

文章编号: 1671-0444(2018)04-0520-06

# 基于深度学习的高分辨率遥感图像车辆检测

孙秉义<sup>1</sup>, 文珊珊<sup>2</sup>, 吴昊<sup>3</sup>, 蔡鸿明<sup>1</sup>

(1. 上海交通大学 软件学院, 上海 200240;  
2. 上海航天设备制造总厂, 上海 200240; 3. 南京润辰科技有限公司, 江苏 南京 210021)

**摘要:** 针对公路交通量的传统检测方法存在周期较长且需要人工辅助等问题, 利用深度学习在目标检测领域的优势, 提出基于深度学习的高分辨率遥感图像车辆检测方法。首先对遥感图像进行预处理和分割, 提取所需的道路区域, 减少其他区域的干扰; 再利用高分辨率遥感图像数据集对深度卷积神经网络进行训练, 得到用于车辆检测的模型, 并最终得到有效的车辆检测系统。经过试验验证, 该方法可有效地检测遥感图像道路区域中的车辆, 并有较高的准确率。

**关键词:** 图像分割; 特征提取; 卷积神经网络; 车辆检测; 遥感图像

**中图分类号:** TP 391      **文献标志码:** A

## Vehicle Detection in High-Resolution Remote Sensing Images Based on Deep Learning

SUN Bingyi<sup>1</sup>, WEN Shanshan<sup>2</sup>, WU Hao<sup>3</sup>, CAI Hongming<sup>1</sup>

(1. School of Software, Shanghai Jiao Tong University, Shanghai 200240, China;  
2. Shanghai Aerospace Manufacture (Group) Co. Ltd., Shanghai 200240, China;  
3. Nanjing Runchen Technology Co. Ltd., Nanjing 210021, China)

**Abstract:** The traditional detection method for highway traffic volume needs to take long time and manual assistance. Based on the advantages of deep learning in target detection area, a vehicle detection method for high-resolution remote sensing images was proposed. Firstly, the remote sensing images were preprocessed and segmented to extract the road area and to reduce the interference in other regions. Then, the deep convolution neural network was trained by high resolution remote sensing image data set to obtain the detection model and vehicle detection system. The experimental results show that the method can detect the vehicles in the road area of remote sensing images effectively and has a high accuracy rate.

**Key words:** image segmentation; feature extraction; convolutional neural network; vehicle detection; remote sensing image

公路交通在我国交通体系中占据主导地位。近些年来, 随着公路交通建设和路网规模的不断扩张, 公路交通的管理要求不断提高, 即不仅要提升公路交通质量和效益, 更要提升公路交通服务水平。

其中, 信息化是公路交通现代化管理的核心, 快速准确地收集道路信息和车辆情况是交通管理的基础。

目前我国道路交通信息数据获取主要通过线圈监测技术、红外线车辆检测技术、GPS 定位、手机定位技

收稿日期: 2018-04-15

基金项目: 国家自然科学基金资助项目(61373030)

作者简介: 孙秉义(1992—), 男, 山东潍坊人, 硕士研究生, 研究方向为信息可视化, E-mail: 5111619021@sjtu.edu.cn

蔡鸿明(联系人), 男, 教授, E-mail: hmcai@sjtu.edu.cn

术、视频检测技术等。文献[1]提出一种基于 Gabor 滤波器和支持向量机(SVM)分类器的红外车辆检测方法,通过选定候选区域和确认车辆,对红外图像中的车辆进行识别。文献[2]对基于视频图像处理的车辆检测和跟踪方法进行概述,该方法首先从视频图像中提取感兴趣区域,再输入到车辆检测和跟踪模块,即可得到交通流量、车流密度等信息。文献[3]提出了基于手机信令的交通量识别和预测方法,该方面以手机为探针,通过手机信令获取车辆位置和速度信息,从而反映出道路拥堵状态。这些方法的缺点在于成本高、检测周期长、检测速度慢、覆盖范围小,而且需要人工的辅助,并在一定程度上受到天气等自然因素的制约。因此,传统技术无法满足公路交通网络高速发展的现状,严重影响我国交通管理和服务质量的提升。

随着遥感技术的不断发展和商业卫星分辨率的提高,将高分辨率的遥感图像应用于交通道路管理成为可能,用遥感图像进行物体检测也成为新的研究方向。目前有多种技术可以从遥感图像中识别车辆。文献[4]提出结合对象几何形状、灰度值以及图像的特征矩来进行对象识别的方法。文献[5]结合方向梯度直方图和哈尔小波转换描述图像特征,并以多核 SVM 作为分类器进行车辆检测。文献[6]提出区域相关的方法即从低像素级别的多光谱图中自动化提取车辆位置,然后通过遥感图和多光谱图成像的时间差来确定车辆移动速度。文献[7]提出了基于空间稀疏编码的词袋模型,通过滑动窗口和特征提取来选择处理单元,建立一个新的空间映射策略来进行几何信息的编码。该方法不仅能表示目标物体各个部分的相对位置,也可以处理旋转变换,同时引入稀疏编码以降低重建错误率。

由以上研究可知,在目前的方法中,图像的特征大多依靠人工提取得到,需要不断尝试来得到最适合的特征,因此成本比较高。而且不同来源的遥感图像分辨率不相同,可能某些特征在某个来源的图像上有较好的效果,但是在其他来源的遥感图像却未必有好的效果,那就需要重新选择特征来进行识别。同时,这种手工提取的图像特征在较复杂场景的车辆检测中精度也不能满足实用要求。因此如何自动地提取图像特征以及提高车辆检测的精度,目前仍是一项高难度的任务。

深度卷积神经网络(DNN)能够自动获取图像特征并训练分类器,在图像识别领域有非常好的表现。针对上述问题,本文提出了基于深度学习的遥感图像车辆检测方法,将结合遥感图像特点和深度卷积神经网络的优点对车辆检测方法进行改进,并提高检测准确度。

## 1 遥感图像道路车辆检测平台框架

考虑到不同设备的计算能力不同,为了降低客户端系统构建的复杂性以及资源使用,本文将遥感道路车辆检测平台构建在云平台上,客户端只需要将获取的图像上传到云平台即可进行识别。采用这种架构不仅节省了客户端计算资源,而且适用于各种移动终端设备。

本文设计并实现了一个遥感图像处理平台,针对遥感图像道路区域中的车辆检测提出了一套完整的解决方案,系统架构如图1所示。平台分为客户端和服务端两部分,用户通过客户端将要检测的图像上传至服务器,服务器通过已经训练好的模型和方法进行检测,将车辆检测结果反馈给用户。

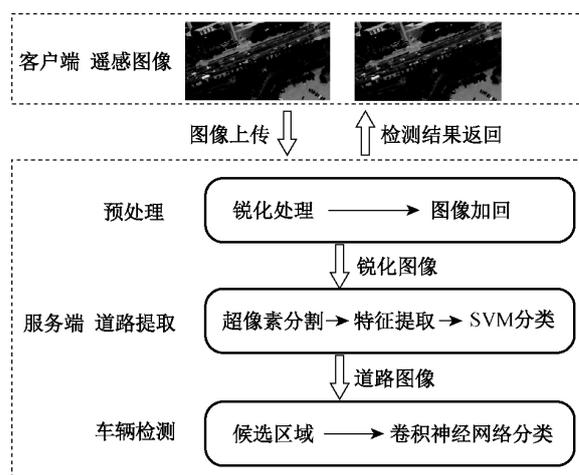


图1 遥感图像车辆检测系统架构图

Fig. 1 System framework of vehicle detection in remote sensing images

其中,服务端处理框架主要分3个部分:

(1) 图像预处理。主要任务是锐化图像,突出边缘信息,为后续道路区域提取和车辆检测增强图像质量。本部分输入为客户端上传的原始遥感图像。

(2) 道路区域提取。本文只关注道路区域中的车辆,而遥感图像中包含许多无关区域的车辆,如城区、停车场等,因此通过特征提取,识别出遥感图像中的道路区域。图像预处理的结果作为本部分输入。

(3) 车辆检测。运用深度学习算法训练网络模型,识别并标注出图像中的车辆位置。本部分输入为只包含道路区域的图像,输出为带车辆标注的完整遥感图像。

## 2 道路区域提取

### 2.1 图像预处理

在遥感图像形成和传输的过程中,可能会由于

天气、光线、地物阴影以及系统噪声等干扰因素导致图像质量下降。因此在进行道路区域提取之前需要对图像进行预处理,以保证获得更加准确的图像分割和车辆标注。

图像预处理主要进行图像增强以突出目标的细节特征。在道路区域提取中,清晰的道路边缘能够保证区域提取效果的准确。同样,车辆标注也需要突出车辆的几何形状和边缘信息。本文采用索贝尔算子对图像进行锐化处理,索贝尔算子由 2 个  $3 \times 3$  的矩阵组成,分别对图像进行卷积操作,得到横向和纵向的灰度值差,最终计算得到每个像素的灰度值。索贝尔算子对图像的高频部分进行了增强,突出了边缘特征,同时对噪声有平滑作用。

对图像进行滤波操作时,定义一个加回值(add back),该值表示原始图像中加回到滤波结果部分的大小,保证图像的空间连续性。

## 2.2 道路区域提取

遥感图像中通常不只包含道路区域,还有树木、建筑等无关区域,这些无关区域中的车辆会影响最终的车辆检测结果,造成一定程度的虚警,因此在进行车辆检测之前,需要对交通道路区域进行提取。道路区域提取过程中通过剔除其他无关区域,排除其他区域中车辆对检测结果的干扰,而且大大降低了输入图像的大小,从而减少了车辆检测算法所需处理时间。

本文采用特征提取和 SVM 对道路区域进行分割,具体流程分为以下 3 个步骤:

- (1) 对遥感图像进行超像素区域划分;
- (2) 对每个超像素块进行特征提取;
- (3) 将提取得到的特征通过 SVM 进行分类,最终得到只包含道路区域的图像。

在对道路进行特征提取之前,本文通过超像素分割技术对图像进行处理。超像素分割技术通过像素之间的相似性,把像素划分为超像素块,在进行分类计算时,不需要对每个像素的特征进行分类,而是对超像素块的特征进行分类,从而大大降低了图像处理的复杂度,减少运算时间。超像素分割现在有多种算法,本文选用 SLIC(simple linear iterative clustering)算法<sup>[8]</sup>,其是对  $k$ -means 算法的改进,主要有速度快、节约内存、边缘信息保留度好等优点,具体算法如下:

Initialization:

初始化步长为  $S$  的像素区域的聚类中心  $C_k = [l_k, a_k, b_k, x_k, y_k]^T$

将聚类中心移动至  $3 \times 3$  邻域内梯度最小的位置  
设置每个像素的标签  $l(i) = -1$

```

设置每个像素的距离  $d(i) = \infty$ 
repeat:
  for 每个像素中心 do
    for 每个聚类中心  $C_k$  周围  $2S \times 2S$  的区域中的像素  $i$  do
      计算  $C_k$  与  $i$  的距离  $D$ 
      if  $D < d(i)$  then
        设  $d(i) = D$ 
        设  $l(i) = k$ 
      end if
    end for
  end for
  计算新的聚类中心
  计算残留误差  $E$ 
until  $E \leq$  阈值

```

由 SLIC 算法完成图像的超像素划分之后,再对各个超像素块进行特征提取,提取的结果将输入到 SVM 分类器中进行分类,最终形成只包含道路部分的图像。超像素块由特征相似的元素组成,需对其中的每个像素的颜色、纹理进行特征提取,从而计算出每个超像素块的特征。

通过 YUV 颜色空间描述图像的强度和颜色特征,其中,  $Y$  表示图像的强度特征,  $U$  和  $V$  表示图像的色彩度信息。YUV 模型将图像的颜色和强度分开表示,以方便特征提取。将图像的 RGB 值转换为 YUV 值,并以直方图形式统计图像的强度和颜色信息,作为最终超像素块特征的一部分。颜色直方图反映了超像素块内颜色的分布和色调,但其丢失了像素的位置信息,不同超像素块可能有相同的颜色直方图,但是它们的颜色空间信息可能不相同。

超像素块的纹理特征通过 LBP(local binary pattern)特征进行描述,即针对每个像素,对比其周围  $3 \times 3$  区域内像素值与其像素值的大小。如果周围像素值大于中心像素值,则标记为 1;如果周围像素值小于中心像素值,则标记为 0。最终周围 8 个方位标记可以表示为二进制像素值,转为十进制后作为其 LBP 特征值。对每个超像素块的 3 个通道计算其 LBP 特征值,以统计直方图的形式表示,并对数据进行归一化处理。

SVM 属于线性分类器,能够将所提取的特征向量进行分类。将上述每个超像素块提取的特征输入到已训练的 SVM 分类器中,判断其是否属于道路区域,最终输出仅包含道路区域的图像。

## 3 基于深度学习的车辆检测

### 3.1 候选区域

遥感图像的尺寸通常较大,虽然经过道路区域提

取之后,图像尺寸会大大减小,但是仍包含着很多背景区域。在进行分类之前,需要选取部分窗口作为候选区域,然后将这些区域输入到分类器中判断是否包含车辆目标。传统方法通常包含以下3个步骤:

(1) 通过不同尺寸的滑动窗口框选一部分区域作为候选区域;

(2) 提取候选区域的特征,如 HOG 特征、Harr 特征等;

(3) 将特征输入到分类器中,如 SVM 分类器。

传统方法的第一步本质上是穷举的策略,由于目标可能出现在图中的任意位置,所以需要设计不同长宽比的滑动窗口来对图像进行遍历搜索。虽然此方法能够在一定程度上保证候选区域包含所有的目标,但是由于窗口数量太多以及冗余等会影响后序的特征提取和分类,造成耗时较长。

近些年随着深度学习的发展,卷积神经网络在图像识别和目标检测领域有着越来越广泛的应用。其中,R-CNN<sup>[9]</sup>(regions with convolutional neural network features)方法解决了传统方法中滑动窗口针对性差、窗口冗余的问题,通过选择性搜索算法<sup>[10]</sup>找出可能存在的目标区域,将窗口数量控制在 1 000~2 000 个,这种方法在减少了窗口数量的同

时提高了窗口质量,并且保证了一定的召回率。

选择性搜索算法原理是:计算相邻区域的相似度,将相似度较高的区域进行合并,直到区域合并为整张图,最后将合并过程中所有出现过的区域作为候选区域。本文应用选择性搜索算法对输入的道路区域图像提取候选区域。

在进行道路区域提取时已经分别得到了区域颜色特征和纹理直方图,其中,颜色特征在 YUV 颜色空间提取,纹理直方图用 LBP 特征表示。结合颜色、纹理和大小特征来计算区域的相似度,并将相似度高的区域进行合并,最后输出候选区域。根据候选区域得到边界框,将框选的内容输入到深度神经网络中进行分类。

### 3.2 基于卷积神经网络的车辆检测

本文采用的神经网络结构借鉴 VGG-16 卷积神经网络<sup>[11]</sup>,并根据数据集类别对其进行修改,如图 2 所示。该网络结构主要包含 5 段卷积,每段包含多个卷积层,每段卷积后接一层最大池化层,最后接 3 层全连接层输出分类结果。每个卷积层所采用的卷积核大小均为  $3 \times 3$ ,步长均为 1;池化层卷积核大小为  $2 \times 2$ ,步长为 2。根据本文采用的数据集,将最后一层全连接层输出类别大小修改为 15。

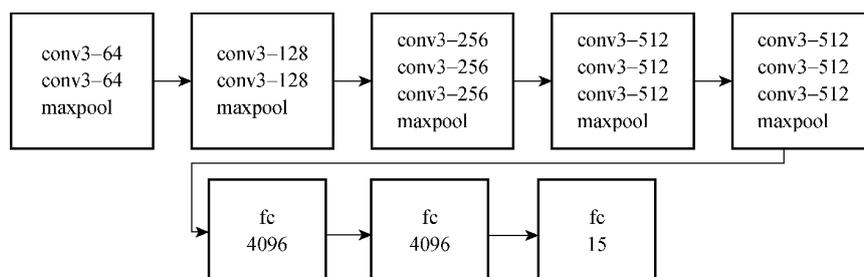


图 2 VGG-16 结构图

Fig. 2 Framework of VGG-16

与其他类型的卷积神经网络相比,VGG-16 神经网络的深度更深,并且采用了更小的卷积核( $3 \times 3$ )和池化核( $2 \times 2$ )。小卷积核减少了卷积层的参数量,并能得到更丰富的特征。每层卷积后都有一个 ReLU 函数,多卷积层可以增强决策函数的辨别能力,更深的深度使得网络容量更大,区分能力更强。

经过选择性搜索算法得到的候选框是矩形框,而且大小、长宽比各不相同,而卷积神经网络要求输入的图像大小是固定的,因此需要对候选框进行处理,以满足输入大小的要求。本文通过对候选区域进行各向异性缩放,将区域大小调整为  $224 \text{ 像素} \times 224 \text{ 像素}$ ,将其作为神经网络的输入。经过神经网络识别之后,可能有多个互相重叠的候选

矩形框都被识别为同一目标,而对某个目标,只需要保留一个最准确的矩形框即可。本文采用非极大抑制算法来消除多余的候选区域框。

虽然深度学习算法在 ImageNet 等常规数据集上有着不错的效果,但是在遥感图像上应用还不是很广泛,原因在于遥感图像和常规的图像有着较大区别。一般图像的拍摄角度大多位于同一个平面,而遥感图像大多是高空视角下拍摄,以俯拍图像居多。此外,遥感图像中的物体所占区域较小,可能只包含几十个像素,目标信息量很小,相比一般图像,其能够获取的特征也较少。而且遥感图像视野范围较大,背景区域复杂多样,对目标检测干扰较强。因此,为了增强车辆检测的准确度,本文不能用常规数

数据集来进行训练,而需要用大规模标注的遥感图像数据集来进行网络的训练和测试。

## 4 系统实现及验证

### 4.1 系统实现

本文选用 DOTA(dataset for object detection in aerial images)高清遥感图像数据集<sup>[12]</sup>来进行网络的训练,此数据集中对大小型车辆、建筑等 15 类目标进行了标注,共包含 2 806 张图像、188 282 个目标,是目前质量较高、数量较大的遥感数据集。

基于深度学习的遥感图像车辆检测系统如图 3 所示。由图 3 可知,从客户端将原始的遥感图像上传至服务器,图像在服务器中经过预处理、道路区域提取、车辆检测等步骤,检测结果分别输出图像中大型车和小型车数量,并框选出检测到的车辆。交通量是公路交通信息的基础数据,根据交通量信息可以分析该路段的车流量、拥堵情况。从本系统得到的车型和数量可以作为现有交通量采集方法的补充,在未部署交通调查设备的路段作为数据参考。与现有采集方法比较,本文方法成本低、速度快、干扰因素少,有较好的应用价值。

### 4.2 试验结果对比与讨论

本文方法与其他方法对遥感图像中车辆检测的试验结果对比如表 1 所示。由表 1 可知,本文方法在遥感图像道路区域车辆检测上有着较好的效果。此方法的关注点主要在道路区域车辆,先采用自动提取方法截取道路区域,再通过卷积神经网络对车辆进行检测,有效地提高了识别的准确率和速度,在不同来源的遥感图像上都有着不错的效果,应用范围较广泛。本文结合大规模标注的数据集进行神经网络的训练,一定程度上解决了人工标注数据的繁琐和局限性。

表 1 不同方法的试验结果对比

Table 1 Comparison of experimental results with different methods

框架	本文方法	文献[5]	文献[13]	文献[14]
准确率/%	95	91	95.3	94.4
检测对象	遥感图像道路区域中的车辆	遥感图像中的车辆	遥感图像城区中的车辆	遥感图像道路区域中的车辆
方法	自动道路提取和卷积神经网络特征提取分类	人工特征提取	人工特征提取	手动裁切道路和人工特征提取
数据集	DOTA 数据集和谷歌地图遥感图像	CLIF(columbus large image format)数据集	谷歌地图遥感图像	谷歌地图遥感图像
存在问题	训练时间长	需要预先选定特征,不适用于大规模数据	需要预先选定特征,测试集较小,背景清晰,场景较单一	自动化程度不高,针对不同来源的遥感图像存在较大误差,测试集小



(a) 遥感图像上传界面



(b) 遥感图像处理过程界面



(c) 车辆检测结果界面

图 3 基于深度学习的遥感图像车辆检测系统界面实现图  
Fig. 3 Implementation of vehicle detection in remote sensing images based on deep learning

## 5 结 语

本文主要关注高分辨率遥感图像中道路区域车辆的识别,从而获得交通量、流量等信息,以帮助提高交通服务水平。首先采用速度较快的传统特征提取方法对道路区域进行提取,以减少其他区域的干扰;然后采用选择性搜索算法得到候选区域,将这些

候选区域输入到训练好的深度卷积神经网络中,实现对道路区域车辆目标的自动化提取。

下一步工作将尝试采用不同的图像分割算法,提高有树木遮挡和深色车辆的检测准确率;尝试更高效的特征算法和神经网络模型,以提高识别准确率和处理速度。

## 参 考 文 献

- [1] 赵英男,杨静宇. 基于 Gabor 滤波器和 SVM 分类器的红外车辆检测[J]. 计算机工程, 2005, 31(10): 191-192.
- [2] 王圣男,郁梅,蒋刚毅. 智能交通系统中基于视频图像处理的车辆检测与跟踪方法综述[J]. 计算机应用研究, 2005(9): 9-14.
- [3] 胡斌杰,詹益旺. 基于手机信令的道路交通流量状态识别及预测[J]. 移动通信, 2015, 39(10): 17-21.
- [4] EIKVIL L, AURDAL L, KOREN H. Classification-based vehicle detection in high-resolution satellite images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2009, 64(1): 65-72.
- [5] LIANG P, TEODORO G, LING H, et al. Multiple kernel learning for vehicle detection in wide area motion imagery[C]// International Conference on Information Fusion. IEEE, 2012: 1629-1636.
- [6] LIU W, YAMAZAKI F, VU T T. Automated vehicle extraction and speed determination from QuickBird satellite images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2011, 4(1): 75-82.
- [7] SUN H, SUN X, WANG H, et al. Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model[J]. IEEE Geoscience and Remote Sensing Letters, 2012, 9(1): 109-113.
- [8] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels compared to state-of-the-art superpixel methods [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282.
- [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] UIJLINGS J R R, SANDE KEAV D, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.
- [11] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[DB/OL]. (2015-4-10)[2018-04-05]. <https://arxiv.org/abs/1409.1556v6>.
- [12] XIA G S, BAI X, DING J, et al. DOTA: A Large-scale dataset for object detection in aerial images[DB/OL]. (2018-1-27) [2018-04-05]. <https://arxiv.org/abs/1711.10398v2>.
- [13] LEITLOFF J, HINZ S, STILLA U. Vehicle detection in very high resolution satellite images of city areas [J]. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48(7): 2795-2806.
- [14] 刘超超. 基于光学卫星影像的车辆识别和速度估算研究[D]. 北京:北京交通大学交通运输学院, 2015.

(责任编辑:徐惠华)