

文章编号: 1004-4353(2020)04-0359-07

# 基于多尺度卷积神经网络的 交通标志识别方法

仲会娟, 蔡清泳

(阳光学院 人工智能学院, 福建 福州 350015)

**摘要:** 为了提升交通标志自动识别的精度, 提出一种基于多尺度 CNN 的交通标志识别方法(TSR-MSCNN 算法)。该方法采用三阶段卷积神经网络, 融合了低阶、中阶和高阶 3 种不同尺度的特征, 并串联了多个小卷积层用以代替单个较大卷积层。通过对全连接层的神经元个数、Dropout 参数、卷积核尺寸等网络超参数进行选比实验, 获得了最佳的网络超参数。利用德国交通标志基准数据库(GTSRB)对不同算法进行测试表明, 本文提出的算法在较小的网络参数量下能够有效提取交通标志特征, 获取的识别准确率达到 99.76%, 且显著优于传统卷积神经网络方法和多尺度特征方法的识别准确率, 因此本文算法在图像识别领域有良好的应用价值。

**关键词:** 交通标志识别; 卷积神经网络; TSR-MSCNN; 多尺度特征

中图分类号: TP391

文献标识码: A

## Traffic sign recognition method based on multi-scale convolutional neural network

ZHONG Huijuan, CAI Qingyong

(College of Artificial Intelligence, Yangou University, Fuzhou 350015, China)

**Abstract:** In order to improve the accuracy of automatic traffic sign recognition, we propose a traffic sign recognition algorithm based on multi-scale CNN. This method uses a three-stage convolutional neural network to fuse features of three different scales: low-order, medium-order, and high-order; and concatenates multiple small convolutional layers to replace a single larger convolution layer. In addition, the network hyperparameters such as the number of neurons in the fully connected layer, the dropout parameters, and the sizes of the convolution kernels are investigated and compared to obtain the best hyperparameter set. Different algorithms are tested on the German traffic sign recognition benchmark (GTSRB). Experimental results show that the proposed algorithm in this paper can effectively extract traffic sign features and obtain recognition accuracy of 99.76% under a small amount of network parameters, which is obviously superior to the traditional convolutional neural network method and multi-scale feature recognition method. Therefore, the algorithm proposed in this paper has good usability in the field of image recognition.

**Keywords:** traffic sign recognition; convolutional neural network; TSR-MSCNN; multi-scale features

### 0 引言

交通标志识别(traffic sign recognition, TSR)是智能驾驶的重要组成部分。但由于交通标志图

像取自自然场景之中, 容易受气候、光照、腐蚀、损坏、遮盖、运动模糊、旋转倾斜等因素的影响, 因此如何提高图像的识别准确率受到学者们的关

收稿日期: 2020-07-26

作者简介: 仲会娟(1985—), 女, 讲师, 研究方向为图像与信号处理、无线通信技术。

基金项目: 福建省中青年骨干教师教育科研项目(JT180724); 电子信息与通信技术慕课应用型团队项目(2019sjtd01)

注<sup>[1]</sup>.传统的交通标志识别算法有基于模板匹配、支持向量机(SVM)、随机森林等,但这些方法需要人为定义特征信息,如 SIFT、HOG、SURF、LBP 等特征<sup>[2-5]</sup>.由于人为定义的特征容易忽略图像的很多内在有用特征,因此上述方法的识别精度有待进一步提高.近年来,随着深度学习的兴起,卷积神经网络因在图像检测与识别中能够获得较强的鲁棒性和较高的识别准确率而受到学者的关注.例如:文献[6]提出了一种二阶段多尺度特征分类算法,该算法在德国交通标志基准数据集 GTSRB (German traffic sign recognition benchmark)<sup>[7]</sup>上的识别准确率为 98.97%.文献[8]提出了一种基于支持向量机的交通标志图像分类算法,该算法在 GTSRB 数据集上的识别准确率为 98.24%.文献[9]提出了一种三阶段多尺度特征分类算法,该算法在 GTSRB 数据集上的识别准确率为 98.82%.

研究表明:图像的多尺度特征在人脸识别、数字识别、图像分割等领域具有良好的适用性<sup>[10-11]</sup>;VGG 卷积神经网络模型因采用了多个小卷积层代替了单个较大卷积层,因此其网络深度得到增加,能够学习更为复杂的图像特征,且在感受野不变的情况下能够提高模型训练与图像识别的速度<sup>[12]</sup>.基于上述研究结果,本文结合多尺度分类算法和 VGG 模型的小卷积核特性,提出了一种基于多尺度卷积神经网络的交通标志识别方法(traffic sign recognition method based on multi-scale convolutional neural network, TSR-MSCNN),并在自建的数据集和 GTSRB 数据集中验证了算法的有效性.

## 1 数据集介绍

### 1.1 自定义交通标志数据集

为减少网络训练时间以及能够验证算法的泛化能力,本文采用自定义数据集用于后期的参数对比实验.自定义数据集中有 40 类交通标志,共 3 250 张图片.图片分别取自比利时交通标志数据集及网络,格式为 ppm,其中包含了光线遮挡、拍摄角度、拍摄模糊等复杂情况.由于 ppm 格式的图片无法直接看到,因此将图片转换为了 jpg 格

式,如图 1 所示.同时,为了方便后期验证,对转换后的 40 类交通标志的图像集合进行了重新命名,如图 2 所示.



图 1 自定义的数据集图像

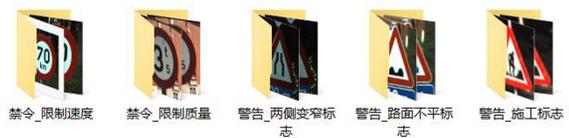


图 2 部分交通标志图像集合

### 1.2 GTSRB 交通标志数据集

GTSRB 数据集中的标志图像由车载高清摄像头在自然环境下拍摄得到.数据集分为 5 大类和 43 小类,共 51 839 幅图片,其中 12 630 幅为测试集,39 209 幅为训练集,每张图片只包含一种交通标志,如图 3 所示.数据集中每类至少包含 9 种不同角度的图像,且每个角度包含从小到大的 30 张图片.

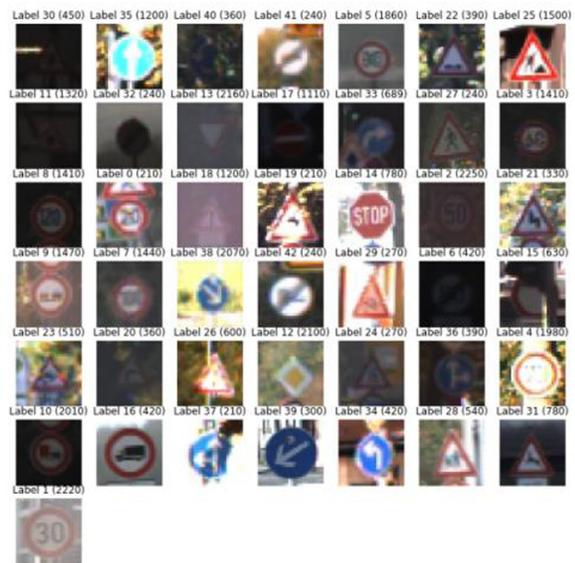


图 3 GTSRB 数据集中的部分图像

## 2 数据集预处理

因数据集中的交通标志图片是在多种自然条件下拍摄得到的,其图像质量和尺寸大小存在较大差别,因此需对图像进行预处理,以便更好地进

行特征提取,提高模型精度。

### 2.1 尺寸归一化

GTSRB 数据库中的交通标志图像的尺寸范围在  $15 \times 15$  像素至  $250 \times 250$  像素之间,因此需对其进行尺寸归一化处理。目前,常用于图像尺寸变换的算法主要有最近邻插值、双线性插值和双三次插值 3 种算法<sup>[13]</sup>。其中,双线性插值法不但能够克服最近邻插值法不连续的问题,而且计算量小于双三次插值法;因此,本文采用双线性插值法对图像的尺寸进行归一化处理,归一化的尺寸统一设定为  $32 \times 32$  像素。

### 2.2 像素归一化

图像像素的取值通常为  $0 \sim 255$  范围的整数。为了保持数据之间的相对关系和提高后续处理数据的速度,本文对像素进行了归一化处理,即将像素值压缩到  $0 \sim 1$  之间。

### 2.3 图像数据增广

TSR-MSCNN 为三阶段卷积神经网络模型,具有识别准确率高、网络参数量低、识别速率快等优点,但该方法的模型训练需要大量的图像。为此,本文采用图像增广技术<sup>[14]</sup>来扩充训练数据规模。由于获取的交通标志图像常存在变形、角度和大小不一等问题,因此本文将原始图像以随机旋转  $10^\circ$ 、随机缩放 0.2 倍、随机平移(幅度为图像尺寸的 0.08 倍)和随机翻转的方式进行增广,增广的图像如图 4 所示。



图 4 数据增广图像

## 3 基于多尺度卷积神经网络的交通标志识别法

### 3.1 二阶段多尺度卷积神经网络

传统的卷积神经网络是利用交替的卷积层和池化层获得图像特征信息,并将最后一层特征送

入全连接层中进行分类识别。由于该方法识别图像的特征尺度单一,因此将该方法直接用于交通标志识别其效果并不理想。M. Zeiler 等<sup>[15]</sup>研究表明,卷积神经网络能够提取出图像的多尺度特征信息,其中提取出的低阶特征反应的是图像中角、边缘等简单的局部特征,提取出的高阶特征反应的是图像的整体特征。二阶段多尺度卷积神经网络结构<sup>[6]</sup>如图 5 所示。

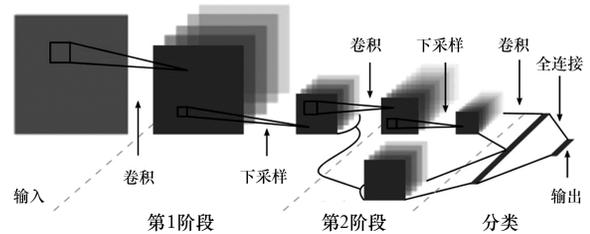


图 5 二阶段多尺度卷积神经网络结构示意图

由图 5 可以看出,网络结构中包含 2 个阶段的卷积层,2 个池化层,1 个全连接层和 1 个分类输出层。其中,第 1 阶段卷积层的输出特征在池化后产生分支,该分支的输出特征经过第 2 个下采样点后与第 2 阶段卷积层的输出特征一同被送入分类器。

### 3.2 MS-TSRCNN 模型

本文在文献<sup>[6]</sup>的基础上通过对网络结构和参数设置进行改进,提出了一种应用于交通标志识别领域的 TSR-MSCNN 模型,模型结构如图 6 所示。图 6 中:第 1 个子卷积网络 Conv1 由卷积层 Conv11 和 Conv12 构成,第 2 个子卷积网络 Conv2 由 Conv21 和 Conv22 构成,第 3 个子卷积网络 Conv3 由 Conv31、Conv32 和 Conv33 构成。Max\_11 和 Max\_22 为 2 个尺寸统一的池化层,分别连接池化层 Maxpool1 及 Maxpool2,其作用是将第 1 和第 2 阶段输出的特征图的尺寸统一为  $4 \times 4$  像素。模型将尺寸统一的 3 个阶段的特征融合为交通标志的多尺度特征,并将多尺度特征送入全局平均池化层、全连接层进行交通标志分类。全局平均池化层的输入为:

$$\text{GlobalAveragePooling merge\_input} = [\text{Max\_11}, \text{Max\_22}, \text{Maxpool3}]$$

其中,[]表示将 Max\_11、Max\_22 和 Maxpool3 的输出信号串联形成一个张量。

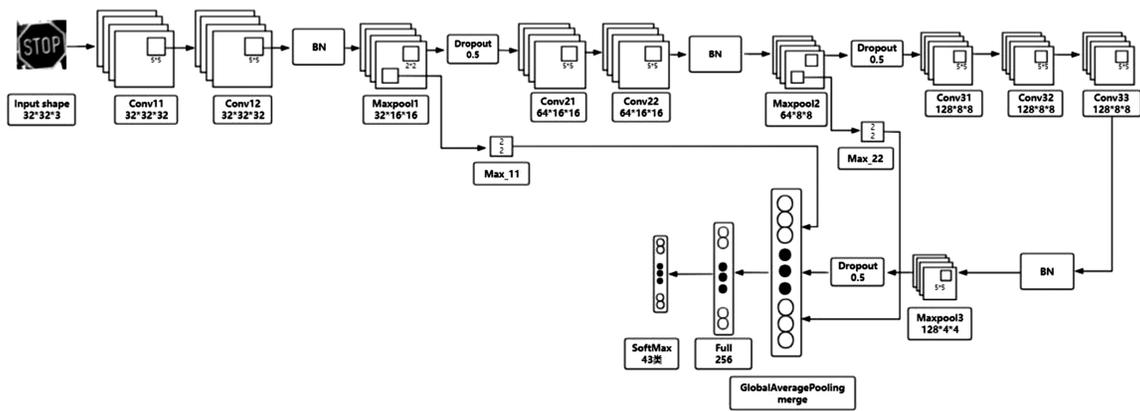


图 6 TSR-MSCNN 模型的结构

相比文献[6]的模型结构,本文提出的 TSR-MSCNN 模型做了如下改进:

1)改进了网络基本结构.①采用 3 阶段卷积神经网络结构代替 2 阶段卷积网络结构,以此增加网络深度,提高网络性能.②使用多个小卷积层串联代替单个较大卷积层,即在第 1 和第 2 阶段,用 2 个直接串联的  $5 \times 5$  小卷积网络替换  $9 \times 9$  的较大卷积层,在第 3 阶段用 3 个直接串联的  $5 \times 5$  小卷积网络替换  $13 \times 13$  的大卷积层,以此在保证算法感受野范围的同时,进一步增加网络深度以及减少网络的训练时间.③在每个子卷积网络中加入 Batch Normalization(BN)算法,以此克服内部协变量偏移的问题,加快网络收敛速度.④使用全局平均池化(global average pooling)代替平坦层(flatten),以此减少空间参数数量,使模型更加健壮,提高拟合效果.⑤在模型训练阶段,在保持输入输出神经元数量不变的情况下加入 Dropout 策略(每次随机关闭或忽略 50%隐藏层神经元),以此防止过拟合,提升网络泛化能力.

2)通过选比确定网络最佳参数.基本确定网络结构后,分别对全连接层神经元个数、Dropout 参数、卷积核尺寸等网络参数进行选比实验,由此最终确定的全连接层神经元个数为 256,Dropout 参数为 0.5,卷积核尺寸为  $5 \times 5$  像素.

#### 4 实验结果与分析

实验使用的计算机为 Windows10 操作系统,硬件环境为 Inter(R) Core(TM) i5-7300HQ CPU @2.50 GHz,内存为 8 G,显卡为 4 G.本文的实验

是基于 Keras 框架和 TensorFlow 开源库,并在 Jupyter NoteBook 开发环境中编程实现.

##### 4.1 数据增广的有效性验证

在保持 TSR-MSCNN 网络结构与参数不变的情况下,对数据集增广前后的模型准确性进行对比.模型采用 Adam 优化器和 Relu 激活函数,初始学习率为 0.001,每个实验中训练集迭代次数为 50,实验结果如表 1 所示.由表 1 中的数据可知,增广数据集后,TSR-MSCNN 模型的识别准确率达到 99.76%(原始数据集的识别准确率为 99.32%),由此说明增广数据集可大幅提高 TSR-MSCNN 模型的识别准确率.

表 1 数据增广前后的实验结果

数据集	损失	准确率 / %	训练时间/s	训练次数
增广数据集	0.005 2	99.76	752	50
原始数据集	0.005 6	99.32	736	50

##### 4.2 多尺度特征的有效性验证

在保持网络参数不变的情况下,分别采用单尺度特征和多尺度特征对图像分类特征的种类进行对比,结果如表 2 所示.由表 2 中的数据可知,多尺度特征的损失和识别准确率优于单尺度特征,且二者的训练时间相同.由此表明采用多尺度特征对交通标志进行识别是有效可行的.

表 2 不同分类特征的性能

特征类型	损失	准确率 / %	训练时间/s	训练次数
单尺度特征	0.007 1	99.64	752	50
多尺度特征	0.005 2	99.76	752	50

### 4.3 小卷积核级联的有效性验证

1)在其他网络参数不变的情况下,对单个大卷积核模型与 TSR-MSCNN 模型进行对比,即用 2 个级联的  $5 \times 5$  卷积核网络代替  $9 \times 9$  卷积核,3 个级联的  $5 \times 5$  卷积核网络代替  $13 \times 13$  卷积核,实验数据如表 3 所示.由表 3 中的数据可知,用小卷积核级联的 TSR-MSCNN 模型代替单个较大卷积核后,系统识别准确率得到有效提升,且系统总参数数量和网络训练时间大幅减小.由此表明,采用串联的小卷积核代替单个大卷积核对交通标志进行识别是有效可行的.

表 3 不同卷积层的性能

卷积层类型	损失	准确率/%	训练时间/s	训练次数	参数总量
大卷积核	0.006 6	99.62	653	50	1 627 435
小卷积核级联	0.005 2	99.76	752	50	1 250 699

2)在保持其他网络参数不变的情况下,通过改变网络中的卷积核尺寸进行尺寸选择实验.实验设置的卷积核尺寸分别为  $3 \times 3$ 、 $5 \times 5$ 、 $7 \times 7$ ,实验所得数据如表 4 所示.由表 4 可以看出:在  $3 \times 3$  卷积核网络中,模型的训练时间最少,但网络的损失最大,且算法的准确率最低;在  $7 \times 7$  卷积核网络中,模型训练花费的时间最多,但算法的准确率并不是最高;在采用  $5 \times 5$  卷积核网络中,模型

的训练时间虽然低于  $7 \times 7$  卷积核网络和高于  $3 \times 3$  卷积核网络,但网络的损失最小,识别准确率最高.因此本文在设计中选用  $5 \times 5$  像素的卷积核.

表 4 不同卷积核尺寸的性能

网络模型	损失	准确率/%	训练时间/s	训练次数
$3 \times 3$ 卷积核级联模型	0.006 7	99.69	456	50
$5 \times 5$ 卷积核级联模型	0.005 2	99.76	752	50
$7 \times 7$ 卷积核级联模型	0.005 6	99.72	1 162	50

### 4.4 BN 层的有效性验证

在保持相关网络参数不变的情况下,采用 3 种网络结构进行实验,结果如表 5 所示.由表 5 可知,每个阶段都加入 BN 层的网络结构,其训练时间虽然比未加入 BN 层的网络结构的训练时间有所增加,但其准确率最高.

图 7 为表 5 中 3 种网络结构训练的准确率和损失函数的变化曲线.由图 7(c)可知,在 3 个卷积阶段分别加入 BN 层后,其识别准确率和损失函数的变化程度低于图 7(a)和图 7(b).由此表明,在每个阶段分别加入 BN 算法对交通标志进行识别是有效可行的.

表 5 有无加入 BN 层的网络性能

有无加入 BN 层			网络结构	损失	准确率/%	训练时间/s	训练次数
Conv1	Conv2	Conv3					
无	无	无	1	0.005 4	99.69	725	50
有	有	无	2	0.005 3	99.71	752	50
有	有	有	3	0.005 2	99.76	752	50

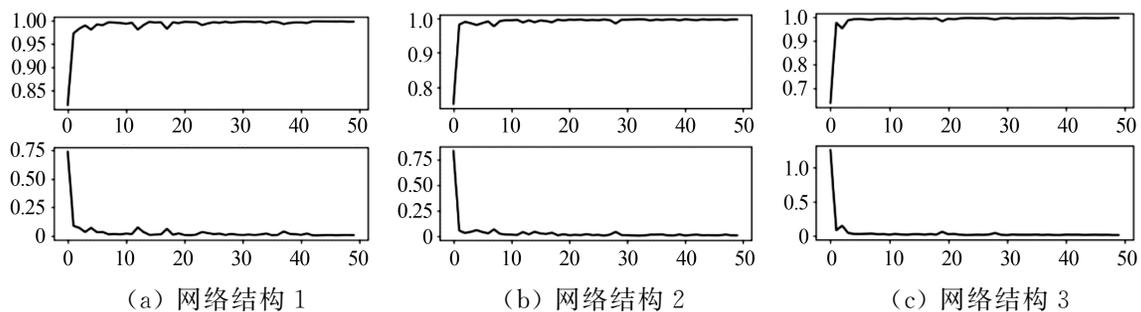


图 7 3 种网络结构训练的准确率和损失函数的变化曲线

#### 4.5 全局平均池化的有效性验证

在保持网络参数及其他网络结构不变的情况下,将模型中的 Flatten 层替换成全局平均池化层进行有效性对比,结果如表 6 所示.由表 6 中的数据可知,全局平均池化模型在损失、准确率、训练时间方面均优于 Flatten 模型,且二者的训练次数相同.由此表明,采用全局平均池化层来代替常规网络的平坦层对交通标志进行识别是有效可行的.

表 6 全局平均池化模型和 Flatten 模型的性能

模型	损失	准确率/ %	训练 时间/s	训练 次数
Flatten 模型	0.010 3	99.63	754	50
全局平均 池化模型	0.005 2	99.76	752	50

#### 4.6 网络参数选比实验

为了验证和提高算法的泛化能力,减少系统训练所需消耗的时间,本文在自定义数据集中完成网络参数选比实验,实验包括 Dropout 最佳参数的选比实验和全连接层参数的选比实验.

1) Dropout 最佳参数的选比实验.在 TSR-MSCNN 网络结构不变的前提下,改变 Dropout 参数设置进行选比实验,实验结果如表 7 所示.由表 7 中的数据可知:添加 Dropout 可以有效减少网络的训练时间.当将每层 Dropout 参数均设置为 0,即不添加 Dropout 层时,网络的识别准确率为 98.46%,但模型的训练时间最长(351 s);当将每层 Dropout 参数均设置为 0.25 时,网络的识别准确率最低;当将每层中 Dropout 参数均设置为 0.5 时,网络识别的准确率最高(98.85%),损失最小,且模型的训练时间最短(101 s).因此,本文在设计中将各层的 Dropout 参数均设置为 0.5.

表 7 Dropout 最佳参数的选比实验结果

Dropout 参数				损失	准确率/%	训练时间/s	训练次数
第 1 层	第 2 层	第 3 层	全连接层				
0	0	0	0	0.051 3	98.46	351	50
0	0	0	0.25	0.036 1	98.62	101	50
0	0	0	0.50	0.078 9	97.77	101	50
0.25	0.25	0.25	0.25	0.048 5	97.69	101	50
0.50	0.50	0.50	0.50	0.022 4	98.85	101	50

2)全连接层参数选比实验.在 TSR-MSCNN 网络结构不变的前提下,通过改变全连接层神经元的个数进行选比实验,实验结果如表 8 所示.由表 8 中的数据可知:全连接层神经元个数为 1024 时,所花费的训练时间最长;全连接层神经元个数为 512 时,损失最大,识别准确率最低;全连接层神经元个数为 256 时,算法的损失最小,识别准确率最高,且训练时间最短.因此,本文在设计中全连接层选用 256 个神经元.

表 8 全连接层最佳参数的选比实验结果

全连接层 神经元/个	训练时间 /s	训练 次数	损失	准确率/ %
128	101	50	0.051 3	98.46
256	101	50	0.022 4	98.85
512	101	50	0.065 1	98.00
1024	114	50	0.060 3	98.31

#### 4.7 网络结构对比实验

在 GTSRB 数据集上测试了传统三阶段模型(模型 1)、二次卷积特征连接模型(模型 2)、TSR-MSCNN 模型(模型 3)、文献[6]的二阶段多尺度卷积神经网络模型(模型 4)和文献[9]的三阶段多尺度特征分类算法(模型 5) 5 种网络结构的识别性能,结果见表 9.其中:模型 3 采用  $2 \times 2$  的最大池化层对不同阶段的卷积特征进行尺寸统一;模型 2 采用  $5 \times 5$  的二次卷积层对各分支的输出进行尺寸统一;模型 5 采用单一的  $5 \times 5$  卷积层提取 3 个阶段的卷积特征.

由表 9 可知,虽然模型 3 的训练参数数量略高于模型 5 和模型 1,但模型 3 在 5 个模型中的识别准确率最高(99.76%),损失最小,且训练时间最少;因此,综合来看,模型 3 的整体性能最优.

表9 不同网络模型的识别性能

网络模型	损失	准确率/%	训练时间/s	训练次数	训练参数
模型1	0.0071	99.64	752	50	1 250 699
模型2	0.0104	99.69	803	50	1 403 371
模型3	0.0052	99.76	752	50	1 275 275
模型4	—	98.97	—	—	1 437 791
模型5	0.0673	98.82	—	—	1 187 435

## 5 结论

本文以轻量级卷积神经网络为基础,结合多尺度特征概念和串联的小卷积核,提出了一种应用于交通标志识别的多尺度卷积神经网络模型TSR-MSCNN.经实验表明,本文算法在自定义数据集和GTSRB数据集上的识别准确率分别为98.85%和99.76%,且整体性能优于文献[6]的方法,因此本文方法在交通标志识别领域具有良好的潜在应用价值.本文算法的识别速率还有待进一步改善.在今后的工作中,我们将尝试在卷积层中加入模型压缩算法来提高图像的识别速度,以进一步提高本文算法的适用性.

## 参考文献:

- [1] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(6):1229-1251.
- [2] 李新,禹翼.基于SIFT算法的交通标志识别[J].制造业自动化,2012,34(5):10-12.
- [3] DOUVILLE P. Real-time classification of traffic signs[J]. Real-time Imaging, 2000,6(3):185-193.
- [4] MALDONADO-BASCON S, LAFUENTE-ARROYO S, GIL-JIMENEZ P, et al. Road-sign detection and recognition based on support vector machines[J]. Intelligent Transportation Systems, 2007, 8(2): 264-278.
- [5] 甘露,田丽华,李晨.基于融合特征和BP网络的交通标志识别方法[J].计算机工程与设计,2017,(38) 10:2783-2813.
- [6] SERMANET P, LECUN Y. Trafficsign recognition with multi-scale convolutional networks[C]//The 2011 International Joint Conference on Neural Networks (IJCNN). Washington DC: IEEE Computer Society, 2011:2809-2813.
- [7] STALLKAMP J, SCHLIPSING M, SALMEN J, et al. The German traffic sign recognition benchmark[EB/OL]. [2012-02-06]. <http://benchmark.ini.rub.de/?section=gtsrb&subsection=news>.
- [8] 王晓斌,黄金杰,刘文举.基于优化卷积神经网络结构的交通标志识别[J].计算机应用,2017,37(2): 530-534.
- [9] 宋青松,张超,田正鑫,等.基于多尺度卷积神经网络的交通标志识别[J].湖南大学学报(自然科学版),2018,45(8):131-137.
- [10] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2014-04-10]. <https://arxiv.org/abs/1409.1556>.
- [11] 仲会娟.基于CNN的多尺度特征在手写数字识别中的应用[J].绵阳师范学院学报,2019,11(5):22-26.
- [12] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2015:3431-3440.
- [13] 冀晓兵.图像尺寸变换的算法研究[D].西安:建筑科技大学信息与技术控制学院,2017.
- [14] 陈清江,李毅,柴昱洲.一种基于深度学习的多聚焦图像融合算法[J].激光与光电子学进展,2018, 55(7):246-254.
- [15] ZEILER M, FERGUS R. Visualizing and understanding convolutional networks[C]//13th European conference on Computer Vision. Zurich: Springer, 2014:818-833.