

多尺度特征融合与极限学习机结合的交通标志识别

马永杰, 程时升, 马芸婷, 陈敏

引用本文:

马永杰, 程时升, 马芸婷, 等. 多尺度特征融合与极限学习机结合的交通标志识别[J]. 液晶与显示, 2020, 35(6): 572–582.

MA Yong-jie, CHENG Shi-sheng, MA Yun-ting, et al. Traffic sign recognition based on multi-scale feature fusion and extreme learning machine[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(6): 572–582.

在线阅读 View online: <https://doi.org/10.3788/YJYXS20203506.0572>

您可能感兴趣的其他文章

Articles you may be interested in

基于深度可分离卷积的交通标志识别算法

Traffic sign recognition algorithm based on depthwise separable convolutions

液晶与显示. 2019, 34(12): 1191–1201 <https://doi.org/10.3788/YJYXS20193412.1191>

基于LeNet-5的卷积神经网络图像识别算法

Convolutional neural image recognition algorithm based on LeNet-5

液晶与显示. 2020, 35(5): 486–490 <https://doi.org/10.3788/YJYXS20203505.0486>

基于深度卷积神经网络的多特征融合的手势识别

Multi-feature fusion gesture recognition based on deep convolutional neural network

液晶与显示. 2019, 34(4): 417–422 <https://doi.org/10.3788/YJYXS20193404.0417>

基于深度学习的航空对地小目标检测

Detection of small target in aerial photography based on deep learning

液晶与显示. 2018, 33(9): 793–800 <https://doi.org/10.3788/YJYXS20183309.0793>

基于多尺度残差网络的CT图像超分辨率重建

CT image super-resolution reconstruction based on multi-scale residual network

液晶与显示. 2019, 34(10): 1006–1012 <https://doi.org/10.3788/YJYXS20193410.1006>

文章编号:1007-2780(2020)06-0572-11

多尺度特征融合与极限学习机结合的交通标志识别

马永杰*,程时升,马芸婷,陈 敏

(西北师范大学 物理与电子工程学院,甘肃 兰州 730070)

摘要:针对卷积神经网络只能使用相同尺寸图像和卷积核的网络进行特征提取,导致提取的特征不全面,在交通标志识别中因车载摄像头与交通标志的位置不断变化影响交通标志的识别精度等问题,提出了一种基于多尺度特征融合与极限学习机结合的交通标志识别方法。首先,将预训练适应 3 种不同尺寸图像的网络模型作为实验的初始模型;然后,融合 3 个网络模型构建多尺度卷积神经网络,将 3 个预训练网络的参数级联到融合模型的全连接层,对融合模型的全连接层进行训练,采用随机梯度下降算法更新网络参数;最后,将融合后的模型作为特征提取器提取特征,把提取到的多尺度特征送入极限学习机,实现交通标志识别。实验采用德国交通标志数据库(GTSRB)对算法性能进行测试,实验结果显示,多尺度特征融合与极限学习机结合的网络识别精度为 99.23%,识别速度为 46 ms。相对于预训练的网络,网络的分类精度分别提高了 2.35%,3.22%,3.74%。多尺度特征融合能够有效提取交通标志图像的特征信息,极限学习机可以提高分类精度和分类时间,该方法能满足交通标志识别的准确性和实时性的要求。

关键词:卷积神经网络;交通标志识别;多尺度融合;智能交通;极限学习机

中图分类号:TP391.7 **文献标识码:**A **doi:**10.3788/YJYXS20203506.0572

Traffic sign recognition based on multi-scale feature fusion and extreme learning machine

MA Yong-jie*, CHENG Shi-sheng, MA Yun-ting, CHEN Min

(School of Physics and Electronic Engineering, Northwest Normal University,
Lanzhou 730070, China)

Abstract: Feature extraction is performed on a convolutional neural network using only the same size image and convolution kernel, which results in incomplete extraction features. In the traffic sign recognition, the continuous change of the position between the vehicle camera and the traffic sign affects the recognition accuracy of the traffic sign. A traffic sign recognition method based on multi-scale feature fusion and extreme learning machine is proposed. Firstly, the network model adapted to three different size images is pre-trained as the initial model of the experiment. Then, the three network models are combined to construct a multi-scale convolutional neural network, and the parameters of the three pre-trained networks are cascaded to the full connection of the fusion model to train the fully

收稿日期:2019-11-05;修订日期:2020-02-02.

基金项目:国家自然科学基金(No.41461078)

Supported by National Natural Science Foundation of China(No.41461078)

* 通信联系人, E-mail: myjmyj@163.com

connected layer of the fusion model, and the network parameters are updated by the random gradient descent algorithm. Finally, the fusion model is used as the feature extractor to extract features, which are sent to the extreme learning machine(ELM) to realize traffic sign recognition. The experimental results show that the recognition accuracy of the multi-scale feature fusion and extreme learning machine combined network is 99.23% and the recognition speed is 46 ms. Compared with the pre-trained network, the classification accuracy of the network is increased by 2.35%, 3.22% and 3.74%, respectively. Therefore, the multi-scale feature fusion method can effectively extract the feature information of traffic sign images, ELM classifier can improve classification accuracy and classification time, and meet the requirements of traffic sign recognition accuracy and real-time.

Key words: convolutional neural network; identification of traffic signs; multi-scale fusion; intelligent transportation; extreme learning machine

1 引言

在智能交通系统(Intelligent traffic system, ITS)中,交通标志识别是其重要研究内容。由于交通标志识别是基于路面实况的复杂自然场景中的识别,很容易受到复杂光照、运动模糊、旋转倾斜、人为破坏、复杂环境背景干扰等条件的影响。

为了解决这些问题,大量的算法被提出。交通标志的识别主要由特征提取和分类器设计两部分组成,传统的特征提取算法主要依赖人工提取特征,如 HOG 特征(Histogram of Orientated Gradient)^[1]、Haar 特征^[2]、SIFT 特征(Scale-invariant feature transform)^[3]、LBP 特征(Local Binary Pattern)^[4]等。上述算法运用人工设计的特征提取算法,提取的特征都很有限,无法得到图像深层次的特征,在分类识别上有一定的局限性。

近年来随着硬件设备的发展,深度学习算法在计算机视觉领域得到很大的发展,很多分类方法不再使用人工特征提取算法,而是将彩色图像或者灰度图像直接输入到卷积神经网络(Convolutional Neural Network, CNN)当中,通过 CNN 自主学习图像特征,调整参数。刘占文^[5]等提出了一种基于图模型与卷积神经网络的交通标志识别,在对限速标志的识别分类中有很好的表现,取得了较好的识别效果。Zeng 等^[6]使用深度神经网络和极限学习机对交通标志进行识别,用深度神经网络进行特征提取,采用极限学习机(Extreme Learning Machine, ELM)对学习到的特征进行分类识别,识别准确率有所提高,但由于

仅利用了最后一层特征图,没有考虑到不同尺寸卷积核提取到的不同特征信息对分类的影响,对特征的表达能力不足。

传统 CNN 特征提取方法需要输入固定大小的图片,但实际上识别目标的尺寸是不相同的,且单一尺寸的卷积核容易丢失一些重要信息。在深度学习中,多尺度信息的输入可以防止设计的卷积神经网络陷入局部最优解,促进网络参数的更新,大量 CNN 研究工作及其应用都是基于多尺度方向展开的^[7-10]。

Liu 等^[11]提出了一种抽样轮廓波变换(Non-Subsampled Contourlet Transform, NSCT)和卷积神经网络的 X 射线图像多尺度融合框架,通过执行 NSCT 对输入图像进行预处理,提取丰富的特征集合,获得图像多尺度、多方向的表示。该方法与相应的空间域方法相比具有明显优势,在该领域的未来研究中具有很大的潜力。Xiang^[12]等设计了一个非对称多尺度卷积神经网络用于人蛋白图谱(Human Protein Atlas, HPA)分类,通过输入不同尺度的图像,设计了 3 组卷积层个数相同但卷积核大小不同的非对称卷积神经网络,该方法可以自动提取图像深度特征,实现多标签 HPA 识别。实验表明,相比于单一尺度,采用多尺度特征提取方法分类效果得到明显改善。Sérgio^[13]提出了一种基于卷积神经网络和传统特征提取器的手部姿态识别方法,并使用二进制、灰度和深度数据以及两种不同的验证技术进行了大量的实验,实验表明基于特征融合的卷积神经网络在验证技术和图像表示的组合中表现得更好。

在交通标志的识别中,车辆行驶中所采集到

的交通标志图像,由于相对位置的不断变化,图像尺寸也在不断地改变。因而本文提出了一种多尺度特征融合的 CNN 结构,通过将不同尺度的图像输入到适合图像尺寸的训练网络,提取不同尺度的特征并将其进行融合为一个新的网络,最后将融合网络的全连接层进行训练,得到具有更强表达能力的特征向量,再将得到的特征向量送入 ELM 分类器实现交通标志识别。

2 多尺度特征融合与 ELM 结合的交通标志识别方法

2.1 网络主体结构

交通标志图像是自然场景中的图像,容易受到光照、形变、拍摄角度以及人为涂抹破坏的影响,对交通标志识别研究具有一定的挑战。CNN 对于输入图像的平移、缩放、倾斜等形变具有较好的鲁棒性,但是 CNN 模型最终的识别效果会因为输入图像尺寸的改变而发生变化^[14]。图像金字塔是把图像表示为一系列图像集合,并且分辨率是逐渐降低的,可用于对数据集的扩增,实际也是一种多尺度的一种体现^[15]。受图像金字塔的启发,采用不同尺寸交通标志图像作为输入,可以使模型具有更强的鲁棒性。在深度学习中,多尺度的信息输入再融合,可以防止算法求解过程中陷入局部最优解,有利于网络参数的更新,并且可以获得图像更多的信

息,增强模型的鲁棒性^[16]。融合模型的网络参数较为庞大,ELM 算法在训练时需要调整的参数小,训练时间会减少^[17]。

因此,本文提出多尺度特征融合与 ELM 结合的交通标志识别算法,首先将交通标志图像进行预处理,增强图像的局部特征,然后将图像分为 3 个尺寸输入到对应的网络之中,并在 3 个网络中分别进行适合图像尺寸的卷积和下采样操作,将 3 个网络上得到的特征在全连接层进行融合,最后将融合网络得到的特征向量送入 ELM 分类器进行识别。为适应不同尺寸图像的特征,所设计网络的卷积核大小不同,网络深度也不同,大尺寸的卷积核通过卷积操作可以得到更加边缘和全局的特征,而小尺寸的卷积核可以得到纹理更加突出的局部特征。

多尺度卷积神经网络的算法步骤包括:

(1)对交通标志图像进行预处理,增强图像的局部特征,然后归一化为 3 个尺寸图像输入网络,预训练网络模型。

(2)根据输入图像尺寸大小的不同设计适合交通标志识别的卷积核尺寸和网络深度。

(3)在预训练网络模型的第一个全连接层,将 3 个网络生成的特征图利用串行融合的方法进行融合,再将融合后的网络进行微调,训练融合网络的全连接层。

(4)将融合网络得到的特征向量输入 ELM 分类器进行交通标志识别,网络结构如图 1 所示。

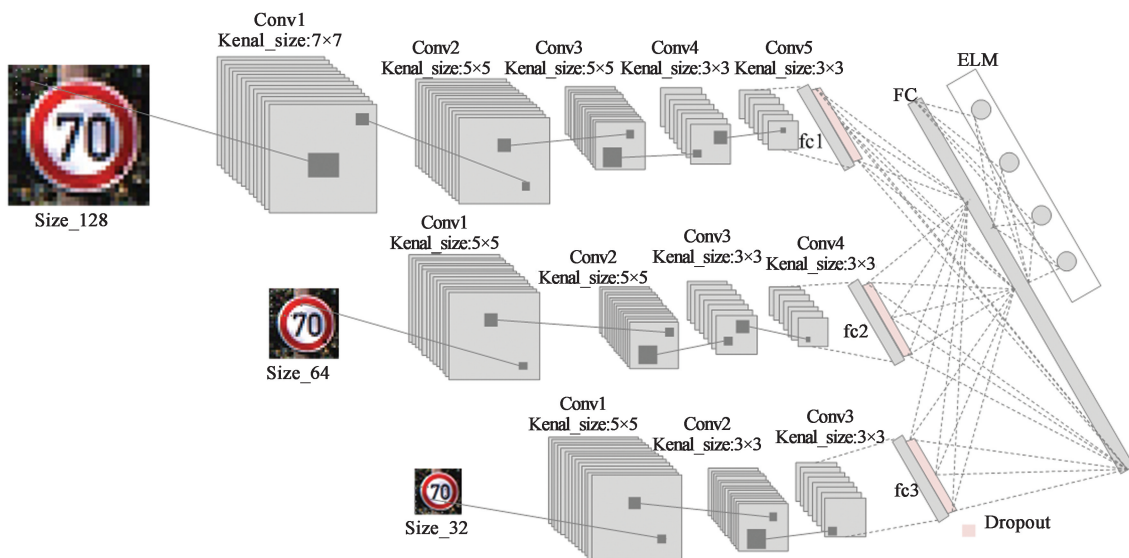


图 1 多尺度卷积神经网络的结构

Fig.1 Structure of multi-scale convolution neural networks

2.2 数据预处理

实验数据集采用的是 2011 年在神经网络国际联合会议上,波鸿大学公布的由 INI-RTCV 组织建立的德国交通标志数据库(GTSRB)。该数据库是车载高清摄像头在驾驶场景的视频中截取出来的,包含 43 类交通标志,如图 2 所示。其中训练集有 39 029 幅,测试集有 12 630 幅,总计为 51 839 幅图像。图片质量参差不齐,受到自然场景光照、图片采集角度变化、人为涂抹污染等诸多条件的影响,可以检验算法的实时性与鲁棒性。数据预处理包含数据增强和图像尺寸归一化两部分内容。



图 2 GTSRB 数据库的 43 类交通标志

Fig.2 43 classes traffic signs in GTSRB dataset

2.2.1 数据增强

数据增强的目的是凸出图像的细节部分,提高图像的清晰度,直接对彩色图像进行数据增强可以保持图像更多的梯度特征,使网络提取到的特征更加全面。直方图变换增强可以避免使用灰度化之后的图像丢失图像颜色特征,因此本文选择使用直方图均衡化进行图像增强,如图 3 所示。

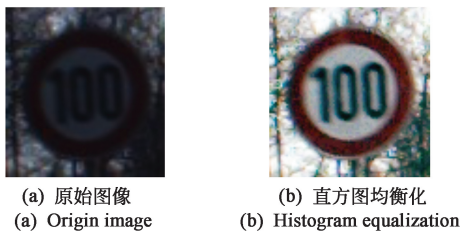


图 3 原始图像与直方图均衡化对比

Fig.3 Original image and histogram equalization

2.2.2 图像尺寸归一化

GTSRB 数据库中图片从自然场景中采集,由于拍摄角度的变化,使交通标志图像尺寸大小参差不齐,而卷积神经网络在输入图片时需要将图像尺寸归一化,不同尺寸的图像对于网络的分类能力会有影响,考虑运算量并尽量保留更多的

特征,本文选择双线性插值法对图像进行归一化处理。把图像归一化为 32×32 , 64×64 , 128×128 三组不同尺寸的图像作为数据集输入相应的预训练网络。

2.3 多尺度 CNN 特征提取

卷积神经网络主要由可见层和隐藏层组成。可见层指的是输入、输出层,隐藏层包括卷积层、下采样层和全连接层。本文算法进行多尺度融合卷积,所以应该着重考虑输入层尺寸的选择和卷积层设计,为适应不同尺寸图片的特征提取,选择不同大小的卷积核。

(1)在模型设计中,首先把图像预处理成大小不同的 3 种尺寸,即 32×32 , 64×64 , 128×128 , 以不同的尺寸分别输入到 3 个预训练 CNN 之中,并且根据尺寸大小的不同设计相应的卷积层和网络深度。 32×32 的图像用 3 个卷积层进行特征提取,第一层卷积层采用 5×5 的卷积核,后两层采用 3×3 的卷积核。 64×64 的图像用 4 个卷积层进行提取特征,前 2 层卷积层采用 5×5 的卷积核,后 2 层采用 3×3 的卷积核。 128×128 的图像用 5 个卷积层提取特征,第一层卷积层采用 7×7 的卷积核,中间两层采用 5×5 的卷积核,最后两层采用 3×3 的卷积核。卷积层上的每个特征图都代表提取到的一组特征,并且每个特征图的每个神经元都与上一层的局部感受野相连接,实现权值共享。卷积层的计算公式为:

$$x_j^l = f(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l) \quad (1)$$

式中, M_j 指的是所有图像特征的集合, k_{ij} 是卷积核, b_j 是每个图像特征的偏移量。不同深度的网络对特征提取的能力不同,并且不同尺寸的卷积核所提取到的图像信息也不同,具体网络参数如表 1 所示。

(2)在卷积层得到特征响应图之后用 ReLU 激活函数进行非线性表达映射到下一层。ReLU 是一种从底部进行半修正的函数,其数学公式是:

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

当输入 $x \leq 0$ 时,输出为 0 值,当 $x > 0$ 时,输出为输入值 x 。该函数为非饱和函数,解决了梯度消失问题。

(3)由于下采样层具有降低特征维度的作用,但是又不改变特征图的数目,并且下采样层一定程度增强了对图像位移、缩放和扭曲等形

表 1 多尺度卷积神经网络模型参数设置

Tab.1 Parameter setting of multi-scale convolution neural network model

Sub-network	Type	Filter size/stride	Pooling size/stride	Pooling model	Layer output
Size_32	Conv1/pooling1	$5 \times 5/1$	$2 \times 2/2$	Max	$48 \times 15 \times 15$
	Conv2/pooling2	$3 \times 3/1$	$2 \times 2/2$	Max	$64 \times 8 \times 8$
	Conv3	$3 \times 3/1$	$2 \times 2/2$	Max	$84 \times 4 \times 4$
Size_64	Conv1/pooling1	$5 \times 5/1$	$2 \times 2/2$	Max	$84 \times 31 \times 31$
	Conv2/pooling2	$5 \times 5/1$	$2 \times 2/2$	Max	$128 \times 15 \times 15$
	Conv3	$3 \times 3/1$	--	--	$384 \times 13 \times 13$
	Conv4	$3 \times 3/1$	$2 \times 2/2$	Max	$384 \times 6 \times 6$
	Conv1/pooling1	$7 \times 7/2$	$3 \times 3/2$	Max	$96 \times 31 \times 31$
Size_128	Conv2/pooling2	$5 \times 5/1$	$2 \times 2/2$	Max	$128 \times 15 \times 15$
	Conv3	$5 \times 5/1$	--	--	$256 \times 11 \times 11$
	Conv4	$3 \times 3/1$	--	--	$384 \times 9 \times 9$
	Conv5/pooling3	$3 \times 3/1$	$2 \times 2/2$	Max	$384 \times 4 \times 4$

变的鲁棒性,所以在经过非线性映射之后将特征图输入下采样层。下采样操作的方式有最大池化法和平均池化法两种,如图 4 所示。本文选择最大池化法进行下采样操作。最大池化法是提取图像区域中最大值作为该区域池化后的值,并且在进行池化之后,网络的输出深度不变。下采样的公式为:

$$x_j^l = f(\beta_j^l d(x_j^{l-1}) + b_j^l) \quad (3)$$

式中, d 表示一个下采样函数, β_j 为权重系数, b_j 为偏置系数。

(4) 在 3 个预训练网络的第一个全连接层之后添加一个 dropout 层,随机删除网络一部分神经元,可以降低网络的参数,并预防和降低网络的过拟合。

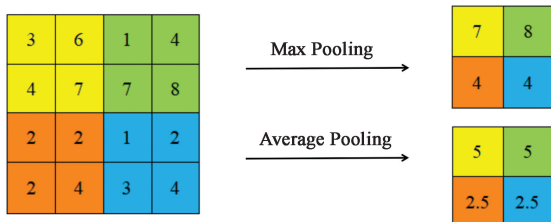


图 4 最大池化和平均池化

Fig.4 Max pooling and average pooling

2.4 多尺度特征融合

多尺度融合卷积神经网络的具体参数如表 1

所示,其中 Conv 表示该层只有一个卷积层,Conv/pooling 表示该卷积层之后还包含一个池化层,每个卷积层之后都有一个 ReLU 激活函数层,多尺度特征级联方式用公式(4)实现:

$$F_{MCNNFC} = \sum_{i=1}^n F_i \quad (4)$$

式中: F_i 为不同的预训练网络的全连接层的特征, n 为用于特征级联的预训练 CNN 的个数。特征融合在 3 个预训练网络的第一个全连接层进行,融合的方法为特征层叠加,每两个预训练网络的全连接层之间用一个 Concat 层进行特征拼接,本文选择在通道上进行拼接,拼接实现方式如下:

$$Z_{concat} = \sum_i^c X_i * K_i + \sum_i^c Y_i * K_{i+c} \quad (5)$$

对于两路输入来说,如果通道数相同且后面带卷积,则式中 X_i 和 Y_i 分别代表两个不同网络的其中任意的一个通道, K 代表卷积核, Z_{concat} 代表单个输出通道的拼接。

$$\begin{cases} F_1 = N \times C_1 \times H \times W \\ F_2 = N \times C_2 \times H \times W \\ F_3 = N \times C_3 \times H \times W \\ F_{MCNN} = N \times (C_1 + C_2 + C_3) \times H \times W \end{cases} \quad (6)$$

式中 N 为特征图的数量, F_1 、 F_2 、 F_3 分别表示本

文3个预训练网络的全连接层, C_1 、 C_2 、 C_3 分别代表3个预训练网络中的某一个通路, F_{MCNN} 代表在通道上进行拼接得到的全连接层的特征图, H 和 W 分别代表特征图的高和宽。

将拼接后的特征输入到融合网络的全连接层,并微调融合网络的全连接层,最后将融合的特征送入ELM分类器进行分类。多尺度特征融合网络有很多优点,不同尺寸图像设计适合的卷积核和网络深度,大尺寸的卷积核提取图像的粗粒度特征,可以保留交通标志的轮廓特征;小尺度卷积核有助于细粒度特征的提取,局部纹理特征在特征图中也可以很好地展现。不同尺寸图像特征图的组合输入,粗粒度与细粒度的结合,提高了网络的识别精度和表达能力。

2.5 ELM分类器

通常卷积神经网络使用 Softmax 分类器对图像进行分类,在易混淆的图像中表现不佳。多尺度特征融合网络参数较大,特征向量维度高于单一尺度特征。考虑到模型的高效性,需要选择一种计算量小、训练时间短、满足模型高效性的分类器。ELM是一种新型的快速学习算法,对于单隐层神经网络,ELM可以随机初始化输入权重和偏置并得到相应的输出权重,不需要重复调整权重和阈值,降低了计算复杂度,训练时间更快,保证了模型的高效性,其结构如图5所示。

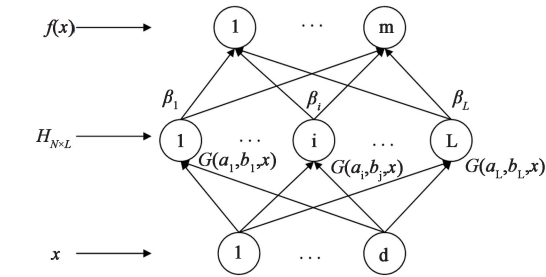


图5 ELM分类器结构

Fig.5 ELM classifier structure

由图5可以看出,ELM由输入层、隐含层、输出层3层组成,其中输入 x 为多尺度特征融合网络得到的特征向量, d 为输入的维度($x \in \mathbf{R}^d$, $x = (x_1, \dots, x_d)^T$),特征向量 x 在隐层被映射为向量 $(G(a_1, b_1, x), \dots, G(a_L, b_L, x))^T$ 。其中 $G(a_i, b_i, x)$ 是第 i 个加性隐节点的输出,其计算公式如下:

$$G(a_i, b_i, x) = g(a_i \cdot x + b_i),$$

$$a_i \in \mathbf{R}^d, b_i \in \mathbf{R}, \quad (7)$$

式中: g 为激活函数, a_i 表示第 i 个隐形节点与所有输入节点之间的权重, b_i 表示偏置, $i = 1, 2, \dots, L$ 。输入的交通标志特征向量在隐层经过线性变换后,输出一个 m 维的向量 $f(x)$,其公式如下, m 相当于交通标志识别的43个类别。

$$f(x) = \sum_{i=1}^L \beta_i G(a_i, b_i, x), \beta_i \in \mathbf{R}^m, \quad (8)$$

ELM的训练集合为 $\{(x_j, t_j) \mid x_j \in \mathbf{R}^d, t_j \in \mathbf{R}^m, j = 1, \dots, N\}$,式中 x_j 为融合网络输出的多尺度交通标志特征向量, t_j 为 x_j 的标签,在对ELM分类器进行训练时,每个交通标志的特征向量 x_j 传入ELM分类器中, t_j 为期待输出结果,并且每个标签向量 t_j 都有一个特征向量 x_j 对应。对于回归分类算法,标签 t_j 直接代表了模型对于输入向量 x_j 的响应,随机初始化输入权重和偏置($a_i, b_i, i = 1, \dots, L$),并得到相应的输出权重,隐含层和输出层之间的连接权值($\beta_1, \dots, \beta_L$)不需要迭代调整,而是通过解方程组一次性确定,公式如下:

$$H\beta = T, \quad (9)$$

其中 H 和 β 分别为:

$$H = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix} = [h(x_1), \dots, h(x_N)]_{L \times N}^T,$$

$$h(x) = \begin{bmatrix} G(a_1, b_1, x) \\ \vdots \\ G(a_L, b_L, x) \end{bmatrix}_{L \times 1},$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}, \quad (10)$$

式中: H 是隐层节点的输出, β 为输出权重, T 为期望输出。

β^* 的最小二乘解可以利用MP广义逆解析得到,并且具有最小范数:

$$\beta^* = H^\dagger T, \quad (11)$$

其中, $H^\dagger$ 是矩阵 H 的Moore-Penrose广义逆。最终可以得到ELM的输出方程:

$$f(x) = \beta^{*T} h(x), \quad (12)$$

综上所述,ELM 的训练与测试过程如图 6 所示。

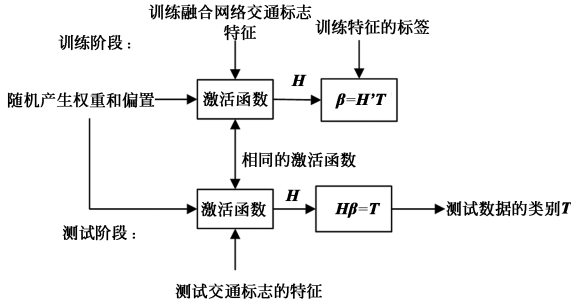


图 6 ELM 的训练与测试过程

Fig.6 ELM training and testing process

3 实验结果与分析

3.1 实验方案

(1)将不同尺寸图像输入网络进行预训练,画出各个预训练网络的训练曲线,确保该预训练网络的合理性,并可用于模型融合。

(2)多尺度特征级联。将预训练好的 3 个网络的全连接层进行特征级联,并进行比选实验,设计不同的特征级联方式,选择出分类性能最好的一组。

(3)将得到的特征向量分别输入到支持向量机,Softmax 分类器中进行分类识别,并与本文使用的 ELM 分类器进行对比。

(4)针对 GTSRB 数据集,采用不同的算法和本文算法进行分类,比较总体分类精度和分类时间,评估算法性能的优劣。

3.2 实验环境

实验所用的计算机配置是 I5-7500 处理器,3.4 GHz 主频,16 GB 内存,Nvidia GeForceGT730 显卡,Ubuntu 16.04 版本的 Linux 操作系统,深度学习框架为 Caffe^[18],软件编程环境为 Python 2.7。

在实验设施配置差别不大、识别速度相近的情况下,识别准确率 K_{Acc} 作为评价算法性能的主要指标:

$$K_{Acc} = \frac{\sum_{i=1}^N n_i}{N} \times 100\% \quad , \quad (13)$$

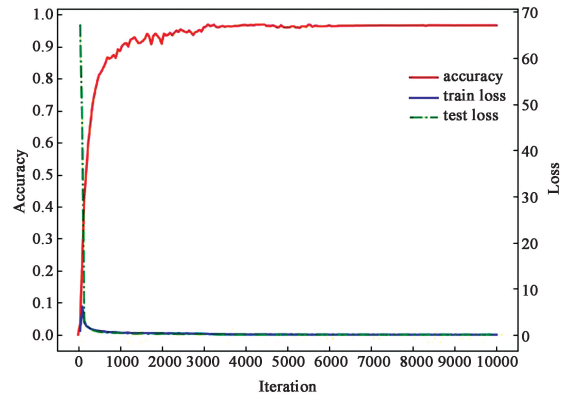
式中: n_i 代表识别正确的样本, N 代表测试的样本总量。

3.3 实验结果及分析

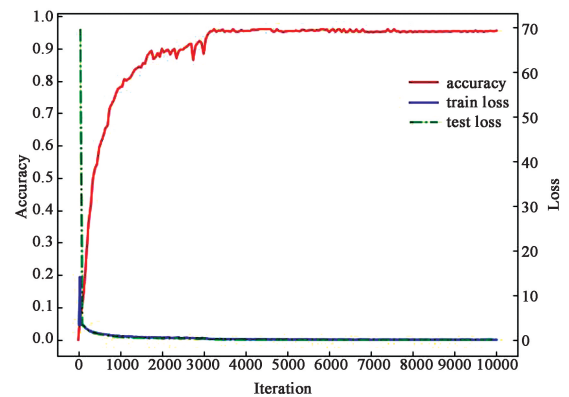
3.3.1 网络模型的预训练

用德国交通标志数据库(GTSRB)对 3 个网络进行预训练,3 个网络均使用随机梯度下降算法进行参数更新,训练网络模型的参数配置相同,基础学习率为 0.01,防止过拟合的权重衰减项为 0.000 5,学习率变化因子为 0.1,最大迭代次数为 10 000,学习率调整策略为 step,调整策略参数 step_size 设置为 3 000,进行 3 次学习率的调整,通过训练损失值和测试损失值来判断预训练的网络是否存在过拟合现象,以保证融合模型的精度。图 7 显示了 4 个模型的分精度、训练损失值和测试损失值变化曲线,其中图 7(a)为输入图片尺寸为 32×32 的网络性能曲线,图 7(b)为输入图片尺寸为 64×64 的网络性能曲线,图 7(c)为输入图片尺寸为 128×128 的网络性能曲线,7(d)为融合模型的网络性能曲线。

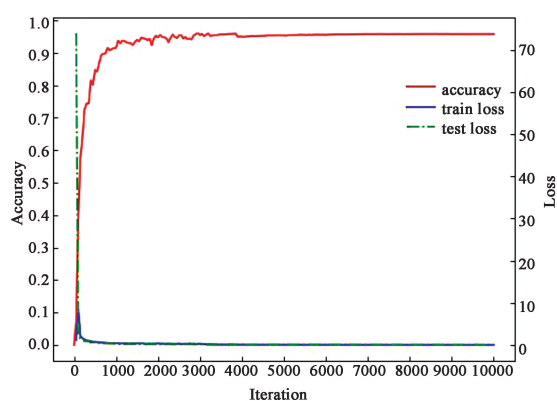
由图 7 可以看出,预训练模型随着迭代次数的不断增加,分类精度都是先快速增长,再逐渐趋于平稳,表明预训练的模型有较好的泛化能力,可



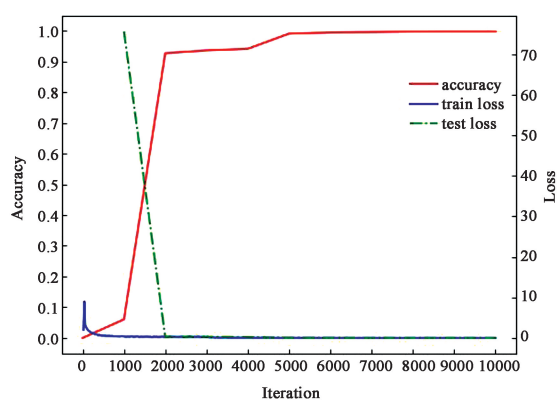
(a) Size_32



(b) Size_64



(c) Size_128



(d) Multi-scale

图7 损失值和分类精度

Fig.7 Cross entropy loss and classification accuracy

用于模型的融合。损失值也是在前几次迭代过程中快速降低之后再趋于平稳,最终基本保持不变,表明模型的学习能力逐渐下降,分类精度提升趋于平稳。融合后的模型,测试损失值的下降速度虽然不如预训练模型,但是分类精度更高,达到99.32%后逐渐趋于平稳。

3.3.2 特征级联

由上文所述的级联公式(4),将网络进行级联,根据级联 CNN 的数量分为二网络级联和三网络级联,再将级联网络的全连接层进行二次训练。表2为不同级联方法分类结果。

从表中可以看出,在单独的网络中输入尺寸为 32×32 的图像时识别精度较高,这是因为该交通标志数据集的图像分辨率在 15×15 到 222×193 像素。统计结果显示,像素值大小的中位数为 41×40 ,所以网络在该尺寸的识别率较高。融合的网络相对于单一网络分类的准确率都有所提高,表明不同网络之间的相互组合可

表2 不同级联方法分类结果

Tab.2 Classification results of different cascading method

Type of network	Classification Accuracy/%	Classification time/ms
Size_32	96.88	40
Size_64	96.01	40
Size_128	95.49	40
Size_32+Size_64	97.88	40
Size_32+Size_128	96.98	42
Size_64+Size_128	97.24	44
Size_32+Size_64+Size_128	99.23	46

以提高识别精度。

3.3.3 与其他分类器的对比

为了验证本文所示用的分类器的优势,设计了对比实验,将融合后得到的特征向量,分别输入到 SVM 分类器和 Softmax 分类器中,与本文 ELM 分类器从分类精度和分类时间两个方面进行对比,实验结果如表3所示。

表3 不同分类器性能比较

Tab.3 Performance comparison of different classifiers

分类器	分类精度/%	分类时间/ms
Softmax	99.12	52
SVM	98.26	60
ELM	99.23	46

从表3可以看出,ELM 分类器的分类精度相较于 SVM 和 Softmax 分类器有所提高。这是由于 ELM 模型具有较低的复杂度,在相同时间内可处理更多的数据,增强了模型的泛化能力。此外,由于 ELM 分类器输入层和隐含层之间的参数采用随机初始化,且之后无需再进行调整,因此,ELM 分类器学习速度更快,使得 ELM 分类器具有更快的分类速度。

3.3.4 与其他分类算法的对比

由表4可知,本文模型相对于 ANN 和 Random forests 两种算法的识别精度都有一定提高,并且分类时间都大幅度下降。相对于文献[19]和文献[20],分类精度提升不多,但是在分类时间上仍然具有一定优势。

表 4 不同方法在 GTSRB 数据集识别结果对比

Tab.4 Comparison of different methods in GTSRB dataset recognition results

Algorithm	Classification time/ms	Accuracy rate / %
ANN	178	89.63
Random forests	211	96.14
Algorithm of Ref.[19]	93	98.82
Algorithm of Ref.[20]	84	98.13
本文算法	46	99.23

文献[19]为验证算法的鲁棒性和在复杂环境下的分类性能,在原测试集中选取了光照不足、受到遮挡和图像模糊的图像各 30 幅组成新的测试集,并且在单层特征和多层特征下对新的测试集进行了验证。本文用相同的方法选取测试集对算法进行验证,并且对新的测试集不进行图像增强处理,与文献[19]的对比,结果如表 5 所示。

表 5 复杂环境下识别分类精度对比

Tab.5 Comparison of recognition classification accuracy in complex environment

Algorithm	Insufficient illumination/ %	Obscured/ %	Blurred image/ %
SCF-ELM of Ref.[19]	95.01	93.72	94.63
MCF-ELM of Ref.[19]	97.15	96.01	96.68
Algorithm of this paper	97.62	97.18	98.01

与表 4 中本文算法的分类准确率相比,在数据集都是复杂环境下的图像时,分类精度有所下降(表 5),并且在受到遮挡的条件下精度下降较为明显,但是在整体情况下相对于文献[19]中的单层特征提取和多层特征提取算法都有所提升,特别在图像模糊的条件下提升较为明显。

本文采用多尺度图像输入,间接地增加了数据集的数量,并使用不同的 3 个网络对不同尺寸的图像进行预训练,卷积核大小也是根据数据集输入图片大小所设计,最后在融合网络进行分类。

多尺度的信息输入再融合,可以防止算法求解过程中陷入局部最优解,有利于网络参数的更新,并且可以获得图像更多的信息,增强模型的鲁棒性^[16,21-23]。在交通标志图像受环境影响较大时,特别是对于拍摄形变造成模糊的图像,本文设计的多尺度卷积融合网络在复杂环境下的鲁棒性相比于文献[19]中的多层特征融合更加优越。

表 6 为本文算法与其他多尺度 CNN 方法和单一尺度 CNN 算法的性能对比结果。文献[20]采用的多尺度交通标志识别方法是将每一个池化层的输出都施加到后端全连接层,再利用 ELM 分类器进行分类。

表 6 与其他多尺度方法的对比

Tab.6 Comparison with other multi-scale methods

Method	Parameters/ million	Loss	Accuracy rate / %
Algorithm of Ref.[20]	1.18	0.067 3	98.82
Single-scale CNN of Ref.[20]	0.92	0.093 5	98.38
Algorithm of this paper	6.8	0.008 3	99.23

从表 6 可以看出,由于本文使用 3 个预训练网络进行融合,所以参数规模要比文献[20]中的多尺度方法和单一尺度方法大,但是本文算法的损失值要远比文献[20]中的方法小很多,说明本文的多尺度融合模型的泛化能力和交通标志的特征表达能力要优于文献[20]中的多尺度模型,并且在准确率上也高出 0.41 %。

本文设计的多尺度卷积融合算法相对于文献[20]中的多尺度方法可以得到更高的准确率和网络模型的泛化能力,原因在于:

(1)在输入上使用 3 种不同尺寸图像,CNN 模型最终的识别效果会因为输入图像尺寸的改变而发生变化^[14],相对于文献[20]中把图片归一化成为 32×32 尺寸的输入,多尺度信息的输入可以增强模型的性能。

(2)在卷积层的设计上,文献[20]设计了 3 层卷积,并且每个卷积核的大小都是 5×5。本文所设计网络的卷积核大小不同,每个预训练

网络深度也不同。大尺寸的卷积核提取图像的粗粒度特征,可以保留交通标志的轮廓特征。小尺度卷积核有助于细粒度特征的提取,可以提取交通标志的局部纹理特征。不同尺寸图像特征图的组合输入,粗粒度与细粒度的结合,相比于单个网络的多尺度提取特征,融合网络多尺度的性能更好。

(3)在网络设计上,本文虽然是由3个预训练网络模型组合得到最后的特征图,但最终识别是在融合网络上完成,并且在融合网络的全连接层进行二次训练,再输入ELM分类器进行分类。相比于每一个池化层的输出都施加到后端全连接层,再利用Softmax分类器进行分类,本文得到的特征图数目更多,信息也更全面,并且选择ELM作为分类器对于分类精度的提高也起到了一定作用。

因此,本文多尺度特征融合与ELM结合的

算法相较于文献[20]性能有一定的提升。

4 结 论

交通标志识别是智能交通系统的重要组成部分,自然场景中的交通标志受环境影响较为复杂,采集到的交通标志尺寸大小也是多样化的,因此本文提出一种多尺度特征融合与ELM结合的交通标志识别方法。将单一尺度的图像经过图像增强,使图像细节特征更加明显,再分别将尺寸归一化成3种不同的尺寸,然后构建适应图像尺寸的3个网络进行预训练,最后将3个网络中学习到的特征进行融合,对融合网络模型进行微调,将融合模型得到的多尺度特征输入ELM分类器进行分类。图像提取到的特征更加全面,分类识别的准确率大幅提升。实验通过在GTSRB交通标志数据集上进行实验,识别率达到了99.23%。

参 考 文 献:

- [1] DALAI N, TRIGGS B. Histograms of oriented gradients for human detection [C]//*IEEE Computer Society Conference on Computer Vision & Pattern Recognition*. San Diego, CA: IEEE, 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*. Kauai, HI, USA: IEEE, 2001: 511-518.
- [3] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [4] OJALA T, PIETIKÄINEN M, HARWOOD D. A comparative study of texture measures with classification based on featured distributions [J]. *Pattern Recognition*, 1996, 29(1): 51-59.
- [5] 刘占文,赵祥模,李强,等.基于图模型与卷积神经网络的交通标志识别方法 [J]. *交通运输工程学报*, 2016, 16(5): 122-131.
- LIU Z W, ZHAO X M, LI Q, *et al.* Traffic sign recognition method based on graphical model and convolutional neural network [J]. *Journal of Traffic and Transportation Engineering*, 2016, 16(5): 122-131. (in Chinese)
- [6] ZENG Y, XU X, FANG Y, *et al.* Traffic sign recognition using deep convolutional networks and extreme learning machine [C]. 5th International Conference on Intelligent Science and Big Data Engineering. Suzhou, China: Springer, 2015: 272-280.
- [7] LIU S, QI L, QIN H F, *et al.* Path aggregation network for instance segmentation [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 8759-8768.
- [8] LIN D, SHEN D G, SHEN S T, *et al.* ZigZagNet: Fusing top-down and bottom-up context for object segmentation [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 7482-7491.
- [9] QIN Z, LI Z M, ZHANG Z N, *et al.* ThunderNet: Towards real-time generic object detection [J]. *ArXiv*, 2019: 1903.11752.
- [10] ZHAO Q J, SHENG T, WANG Y T, *et al.* M2Det: A single-shot object detector based on multi-level feature pyramid network [J]. *ArXiv*, 2019: 1811.04533.
- [11] LIU Y, ZHANG C, CHENG J, *et al.* A multi-scale data fusion framework for bone age assessment with convolutional neural networks [J]. *Computers in Biology and Medicine*, 2019, 108: 161-173.

- [12] XIANG S, LIANG Q K, HU Y C, *et al.* AMC-Net: Asymmetric and multi-scale convolutional neural network for multi-label HPA classification [J]. *Computer Methods and Programs in Biomedicine*, 2019, 178: 275-287.
- [13] CHEVTCHEENKO S F, VALE R F, MACARIO V, *et al.* A convolutional neural network with feature fusion for real-time hand posture recognition [J]. *Applied Soft Computing*, 2018, 73: 748-766.
- [14] 张俊. 基于 AlexNet 融合特征的图像检索研究[D]. 重庆: 重庆邮电大学, 2016.
ZHANG J. Image retrieval based on AlexNet fusion feature [D]. Chongqing: Chongqing University of Posts and Telecommunications, 2016. (in Chinese)
- [15] 林志洁, 罗壮, 赵磊, 等. 特征金字塔多尺度全卷积目标检测算法[J]. 浙江大学学报: 工学版, 2019, 53(3): 533-540.
LIN Z J, LUO Z, ZHAO L, *et al.* Multi-scale convolution target detection algorithm with feature pyramid [J]. *Journal of Zhejiang University: Engineering Science*, 2019, 53(3): 533-540. (in Chinese)
- [16] 余焯, 金强, 傅云翔, 等. 基于 Fg-CarNet 的车辆型号精细分类研究[J]. 自动化学报, 2018, 44(10): 1864-1875.
YU W, JIN Q, FU Y X, *et al.* Fine-grained classification of car models using Fg-CarNet convolutional neural network [J]. *Acta Automatica Sinica*, 2018, 44(10): 1864-1875. (in Chinese)
- [17] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: Theory and applications [J]. *Neurocomputing*, 2006, 70(1/3): 489-501.
- [18] 欧先锋, 向灿群, 郭龙源, 等. 基于 Caffe 深度学习框架的车牌数字字符识别算法研究[J]. 四川大学学报: 自然科学版, 2017, 54(5): 971-977.
OU X F, XIANG C Q, GUO L Y, *et al.* Research of recognition of digital characters on vehicle license based on Caffe deep learning framework [J]. *Journal of Sichuan University: Natural Science Edition*, 2017, 54(5): 971-977. (in Chinese)
- [19] 孙伟, 杜宏吉, 张小瑞, 等. 基于 CNN 多层特征和 ELM 的交通标志识别[J]. 电子科技大学学报, 2018, 47(3): 343-349.
SUN W, DU H J, ZHANG X R, *et al.* Traffic sign recognition method based on multi-layer feature CNN and extreme learning machine [J]. *Journal of University of Electronic Science and Technology of China*, 2018, 47(3): 343-349. (in Chinese)
- [20] 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.
SONG Q S, ZHANG C, TIAN Z X, *et al.* Traffic sign recognition based on multi-scale convolutional neural network [J]. *Journal of Hunan University: Natural Sciences*, 2018, 45(8): 131-137. (in Chinese)
- [21] 梁浩, 刘克俭, 刘康, 等. 引入再检测机制的孪生神经网络目标跟踪 [J]. 光学 精密工程, 2019, 27(7): 1621-1631.
LIANG H, LIU K J, LIU K, *et al.* Siamese network tracking with redetection mechanism [J]. *Optics and Precision Engineering*, 2019, 27(7): 1621-1631. (in Chinese)
- [22] 王军, 何昕, 魏仲慧, 等. 基于多特征匹配的快速星图识别算法 [J]. 光学 精密工程, 2019, 27(8): 1870-1879. Jun
WANG J, HE X, WEI Z H, *et al.* Fast star identification algorithm based on multi-feature matching [J]. *Optics and Precision Engineering*, 2019, 27(8): 1870-1879. (in Chinese)
- [23] 王广龙, 田杰, 朱文杰, 等. 特征融合和自适应权重更新相结合的运动模糊目标跟踪 [J]. 光学 精密工程, 2019, 27(5): 1158-1166.
WANG G L, TIAN J, ZHU W J, *et al.* Feature fusion and weight adaptive updating based motion blur object tracking [J]. *Optics and Precision Engineering*, 2019, 27(5): 1158-1166. (in Chinese)

作者简介:



马永杰(1967—),男,甘肃灵台人,博士,教授,2011 年于兰州交通大学获得博士学位,主要从事图像处理、人工智能、测控技术等方面的研究。E-mail:myjmyj@163.com