

DOI:10.20033/j.1003-7241.(2026)02-0174-05

# 一套英语试卷自动批阅系统的设计与实现

李静莹

(咸阳师范学院, 陕西 咸阳 712000)

**摘要:**为提升教育机构英语试卷批阅的自动化水平,研究基于卷积神经网络和 LeNet-3 模型两项技术提出了一套英语试卷自动批阅系统,并通过 MNIST 手写数据集对该系统进行训练。为验证英语试卷自动批阅系统的有效性,专门为该系统建立软硬件运行环境并通过 accuracy\_avg 指标对系统模型图像识别的准确率进行评价。经试验研究发现,该系统所使用的 LeNet-3 模型能够在迭代 300 轮之后实现收敛,图像识别准确率达 98.677%。最后采用 Python 软件为自动批阅系统建立可视化的图形操作界面,并展示了操作界面设计效果。

**关键词:**卷积神经网络;图像预处理;模型试验;系统实现;自动批阅系统

中图分类号: TP319

文献标志码: A

文章编号: 1003-7241(2026)02-0174-05

## Design and implementation of an automatic English test paper marking system

LI Jingying

(Xianyang Normal University, Xianyang 712000, Shaanxi, China)

**Abstract:** In order to improve the automation level of English test paper grading in educational institutions, this study proposes an automatic English test paper grading system based on two technologies: Convolutional Neural Network and LeNet-3 model, and trains the system using the MNIST handwritten dataset. To verify the effectiveness of the automatic grading system for English test papers, a software and hardware operating environment is specifically established for the system, and the accuracy of the system model image recognition is evaluated through the accuracy\_avg index. Through experimental research, it is found that the LeNet-3 model used in the system can achieve convergence after 300 iterations, with an image recognition accuracy of up to 98.677%. Finally, a visual graphical operation interface is established for the automated grading system using Python software, and the design effect of the operation interface is demonstrated.

**Keywords:** convolutional neural network; image preprocessing; model testing; system implementation; automatic marking system

随着我国计算机技术的不断发展,“深度学习”在图像处理工作中的应用越来越广泛,如何选择数据处理算法,提高图像处理工作效率已经成为相关单位十分重要的研究课题之一。在英语教学领域中,越来越多的教育机构已经实现了无纸化阅卷,将试卷图像输入卷积神经网络能够实现端到端的数据处理,摆脱了传统的 OMR 机器阅卷,初步具备了自动化阅卷的实现基础<sup>[1-3]</sup>。叶佩等<sup>[4]</sup>提出了一套基于 MVC 框架的英语在线学习资源管理系统,用户和资源双角度分析,建立用户特征模型,凭借密度用户项目矩阵及模糊类似度的优先比方法计算用户间相似性,找出用户间最类似项目以及其次项目,显著提升了英语学习资料推送的准确度;胡湘雨等<sup>[5]</sup>针对现行大学英语教学质量测评系统存在测评结果斯皮尔曼相关系数数值较低的问题,提出基于数据挖掘的大学英语教学质量测评系统设计及应用,该系统测评结果斯皮尔曼相关系数在 0.9

以上,测评结果可以精准地反映大学英语教学质量情况。为进一步提高英语试卷的自动化批阅效率,本次研究提出了一项基于 LeNet-3 网络的试卷字符识别算法,有效提高了图像字符的识别准确率,为英语试卷的自动化批阅提供了有力的技术支持。

### 1 卷积神经网络

本次研究通过卷积神经网络对英语试卷图像数据进行处理,该网络主要由卷积层、池化层和激活函数所组成。

#### 1.1 卷积层

在卷积神经网络中,卷积层是用来提取特征的核心部分,以规格为 3×3 的卷积核为例,该卷积核步长为 1,其卷积操作如图 1 所示。由于主流网络模型通常倾向于使用较大的卷积核,计算参数量较为庞大,不适用于英语试卷自动批阅系统。因此,本次研究用规格为 3×3 的卷积核

收稿日期:2024-05-28;录用日期:2024-06-03

基金项目:2021 年度陕西省教育科学“十四五”规划课题(SGH21Y0203)

作者简介:李静莹(1978—),女,博士,讲师,研究方向:英语教学与计算机应用。

引用本文:李静莹.一套英语试卷自动批阅系统的设计与实现[J].自动化技术与应用,2026,45(2):174-178.(LI Jingying. Design and implementation of an automatic English test paper marking system[J]. Techniques of Automation and Applications, 2026,45(2):174-178.)

替代规格为5×5的卷积核。

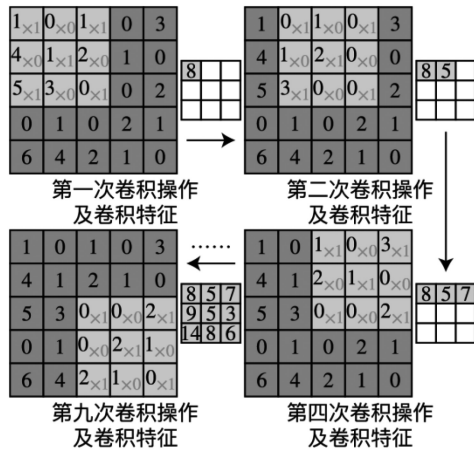


图1 卷积数据处理流程

Fig. 1 Convolutional data processing flow

### 1.2 池化层

本次研究所采用的池化层为最大池化层,具体形式如图2所示。以规格为2×2的最大池化层为例,该池化层步长为1,可在池化窗中筛选出最大值。最大池化层具有平移不变性,计算公式为

$$\mu_k = \max_{(i,j) \in \Omega} M_{i,j}^k \quad (1)$$

式中,  $M$  为特征图,  $k$  为卷积核数量,  $\Omega$  为最大池化的局部大小,即子采样大小,  $M_{i,j}^k$  为卷积后的第  $k$  个特征图<sup>[6]</sup>。

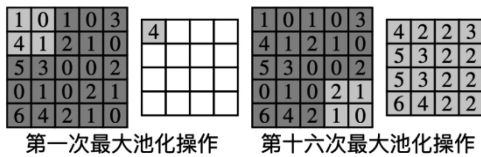


图2 最大池化数据处理流程

Fig. 2 Maximum pooling data processing flow

### 1.3 激活函数

激活函数负责通过数值转换的方式在模型中添加非线性因素,即采用非线性函数来强化模型的泛化性。本次研究所采用的激活函数为 ReLU 函数,具体形式为

$$f_{\text{relu}}(x) = \max\{0, x\} \quad (2)$$

对于 ReLU 函数来说,在输入大于 0 的情况下,输出即为输入;在输入小于 0 的情况下,输出即为 0。这种简单的计算方式收敛速度较快,而且能够一定程度上应对网络模型在训练过程中所存在梯度消失现象。

## 2 英语试卷图像的预处理

首先需要在录入英语试卷图像后将其转换为灰度图,在此基础上通过 Canny 边缘检测算法对图像进行边缘检测处理,再根据图像的灰度值分别设定高、低两个阈值,将高、低两个阈值分别记为  $H$  和  $L$ <sup>[7-9]</sup>。基于以往的研究经验,本次研究分别将  $H$  阈值和  $L$  阈值设定为 210 和 70。

设图像水平方向像素点为  $x$ ,垂直方向像素点为  $y$ ,在

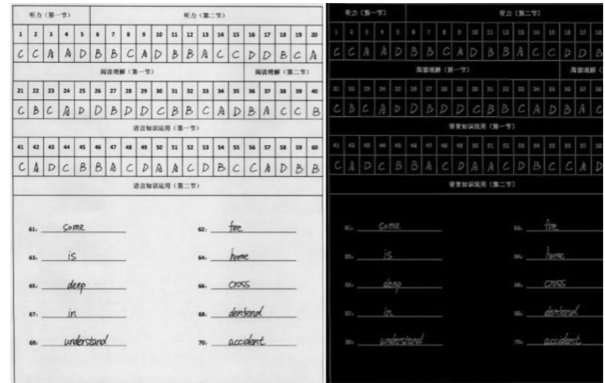
此基础上计算像素点的梯度强度和方向,具体方法为

$$G = \sqrt{G_x^2 + G_y^2} \quad (3)$$

$$\theta = \alpha \tan 2(G_y, G_x) \quad (4)$$

式(3)和式(4)中,  $G$  代表像素点的梯度强度,  $\theta$  代表像素点的方向,  $G_x$  代表水平方向的差分,  $G_y$  代表垂直方向的差分。

在此基础上,若当前像素点的梯度小于沿正负梯度方向像素的梯度,则抑制该像素,反之则保留该像素。最终所得到的边缘检测结果如图 3 所示。



(a) 图像 (b) 边缘检测结果

图3 图像边缘检测结果

Fig. 3 Image edge detection results

### 2.1 字符的截取与识别网络

对于英语试卷来说,截取单词图像的关键在于消除答案正文的横线。首先将图像的宽记为 width,将横线所在行设定为 0 灰度值像素数量大于 0.3 倍 width 的行。分别将行坐标的最大值和最小值记为 max 和 min,对于  $[\max + 2, k]$  内的某值  $k$ ,若其二值化结果的  $[\max + 2, k]$ 、 $[\max + 1, k]$ 、 $[\min - 1, k]$ 、 $[\min - 2, k]$  像素值均为 0,则将横线所在行中的  $k$  列 0 灰度值像素全部保留,否则将该像素赋值为 225,在此基础上对该像素实施二值化和中值滤波处理,将横线残留去除干净<sup>[10-12]</sup>。在此基础上,记录图像纵横两个方向中灰度值为 0 的像素数量,进而显现出该图像四个边缘,最终得到如图 4 所示的单词图像截取结果。



图4 单词图像截取结果

Fig. 4 Word image extraction results

### 2.2 字符识别网络

#### 2.2.1 LeNet-5 模型

LeNet-5 是一个应用范围较为广泛的轻量化模型,本次研究在该模型的基础上加以改进。首先对 LeNet-5 模型进行微调,即通过 AlexNet 中的 dropout 函数和 ReLU 函数来解决由数据集过小所造成的过拟合问题。在此基础上,通过“VALID”padding 来削弱边缘信息的参考权重,充分保留特征图中的有效特征信息。最后在模型中适当

增加全连接层和卷积核的数量,进一步强化特征图的可用性。

### 2.2.2 LeNet-3 模型

本次研究用 2 个 3×3 规格的卷积核替代 LeNet-5 网络模型中 5×5 规格的卷积核,进而引入更多的非线性并增加网络深度,以较少的参数实现较强的模型判决力,该过程的替换原理如图 5 所示。

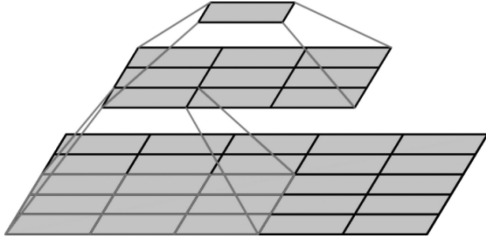


图 5 LeNet-5 模型卷积核替代原理

Fig. 5 Principle of convolutional kernel substitution in LeNet-5 Model

## 3 模型试验

### 3.1 试验准备

模型试验所选用的数据集为 MNIST 手写数据集,该数据集包含 10 000 张测试图片和 60 000 张训练图片,所有图片均为灰度化的手写内容,图片规格均为 28×28 像素。

模型试验的硬件环境如下: Intel® Core™ i7 processor 14650HX 中央处理器 (30M Cache, up to 5.20 GHz)、NVIDIA Quadro RTX 5000 图形处理器 (16 GB GDDR6)、Kingston DDR4 16 G 2133 MHz 内存、三星 990 EVO PCIe 4.0 NVMe M.2 固态硬盘。

模型试验的软件环境如下: Windows 10、Numpy1.16.2、Opencv3.4.1、CUDNN7.5.0、CUDA10.0、Python3.7、TensorFlow1.13。

本次研究通过平均准确率 accuracy\_avg 指标来评价模型的准确率水平,模型的性能与 accuracy\_avg 值成正比。

### 3.2 更新网络权重

对于网络训练环节来说,前向传播与反向传播是两个重要的训练过程,分别用于计算损失和调整参数。

#### 3.2.1 前向传播

首先通过正态分布的方式为偏置  $b$  和权重参数  $w$  设定一个初始值,将经过预处理的图像输入至模型中,通过卷积神经网络逐层提取特征,具体方式为

$$z^{l+1} = f(u^{l+1}) \quad (5)$$

$$u^{l+1} = w^{l+1}z^l + b^{l+1} \quad (6)$$

式(5)和式(6)中,  $z^l$  为  $l+1$  层的输入且为  $l$  层的输出,  $f(u^{l+1})$  为 ReLU 激活函数,  $u^{l+1}$  为  $l+1$  层的输出,  $b^{l+1}$  为  $l+1$  层的偏置,  $w^{l+1}$  为  $l+1$  层的权重<sup>[13-14]</sup>。

#### 3.2.2 反向传播

本次研究通过交叉熵损失函数对真实标签与预测值之间的误差进行计算,该函数具体形式为

$$E = -\frac{1}{N} \sum_{n=1}^N [y_n \ln Y_n + (1 - y_n) \ln(1 - Y_n)] \quad (7)$$

式中,  $Y_n$  为预测值,同时也是 softmax 层输出的概率向量,  $E$  为损失值,  $N$  为图片数量,  $y_n$  为样本标签,同时也是 one-hot 编码<sup>[15]</sup>。

在此基础上,通过梯度下降算法实现针对损失函数的最小化处理,确保模型的偏置和权重参数与数据特征保持一致。在反向传播的过程中,卷积网络的池化层负责将更新后的权重传递给卷积层,具体方式为

$$w^l = w^l - \eta \frac{\partial E^l}{\partial w^l} \quad (8)$$

式中,  $\eta$  代表学习率,  $E^l$  为第  $l$  层的损失。

### 3.2.3 设置网络参数

本次研究通过边测试边训练的方式来优化网络模型,即每次迭代过程中都要向模型输入所有的测试图片,根据类别得分对字符的类别属性加以判断,再通过标签对比的方式获得准确率数值。在完成训练后,采用 PowerBI 数据分析软件输出模型的损失曲线,根据该曲线对模型的收敛状况进行判断。

为了使测试结果更加具有说服力,本次研究统一规范模型的训练参数,采用 Adam 优化器优化模型,将 Batch\_size 设定为 1 024,将学习率设定为 0.000 1,迭代次数设定为 800。

相较于其他优化手段来说,Adam 优化器能够对 RMSProp 和 AdaGrad 两种算法加以整合,有效解决模型训练过程中所遇到的噪声和梯度稀疏问题。为防止出现梯度消失和过拟合现象,本次研究在模型中引入 ReLU 激活函数,并将 dropout 技术应用于全连接层中,设定元素保留概率为 0.6。

在完成训练后,LeNet-3 的损失曲线如图 6 所示。根据图 6 可知,在经过 700 轮迭代后,模型的损失值开始收敛,有效保证了试验结果的可靠性与真实性。

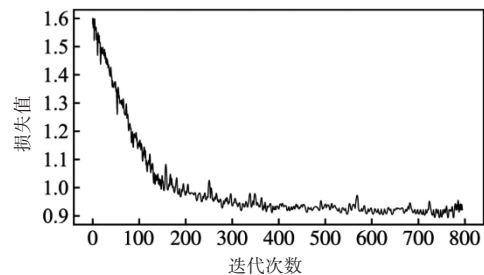


图 6 模型收敛情况

Fig. 6 Convergence of the model

### 3.2.4 试验结果

#### 1) 网络性能对比

由于测试图像规格为 28×28 像素,图像清晰度较低,虽然可以通过重新调整图像大小的方式获得更高级的特征,但在网络较深的情况下,原始图像的失真会大大降低特征的可靠性,进而对模型性能造成较为严重的影响。因此,本次研究未采用 ResNet 等深层网络,而是集中测试 VGG-16、AlexNet、微调 LeNet-5 以及 LeNet-5 等四种网

络的图像识别性能。

由于每一种网络在图像规格方面的要求存在差异,因此本次研究在训练和测试的过程中对图像进行了统一调整。四种网络模型在收敛状态下的图像识别准确率如表1所示。

表1 各网络模型的图像识别平均准确率

Tab.1 Average accuracy of image recognition for each network model

模型	准确率/%
VGG-16	97.933 91
AlexNet	97.933 91
微调 LeNet-5	98.206 76
LeNet-5	98.070 33

根据表1可知,由于 LeNet-5 网络模型体量较轻,在识别小尺寸图像的过程中体现出了较为理想的准确性效果。AlexNet 网络模型所要求的输入图像尺寸为 227×227 像素,对网络性能造成了较为严重的影响,VGG-16 网络模型在运行过程中也出现了一定程度的网络退化问题,即使是在堆叠网络层数的情况下也无法提升准确率。本次研究所提出的微调 LeNet-5 网络模型在准确率方面高于原始 LeNet-5,说明该模型的微调方案具有较为充分的可行性。

2) 3×3 卷积核堆叠结构的性能

本次研究通过对比试验的方式来判断网络模型中 3×3 卷积核堆叠结构的性能表现,分别对 LeNet-3、微调 LeNet-5 和 LeNet-5 三种网络模型进行训练,同时使超参数和优化方式保持一致,绘制网络模型在完成训练后的准确率曲线,试验结果如图7和表2所示。

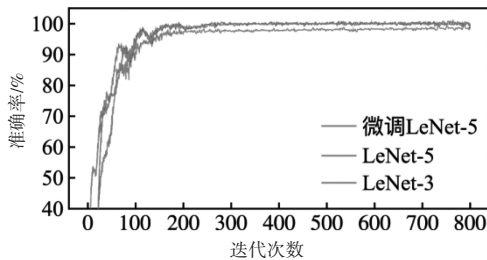


图7 迭代过程中三种网络模型的准确率变化曲线

Fig.7 Accuracy variation curves of three network models during the iteration process

表2 三种网络模型的平均准确率测试结果

Tab.2 Average accuracy test results of three network models

模型	平均准确率/%
LeNet-3	98.677 0
微调 LeNet-5	98.465 2
LeNet-5	97.829 5

由于本次研究用 2 个 3×3 卷积核堆叠结构替换了 LeNet-5 中的 5×5 卷积核,因此卷积层数量由 2 层增加至 4 层。第一和第二卷积层分别使用了 8 个 3×3 卷积核和 32 个 3×3 卷积核,第三和第四卷积层分别使用了 32 个 3×

3 卷积核和 64 个 3×3 卷积核。若将偏置项忽略不计,则 LeNet-3 和 LeNet-5 的卷积部分参数则分别有 30 168 个和 53 600 个,LeNet-3 网络模型的参数数量较 LeNet-5 下降了 43%,有效提升了模型的运行效率和训练效率。根据表2可知,在 MNIST 所提供的测试集中,LeNet-3 网络模型的图像识别准确率高达 98.677%,显著高于 LeNet-5 和微调 LeNet-5。根据图7可知,LeNet-3、LeNet-5 和微调 LeNet-5 在具有较高准确率的同时,也能够经历 300 轮迭代后达到较为稳定的准确率水平,相较于 VGG-16 和 AlexNet 来说准确率波动较小,稳定性优势显著。

4 自动批阅系统的实现

本次研究主要通过 Python 软件来组建英语试卷自动批阅系统。在实际操作方面,采用 PyQt5 设计图形操作界面,通过 TensorFlow 训练 LeNet-3 网络模型,用 Opencv 完成图像处理任务。以“查看试卷”界面为例,英语试卷自动批阅系统的最终设计效果如图8所示。

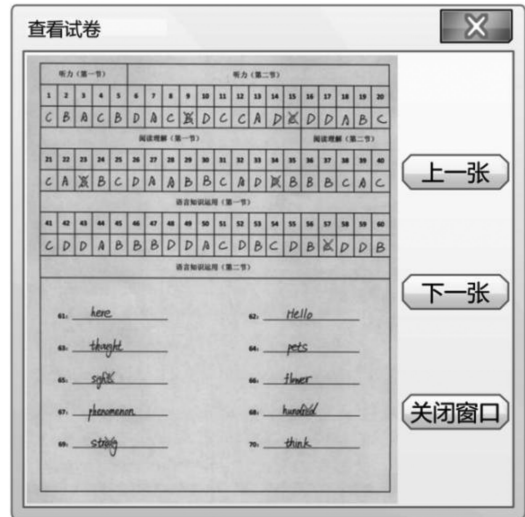


图8 查看试卷界面

Fig.8 View exam paper interface

5 结论

本文详细介绍了英语试卷自动批阅系统的设计方法,具体阐述了该系统所采用的网络模型基本结构以及图像预处理的实现策略。并通过模型试验的方式对该系统的图像识别性能加以验证。经试验研究发现,英语试卷自动批阅系统的图像识别准确率高达 98.677%,且收敛性能良好。系统操作界面美观友好,有助于提升教育机构英语试卷批阅的自动化水平,具有一定的应用价值。

参考文献

[1]周博,葛洪武,李珩,等. 基于特征重加权的小样本遥感图像目标检测算法[J]. 计算机测量与控制, 2024, 32(2):283-290.

[2]廖欢,朱文球,雷源毅,等. 融合全局上下文注意力的遥感图像检测方法[J]. 兵器装备工程学报, 2024, 45(2):278-283.

[3]禹鑫淼,林密,卢江平,等. 基于向量叉乘标签分配的遥感图像目标

检测算法[J]. 高技术通讯, 2024, 34(2):132-142.

[4]叶佩. 基于MVC框架的英语在线学习资源管理系统[J]. 自动化技术与应用, 2024, 43(4):89-92.

[5]胡湘雨. 基于数据挖掘的大学英语教学质量测评系统设计及应用[J]. 中国新技术新产品, 2024(2):51-53.

[6]张立国, 雷璇瑞, 金梅, 等. 基于图像处理的电路板缺陷检测系统设计[J]. 高技术通讯, 2024, 34(2):209-217.

[7]王照, 葛馨远, 饶毅. 基于多算法融合的破损陶瓷绝缘子图像识别与检测方法[J]. 自动化技术与应用, 2024, 43(2):40-44, 88.

[8]蔡永洪, 马婷婷, 杨叶花. 基于频谱遮罩的玻璃量器图像处理与围线检测方法[J]. 机电工程技术, 2024, 53(2):35-38, 88.

[9]徐健康, 朱正兵, 杨德志, 等. 基于图像识别技术的电动工具视觉检测系统应用[J]. 电动工具, 2024(1):1-3, 22.

[10]张一民. 基于PLC的泵站自动化系统节能改造设计[J]. 自动化技术与应用, 2024, 43(9):185-188.

[11]李相格. 基于改进LOG算子与Otsu算法结合的物体表面图像残损检测方法[J]. 兰州交通大学学报, 2024, 43(1):59-63.

[12]程智慧. 三维实时图像声呐在风电桩基础检测中的应用[J]. 中国水运, 2024(2):76-77.

[13]吴鑫, 王莉, 徐连明, 等. 基于多模态交互引导网络的遥感图像密集车辆检测[J]. 指挥与控制学报, 2024, 10(1):1-8.

[14]杨轲, 董兵, 吴悦, 等. 基于轻量化YOLOv4机场场面遥感图像目标检测方法[J]. 计算机与现代化, 2024, (2):93-99.

[15]崔海花, 赵英凯. 基于热红外图像的船舶电气设备状态异常检测研究[J]. 舰船科学技术, 2024, 46(3):147-150.

(上接第72页)

### 3 结论

为了明确GIS内部缺陷的情况,研究搭建该设备的3D模型,利用传感器检测设备的振动情况。同时,通过X射线成像分析设备内部存在的缺陷情况,检测内部气体分析实际情况。结果表明,X射线照射后, $\text{SF}_6$ 气体中的 $\text{SO}_2$ 和 $\text{H}_2\text{S}$ 含量均为0,未发生明显变化,表明X射线照射并未引起 $\text{SF}_6$ 气体的分解或产生其他有害气体,保持了 $\text{SF}_6$ 气体的稳定性。针对螺栓缺陷情况,通过振动信号的分析,确定了不同频率下振幅的变化情况。当频率为170 Hz时,振幅最高,为 $6.0 \text{ m/s}^2$ ,200 Hz时振幅出现较大波动变化,370 Hz时振幅为 $2.8 \text{ m/s}^2$ 。该方法具有一定可行性,有助于准确识别GIS设备内部可能存在的螺栓松动缺陷。这为维护人员提供了清晰明了的内部情况,有助于提高检修效率。但实验室环境条件的温度、湿度等环境因素的变化可能影响传感器的测量精度,未来可以设置实验环境保持智能恒温恒湿条件,以确保实验条件的稳定性和可控性。

### 参考文献

[1]杨为, 柯艳国, 赵恒阳, 等. 基于轻量级卷积神经网络的GIS绝缘和机械故障诊断方法[J]. 高压电器, 2023, 59(9):201-210.

[2]MUDHIGOLLAM U K, TIWARI N, RAO M M. Transient thermal analysis of gas insulated switchgear modules using thermal network approach[J]. International Journal of Emerging Electric Power Systems,

2024, 25(2):163-174.

[3]吴细秀, 冷宇宽, 庞文龙, 等. GIS母线潜伏性故障电场劣化特性与故障概率预测研究[J]. 电网技术, 2023, 47(7):3018-3028.

[4]张芳胜, 王妙龄, 季嘉辉, 等. 基于语音识别的多资源组合应急调度指挥系统[J]. 自动化技术与应用, 2024, 43(3):155-159.

[5]董冰冰, 高常胜, 许渊, 等. 基于修正时频聚类分界的GIS设备特高频偶发局放识别算法[J]. 高电压技术, 2023, 49(11):4736-4746.

[6]李婷. 基于数据可视化技术的竞争对手数据分析—以9家网店数据为例[J]. 自动化技术与应用, 2025, 44(9):169-174.

[7]吴旭涛, 赵晋飞, 马云龙, 等. 基于多频激励下振动响应的GIS机械缺陷诊断方法[J]. 电力电容器与无功补偿, 2022, 43(4):108-115.

[8]唐志国, 李阳. 恒压下GIS固体绝缘界面气隙缺陷放电严重程度评估[J]. 高压电器, 2024, 60(2):62-68.

[9]沙伟燕, 罗艳, 马鹏欢, 等. 不同金属微粒缺陷下GIS盆式绝缘子表面放电与电场特性研究[J]. 绝缘材料, 2023, 56(9):89-93.

[10]王琼苑, 褚继峰, 李秋霖, 等. 基于微型气体传感阵列的空气绝缘设备放电故障识别[J]. 电工技术学报, 2023, 38(23):6494-6502.

[11]吕玮, 张立红. 特高压气体绝缘金属封闭输电线路地震反应分析中建模方法研究[J]. 建筑结构, 2023, 53(增刊1):730-733.

[12]范展滔, 吴小刚, 吕耀棠, 等. GIS设备刀闸振动信号噪声分离[J]. 电力系统及其自动化学报, 2023, 35(7):29-35.

[13]梁基重, 晋涛, 牛曙, 等. 基于EMD-FFT特征提取的GIS机械缺陷诊断方法研究[J]. 电力科学与技术学报, 2023, 38(3):216-223.

[14]王森, 罗林根, 钱勇, 等. 基于模拟故障数据库的GIS局部放电反演方法[J]. 高电压技术, 2022, 48(5):1663-1672.

[15]张语桐, 吴泽华, 徐家忠, 等. 特高压GIS用单支撑绝缘子绝缘结构优化设计[J]. 电工技术学报, 2023, 38(1):258-269.