Static Sign Language Recognition Method Based on Wireless Sensor Technology and Convolutional Neural Network*

LÜ Jun[1*], QIANG Yan[2]
(1.Department of Computer Science and Technology, Lüliang University, Lüliang Shanxi 033000, China;
2.College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan Shanxi 030024, China)

Abstract: The complexity of gestures increases the difficulty of sign language recognition. Therefore, in order to obtain more accurate static sign language recognition results, a static sign language recognition method based on wireless sensor technology and convolutional neural network is proposed. The static sign language image is filtered by using wireless sensor technology. The static sign language features are extracted by using feature extraction method and fused. The fused continuous two-dimensional feature image is direct transformed into one-dimensional vector, which is used as the input of convolutional neural network, and the static sign language recognition is completed through iterative training. The experimental results show that the accuracy of the proposed method is higher than 94.11% in different scenes, and the recognition speed is fast. Therefore, it is verified that the proposed method can complete the static sign language recognition quickly and accurately.

Key words: wireless sensor technology; convolutional neural network; static sign language; image filtering; gesture recognition
EEACC: 7230 **doi**: 10.3969/j.issn.1004-1699.2023.04.017

基于无线传感技术与卷积神经网络的静态手语识别方法*

吕 军[1*], 强 彦[2]
(1.吕梁学院计算机科学与技术系, 山西 吕梁 033000; 2.太原理工大学计算机科学与技术学院, 山西 太原 030024)

摘 要: 基于传感网络的手语识别随着手势复杂程度的提升, 识别的难度增大, 因此, 为了获取更加精准的静态手语识别结果, 提出了基于无线传感技术与卷积神经网络的静态手语识别方法。通过无线传感技术对静态手语图像进行滤波处理, 采用特征提取方法提取静态手语特征并进行融合处理, 将融合后的连续二维特征图像直接转换为一维向量, 将其作为卷积神经网络的输入, 通过迭代训练完成静态手语识别。实验结果表明, 所提方法在不同场景下静态手语识别精度高于 94.11%, 且识别速度较快, 由此验证了所提方法可以快速准确完成静态手语识别。

关键词: 无线传感技术; 卷积神经网络; 静态手语; 图像滤波; 手语识别
中图分类号: TP391 **文献标识码**: A **文章编号**: 1004-1699(2023)04-0623-06

随着科学技术和互联网技术的飞速发展, 在人机交互过程中, 手语占据越来越重要的社会地位, 尤其是和聋哑人的交流。健全人对手语相关知识了解比较少, 所以两者的交流存在诸多不便。静态手语识别的主要目的是采用先进的计算机技术将手语信息转换为可以理解的文本信息, 保证所有人均可以理解手语所表达的含义。现阶段, 通过简单的手语就可以完成人机交互, 所以手语识别具有十分重要的意义。

张淑军等[1]从孤立词与连续语句两个分支阐述了近年来基于深度学习的手语识别技术, 并以目前国内外常用手语数据集为研究对象, 探讨了手语识别技术的研究挑战与发展趋势, 但是该方法尚未应用于实际中进行检验。何坚等[2]在双流融合网络中引入有效通道注意力, 并利用空间卷积网络和时间卷积网络提取的人体姿态特征, 将时序特征融合后识别手势, 但是该方法受通道注意力的影响较高。缪永伟等[3]优先在手语图像数据库提取不同类型手语的关节信息, 采用手部关节点位置提取手指相关特征信息, 并将其输入到支持向量机多分类器中完成手语识别。该方法具有分类识别准确率较高的优势, 但是该方法受手语动作所处环境影响较大。Arun 等[4]通过边缘和肤色检测将手部从图像中分割出来后, 利用离散余弦变换等方法从分割后的手部图像中获得鲁棒特征, 引入 Deep-Elman 递归神经网络分类器完成静态手语识别。该方法的识别精度受手语动作分割精度影响较大。米娜瓦尔·阿不拉等[5]主要通过计算视觉技术

项目来源: 虚拟现实技术与系统国家重点实验室开放基金项目(BUAA-VR-17KF-14); 虚拟现实技术与系统国家重点实验室开放基金(VRLAB2018A08)
收稿日期: 2022-04-12 **修改日期**: 2022-06-06

和深度学习技术完成手语识别,但是该方法尚未应用于实际中进行检验。Bao 等[6]为了进一步提高卷积神经网络分类器在表面肌电图信号中识别手势的可靠性,建立了基于 CNN 后验概率的置信度估计模型来生成置信度得分,并设计了一个目标函数来训练该模型的参数。

本文在识别手语前,利用无线传感技术对手语图像进行了滤波处理,剔除了其中的噪声,以免噪声对最终的静态手语识别质量造成影响,出现沟通障碍的问题。并在此基础上结合卷积神经网络方法实现静态手语识别。从不同输入图像尺寸和不同环境下的手语识别精度以及识别速度三个方面进行了实验,实验结果表明,所提方法不仅可以有效提升识别速度,同时还可以获取比较高的静态手语识别精度。

1 静态手语识别方法设计

1.1 基于无线传感技术的静态手语图像滤波处理

为了有效消除手语图像中的干扰噪声[7],提升静态手语图像的质量。将传感器噪声模式设定为一种和高斯白噪声近似的随机信号,使其在频域可以获取比较平坦的频率。考虑到传统噪声是由图像采集过程中或者采集设备自身因素而导致的,而上述两种噪声则是随机产生的,因此,静态手语图像中的噪声属于一种随机信号,拥有比较平坦的频谱,当出现插值或者压缩的情况,则会出现多峰现象。为了有效剔除静态手语图像中的噪声[8-9],本文引入无线传感技术完成图像滤波处理,其详细的操作步骤如图 1 所示。

设定静态手语图像中的噪声是一种类似高斯白噪声的信号,将噪声经过处理转为线性。利用高斯白噪声对静态手语图像进行平滑处理,不仅可以保留图像细节信息,同时还可以剔除无利用价值的信息,最终获取相对平坦的噪声频率。因此依据图 1 对静态手语图像滤波处理。

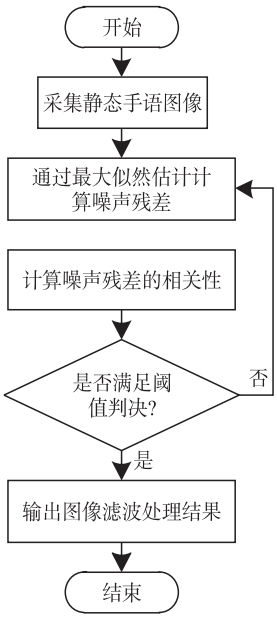


图 1 基于无线传感技术的静态手语图像滤波处理流程图

通过单摄像头采集不同类型的静态手语图像,确定对应的噪声模式,经过计算获取图像对应的噪声残差,同时引入最大似然估计方法计算相机传感器模式噪声,如式(1)所示:

$$\begin{cases} H_i = O_i - G(O_i) \\ S = \dfrac{\sum_{i=1}^{n} H_i O_i}{\sum_{j=1}^{n} H_i (G)^2} \end{cases} \tag{1}$$

式中:H_i 代表噪声残差;O_i 代表静态手语图像频谱;G 代表平坦频谱;S 代表相机传感模式噪声;i 和 j 分别代表第 i 幅和第 j 幅图像。

通过传感器模式噪声可以获取全部待检测图像的噪声残差,同时计算噪声残差的相关性。同时采用噪声残差相关性实行阈值判决,假设大于阈值,则说明静态手语图像来自相机;反之,则说明不是。

在实际应用过程中,由于受到引导滤波等相关理论的启发,需要通过选取相同尺寸的噪声模式对静态手语图像进行平滑处理。通过以上操作就可以获取高斯白噪声的输出,如式(2)所示:

$$L = T(F, Y) \tag{2}$$

式中:L 代表输出的高斯白噪声;T 代表传感器模式噪声;F 代表高斯白噪声;Y 代表平滑处理过程。

在具体平滑处理过程中,从局部出发,设定局部矩形窗区域内全部像素满足线性变换需求,如式(3)所示:

$$L = e_k H_i(G) + (c_k m_k)^2 \tag{3}$$

式中:e_k、c_k 和 m_k 代表线性系数。

通过上述分析可知,静态手语图像在经过处理之后,需要有效保留自身特性。所以,以下主要通过最小化处理后的噪声求解对应的线性系数,如式(4)所示:

$$\begin{cases} e_k = \dfrac{1}{2} H_i(G) \mid L - O_i \mid \\ c_k = \dfrac{1}{2} \sum_{i=1}^{n} [O_i - G(O_i)]^2 \\ m_k = \sum_{j=1}^{n} \dfrac{O_i - G(O_i)}{L} \end{cases} \tag{4}$$

使用高斯白噪声平滑处理的主要目的是确保输出的静态手语图像中不仅可以有效保留高斯白噪声,同时还可以获取比较完整的高斯频谱。为了满

足以上需求,需要考虑像素点和周围邻域像素的影响,所以需要在平滑处理过程中设定窗口大小,主要是为了避免得到的频谱取值比较小。

对于凸函数 $B(a_t,b_t)$ 而言,在偏导取值为 0 的情况下,可以采用式(5)表示:

$$\begin{cases}\alpha_k=\dfrac{c_kB(a_t,b_t)}{\omega}\\ \beta_k=\dfrac{m_kB(a_t,b_t)}{\omega}\end{cases}\tag{5}$$

式中:α_k 和 β_k 代表局部窗口区域系数;ω 代表均值。在求解和统计局部窗口区域系数过程中仍然存在大量未知的变量,需要通过窗口尺寸[10]和惩罚因子[11]调整得到最优变量。另外,还需要确定各个变量的取值范围。

通过上述分析,完成无线传感技术对静态手语图像的滤波处理,最终得到滤波处理的图像。

1.2　基于卷积神经网络的静态手语识别

当今社会人工智能的大部分成果离不开图像识别领域的贡献,在图像识别领域内,深度学习占据十分重要的地位。尤其是卷积神经网络,通过卷积运算原理,采用比较大的卷积核完成卷积运算的计算量和参数量,利用特征图完成卷积运算,则卷积运算的计算量 Q 和参数量 P 可以表示为式(6):

$$\begin{cases}Q=A_d\times B_d\times C_d\times E_d\\ P=M\times B_d\times C_d\times E_d\end{cases}\tag{6}$$

式中:M 代表卷积核尺寸;A_d、B_d、C_d 和 E_d 分别代表不同的特征图。

在使用卷积神经网络[12-13]展开静态手语识别前,需要对全部特征展开卷积以及池化等相关操作,最终输出分类结果,具体过程如图 2 所示。

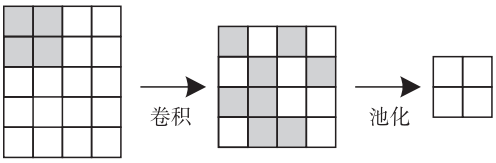


图 2　单层卷积神经网络

卷积神经网络(Convolutional Neural Networks, CNN)中参数共享是通过核确保运算过程中只包含一个学习参数,进而通过共享组成特征图,有效降低网络维数以及参数数量。

卷积层的输入是经卷积核卷积后的图像。卷积核的主要目的是对构建的矩阵进行求和处理,同时将权值经过一系列卷积操作求解对应的特征值,将特征值融合处理组建特征图,并将其传递给池化层。其中,卷积层对应的计算结果为式(7):

$$d_j^l=f\left(\sum_{i=1}^{n}d_i^{(l-1)}\otimes h_{ij}^l+c_j^l\right)\tag{7}$$

式中:d_j^l 代表卷积层计算结果;f 代表激活函数;h_{ij}^l 代表偏置项;c_j^l 代表输入图像;$\otimes$ 代表卷积运算。

在实际应用过程中,多种特征是通过不同的卷积核形成的结果。一般情况下,需要采用卷积核处理之后获取手语特征图,完成特征分类处理。另外,还需要注意计算时间较长以及过拟合情况的发生。

卷积神经网络的全连接层包含 Softmax 分类器[14-15],Softmax 分类器存在的主要目的是解决各种类型分类问题。设定分类结果对应的计算概率为 R,样本数量为 r,则对应的函数可以表示为式(8):

$$\boldsymbol{h}_{(x,y)}(i)=\begin{bmatrix}p(y^{(i)}=1\mid x^{(1)},\theta)\\ p(y^{(i)}=2\mid x^{(2)},\theta)\\ \vdots\\ p(y^{(i)}=r\mid x^{(i)},\theta)\end{bmatrix}\tag{8}$$

式中:矩阵 $\boldsymbol{h}_{(x,y)}(i)$ 代表分类函数;其含义是将输出的分类结果映射到(0-1)之间,将神经网络的分类结果转化成对应的概率。p 代表分类数量;θ 代表学习参数。

式(9)是基于式(8),为了防止过拟合而给出损失函数对应的表达式:

$$N(\theta)=\frac{1}{r}\left\{\sum_{i=1}^{n}\sum_{j=1}^{r}\left[\lg\left(\sum_{i=1}d_i^{(l-1)}\otimes h_{ij}^l+c_j^l\right)\right]\right\}\tag{9}$$

式中:$N(\theta)$ 代表损失函数。

通过梯度下降法求解损失函数的最优解 $\nabla N(\theta)$,求解结果为式(10):

$$\nabla N(\theta)=-\frac{1}{r}\left\{\sum_{i=1}^{n}\sum_{j=1}^{r}\left[\lg\left(\frac{c_km_kB(a_t,b_t)}{\omega}\right)\right]\right\}\tag{10}$$

在上述分析的基础上,需要提取不同静态手语图像的特征,同时将全部提取到的特征融合处理,将经过融合处理之后的二维图像特征直接转换为一维向量。

在二维图像特征转换过程中,需要 CNN 网络中的池化层充分发挥池化作用,使用最大池化功能实现一维向量转换,式(11)为池化层使用最大池化功能:

$$a_j^l=\nabla N(\theta)\times\mathrm{down}(a_j^{l-1})\tag{11}$$

式中:a_j^{l-1} 代表上一个卷积层的输出结果,down(·)代表下采样。通过分析最大池化功能,将大小为 $k\times k$ 的结果中的上限值重复抽取,并获得新的一维向量特征。

根据上述过程获取一维向量,并且将其设定为 CNN 网络的输入,实现静态手语识别,具体流程如图 3 所示。

根据图 3 所示的过程,获取手语一维向量特征,并通过卷积神经网络训练达到静态手语识别的目的。

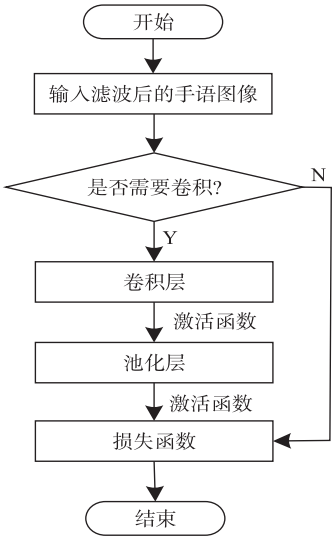


图 3 静态手语识别流程

2 实验及结果分析

为了验证所提方法的有效性,将文献[3]和文献[4]作为对比方法进行静态手语识别实验。

2.1 实验准备

实验前通过摄像机采集数据,同时构建手语数据库,并从中选取 50 个使用频率比较高的手语作为测试样本集,如图 4 所示。

将每个手语做成 10 个时长为 3 s 的视频。随机在每个视频中选择 10 张手势图像,设置 CNN 模型输入的手语图像大小为 512×512,其中卷积层采用 5×5 卷积核,步长为 1;激活函数采用 ReLU 函数;池化层采用 Max-Pooling,且窗口大小为 4×4,步长为 1,输出结果为一维向量。按照实验数据的占比不同将作出两种分配方式。实验数据配置如表 1 所示。

图 4 部分静态手语图像测试样本集

除测试样本集和图像训练准备外,为了保证所提方法的有效性,首先通过迭代训练获取所提方法的最优训练参数值,结果如表 2 所示。

表 1 实验数据配置

	训练集	验证机	样本集
数据比例	40	10	50
个数/组	140	20	160

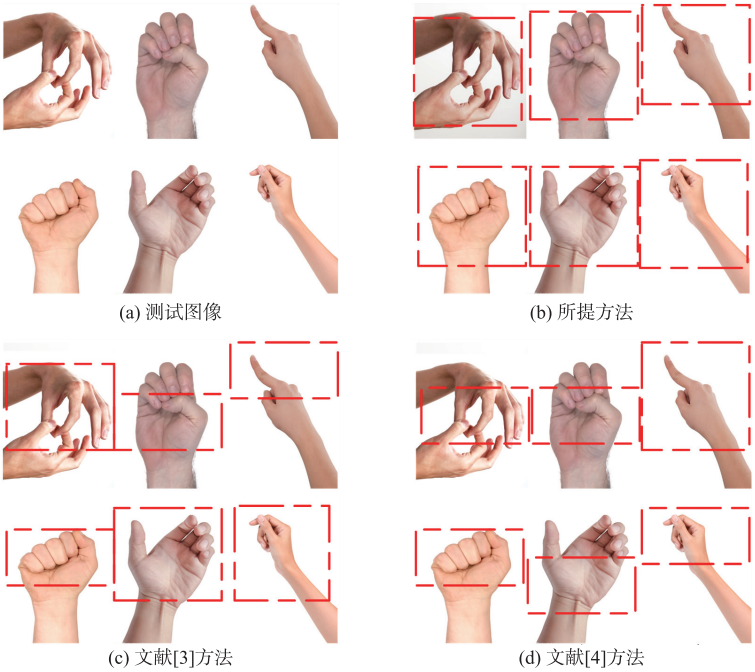
表 2 最优训练参数值

参数	数值	参数	数值
基础学习率	0.003	权重衰减	0.000 8
学习率调整策略	Step	学习率变化比率	0.1
学习率变化时迭代次数	1 500	总迭代次数	35

以上述参数取值为基础,以图 4 为目标识别图像开展下述实验。

2.2 结果分析

为了验证所提方法的识别性能,采用三种不同的方法对图 3 中的手语图像识别,测试结果如图 5 所示。



(a) 测试图像 (b) 所提方法

(c) 文献[3]方法 (d) 文献[4]方法

图 5 不同方法的静态手语图像识别结果对比分析

分析图 5 可知,对同一图像进行识别时,所提方法可以准确识别不同的静态手语图像,而文献[3]和文献[4]方法在识别过程中,存在部分关节特征缺失的情况,由此验证了所提方法的优越性。

为了进一步分析所提方法的识别精度,在不同图像尺寸下测试三种方法的识别精度,测试结果如图 6 所示。

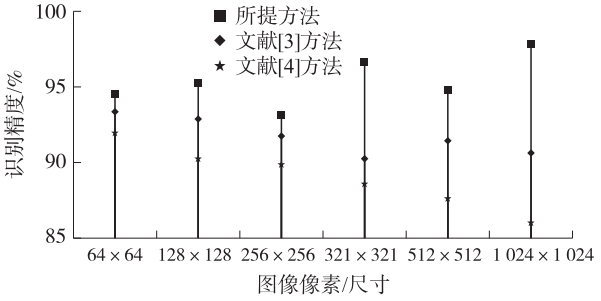


图 6　不同输入图像尺寸的识别精度对比结果

分析图 6 可知,三种方法的识别精度均高于 86%,其中,所提方法的识别精度最高,最高识别精度为 97.85%;文献[3]方法的识别精度次之,文献[4]方法的识别精度最低。

进一步分析灰暗、背景简单、明亮且复杂、灰暗且简单、灰暗且复杂和明亮且简单等不同环境下各个方法的识别精度变化情况。其中,灰暗环境为图像饱和度趋近于 0;背景简单为图像中仅包含手势动作,无其他背景干扰因素;明亮且复杂为图像对比度高且背景中包含了除手势动作外因素的图像;灰暗且简单为饱和度趋近于 0 且仅包含手势动作,无其他背景干扰因素的图像;灰暗且复杂为饱和度趋近于 0 且包含了除手势动作外因素的图像;明亮且简单为图像对比度高且仅包含手势动作,无其他背景干扰因素的图像。结果如表 3 所示。

表 3　不同环境下各个方法的识别精度测试结果对比

测试环境	识别精度/%		
	所提方法	文献[3]方法	文献[4]方法
灰暗	95.85	93.25	91.14
背景简单	94.74	92.85	89.01
明亮且复杂	96.36	90.77	87.20
灰暗且简单	94.11	91.20	86.25
灰暗且复杂	96.85	90.36	85.14
明亮且简单	97.75	92.55	84.45

分析表 3 可知,所提方法在识别手语过程中,受测试环境的影响较小,均可以保持较高精度的识别结果,文献[3]方法次之,文献[4]方法相对受环境影响较大。本文方法受环境影响较小的原因是,在识别手语动作前,利用无线传感技术对图像进行滤波处理,最大程度地降低了图像噪声的干扰。

分析三种不同方法的识别速度变化情况,结果如图 7 所示。

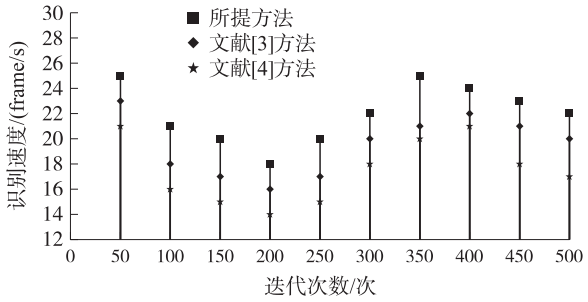


图 7　不同方法的识别速度测试结果对比

分析图 7 可知,所提方法的识别速度明显优于文献[3]和文献[4]方法,识别速度高于 18 frame/s,其主要原因是,本文在识别静态手语前,利用无线传感技术对图像进行了滤波处理,减少了噪声的干扰,并通过卷积神经网络训练优化,提高了静态手语识别效率。

3　结束语

本文结合无线传感技术和卷积神经网络的优势,提出了静态手语识别方法。实验结果表明,所提方法可以获取比较满意的静态手语识别结果,同时可以加快识别速度。在后续研究过程中,将进一步扩大研究范围,使其可以被广泛应用于不同的研究领域中,重点针对以下几方面内容展开研究:

①手语是一种比较复杂的人体语言,所提方法在研究过程中只是简单展开了研究,后续将改进和完善分类器算法,确保得到的识别结果更加理想。

②后续可以将多传感器融合技术引入到静态手语识别方法中,使其可以更高效完成不同类型的手语识别。

参考文献:

[1]　张淑军,张群,李辉. 基于深度学习的手语识别综述[J]. 电子与信息学报,2020,42(4):1021-1032.

[2]　何坚,刘炎,祖天奇. 时空特征与通道注意力融合的视觉手语识别技术[J]. 北京工业大学学报,2021,47(8):824-832.

[3]　缪永伟,李佳颖,刘家宗,等. 融合关节旋转特征和指尖距离特征的手语识别[J]. 计算机学报,2020,43(1):78-92.

[4]　Arun C, Gopikakumari R. Optimization of both Classifier and Fusion Based Feature Set for Static American Sign Language Recognition[J]. IET Image Processing,2020,14(22):2101-2109.

[5]　米娜瓦尔·阿不拉,阿里甫·库尔班,解启娜,等. 手语识别方法与技术综述[J]. 计算机工程与应用,2021,57(18):1-12.

[6]　Bao T,Zaidi S A R,Xie S Q,et al. CNN Confidence Estimation for Rejection-Based Hand Gesture Classification in Myoelectric Control

[J]. IEEE Transactions on Human-Machine Systems, 2022, 52(1):99-109,.

[7] Liu B, Wu L, Hao H, et al. Interferometric Phase Image Denoising Method via Residual Learning[J]. Journal of Electronic Imaging, 2021, 30(2):023013.

[8] Wang S, Xia K, Wang L, et al. Improved RPCA Method via weighted non-Convex Regularization for Image Denoising[J]. IET Signal Processing, 2020, 14(5):269-277.

[9] 陈波,鄢爱兰. 改进型多项式匹配图像去噪算法的研究与应用[J]. 电子测量与仪器学报,2021,35(4):177-186.

[10] 曹林,于威威. 基于图像分割的自适应窗口双目立体匹配算法研究[J]. 计算机科学,2021,48(S2):314-318.

[11] 李彦,赵其峰,闫河,等. Canny 算子在 PCBA 目标边缘提取中的优化应用[J]. 光学精密工程,2020,28(9):2096-2102.

[12] 朱幸辉,周勇. 基于卷积神经网络的番茄叶片病斑识别仿真[J]. 计算机仿真,2021,38(7):481-485.

[13] 邓棋,雷印杰,田锋. 用于肺炎图像分类的优化卷积神经网络方法[J]. 计算机应用,2020,40(1):71-76.

[14] 胡宏宇,刘家瑞,高菲,等. 基于一维卷积神经网络的驾驶人身份识别方法[J]. 中国公路学报,2020,33(8):195-203.

[15] 吴雪,宋晓茹,高嵩,等. 基于数据增强的卷积神经网络火灾识别[J]. 科学技术与工程,2020,20(3):1113-1117.

吕　军(1982—),男,汉族,山西兴县人,硕士,副教授,研究方向:计算机网络、人工智能,lvjun13215@163.com;

强　彦(1969—),男,汉族,山西古交人,博士,教授,研究方向:图像识别以及高级人工智能。