文章编号：2095-2163(2021)07-0086-05 中图分类号：TP183 文献标志码：A

基于 Transformer 改进 YOLO v4 的火灾检测方法

王国睿

（山东科技大学 计算机科学与工程学院，山东 青岛 266590）

摘 要：针对火灾检测算法检测多尺度火焰和烟雾精度低，且实时性差的问题，提出了一种基于 Transformer 改进 YOLO v4 的火灾检测方法。首先，结合 MHSA（Multi-Head Self-Attention）改进了 CSPDarknet53 主干网络，建模全局依赖关系以充分利用上下文信息。此外，基于 MHSA 改进了 PANet 模块进行多尺度特征图融合，获取更多的细节特征。为验证改进方法的有效性，与 YOLO v4、YOLO v3 等算法进行比较。实验证明，不仅能够检测多尺度目标，且视频监控场景下达到实时性，具有准确率高、误报率低、检测实时性等优点，满足监控视频场景下的火灾检测任务。

关键词：深度学习；注意力机制；YOLO v4 算法；火灾检测

Fire detection method based on Transformer improved YOLO v4

WANG Guorui

（College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao Shandong 266590, China）

【Abstract】 Aiming at the problem of low accuracy and poor real-time performance of the fire detection algorithms in detecting multi-scale flames and smoke, a fire detection method based on Transformer improved YOLO v4 is proposed. First, combined with MHSA (Multi-Head Self-Attention) to improve the CSPDarknet53 backbone network, global dependencies is modeled to make full use of context information. In addition, based on MHSA, the PANet module is improved to perform multi-scale feature map fusion to obtain more detailed features. In order to verify the effectiveness of the improved method, it is compared with YOLO v4, YOLO v3 and other algorithms. Experiments have proved that it can not only detect multi-scale targets, but also achieve real-time performance in video surveillance scenarios. It has the advantages of high accuracy, low false alarm rate, and real-time detection, which can meet the fire detection tasks in surveillance video scenarios.

【Key words】 deep learning; attention mechanism; YOLO v4 algorithm; fire detection

0 引 言

随着社会的不断发展，各类灾害对公共安全与社会财富的危险性也相应地有所增加，其中火灾较为常见，防范与及时发现火灾越来越受到重视。传统的火灾检测方法，通常是采集温度、烟雾传感器数据进行火灾检测，缺点是误报率比较高、实时性较差。基于图像识别的火灾检测方式，因其具有响应快、事后追溯直观等特点，被广泛应用于监控视频场景下的火灾检测与实时报警任务。

近年来，深度学习技术在图像分类、目标检测等计算机视觉领域得到广泛应用，并取得丰硕的研究成果。基于深度学习的火灾检测方法主要通过 CNN 进行特征提取获取火灾图像特征，然后进行分类与回归获得检测结果。文献[1]提出基于改进 YOLO v3[2] 的火灾检测与识别方法，通过改进 YOLO v3 解决小目标识别性能不足的问题。文献[3]提出嵌入 DenseNet[4] 结构和空洞卷积模块改进 YOLO v3 的火灾检测方法，通过在 Darknet-53[5] 中嵌入空洞卷积模块来扩展感受野，提升对多尺度目标火灾的特征提取效果，其本质是充分利用上下文信息。文献[6]采用 Anchor-Free 网络结构的实时火灾检测算法，优点是避免了 Anchor 方法中超参数过多、网络结构复杂的缺点，主干网络选取 MobileNetV2[7]，同时引入了特征选择模块。上述火灾检测方法存在以下问题：

（1）主干网络多为图像分类任务设计的，未针对目标检测任务对进行优化，导致算法缺乏鲁棒性。

（2）通过堆叠卷积模块扩展网络深度，虽然获得良好的检测效果，但难以达到实时性。

（3）针对火灾小尺度目标检测任务性能不足。

在此基础上，通过借鉴 Bottleneck Transformer[8] 算法设计思想，提出了一种改进 YOLO v4[9] 的火灾检测方法，主要改进点如下：

（1）在原 CSPDarknet53[10] 中引入了 MHSA（Multi-Head Self-Attention）层，有效地将目标之间

作者简介：王国睿（2000-），男，本科生，主要研究方向：图像视觉。

收稿日期：2021-04-27

的信息与位置感知相关联，增强网络全局依赖关系建模的能力，充分利用多尺度上下文信息，提升火灾小目标的检测能力。

（2）采用同样的方式对 PANet[11] 模块进行优化，改善多尺度特征融合能力，获取更多特征细节。

实验表明，改进的 YOLO v4 算法在监控视频场景下检测精度达到 94%，检测速度达到 26 帧/s，优于现有的其他火灾检测算法，满足监控视频场景下的火灾检测。

1　YOLO v4 与 MHSA 原理

1.1　YOLO v4 算法原理

YOLO v4 算法是一种端到端的实时目标检测框架，其网络结构如图 1 所示，该网络主要包括 CSPDarknet53、SPP 附加模块[12]、PANet 路径聚合模块、YOLO v3 头部。

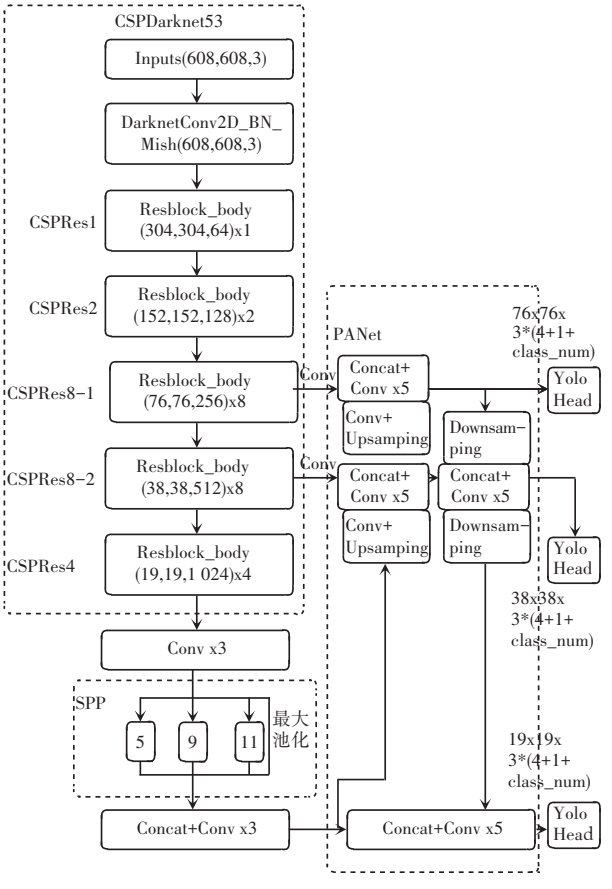

图 1　YOLO v4 网络结构

Fig. 1　YOLO v4 network structure

在 Darknet53 基础上引入 CSP 结构，减少了计算量并增强梯度表现，主要思想：在输入 block 之前，分为 2 个部分。其中，一个部分直接通过一个短路进行连接，该方式降低了 20% 的计算量，提高了计算能力。同时使用 Mish[13] 激活函数，在 PANet 中使用了 Leaky relu 激活函数，通过上述方式使得 YOLO v4 的检测精度更高。

SPP 附加模块与 PANet 路径聚合网络称为 Neck 结构，优化了多尺度特征融合的能力。研究中，SPP 附加模块采用 5×5、9×9、13×13 三种不同尺度的最大池化操作，扩展了感受野。PANet 路径聚合网络主要通过从底向上的路径增强、自适应特征池化、全连接融合的方式形成新的不同尺度特征图。

1.2　MHSA 模块网络结构

近年来，Transformer 不仅在 NLP 领域取得可观成果，同时在 CV 领域获取巨大成功，比如图像分类任务的 ViT[14]、目标检测任务的 DETR[15] 和 Deformable DETR[16] 模型，均是基于 Transformer 思想设计的。UC Berkeley 和 Google 基于 Transformers 结构设计了 BoTNet[8]，是一种简单且功能强大的 Backbone。通过仅在 ResNet 的最后 3 个 bottleneck blocks 中用多头注意力层（Multi－Head Self－Attention，MHSA）替换 3×3 空间卷积，如图 2 所示。MHSA 层如图 3 所示，引入相对位置编码不仅考虑内容信息，而且考虑不同位置的要素之间的相对距离，有效地相关联物体之间的信息与位置感知。

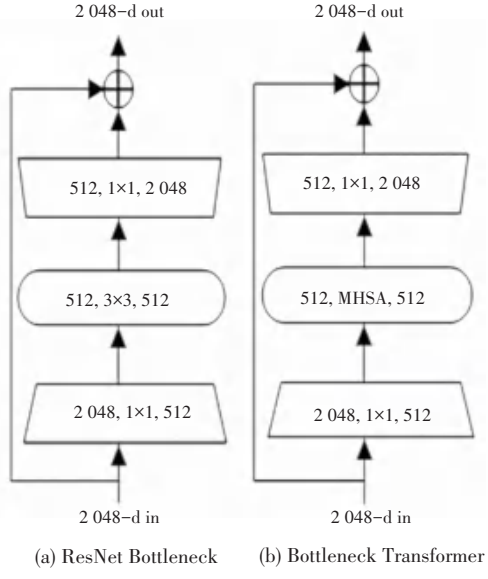

图 2　ResNet Bottleneck 与 BoTNet 网络结构

Fig. 2　ResNet Bottleneck and BoTNet network structure

2　改进的 YOLOv4 火灾检测方法

2.1　网络结构改进

2.1.1　特征提取主干网络的改进

主干网络由 5 个采用 CSP 单元模块组成，分别为 CSPRes1、CSPRes2、CSPRes8－1、CSPRes8－2、

CSPRes4,每个模块中有多个残差单元构建,参见图 1。引入 CSP 结构单元,一定程度降低计算量,但难以建模全局依赖关系。本文借鉴了 Bottleneck Transformer 结构对主干网络改进,采用 MHSA 层替换 3×3 空间卷积层。通过上述方式不仅增强网络全局依赖关系建模的能力,同时减少了参数,降低了计算时延。

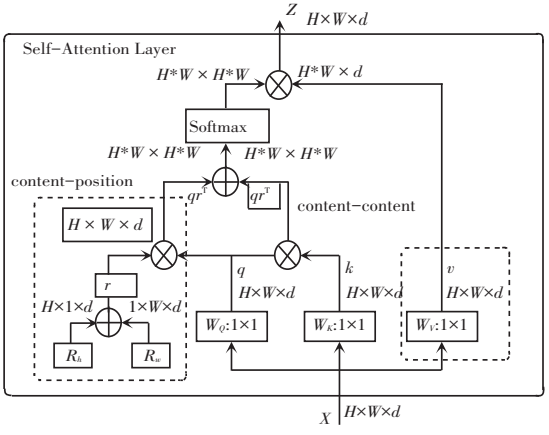

图 3 Multi-Head Self-Attention 网络结构

Fig. 3 Multi-Head Self-Attention network structure

对主干网络的改进主要思路为 2 点:

(1)使用卷积从大图像中学习抽象和低分辨率的特征图。

(2)使用全局(all2all)Self-Attention 来处理和聚合卷积捕获高层语义信息。

采用这种混合设计的方式,通过使卷积进行空间下采样并结合注意力模型集中在较小的分辨率上,同时可以有效地处理大尺度图像。具体改进思路如下:

(1)首先改进网络中 CSPRes8-1 与 CSPRes8-2,CSPRes8-x 模型,输入经过一层 3 × 3 卷积层处理后分成 2 个分支,第一分支仅经过一层 1 × 1 点卷积层处理,第二分支先经过一层 1 × 1 点卷积层处理以及循环经过 8 个 ResBlock Bottleneck 模块,紧接着经过一层 1 × 1 点卷积层,并与第一分支输入的特征图进行拼接,再将拼接后的特征图经过 1 × 1 点卷积处理后输出。将模块中 3 × 3 卷积层替换为 MHSA 层,如图 4 所示。

(2)主干网络中 CSPRes4 与 CSPRes8-x 模块结构相似,主要区别在于 ResBlock Bottleneck 结构不同,CSPRes4 模块中 ResBlock Bottleneck 模块先经过 3 × 3 卷积层,然后是 1 × 1 点卷积处理。其次,CSPRes4 经过 4 个 ResBlock Bottleneck 模块循环。具体改进方式将 ResBlock Bottleneck 模块中 3 × 3 卷积层替换为 MHSA 层,如图 5 所示。

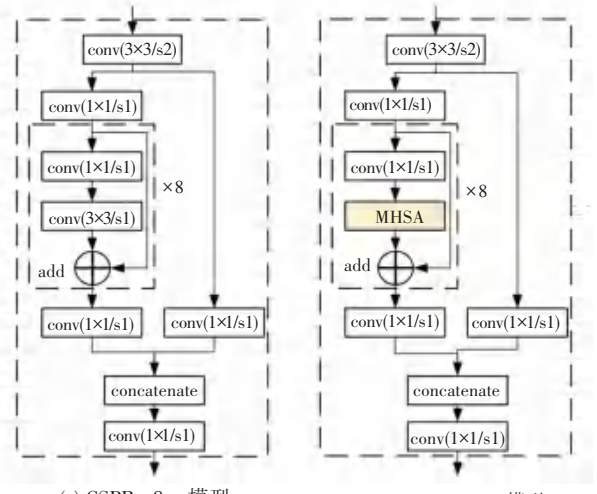

图 4 CSPRes8-x 模块与 BoT-CSPRes8-x 模块

Fig. 4 CSPRes8-x module and BoT-CSPRes8-x module

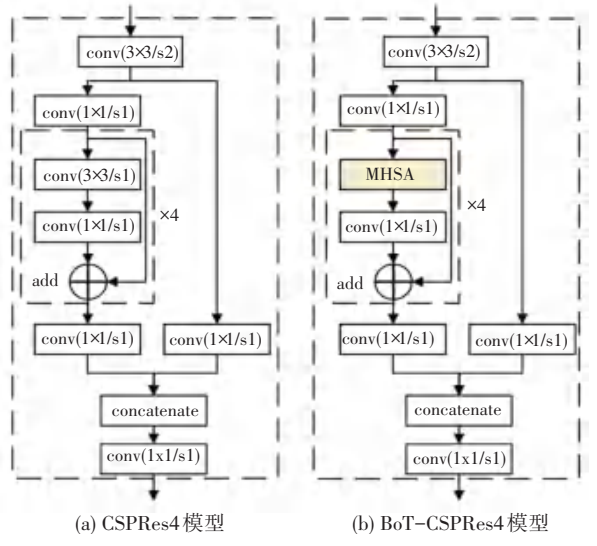

图 5 CSPRes4 模块与 BoT-CSPRes4 模块

Fig. 5 CSPRes4 module and BoT-CSPRes4 module

2.1.2 PANet 模块改进

PANet 路径聚集模块为 YOLO v4 的 Neck,参见图 1。对 PANet 的改进,同样借鉴 Bottleneck Transformer 设计思想,将网络中部分 3×3 CBL 单位替换为 MHSA 层,如图 6 所示。

2.2 火灾检测方法流程

火灾检测方法以改进的 YOLOv4 网络结构为基础,火灾检测的主要流程如下:

(1)对构建的火灾检测训练集进行预处理,标签转换为 YOLOv4 标准训练集格式。

(2)将经过预处理的训练集图像输入到改进的 CSPDarknet53 网络进行特征提取。

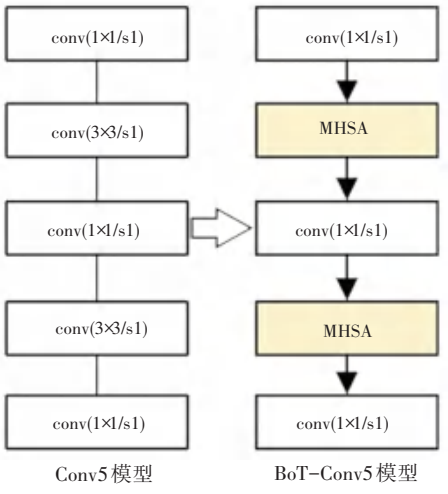

图 6　改进 PANet 中 Conv5 模块

Fig. 6　Improving the Conv5 module in PANet

(3)获取 CSPRes8-1 层、CSPRes8-2 层为输出第一、第二尺度的特征,CSPRes4 层经过 SPP 处理获取第三尺度的特征。

(4)上述三种尺度特征经过 PANet 层进行特征融合,获取 76×76、38×38、19×19 三种尺度的最终输出特征。

(5)分别将 3 种尺度特征输入的 YOLOv4 检测层,经过多轮训练生成最终的网络权值。

(6)测试阶段,将测试图像输入到 YOLOv4 网络中,调用训练得到的网络权值进行预测,并输出火灾检测结果。

3　实验结果与分析

3.1　火灾检测数据集

由于公开火灾数据集较少,通过采集互联网数据与视频监控数据两种方式,构建涵盖室内、野外、工厂、城市高楼、隧道等多个场景的火灾检测数据集。采集约 5 万张图片,通过数据清洗,12 886 张用于构建数据集,如图 7 所示。

图 7　火灾检测训练样本

Fig. 7　Fire detection training samples

3.2　实验环境与模型训练

基于 Ubuntu 18.04 操作系统,硬件配置为 2 块 Intel 至强 E5 CPU,显卡为 6 块 16GB NVIDIA Tesla P100,内存 500 GB。采用 python 与 PyTorch 深度学习框架搭建模型。

训练参数:初始学习率为 0.001、动量初始值为 0.9、权重衰减率为 0.000 5,批处理大小为 64,迭代次数为 8 000,采用步阶衰减学习率调度策略。

3.3　实验结果分析

改进的 YOLO v4 分别与 YOLO v3、YOLO v4 对比实验,主要对比精确率、召回率、平均精度(*mAP*)和检测时间,见表 1。

表 1　算法测试结果

Tab. 1　Algorithm detection results

模型	*P*/ %	*R*/ %	*mAP*/ %	*FPS*
YOLO v3	83.4	76.7	78.3	19
YOLO v4	87.2	79.3	82.5	21
改进的 YOLO v4	94.6	85.6	87.3	26

分析可知,改进 YOLO v4 算法相比 YOLO v3、YOLO v4,精确率方面提升 11.2%、7.4%,召回率方面提升 8.9%、6.3%, *mAP* 提升 9%、4.8%。改进后的 YOLO v4,检测速度比 YOLO v3 与 YOLO v4 均有大幅度提升,检测速度达到 27 帧/s。火灾检测结果如图 8 所示。

图 8　改进的 YOLO v4 算法检测结果

Fig. 8　The detection results of the improved YOLO v4 algorithm

实验表明改进的火灾检测算法能够检测大尺度与小尺度的火焰与烟雾目标,既是在存在干扰目标、目标遮挡的复杂场景下,依然能够有效检测目标,具有检测精度高、误检率低、鲁棒性等优点。

4 结束语

针对 YOLO v4 火灾检测性能不足的问题,借鉴 Bottleneck Transformer 结构设计思想,引入 MHSA 层对 YOLO v4 主干网络 CSPDarknet 和 PANet 模块进行改进。由于火灾检测数据集较少,采集了大量图片与视频火灾数据,构建多场景火灾检测数据集。

通过对比 YOLO v4、YOLO v3 火灾检测方法表明,本文改进后的方法比现有的火灾检测方法具有更好的鲁棒性、更低的误检测率,检测精度与实时性均有良好的性能。测试集上达到 94.6%的准确率、85.6%的召回率、87.3%的 mAP。未来的研究工作中,重点研究结合 Transformer 改进网络进行优化,提升检测效果与实时性,以及扩展现有的火灾检测数据集,增加火灾样本的多样性,提升检测算法的泛化能力。

参考文献

[1] 任嘉锋, 熊卫华, 吴之昊,等. 基于改进 YOLOv3 的火灾检测与识别[J]. 计算机系统应用, 2019, 28(12):171-176.

[2] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. arXiv preprint arXiv:1804.02767, 2018.

[3] 张为, 魏晶晶. 嵌入 DenseNet 结构和空洞卷积模块的改进 YOLO v3 火灾检测算法[J]. 天津大学学报(自然科学与工程技术版), 2020, 53(9):100-107.

[4] HUANG G, LIU Z, LAURENS V, et al. Densely Connected Convolutional Networks [J]. arXiv preprint arXiv: 1608.06993, 2016.

[5] REDMON J. Darknet: Open source neural networks in C[EB/OL]. [2013-2016]. http://pjreddie.com/darknet/.

[6] 晋耀,张为. 采用 Anchor-Free 网络结构的实时火灾检测算法[J]. 浙江大学学报(工学版), 2020, 54(12):163-169.

[7] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks [J]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE, 2018:4510-4520.

[8] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition[J]. arXiv preprint arXiv:2101.11605, 2021.

[9] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv: 2004.10934,2020.

[10] WANG C Y, LIAO H, YEH I H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA :IEEE, 2019:1571-1580.

[11] CHEN Yunian, WANG Yanjie, ZHANG Yang, et al. PANet: A context based predicate association network for scene graph generation [C]// 2019 IEEE International Conference on Multimedia and Expo (ICME). Shanghai, China: IEEE, 2019: 508-513.

[12] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep Convolutional Networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904-1916.

[13] MISRA D. Mish: A self regularized non-monotonic neural activation function[J]. arXiv preprint arXiv:1908.08681,2019.

[14] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. ICLR2021,Vienna, Austria:[s. n.], 2020:1-21.

[15] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[M]//VEDALDI A, BISCHOF H, BROX T, et al. Computer Vision-ECCV 2020. ECCV 2020. Lecture Notes in Computer Science. Cham : Springer, 2020, 12346:213-229.

[16] ZHU Xizhou, SU Weijie, LU Lewei, et al. Deformable DETR: Deformable transformers for end-to-end object detection [J]. arXiv preprint arXiv:2010.04159, 2020.

[17] ZHENG Zhaohui, WANG Ping, LIU Wei, et al. Distance-IoU loss: Faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020,34(7):12993-13000.

(上接第 85 页)

[3] LIN J, NIEMEIER D A. An exploratory analysis comparing a stochastic driving cycle to California´s regulatory cycle [J]. Atmospheric Environment, 2002, 36(38):5759-5770.

[4] PACHECO A F, MARTINS M, HUA Z. New European Drive Cycle (NEDC) simulation of a passenger car with a HCCI engine: Emissions and fuel consumption results[J]. Fuel, 2013, 111(9): 733-739.

[5] 彭育辉,杨辉宝,李孟良,等. 基于 K-均值聚类分析的城市道路汽车行驶工况构建方法研究[J]. 汽车技术,2017(11):13-18.

[6] 刘燕. 基于抽样和最大最小距离法的并行 K-means 聚类算法[J]. 智能计算机与应用,2018,8(6):37-39,43.

[7] 高建平,高小杰. 改进模糊 C 均值聚类法的车辆实际行驶工况构建[J]. 河南科技大学学报(自然科学版),2017,38(6):21-27,4-5.

[8] AMIRJAMSHIDI G, ROORDA M J. Development of simulated driving cycles for light, medium, and heavy duty trucks: Case of the Toronto Waterfront Area[J]. Transportation Research Part D, 2015, 34(1):255-266.

[9] 宋怡帆. 基于聚类和 Python 语言的深圳市城市道路车辆行驶工况构建[D]. 西安:长安大学,2018.

[10] 刘子谭,朱平,刘旭鹏,等. K 均值聚类改进与行驶工况构建研究[J]. 汽车技术,2019(11):57-62.

[11] 于仲安,褚彪,葛庭宇. 基于 HPSO-BP 神经网络融合的锂电池 SOC 预估研究[J]. 汽车技术,2019(6):20-24.

[12] 谢秀华,李陶深. 一种基于改进 PSO 的 K-means 优化聚类算法[J]. 计算机技术与发展,2014,24(2):34-38.

[13] 王盟,余粟,冯益林. 改进小波阈值对热泵电机振动信号的去噪研究[J]. 智能计算机与应用,2020,10(4):17-21.

[14] 石琴,郑与波,姜平. 基于运动学片段的城市道路行驶工况的研究[J]. 汽车工程,2011,33(3):256-261.

[15] 郑与波,石琴,王世龄. 合肥市汽车行驶工况的研究[J]. 汽车技术,2010(10):34-39.