收敛损失. 检测结果如表 2 第 6 行所示,mAP 达到 97.1%,每幅图像的检测速度 0.07 s.所提模型比以 ResNet50 为基础网络的 SSD 模型 mAP 高出 1.1%,比原始 SSD 模型高出 2.6%.同时开展本文算法和 Faster R-CNN[25]和 FPN 检测比对实验,结果见表 2 第 2 行和第 3 行,均表明所提模型检测性能得到明显改善.

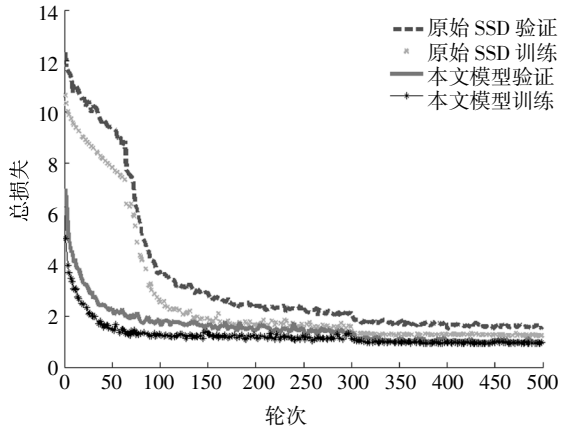


图 7 损失变化曲线

Fig.7 Loss variation curve

表 2 检测结果

Tab.2 Test results

方法	数据集	禁止标志(AP)/%	指示标志(AP)/%	危险标志(AP)/%	mAP/%	Time/(s·幅$^{-1}$)
Faster R-CNN	GTSDB	86.5	84.4	81.1	84.0	0.08
FPN		98.4	89.89	90.8	93.0	0.11
原始 SSD 模型		99.1	91.5	92.9	94.5	0.05
ResNet50+SSD 模型		99.2	95.4	93.2	96.0	0.07
本文所提模型		99.5	96.9	94.8	97.1	0.07
本文所提模型	CTSD	88.2	85.5	95.3	89.7	0.08

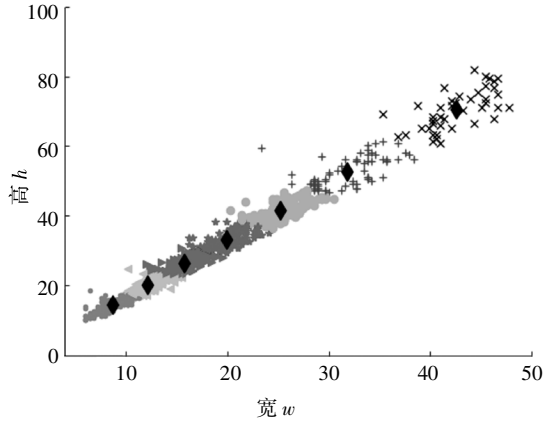


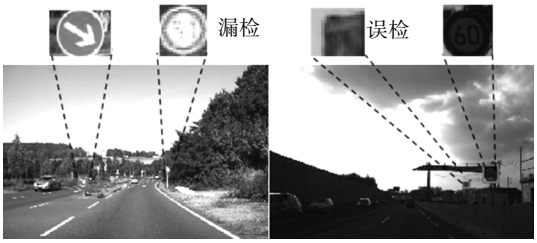
图 8 聚类结果

Fig.8 Clustering results

图 9 给出了不同天气、不同交通场景下 2 幅典型图像的交通标志小目标检测结果. 图 9(a)中所有

的目标都被检测出来,图 9(b)中第 1 幅图像有一个漏检的目标，第 2 幅图像有一个误检的目标. 所提模型对天气变化和交通场景改变具有更好的鲁棒性.

(a)所提模型的检测结果



(b)原始 SSD 模型的检测结果

图 9　GTSDB 图像检测结果

Fig.9　Detection results on GTSDB images

3.3　实验分析

表 2 给出了同一数据集(GTSDB)下不同算法的实验结果. 可以看出,在检测准确率方面,本文算法达到最优 mAP 为 97.1%，比原始 SSD 模型有 2.6%的提升,比 FPN 算法有 4.1%提升;同时检测效率并没有受到明显的影响,仅比原始的 SSD 模型减少 0.02 s. 单类检测准确率方面，禁止标志的 AP 可以达到 99.5%,优于其他所有算法;指示标志和危险标志的 AP 分别为 96.9%和 94.8%,检测准确率优于其他所有算法,但并没有达到和禁止标志相当的检测准确率. 针对这一问题,对数据集中禁止标志、指示标志、危险标志进行统计,并以直方图的方式可视化,如图 10 所示,可以看出指示标志和危险标志的数量都较少,存在数据不均衡问题,影响了模型对该类标志的检测性能. 后续研究中，考虑适当增加数据,或收集更丰富的数据集进行研究,以提升不同场景图像检测的鲁棒性.

3.4　算法验证

在相同实验条件下，针对 CTSD 验证所提模型的有效性. 根据 K-均值算法得到默认窗口的基准大小分别为 [12.95,19.65]，[21.00,32.21]，[31.65,48.29]，[44.19,68.46]，[62.05,97.23]，[87.02,129.97]和 [152.35,233.55]. 其他训练和测试过程与 GTSDB 的相同. 检测结果如表 2 第 7 行所示,本文所提模型获得了 89.7% mAP 和每幅图像 0.08 s 的检测速度.

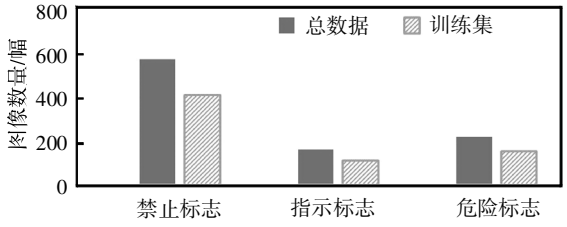


图 10　3 类标志的数量分布

Fig.10　The number of distribution map of three class markers

图 11 给出了 CTSD 中两幅典型的自然场景交通图像,左图是晴朗天气下城市道路场景,右图是阴雨天气下高速公路场景. 本文所提模型将图中的交通标志小目标全部正确检出.

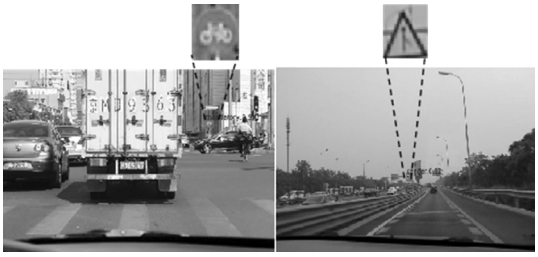


图 11　CTSD 图像检测结果

Fig.11　Detection results on CTSD images

4　总　结

针对自然场景中交通标志小目标检测问题,以及原始 SSD 模型用于小目标检测时特征表征能力和检测效率两方面的不足，提出一种聚类残差 SSD 模型. 一方面将原始 SSD 模型的基础网络 VGG16 替换为特征表征能力更强的 ResNet50 深度残差网络;另一方面采用 K-均值聚类算法发现小目标默认窗口的大小,实现默认窗口大小的优化选择,改善了原始 SSD 模型中盲目搜索默认窗口大小的缺陷. 针对 GTSDB 基准数据集的测试获得了 97.1% mAP 和每幅图像 0.07 s 的检测速度，针对 CTSD 基准数据集的测试获得了 89.7% mAP 和每幅图像 0.08 s 的检测速度,表明所提模型求解交通标志小目标检测问题的有效性.

参考文献

[1]　DALAL N,TRIGGS B. Histograms of oriented gradients for human

detection [C]//IEEE Conference on Computer Vision & Pattern Recognition.Washington DC:IEEE Computer Society,2005:886—893.

[2] VIOLA P,JONES M. Rapid object detection using a boosted cascade of simple features [C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society,2001:511—518.

[3] LOWE D G. Distinctive image features from scale-invariant key points [J]. International Journal of Computer Vision,2004,60(2):91—110.

[4] OJALA T,HARWOOD I. A comparative study of texture measures with classification based on feature distributions [J]. Pattern Recognition,1996,29(1):51—59.

[5] BOI F,GAGLIARDINI L. A support vector machines network for traffic sign recognition[C]//International Joint Conference on Neural Networks. Washington DC: IEEE Computer Society,2011:2210—2216.

[6] WANG G Y,REN G H,WU Z L,*et al*. A hierarchical method for traffic sign classification with support vector machines[C]//The 2013 International Joint Conference on Neural Networks (IJCNN). Washington DC: IEEE Computer Society,2013:1—6.

[7] TANG S S,HUANG L L. Traffic sign recognition using complementary features [C]//2013 2nd IAPR Asian Conference on Pattern Recognition (ACPR). Washington DC: IEEE Computer Society,2013:210—214.

[8] SUGIYAMA M. Local fisher discriminant analysis for supervised dimensionality reduction[C]//Proceedings of the 23rd International Conference on Machine Learning. New York : ACM,2006: 905—912.

[9] CHEN T,LU S J. Accurate and efficient traffic sign detection using discriminative AdaBoost and support vector regression [J].IEEE Transactions on Vehicular Technology,2016,65(6):4006—4015.

[10] GIRSHICK R,DONAHUE J,DARRELL T,*et al*. Rich feature hierarchies for accurate object detection and semantic segmentation[C] // Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Ohio: IEEE,2014: 580—587.

[11] LECUN Y,BOTTOU L,BENGIO Y,*et al*. Gradient-based learning applied to document recognition [J].Proceedings of the IEEE,1998,86(11):2278—2324.

[12] ZHU Y Y,ZHANG C Q,ZHOU D Y,*et al*. Traffic sign detection and recognition using fully convolutional network guided proposals [J].Neurocomputing,2016,214:758—766.

[13] SHELHAMER E,LONG J,DARRELL T,*et al*. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(4): 640—651.

[14] REDMON J,DIVVALA S,GIRSHICK R,*et al*. You only look once: unified,real-time object detection [C] // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern. Nevada: IEEE,2016:779—788.

[15] MENG Z B,FAN X C,CHEN X,*et al*. Detecting small signs from large images [C] // Proceedings of the 2017 IEEE International Conference on Informatiorn Reuse and Integration. California: IEEE,2017: 217—224.

[16] HE K M,ZHANG X Y,REN S Q,*et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,37(9):1904—1916.

[17] TSUNG Y L,PIOTR D,ROSS G,*et al*. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu:IEEE,2007:936—944.
Facebook AI Research (FAIR),Cornell University and Cornell Tech,2017.

[18] LIU W,ANGUELOV D,ERHAN D,*et al*. SSD: single shot multibox detector [C]//Proceedings of the 2016 European Conference on Computer Vision. Amsterdam: ECCV,2016:21—37.

[19] HE K M,ZHANG X Y,REN S Q,*et al*. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas:IEEE,2016:770—778.

[20] SIMONYAN K,ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]//International Conference on Learning Representations. San Diego:Oxford,2015:1—14.

[21] 陈建廷，向阳. 深度神经网络训练中梯度不稳定现象研究综述[J]. 软件学报,2018,29(7):2071-2091.
CHEN J T,XIANG Y. Research review on gradient instability in deep neural network training [J]. Journal of Software,2018,29(7):2071—2091.(In Chinese)

[22] HOUBEN S,STALLKAMP J,SALMEN J,*et al*. Detection of traffic signs in real-world images:The German traffic sign detection benchmark[C]//The 2013 International Joint Conference on Neural Networks(IJCNN). Dallas:IEEE,2013:1—8.

[23] YANG Y,LUO H L,XU H R,*et al*. Towards real-time traffic sign detection and classification [J]. IEEE Transactions on Intelligent Transportation Systems,2016,17:2022—2031.

[24] EVERINGHAM M,GOOL L,WILLIAMS C K,*et al*. The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision,2010,88(2): 202—228.

[25] REN S Q,HE K M,GIRSHICK R,*et al*. Faster R-CNN: towards real-time object detection with region proposal networks [C]// Advances in Neural Information Processing Systems 28. Montreal: NIPS,2015:91—99.