

# 基于深度属性学习的交通标志检测

王方石<sup>1</sup>, 王 坚<sup>1</sup>, 李 兵<sup>2,3</sup>, 王 博<sup>2,3</sup>

(1. 北京交通大学 软件学院, 北京 100044; 2. 中国科学院 自动化研究所, 北京 100190; 3. 中国科学院 模式识别国家重点实验室, 北京 100190)

**摘 要:** 为了弥补交通标志底层图像到高层语义之间的鸿沟, 本文引入交通标志的形状、颜色、图案内容三种视觉属性, 在卷积神经网络(Convolutional neural network, CNN)中加入属性学习(Attribute learning)约束, 同时进行交通标志属性学习和分类学习, 提出了一种基于深度属性学习的交通标志检测方法。并在公开数据集 Sweden traffic sign detection dataset (STSD)和 German traffic sign detection dataset(GTSD)上进行的实验结果表明, 该方法能够有效地提高交通标志检测的准确率和召回率。

**关键词:** 信息处理技术; 交通标志检测; 深度属性学习; 卷积神经网络

**中图分类号:** TN911.73 **文献标志码:** A **文章编号:** 1671-5497(2018)01-0319-11

**DOI:** 10.13229/j.cnki.jdxbgxb20161120

## Deep attribute learning based traffic sign detection

WANG Fang-shi<sup>1</sup>, WANG Jian<sup>1</sup>, LI Bing<sup>2,3</sup>, WANG Bo<sup>2,3</sup>

(1. School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China; 2. Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China; 3. National Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** A traffic sign detection method based on deep attribute learning was proposed. To make up the gap between raw image and high level semantics, three visual attributes, including shape, color and pattern, were introduced. Attribute learning was added to train Convolutional Neural Network (CNN), where attribute learning and classification learning were carried out simultaneously. Experimental results on datasets Sweden Traffic Sign Detection Dataset (STSD) and German Traffic Sign Detection Dataset (GTSD) show that the proposed method can effectively improve the precision and recall in terms of traffic sign detection.

**Key words:** information processing technology; traffic sign detection; deep attribute learning; convolutional neural network(CNN)

## 0 引 言

近年来,高级驾驶辅助系统(Advanced driver

assistance systems)和无人驾驶技术发展迅速,其中,交通标志的检测和识别是其重要组成部分<sup>[1]</sup>。常见的交通标志有速度限制、减速、停止、人行横

收稿日期:2016-10-14.

基金项目:国家自然科学基金项目(61370038, 61472421).

作者简介:王方石(1969-),女,教授,博士. 研究方向:视觉信息处理. E-mail: fshwang@bjtu.edu.cn

通信作者:李兵(1983-),男,副研究员,博士. 研究方向:模式识别. E-mail: bli@nlpr.ia.ac.cn

道和车道限制等。无人驾驶车辆需要通过检测交通标志以获取前方道路状况信息,才能减少事故发生的可能性,增加驾驶的安全性。在现实环境中,交通标志的检测是一个具有挑战性的任务。人类驾驶员可以很容易地找出并识别交通标志,是因为交通标志的设计使用了颜色、形状和文本等语义性强、适合人类理解的元素的组合。但对于机器而言,由于交通标志的种类多、类间区分度不大,各个国家地区的交通标志不尽相同,以及因光照、模糊、部分遮挡和背景复杂等因素,使得难以从复杂场景中定位、准确识别交通标志区域<sup>[2-5]</sup>。

本文利用交通标志中不受客观条件影响的共同特性,设计并实现一个简单、通用和准确的交通标志的检测方法,对于提高辅助驾驶和无人驾驶安全性具有重要意义。在充分挖掘交通标志牌表面特性和考虑交通标志图案视觉属性的基础上,本文提出了一种在 HSV 颜色空间上提取最大稳定极值区域(MSER)以提取交通标志候选区域<sup>[1,4-8]</sup>、利用加入了属性学习约束的卷积神经网络<sup>[9-11]</sup>识别交通标志区域的交通标志检测方法,简称为 MH-CAL(MSER on HSV & CNN with attribute learning)。

## 1 交通标志检测方法

### 1.1 整体结构框架

本文提出的 MH-CAL 方法的整体结构框架

如图 1 所示。该方法主要流程如下:

(1)提取交通标志候选区域。在图像的 HSV 颜色空间的三个分量形成的灰度图上,分别使用 MSER 提取交通标志候选区域;

(2)交通标志候选区域分类。使用加入了属性学习约束的卷积神经网络提取候选区域特征,然后通过分类网络(共  $N+1$  个输出类,包括交通标志  $N$  个类及非交通标志类)以判断候选区域是否是交通标志区域,以及交通标志的类别。

(3)交通标志区域确定。在对交通标志区域分类后,使用非极大抑制算法(Non-maximum suppression, NMS)从重叠候选区域中选出一个最合适的候选区域确定为最终交通标志区域。

需要说明的是,属性约束网络旨在训练网络时,其损失(Loss)与分类网络损失共同作用于反向传播过程以优化网络参数。而在测试时,属性约束网络将被移除。

### 1.2 候选区域提取

由于交通标志牌的色调一致、饱和度更高和比周围区域更亮<sup>[12,13]</sup>,导致在 HSV 颜色空间的三个分量(色调 Hue, 饱和度 Saturation, 亮度 Value)分别形成的三个灰度图上,组成交通标志的图案内部灰度值相近,且相较于其周围灰度值更高或更低,而这正是连通域的特点。因此,可以检测连通域以提取候选区域。为了增加鲁棒性,本文提出同时在该三个灰度图上检测连通域以提取交通标志候选区域。

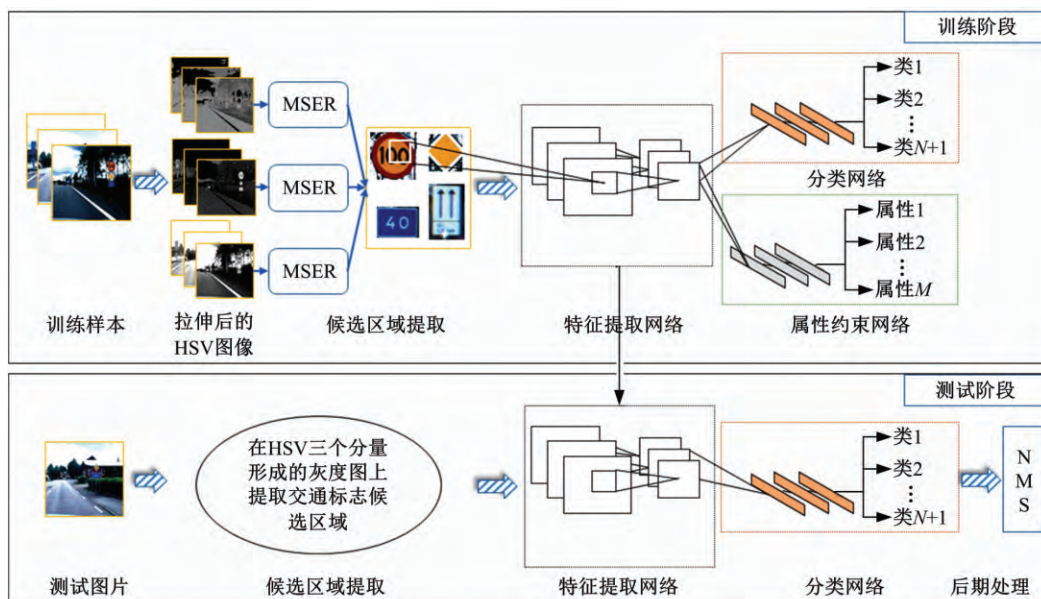


图 1 本方法整体结构框架

Fig. 1 Overview of the proposed method

最大稳定极值区域(MSER)<sup>[14]</sup>算法旨在检测灰度图中内部具有相近的灰度值、形状任意的区域。通过使用多个阈值将灰度图像二值化后,提取每幅二值化图像中的连通域,那些在多个二值化图像中形状基本保持不变的连通域即为MSER。因此,本文提出在HSV颜色空间的三个分量形成的三个灰度图上,利用MSER算法检测连通域以定位交通标志候选区域。

本文提出的候选区域提取算法如算法1所示。首先,将图像 $I_{RGB}$ 从RGB颜色空间转换到HSV颜色空间得到图像 $I_{HSV}$ ;接着,为了得到图像 $I_{HSV}$ 的三个通道形成的灰度图 $I_H$ 、 $I_S$ 、 $I_V$ ,需要对其三个通道分别使用线性对比度拉伸(Linear contrast stretching),使三个通道像素值的取值范围处于相同的区间 $[0,1]$ ;然后,分别在灰度图 $I_H$ 、 $I_S$ 、 $I_V$ 上利用MSER检测连通域;最后,根据经验,交通标志长宽比通常不会过大或过小,所以过滤掉长宽比过大或者过小的连通域,剩下的连通域对应原图中的区域作为交通标志候选区域,由分类器识别这些候选区域。

可以看出,上述方法在HSV颜色空间的三个分量上利用MSER检测连通域以提取交通标志候选区域。该方法并没有考虑标志的具体颜色种类和形状对称等先验信息,因此方法简单且通用性强。

#### 算法1 候选区域提取算法

Input: RGB 图像  $I_{RGB}$ , 长宽比阈值  $t_1, t_2$ , 满足  $t_1 > t_2$ .

Output: 交通标志候选区域框  $Bbox_{ROI}$ .

```

1  begin
2  将图像  $I_{RGB}$  由 RGB 颜色空间转换到 HSV 颜色空间得到图像  $I_{HSV}$ , 设  $I_{HSV}$  三个通道形成的图像分别为  $I_H, I_S, I_V$ ;
3   $I_1 = I_H/360; I_2 = I_S/100; I_3 = I_V/100$ ;
4   $ROI = \emptyset$ ;
5  for  $i=1, 2, 3$  do
6   $ROI = ROI \cup$  在灰度图  $I_i$  上使用 MSER 检测到的连通域;
7  end
8  for each  $roi \in ROI$  do
9    if  $Width_{roi}/Height_{roi} > t_1$  or  $Width_{roi}/Height_{roi} < t_2$ 
      then  $ROI = ROI \setminus roi$ ;
10 end
```

11 获取 ROI 中的候选区域的位置信息, 保存至  $Bbox_{ROI}$ ;

12 end

### 1.3 候选区域分类

在提取候选区域后,需要对候选区域分类以判断候选区域是否是交通标志区域,以及交通标志的类别。候选区域中通常包含了真实交通标志区域和非交通标志区域。真实交通标志区域通常会有一定的遮挡或形变;非交通标志区域通常数量较多并且与真实的交通标志区域具有相似的表现特征。因此需要泛化能力强(Strong generalization ability)和容量大(Higher capacity)的分类器,在准确识别真实交通标志区域的同时,过滤非交通标志区域。

卷积神经网络由于能够进行特征学习,且泛化能力强、容量大,已经被证实是能够理解图片内容、在图像识别、分割、检测、检索上达到最好效果的一类有效的模型<sup>[15-20]</sup>。其中,作为卷积神经网络的一个典型, AlexNet 在 ImageNet 1000 类目标分类任务上取得了当时最好的效果<sup>[15]</sup>。

然而,简单的卷积神经网络(如 AlexNet)只能学习目标类别这一高层语义信息。对于交通标志分类任务来说,由于交通标志具有特定的形状、颜色和图案内容,因此在训练卷积神经网络时,可以加入这三种视觉属性约束以弥补底层图像到高层语义之前的鸿沟,提升卷积神经网络的分类性能。故本文引入了交通标志形状、颜色和图案内容三种视觉属性,并针对不同的交通标志数据集设计了不同的属性标签。如对于一个交通标志数据集  $D$ ,若组成交通标志的形状、颜色和图案内容各有  $N_s^D, N_c^D, N_p^D$  种,则形状属性标签定义为:  $S^D = (S_1^D, S_2^D, \dots, S_{N_s^D}^D)$ , 颜色属性标签定义为:  $C^D = (C_1^D, C_2^D, \dots, C_{N_c^D}^D)$ , 图案内容属性标签定义为:  $P^D = (P_1^D, P_2^D, \dots, P_{N_p^D}^D)$ 。其中,  $S_i^D, C_j^D, P_k^D$  分别表示数据集  $D$  中组成交通标志的第  $i$  种形状、第  $j$  种颜色、第  $k$  种图案内容。那么,该数据集的属性标签则定义为:

$T^D = (S_1^D, S_2^D, \dots, S_{N_s^D}^D, C_1^D, C_2^D, \dots, C_{N_c^D}^D, P_1^D, P_2^D, \dots, P_{N_p^D}^D)$  若数据集  $D$  中某一类别的交通标志具有某一项属性,就将该属性标签的值置 1, 否则,置 0, 则每个类别的交通标志将具有唯一的属性向量,该属性向量将作为属性约束的学习目标,参与卷积神经网络的训练过程。

本文设计的卷积神经网络的训练网络结构如图 2 所示。该网络由特征提取网络(AlexNet 网络

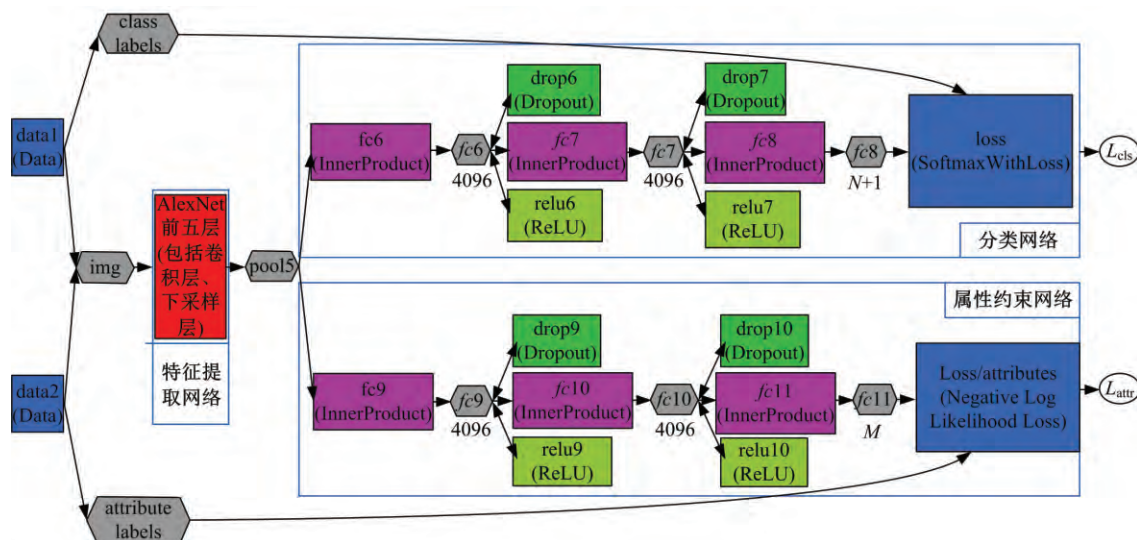


图 2 本文设计的卷积神经网络的训练网络结构

Fig. 2 The proposed CNN structure for training

前五层,包括卷积层、下采样层等)、分类网络、属性约束网络构成。其中,分类网络和属性约束网络各由三层全连接层组成,两者共享特征提取网络的输出(即图像的高层特征 pool5)。该网络拥有两个损失,一个是分类网络损失,另一个是属性约束网络损失。在训练卷积神经网络时,属性约束网络的损失将和分类网络的损失共同作用于反向传播过程以优化网络各层参数,通过参数共享的方式同时进行属性学习和分类学习。

训练网络时,假设训练样本集  $D$  中共有  $m$  个样本、 $N+1$  个类别(交通标志  $N$  个类及非交通标志类)、 $M$  维属性标签。样本  $i$  的真实类别为  $y^{(i)}$ ,形状属性向量为  $S_{y^{(i)}}^D$ ,颜色属性向量为  $C_{y^{(i)}}^D$ ,图案内容属性向量为  $P_{y^{(i)}}^D$ ,则该样本的属性向量为  $T_{y^{(i)}}^D = (S_{y^{(i)}}^D, C_{y^{(i)}}^D, P_{y^{(i)}}^D)$ 。本文设计的具有属性约束的卷积神经网络的目标函数定义为:

$$\min L = \min(L_{\text{cls}} + \lambda L_{\text{attr}}) \quad (1)$$

$$L_{\text{cls}} = -\frac{1}{m} \sum_{i=1}^m \sum_{j=0}^N (1\{y^{(i)} = k\} \log p_j^{(i)}) + R_{\text{cls}} \quad (2)$$

$$L_{\text{attr}} = -\frac{1}{m} \sum_{i=1}^m T_{y^{(i)}}^D \log t^{(i)} + R_{\text{attr}} \quad (3)$$

其中,分类网络的损失函数  $L_{\text{cls}}$  为 Softmax 损失函数,属性约束网络的损失函数  $L_{\text{attr}}$  为负对数似然损失函数(Negative log-likelihood), $1\{\cdot\}$  为指示函数,  $R_{\text{cls}}$  和  $R_{\text{attr}}$  为  $L_2$  正则化项;  $p^{(i)} = (p_0^{(i)}, \dots, p_N^{(i)})$ ,  $p_j^{(i)} \in (0, 1)$  为样本  $i$  的预测类别的概率分布;  $t^{(i)} = (t_0^{(i)}, \dots, t_{M-1}^{(i)})$ ,  $t_j^{(i)} \in (0, 1)$  为样本  $i$  的预测属性向量。式(1)中的超参数  $\lambda$  控

制了属性学习和分类学习在训练过程中的重要程度,在实验中,  $\lambda = 1$ , 使用反向传播算法训练整个网络。

从该目标函数中可以看出,在训练时,该网络不仅考虑了交通标志类别信息,同时也考虑了交通标志的形状、颜色、图案内容三种视觉属性信息,如此训练后得到的卷积神经网络将具有属性约束特性,能够有效弥补交通标识底层图像到高层语义之间的鸿沟。需要注意的是,在测试时,属性约束网络将被移除,保留特征提取网络和分类网络,进而直接得出输入图像的分类结果。

#### 1.4 交通标志区域确定

在提取交通标志候选区域过程中,会出现同一个交通标志被检测出多个候选区域的情况,且这些候选区域高度重叠。因此,在对每个候选区域分类后,需要从重叠的候选区域中,选出一个最合适的候选区域确定为最终的交通标志区域,去除其他的候选区域。

##### 算法 2 针对交通标志的 NMS 算法

Input: 已分类的候选区域  $R$ ,  $n$  是  $R$  中区域的数量,  $op$  是覆盖率阈值。

Output: 交通标志区域  $R_{pr}$ 。

Initialization:  $keep[1 \cdots n] = \text{true}$ , 表示  $R$  中的区域是否被保留。

- 1 begin
- 2 将  $R$  中区域按照区域面积降序排列。
- 3 将  $R$  中区域按照区域分类得分降序排列。
- 4 for  $i=1, \dots, n$  do

```

5   for j=i+1,...,n do
6       if overlap(R[i],R[j]) > op
           then keep[j]=false;
7   end
8   end
9   Rpr=R[index(keep==true)];
10  end

```

本文根据每个候选区域的分类得分采用 NMS 算法确定最终的交通标志区域。即对于每个候选区域,找出与其重叠且 IoU 覆盖率大于一定阈值的候选区域,保留分类得分最高的候选区域,去除其他候选区域。然而,当两个候选区域处于包含关系时,若两者的分类得分相同,标准的 NMS 可能会保留被包含的区域,去除包含的那个区域。但是对于交通标志来说,经常出现同一个标志内,有多个图案处于包含关系。限速标志 70 内部黄色区域外嵌套着红色的区域,由于黄色区域和红色区域都是连通域,因此两者覆盖的区域都会被选为交通标志候选区域。显然,若这两个候选区域分类得分相同,由于面积较大者(红色区域对应的候选区域)定位更加准确,此时算法应该保留面积较大的候选区域,去除面积较小的候选区域(黄色区域对应的候选区域),然而,标准的 NMS 可能会保留面积较小的候选区域,去除面积较大的候选区域。为了解决该问题,本文针对交通标志检测改进了 NMS 算法,如算法 2 所示,在标准的 NMS 的基础上,对 NMS 的输入做了按照候选区域的面积递减排序的预处理。

## 2 实 验

### 2.1 实验数据集

本文在两个公开的数据集上对本文提出的交通标志检测方法进行了评测:STSD(Sweden

traffic sign detection dataset)<sup>[21]</sup>,GTSD(German traffic sign detection dataset)<sup>[22]</sup>。STSD 中交通标志共有 20 个类别,有 20 000 多张图片,其中 20% 被人工标注,共 3488 个交通标志;GTSD 中交通标志共有 43 个类别,有 600 张训练图片,其中包含 846 个交通标志,有 300 张测试图片,其中包含了 360 个交通标志。

对于 STSD,本文按照文献[21]中的评测方法,只考虑 Pedestrian crossing、Designated lane right、No standing or parking、50 kph、Priority road 和 Give way 这 6 个类别。对于 GTSD,本文跟其他文献一样,只评测 43 个类别所属的三个超类(Superclass),即 Mandatory、Danger 和 Prohibitory。

### 2.2 候选区域提取分析

#### 2.2.1 实验设计

本小节设计了 6 种提取交通标志候选区域的方法,并在公开数据集 STSD 和 GTSD 上进行了评估。该实验旨在评估不同颜色空间对交通标志候选区域提取效果的影响,该实验分别在 6 种颜色空间的分量形成的灰度图上提取 MSER 作为交通标志候选区域,6 种颜色空间分别为 Ycber, RGB, YUV, Lab 和 HSV(本文选择的颜色空间),以及文献[1]中提出的 Enhanced RGB 颜色空间,该方法在 RGB 颜色空间中增强了 R 和 B 通道,使得交通标志区域更加显著。

#### 2.2.2 实验结果及分析

表 1 列出了在 Ycber、RGB、YUV、Lab 各自的三个通道上分别使用 MSER 提取候选区域、文献[1]在 Enhanced RGB 颜色空间中的 R、B 两个通道上使用 MSER 提取候选区域和本文中提出的方法(HSV + MSER)的比较,其中,算法 MSER 中的参数  $\Delta$  取值为 2。本小节实验采用平

表 1 不同候选区域提取方法在 STSD、GTSD 数据集上的实验结果

Table 1 Experimental results on dataset STSD and GTSD using different methods to extract candidate regions

颜色模型	STSD			GTSD		
	MABO	# win	MR	MABO	# win	MR
Ycber	0.647	310	0.758	0.727	823	0.767
RGB	0.661	994	0.749	0.750	2464	0.847
RGB <sup>[1]</sup>	0.684	153	0.763	0.827	437	1.000
YUV	0.780	591	0.900	0.748	1087	0.801
Lab	0.821	403	0.955	0.826	997	0.943
HSV(本文方法)	0.838	1157	0.979	0.846	2552	0.987



均召回率(Mean recall, MR)和文献[23]中提出的两个评价指标: MABO (Mean average best overlap)和 # win。其中 MR 和 MABO 被定义为:

$$MR = \text{avg}_{c_k \in C}(\text{recall}(c_k)) \quad (4)$$

$$MABO = \text{avg}_{c_k \in C}(ABO^{c_k}) \quad (5)$$

$$ABO^c = \frac{1}{|G^c|} \sum_{g_i^c \in G^c} \max_{l_j \in L_{g_i^c}} \text{Overlap}(g_i^c, l_j) \quad (6)$$

其中,  $C$  表示所有交通标志类别,  $G^c$  表示第  $c$  个类别中所有真实的标注区域,  $L_{g_i^c}$  表示候选区域  $g_i^c$  所在图片中的所有候选区域,  $\text{Overlap}$  表示两个候选区域的 IoU 覆盖率。

从表 1 中可以看出:

(1)在不同的颜色空间上候选区域提取的效果差距明显。在指标 MABO 和 MR 上,使用 Ycbcr 颜色空间提取候选区域效果最差,在最常见 RGB 颜色空间上效果也不理想。

(2)在指标 MABO 和 MR 上,使用 YUV 颜色空间的效果与本文较为接近,原因是都利用了颜色的色调、饱和度和亮度信息;在 Lab 颜色空间上的效果与本文的结果最近,原因是该颜色空间以数字化方式来描述人的视觉感应,能够表达的颜色更多,因此能够更好地区分目标区域和背景。

(3)在公开库 STSD 上,在不利用对称信息的情况下,在指标 MABO 上,本文提出的方法比文献[1]高出约 0.15,比其他方法高出 0.02~0.19;在指标 MR 上,本文比文献[1]上高出约 0.22,比其他方法高出 0.02~0.23。

(4)在公开库 GTSD 上,使用文献[1]中的方法提取候选区域具有最高的 MR=1,略高于本文提出的方法,但是在指标 MABO 上比本文低;在 Lab 颜色空间上的 MABO 和 MR 的值仅次于文献[1]和本文提出的方法;相比较其他颜色空间,本文在指标 MABO 上高出 0.10~0.12,在指标 MR 上高出 0.14~0.22。

(5)在指标 # win 上,本文提出的方法高于其他方法,这会导致在候选区域分类过程中在速度上不占优势。

从在两个公开库上的比较可以看出,使用 HSV、Lab 颜色空间提取交通标志候选区域具有较好的效果。使用 HSV 颜色空间平均覆盖率和平均召回率较高,使用 Lab 颜色空间由于 # win

较小,故检测算法整体速度更快。因此,从实验对比可知,色调、对比度和亮度信息是提取交通标志候选区域的重要信息。本文提出的方法正是利用了上述重要信息,即在 HSV 颜色空间的三个通道形成的灰度图上使用 MSER 算法检测标志候选区域,才能具有较高的平均覆盖率和平均召回率,即交通标志定位更准和找到的交通标志更多。由在不同的库上实验对比得出,本文提出的交通标志候选区域的方法效果最好;然而,本文提出的方法在每张图片平均候选区域数量上相比较其他方法略高。

## 2.3 交通标志检测实验

### 2.3.1 实验设计

在 STSD 数据集上,本文对比了 4 种检测算法,分别为文献[21]提出的使用傅里叶描述子描述交通标志的轮廓并分类的方法、文献[8]提出的方法、本文提出的 MH-CAL 方法和本文提出的去除属性约束网络的方法 MH-CNN。

在 GTSD 数据集上,本文对比了 6 种检测算法,分别为文献[22]提出的 HOG 特征结合颜色直方图(Color Histogram)特征描述交通标志候选区域并用 SVM 对候选区域分类的方法、文献[24]使用 ChannelFeature 检测子的检测方法、文献[1]、文献[8]提出的方法、本文提出的方法 MH-CAL 以及本文提出的方法去除属性约束网络的方法 MH-CNN。

### 2.3.2 属性标签设计

本文在充分观察、总结数据集 STSD 和 BTSD 中交通标志图案的特点后,分别为 STSD 和 BTSD 设计了属性标签,表 2 列出了在 STSD 数据集上需要评测的 6 个交通标志类别和背景类别的属性标签和属性向量,表 3 列出了在 GTSD 数据集上需要评测的 3 个超类和背景类别的属性标签和属性向量。本文从标志整体形状、标志颜色、标志图案内容这三个方面对属性进行了划分,各类交通标志如果含有某项属性,则该属性标签值置为 1,否则置为 0。由于 GTSD 数据集上只需评测 3 个超类,而不是具体交通标志类别,因此只需使用形状和颜色这两个方面设计属性标签。最后,每个交通标志类(或超类)会生成一个唯一的属性向量,在训练卷积神经网络时,与类别信息共同作为训练目标。

表 2 数据集 STSD 的属性标签和属性向量

Table 2 Attribute labels and attribute vectors for dataset STSD

属性	标签	Background	Pedestrian crossing	Designated lane right	No standing or parking	50 km/h	Priority road	Give way
形状	圆形	0	0	1	1	1	0	0
	三角形	0	1	0	0	0	0	1
	菱形	0	1	0	0	0	1	0
	八边形	0	0	0	0	0	0	0
颜色	红色	0	0	0	1	1	0	1
	白色	0	1	1	0	0	1	0
	黑色	0	1	0	0	1	0	0
	蓝色	0	1	1	1	0	0	0
	黄色	0	0	0	0	1	1	1
图案内容	30	0	0	0	0	0	0	0
	50	0	0	0	0	1	0	0
	60	0	0	0	0	0	0	0
	70	0	0	0	0	0	0	0
	80	0	0	0	0	0	0	0
	90	0	0	0	0	0	0	0
	100	0	0	0	0	0	0	0
	110	0	0	0	0	0	0	0
	120	0	0	0	0	0	0	0
	Lane left	0	0	0	0	0	0	0
	Lane right	0	0	1	0	0	0	0
	Pedestrian	0	1	0	0	0	0	0
	Stop	0	0	0	0	0	0	0
		0	0	0	0	0	0	0
		0	0	0	1	0	0	0
	无	0	0	0	0	0	1	1

表 3 数据集 GTSD 的属性标签和属性向量

Table 3 Attribute labels and attribute vectors for dataset GTSD

属性	标签	Background	Mandatory	Danger	Prohibitory
形状	圆形	0	1	0	1
	三角形	0	0	1	0
	菱形	0	0	0	0
	八边形	0	0	0	0
颜色	红色	0	0	1	1
	白色	0	1	1	1
	黑色	0	0	1	1
	蓝色	0	1	0	0
	黄色	0	0	0	0

2.3.3 训练过程

由于两个公开数据集数据量不足,因此本文中卷积神经网络的训练采用基于归纳式迁移学习

(Inductive transfer learning)的方法。本文使用在数据集 ImageNet 上训练好的 Alex 网络模型作为预训练模型(Pre-trained model),使用交通标志样本对网络进行有监督的微调(Supervised fine-tuning)。在训练时,特征提取网络和分类网络使用预训练模型中的参数,属性约束网络中的参数使用随机初始化参数。

由于现有的交通标志数据集训练样本并不充足,为了获取更多的训练样本,本文采取以下方法。

首先,将样本库中的所有图像按顺时针和逆时针方向各旋转 5 度形成新的训练集;然后,在新的训练集上提取交通标志候选区域,根据经验知识,交通标志的长宽比不会过大,因此,本文只保

留长宽比( $w/h$ )满足  $1/3.5 \leq w/h \leq 1.4$  的候选区域;最后,将与真实区域 IoU 覆盖率大于 0.7 的候选区域作为交通标志样本,覆盖率小于 0.3 的候选区域作为非交通标志样本,其他的候选区域不作为训练样本。本文使用随机梯度下降法以 0.001 的学习率(Learning rate)开始训练网络。

在增加交通标志样本的同时,也增加了大量非交通标志的样本。非交通标志样本和交通标志样本的比例大约为 400 : 1。为了处理数据不均衡的问题,在训练模型时,本文采用如下训练方法:首先,从非交通标志样本中随机挑选 4000 个样本和所有的交通标志样本作为训练集  $S_{\text{train}}$ ,训练模型 M;然后,用  $S_{\text{train}}$  训练好的模型 M 测试  $S_{\text{train}}$  中所有非交通标志样本,将分错的非交通标

志样本随机选取 10% 加入到  $S_{\text{train}}$  中,重新训练模型 M;以此类推,重复几次,直到非交通标志样本的误检率降低到目标值。

#### 2.3.4 实验结果及分析

在 STSD 数据集上,本文按照文献[21]中要求,只考虑 20 个交通标志类别中的 6 个类别: Pedestrian crossing、Designated lane right、No standing or parking、50 km/h、Priority road 和 Give way,使用准确率(Precision)和召回率(Recall)评价方法。表 4 是在 STSD 公开数据集上与文献[8,21]中的结果的对比,其中 MH-CAL 是本文提出的方法,MH-CNN 指本文提出的方法去除属性约束网络后的方法。

表 4 STSD 数据集上的实验对比

Table 4 Comparison of experiments on dataset STSD

交通标志类别	Precision/%				Recall/%			
	文献[21]方法	文献[8]方法	MH-CAL	MH-CNN	文献[21]方法	文献[8]方法	MH-CAL	MH-CNN
Pedestrian crossing	96.03	98.52	100.00	99.12	97.77	93.45	90.32	88.89
Designated lane right	100.00	100.00	100.00	100.00	95.33	97.53	96.84	95.16
No standing or parking	97.14	99.20	100.00	100.00	77.27	81.46	85.71	84.58
50 km · h <sup>-1</sup>	100.00	100.00	100.00	96.77	76.12	80.56	96.88	93.75
Priority road	95.66	97.89	98.34	98.00	74.24	79.68	88.54	85.43
Give way	59.26	71.50	97.42	94.80	47.26	52.39	86.11	84.12
Overall	91.35	94.52	99.29	98.12	77.00	80.85	90.73	88.67

注:MH-CAL 是本文提出的方法,MH-CNN 指的是 MH-CAL 去除属性约束网络后的方法

(1)在准确率方面,在每一个类别上,本文提出的检测方法都优于文献[21]和文献[8]中提出的方法,尤其是在类别 Give way 上,较文献[8]提升了 25.92%,较文献[21]提升了 38.16%,这有赖于卷积神经网络提取的有效特征;

(2)在召回率方面,对于类别 No standing or parking、50 km/h、Priority road、Give way,本文提出的方法超过了文献[8]4%~34%,尤其是在 50 km · h<sup>-1</sup> 和具有挑战性的 Give way 这两个类别上有了大幅提升,说明本文提出的候选区域提取方法能够更好地定位交通标志区域。

(3)在类别 Designated lane right 上,本文在召回率上稍落后于文献[8]的方法;对于类别 Pedestrian crossing,本文的方法在召回率上较低,原因是其内部图案较为复杂,当该标志在图片中较小时,标志表面的连通域不够大,导致在候选区域提取步骤中无法提取到该区域。如图 3 所

示,黑方框标出的是本文提出的算法提取出的候选区域,可以看到,整个交通标志区域没有被提取出来。

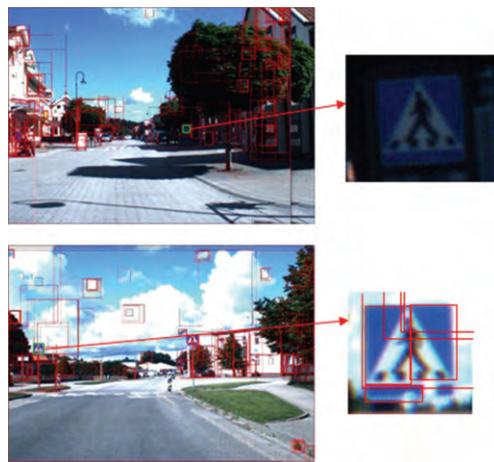


图 3 未检测到整个 Pedestrian crossing 区域

Fig. 3 Fail to detect the whole area of Pedestrian crossing

(4)较 MH-CNN 方法,在加入属性学习约束后,准确率和召回率都有所提升,说明本文提出的



属性学习约束方法确实有效。使用卷积神经网络的 MH-CNN 方法在准确率和召回率上普遍优于使用传统分类器的方法<sup>[8,21]</sup>,说明在交通标志分类任务上,深度学习方法要优于传统学习方法。

(5)整体来看,在平均准确率和平均召回率上,较文献[8]分别提高了 4.77%和 9.88%,较文献[21]分别提高了 7.94%和 13.73%。

同时,本文也在 GTSD 数据集上用了同样的方法和参数对本文提出的检测方法和其他文献中的方法做了对比,在该库上,本文和其他文章一样,只评测 43 个类别所属的三个超类:

表 5 GTSD 数据集上的实验对比

Table 5 Comparison of experiments on dataset GTSD

交通标志类别	AUC/%		
	Mandatory	Danger	Prohibitory
HOG+ColorHist+SVM <sup>[22]</sup>	92.00	98.85	100
Channel features <sup>[24]</sup>	96.98	100	100
HOG+SVM+Context aware filter <sup>[1]</sup>	95.01	97.22	99.43
Saliency+Ada+SVR <sup>[8]</sup>	99.87	100	100
MH-CNN	95.95	98.80	100
MH-CAL(本文方法)	96.10	100	100

从上述两个实验可以看出,本文提出的交通标志检测方法都取得了较高的召回率,说明了在 HSV 颜色空间上提取交通标志候选区域的有效性,由此可知,色调、对比度和亮度信息是提取交通标志候选区域的重要信息;本文提出的方法也取得了较高的准确率,说明卷积神经网络相比于

Mandatory、Danger、Prohibitory。在这个公开数据集上训练时,也是采用归纳式迁移学习的方法,与 STSD 不同的是,在 GTSD 上用的预训练模型是在 STSD 数据集上训练好的模型。表 5 使用了 AUC(Area Under PR Curve)评价标准在 GTSD 上对比了不同的检测方法。从表中可以看出,本文提出的算法在类别 Danger、Prohibitory 上优于文献[1]和文献[22]中提出的方法,与文献[24]和文献[8]中提出的方法持平;在类别 Mandatory 上也取得了较高的结果。

传统分类器,泛化能力更强、容量更大;在加入属性学习约束后,准确率和召回率都得到了提升,说明本文提出的属性学习约束学习确实有效。然而,本文提出的方法有一定的不足,当交通标志表面图案较为复杂、元素较多时,如图 4 所示,候选区域提取的性能会有一定的损失。

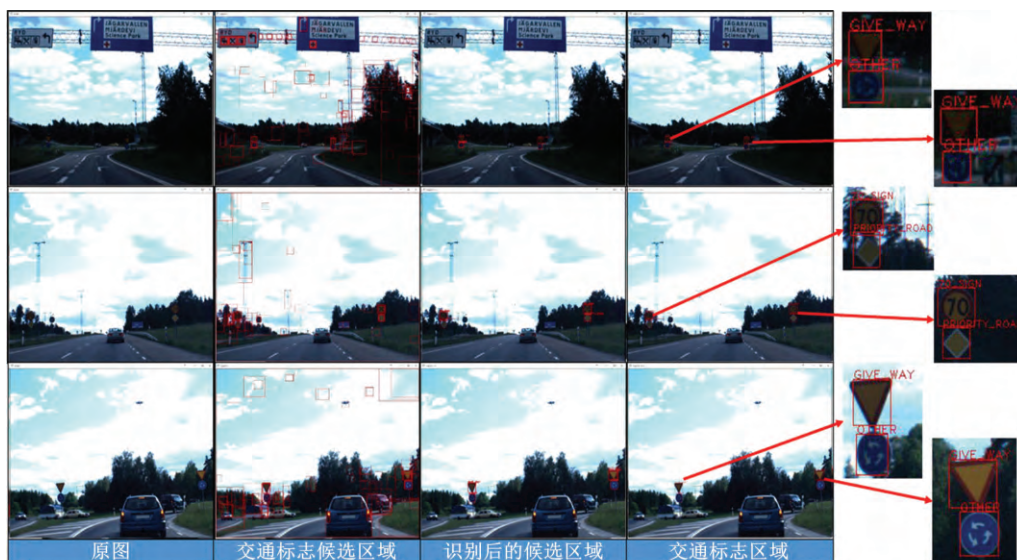


图 4 本文提出的交通标志检测方法检测效果示例

Fig. 4 Samples of the proposed traffic sign detection method

### 3 结束语

现有的大部分交通标志检测方法都没有考虑交通标志牌表面的特性和交通标志图案的视觉属性。本文提出了一种基于深度属性学习的交通标志检测方法。该方法首先在 HSV 颜色空间上使用 MSER 提取交通标志候选区域;然后使用加入了属性学习约束的卷积神经网络提取候选区域特征并识别。本文方法不但充分利用了交通标志牌表面不同颜色的共同特性、制作材料的特殊性,而且将交通标志的形状、颜色、图案内容等一系列视觉属性作为约束条件融入到了分类器中。通过在两个公开数据集上的实验证明,本方法不但能够在复杂环境下准确定位交通标志区域,而且能够准确地识别交通标志区域和过滤非交通标志区域。

#### 参考文献:

- [1] Salti S, Petrelli A, Tombari F, et al. Traffic sign detection via interest region extraction[J]. Pattern Recognition, 2015, 48(4): 1039-1049.
- [2] Gómez-Moreno H, Maldonado-Bascón S, Gil-Jiménez P, et al. Goal evaluation of segmentation algorithms for traffic sign recognition[J]. IEEE Transactions on Intelligent Transportation Systems, 2010, 11(4): 917-930.
- [3] Shadeed W G, Abu-Al-Nadi D I, Mismar M J. Road traffic sign detection in color images[C]//IEEE International Conference on Electronics, Circuits and Systems, Sharjah, United Arab Emirates, 2003: 890-893.
- [4] Ruta A, Li Y, Liu X. Real-time traffic sign recognition from video by class-specific discriminative features[J]. Pattern Recognition, 2010, 43(1): 416-430.
- [5] Li H, Sun F, Liu L, et al. A novel traffic sign detection method via color segmentation and robust shape matching[J]. Neurocomputing, 2015, 169: 77-88.
- [6] Loy G, Barnes N. Fast shape-based road sign detection for a driver assistance system[C]//IEEE International Conference on Intelligent Robots and Systems, Sendai, Japan, 2004: 70-75.
- [7] Barnes N, Zelinsky A, Fletcher L S. Real-time speed sign detection using the radial symmetry detector[J]. IEEE Transactions on Intelligent Transportation Systems, 2008, 9(2): 322-332.
- [8] Chen T, Lu S. Accurate and efficient traffic sign detection using discriminative adaboost and support vector regression[J]. IEEE Transactions on Vehicular Technology, 2015, 65(6): 4006-4015.
- [9] Wu Y, Liu Y, Li J, et al. Traffic sign detection based on convolutional neural networks[C]//IEEE International Joint Conference on Neural Networks, Dallas, TX, USA, 2013: 1-7.
- [10] Qian R, Zhang B, Yue Y, et al. Robust Chinese traffic sign detection and recognition with deep convolutional neural network[C]//IEEE International Conference on Natural Computation, Zhangjiajie, China, 2015: 791-796.
- [11] Zang D, Zhang J, Zhang D, et al. Traffic sign detection based on cascaded convolutional neural networks[C]//IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, Shanghai, China, 2016: 201-206.
- [12] Farhadi A, Endres I, Hoiem D, et al. Describing objects by their attributes[C]//IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA, 2009: 1778-1785.
- [13] Berg T L, Berg A C, Shih J. Automatic attribute discovery and characterization from noisy web data [C]//European Conference on Computer Vision, Heraklion, Crete, Greece, 2010: 663-676.
- [14] Matas J, Chum O, Urban M, et al. Robust wide-baseline stereo from maximally stable extremal regions[J]. Image and Vision Computing, 2004, 22(10): 761-767.
- [15] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems, Lake Tahoe, Nevada, USA, 2012: 1097-1105.
- [16] Farabet C, Couprie C, Najman L, et al. Learning hierarchical features for scene labeling[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1915-1929.
- [17] Ciresan D, Giusti A, Gambardella L M, et al. Deep neural networks segment neuronal membranes in electron microscopy images[C]//Advances in neural information processing systems, Lake Tahoe, Nevada, USA, 2012: 2843-2851.
- [18] Sermanet P, Eigen D, Zhang X, et al. Overfeat: in-

- tegrated recognition, localization and detection using convolutional networks [J]. arXiv preprint arXiv:1312.6229, 2013.
- [19] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014:580-587.
- [20] Sharif R A, Azizpour H, Sullivan J, et al. CNN features off-the-shelf: an astounding baseline for recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014:806-813.
- [21] Larsson F, Felsberg M. Using Fourier descriptors and spatial models for traffic sign recognition[C]//Scandinavian Conference on Image Analysis, Ystad, Sweden, 2011: 238-249.
- [22] Houben S, Stallkamp J, Salmen J, et al. Detection of traffic signs in real-world images: the German traffic sign detection benchmark[C]//IEEE International Joint Conference on Neural Networks, Dallas, TX, USA, 2013:1-8.
- [23] 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.
- [24] Mathias M, Timofte R, Benenson R, et al. Traffic sign recognition—how far are we from the solution? [C]//IEEE International Joint Conference on Neural Networks, Dallas, TX, USA, 2013:1-8.