Journal of Computer Applications
计算机应用, 2020, 40(2): 553 - 560
ISSN 1001-9081
CODEN JYIIDU
2020-02-10
http://www.joca.cn

文章编号:1001-9081(2020)02-0553-08
DOI:10.11772/j.issn.1001-9081.2019101795

面向特定目标自识别的交通图像语义检索方法

赵　一[1,2],段　兴[1*],谢仕义[1],梁春林[1,2]
(1. 广东海洋大学 数学与计算机学院,广东 湛江 524000; 2. 湛江湾实验室 南海渔业大数据中心,广东 湛江 524000)
(*通信作者电子邮箱279153621@qq.com)

摘　要:为了从海量的道路交通图像中检索出违反交通法规的图像,提出了一种特定目标自识别的语义图像检索方法。首先,通过交通领域专家建立交通领域本体及道路交通规则描述;然后,通过卷积神经网络(CNN)对交通图像的特征进行提取,并结合改进的支持向量机决策树(SVM-DT)算法对图像特征进行分类的策略,对交通图像中的特定目标及目标间空间位置关系进行自动识别,并映射成为相应的本体实例及其对象之间的关联关系(规则实例);最后,利用本体实例和规则实例,通过推理得到语义检索结果。实验结果表明,相比关键字和本体交通图像语义检索方法,所提方法具有更高的准确率、召回率和检索效率。

关键词:交通领域本体;图像语义检索;语义推理;支持向量机决策树分类;目标识别
中图分类号:TP751　　**文献标志码:**A

Traffic image semantic retrieval method based on specific object self-recognition

ZHAO Yi[1,2], DUAN Xing[1*], XIE Shiyi[1], LIANG Chunlin[1,2]
(1. *College of Mathematics and Computer Science, Guangdong Ocean University, Zhanjiang Guangdong* 524000, *China*;
2. *Guangdong Zhanjiang Bay Laboratory, Fisheries Big Data Center of South China Sea, Zhanjiang Guangdong* 524000, *China*)

Abstract: In order to retrieve images of traffic violations from a large number of road traffic images, a semantic retrieval method based on specific object self-recognition was proposed. Firstly, road traffic domain ontology as well as road traffic rule description were established by experts in traffic domain. Secondly, traffic image features were extracted by Convolutional Neural Network (CNN), and combined with the strategy for classifying image features which is based on the proposed improved Support Vector Machine based Decision Tree (SVM-DT) algorithm, the specific objects and the spatial positional relationship between the objects in the traffic images were automatically recognized and mapped into the association relationship (rule instance) between the corresponding ontology instance and its objects. Finally, the image semantic retrieval result was obtained by reasoning based on ontology instances and rule instances. Experimental results show that the proposed method has higher accuracy, recall and retrieval efficiency compared to keyword and ontology traffic image semantic retrieval methods.

Key words: traffic ontology; image semantic retrieval; semantic reasoning; Support Vector Machine Decision Tree (SVM-DT) classification; object recognition

0　引言

截止2015年,北京市已经安装100多万个摄像头,对2 174多万人口及700多万车辆实施交通监控[1],交通监管部门需要从海量交通监控图像中检索出违反交通规则的图像。这些道路交通监控图像并没有描述文本,无法利用现有的文本检索方法,如果仅通过人工去识别图像所包含的语义,则费时费力,因此,交通图像语义检索工具的支持是非常必要的。

图像语义检索框架主要包括两大模块:语义分析模块和语义检索模块[2]。语义分析模块将图像中的底层特征抽取出来,并与高层的本体概念系统进行匹配。图像的底层可视特征与高层语义特征之间存在着一道鸿沟,在将可视化数据(visual data)导入语义分析模块进行检索之前,必须要将可视化数据转换成为语义检索模块能接受的本体实例及实例间关联。语义图像检索的性能极大地依赖于可视化数据特征的提取和描述。基于内容的图像语义检索面临14个语义鸿沟,包括图像内容、图像特征、图像语义等[3]。这些鸿沟将人类对图像的高层场景理解和底层的计算机像素分析分割开来,图像处理技术是弥补语义鸿沟的手段。关于语义检索模块,目前已有将颜色、位置、形状等底层次图像特征和图像中二维区域的空间位置关系特征,映射到本体概念及本体概念间关联的方法[4]。在抽取图像内容(visual content)方面,可以通过尺寸不变特征变换(Scale-Invariant Feature Transform, SIFT)技术来抽取视觉内容并转成为矢量模型,然后通过对矢量聚类得到视觉词汇(visual word),并映射到领域本体中,从而实现对图像的检索[5]。然而,对道路交通图像的语义检索主要关注

收稿日期:2019-08-30;**修回日期:**2019-10-24;**录用日期:**2019-10-24。
基金项目:南方海洋科学与工程广东省实验室自主立项重大项目(ZJW-2019-08);广东海洋大学创新强校重大科研项目(GDOU2017052501)。
作者简介:赵一(1984—),男,湖北武汉人,讲师,博士,CCF会员,主要研究方向:软件工程、人工智能; 段兴(1987—),男,湖南娄底人,助教,硕士,主要研究方向:软件工程、机器学习; 谢仕义(1963—),男,广东湛江人,教授,硕士,主要研究方向:软件工程; 梁春林(1975—),男,广东湛江人,副教授,硕士,主要研究方向:软件工程。

的是特定的交通目标，而非提取图像中所有目标，关注的也不是图像的纹理、颜色等底层特征。

基于领域本体搜索是利用语义检索模块和语义分析对图像内容标注，然后进行内容关系推理得出检索结果。本体经常用于信息系统中，作为一种有效的语义表征（semantic representation）[6]。在知识结构体中，本体也常用于图像检索系统，支持精确地发现、分类、检索以及标注图像[7]。当前，已有不少利用本体来辅助图像检索的相关工作。比如：本体结合物种分类知识图谱已用于动物图像的语义检索[8]；同时，本体结合情感表征也已用于艺术图像[9]及博物馆图像的语义检索。实验表明，通过本体，可以检索到一些语义相似的图像，从而提高图像检索的查准率和查全率。本体还被用于图像查询的扩展，提高图像检索的精度。目前，本体图像检索较新的研究方向为多模式本体（Multi-Modal Incompleteness Ontology，MMIO）模型，它既包含了某领域核心词汇，也包含了文本描述概念和视觉特征概念，通过多元信息融合提高检索精度[10]。

然而，已有方法在使用本体进行图像语义检索时，主要还是利用本体概念间继承关系进行的推理，如果图像目标之间存在更加复杂的语义关联，如在确定图像目标中的空间位置关系时，若仍然使用当前的语义图像检索方法，便无法自动检索到更加完整的图像与目标之间的空间位置关系。

针对这些问题，本文提出了一种改进目标自识别的交通图像语义检索方法。该方法能根据交通违规查询请求，自动识别出交通图像中的特定图像目标及目标间的空间位置关系，并通过语义推理得到违反道路交通的查询结果。本文的主要工作包括：

1）基于卷积神经网络（Convolutional Neural Network，CNN）抽取图像特征，然后利用模拟退火结合遗传算法（Simulated Annealing Genetic Algorithm，SA+GA）改进的支持向量机决策树（Support Vector Machine based Decision Tree，SVM-DT）算法对特征进行分类，最后利用生成的决策树对交通图像中的特定目标进行自动识别，并映射到相应领域本体中的实例。与其他分类器相比，通过改进算法得到的SVM-DT对交通目标进行自动识别时，具有更高的精确度。

2）在描述目标之间空间位置关系时，利用了基于方向关系矩阵模型和一阶逻辑（first-order logic）的本体规则进行建模和推理，使得逻辑推理能力更强，检索结果更加准确。

3）构建了原型系统，实现了本文所述的目标自识别的道路交通图像语义检索方法。

1 相关工作

近年来，图像的关系语义识别主要包括如下方面的研究：

1）基于神经网络[11]的图像语义识别。神经网络最初应用在图像的情感语义识别上，它利用底层特征映射到情感语义空间，在构建的二维图像情感因子空间，通过机器学习的BP神经网路实现了图像的底层特征的语义因子空间的映射。

2）基于特征融合[12]的图像语义识别。文献[12]采用了一种加权值的图像特征融合算法，并应用于图像情感语义识别，该方法根据不同特征对情感语义的影响提取颜色、纹理和形状特征后通过加权融合为新特征输入量，并用SVM来实现情感语义的识别。

3）基于本体的图像情感语义识别[13]。文献[13]针对人类在图像情感语义理解过程中存在的个性化和群体化现象，系统使用多层次的情感推理模型来分析图像情感语义。

4）基于Zernike矩阵的图像识别[14]。文献[14]针对图像的旋转、尺度和平移不变性识别问题，采用了一套旋转不变特征。使用正则几何矩阵图像的比例和平移参数不变特征，重新利用Zernike矩阵的正交性，简化了图像重建的过程，使图像的特征选择方法切实可行。

5）基于卷积网络的图像识别[15]。文献[15]采用了一个深入的评估网络精度的方法，构建了ConvNet模型，用于解决计算机视觉中深层视觉表象的研究。

6）基于非线性图像变形模型的图像识别[16]。文献[16]采用了经常出现在图像对象变化时存在相似的模块，并结合模型定位局部变化点，从而提高了MNIST标准的识别效果。

7）基于多方式局部集中的图像识别方法[17]。研究针对目标识别系统中集中的特征向量空间的局部不变的特征，提出了一种空间金字塔框架（spatial pyramid framework）来提高图像识别效率。

目前，利用分类模型对图像关系语义进行识别的研究，主要有基于Nasent（Neural analysis of sentiment）[18]改进的各种模型。首先，Nasent模型在深度学习的基础上开发了一种无监督情况下运行的模型，但是它针对图像的感知分析主要聚焦于模型本身，忽略了词序，并且只适合用于简单的例句，还远达不到人类的理解能力。

其次，利用Nasent模型对图像进行语义分析时，所描述的语义会随图像中各种参照物的不同而变化，Nasent模型虽然会为每个图像的描述构造语法树，但是在分析描述图片的语句时，根据的是关系树构造图像之间的语义联系。Nasent的准确率达到了85%，但遇到没有被统计的图像描述词汇或短语时，这个系统就会失效。

基于传统的SVM决策树多分类的图像目标语义识别方法是一种启发式搜索算法，因此，不能保证在任何情况下生成的决策树一定是最优。而本文所提出的SVM-DT是基于测度的一对多比较，能确保每次比较都是最优，而且能够按照测度最优值分配决策树划分结构。通过对传统的SVM一对多（one-versus-rest）分析与比较：我们发现传统的一对多分类算法训练时依次把某个类别的样本归为一类，其他剩余的样本归为另一类，这样k个类别的样本就构造出了k个SVM。其优点是训练k个分类器，个数较少，分类速度相对较快；缺点是每个分类器的训练都是将全部的样本作为训练样本，这样在求解二次规划问题时，训练速度会随着训练样本的数量的增加而急剧减慢。

本文改进的SVM-DT分类因为采用了分离性测度，所以其分类判断的最坏情况等于算法所需的SVM个数，则需要训练SVM个数为$k-1$；且每一次利用测量函数后，就会得到一个最优解，即：两个分类所需要遍历SVM个数小于等于$k(k-1)(k+2)/2(k+1)$。

本文关注的图像关系语义挖掘与已有图像语义识别方法关注点不同，在于详细分析图像中各实例的位置关系，并利用SWRL规则准确判定语义规则。且已有的算法研究并没有充分考虑到图像的实例关系，而是统一将整幅图像进行分析，这将不能对图像中各个实例的关系进行解释，无法实现像人脑那样对一幅图像进行深层次的理解。本文利用相应的工具对图像进行实例化标注，以有效地挖掘出视频图像中实例的位置关系，能较完整地解决机器自动判断道路上机动车、非机动

车及行人违法的问题。

2　基于目标自识别的交通图像语义检索

道路交通监管部门需要从海量的交通图像中检索出违反交通规则的图像,如果用人工对海量交通图像进行甄别,则十分费时费力。根据道路交通监管部门的查询请求,本文方法首先利用机器学习方法识别出特定目标,将其标注成为交通领域本体中的实例,继而分析得到特定目标之间的空间位置关系;然后,通过语义推理技术,判断自动标注出来的本体实例之间是否满足交通规则的约束;最后,将违反交通规则的图像返回给交通监管部门。本文方法可实现完全自动化,为交通监管部门节省大量人力,且提供更加直观的执法依据。本文提出的目标自识别的交通图像语义检索方法包括三个主要步骤:

1) 道路交通领域知识构建。领域专家构建道路交通领域本体,包括概念、实例、继承、实例化、属性关联等,并构建基于一阶逻辑的道路交通规则公式。

2) 图像目标自动挖掘。从图像数据中自动识别出查询请求的特定目标,并得到目标之间的空间位置二元关系。

3) 基于领域本体推理的图像检索。根据步骤1)建立的领域知识和步骤2)中自动挖掘出的特定目标及其关联进行逻辑推理,得到图像检索结果。

方法总体框架如图1所示。

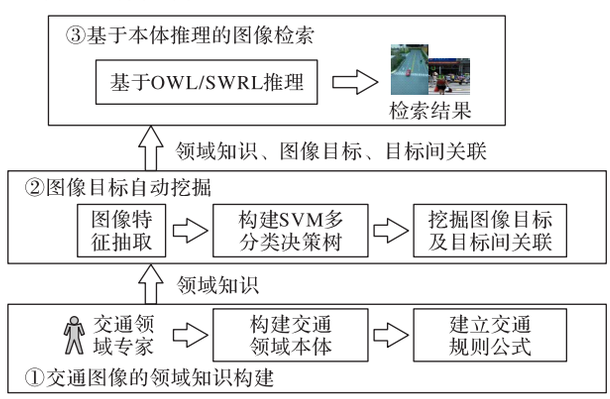


图1　基于领域知识的交通图像语义检索框架

Fig. 1　Traffic image semantic retrieval framework based on domain knowledge

2.1　路交通领域知识的构建

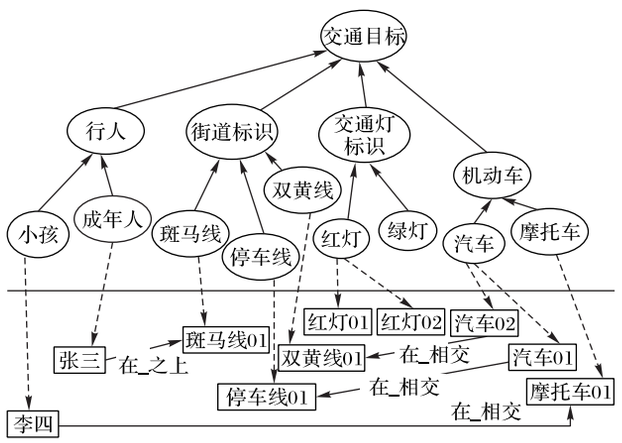
道路交通领域知识的构建,包括构建道路交通领域层次概念结构的本体以及道路交通规则公式。通过图像目标自动识别算法,识别出的目标映射为交通领域本体的实例,图像目标之间的空间位置关系对应于交通领域本体中实例间的对象关联属性。

道路交通领域本体提供了道路交通领域词汇和背景知识,如:行人、交通标识及各种机动车等。本体中的类(Class)和关联关系(Relationship)的定义以及交通领域法规,来自于《中华人民共和国道路安全交通法》(http://www.bjjtgl.gov.cn/jgj/fl/205308/index.html)。该法规提供了我国道路交通安全领域标准术语及其基本的道路安全交通法规。道路交通法规例子如表1所示。本文使用语义网规则语言(Semantic Web Rule Language, SWRL)来描述交通法规。图2示例了道路交通领域知识,包括道路交通领域本体和SWRL描述的道路交通法规公式,其中:椭圆形表示本体概念,矩形表示本体实例,本体实例间有对象属性关联相连。

表1　道路交通领域法规

Tab. 1　Rules in the domain of road traffic

法规#	法规描述
Rule1	当斑马线前方红灯亮起时,如果行人还处在斑马线上,则判定行人违规
Rule2	当机动车道前方红灯亮起时,如果机动车压在停车线上,则判定机动车违规
Rule3	非机动车驶入机动车道行驶,则判定非机动车违规
Rule4	驾驶摩托车时后座乘坐不满12岁小孩,则判定摩托车违规
Rule5	机动车压在双黄线上行驶,则判定机动车违规



R1 红灯(?r)∧行人(?x)∧斑马线(?z)∧在上面(?x,?z)=>违规(?x)
R2 红灯(?r)∧机动车(?a)∧停车线(?s)∧在上面(?a,?s)=>违规(?a)
R3 非机动车(?n)∧机动车道(?r)∧在上面(?n,?r)=>违规(?a)
R4 摩托车(?m)∧小孩(?c)∧在上面(?m,?c)=>违规(?m)
R5 机动车(?a)∧双黄线(?y)在上面(?a,?y)=>违规(?a)

图2　道路交通领域知识

Fig. 2　Road traffic domain knowledge

2.2　图像目标自动识别

图像目标自动识别的目的,是从图像的可视化数据(Visual data)中识别出用户查询请求中规定的特定交通目标,如:机动车、非机动车、行人或斑马线等。交通图像中目标识别是一个典型的高维数、小样本问题。文献[20]卷积神经网络-支持向量机(Convolutional Neural Networks- Support Vector Machine, CNN-SVM)在解决高维数、小样本识别问题中表现出特有的优势。文献[21]基于训练样本分布定义了类分离度量,并引入到SVM-DT构成的过程中,得到了一种改进的SVM-DT训练算法,再将该方法用于雷达目标的高分辨一维距离图像的目标识别中。文献[22]以最大分类间隔为准则,利用遗传算法对传统的SVM-DT进行优化,提出了一种最优(或近优)决策二叉树训练算法,且实验证明了该方法有更高的分类精度;但是遗传算法存在"过早收敛"的问题,导致迅速收敛的结果未必是全局最优。本节提出了一种基于模拟退火遗传算法的SVM-DT多分类策略,使得收敛结果能更好地接近全局最优解,具体如下所述:

为了自动从图像中识别出特定目标,首先,进行图像预处理,将图像转换成为特征向量。本文选择CNN卷积神经网络对待识别交通目标作为特征提取,然后全链接层得到图像特征通过一种改进的SVM-DT构建算法训练出最(近)优决策树进行分类。最后获取了决策二叉树对图像进行测试,挖掘出

被测试图像中的特定交通目标。

2.3 利用CNN提取交通图像特征

本文所使用的CNN提取交通图像特征借鉴了经典的R-CNN算法[19]，在此基础上改进了SVM分类器，提高了分类精度。图像特征抽取是通过CNN局部连接网络的特性，将图像分成小的连通区域，然后根据自然图像的统计特性，某个区域的权值也可用于另一区域（权值共享性），权值共享表征为卷积核共享，对于一个卷积核将其与给定的图像做卷积，就可得到一种图像的特征，不同的卷积可以提取不同的图像特征。具体流程为：

1）输入一张图片，通过生成候选窗口（selective search）算法定位2 000个物体候选框（bounding box），这些框中有本文需要的物体特征。

2）非极大值抑制（Intersection Over Union，IOU）。由于选定框是矩形，而且大小各不相同，而CNN要求输入图片的大小是固定值，所以必须对矩形候选框作缩放处理，本文选取的是各向同性缩放处理。接着对候选框进行非极大值抑制处理，目的是为了对候选框与真实框（ground truth）进行回归，校正微调提取框（region proposal）的大小，以提高最终的样本检测精度。

3）CNN特征提取。本文选取的神经网络架构是经典的Alexnet，其特点是卷积核比较小、跨步小、精度较高。Alexnet特征提取部分包含了5个卷积层、2个全连接层，其每层神经元个数都是4 096，保证最后提出每个候选框图片都有4 096维特征向量。首先进行网络有监督的预训练阶段，得到一个初始模型；然后对初始模型进行fine-runing训练，其目的是针对特定的任务来缩小训练的数据集，以提高SVM训练精度；最后假设要检测的物体类别有N类，将预训练阶段的CNN模型的最后一层替换成$N+1$个输出的神经元（“+1”表示还有一个背景），直接采用参数随机初始化的方法，其他网络层的参数不变，开始随机梯度下降（Stochastic Gradient Descent，SGD）训练。开始时，SGD学习率选择0.001，在每次训练的时候，batchsize大小选择128，其中32个是正样本，96个是负样本。

4）把CNN提取的特征输入到本文改进的SVM中进行分类，如果这个特征向量feature vector所对应的候选框是需要的物体则分为同一类，否则分为其他类别。

CNN获取图像实例特征的算法流程如下：

输入　样本图像数据；理想输出；

输出　实际输出：

```
1)  取得卷积核个数：ConvolutionLayer::f_prop(input, output)
2)  For(int i = 0; i < n; i++)
3)    用第i个卷积核和输入层第a个特征映射做卷积
4)    convolution = Conv(input[a], kernel[i]);
5)    把卷积结果求和：sum[b] += convolution;
6)  End For
7)  for (i = 0; i <(int)bias.size(); i++)
8)    加上偏移量：sum[i] += bias[i];
9)    调用Sigmoid函数：output = Sigmoid(sum);
10)  梯度通过DSigmoid反传
11)  sum_dx = DSigmoid(out_dx);
12)  计算bias和coeff的梯度      //coeff是回归系数，bias是偏置
13)  coeff_dx[i] += sub[j][k]*sum_dx[i][j][k];
14)  bias_dx[i] += sum_dx[i][j][k]);
15)  调整权矩阵
16)  End For
```

2.4 基于模拟退火遗传算法改进的SVM-DT多分类策略

采用基于模拟退火遗传算法改进的SVM-DT分类方法生成决策二叉树，并对CNN传入的图像特征进行改进的SVM-DT分类，这样做的目的是为了解决CNN的Softmax层训练时，正负样本阈值无法调整的问题。因为理想状态是：当候选框把整辆车都包含在内，称为正样本；候选框没有包含到车辆，就称为负样本。如果遇到候选框部分包含车辆的情况，就可以通过调整IOU阈值来获取最好的分类效果。本文通过实验发现，设定IOU为0.4划分汽车效果最好。通过设定的值训练出来的SVM分类器能够比Softmax层分类的结果更准确。

本文的目标是在每个决策节点将原始多类训练样本集划分为两类，并且使分类间隔最大，所以选择SVM分类算法的分类间隔作为模拟退火遗传算法适应度函数。引入模拟退火算法的控制参数T来控制变异时子结点染色体替换父节点染色体的概率，优化了遗传算法中生存策略，从而能找到接近全局最优解，得到一棵分类间隔最大的决策二叉树。

线性分类器的学习目标便是要在n维的数据空间中找到一个超平面：$H(x_i, y_i), i = 1, 2, \cdots, n, \boldsymbol{x} \in \mathbf{R}^n, y \in (+1, -1), i$为样本数，$\mathbf{R}^n$为输入维度，若是线性可分情况，将两样本完全分开的超平面为H：

$$\boldsymbol{w}^H \boldsymbol{x} + \boldsymbol{b} = 0 \tag{1}$$

若使分类间隔最大的超平面为最优分类面，则

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2$$

$$\text{s.t.} \quad y_i(\boldsymbol{w}^H x_i + \boldsymbol{b}) \geqslant 1;\ i = 1, 2..., n \tag{2}$$

但不是所有的数据集都能线性可分，所以引入松弛变量$\xi_i (i = 1, 2, \cdots, n)$：

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n} \xi_i \tag{3}$$

$$\text{s.t.} \quad y_i(\boldsymbol{w}^H x_i + \boldsymbol{b}) \geqslant 1 - \xi_i, \xi_i \geqslant 0;\ i = 1, 2, ..., n$$

其中：C为惩罚系数，表示对分错的点加入惩罚。C越大，错分点更少，但是过拟合的情况可能会更严重；当C很小时，分错点会越多，所以得到的模型也会不正确，因此们引入拉格朗日乘子，用条件极值求解最优分界面。

$$f(\alpha) = L(\boldsymbol{w}, \boldsymbol{b}, \alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^H x_j \tag{4}$$

根据SVM理论，两类样本的最大分类间隔是：

$$margin = 2 / \|\boldsymbol{w}^H\|$$

其中：

$$\boldsymbol{w}^H = \sum_{i=1}^{n} \alpha_i^H y_i x_i \tag{5}$$

本文提出的基于模拟退火遗传算法改进的SVM-DT多分类策略的步骤描述如下：

步骤1　将全部训练样本集所属类别按实值编码策略进行编码，决策树根节点的染色体的编码为$\{1, 2, \cdots, N\}$，其中$N \geqslant 3$为原样本训练集的类别总数，染色体中每个基因对应原训练样本集的类别编号；并在根节点调用GA将原始训练样本所属类别划分为两类。

步骤2　判断各子节点是否只包含一类样本，若是转步

骤4,反之转步骤3。

步骤3　设新产生的适应性(flexibility)函数为$f(a_i)$, a_i为个体。变动的阈值为$\bar{f}(a_i)$,当$f(a_i)>\bar{f}(a_i)$时,接受新个体;否则,以一定概率接受新个体$P=\exp((f(a_i)-\bar{f}(a_i))/T)$($T$是控制参数,即模拟退火中的热力学温度);从群中选择n对个体,作为父类,对每一父类,由父类p_1、p_2使用交叉、突变算子生成子代a_1、a_2,计算a_1、a_2的适应性。

步骤4　设群平均适应性为$f(a_{\text{avg}})$,最低适应性为$f(a_{\text{weakest}})$,则$\bar{f}(a)=f(a_{\text{avg}})-f(a_{\text{weakest}})$。对于每个新产生的个体$f(a_i)>\overline{f(a)}$,则在群随机选择一个适应性低于$\bar{f}(a)$的个体替换;否则,以概率$p=\exp((f-\bar{f})/T)$替换其父样本。

步骤5　结束循环,生成接近全局最优决策树。

最后,使用上述步骤生成的SVM-DT对道路交通图像进行测试,识别出图像中相应交通目标,并返回目标对应的视觉词汇。

2.5　基于本体规则推理的图像检索

本节将阐述如何进一步识别出交通目标之间的空间位置关系,并映射到领域本体实例之间的对象属性关联关系上;最后,基于道路交通领域知识进行推理,得出结论。

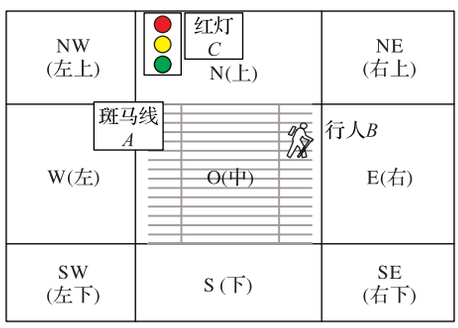
在图像中识别出交通目标实例后,应用Python工具中matplotlib库(http://matplotlib.org/)给出目标实例的边界区域,然后,调用minSize函数给出了每个目标实例活动窗口的4个参数,分别为:四边形X轴、Y轴位置、宽度W和高度h。通过以上4个参数调用rectangle函数得到该目标实例的活动窗口。与此同时,调用minNeighbors函数找到当前目标实例邻近的其他目标实例,并同理得到其他目标实例的活动窗口。

窗口位置关系判断算法:假设图像中两个目标实例的活动窗口为四边形R_1、R_2,分别对应于交通领域本体中的本体实例A、B,且A和B满足方向位置矩阵模型时,则返回关联关系。具体例子根据图3(a)所示,行人B相对于参考目标斑马线A的方向关系:$Dir(A,B)=\{\text{O},\text{E}\}$。若是交通灯为红灯时,行人$B$与斑马线$A$的位置关系只有满足$Dir(A,B)=\{\text{NW}\}$;$Dir(A,B)=\{\text{N}\}$;$Dir(A,B)=\{\text{NE}\}$;$Dir(A,B)=\{\text{SW}\}$;$Dir(A,B)=\{\text{S}\}$;$Dir(A,B)=\{\text{SE}\}$时,表示行人在斑马线的两边站立或行走,没有闯红灯违规现象,其余的关系都判定为违规。

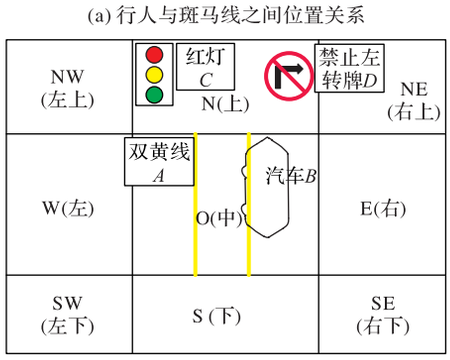
图4描述的是道路交通图像“行人闯红灯”中目标和目标间关系与图2所示的道路交通领域本体的映射。如图4左方所示,“红灯01”“张三”和“斑马线01”是该图像中识别出来的目标实例,这些目标分别映射到图4右方所示道路交通领域本体中的实例“红灯01”“张三”和“斑马线01”。目标实例“张三”和“斑马线01”对应的活动窗口的重合区域可映射到道路交通领域本体中实例“张三”和“斑马线01”之间的对象属性关联“相交”。

最后,将映射到道路交通领域本体中的实例及实例间关联代入道路交通法规公式中,如果每项原子公式都为真,则推出违反道路交通法规的结论。如图4下方道路交通法规公式所示,可推出成年人“张三”违反交通法规,并根据行人、红灯、方向位置矩阵计算结果,输入SWRL三元组规则编辑器中进行推导,得出推导后的最终结果,然后查询表1中列出的规则关系,最终结论“行人违规闯红灯”。该图像满足检索条件。

如图5所示,汽车与双黄线之间的关系为$Dir(A,B)=\{\text{O}\}$表示在图5中,机动车辆违反了交通规则Rule5,机动车压在双黄线上行驶,则判定机动车违规。汽车实例与双黄线实例的位置关系示意图如3(b)所示,可以得知汽车B与双黄线A之间的关系为典型的相交例子,这种关系就可以判断出汽车是否压了双黄线,为以后的图像搜索提供了依据。



(a) 行人与斑马线之间位置关系



(b) 汽车与双实线之间位置关系

图3　交通图像实例的位置关系

Fig. 3　Local relationship of traffic graph instance

表2　图像目标实例与本体关联属性的映射

Tab. 2　Mapping of image object instances and ontology-related attributes

序号	图像目标实例	本体实例间关联属性
1	R_1处于R_2上面	在上面(?A,?B)
2	R_1处于R_2下面	在下面(?A,?B)
3	R_1处于R_2左边	在左边(?A,?B)
4	R_1处于R_2右边	在右边(?A,?B)
5	R_1处于R_2中	相交(?A,?B)
6	R_1与R_2平行	平行(?A,?B)
7	R_1处于R_2左上	在左上(?A,?B)
8	R_1处于R_2右上	在右上(?A,?B)
9	R_1处于R_2左下	在左下(?A,?B)
10	R_1处于R_2右下	在右下(?A,?B)

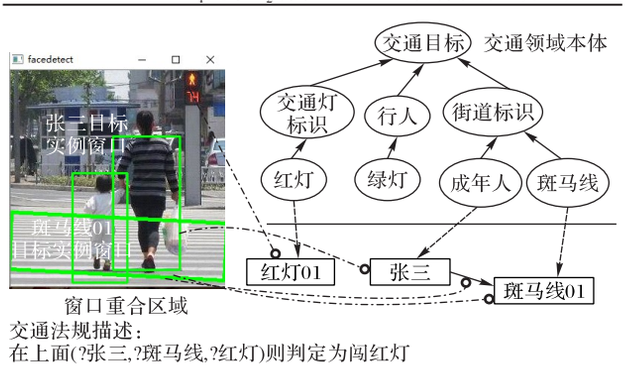


图4　图像目标及其关联与领域本体的映射

Fig. 4　Mapping of image objects and their associations with domain ontologies

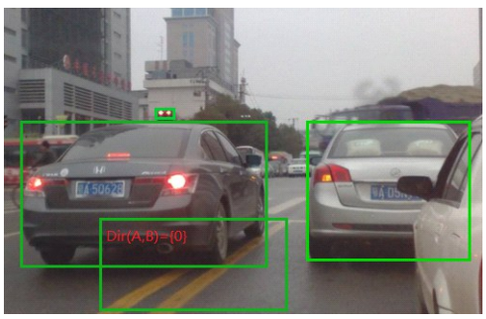


图5 汽车与双实线之间位置关系实例

Fig. 5 Example of positional relationship between cars and double solid lines

3 工具实现

基于前述理论和方法，设计并实现了目标自识别的图像语义检索原型系统，该系统由以下几个模块组成（如图6所示）：

1）领域知识构建模块：道路交通领域专家可通过该模块提供的可视化本体建模工具对交通领域本体和交通规则公式建模，建立OWL描述的道路交通领域本体和SWRL描述的交通规则公式，并存储在SPARQL数据库中。

2）图像目标识别模块：该模块基于模拟退火遗传SVM-DT多分类策略识别图像中特定交通目标。该模块的输入是待识别的交通图像及特征训练集；输出是目标实例（如：行人、斑马线、交通信号灯）已标注的道路交通图像。

3）目标间空间位置关联识别模块：该模块通过图像中目标对应活动窗口挖掘出目标对应的空间位置关系。

4）Hermit推理机模块：该模块基于交通领域知识，结合识别出的图像目标实例和目标间关联关系进行推理。它的输入是道路交通领域本体、道路交通规则公式、图像中的目标实例和目标间关联关系，输出是该图像的检索结果。

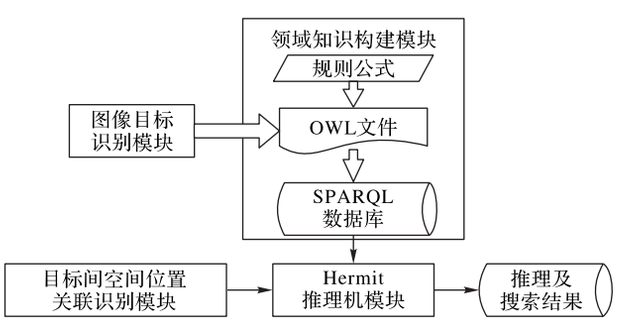


图6 目标自识别的交通图像语义检索工具框架

Fig. 6 Traffic image semantic retrieval tool framework based on target self-recognition

在自动识别出图像中的交通目标后，结合领域专家建立的道路交通领域知识，系统将读取标注并对应生成类和实例，按着点击“规则”按键进入规则推导。系统通过载入语义标注的图像，自动生成由所识别特定目标的本体实例和领域专家所构建的交通领域本体所结合而成的本体OWL文件。在读取该文件后，使用Hermit推理机对实例之间的关系进行分析，得到最终的判定结果，从而智能地判断图像有无涉及违反交通规则的内容。

4 实验分析

实验分析采用三种数据集的图像数据：

1）CVC数据（http://www.cvc.uab.es/adas/site/）的CVC-02-Classification数据集。

2）google交通图库。

3）ImageNet交通类数据集。它是一个拥有超过1 500万张带标签的高分辨率图像的数据集，这些图像分属于大概22 000个类别。这些图像是从网上收集的，并使用Amazon Mechanical Turk众包方式进行人工贴标签。本文选择其中与道路交通相关的图像，共有8 300个训练样本，6 640张测试样本。第一类图像是二分类（其中训练样本1 900张，测试样本1 520张），有行人和交通信号灯；第二类图像是三分类（其中训练样本为1 700张，测试样本为1 360张），有行人、交通信号灯和斑马线；第三类图像是四分类（其中训练样本为2 200张，测试样本为1 760张），有行人、交通信号灯、斑马线和汽车；第四类图像是五分类（其中训练样本为2 500张，测试样本为2 000张），有行人、交通信号灯、斑马线、汽车和自行车。

4.1 改进的SVM-DT目标识别性能评估

本节通过实验对SVM的1对1分类（记作1-a-1）、1对多分类（记作1-a-r）、基于遗传算法的SVM-DT（记作GADT）多分类及本文提出的SVM-DT方法进行对比。当测试图像数量分别为1 000、2 000、3 000、4 000时，比较关键字搜索方法、本体搜索方法和本文所提的SWRL+本体搜索方法在图像搜索的准确率、召回率等两个维度上性能进行分析与比较，实验结果如图7所示。

由图7（a）可见，本文方法的查询准确率要高于其他方法，准确率相对于关键字搜索提高了约19个百分点；相对于MMIO本体搜索提高了约12个百分点，这是因为本文“SWRL+本体”推理不仅能就视觉词汇的上下位关系推理，还能结合空间位置关系的描述进行规则推理。而MMIO等本体检索方法没有使用规则判断，因此在准确率上比本文采用方法低。如图7（b）所示，在召回率上，本文方法相较于关键字搜索提高了约3个百分点，与MMIO等本体检索算法基本相同。然而，可以从图7（b）得知，当测试图像数量在1 000到3 000时，本文方法都是略优于本体搜索的，但是当样本数变为4 000时，本文方法检索的召回率略低于本体搜索，其原因是新加入的样本中含有使用SWRL规则较难判断的图像，如汽车红灯时只是压线，但并没有越过线，在标准集中判断为没有违规，而使用本文方法判定为违规，所以检索结果中把正样本预测为负类的数量变多，从而导致召回率偏低。

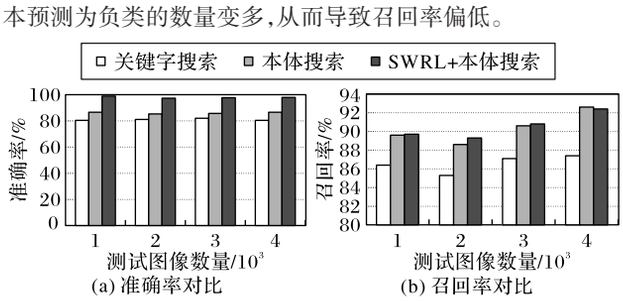


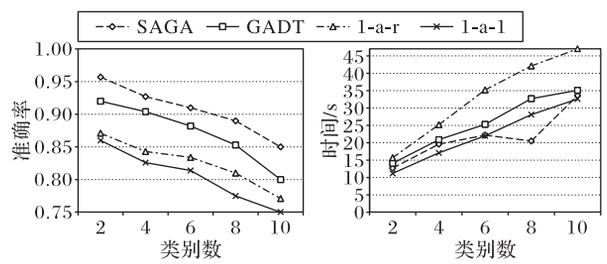
图7 不同方法搜索交通违规图像的结果

Fig. 7 Results of different methods for searching traffic violation images

目标实例的分类准确率如图8（a）所示。可以看出，随着

分类类别数的增加,所有SVM多分类方法的分类精度都呈下降趋势。基于模拟退火遗传算法改进的SVM最优决策树在2分类、3分类、4分类时,分类精度高于经典的1-a-1、1-a-r方法,略高于GADT方法的分类精度;在5分类时分类精度明显高于其他的三种方法。实验证明本文方法随着分类类别增多精度增高。

分类耗时如图8(b)所示:1-a-r由于每次训练都需要所有的样本参与,故其训练时间最长;1-a-1虽训练复杂,但每次训练只需要两类样本参与,相较于前者耗时最短,且与类别的变化关系不大;本文方法训练时间仅次于1-a-1,略快于GADT。



(a) 不同类别数对应的分类精度　(b) 不同类别数对应的分类耗时

图8　各算法在不同指标上对比结果

Fig. 8　Result comparison of each algorithm on different indicators

4.2　基于本体规则推理实验的违规实例

从四个规则搜索的返回结果图9可以看出,本文方法返回结果的图像都是违规的,说明准确率较高,因此该方法可以应用于司法部门处理交通违法、刑事侦查、司法调查等,并能为其提供准确和可靠的执法依据。

(a) 搜索Rule1行人在斑马线上结果(例)　(b) 搜索Rule2机动车闯红灯违规结果(例)

(c) 搜索Rule3非机动车驶入机动车道违规结果(例)　(d) 搜索Rule5机动车在双黄线上行驶违规结果(例)

图9　四个规则搜索的返回结果示例

Fig. 9　Results of retrieval by four rules

5　结语

本文针对道路交通领域,提出了一种基于目标自识别的图像语义检索方法。首先,建立道路交通领域知识;然后,通过模拟退火遗传算法训练出SVM-DT,对道路交通图像中的特定目标进行识别,并映射为领域本体实例;再进一步识别出特定目标之间的空间位置关系,并映射为领域本体实例间的对象属性关联关系;最后,利用规则推理判断图像是否满足查询条件。实验结果表明,当处理多目标时,因为本文使用了改进的空间位置识别算法,所以可以准确地检测出多目标在复杂空间中的位置关系,通过运用本文的方法进行语义图像检索,在图像目标自识别和语义推理两阶段的性能均有所提升,并能得到更加精确的检索结果。未来工作包括尝试其他更加高效的机器学习的方法来自动识别图像目标,以及通过时序逻辑算子描述更加复杂的道路交通语义场景。

参考文献 (References)

[1] 郭雪梅. 中国工程院院士高文:城市多媒体大数据高效存储与处理技术[EB/OL]. [2019-02-20]. http://www.csdn.net/article/2015-06-05/2824873. (GUO X M. Academician of Chinese Academy of Engineering GAO Wen: high efficient storage and processing technology of urban multimedia data[EB/OL]. [2019-02-20]. http://www.csdn.net/article/2015-06-05/2824873.

[2] 高隽,谢昭,张骏,等. 图像语义分析与理解综述[J]. 模式识别与人工智能, 2010, 23(2):191-202. (GAO J, XIE Z, ZHANG J. et al. Image semantic analysis and understanding: a review[J]. Pattern Recognition and Artificial Intelligence, 2010, 23(2): 191-202.)

[3] DESEMO T M, ANTANI S, LONG R. Ontology of gaps in content-based image retrieval[J]. Journal of Digital Imaging, 2009, 22(2): 202-215

[4] SCHOBER J P, HERMES T, HERZOG O. Content-based image retrieval by ontology-based object recognition[C/OL]// Ki-2004 Workshop on Applications of Description Logics. [2019-02-20]. http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-115/01-schober.pdf.

[5] ZHOU W, LI H, SUN J, et al. Collaborative index embedding for image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(5):1154-1166.

[6] MEERSMAN R A. Semantic ontology tools in IS design[C]// Proceedings of the 1999 International Symposium on Methodologies for Intelligent Systems, LNCS 1609. Berlin: Springer, 1999: 30-45.

[7] KANIMOZHI T, CHRISTY A. Applications of ontology and semantic Web in image retrieval and research issues[J]. International Journal of Computer Science and Information Technologies, 2014, 5(1): 763-769

[8] SCHREIBER A T, DUBBELDAM B, WIELEMAKER J, et al. Ontology-based photo annotation[J]. IEEE Intelligent Systems, 2001, 16(3):66-74.

[9] HOLLINK L, SCHREIBER G, WIELEMAKER J, et al. Semantic annotation of image collections[EB/OL]. [2019-02-20]. https://www.math.vu.nl/~laurah/1/papers/Hollink03_saic.pdf.

[10] POSLAD S, KESORN K. A Multi-Modal Incompleteness Ontology model (MMIO) to enhance information fusion for image retrieval[J]. Information Fusion, 2014, 20:225-241

[11] HAUBOLD A, NATSEV A, NAPHADE M R. Semantic multimedia retrieval using lexical query expansion and model-based reranking[C]// Proceedings of the 2006 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE, 2006: 1761-1764.

[12] DASIOPOULOU S, MEZARIS V, KOMPATSIARIS I, et al. An ontology-based framework for semantic image analysis and retrieval[M]// Semantic-Based Visual Information Retrieval. Hershey, PA: IRMA Press. 2007: 208-246.

[13] KESORN K, POSLAD S. An enhanced bag-of-visual word vector space model to represent visual content in athletics images [J]. IEEE Transactions on Multimedia, 2012, 14(1): 211-222.

[14] NATSEV A, HAUBOLD A, TEŠIĆ J, et al. Semantic concept-based query expansion and re-ranking for multimedia retrieval [C]// Proceedings of the 15th ACM International Conference on Multimedia. New York: ACM, 2007: 991-1000.

[15] 万华林, CHOWDHURY M U. 基于支持向量机的图像语义分类[J]. 软件学报, 2003, 14(11): 1891-1899. (WAN H L, CHOWDHURY M U. Image semantic classification by using SVM [J]. Journal of Software, 2003, 14(11):1891-1899.)

[16] DENG L. The MNIST database of handwritten digit images for machine learning research [best of the Web] [J]. IEEE Signal Processing Magazine, 2012, 29(6): 141-142.

[17] MEZARIS V, KOMPATSIARIS I, STRINTZIS M G. An ontology approach to object-based image retrieval [C]// Proceedings of the 2003 International Conference on Image Processing. Piscataway: IEEE, 2003, 2: 511-514.

[18] AL-SMADI M, TALAFHA B, AL-AYYOUB M, et al. Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews [J]. International Journal of Machine Learning and Cybernetics, 2019, 10(8):2163-2175.

[19] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[20] KANG J, PARK Y-J, LEE J. Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems [J]. IEEE Transactions on Industrial Electronics, 2018, 65(5): 4279-4289.

[21] WANG X, WU C, QI Y, et al. HRRP classification by using improved SVM decision tree [C]// Proceedings of the 6th World Congress on Intelligent Control and Automation. Piscataway: IEEE, 2006: 10096-10100

[22] 连可,黄建国,王厚军,等. 一种基于遗传算法的SVM决策树多分类策略研究[J]. 电子学报, 2008, 36(8):1502-1507. (LIAN K, HUANG J G, WANG H J, et al. Study on a GA-based SVM decision-tree multi-classification strategy [J]. Acta Electronica Sinica, 2008, 36(8): 1502-1507.)

This work is partially supported by the Guangdong Laboratory Autonomous Set Up Major Project for Southern Marine Science and Engineering (ZJW-2019-08), the Major Scientific Research Projects of Innovative and Strong University of Guangdong Ocean University (GDOU2017052501).

ZHAO Yi, born in 1984, Ph. D., lecturer. His research interests include software engineering, artificial intelligence.

DUAN Xing, born in 1987, M. S., teaching assistant. His research interests include software engineering, machine learning.

XIE Shiyi, born in 1963, M. S., professor. His research interests include software engineering.

LIANG Chunlin, born in 1975, M. S., associate professor. His research interests include software engineering.