▶ 现代计算机

XIANDAI JISUANJI 第30卷第6期(总第798期) 半月刊(1984年创刊) 2024年3月25日出版

主管単	位	中山大学
主办单	位	广州中山大学出版社有限公司
出版单	位	广东现代计算机杂志社有限公司
发	行	广东省报刊发行局(全国公开发行)
印	刷	广州市友盛彩印有限公司
社	长	黄少伟
主	编	石玉珍
编	委	邹岚萍 熊锡源 李 文 石玉珍 梁嘉县
地	址	广州市海珠区新港西路135号
		中山大学内(510275)
电	话	020-84112089(编辑部)
网	址	www.moderncomputer.cn
电子曲	『箱	tougao@moderncomputer.cn

ISSN 1007-1423 CN 44-1415/TP 邮发代码: 46-121 定价: 30.00元





	20	24 年	3月	第30
	第	5)期	(总第	3798期
0				
0				
ъ С				
		甲山大	子出版社	主办

M

现代计算机

第30卷

第6期

(总第798期

 \smile

2024年3月





MC

TER

ISSN 1007-1423 CN 44-1415/TP

中国期刊数据库CNKI全文收录期刊 中国学术期刊(光盘版)收录期刊 中文科技期刊数据库全文收录期刊 中国核心期刊(遴选)数据库收录期刊 中国学术期刊综合评价数据库收录期刊

- ◆研究与开发: 计算机发展和软、硬件开发的理论研究
- ◆ 图形图像: 重点为与图形图像相关的理论及实践研究
- ◆ 开发案例: 基于某方面的计算机开发案例研究与分析
- ◆ 实践与经验: 计算机应用的实例及心得

版权声明

1. 本刊版权属于杂志社所有,其他报刊或网站如需转载,须经本刊 同意,注明转载自本刊并付作者稿酬。

2. 本刊来稿恕不退还,请自留底稿。请勿一稿多投。来稿文责自负, 严禁抄袭。对侵犯他人版权或其他权利的稿件,本刊概不承担连带责任。

3. 对所投稿件,本刊编辑有权根据刊物的需要进行删改或调整。

4. 凡是刊登在本刊的稿件,即表示作者同意稿件在《现代计算机》 网站、中国期刊数据库CNKI、中国学术期刊(光盘版)、中文科技期刊 数据库、中国科技期刊(遴选)数据库、中国学术期刊综合评价数据库 等媒体发布。

现代计算机

XIANDAI JISUANJI

第30卷 第6期(总第798期)2024年3月25日出版

目 次

研究与开发

改进YOLOv5的路面裂缝识别方法 ······	杨景维,	黎远松	(1)
基于双尺度时间特征的步态识别方法 魏永超,徐未其,朱泓超,	朱姿翰,	刘伟杰	(8)
基于 BERT 和图注意力网络的篇章级事件论元识别 ······	王 凯,	廖涛	(14)
适用于边缘设备的轻量级人体检测算法周 宁,	陶青川,	彭勃兴	(20)
基于 DBNet 改进的检务场景文本检测算法研究	于晓,	林世基	(26)
联合 Swin Transformer 和 UNet 的 GAN 人脸修复算法 ······	•••••	张梦澜	(32)
基于LSTM-Attention的空间目标分类研究 杨礼友,	余显冰,	李 智	(38)
基于受扰判断的智慧网络移动终端数据链多层加密算法 梁正华,	温权波,	邹立朋	(44)
基于 Bi-LSMT 结合注意力机制的钓鱼网站识别 ······	尚培文,	李东帅	(50)
基于图像亮度均衡技术的霍夫圆检测研究	任建松,	来枫璟	(56)
基于改进YOLOv5的路面坑洞检测设计 ·······周研逸,周月娥,沈琳芸,	沈 立,	赵远东	(61)
基于 Transformer 的蛋白质相互作用预测研究	靳晓宏,	韦文山	(65)
基于改进的LSD线特征提取算法研究 谭帅奇,	孙皓,	李健铭	(69)

实践与经验

基于不确定性的机坪管制系统风险评估 廖 勇, 黄 杰,	方	娜,	赵一	阳	(73)
基于CBM的共用信息系统装备维修保障体系分析与架构设计					
庞军伟,	张	鹏,	侯兴	明	(80)
知识增强大模型在信息系统故障分析中的应用研究 张海龙, 黄文锋,	路	翔,	张	磊	(87)
改进哈里斯鹰算法用于阵列天线方向图赋形	朱智	坚,	华	伟	(94)
智慧城市下基于属性加密的隐私数据保护	时	豪,	田	野	(99)
基于神经网络的OCR技术在自动阅卷系统中的应用研究 ······	王	睿,	林	凯	(103)
基于GA-XGBoost算法的河南省粮食产量预测研究	付金	·鹏,	王	哲	(107)

开发案例

基于PyQt的双目图像采集软件设计	•••••	杨昌其,	翟梓希,	惠国腾,	刘晓东	(111)
基于 ZigBee 的智慧园区 LED 路灯无线	浅测控系统设计	•••••	•••••	柴西林,	邵照勇	(117)

Modern Computer

(Vol. 30, No. 6; Mar. 25, 2024)

CONTENTS

Research and Development

Road crack detection method based on YOLOv5	(1)
Gait recognition method based on dual-level temporal features	(8)
Document level event argument recognition based on BERT and graph attention network	(14)
Lightweight human body detection algorithm suitable for edge devices ······	(20)
Research on improved text detection algorithm for prosecutorial scenarios based on DBNet	(26)
Combined Swin Transformer and UNet for GAN face inpainting	(32)
Research on space object classificationbased on LSTM-Attention	(38)
A multilayer encryption algorithm for smart network mobile terminal data link based on	
disturbance judgment	(44)
Phishing website identification based on Bi-LSTM with attention mechanism	(50)
Hough circle detection based on image brightness equalization technology	(56)
Design of road pothole detection based on improved YOLOv5	(61)
Transformer-based protein interaction prediction research	(65)
Research on improved LSD line feature extraction algorithm	(69)

Practice and Experience

Risk assessment of apron control system based on uncertainty	(73)
A maintenance and support mode for shared information system equipment framework based on CBM	
	(80)
Research on the application of knowledge enhanced LLM in fault analysis of information system	(87)
Improved Harris hawks algorithm for pattern shaping of linear antenna arrays	(94)
A privacy protection scheme based on attribute encryption in smart cities	(99)
Research on the application of OCR technology based on neural network in automatic marking system	
	(103)
Grain yield prediction based on GA-XGBoost algorithm in Henan Province	(107)

Development Solution

Design of binocular image acquisition software based on PyQt	(111)
Design of wireless measurement and control system for LED street lights in smart park based on ZigBee	
	(117)

研究与开发

文章编号:1007-1423(2024)06-0001-07

DOI: 10.3969/j.issn.1007-1423.2024.06.001

改进YOLOv5的路面裂缝识别方法

杨景维,黎远松*

(四川轻化工大学计算机科学与工程学院, 宜宾 644000)

摘要:针对于深度学习在道路路面裂缝目标检测上检测精度低、错检、漏检等问题,提出一种基于改进YOLOv5的路面裂缝检测方法。首先在YOLOv5的主干网络中加入S2-MLPv2注意力模块,提高模型的精准定位能力。其次在颈部网络部分,使用GSConv+Slim-neck组合结构减轻模型复杂度的同时提升精度。最后将激活函数换为FReLU,在ReLU和PReLU的基础上进行改进,解决了激活函数中的空间不敏感问题。实验结果表明,与传统YOLOv5相比,本模型的mAP50增加了3个百分点,检测效果较好,能达到准确检测的要求。

关键词:裂缝检测;深度学习;YOLOv5;注意力机制

0 引言

道路路面裂缝检测对交通安全、道路损坏 的维护和汽车智能辅助驾驶都有很重要的意义。 裂缝是沥青混凝土路面最常见的病害之一,不 同地区气候条件和交通量不同,导致沥青路面 裂缝的原因也各不相同。路基结构不稳定、路 面基层结构破坏、沥青面层结构破坏、沥青表 面层性能衰减、路面整体强度不足、交通荷载 的增加,都会导致沥青路面出现各种裂缝^[1]。 传统的人工检测方法往往需要封路影响交通, 且受天气等场外条件限制,检测成本较高,检 测效率较低,其检测结果受检测人员等主观因 素影响,准确率较低[2]。随着科技的发展,道 路检测逐渐开始大量使用全自动化检测技术。 全自动化检测技术在速度、精度方面都远远超 过了传统检测技术,而且检测内容也更加丰富 细致,如无人机检测法和红外热像法,无人机 检测法利用无人机巡航实时拍摄影像画面,其 中高清相机拍摄裂缝形态,激光雷达、测量模 块来获取点云数据,进而建立路面裂缝三维模 型。红外热像法检测路面裂缝的原理就是利用 路面裂缝处的热传导能力与路面其他部位不同, 两处温差较大,会在图像上形成强烈对比,即 路面存在裂缝时, 红外图像上的整体温度会存 在异常,此异常值可能为极大值或极小值[3]。 随着深度学习的快速发展,计算机视觉领域的 学者们对基于深度学习的路面病害检测进行了 研究。任安虎等^[4]在 YOLOv5s 的基础上引入 GAM注意力、SPSF网络和空洞深度可分离卷 积,提升了道路裂缝检测的准确性和对不同背 景下裂缝识别的泛化能力。冯卉^[5]通过在卷积 神经网络模型上应用多个不同尺寸的卷积核, 结合图像注意力机制,赋予不同尺度的特征图 以不同的权重,从而实现对道路裂缝的有效识 别。孙朝云等^[6]通过对Faster-RCNN模型进行优 化,并结合了VGG16、ZFNet和Resnet50网络的 优势,提出了一种基于改进Faster-RCNN模型的 路面灌封裂缝检测方法,从而显著提高了定位 精确度和测试结果的准确性。顾书豪等[7]提出 了一种裂缝自动检测算法,通过增强语义信息 与多通道特征的融合,引入扩张卷积模块和注 意力机制,以提升模型对特征细节的提取能力。 深度学习在图像识别领域表现出绝对优势,深度 卷积神经网络能够提取图像的高级语义特征,无 需对输入特征做预处理即可实现对输入图像进行 自动识别,较传统图像识别方法,效果更佳。

收稿日期: 2023-10-11 修稿日期: 2023-11-21

作者简介:杨景维(1998—),男,四川遂宁人,在读硕士研究生,研究方向为计算机视觉;*通信作者:黎远松(1970—), 男,四川自贡人,硕士,教授,研究方向为自然语言处理、机器学习等,E-mail:530653481@qq.com 本文提出一种基于改进 YOLOv5s 的路面裂 缝检测方法:

(1) 在传统 YOLOv5s 模型基础上,在 Backbone 部分引入 S2-MLPv2 注意力机制^[8],提高目 标的精准定位能力。

(2) 在 Neck 加入 Slimneck+GSConv^[9]组合结构,在减轻模型复杂度的同时提升精度。

(3)使用 FReLU^[10]激活函数代替 SiLU 激活 函数,解决激活函数中的空间不敏感问题。

1 YOLOv5算法及原理

YOLO^[11]系列算法是一种单阶段目标检测算 法,其独特之处在于省略了候选区域生成的步 骤,能够直接获得目标的分类和位置坐标信息。 YOLO算法将目标检测任务视为回归问题,只需对 输入图像进行一次处理,即可同时获取目标的位 置和类别信息。在YOLOv4的基础上,YOLOv5引 入了focus结构,对输入特征图像进行切片操作并 在深度上进行拼接,此外,在网络结构的neck部 分简化了CSP结构,并将Mish激活函数改进为 SiLU激活函数。此外,YOLOv5还采用了CloU损 失函数来进行预测框的定位损失函数计算。

YOLOv5 是一个由 Input、Backbone、Neck 和 Head 四个主要组件构成的目标检测模型。在 Input 部分,模型应用了 Mosaic 变换、自适应锚 框以及自适应图像处理等技术,这些操作不仅 能够丰富检测数据,还能提升模型的鲁棒性。 Mosaic变换通过将多个图像进行拼接,创造更 多样化和丰富的训练样本。这种数据增强技术 能够扩充训练数据规模,并提高模型对于复杂 场景和遮挡物的处理能力。

自适应锚框是一项关键技术,它能够自动 适应不同尺度和长宽比的目标,无需手动设置锚 框。这种自适应性能够提高模型在各种目标上的 检测精度和适应性。另外,自适应图像处理技术 通过动态调整输入图像的亮度、对比度和饱和度 等参数,提供更好的环境适应性。这样的处理手 段使得模型在面对不同环境条件下的图像时能够 更好地捕捉目标。Neck部分由特征金字塔网络 (feature pyramid network, FPN)和路径聚合网络 (path aggregation network, FAN)构成,实现多尺 度语义信息的融合。Head部分将获取到的特征通 过三种不同大小的特征图分别检测大、中、小目 标,再通过 CloU Loss 计算定位损失,并经过非 极大值抑制 NMS(non-maximum suppression)得到 最优预测框,提高预测精度。

2 YOLOv5算法改进

根据道路裂缝的特点,为了提高道路裂缝 检测精度以及保证模型实时性,提出改进的 YOLOv5s算法,其结构如图1所示。



图 1 改进 YOLOv5s 网络结构

如图1所示,在Backbone中SPPF前面嵌入 S2-MLPv2注意力模块,加强模型的特征提取能力。在Neck部分,为了加快预测的计算速度, 引入轻量级卷积方法GSConv代替普通的Conv模块,使用VoV-GSCSP模块代替C3模块,降低了 模型的复杂度,同时维持了足够的准确性。使 用非线性激活函数FReLU替换SiLU,将ReLU 和PReLU扩展成2D激活函数,解决激活函数的 空间不敏感问题。

2.1 S2-MLPv2注意力机制

本文通过引入S2-MLPv2注意力模块,使模 型在检测过程中加大对裂缝的关注程度,减少 其他因素的影响,从而提高目标检测的精度。 S2-MLPv2是在S2-MLP模型上进行改进得到的, MLP block 和 classification head 组成, 它将尺寸 为 $W \times H \times 3$ 的输入图片划分为 $w \times h$ 数量的 patch,所有 patch 的尺寸均为 $p \times p \times 3$, p = $W/w = H/h_{\circ}$ 经过这一步骤后,它经过一个全连 接层将每个patch映射到一个d维向量中。通道平 均划分成四个区域并向不一样的方位进行位移, 最后实现 MLP 操作。spatial-shift 层的输入特征是 大小为 $w \times h \times c$ 的X, X在通道上被平均划分为 四个模块,在每一模块进行四个方位的shift: $[X [2:h, :, 1:c/4] \leftarrow X [1:h - 1, :, 1:c/4]$ $X[1:h-1,:,c/4+1:c/2] \leftarrow X[2:h,:,c/4+1:c/2]$ $X[:, 2:w, c/2:3c/4] \leftarrow X[:, 1:w - 1:c/2:3c/4]$ $X\left[:, 1:w - \frac{1, 3c}{4}:c\right] \leftarrow X\left[:, 2:w, \frac{3c}{4}:c\right]$ (1)

其整体结构如图2所示。

在S2-MLP的基础上,对视觉骨干网络进行 优化,沿着通道方向扩展特征图,然后将扩展 的特征图划分为若干块,对这些划分出来的部 分执行不同的空间位移操作,同时,使用分割 注意力手法来整合这些被划分的部分。此外, 我们同样采用了小型补丁,并且用金字塔结构 来提升图像识别的准确度。这个优化后的空间 位移 MLP视觉骨干网络被命名为S2-MLPv2。

对于X, S2-MLPv2将X在通道上扩张成原 特征的三倍,如下:

$$\hat{\chi} = MLP_1(X) \in \mathbf{R}^{w \times h \times 3c} \tag{2}$$

然后将
$$X$$
划分成 X_1 , X_2 和 X_3 :

$$\begin{cases} X_1 = \hat{\chi} [:, :, 1:c] \\ X_2 = \hat{\chi} [:, :, c+1:2c] \\ X_3 = \hat{\chi} [:, :, 2c+1:3c] \end{cases}$$
(3)

 X_1 和 X_2 分别用 $SS_1(\cdot)$ 和 $SS_2(\cdot)$ 两种方式进行 spatial-shift,这两种方式不是对称的,可以 互相补充信息,而 X_3 则不需要移动。 $SS_2(\cdot)$ 表 达式如式(4)所示:

 $\begin{cases} X_2[:, 2:w, 1:c/4] \leftarrow X_2[:, 1:w - 1, 1:c/4] \\ X_2[:, 1:w - 1, c/4 + 1:c/2] \leftarrow \chi_2[:, 2:w, c/4 + 1:c/2] \\ X_2[2:h, :, c/2:3c/4 \leftarrow X_2[1:h - 1, :, c/2:3c/4] \\ X_2[1:h - 1, :, 3c/4:c \leftarrow X_2[2:h, :, 3c/4:c] \end{cases}$

使用 split attention 将三组特征进行聚合:

$$\hat{X} = SA\left(\left\{X_k\right\}_{k=1}^3\right) \tag{5}$$

将得到的特征送入另一个MLP中增强:

$$\overline{X} = MLP_2(\hat{X}) \tag{6}$$

S2-MLPv2结构如图3所示。



图2 S2-MLP结构图



2.2 GSConv+SlimNeck模块

通常,为了提高推理速度,卷积神经网络中的图像要向通道逐级传递空间信息。GSConv以较低的成本尽量地保留每个通道之间的连接,保护特征图的语义信息。GSConv先是通过一个常规卷积下采样,然后经过DWConv深度卷积,将两者处理后的结果进行拼接操作,最后利用shuffle使两个卷积的对应通道相邻。GSConv结构如图4所示。



图 4 GSConv 结构图

我们使用GSConv来代替标准卷积,能够有效地降低计算成本,并且保持模型的学习能力。 然后,在GSConv的基础上构造GSbottleneck模块,其结构如图5所示。



图 5 GSbottleneck结构图

在 GSbottleneck 模块的基础上引入 VoV-GSCSP。VoV-GSCSP模块减小了计算和网络模型的复杂度,使模型轻量化,并且能保证足够的准确性。使用 VoV-GSCSP 代替 Neck 的 CSP。VoV-GSCSP 结构如图6所示。

2.3 FReLU激活函数

FReLU通过添加可忽略的空间条件开销将

ReLU和PReLU扩展到2D激活,在激活函数阶 段实现像素级的空间信息建模能力,简单又 高效。ReLU和PReLU的形式分别为 $y = \max(x, 0)$ 和 $y = \max(x, px)$,而FReLU的形式为 $y = \max(x, T(x))$,其中 $T(\cdot)$ 为二维空间条件。 利用T(x)和max函数的特性,可以提供像素级 的建模能力或空间布局能力,可以很自然地提 取物体的空间结构。FReLU在非线性激活的同 时产生空间依赖,激活依赖于空间条件,解决 了激活函数空间不敏感的问题。FReLU结构图 如图7所示。







图 7 FReLU 结构图

3 实验结果与分析

3.1 实验环境与参数设置

本文实验操作系统为Windows10,显卡为 RTX 2080Ti, CPU为 Intel Xeon Platinum 8255C CPU@2.50GHz,实验环境使用 Pycharm 作为 IDE, PyTorch 作为深度学习框架, Python 版本 为3.8。本文实验参数配置见表1。

· 5 ·

表 1 实验参数配置

实验参数	参数设置
图片分辨率	640×640
优化器	SGD
学习率	0.01
权重衰减系数	5×10^{-4}
动量因子	0.937
Epoch	300
Batch-size	32

3.2 数据集的构建

实验使用的裂缝图像来自公开数据集 crack-detection^[12]、CFD等以及自行搜集的图像。 为了丰富数据,增强模型在不同干扰条件下的 鲁棒性,利用数据增强技术,将数据图片进行 旋转、噪点、亮度的变化,如图8所示,经过数 据增强,最终获得3163张图片。



图 8 数据增强处理

数据集按照7:1:2划分为训练集、验证集 和测试集。使用 Labeling 对图片中的裂缝进行 标注,将目标裂缝标记为 crack,将标注好的图 片标注信息保存为.xml格式文件,其内容包括 目标的类别和位置坐标信息。部分数据集图像 如图9所示。



图 9 部分数据集图像

3.3 评价指标

本文采用精确率(Precision, P)、召回率 (Recall, R)、平均精确率均值(mAP)、参数量 和计算量来作为模型评价指标。精确率是指在 所有检测出的目标中检测正确的比例;召回率 是指模型正确识别的有目标判例占所有正确判 例的比例;mAP是多个类别的平均精确率均值, 用来评估模型在所有类别上的好坏。评价指标 的计算公式如下:

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$AP = \int_{0}^{1} P(R) \tag{9}$$

$$mAP = \frac{1}{n} \sum_{i=1}^{c} AP_i \tag{10}$$

式中: TP表示检测正确的正样本数量; FP表示 检测错误的正样本数量; FN表示检测错误的负 样本数量; n为类别数量。

3.4 实验结果

图 10 是本文改进算法与 YOLOv5s 算法的 mAP50 曲线图,由图 10 可以看出,两种算法的 mAP50 均在前 100 轮次增长很快,在 100~200 轮次间,mAP50 的增长速度开始逐渐变得缓慢, 在 200~300 轮次间,开始趋于平稳。从图中还 可以看出,本文算法的 mAP50 曲线基本上在 YOLOv5s之上,并且波动起伏相对来说更加平 缓,表明本文算法在裂缝检测的准确性和特征 学习效果上优于 YOLOv5s。



图 10 两种算法 mAP50 曲线图

使用 YOLOv5s 和本文改进模型在自建裂缝数据集上训练,其损失函数曲线如图 11 所示。 由图 11 可以看出,在 0~50 轮次中,损失值下降较快,随着训练轮次的增加,损失值下降速度逐渐变慢,在接近 300 轮次时,曲线趋于平缓,模型收敛,训练过程并未出现过拟合现象,且改进模型算法曲线基本位于 YOLOv5s 模型之上,表明本文算法定位能力更优。



图 11 两种算法损失函数曲线

为了更直观地体现本文改进YOLOv5s模型的检测性能,图12展示了三组同一张图像在两种算法下的检测效果对比。在第①组图像中,对于同一个简单目标,本文改进模型相较于原始YOLOv5s模型呈现出置信度更高。在第②组图像中,对于多个目标,YOLOv5s检测时出现了漏检的情况,而漏检的目标被本文改进模型成功检测出,体现出模型在复杂背景下检测能力的稳定性。在第③组图像中,对于多个目标,YOLOv5s模型出现了重复检测的情况,而本文算法有所优化。



图 12 两种算法检测效果对比

本文算法模型的设计过程中,在YOLOv5s的基础上,在骨干网络中引入S2-MLPv2注意力机制,在Neck部分中使用GSConv+SlimNeck组合模块,最后将激活函数SiLU替换成FReLU。为了验证各模块设计的合理性和有效性,在自制数据集上设计8组消融实验,结果见表2。

表 2 各模块改进的消融实验结果

序号	S2-MLPv2 注意力	GSConv- SlimNeck	FReLU	参数量/M	Precision/%	Recall/%	mAP50/%
1				7.01	84.9	77.3	80.2
2	\checkmark			9.11	84.6	78.6	81.3
3		\checkmark		6.73	81.3	80.3	81.7
4			\checkmark	7.12	84.0	79.1	82.6
5	\checkmark	\checkmark		8.83	83.4	80.4	82.8
6	\checkmark		\checkmark	9.22	85.4	77.4	82.1
7		\checkmark	\checkmark	6.79	85.8	78.3	82.7
8	\checkmark	\checkmark	\checkmark	8.89	86.0	79.3	83.2

由表2可以看出,每一个改进模块都对模型 起到了一定的促进作用。将S28-MLPv2注意力机 制融入模型骨干网络,相较于原模型,mAP50提 升了1.1个百分点,但参数量有所增长。将 GSConv+SlimNeck组合模块融入Neck部分, mAP50提升1.5个百分点的同时,参数量减少了 0.28M。将激活函数替换成FReLU,参数量仅增 加0.11 M的同时,mAP50提升了2.4个百分点。 最后,将S2-MLPv2注意力和GSConv+SlimNeck 模块同时融入模型,并使用FReLU激活函数,在 模型参数量增长1.88 M的同时,模型的各项指标 得到了很大的提升,其中,Precision、Recall、 mAP50分别提升了1.1、2.0以及3.0个百分点。

为了验证本文改进的YOLOv5s网络模型的 性能,将其与经典网络YOLOv4、YOLOv5s、 YOLOX、SSD进行对比,结果见表3。

表 3 经典目标检测网络对比实验

序号	算法	参数量/M	Precision/%	Recall/%	mAP50/%
1	YOLOv4	64.36	82.46	58.59	73.94
2	YOLOX-s	8.97	83.60	75.70	81.40
3	YOLOv5s	7.01	84.90	77.30	80.20
4	SSD	26.29	85.81	69.35	81.28
4	Ours	8.89	86.00	79.30	83.20

由表3可以看出,相较于几种经典的一阶段 目标检测算法,YOLOv4算法模型体积过于庞大, 参数量高达64.36 M,但并没有达到较高的检测 精度,mAP50 仅达到 73.94%。而 YOLOX-s 模型 各项指标均低于本文改进算法。SSD 算法参数量 较大,并且其他指标低于本文算法。本文改进算 法由于引入 S2-MLPv2 注意力模块,参数量略大 于 YOLOv5s 模型,但是在检测精度方面有所提 高,mAP50 达到了 83.2%,相比于 YOLOv5s, mAP50 提高了 3 个百分点,并且其余几项评估 指标均取得了最优值。

4 结语

本文针对道路裂缝检测任务中精度低、错 检、漏检等问题,提出一种改进YOLOv5s的网 络模型。通过引入S2-MLPv2注意力机制、用 SlimNeck+GSConv组合模块构建模型Neck部分、 使用FReLU激活函数代替原本的SiLU激活函数 等方式进行网络优化。实验结果表明,本文改 进算法的检测精度达到了86.0%,召回率达到了 79.3%,mAP50达到了83.2%,性能优于原始 YOLOv5s模型,能够有效提取特征信息,满足 对道路裂缝准确检测的要求。

参考文献:

- [1] 周伟红.高速公路沥青路面养护中裂缝病害的评价
 [J].科技资讯,2023,21(15):117-121.
- [2] 庄一舟,万明.深度卷积神经网络在道路病害检测中的应用研究[J].信息与电脑(理论版),2022,34
 (20):175-178,182.
- [3] 王兆昌.公路沥青路面病害检测与养护[J]. 交通世 界,2023(21):72-74.

- [4] 任安虎,美子渊,马晨浩.基于改进YOLOv5s的道路裂缝检测算法[J/OL].激光杂志:1-7
 [2023-09-22]. http://kns. cnki. net/kcms/detail/50.1085.TN.20230626.1756.012.html.
- [5] 冯卉.基于深度学习的道路裂缝识别算法研究与实现[D].北京:北京邮电大学,2019.
- [6] 孙朝云,裴莉莉,李伟,等.基于改进Faster R-CNN 的路面灌封裂缝检测方法[J]. 华南理工大学学报 (自然科学版),2020,48(2):84-93.
- [7] 顾书豪,李小霞,王学渊,等.增强语义信息与多通 道特征融合的裂缝检测[J].计算机工程与应用, 2021,57(10):204-210.
- [8] YU T, LI X, CAI Y, et al. S2-MLP: spatial-shift MLP architecture for vision[C]//Proceedings of the IEEE/ CVF Winter Conference on Applications of Computer Vision, 2022:297-306.
- [9] LI H, LI J, WEI H, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles [EB/OL]. arXiv: 2206.02424, 2022.
- [10] MA N, ZHANG X, SUN J. Funnel activation for visual recognition [C] // Proceedings of the 16th European Conference on Computer Vision (ECCV 2020), Glasgow, UK, 2020: 351-368.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [12] XU H Y, SU X, WANG Y, et al. Automatic bridge crack detection using a convolutional neural network[J]. Applied Sciences, 2019, 9(14):2867.

Road crack detection method based on YOLOv5

Yang Jingwei, Li Yuansong*

(College of Computer Science and Engineering, Sichuan University of Science and Engineering, Yibin 644000, China)

Abstract: Aiming at the problems of low detection accuracy, error detection and missing detection in road surface crack target detection by deep learning, a road surface crack detection method based on improved YOLOv5 is proposed. Firstly, the S2-MLPv2 attention module was added to the backbone network of YOLOv5 to improve the model's accurate positioning ability. Secondly, in the neck part, GSConv+Slim-neck combined structure is used to reduce the complexity of the model and improve the accuracy. Finally, the activation function is changed to FReLU, and ReLU and PReLU are extended to 2D activation function, which solves the space insensitivity problem in activation function. The experimental results indicate that compared to the traditional YOLOv5, this model achieves a 3-percentage-point increase in mAP50, demonstrating better detection performance and meeting the requirements for accurate detection.

Keywords: crack detection; deep learning; YOLOv5; attention mechanism

文章编号:1007-1423(2024)06-0008-07

DOI: 10.3969/j.issn.1007-1423.2024.06.002

基于双尺度时间特征的步态识别方法

魏永超1,徐未其2*,朱泓超2,朱姿翰3,刘伟杰2

(1. 中国民用航空飞行学院科研处,德阳 618307; 2. 中国民用航空飞行学院民航安全工程学院,德阳 618307;3. 中国民用航空飞行学院航空电子电气学院,德阳 618307)

摘要:当前,多数步态识别方法关注于步态序列单一时间尺度建模,忽略了不同时间尺度的信息交互。基于此, 提出了一种双尺度时间特征表示网络。该方法聚合两个时间尺度特征来获取步态的运动表示,并将两个时间尺度上特 征进行融合,实现信息交互。通过多视角识别实验验证,该方法在数据集 CASIA-B上的性能超越了主流的步态识别方 法,在NM、BG和 CL条件下 Rank-1 准确率分别达到 97.8%、93.1% 以及 80.6%。

关键词:步态识别;时间尺度;空间特征;多视角识别

0 引言

近些年,随着生物识别技术的发展,具有 远距离识别、配合度低等特性的步态识别技术, 已经成为计算机视觉领域的研究热点,且在社 会安全和身份识别等方面展现出巨大应用前 景^[1-3]。步态识别根据人行走的方式来进行远距 离身份验证^[4],其本质是以人行走姿态作为主 要特征依据,再结合时空关系表示方法实现识 别。但在实际场景中,受到行人衣着、携带状 态及采集视角等因素的影响,如何获得判别性 步态特征表示是当前面临的挑战。

为了应对挑战,众多学者进行了相关研究。 依据不同时间建模方式分为隐式方法和显式方 法。隐式方法通常是以集合化方式提取步态特 征,例如GaitSet^[5]和GaitTB^[6]在判别性特征表示 上,均使用池化操作把连续步态序列聚合为集 合特征。隐式方法虽然快速、灵活,但损失了 步态序列前后时间关联性,包含的信息有限。

基于隐式方法的问题,显式方法则是对步

态轮廓序列建立模型,对信息进行时间编码, 以提高步态特征的前后时间关联性。据时间序 列不同尺度划分,可分为短时相关性特征编码 和长时相关性特征编码。短时特征编码的是局 部帧级序列时空关系的特征表示,例如Gait-Part^[7]从人体外观上的局部细节出发,利用微动 作捕捉模块对时间的依赖性进行建模。Sun 等^[8] 提出了短时步态识别网络,对身体不同区域内 的关节运动数据进行独立时间序列编码。短时 特征编码能获取序列内局部帧的时间相关性, 提升识别性能, 但忽略了整体序列的时间相关 性。长时特征编码关注的是步态在整个序列中 的变化情况,例如Thomas应用3DCNN^[9]提取整 个序列的时空信息,利用提取的时空信息来描 述人类步态。GaitNet^[10]是从步态图像中提取相 关特征,然后用LSTM 对步态序列的时间变化进 行建模。长时特征编码将人体作为一个单元提 取序列中的时空信息,用来表示步态序列全局 描述,在性能上有所提升。但长序列时间建模 保留了冗余的信息,约束了步态识别的灵活性。

收稿日期: 2024-01-29 修稿日期: 2024-03-01

基金项目:西藏科技厅重点研发计划(XZ202101ZY0017G);四川省科技厅重点研发项目(2022YFG0356);中国民用 航空飞行学院科研基金(J2020-040、CJ2020-01)

作者简介: 魏永超(1981—),男,河南许昌人,博士,教授,研究方向为计算机视觉、机器视觉; *通信作者:徐未其(1994—),男,四川成都人,硕士研究生,研究方向为深度学习、步态识别,E-mail:539892287@qq.com; 朱泓超(1998—), 男,四川成都人,硕士研究生,研究方向为机器学习、图学习; 朱姿翰(1998—),女,山西临汾人,硕士研究生,研究方向为深 度学习、多模态融合; 刘伟杰(1993—),男,四川成都人,硕士研究生,研究方向为计算机视觉 综合上述研究,隐式时间建模着重关注空间中的静态信息,但丢失了时间序列上的特征 表;显式时间建模关注步态在时间上的前后关 联性表达,而不同时间尺度的时空特征表示都 有限,未能对步态特征进行多样化表达。针对 上述问题,本文提出了双尺度时间特征表示学 习方法,即帧级特征和长短期时间特征。其中, 帧级特征是每个时刻的帧特征,长短期时间特 征则是融合了短序列和长序列的特征,增加步 态时空信息。

1 步态识别模型

本文提出的模型总体如图1所示,该网络由 浅层卷积网络、双尺度时间提取模块、双尺度 空间学习模块三个模块组成。采用与文献[5]相 同的方式,将一批N帧的M个步态样本作为网 络的输入,表示为 $F_{input} \in \mathbf{R}^{T \times H \times W}$ 。首先,通过 4层卷积网络和特征复用对步态序列进行操作, 使得到的浅层全局特征包含更多的空间特征信 息,产生特征 $F \in \mathbf{R}^{C \times T \times H_i \times W_i}$,其中H,W,T, C分别表示为步态帧的高度、宽度、帧数和通道 数。然后,以F作为双尺度时间提取模块的输 入,生成具有两种不同尺度的时间特征,即帧 级和长短期。将人体步态特征划分成对应身体 部位的K份,此时两个尺度时间特征 T_f 、 T_s , T_s 大 小为 $\mathbf{R}^{T \times C \times K}$ 。接下来将两个时间尺度特征聚合, 得到时间特征 $T \in \mathbf{R}^{M \times C \times K}$ 。另外,将两个时间 尺度特征作为双尺度空间学习模块的输入,重 组空间特征 $S \in \mathbf{R}^{M \times C \times K}$,其中M表示步态样本 的数量。最后,时间特征和空间特征在通道维 度上通过全连接作为输出。





1.1 双尺度时间特征提取模块

正如GaitPart所提的思想,步态是一种与时间有关的运动模式。不同的识别对象,其运动模式特征在显著性上各有异同,特征变化显著的识别对象能在短时间周期内被识别;而特征变化细微的识别对象则需要在较长时间周期内才能被识别。

基于此特性,本文提出了双尺度时间提取 模块,如图2所示,该方法是将短时间特征和长 时间特征进行整合,与帧级特征结合,实现双尺 度时间之间的信息交互,从而包含更多的时空特 征。采用两个串行的一维卷积提取短时间周期特 征,获得步态序列中局部时间线索,这些线索对 时间变换上的位置信息敏感,不利于提取细微运 动模式。因此,在得到步态序列中局部时间信息 后,将序列中各个判别性特征组合,形成长时间 特征,从而增加特征信息量。



经过基础卷积后,得到了特征F。将人体步态整体特征F划分为K部分,在其基础上做全局 池化,得到 $F \in \mathbf{R}^{\tau \times c \times \kappa}$ 。其中 F_m^* 表示部分层级 上第*m*个样本中第*n*帧的特征。如图2所示,直接将*F*作为帧级特征。

为了捕获短期时间特征,应用两个核大小为3的串行一维卷积,并将前一个卷积的结果与 后一个卷积的结果相加得到短尺度时间特征*T*_s。如式(1)所示,*T*ⁿ_s表示第*m*个样本的第*n*帧的短 时间特征。

$$T_{s} = Conv_{1\times 1} \oplus Conv_{1\times 1} (Conv_{1\times 1}(T_{f}))$$
(1)

长时间尺度特征是在短时间尺度特征基础 上来提取。首先对L_s应用多层感知机(MLP)和 Sigmoid 函数来评估不同帧在短时间特征上的重 要性,再对所有帧的短时间重要性分数进行加权 求和,作为双时间尺度特征L_{sl}。如式(2)所示:

$$T_{sl}^{m} = \frac{\sum_{n=1}^{N} Sigmoid \left(MLP(T_{m}^{n}) \right) \odot T_{m}^{n}}{\sum_{n=1}^{N} Sigmoid \left(MLP(T_{m}^{n}) \right)}$$
(2)

式(2)中: \odot 表示点积, T_{sl}^{m} 表示第m个样本的全局运动线索。

1.2 双尺度时间特征融合模块

得到不同尺度时间特征后,为了实现特征 信息交换,时间特征进行融合。融合方式如图3 所示。



图 3 时间特征融合

利用帧级和双时间尺度的信息学习时间重要性权重。通过两个全连接层和一个Sigmoid函数来实现,时间重要性权重 W_r 如式(3)所示:

 $W_{T} = Sigmoid\left(FC\left(FC(T_{f} \odot T_{sl})\right)\right)$ (3)

式(3)中: ©表示串联运算操作, W_r 结合了两个时间尺度上的重要性权重, $W_r \in \mathbf{R}^{T \times 2 \times C \times K}$, 而 $W_r^{m,n}$ 表示第m个样本中的第n帧的时间重要性权重。然后在两个时间尺度上引入注意力机制获得注意力时间特征, 如式(4)所示:

 $T_{A}^{m,n} = T_{f}^{m,n} \odot W_{T,1}^{m,n} + T_{sl}^{m,n} \odot W_{T,2}^{m,n}$ (4) 对于跨尺度时间融合,本文用加权求和的 方式获得第m个样本在序列上的表示 T_{m} ,如式 (5)所示。其中 $T = \{T_{m} | m = 1, \dots, M\}$ 。时间关系 上的建模使网络能选择时间上的运动特征,学 习运动模式。

$$T_{m} = \frac{\sum_{n=1}^{N} T_{A}^{m,n}}{\sum_{n=1}^{N} \sum_{i=1}^{2} W_{T,i}^{m,n}}$$
(5)

1.3 双尺度空间特征学习

本文方法将 MLP 和 Sigmoid 函数应用于两个时间尺度特征,用来产生每帧的部分分数 $T_s^{m,n}$,其定义如式(6)所示:

$$P_{s}^{m,n} = Sigmoid\left(MLP(T_{f}^{m,n} \otimes T_{sl}^{m,n})\right)$$
(6)
$$\tilde{P}_{s}^{m,n} = \frac{P_{s}^{m,n}}{\sum_{s}^{N} P_{s}^{m,n}}, \quad \ddagger \stackrel{\sim}{\to} \tilde{P}_{s}^{m,n} \in \mathbb{R}^{1 \times K} \\ \ddagger \stackrel{\sim}{\to} \tilde{R}^{m,n}$$

中第n帧的部分分数,而 $\tilde{P}_{s}^{m,nk}$ 表示第m个样本 中第n帧的第k部分分数。部分分数代表了局部 特征判别行的高低,得分越高表示空间特征越 明显。为了监督特征的正确性,本文在 T_f 和 \tilde{P}_s 的加权和上使用了具有交叉熵损失的全连接层。 首先,第m个全连接层样本的加权部分特征表 示如式(7)所示:

$$P_w^m = FC\left(\sum_{n=1}^N T_f^{m,n} \odot \tilde{P}_s^{m,n}\right)$$
(7)

式(7)中: $P_w^m \in \mathbf{R}^{C_t \times K}$, C_t 表示训练样本数量。 然后,对 P_w^m 应用交叉熵损失产生式(8):

$$L_{ce} = -\sum_{m=2}^{M} \sum_{c=1}^{C_{i}} y_{b,c} \operatorname{Log}\left(\operatorname{softmax}\left(p_{w}^{m}\right)\right)_{c} \quad (8)$$

式(8)中: *y_{b,c}*表示第*b*个样本的身份信息,其值为0或1。然后得到在时间上的维度分数最高的部分索引,如式(9)所示:

$$x_m^k = \arg\max_s P_s^{m,n,k} \tag{9}$$

式(9)中: x_m^k 表示在M个样本序列中第m个样本中选定的第k个部分的时间索引。然后通过硬注意力机制在时间索引 $\{x_m^k | k = 1, 2, \dots, k\}$ 的引导下获得重组帧特征 R_m ,如式(10)所示:

$$Loss = L_{ce} + L_{tri} \tag{10}$$

2 实验与分析

2.1 数据集简介与划分

CASIA-B^[11]是中国科学院自动化研究所提 供的步态数据集,是目前最标准、最主流和应 用最广泛的数据集。该数据集是由124个受试者 的步态数据组成,每个受试者的数据有3种行走 条件,即正常行走(NM)、背包行走(BG)和穿外 衣行走(CL),每种行走条件包含了11个视图 (从0°~180°,每个角度的间隔为18°)。对应上 述条件,每名受试者总计110个序列。

CASIA-B数据集不存在官方规定的训练集 和测试集,本文采用文献[5]中的三种划分方 式,见表1,分别为小样本训练(ST)、中样本训 练(MT)和大样本训练(LT)。在ST中,前24个 受试者用于训练,其余100个受试者用于测试; 在MT中,前62个受试者用于训练,其余62个 受试者用于测试;在LT中,前74个受试者用于 训练,其余50个受试者用于测试。每种划分方 式下均采用正常行走状态的前4个序列(nm-01, nm-02, nm-03, nm-04)作为注册库,剩余的6 个序列(nm-05, nm-06, bg-01, bg-02, cl-01, cl-02)作为查询集。

表 1 数据集划分

Division	Train	Test
Small-Sample(ST)	24	100
Mid-Sample(MT)	62	62
Large-Sample(LT)	74	50

2.2 实验设置

本文实验基于深度学习框架 PyTorch 以及 RTX 3090 搭建训练测试平台。为验证算法性能, 控制实验变量,本文统一实验中网络模型的相 关超参数,具体超参数见表2。

模型训练的批次大小为 8 × 8 × 30,即随机 选取 8 个不同的人,每个人随机选取不同穿戴条件 及视角下的 8 个步态序列,然后从每个步态序列中 抽取 30 帧图像。每帧图像裁剪大小为 64 × 64, *K* 为步态特征划分数量,设置为 32。模型采用 Adam 优化器,学习率设为 1E-2。此外,在三元 组损失函数的超参数设置中,Margin设置为 0.2, 用于控制三元组样本对的类内和类间距离。程 序总迭代数为150000次。

表 2 超参数

Hyper-Parameters	Value
Batch size	(8, 8)
Resolution	64×64
Sequence Length	30
Optimizer	Adam
Lr	1E-2
Iterations	150000
Loss	Triplet
Margin	0.2
Κ	32

2.3 评价指标

在步态识别任务中, Rank-1准确率表示图 像的第一检索目标的准确率,并作为模型泛化 性能评价指标。首先定义一个判定函数,其计 算表达式如下:

$$J(Label_{q}, Label_{g}) \triangleq \begin{cases} 1, & \text{if } Label_{q} = Lab \\ 0, & \text{if } Label_{g} \neq Lab \end{cases} (11)$$

其中: Label_q与Label_g分别代表查询集标签以及 注册集标签。查询集标签通过网络预测特征距 离得到。而后续计算Rank-1准确率时,只需要 统计查询结果的正确数与查询总数,公式如下:

$$Rank-1 = \frac{1}{\|Q\|} \sum_{q=Q} J(Label_q, Label_g) \quad (12)$$

其中: Q代表查询集数量, 即检索总次数。

2.4 实验结果

为验证本文算法模型性能,在CASIA-B上 将本文所提出方法与主流方法进行比较,主流 方法包括GaitSet^[5]、GaitTB^[6]、GaitPart^[7]、Gait-Net^[10]和MGAN^[12],实验结果见表3。

依据数据集的不同划分方式进行分析,在 ST划分方式下,本文方法与GaitSet相比,在跨 视角和行走条件下都有显著提升。与GaitTB相 比,仅在CL行走条件下性能最佳。GaitTB使用 了整体加局部的双分支结构,在NM和BG条件 下能获得更多步态特征,但在CL条件下的特征 表示能力有限。在MT和LT划分方式下,本文

	Gallerv NM	11-4						$0 \sim 180^{\circ}$						
	Probe		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	- Mean
		GaitSet	64.6	83.3	90.4	86.5	80.2	75.5	80.3	86.0	87.1	81.4	59.6	79.5
	NM5-6	GaitTB	76.2	87.8	93.8	91.8	83.6	78.5	83.4	90.4	93.1	86.4	72.7	85.2
		Ours	72.7	85.4	92.4	89.4	79.1	75.8	80.7	88.1	89.3	83.5	74.2	82.8
		GaitSet	55.8	70.5	76.9	75.5	69.7	63.4	68.0	75.8	76.2	70.7	52.5	68.6
ST	BG1-2	GaitTB	68.8	79.1	85.5	82.6	76.8	71.0	76.9	83.2	84.8	78.8	65.3	77.5
		Ours	63.3	73.3	80.3	78.1	69.2	63.7	70.3	77.0	79.5	72.2	62.0	71.7
		GaitSet	29.4	43.1	49.5	48.7	42.3	40.3	44.9	47.4	43.0	35.7	25.6	40.9
	CL1-2	GaitTB	40.9	50.0	54.5	52.3	49.1	46.3	47.6	50.6	44.7	39.9	32.4	46.2
		Ours	38.3	50.4	54.3	53.8	51.9	49.8	52.5	55.3	52.7	46.8	38.5	49.5
		MGAN ^[12]	54.9	65.9	72.1	74.8	71.1	65.7	70.0	75.6	76.2	68.6	53.8	68.1
	NME (GaitSet	86.8	95.2	98.0	94.5	91.5	89.1	91.1	95.0	97.4	93.7	80.2	92.0
	NM3-6	GaitTB	90.5	96.4	99.0	97.5	93.7	90.5	93.9	97.9	98.5	94.8	85.4	94.4
		Ours	94.0	97.9	98.9	97.8	94.6	92.5	95.2	98.4	99.2	97.1	91.0	96.1
		MGAN ^[12]	48.5	58.5	59.7	58.0	53.7	49.8	54.0	61.3	59.5	55.9	43.1	54.7
МТ	PC1 2	GaitSet	79.9	89.9	91.2	86.7	81.6	76.7	81.0	88.2	90.3	88.5	73.0	84.3
MI I	DG1-2	GaitTB	85.4	92.7	95.2	93.8	87.7	84.1	88.2	94.1	94.5	91.2	81.1	89.8
		Ours	88.2	93.9	95.5	93.3	87.5	81.6	88.2	93.2	96.0	92.9	86.3	90.6
		MGAN ^[12]	23.1	34.5	36.3	33.3	32.9	32.7	34.2	37.6	33.7	26.7	21.0	31.5
	CL1-2	GaitSet	52.0	66.0	72.8	69.3	63.1	61.2	63.5	66.5	67.5	60.0	45.9	62.5
		GaitTB	64.6	76.6	82.7	76.7	70.2	64.6	71.3	74.7	73.3	69.6	54.1	70.8
		Ours	70.4	80.6	84.3	81.6	75.3	71.1	74.8	81.5	82.7	78.6	63.7	76.8
		GaitNet	91.2	92.0	90.5	95.6	86.9	92.6	93.5	96.0	90.9	88.8	89.0	91.6
		GaitSet	90.8	97.9	99.4	96.9	93.6	91.7	95.0	97.8	98.9	96.8	85.8	95.0
	NM5-6	GaitTB	93.5	96.4	99.0	97.5	93.7	90.5	93.9	97.9	98.5	94.8	85.4	96.7
		GaitPart	94.1	98.6	99.3	98.5	94.0	92.3	95.9	98.4	99.2	97.8	90.4	96.2
		Ours	96.0	99.5	99.6	98.0	96.0	96.1	97.4	99.4	99.9	98.5	95.5	97.8
		GaitNet	83.0	87.8	88.3	93.3	82.6	74.8	89.5	91.0	86.1	81.2	85.6	85.7
		GaitSet	83.8	91.2	91.8	88.8	83.3	81.0	84.1	90.0	92.2	94.4	79.0	87.2
LT	BG1-2	GaitTB	91.0	95.6	96.3	95.5	91.8	87.7	91.5	95.0	96.9	94.6	85.5	92.8
		GaitPart	89.1	94.8	96.7	95.1	88.3	94.9	89.0	93.5	96.1	93.8	85.8	91.5
		Ours	91.6	96.4	96.3	94.0	89.4	85.6	89.2	94.9	97.5	97.0	91.5	93.1
		GaitNet	42.1	58.2	65.1	70.7	68.0	70.6	65.3	69.4	51.5	50.1	36.6	58.9
		GaitSet	61.4	75.4	80.7	77.3	72.1	70.1	71.5	73.5	73.5	68.4	50.0	70.4
	CL1-2	GaitTB	72.4	84.2	87.2	81.8	77.2	75.1	78.6	81.4	81.2	78.3	62.3	78.1
		GaitPart	70.7	85.5	86.9	83.3	77.1	72.5	76.9	82.2	83.8	80.2	66.5	78.7
		Ours	74.9	85.7	85.8	82.4	79.4	76.8	78.9	81.4	85.5	84.5	71.0	80.6

表 3 CASIA-B数据集 3 种划分方式下的平均 Rank-1 准确率(%)

方法在多视角和行走条件下都具有明显优势, 尤其是在行走条件更加复杂的CL子集下优势更 为明显。由此可证明,不同数据集划分方式, 本文方法在CL条件下泛化性能更佳。依据行走 条件进行分析,NM子集为无干扰的正常行走条 件,与GaitPart相比,本文方法在前视角、侧视 角和后视角(0°、90°、180°)的识别性能更高。 前者是基于短时空特征表征,所包含的步态时 空多样化表达有限,整体性能与本文算法在三 个行走条件下有1.6%的差距。而与其他方法相 比,本文方法在不同行走条件下的所有视角都 达到了最佳识别精度,证明了本文方法的双时 间尺度特征提取器的优势。在有干扰的BG与 CL行走条件下,GaitPart的性能都显著下降,但 在这两个条件下的性能分别提升了 1.6% 和 1.9%。在相同条件下,与其他方法相比,本文 算法性能提升更加显著。

2.5 时间尺度验证实验

为了探究不同时间尺度的时空特征性能表现,本文以帧级特征作为基准,在此基础上分别引入短尺度、长尺度和双尺度时间特征,验证在三个行走条件下的性能。具体结果见表4。

Multi-scale Features				Rank-1	Accurac	y	
Frame- level	Short- term	Long- term	Dual- level	NM	BG	CL	Mean
\checkmark				96.7	91.1	78.0	88.6
\checkmark	\checkmark			97.5	93.1	80.1	90.2
\checkmark		\checkmark		97.1	92.3	79.5	89.6
\checkmark			\checkmark	97.8	93.1	80.6	90.5

表 4 不同时间尺度下的平均准确率(%)

从表4可知:①通过在帧级特征基础上引入 不同尺度时间特征,对比前三组实验结果,短尺 度和长尺度在帧级特征学习的基础上提高了识别 性能,证明了时间信息建模的有效性;②对比后 三组实验可知,双尺度时间特征学习的性能优 于单一尺度时间特征,这表明了两个尺度上的 特征能在时间关系上形成互补;③在NM条件 下,不同时间尺度在整体性能上相差不大。但 在BG和CL条件下的性能,则有2%与2.6%的 提升。由此可以证明,在复杂行走条件下,双 尺度时间特征更适用于步态的多样化表达。

3 结语

本文提出了一种双时间尺度特征表示学习 方法,通过双尺度时间建模方法获取更具判别 性的时空信息,再融合两个尺度上的时空步态 特征,实现更高精度的步态识别。实验结果表 明,与隐式时间建模方法相比,本文方法的平 均Rank-1准确率具有显著优势。与显式单一尺 度时间建模方法相比,本文方法的平均Rank-1 准确率提升了1.7%,验证了本文算法的有效性。

参考文献:

[1] BALAZIA M, PLATANIOTIS K N. Human gait recognition from motion capture data in signature poses[J]. IET Biometrics, 2017, 6(2): 129-137.

- [2] MACOVECIUC I, RANDO C J, BORRION H. Forensic gait analysis and recognition:standards of evidence admissibility [J]. Journal of Forensic Sciences, 2019,64(5):1294-1303.
- [3] PREMALATHA G, V CHANDRAMANI P. Improved gait recognition through gait energy image partitioning [J]. Computational Intelligence, 2020, 36 (3):1261-1274.
- [4] 任娜,郭敏,张景虎,等.基于步态特征的人体身份 识别[J].火力与指挥控制,2008(7):5-8,12.
- [5] CHAO H, HE Y, ZHANG J, et al. Gaitset: regarding gait as a set for cross-view gait recognition [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1):8126-8133.
- [6] 张智,常超伟,王雷,等.结合整体和局部特征的步态识别方法[J].火力与指挥控制,2023,48(4): 141-146.
- [7] FAN C, PENG Y, CAO C, et al. Gaitpart: temporal part-based model for gait recognition [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020:14225-14233.
- [8] SUN W, LU G, ZHAO Z, et al. Regional time-series coding network and multi-view image generation network for short-time gait recognition [J]. Entropy, 2023, 25(6):837.
- [9] WOLF T, BABAEE M, RIGOLL G. Multi-view gait recognition using 3D convolutional neural networks
 [C] //Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), 2016: 4165-4169.
- ZHANG Z, TRAN L, YIN X, et al. Gait recognition via disentangled representation learning [C]
 //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 4710-4719.
- [11] YU S, TAN D, TAN T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition [C]//Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06),2006,4:441-444.
- [12] HE Y, ZHANG J, SHAN H, et al. Multi-task GANs for view-specific feature learning in gait recognition
 [J]. IEEE Transactions on Information Forensics and Security, 2018, 14(1):102-113.

(下转第55页)

文章编号:1007-1423(2024)06-0014-07

DOI: 10.3969/j.issn.1007-1423.2024.06.003

基于BERT和图注意力网络的篇章级事件论元识别

王凯*,廖涛

(安徽理工大学计算机科学与工程学院,淮南 232001)

摘要:事件论元识别是事件抽取的子任务之一,其目的在于识别文本中与事件相关的论元及论元对应的论元角色。 研究表明,句子的依存句法关系有助于事件论元任务识别,然而,在构造篇章的依存句法关系时容易引入不相关的论 元产生噪声问题,现有方法对噪声问题处理不佳。针对该问题,提出了一个基于BERT和图注意力网络的篇章级事件论 元识别模型。该模型从两个角度去解决噪声问题,一方面,通过获取充分的篇章语义特征作为辅助,去构建更有效的 篇章依存句法特征;另一方面,采用图注意力网络对不同的论元节点分配不同的权重,从而去除掉无效的论元。在 RAMS语料库上的实验结果表明,该方法有效解决了篇章依存句法关系中存在的噪声问题,取得了较好的篇章级事件 论元识别结果。

关键词: 篇章级事件论元识别; 依存句法关系; BERT; 图注意力网络

0 引言

事件抽取作为信息抽取¹¹的子任务之一,其 目的是识别文本中的事件并以结构化的形式表示 出来。事件抽取一般包含两个子任务:①事件检 测;②事件论元识别。前者旨在识别文本中所 包含的事件,通常以识别文本中触发词并分类 的方式实现。后者旨在识别文本中与事件相关 的事件论元并为其分配对应的论元角色。本文 的主要研究目标是事件论元识别任务。

现有研究大多集中在句子级事件论元识别, 即假定事件的触发词和论元在同一句子中,然 而在现实文本中往往存在事件论元分散在不同 句子中的现象。表1为RAMS数据集中该现象的 一个例子,在该例子中死亡事件的攻击者和受 害者不在同一句子中。对于这样的事件,现有 的句子级事件论元识别方法难以取得较好的抽 取效果。因此,越来越多的研究者开始了篇章 级事件论元识别的研究。

依存句法分析作为自然语言处理的基础任

务之一,其目的是获取句子中词与词的依存关 系。在事件论元识别任务中,触发词与论元、 论元与论元之间通常存在依存关系,融合依存 句法特征可以有效地提升事件论元识别精度。 依存句法分析已在篇章级事件论元任务中取得 了优异的效果,然而由于篇章中存在大量的论 元,在构造篇章的依存句法关系时容易引入与 事件不相关的论元,从而产生噪声问题,现有 方法对于噪声问题的处理不佳。

针对上述问题,本文提出了一个基于BERT 和图注意力网络的篇章级事件论元识别模型。 首先,通过BERT和Bi-GRU的组合获取篇章的 语义特征;其次,采用自然语言处理工具获取 篇章的依存句法关系树,并进行一定的剪枝操 作去掉不相关的论元;然后,将处理后的篇章 依存句法树的邻接矩阵输入图卷积网络中捕获 篇章的依存句法特征,并引入图注意网络,进 一步获取更优异的特征表示;最后,融合篇章 的语义特征和依存句法特征,并输入Softmax分 类器中进行分类。

收稿日期: 2023-10-12 修稿日期: 2023-11-03

基金项目:国家自然科学基金面上项目(62076006);安徽省高等学校自然研究基金资助项目(KJ2016A202);安徽 省高校优秀青年人才支持计划项目(gxyq2017007)

作者简介:*通信作者:王凯(1999—),男,江西景德镇人,硕士研究生,研究方向为自然语言处理,E-mail: 973835226@qq.com;廖涛(1977—),男,安徽淮南人,博士,副教授,研究方向为Web数据挖掘

表 1 事件论元分散示例表

- S1: Transportation officials are urging carpool and teleworking as options to combat an expected flood of drivers on the road.
- S2: A <u>Baltimore</u> prosecutor accused a police detective of "sabotaging" investigations related to the death of <u>Freddie Grav</u>, accusing him of fabricating notes to suggest that the states medical examiner believed the manner of death was an accident rather than a <u>homicide</u>.
- S3: The heated exchange came in the chaotic sixth day of the trial of Baltimore Officer <u>Caesar Goodson Jr</u>.who drove the police van in which Gray suffered a fatal spine injury in 2015.

Event type	life.die.deathcausedbyviolentevents					
Б	killer: Caesar Goodson Jr; victim: Freddie					
Event arguments	Gray; place: Baltimore					

注: 单下划线标注为事件触发词, 双下划线标注 为事件论元。

1 相关工作

事件论元识别作为事件抽取的关键子任务, 其精度影响着事件抽取任务的结果。近年来, 随着深度学习的快速发展,基于深度学习的事 件论元识别方法取得了较大的发展。根据现有 方法处理的文本粒度的不同,事件论元识别方 法可以分为句子级事件论元识别方法和篇章级 事件论元识别方法。

1.1 句子级事件论元识别方法

句子级事件论元识别研究单个句子中识别 事件相关论元,并为其分配对应的论元角色。现 有的大多数研究将论元角色作为独立的个体进行 识别,忽略了论元角色之间的交互。Wang等^[2]提 出了一种层次模块化事件论元识别模型, 该模 型通过为论元角色建立概念层次,从而从概念 的角度获取论元角色的相关性,提升了事件论 元识别任务的精度。Nan等^[3]提出了一种显示角 色交互网络, 该网络可以动态捕获论元角色之 间的相关性,从而获取不同论元角色之间的关 系,提升了事件论元识别任务的精度。You等[4] 提出了一种基于事件图的事件抽取模型,该模 型将事件抽取转换为事件图解析的方式,充分 获取了触发词和事件论元之间的交互信息,提 升了事件论元识别任务的精度。研究表明,句 子的依存句法关系可以获取词与词之间的依存 关系,该关系有助于建立事件论元之间的交互,

对事件论元识别任务有着重要作用。王士浩等^[5] 提出了一种基于门控图卷积与动态依存池化的 事件论元抽取模型,该模型通过图神经网络获 取句子的依存句法特征,并将句子的语义特征 和句法特征相融合,从而获取更丰富的特征, 提升了事件论元识别任务的精度。Amir等^[6]提 出了一种基于图变换网络的事件论元识别模型, 该模型通过图变换网络融合句子的语义和句法 特征,从而获取到更有效的句法结构,提升了 事件论元识别任务的精度。

1.2 篇章级事件论元识别方法

篇章级事件论元识别研究在多个句子中识 别事件的相关论元,并为其分配对应的论元角 色。由于篇章中一般包含多个句子,在识别篇 章中的事件论元时,往往存在事件论元分散在 多个句子中的问题。为了解决这个问题, Yuan 等「7」提出了一种关系增强篇章级事件论元识别模 型,该模型通过构建关系增强注意力转换器, 去获取多层次和多数量的事件论元之间的关系, 从而缓解篇章中多事件和论元分散问题,提升了 篇章级事件论元识别的精度。Du 等^[8]提出了一种 基于全局神经生成的篇章级事件论元识别模型, 该模型通过构建篇章内存存储的方式记录上下文 的事件信息,并利用该信息抽取事件论元,提升 了篇章级事件论元识别的精度。Xu等^[9]提出了一 种双流抽象意义表式增强篇章级事件论元识别模 型,该模型通过双流编码的方式从不同角度对篇 章进行编码,从而获取更加充分的上下文信息, 提升了篇章级事件论元识别的精度。Amir等^[10]提 出了一种基于依存句法关系的篇章级事件论元识 别模型,该模型通过构建篇章的依存句法关系获 取论元之间的依赖关系,并提出了一种新的剪枝 方式,去除了篇章依存句法关系中不相关的论 元,从而获取到更加优异的篇章依存句法结构, 提升了篇章级事件论元识别的精度。Liu 等^[11]提 出了一种新的事件论元识别任务链推理范式,通 过链式的推理方式获取事件论元的长距离依赖, 从而充分获取了论元之间的交互信息,提升了篇 章级事件论元识别的精度。

2 模型结构

本文模型主要由输入编码层、特征抽取层



图 1 基于BERT和图注意力网络篇章级事件论元识别模型图

和论元分类层三个部分组成,模型整体结构如 图1所示。①输入编码层,通过BERT预训练模 型将输入的篇章编码为对应的向量表示;②特 征抽取层,获取输入篇章的语义特征和依存句 法特征并融合;③论元分类层,将上一层获取 的融合特征输入Softmax分类器中进行论元角色 分类。

2.1 输入编码层

本文采用BERT^[12]预训练模型对输入篇章进 行编码,从而获取输入篇章的向量表示。BERT 预训练模型是基于Transformer^[13]模型实现的, 该模型采用多个Transformer编码器堆叠的方式 获取句子的向量表示。由于BERT预训练模型存 在最大输入长度,因此需要对输入的篇章进行 预处理。给定一个最大输入长度max_length,对 篇章中句子长度大于max_length的句子进行切割 处理,小于max_length的句子使用Padding进行 补全。

通过上面的处理可以得到处理后的篇章 $D = \{S_1, S_2, \dots, S_N\}, 其中 S_i 表示篇章中第 i 个句子, Ns 表示篇章中的句子数量。将篇章 D 逐句输入$

到 BERT 预训练模型中,可以得到篇章中每个句 子的向量表示,具体实现如式(1)所示: $H_i(h_1, h_2, ..., h_{max_length}) = BERT(w_1, w_2, ..., w_{max_length})$ (1) 其中, H_i 表示第*i*个句子的向量表示, h_i 表示 S_i

中第*i*个词的向量表示。将篇章中所有句子的向量表示进行拼接,可以得到最终的篇章向量表示 $H = \{H_1, H_2, \dots, H_{N_s}\}_{\circ}$

2.2 特征抽取层

2.2.1 语义特征抽取

为了获取充分的语义特征,本文采用BERT 和 Bi-GRU 结合的方式抽取篇章的语义特征。 GRU 网络是 RNN 网络的一种,该网络可以解决 RNN 网络存在的梯度爆炸以及长距离记忆问题。 与LSTM 网络相比,GRU 网络结构相对简单、参 数较少,使得其消耗的硬件资源较少且同样可 以达到较好的效果,其具体实现见式(2)~(5):

$$Z_{t} = \sigma \left(W_{Z} \cdot \left[h_{t-1}, x_{t} \right] \right) \tag{2}$$

$$r_{t} = W_{r} \cdot \left[h_{t-1}, x_{t}\right] \tag{3}$$

$$\widetilde{h_{\iota}} = \tanh\left(W \cdot \left[r_{\iota} * h_{\iota-1}, x_{\iota}\right]\right) \tag{4}$$

$$h_{t} = (1 - Z_{t})^{*} h_{t-1} + Z_{t}^{*} \widetilde{h_{t}}$$
(5)

其中: Z_i表示更新门计算, r_i表示重置门计算, h_i表示应用门控装置后的输入结果。

经过上一层对篇章的处理可以得到篇章的 向量表示*H*,该向量表示包含BERT预训练模型 所带来的语义信息。为了进一步获取篇章的上 下文语义特征,将*H*输入Bi-GRU网络中,从 正、反两个方向捕获篇章的上下文语义特征, 具体实现如式(6)~(8)所示:

$$\vec{H}\left(\vec{h}_{1},\vec{h}_{2},\cdots,\vec{h}_{N_{i}}\right) = \vec{GRU}\left(H_{1},H_{2},\cdots,H_{N_{i}}\right)$$
(6)

$$\widetilde{H}\left(\widetilde{h}_{1},\widetilde{h}_{2},\cdots,\widetilde{h}_{N_{i}}\right)=\widetilde{GRU}\left(H_{1},H_{2},\cdots,H_{N_{i}}\right)$$
(7)

$$X = concat \left[\vec{H}, \, \tilde{H} \right] \tag{8}$$

其中: \vec{H} 表示正向获取的隐层表示, \vec{H} 表示反向 获取的隐层表示, concat 表示拼接运算,X表示 拼接结果。

2.2.2 依存句法特征抽取

本文采用 stanfordCoreNLP 工具对篇章 D 中 的每个句子进行依存句法分析,得到每个句子 的依存句法树,然后将相邻句子的依存句法树 从根部进行连接,从而得到篇章的依存句法树。 由于篇章中包含多个句子,融合多个句子的依 存句法树容易引入与事件不相关的论元,继而 产生噪声问题。本文参考了 Li等^[14]的方法,采 用基于依赖路径的剪枝策略去除与事件不相关 的论元,处理后的篇章依存句法树为*T*。

为了生成*T*的邻接矩阵方便后续捕获特征,本 文将篇章中的每个词都设定为一个节点。首先,根 据 stanfordCoreNLP工具得到的依存句法分析,判 断*T*中每个词之间是否存在依存关系,若存在则连 接两个节点,不存在则不连接。处理后可以得到*T* 的邻接矩阵*G* = (*V*, *E*), *V* = { e_1, e_2, \dots, e_{Ne} },其中 e_i 表示篇章中第*i*个节点, *Ne*表示篇章中的节点 个数, *E* = {(e_1, e_2), ..., (e_{Ne-1}, e_{Ne} },其中(e_i, e_j) 表示*i*和*j*节点存在依存关系。然后,将*G*输入 图卷积网络^[15]中获取篇章的依存句法特征,其 具体实现见式(9)所示:

$$h_i^{l+1} = \sigma \left(\frac{\boldsymbol{G}_i^t (h_i^l \boldsymbol{W} + \boldsymbol{b})}{\sum \boldsymbol{G}_i} \right)$$
(9)

其中: l表示图卷积层数, G,是图邻接矩阵, W表

示权重系数, b表示偏置向量。

为了进一步去除篇章依存句法结构中事件 不相关的论元,本文引入图注意力网络^[16]对不 同的邻居节点分配不同的权重,从而获取更好 的特征表示,其具体实现如式(10)所示:

$$h'_{i}(K) = \left\| {\scriptstyle K \atop k = 1}^{K} \sigma \left(\sum_{j \in N_{i}} \alpha^{k}_{ij} W^{k} h_{j} \right)$$
(10)

其中: α_{ij} 表示节点i和j的注意力计算, $h'_i(K)$ 表示i节点融合了相邻节点信息的特征输出,K表示注意力头的个数。最后,将 h^{i+1}_i 和 $h'_i(K)$ 进行 拼接融合得到处理后的篇章依存句法特征,我 们记为 h_o

2.2.3 特征融合

由于本文只关注篇章级事件论元识别任务, 将篇章中所包含的事件类型以及触发词的位置 信息作为先验知识与篇章的语义特征相融合, 从而更精确地捕获事件论元。首先,本文对篇 章中所包含的事件类型信息和词的位置信息都 进行随机初始化,得到篇章的触发词位置特征*p* 和事件类型特征*t*。然后,将得到的触发词位置 特征*p*和事件类型特征*t*与篇章的语义特征*X*相 融合,可以得到融合后的特征,我们记为*C*。最 后,将*C*与篇章的依存句法特征*h*进行拼接融 合,其具体实现如式(11)所示:

$$\hat{H} = \begin{bmatrix} C & : & h \end{bmatrix} \tag{11}$$

其中: Ĥ表示融合后的特征隐层向量。

2.3 论元分类层

事件一般包含多个事件论元角色,故可将 事件论元识别任务建模为多标签分类任务。首 先,通过上一层可以得到特征融合后的特征隐 层向量 *Ĥ*,对该向量进行最大池化操作筛选与 事件论元和触发词有直接依存关系的论元,我 们记池化后的向量为*Ĥ*。然后,将该向量映射 到事件论元角色标签空间,其具体实现如式 (12)所示:

$$\gamma = W\tilde{H} + b \tag{12}$$

其中: W为权重矩阵, b为偏置向量。最后,将 该向量输入Softmax分类器中进行事件论元分 类,计算每个候选论元为对应论元角色的概率, 其具体实现如式(13)所示:

$$p(r|a) = \frac{\exp(y_r)}{\sum_{i=1}^{K} \exp(y_i)}$$
(13)

其中:r表示论元角色,a表示候选论元,K表 示事件论元角色的个数。

3 实验结果与分析

3.1 实验数据集

本文使用的数据集为RAMS^[17],该数据集 广泛用于篇章级事件论元识别任务,数据集中 的篇章仅包含单一事件并通过跨句论元标注的 方式标注完整的事件论元,其包含9124个新闻 事件、139种事件类型和65种事件论元角色。 本文采用RAMS数据集官方提供的训练集、测 试集和验证集作为本文实验的训练集、测试集 和验证集。

3.2 实验参数和评价指标

本文采用 BERT 预训练模型初始化词向量, BERT 预训练模型的隐层维度为 768。GRU 网络 的隐层维度为 150, 触发词位置特征和事件类型 特征的维度为 50。句子最大输入长度 max_length 为 128, 最大句子数为 64。本文实验的详细参数 设置见表 2。

参数	数值
GCN Layer	3
GAT Layer	3
Batch_size	64
Learning_rate	$1 \times e^{-4}$
Dropout	0.5
Epoch	100

表 2 实验参数设置

本文模型的评价指标采用准确率(Precision, P)、召回率(Recall, R)和调和平均数F1值,具体实现公式如式(14)~(16)所示:

$$P = \frac{TP}{TP + FP} \times 100\% \tag{14}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{15}$$

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\%$$
(16)

例, FN表示错误的反例。

3.3 实验结果与分析

本文实验采用相同的实验环境在RAMS数据集下进行,对比的基准模型如下:

(1) HMEAE^[2]模型:提出了一种层次模块 化事件抽取模型,将论元角色转换为一种概念, 从概念的角度进行论元角色的交互。

(2) SemSynGTN^[6]模型:通过图变换网络融 合句子的语义和句法特征,并引入了一种归纳 偏差策略增加了模型的泛化能力。

(3) OTEAE^[10]模型:提出了一种新的剪枝 策略,该策略关注不在依赖路径上的论元,并 通过计算其与依存路径上论元的句法语义相关 性,判断是否遗漏关键论元,从而补全遗漏的 论元。

实验结果见表3,从表3可以看出,本文方 法对比基准模型在准确率和F1值上均有所提 升,证明了本文方法的有效性。与最佳基线模 型OTEAE模型相比,在准确率和F1值上分别提 升了3.3和1个百分点。提升原因在于语义特征 获取阶段,采用BERT和Bi-GRU结合的方式获 取了充分的语义特征,增强了模型对篇章语义 信息的感知;在依存句法特征获取阶段,采用 GCN和GAT并行的方式进一步筛选了有效的论 元,解决了篇章依存句法结构中的噪声问题。

表 3 本文模型与对比模型事件论元识别结果(%)

模型	Precision	Recall	F1值	
HMEAE	60.5	53.2	56.6	
SemSynGTN	67.4	54.3	60.1	
OTEAE	75.2	76.1	75.6	
本文模型	78.5	74.8	76.6	

4 结语

本文针对现有方法对篇章依存句法结构中 存在的噪声、处理不佳的问题,提出了基于 BERT和图注意力网络的篇章级事件论元识别模 型。该模型通过获取充分的篇章语义特征作为 辅助,从而更好地诱导出优异的篇章依存句法 结构,并引入图注意网络进一步筛选与事件有 关论元,进而解决噪声问题。实验结果表明, 本文方法具有有效性。然而,本文方法在处理 包含多事件的篇章时无法取得较好的事件论元 识别结果,因此下一步的研究方向为多事件情 况下的篇章级事件论元识别。

参考文献:

- SARAWAGI S. Information extraction [J]. Foundations & Trends in Databases, 2008, 1(3):261-377.
- [2] WANG X, WANG Z, HAN X, et al. HMEAE: hierarchical modular event argument extraction [C] //Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 2019:5777-5783.
- [3] NAN D, CHUNMING H, KAI S, et al. Explicit role interaction network for event argument extraction [C]//Findings of the Association for Computational Linguistics: EMNLP, Abu Dhabi, United Arab Emirates, 2022: 3475-3485.
- [4] YOU H, DAVID S, SAMIA T, et al. Event graph: event extraction as semantic graph parsing [C] //Proceedings of CASE: The 5th Workshop on Challenges and Applications of Automated Extraction of Sociopolitical Events from Text, Abu Dhabi, United Arab Emirates, 2022:7-15.
- [5] 王士浩,王中卿,李寿山,等.基于门控图卷积与动态依存池化的事件论元抽取[J]. 计算机科学, 2021,48(A2):52-56.
- [6] AMIR P, TUAN N, THIEN H. Graph transformer networks with syntactic and semantic structures for event argument extraction [C] //Proceedings of the Findings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020: 3651-3661.
- YUAN L, ZHUO J, DI Y, et al. RAAT: relation-augmented attention transformer for relation modeling in document-level event extraction [C] //Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Seattle, United States, 2022: 4985-4997.
- [8] DU X, LI S, JI H. Dynamic global memory for document-level argument extraction [C] //Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL), Dublin, Ireland, 2022:5264-5275.

- [9] XU R, WANG P, LIU T, et al. A two-stream AMR-enhanced model for document-level event argument extraction [C] //Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Seattle, United States, 2022: 5025-5036.
- [10] AMIR P, MINH V, FRANCK D, et al. Document-level event argument extraction via optimal transport [C] //Findings of the Association for Computational Linguistics (ACL), Dublin, Ireland, 2022:1648-1658.
- [11] LIU J, LIANG C, XU J, et al. Document-level event argument extraction with a chain reasoning paradigm [C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL), Toronto, Canada, 2023:9570-9583.
- [12] JACOB D, MING C KENTON L, et al. BERT: pre-training of deep bidirectional transformers for language under-standing [C] // Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, Minnesota, 2019: 4171-4186.
- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems, California, United States, 2017: 5998-6008.
- [14] LI Q, JI H, HUANG L. Joint event extraction via structured prediction with global features [C] //Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), Sofia, Bulgaria, 2013:73-78.
- [15] NGUYEN T, GRISHMAN R. Graph convolutional networks with argument-aware pooling for event detection [C] //Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI), New Orleans, United States, 2018:5900-5907.
- [16] PETAR V, GUILLEM C, ARANTXA C, et al. Graph attention networks[C]//Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, Canada, 2018:1-12.
- [17] EBNER X, XIA P, CULKIN R, et al. Multisentence argument linking [C] //Proceedings of the Annual Meeting of the Association for Computational Linguistics(ACL), Online, 2020:8057-8077.

(下转第64页)

文章编号:1007-1423(2024)06-0020-07

DOI: 10.3969/j.issn.1007-1423.2024.06.004

适用于边缘设备的轻量级人体检测算法

周 宁,陶青川*,彭勃兴

(四川大学电子信息学院,成都 610065)

摘要:针对现存的人体检测网络都比较复杂,部署到边缘设备上时表现不佳的问题,基于YOLOv7提出一种轻量级人体检测算法。该算法首先使用改进后的ShuffleNev2基本模块替换原网络ELAN模块;接着在主干网络末端添加SE 注意力和SPPF 池化;然后在Neck部分使用改进后的GSConv 替换标准卷积,引入基于GSConv 的 VoVGSCSP 替换 ELAN-W模块。通过在GPU和Sophon SE5上的验证结果表明,该轻量级人体检测算法与YOLOv7相比损失2.6%的精度,但计算量大幅度降低,在Sophon SE5上推理速度达到了54 FPS,相比较YOLOv7提升了39 FPS。

关键词: YOLOv7; 目标检测; ShuffleNet; 边缘计算; Sophon SE5

0 引言

在安防和行为识别等领域,人体检测作为 基础算法有着重要的应用价值。在行为识别场 景下,需要先使用人体检测算法检测到人体, 再进行后续的处理。人体检测作为行为识别的 第一步,其性能好坏对整个系统性能有着重要 影响^[1]。

目前目标检测算法主要分为两大类:一类 是一阶段目标检测算法,检测速度快、精度略 低,代表算法有YOLO系列^[2]等;一类是二阶段 目标检测算法,先进行定位,再进行分类,精 度稍高,但速度慢,代表算法有R-CNN系列^[3]。 在实际工程中,由于一阶段目标检测算法结构 简单、运算速度快,且边缘设备对相关算子支 持较好,得到了广泛的应用。

本文在 YOLOv7 算法^[4]的基础上,参考目前 优秀轻量级的网络结构,对其网络结构进行改 进,提出一种轻量化的人体检测算法,使其保 持较高检测精度的同时,降低网络计算量,提 高推理速度,更适于部署在边缘设备上。

1 相关工作

YOLOv7系列算法在检测速度与精度上取得 了较好的均衡。与之前版本相比,其在网络结 构、卷积计算、添加辅助检测头以及动态标签 分配等方面进行改进,使得其速度和精度达到 较高水平的平衡^[5]。

目前行为识别系统所应用的人体检测方法 也多为YOLO系列算法,通过对检测结果进行 过滤得到人体检测结果。然而这些目标检测算 法为了在多个目标检测上达到最佳,其参数量 都较大,对计算资源要求也较高,一般需要部 署在GPU上运行,在嵌入式、边缘计算等算力 有限的设备上无法达到较好的效果^[6]。针对该 问题,本文基于算法的使用场景和所用边缘设 备的算力,对YOLOv7网络结构进行调整和改 进,相关工作主要有以下三个方面:

(1)对 ShuffleNetv2 的模块^[7]进行改进,使 用改进后的模块对 Backbone 进行重构,以降低 网络推理计算量,加快推理速度。

(2)在Backbone末端添加注意力机制,以提升检测精度;使用SPPF 替换SPP模块^[8],在

收稿日期: 2023-10-17 修稿日期: 2023-11-15

作者简介:周宁(1998—),男,河南周口人,在读硕士研究生,研究方向为智能信息系统;*通信作者:陶青川(1972—), 男,四川南充人,博士,副教授,研究方向为模式识别与智能系统,E-mail:taoqingchuan@scu.edu.cn;彭勃兴(2000—),男, 陕西延安人,在读硕士研究生,研究方向为模式识别与智能系统

实现特征融合的同时加快推理速度。

(3)使用GSConv卷积^[9]对Neck中的标准卷 积进行替换;同时使用VoVGSCSP模块替换 E-ELAN模块,使各层特征得到充分信息交流的 同时,减少特征冗余,降低网络计算量,加快 推理速度。

2 算法细节

2.1 ShuffleNetv2网络模块改进

2.1.1 ShuffleNetv2基本模块

ShuffleNetv2是在ShuffleNetv1^[10]的基础上提出的升级版本。与ShuffleNetv1相比,在同等复杂度下正确率更高,且在参数量和推理速度上做出优化,适用于部署到边缘设备上。

本文通过实验验证,提出设计网络时加快 推理速度的四条指导性原则:①同等通道大小 最小化内存访问量;②过量使用组卷积会增加 内存需求;③网络碎片化会降低并行度;④不 能忽略元素级操作^[11]。基于这四条原则,本文 在 ShuffleNetv1 基础上提出 ShuffleNetv2,其基本 模块如图1所示。



图 1 ShuffleNetv2基本模块

与ShuffleNetv1相比,ShuffleNetv2进行以下 四点改进:①新增Channel Split模块,保持卷积 输入输出通道一致;②使用1×1的标准卷积代 替组卷积;③不再使用池化操作,使用stride为2 的深度卷积完成下采样;④使用Concat代替Add 操作,减少元素级运算,分别从内存占用和网 络碎片化方面进行优化。

2.1.2 不同卷积参数量对比

ShuffleNet 网络中使用了组卷积(group conv, GConv)、深度可分离卷积(depth-wise separable conv, DSConv),用来降低模型的参数量,提高 推理速度,使网络更轻量化。计算每种卷积所 需的参数量与标准卷积(standard convenience, SConv)进行对比,假设输入特征图大小为 $W \times H$,输入通道数为 C_i ,输出通道数为 C_o ,卷 积核大小为 $k \times k_o$ 标准卷积的参数量为

$$P_s = C_i \times k \times k \times C_a \tag{1}$$

组卷积通过将输入特征按通道数分组来实 现标准卷积。组卷积的参数量为

$$P_{g} = \frac{C_{i}}{g} \times k \times k \times C_{o}$$
(2)

深度可分离卷积涉及到两个操作:深度卷积(depth wise, DW)和逐点卷积(point wise, PW),深度卷积在单个通道上完成特征提取,然后通过逐点卷积完成不同通道之间的交流,实现通道数的变换。深度可分离卷积的参数量为

 $P_{dsc} = C_i \times k \times k + 1 \times 1 \times C_i \times C_o \qquad (3)$ $\frac{P_{dsc}}{P_s} = \frac{C_i \times k \times k + 1 \times 1 \times C_i \times C_o}{C_i \times k \times k \times C_o} = \frac{1}{C_o} + \frac{1}{k^2} \approx \frac{1}{k^2} (4)$

对比三种卷积的参数量可以发现,标准卷 积参数量最大,组卷积参数量是其1/g,深度可 分离卷积约为其1/k²,通过使用组卷积和深度 可分离卷积,可以有效降低网络的参数量,加 快推理速度。

2.1.3 改进后 Shuffle_Block

组卷积和深度卷积降低了网络的参数量, 加快了推理速度,但是其通道间缺少信息交流, shuffle操作和逐点卷积都是为了解决通道间信 息交流的问题而存在的。在ShuffleNetv2中,同 时使用以上两种方式完成不同通道间的信息交 流,在获得充分信息交流的同时,也增加了推 理时的计算量,使得网络碎片化,因而本文选 择丢掉 shuffle模块。

有研究表明^[12],在特征信息有限时,残差 模块中的短链接会影响梯度回传。为了获得更 好的梯度回传效果,同时保留更多的空间特征, 本文选择在ShuffleNetv2的基本模块的右边分支 添加深度卷积。改进后结构如图2所示。



图 2 改进 Shuffle_Block 模块

2.2 SE模块

注意力机制在图像、自然语言处理等领域 都取得了重要突破,在一定程度上能提升模型 的性能。SENet^[13]较早地将注意力机制引入到 CNN中,其结构简单,参数量和计算量都较少, 适用于部署到边缘设备上。SENet中的SE模块 通过一个权重矩阵,赋予不同位置的通道不同 的权重系数,关注更有用的特征信息,因此获 得了更好的分类效果。SE模块原理如图3所示。



图 3 SE 模块

SE模块主要有三个操作:压缩(squeeze)、挤压(excitation)和scale操作。其过程首先是特征图X经过标准卷积后得到特征图U,其通道数为C,特征图大小为 $W \times H$;接着对U进行压缩操作,压缩通过池化操作实现,对全部通道的特征图进行池化得到 $1 \times 1 \times C$ 全局信息向量;然后

将该向量通过激励模块得到带有权重的1×1×C 向量;最后将该向量与U相乘,即可得到带有 权重的特征图F,其通道数和特征图的尺寸与U 一致。每个特征图的压缩公式为

$$Z = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{e}(i,j)$$
(5)

其中: $X_{c}(i,j)$ 表示特征图第i行, 第j列的像素 值, W和H为特征图的宽和高。

激励模块主要由全连接层和激活函数组成, 是一个由4个全连接层组成的多层感知机,其结 构如图4所示。



图 4 激励模块

2.3 SPPF模块

SPP模块主要是用来解决输入图像大小不一 致时造成的特征图失真的问题。SPP模块能够对 不同尺度的特征图进行压缩和融合,能够得到 更真实的特征图信息。SPPF在SPP的基础上进 行改进,降低了模型的计算量,加快了推理速 度,因而本文选择使用SPPF来完成不同尺度特 征图的融合。SPPF模块如图5所示。





2.4 GSConv卷积

标准卷积在进行特征提取时,会造成特征 图的冗余,累积的参数也会增加推理时的计算 量。Han等^[14]提出的Ghost Conv卷积,先对特征 图做标准卷积,得到少量的特征图,然后对少 量的特征图进行线性变换,最后将两个操作得 到的特征图拼接到一起,扩充特征图的通道。

Ghost Conv因其出色的性能而广泛应用在轻量化的模型中,但是其在进行拼接的过程中损失了通道信息。为了获得通道相关的信息,Li等^[9]提出了GSConv卷积,如图6所示。其输入输出

通道数分别为*C_i*, *C_o*,首先经过一个卷积通道 变为*C_o*/2,然后经过深度卷积得到同样通道数 为*C_o*/2的特征图,最后将二者进行拼接和通道 混洗,增强通道间信息交流,提升特征的表达 能力。考虑到 shuffle 会使网络碎片化,且其中 已经使用标准卷积,因而本文选择丢掉 shuffle 操作,改进后的 GSConv 如图7所示。



图 7 改进后的GSConv卷积

GSConv可以看成是由标准卷积和深度可分 离卷积拼接而成,在较少计算量的同时能够保 证特征图更接近标准卷积,得到更具表达力的 特征图,保证模型的准确性。本文引入基于 GSConv的VoVGSCSP模块替换网络中的ELAN 结构,降低参数量和计算量,加速推理。 VoVGSCSP模块结构如图8所示。



图 8 VoVGSCSP模块结构

2.5 改进后人体检测网络结构

在以上改进各个结构的基础上,针对目 标检测网络的各部分进行重构,得到如图9所 示的人体检测网络,为方便下文表述,称其



图 9 HumanNet-tiny网络结构图

3 实验对比

3.1 实验平台

本文选择带有 GPU 的 PC 和 Sophon SE5 作为 实验硬件平台。其中 PC 硬件配置见表1。

表1 PC硬件配置

模块	参数
CPU	12 核 Intel(R) 8255C CPU @ 2.50GHz
GPU	RTX 3080(10 GB) * 1
内存	40 GB
软件版本	ubuntu20.04, PyTorch 1.10.0, Cuda 11.3

Sophon SE5 是算能科技推出的新一代边缘计 算盒,其性能优秀、功耗低。其硬件配置见表2。

表 2 Sophon SE5 配置列表

模块	参数	-
CPU	BM1684 8核 A53@2.3GHz	
推理引擎	TPU	
内存	12 GB	

由于公开数据集中包含较多小目标,考虑 到本文算法后续使用场景,实际应用时效果不 佳,本文选择自制数据集,其中包括3246张 校内实景采集、1250张网络图片以及从CUHK Occlusion Dataset 中选取的 2120 张图片, 共6616 张图片,使用labelimg重新标注后,以8:2比 例划分训练集和测试集。

3.3 实验结果

3.3.1 GPU验证结果

为了验证本文所提出模型的检测效果,选 择 YOLOv7 系列和 YOLOv5 系列的部分模型作为 对比,使用相同的硬件配置和软件环境,对上 述几个网络模型进行重新训练。对比多个模型 的计算量(GFLOPs)、模型大小、推理速度和准 确率,实验结果见表3。

模型名称	计算量/ GFLOPs	模型大小/ MB	推理速度/ FPS	mAP/%
YOLOv5s	15.8	14.4	77	96.0
YOLOv5s-Ghost	8.2	7.8	78	93.9
YOLOv7	105.1	74.8	62	96.2
YOLOv7-tiny	13.2	12.3	102	94.7
HumanNet-tiny	7.1	5.2	104	93.6

表 3 不同模型 GPU 验证结果

图 10 所示为不同模型在 GPU 上的实际检测 效果图。

从图 10 的检测结果来看,在 GPU 上各个模 型都能较好地检测出人体。对比各个模型的实验 结果可以发现,本文提出的模型HumanNet-tiny 在保持较高检测精度的同时,模型推理计算量 大幅度降低, 仅7.1 GFLOPs; 模型大小减小到 5.2 MB, 推理速度相比 YOLOv7 提升了 67.9%。

3.3.2 Sophon SE5 验证结果

本文所提出的HumanNet-tiny模型旨在部署 到边缘设备上,利用其功耗低、性价比高的优 势,因此将模型部署到 Sophon SE5 设备上进行 测试,实验结果见表4。

(a) 原图





(c) Yolov5s-Ghost

(d) YOLOv7



(e) YOLOv7-tiny

(f) HumanNet-tiny

图 10 不同模型检测效果图

表4 Sophon SE5上HumanNet-tiny与YOLOv7 验证结果

模型名称	mAP/%	推理速度/FPS
YOLOv7	95.5	15
HumanNet-tiny	93.2	54

图 11 为将 HumanNet-tiny 部署到 Sophon SE5 上时的检测效果图。



图 11 Sophon SE5上HumanNet-tiny检测效果图

从检测效果来看, 部署到Sophon SE5上后, HumanNet-tiny 也能较好地检测到人体。结合 表3中模型转换前在GPU上实验结果,从表4中 实验数据可以看出,模型转换造成了HumanNettiny 精度损失了0.4个百分点,但转换后的模型帧 率达到54 FPS,对比YOLOv7的帧率提升39 FPS, 且Sophon SE5的功耗仅9.6W,远小于GPU百瓦 级别的功耗。

上述实验对比与结果分析表明,本文提出

的HumanNet-tiny算法性能较好,且计算量和参数量都较小;在推理速度上,相比较于在GPU上的推理速度提升,其在Sophon SE5上提升得更为明显,对比YOLOv7提升39FPS,能够充分利用Sophon SE5低功耗、加速推理等优势,适用于部署到边缘设备上。

4 结语

针对边缘设备算力有限,直接移植深度学 习模型推理速度不佳的问题,本文在YOLOv7和 目前一些优秀的轻量级算法的基础上,提出 HumanNet-tiny算法。实验表明,该算法在GPU 和Sophon SE5上都取得了不错的检测效果和推 理性能,其在Sophon SE5上的帧率达到54 FPS, 且功耗仅9.6 W,表明该算法具有较高的应用价 值。为了进一步提升检测精度以及推理速度, 在后续的工作中,将基于Sophon SE5的特性进 行分析研究,探索模型中不同操作对精度和推 理速度的影响,进一步提升算法的检测精度和

参考文献:

- [1] 张阳婷,黄德启,王东伟,等.基于深度学习的目标 检测算法研究与应用综述[J].计算机工程与应用, 2023,59(18):1-13.
- [2] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, United States, 2016:779-788.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, United States, 2014: 580-587.
- [4] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new stateof-the-art for real-time object detectors [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC,

Canada, 2023:7464-7475.

- [5] 戚玲珑,高建瓴.基于改进YOLOv7的小目标检测 [J].计算机工程,2023,49(1):41-48.
- [6] 刘浩翰,樊一鸣,贺怀清,等.改进YOLOv7-tiny的
 目标检测轻量化模型[J].计算机工程与应用,2023,
 59(14):166-175.
- [7] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2:practical guidelines for efficient CNN architecture design[C]//Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018:116-131.
- [8] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37 (9) : 1904-1916.
- [9] LI H, LI J, WEI H, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles [EB/OL]. arXiv: 2206.02424, 2022.
- [10] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: an extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, United States, 2018;6848-6856.
- [11] 徐源,张玉杰. 基于改进 ShuffleNetV2 的敏感内容 识别与应用[J]. 传感器与微系统, 2023, 42(3): 164-168.
- [12] SANKARARAMAN K A, DE S, XU Z, et al. The impact of neural network overparameterization on gradient confusion and stochastic gradient descent [C] //Proceedings of the International Conference on Machine Learning, Vienna, Austria, 2020: 8469-8479.
- [13] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, United States, 2018:7132-7141.
- [14] HAN K, WANG Y, TIAN Q, et al. Ghostnet: more features from cheap operations [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Electr Network, 2020: 1580-1589.

(下转第68页)

文章编号:1007-1423(2024)06-0026-07

DOI: 10.3969/j.issn.1007-1423.2024.06.005

基于DBNet改进的检务场景文本检测算法研究

于 晓*,林世基

(天津理工大学电气工程与自动化学院, 天津 300384)

摘要:针对检务场景文本检测中,现有的检测算法仍存在误检率和漏检率高等问题。通过改进现有的特征提取网络,引入高效通道注意力和空间注意力模块 CBAM,同时改进可微二值化函数,并将改进后的网络应用到检务场景文本检测当中。改进后的算法在 ICDAR 2015 数据集上的准确率、召回率及F值相较于改进前分别提升了2.2、5.4及4.2个百分点,达到了89.2%和63.6%及74.3%。实验数据表明,改进DBNet文本检测算法在收敛速度和检测精度上都有明显的提升。

关键词: 文本检测; 检务场景; 可微二值化; 深度学习; CBAM

0 引言

文字是人类感知世界和信息交流的重要载 体之一,能够记录和传达各种信息和思想。在 当今信息时代,文字在各种领域中发挥着越来 越重要的作用。图片中的文字包含着丰富的语 义信息,这些信息对于计算机场景理解、人工 智能发展、工业自动化生产等行业具有重要意 义^[1]。随着国家信息化的建设,在检察系统的 工作过程中,需要从图片中提取有效的电子文 字信息,完成智慧检务中的辅助办案工作。

智慧检务场景文本检测是自然场景文本检 测的一个重要应用领域,它涉及到对大量文本 数据的自动化分析和处理。现有的自然场景文 本检测方法主要分为以下两类:传统方法的自 然场景文本检测和基于深度学习的自然场景文 本检测。

对于传统的自然场景文本检测方法,一般 是通过分析文本的特征,人工设计相应的特征 提取方法,然后使用传统的分类器进行分类。 Epshtein等^[2]提出笔画宽度变换(SWT)的场景文 本检测方法,其思想是通过 canny 边缘检测算子 提取文字的边缘和梯度特征,然后通过连通域 的颜色方差等信息筛选出文本区域。但是依赖 于人工设计特征提取算法,同时只能检测水平 方向的文本,不能满足复杂的场景文字检测。 Matas 等^[3]提出一种类似于分水岭法的MSER场 景文字检测算法。MSER方法具有较好的稳定性 和仿射不变性^[4],但是在模糊、光照条件不佳 的情况下,算法的准确率并不高。

随着计算机技术的发展,越来越多的学者 将深度学习算法应用到场景文本检测当中。基 于深度学习的场景文字检测方法主要分为基于 候选框回归方法和基于图像分割方法。

受到Faster R-CNN的启发,Zhong等^[5]提出 DeepText方法,把该目标检测算法应用到自然 场景文本检测工作钟。Liao等^[6]针对自然场景文 本的特性提出了TextBoxes算法,该算法采用了 长条形的卷积核,能对水平文本取得不错的检 测性能。Tian等^[7]提出了CTPN文本检测算法, 该算法提出了固定定宽的anchor机制,使用 CNN+RNN能有效地检测出复杂场景的横向分布 的文字。Shi等^[8]提出了SegLink文本检测方法,

基金项目: 国家自然科学基金(61502340);天津市自然科学基金(18JCQNJC01000)

作者简介:*通信作者:于晓(1985—),男,山东临沂人,博士,副教授,研究方向为人工智能、图像处理,E-mail: yx_tjut@163.com;林世基(1998—),男,广西桂平人,硕士,研究方向为图像处理

收稿日期: 2023-10-20 修稿日期: 2023-11-28

该方法融入了CTPN算法的细粒度候选框,使得 模型不再受限于默认比例的人工设定文本框, 但该模型对于弯曲文本和间距较大文本检测结 果较差。以上的文字检测方法,主要是针对水 平方向的文本检测,无法解决自然场景当中多 个方向的文本以及不规则文本的检测问题。 Zhou等^[9]借鉴了DenseBox的架构和U-Net的特性 提出了基于两阶段的文本检测方法EAST,能够 检测除了标准四边形外的其他形状。

由于检务场景文字具有复杂的环境条件, 基于分割的方法往往具有更好的检测效果。 Deng 等^[10]提出基于实例图片分割的文本检测算 法 PixelLink, 该算法采用纯分割的思路, 放弃了 边界框回归的方法,直接进行实例的分割,再生 成边界框。与基于回归的方法相比, PixelLink 可 以在多个基准上获得更好或可比的性能,但该方 法对相邻文本实例分割效果较差。Wang等^[11]提 出 PSENet, 该文本检测网络使用 LSAE 计算像素 之间的特征距离来聚类,然后使用连续尺度扩张 的方式后处理。Liu等^[12]提出了一种基于实例分 割的文本检测算法 PANNet。该网络采用从核心 区域往外逐步扩展的方法解决距离近的文本区域 粘连的问题, 使得网络检测速度快而且精度高。 Long 等^[13]提出了TextSnake 模型,该模型采用圆 环作为基础,更灵活的场景文本表征,能够以水 平、多方向和弯曲的形式有效地表征文本实例。 Liao等^[14]提出一种可微分二值化(DBNet)的方法, 该方法将二值化操作嵌入网络,学习文本区域的 概率图和预测对应的阈值图,大大提高了后处理 的效率和模型前向推理速度。

1 问题分析

现有的深度学习文本检测方法主要由特征 提取、预测网络以及NMS组成^[15]。相对于传统 方法,基于深度学习的文本检测在速度和精度 上都有显著的提升。然而,在实际检务场景中, 检务文本图像存在复杂的干扰因素,主要分为 以下几种情况:

(1)检务图像质量不一。由于司法系统工作的特殊性,检务图像的获取方式不一,从而导致检务图像背景多样,并且大多数图像由手持拍摄获取,极易受到环境光、拍摄设备、遮挡物等因素的影响,导致文本提取速度和提取效果不佳,如图1(a)所示。

(2)检务文本多角度、弯曲、空间扭曲等问题。在检务场景当中,为了追求快速取证可能出现多角度拍摄的情况,同时文字并非始终成规则的矩形区域,而是时常会出现弯曲,或者成不规则的文字区域,进而导致文本检测存在准确率不高及效率不佳的问题,如图1(b)所示。

(3)检务文本排版复杂及语种多样性。检务 文本的排列和版面复杂、排列密集程度不一, 同时文本包含各种语种文本(中英文、阿拉伯数 字等),而且不同语种属性复杂不一,其分析和 处理会具有很大的歧义性,严重影响了文本区 域的定位精度,如图1(c)所示。



(a) 光照不均 (b)空间扭曲 (c)背景、排版 复杂及多语种

图 1 检务文本图像干扰因素

综上,对检务场景中的文本进行准确、高效 的检测依然是当下极具挑战性的研究。如何有效提 取显著性的文本区域特征是提升检测精度的关键, 而骨干网络的特征提取能力是至关重要的^[15]。

本文通过对可微二值化网络结构的改进, 提升网络对文本特征提取的能力,引进高效注 意力机制,提升对文本检测的准确率和效率, 使文字检测网络具有更加鲁棒的检测性能,能 够应对检务场景中各种难以检测的问题。

2 改进DBNet的检务场景文本检测算法

传统的分割算法当中,文本图像经过特征 提取网络后,得到文本分割的概率图,然后根据 经验设置一个概率阈值,将概率图转化成二值图 像,最后采用像素聚类等技术将像素级结果转换 成检测结果。传统的分割方法如图2所示。



图 2 传统分割方法

本文的检务场景文本检测模型基于 DBNet 构建,特征提取网络采用 ResNet-18,并通过特 征金字塔进行特征融合。首先,提出对于 DB-Net特征提取网络的改进,将文本特征提取网络 ResNet 改进为 ResNet-V2S,提升卷积的特征信 息提取能力。其次,为了抑制无效特征,使网 络能够自动学习不同通道的重要程度和关联性, 引入了空间注意力和通道注意力模块 CBAM。 最后,将 tanh 函数引入到可微二值化函数中, 以提高文本检测网络的训练效率。

2.1 ResNet特征提取网络改进

DBNet 文本检测算法采用网络学习对每个 像素点进行二值化。其中,每个二值化都是自 适应的,因此得到的二值图具有很强的鲁棒 性^[16]。简化了复杂的后处理步骤的同时还提升 了文本检测效果。可微二值化文本检测网络结 构如图3所示。



图 3 可微二值化文本检测网络

首先,可微二值化文本检测网络的主干网 采用ResNet。检务图像通过FPN特征金字塔提 取不同尺度的文本特征,得到四张大小为原图 的1/4、1/8、1/16、1/32大小的特征图,再将四 个特征图分别进行上采样至原图的1/4,最后将 四个1/4的特征图进行 Concat,得到特征图 F。 特征图 F用来预测概率图 P和阈值图 T,最后通 过可微二值化计算得到近似二值化图 B^[17]。

在深度学习领域当中,增加网络的层数可 以增强模型的检测性能,因此,增大网络的深 度理论上可以取得更好的检测效果。但是 CNN 网络在层数加深的同时,网络面临过拟合的问 题。本文使用 ResNet 作为检务文本特征提取网 络,该网络增加层数的方式是在浅层网络上叠 加恒等映射层,从而构建出残差学习单元^[18]。 为了提高网络对检务文本特征的提取能力,把 ReLU激活函数置于卷积层之前,实现输入与输 出的直连,最大程度上保留传递的信息,同时 使用批标准化(batch normalization, BN),增强了 模型的正则化能力。改进后的文本特征提取网 络如图4所示。



图 4 改进后的文本特征提取网络结构

2.2 空间注意力和通道注意力机制

注意力机制自提出以来,已经成功地应用 于图像处理、自然语言处理和数据预测等领域。 注意力机制在参与文本特征提取时,将局部感 受野中的空间信息和通道信息进行融合,减弱 对无关特征参数的权重,以加强对重要特征的 提取,因此具有更高的可扩展性和鲁棒性。为 了抑制文本检测网络中的无效特征,本文引入 了高效的注意力机制CBAM,使得网络能够自 动学习不同通道的重要程度和关联性。CBAM 注意力机制网络结构如图5所示。



图 5 CBAM 注意力机制网络结构

CBAM 是一种混合注意力机制模块,结合 了空间(spatial)和通道(channel)注意力机制模 块。相比于仅关注一方面的注意力机制模块, CBAM 能两方兼顾,获得更好的效果。本文将 CBAM 注意力模块引入到 ResNet 中第一层卷积 和最后一层卷积当中,如图6所示。



图 6 ResNet-CBAM 网络结构图

2.3 可微二值化函数的改进

传统的分割网络训练得到概率图后进行二 值化,标准二值化公式如式(1)所示。

$$B_{i,j} = \begin{cases} 1, & \text{if } P_{ij} \ge t \\ 0, & \text{otherwise} \end{cases}$$
(1)

其中: *P_{i,j}*表示图坐标点(*i,j*)处的概率值,它由 文本图像通过分割网络后产生,*t*是人为设定的 阈值。

由于标准二值化并不可微,不能将其放入 网络当中进行训练。因此,传统的方法需要进 行复杂的后处理操作。为了解决标准二值化不 可微的问题,DBNet模型中提出了Differentiable Binarization,一种近似的方法来做二值化,这 种方法在与分割网络一起使用时是完全可微的。 可微二值化函数如式(2)所示。

$$\hat{B}_{i,j} = \frac{1}{1 + e^{-k(P_{i,j} - T_{i,j})}}$$
(2)

其中: $\hat{B}_{i,j}$ 是二值化图像,(i,j)为像素点, $T_{i,j}$ 是 网络中学习到的阈值图,k为方法系数,在网络 训练当中充当放大错误区域梯度的作用,以加 快网络收敛。根据经验设置k = 50。

DBNet中的可微二值化函数使用sigmoid函数 来解决图像二值化不可微的问题。但是,sigmoid 函数的输出具有偏移现象,即输出均为大于0的 实值,因此,将可微二值化放入到网络训练时 BP的训练效率不高。tanh函数关于原点对称, 输出则均匀地分布在y轴两侧,是一个0均值的 函数。tanh对于文本特征相差明显时,网络学 习特征的效果会更好,并且在网络的训练过程 中会不断扩大特征的差异,使得梯度变化更快, 从而收敛速度要比sigmoid快^[19]。因此,改进后 的可微二值化函数可以减少迭代次数,提高文 本检测的性能。改进的激活函数如式(3)所示。

$$\hat{B}_{i,j} = \frac{1 - e^{-k(P_{ij} - T_{ij})}}{1 + e^{-k(P_{ij} - T_{ij})}}$$
(3)

其中: k取经验值50。

3 实验对比与分析

3.1 数据集及实验环境

本文采用公开数据集 ICDAR 2015 进行训 练。ICDAR 2015数据集是 ICDAR于 2015 年举办 的场景文本检测竞赛中使用的官方数据集,被 广大学者用于自然场景文本检测当中。数据集 文本包含多方向和任意形状,主要侧重于弯曲 文本^[20]。其中,数据集中的文本实例全部使用 四边形进行标注。该数据集中包含了 1500 幅带 有文本以及文本标签的图像,其中 1000 幅图像 用于网络训练,500 幅图像用于模型验证,可以 有效地检验算法检测性能的优劣。实验环境的 软硬件配置见表 1。

表 1 实验环境软硬件配置

软硬件	配置
操作系统	Linux
CPU	12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU@ 2.50GHz
GPU	RTX 3080(10 GB)
内存	43 GB
硬盘	25 GB
软件环境	PyTorch1.8.1, PyCharm2023, Python3.8 (ubuntu18.04), Cuda11.1

3.2 消融实验

实验使用 ResNet-18 作为 backbone, 学习率为0.001,采用 Adam 优化器,无预训练, epoch为200。为了验证本文算法的文字检测性能,选择准确率(Precision, P)、召回率(Recall, R)以

及F值(F-score)作为评价指标。准确率和召回率的公式如式(4)~(6)所示。

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{H}{TP + FN} \tag{5}$$

$$F = 2 \times \frac{1}{P + R} \tag{6}$$

消融实验的实验结果见表2。

表 2 消融实验(%	b)
------------	------------

Backbone	DB	V2S	CBAM	tanh	Precision	Recall	F-score
ResNet-18	\checkmark	×	×	×	87.0	58.2	70.1
ResNet-18	\checkmark	\checkmark	×	×	89.0	57.9	70.3
ResNet-18	\checkmark	×	\checkmark	×	89.7	58.9	71.1
ResNet-18	\checkmark	×	×	\checkmark	89.0	63.1	73.8
ResNet-18	\checkmark	\checkmark	\checkmark	\checkmark	89.2	63.6	74.3

从实验结果来看,改进特征提取网络结构 后,召回率轻微下降的同时,准确率有较好提 升,提升幅度为2个百分点。引入注意力机制 CBAM后,网络能提取到更深层次的多层网络 文本特征,增强原始文本特征并获取注意力权 重,使得准确率和F值有较大的提升,分别提高 了2.7和1个百分点,表明引进CBAM注意力机制 的有效性。通过对可微二值化函数进行改进,准 确率、召回率和F值都有较大的提升,分别提高 了2.2、5.4和4.2个百分点。三个指标全部优于 DBNet方法,具有更好的文本检测性能。

3.3 实验定性分析

在检务场景文本检测任务中,改进后的网 络模型在一般场景下的文本检测具有良好的检 测精度,检务场景文本预测区域与检测结果如 图7所示。





本文使用了传统算法中经典的文本检测算 法MSER和深度学习中的文本检测算法EAST进 行对比分析。其中,EAST算法使用以上实验环 境,训练epoch为400,得出算法检测模型并进 行文本检测测试,实验结果如图8所示。



图 8 文本检测结果对比分析

传统的 MSER 算法面对光照不足和文本空 间扭曲的场景时,出现文本漏检的问题,同时 产生文本检测框重叠的现象。在背景及排版复 杂的场景下,MSER 算法产生大量的错误检测 框,表明 MSER 算法在复杂背景场景下,存在 误检率高的缺陷。基于深度学习文本检测算法 EAST在一些光照不均、空间扭曲、背景复杂、 文本行排版不一的场景下均出现一定的漏检。 而改进后的 DBNet 文本检测依然可以准确地区 分出每个字符的位置,检测精度高,证明了该 网络模型对于检务文本检测的有效性。

4 结语

针对检务场景中存在文字图像质量不高、 空间扭曲、背景复杂、文本行排版不一等因素 所导致的文本检测结果误检漏检、定位不准确 的问题。本文对DBNet文本检测网络进行改进, 改进了ResNet特征提取网络和可微二值化函数, 同时引入高效注意力模块CBAM,能够自动增 强文本特征权重,降低误检率和漏检率。本文 提出的改进DBNet文本检测方法的准确率、召回 率和F值依次达到了89%、63.6%和74.3%,比原 DBNet方法提升了2.2、5.4和4.2个百分点。实 验表明,本文提出的改进DBNet方法性能要比 DBNet方法更优,更适用于检务场景文本检测。

参考文献:

- [1] 陈袁宇.基于深度学习的场景文本检测与识别研究
 [D].海口:海南大学,2023.
- [2] EPSHTEIN B, OFEK E, WEXLER Y. Detecting text in natural scenes with stroke width transform [C]//Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010: 2963-2970.
- [3] MATAS J, CHUM O, URBAN M, et al. Robust wide-baseline stereo from maximally stable extremal regions [J]. Image and Vision Computing, 2004, 22 (10):761-767.
- [4] 王大千,崔荣一,金璟璇.基于视觉关注模型与多尺度 MSER 的自然场景文本检测[J].应用科学学报, 2020,38(3):496-506.
- [5] ZHONG Z Y, JIN L W, HUANG S P. Deeptext: a new approach for text proposal generation and text detection in natural images [C] //Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017:1208-1212.
- [6] LIAO M H, SHI B G, Bai X, et al. Textboxes: a fast text detector with a single deep neural network [C] //Proceedings of the AAAI Conference on Artificial Intelligence, California, USA, 2017.
- [7] TIAN Z, HUANG W L, He T, et al. Detecting text in natural image with connectionist text proposal network[C]//Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, October 11-14, 2016:56-72.
- [8] SHI B G, BAI X, BELONGIE S. Detecting oriented text in natural images by linking segments [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017:2550-2558.
- [9] ZHOU X, YAO C, WEN H, et al. East: an efficient

and accurate scene text detector [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 2017:5551-5560.

- [10] DENG D, LIU H, LI X, et al. Pixellink: detecting scene text via instance segmentation [C] //Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, USA, 2018:6773-6780.
- [11] WANG W, XIE E, LI X, et al. Shape robust text detection with progressive scale expansion network [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 9336-9345.
- [12] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018:8759-8768.
- [13] LONG S, RUAN J, ZHANG W, et al. Textsnake: a flexible representation for detecting text of arbitrary shapes[C]//Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018:20-36.
- [14] LIAO M, WAN Z, Yao C, et al. Real-time scene text detection with differentiable binarization [C] //Proceedings of the AAAI Conference on Artificial Intelligence, New York, USA, 2020, 34 (7) : 11474-11481.
- [15] 宋彭彭,曾祥进,郑安义,等.基于DenseNet的自然场景文本检测[J].武汉工程大学学报,2022,44
 (3):309-314.
- [16] 李雨, 闫甜甜, 周东生, 等. 基于注意力机制与深度 多尺度特征融合的自然场景文本检测[J]. 图学学 报, 2023, 44(3):473-481.
- [17] 吴鑫磊,陶青川,张畅.基于DBNet网络的瓶盖文 字目标检测[J].现代计算机,2021(14):47-53.
- [18] 陈娅娅,刘全香,王凯丽,等.基于ResNet和迁移学 习的古印章文本识别[J]. 计算机工程与应用, 2022,58(10):125-131.
- [19] 张慧宇.广电视频文字检测与识别的研究[D]. 郑 州:郑州大学,2021.
- [20] 董维振,陈燕,梁海玲.改进DBNet与CRNN的面标识别方法[J]. 计算机工程与设计,2023,44(1): 116-124.

(下转第72页)

文章编号:1007-1423(2024)06-0032-06

DOI: 10.3969/j.issn.1007-1423.2024.06.006

联合Swin Transformer和UNet的GAN人脸修复算法

张梦澜*

(太原师范学院计算机科学与技术学院,晋中 030619)

摘要:基于GAN的人脸修复技术大都采用CNN进行修复,忽略了人脸修复的全局信息和整体均匀性,从而导致 修复结果不理想。基于此问题,提出一种联合Swin Transformer和UNet的GAN人脸修复算法,进行人脸图像修复。该 方法整体采用GAN生成器-判别器架构,使用Swin Transformer作为主干网络,用于捕捉图像的全局依赖关系;采用 UNet的编码-解码结构,在局部区域进行特征提取和重建。实验结果表明,相较于以往方法,该方法能更好地处理人 脸图像修复任务。

关键词:生成对抗网络;人脸修复;Swin Transformer; UNet

0 引言

图像修复(image inpainting)是计算机视觉的 典型应用^[1]。借助图像修复技术,可以填充图 像中的空白区域或删除其中的一些元素,在照 片编辑、照片修复、字幕消除等场景中有着广 泛的应用^[2-3]。人脸修复是图像修复的一个子 集,其目的是填补人脸图像中的缺失区域。在 修复人脸缺失部分时,需要认真考虑两个主要 问题:首先,修复的区域必须与人脸的其他部 分保持均匀,且与输入图像的周围可用区域高 度相关;其次,必须保持左右两侧的面部对称 性。许多修复方法已被提出,其中一些在修复 自然图像的缺失区域方面取得了优秀的结果。

随着生成对抗网络(GAN)^[4]的快速发展和 应用,图像修复取得了显著的成功。近年来, 生成对抗网络已被用于图像补全和修复。Pathak 等^[5]提出的方法首先将GAN应用于图像修复,并 引入了上下文编码器。Yeh等^[6]将卷积神经网络 (CNN)与GAN相结合,提出了一种深度卷积生成 对抗网络,这种方法能够产生更好的修复结果。 为了处理大面积图像修复,Lizuka等^[7]改进了上 下文编码器,并引入了同时使用全局与局部判 别器的方法。Yu等^[8]通过改进先前的模型引入 了粗略生成和精细修复的网络结构,此外,在 精细修复网络中创新性地引入了上下文注意力 层。Zeng等^[9]提出金字塔上下文编码器用于学 习图像语义信息。Hui等^[10]提出了一种用于细 粒度图像修复的密集多尺度融合网络,该网络 通过密集的扩张卷积组合来获取多尺度图像的 上下文信息,并引入自引导回归损失函数和几 何约束项来增强语义细节和改善语义空间位置, 从而提高生成图像的质量。Guo等^[11]提出了一 种双流网络用于图像修复,将修复任务分为纹 理合成和结构重建两个子任务,以生成更合理 的图像。Lin 等^[12]设计了一个深度多特征协同 学习网络用于图像修复,以提升修复效果。这 些方法在图像修复领域提供了不同的创新和改 进,以改善修复结果的质量和视觉效果。

尽管上述这些深度修复模型带来了一定的 进步,但仍然存在以下不足:一方面,在图像 修复中,必须充分考虑整个图像的上下文信息, 以更精确地预测缺失或损坏的像素。目前的图 像修复方法主要基于卷积神经网络的设计,然 而由于卷积的局部连接性,这些方法难以捕捉 图像的长距离依赖信息。研究表明,实际感受 野往往远小于理论感受野,这限制了充分利用 上下文信息进行特征捕获。虽然可以不断增加

作者简介: *通信作者:张梦澜(1998—),男,山西运城人,硕士,研究方向为计算机视觉,E-mail:1455438933@qq.com

收稿日期: 2023-09-26 修稿日期: 2023-11-14
更深的卷积层,但这会导致模型过于庞大和计 算量过大。另一方面,尽管采用了这些方法可 以保持图像的一致性和连贯性,但仍然存在一 些问题,例如修复后的图像可能会出现模糊、 结构不合理和视觉效果差等情况。这些问题可 能与修复过程中未能充分考虑全局上下文信息 和长距离依赖关系有关。因此,为了进一步提 升修复效果,需要探索更有效的方法来整合全 局信息和处理长距离依赖,以更好地修复图像 并保持其真实性和准确性。

最近,由于Transformer在自然语言处理相 关任务中表现出高效性,研究人员开始将其应 用于视觉任务。在文献[13]中,研究人员使用 Transformer进行图像补全。他们利用Transformer 的全局视野, 在处理大面积遮挡时取得了比其 他方法更好的结果。在文献[14]中,提出了一 种新颖的高分辨率图像修复框架。作者修改了 Transformer块, 以稳定处理大面积遮挡的训练, 由于Transformer在建模长程依赖方面的优势, 该架构能够生成高保真度的图像。因此,该架 构可以通过基于注意力的模型理解图像中不同 区域之间的关键依赖关系。在文献[15]中,提 出了 blind 人脸修复的方法。由于检测不同形状 和尺寸的遮挡掩膜以及恢复逼真遮挡图像的难 度,他们采用了两阶段的盲目人脸修复过程。 在第一阶段,使用了Transformer模型来检测受 损区域。接下来提出了一个网络来逐层恢复不 同级别的特征,从而基于面部未遮挡区域生成 语义一致的内容。

基于此,提出了一种联合Swin Transformer 和UNet并采用GAN思想的深度学习方法用于人 脸图像修复。具体来说,首先,本文综合考虑 人脸修复中全局上下文信息和局部细节之间的 关联,使用Swin Transformer作为主干网络。 Swin Transformer具有较大的感受野和跨层的自 注意力机制,能够捕捉图像的全局依赖关系, 并且在处理大尺度图像时表现出色。这使得它 可以更好地理解整体的图像结构,并且能够在 修复过程中保持图像的整体一致性。此外, Swin Transformer的设计使得它在计算和参数上 更加高效,相对于传统的自注意力机制网络, 具有更少的参数数量,减少了计算量。这对于 复杂的人脸图像修复任务来说是一个重要的优 势,因为人脸图像通常具有较高的分辨率和复 杂的细节。其次,结合UNet的编码-解码结构, 用于在局部区域进行特征提取和重建。UNet能 够有效地恢复缺失的细节和纹理,有助于保持 人脸图像的细微特征,如皮肤颜色和表情。通 过逐步进行特征恢复, UNet 可以更好地将缺失 区域与周围已有的人脸信息融合在一起,产生 自然且逼真的修复结果。最后,在实验方面, 我们在 CelebA 数据集上对模型进行了充分测试 和评估。实验结果表明,相较于以往的方法, 本文在人脸图像修复方面取得了更为优越的效 果。通过引入Swin Transformer 和 UNet 的结合, 不仅在修复结果上呈现出更高的质量,同时还 具备更强大的能力去保持图像的全局一致性和 局部细节准确性。

1 本文方法

如图1所示,所提出的方法被实现为一个生成对抗网络框架,其主要包括一个生成器和一 个判别器。其中生成器的主要功能是接收包含 图像缺失或损坏区域的输入图像,并尝试生成 一个完整的图像,其中缺失区域已经被合理地 填补或恢复。判别器的主要任务是接收两类图 像:原始图像(包含缺失部分)和生成器生成的 修复图像。它的目标是区分这两种类型的图像, 即判断哪些图像是真实的(原始图像),哪些是 伪造的(生成器生成的修复图像)。





1.1 生成器

编码器-解码器在多个图像处理任务中都展示了其强大的功能,因此本文的生成器设计了 一个类似 UNet 的网络结构,在生成器中引入 Swin Transformer,使用多个Swin Transformer block构成UNet网络中的编码-解码结构。首先 将输入图像切分为分块(patch),并进行嵌入, 然后构建了一系列Encoder和Bottleneck层,将 每个patch通过Swin Transformer block,它能够 对输入图像的每个分块进行自注意力机制的计 算,以捕捉分块内部和分块之间的关系。通过 多个Swin Transformer块的堆叠,编码器能够用 于从低级到高级逐层提取图像特征。

在解码器中,Swin Transformer被用于逐层 解码和恢复特征。通过多个Swin Transformer块 的堆叠,解码器能够逆向地从高级到低级的特 征表示进行解码操作,逐渐恢复图像的细节和 结构,最后通过上采样将特征图恢复到原始图 像大小,得到最终的结果。

与具有类似参数数量的卷积模型相比,Swin Transformer block并不会增加太多计算复杂性, 这是因为Swin Transformer 可以计算图像特定部 分的patch之间的相关性。此外,在图像修复任 务中,跳跃连接会降低生成器模型的图像修复 能力,这是由于它允许生成器将输入的可用部 分直接复制到输出中导致的。跳跃连接阻止模 型基于在生成器中提取的高级特征来构建输入 的整个可用和缺失部分,这会导致输出完全不 一致的面部特征。因此,我们从模型中省略了 跳跃连接。

1.2 判别器

为了确保修复后的图像在视觉上具有真实 性,引入了一个判别器网络来监督生成器生成 视觉上一致的图像。判别器网络是基于卷积神 经网络,它将输入的图像压缩成小的特征向量, 并通过比较生成的图像和真实图像的特征向量 来评估它们的相似性。通过这种方式,判别器 网络可以指导生成器生成更加逼真的修复图像。 本文中判别器由多个卷积层和归一化层构成。 将图像压缩成小的特征向量。判别网络的输入 是 256×256像素的图像,输入图像经过第一个 卷积层,用于提取图像的基本特征,包括边缘、 纹理等。该卷积层使用了 4×4 的卷积核,步长 为2,padding为1,在卷积层后面使用了 Leaky-ReLU激活函数,用于增强模型的非线性表达能 力。再使用了多个相似的卷积层,逐层提取图 像的高层次特征。每一层卷积层的输出通道数 是前一层的两倍,但不超过512个。卷积层的卷 积核大小为4×4,步长为2,padding为1,与第 一个卷积层相似。在每一层卷积层后面使用了 归一化层(BatchNorm2d)和 LeakyReLU激活函 数,用于增强模型的稳定性和非线性表达能力。 在最后一个卷积层中,输出通道数为1,用于输 出一个标量值,表示输入图像的真实程度。该 卷积层的卷积核大小和其他卷积层相同,但步 长变成了1,padding仍为1。在这一层卷积层后 面没有使用激活函数,直接输出一个标量值作 为判别结果。

1.3 Swin Transformer block

Swin Transformer block 结构如图 2 所示,由 一个基于移动窗口的 MSA 模块组成,其次是一个 两层 MLP,中间是 GELU 非线性。在每个 MSA 模 块和每个 MLP 模块之前应用一个 LN (LayerNorm) 层,在每个模块之后应用一个残差连接。



图 2 Swin Transformer block

Swin Transformer block 将输入特征图分割为 多个窗口,然后在每个窗口内利用自注意力机 制来建立窗口内元素的依赖关系。这种窗口化 的自注意力机制能够减少计算复杂度和内存消 耗,能够处理较大的图像尺寸。在窗口化的自 注意力中,Swin Transformer block 除了在每个窗 口内进行自注意力计算,还引入了局部窗口和 跨窗口的连接机制。局部窗口连接允许窗口内 元素之间的交互,而跨窗口连接则允许不同窗 口之间的信息传递,从而捕捉更大范围的依赖 关系。每个Swin Transformer 块包含一个多层感 知机(MLP),用于对窗口内的特征进行非线性 变换和映射。MLP由两个全连接层组成,通过 激活函数进行非线性变换,增强了模型的表达 能力。为了加强梯度流和模型的稳定性,Swin Transformer 块采用了残差连接和层归一化。残 差连接将输入特征和经过 MLP 变换后的特征相 加,层归一化则对相加后的特征进行归一化处 理,保证了特征的一致性和稳定性。

1.4 损失函数

在图像修复任务中,主要使用的损失函数 包括像素级别损失、对抗损失、特征匹配损失 和感知损失。在本节中,将讨论每种损失对模 型训练的影响。

1.4.1 像素级别损失

如公式(1)所示,像素级别损失是在修复图 像和真实图像之间计算的。其目标是通过考虑 面部的可见部分,使模型能够修复缺失区域, 使其与真实图像相似。然而,可见部分不能完 全描述图像中的缺失部分。因此,这种损失只 能使修复网络理解缺失部分的低级特征。

$$L_{PW}(x,\hat{x}) = \|x - \hat{x}\|_{1 \text{ or } 2}$$
(1)

式(1)中: x表示修复图像, \hat{x} 表示真实的图像, || $\|_{L_{\infty}}$ 表示的是L1和L2正则化。

1.4.2 对抗损失和特征匹配损失

对抗损失,见公式(2)~(4),试图基于真实 图像和生成图像的分布来检查修复图像的真实 性,使用了一个额外的被称为鉴别器的网络。 鉴别器的架构是基于patch的。鉴别器测试整个 图像的各个图块的真实性。x、*x*、*D*、*G*、*e*和*θ* 分别表示修复图像、真实图像、鉴别器、生成 器、鉴别器参数和生成器参数。另外,sg表示 停止梯度,指示反向传播的梯度在达到特定参 数时停止。最后,特征匹配损失,见公式(5), 是在修复图像和真实图像的中间特征之间计算 的,这些特征从鉴别器的中间层中提取。公式 (5)表示鉴别器的中间特征。对抗损失和特征匹 配损失的思想是,尽管我们无法精确地重建与 真实图像完全相似的缺失区域,但至少可以确 保修复的区域看起来是逼真的。

 $L_{D} = -E_{x} \left[\log D_{\varepsilon}(x) \right] - E_{x} \left[1 - \log D_{\varepsilon}(x) \right]$ (2)

$$L_{g} = -E_{x} \left[\log D_{\varepsilon}(\hat{x}) \right] \tag{3}$$

$$L_{Adv} = sg_{\theta}(L_{D}) + sg_{\varepsilon}(L_{C})$$
(4)

$$L_{FM}(x, \hat{x}) = \left\| MFD_{\varepsilon}(x) + MFD_{\varepsilon}(\hat{x}) \right\|_{1 \text{ or } 2}$$
(5)

1.4.3 感知损失

如公式(6)所示,这个损失是使用预训练分 割网络的编码器来计算的,用于比较生成图像和 真实图像的高级特征。在公式(6)中,x、 \hat{x} 、 ϕ 分 别表示修复图像、真实图像和预训练分割网络 的编码器。这个损失主要考虑了图像中的高级 特征,如边缘。感知损失主要关注图块和边缘 的平滑度,与图像的均匀性和对称性的重要性 无关。

$$L_{PL}(x, \hat{x}) = \left\| \phi(x) + \phi(\hat{x}) \right\|_{1 \text{ or } 2}$$
(6)

基于上述损失,我们定义模型的整体损失为 $L(x, \hat{x}) = \alpha L_{PW}(x, \hat{x}) + \beta L_{Adv}(x, \hat{x}) + \gamma L_{FM}(x, \hat{x}) + \delta L_{PL}(x, \hat{x})$

(7)

其中: α、β、γ和δ被设置为10、10、100和 20。这些值是基于实验得到的,以获得在测试 数据上的最佳结果。

2 实验结果与分析

2.1 数据集与实验设置

在 CelebA-HQ数据集上评估所提的图像修 复模型。该数据集基于 CelebA数据集,包含了 CelebA数据集中的高质量人脸图像,以及从其 他数据集中获取的更高分辨率的人脸图像。 CelebA-HQ数据集包含了30000张人脸图像,是 目前最大的人脸合成数据集之一。本文从数据 集中随机抽取了28000张人脸图像进行实验,其 中 26000张用于训练,2000张用于测试。

该模型的输入是局部缺损的人脸图像,输 出是相应的高质量复原图像。在训练模型时, 将图像的大小调整为256×256,实验中批处理大 小为5,使用 Adam 算法优化损失函数,设置网 络初始学习率为0.001。

2.2 实验环境

本文的实验平台为Windows 11系统,基于 神经网络的PyTorch 1.11.0框架和Python 3.8编译 环境,GPU为NVIDIA GeForce RTX3090。

2.3 评价指标

为了对本文模型的性能进行定量分析,从 而评估其优劣,本文采用图像修复领域常用的 评估指标:峰值信噪比(PSNR)、结构相似性 (SSIM)和均方根误差(RMSE)来评估修复结果的 性能,计算分别如下列公式所示:

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \tag{8}$$

其中: MSE 为均方值误差(Mean Squared Error)。

$$SSIM = \frac{(2\mu_x\mu_y + C1)(2\sigma_{xy} + C2)}{(\mu_x^2 + \mu_y^2 + C1)(\sigma_x^2 + \sigma_y^2 + C2)}$$
(9)

其中: x, y表示对比的两个相似图片, μ 、 σ 、 σ_{xy} 分别表示图片的均值、标准差和协方差。

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[O(i,j) - C(i,j) \right]^2 \quad (10)$$

其中: m,n表示图像的尺寸大小, O和C分别代表原始图和修复图。i,j分别代表图像上第i行, 第i列像素。

2.4 评价算法修复效果对比(CelebA-HQ)

为了证明模型的有效性,将本文提出的算法 与 CE^[5]、GLGIC^[7]、DMFN^[10]、CTSDG^[11]和 MFCL^[12]基准算法在同一数据集上(CelebA-HQ) 进行比较,以居中矩形掩膜作为缺失区域。

表1是用 PSNR、SSIM 和 RMSE 三个指标来 评价六种模型的性能,其中加粗字体为最佳性 能指标,可以得出:在六个模型中,使用本文方 法修复的图像的平均 PSNR为 27.225 dB,SSIM 为 0.9154, RMSE为11.475,与表中其它方法对 比,三项指标都有不同程度的提升,充分证明 本文方法具有更高的图像修复结果。

表	1	不同方法的人脸修复指标比较	ζ

Method	PSNR/dB	SSIM	RMSE
CE	25.770	0.8976	13.154
GLCIC	24.312	0.8697	16.027
DMFN	25.326	0.8814	14.294
CTSDG+	26.1691	0.9043	13.048
MFCL	26.263	0.9003	12.9649
本文方法	26.805	0.9154	11.475

2.5 定性评价

定性评价以主观观察为主,观察图像修复的

效果。图3显示了本文算法在数据集 CelebA-HQ 上的部分修复结果。图7(A)~(F) 六组图显示了 居中矩形掩膜作为缺失区域作用于 CelebA-HQ上 的修复结果,可以看到大部分人脸特征都因被 掩膜覆盖而丢失,但本算法也能得到全局结构和 细节纹理合理的修复结果,且图7(A)~(F)的修 复结果都与原图相似,图7(C)的修复结果与原 图人脸表情存在差异,但也满足视觉感知需求。 总体来看,本文算法在 CelebA-HQ数据集上的 修复结果都能有较为逼真且合理的语义。





图 3 本文算法的部分修复结果

3 结语

本文从改进模型的角度提高了图像修复的 质量,分析了现有方法在图像修复任务中的局 限性,提出了一种联合Swin Transformer和UNet 的GAN人脸修复算法。该算法在生成器中引入 了Swin Transformer结构,采用Swin Transformer 替换了广泛使用的CNN,解决了卷积网络无法 整合长距离信息的问题,充分结合 Swin Transformer 和 UNet 二者的优势。同时利用多个损失 函数优化网络,实现了修复性能更好的图像修 复效果。实验结果表明,本文提出的方法修复 的图像不仅具有较高的客观评价指标,而且主 观视觉效果更好,与原始图像非常逼近,在实 际场景中具有较高的应用价值。

参考文献:

- [1] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting [C] //Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, New York: ACM, 2000:417-424.
- [2] JAM J, KENDRICK C, DROUARD V, et al. R-MNet: a perceptual adversarial network for image inpainting [C] // Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2021: 2713-2722.
- ZHOU Y Q, BARNES C, SHECHTMAN E, et al. TransFill: reference-guided image inpainting by merging multiple color and spatial transformations [C] //Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021: 2266-2267.
- [4] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [J]. Commun ACM, 2020, 63(11):139-144.
- [5] PATHAK D, KRÄHENBÜHL P, DONAHUE J, et al. Context encoders: feature learning by inpainting [C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Vegas, NV, USA, 2016: 2536-2544.
- [6] YEH R A, CHEN C, LIM T Y, et al. Semantic image inpainting with deep generative models[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:

6882-6890.

- [7] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion [J].
 ACM Trans Graph, 2017, 36(4):1-14.
- [8] YU J H, LIN Z, YANG J M, et al. Generative image inpainting with contextual attention [C] //Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Lake City, UT, USA, 2018:5505-5514.
- [9] ZENG Y H, FU J L, CHAO H Y, et al. Learning pyramid context encoder network for high quality image inpainting[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, 2019: 1486-1494.
- [10] HUI Z, LI J, WANG X M, et al. Image fine-grained inpainting[EB/OL]. arXiv:2002.02609,2020.
- [11] GUO X F, YANG H Y, HUANG D. Image inpainting via conditional texture and structure dual generation [C] // Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2022:14114-14123.
- [12] LIN J Y, WANG Y G, TANG W Z, et al. Multi-feature co-learning for image inpainting [EB/ OL]. arXiv:2205.10578,2022.
- [13] WAN Z, ZHANG J, CHEN D, et al. High-fidelity pluralistic image completion with transformers [EB/ OL]. arXiv:2103.14031,2021.
- [14] LI W, LIN Z, ZHOU K, et al. MAT: mask-aware transformer for large hole image inpainting[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), New Orleans, LA, USA, 2022:10748-10758.
- [15] WANG J, CHEN S, WU Z, et al. FT-TDR: frequency-guided transformer and top-down refinement network for blind face inpainting [J]. IEEE Transactions on Multimedia, 2021, 25:2382-2392.

Combined Swin Transformer and UNet for GAN face inpainting

Zhang Menglan^{*}

(College of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China)

Abstract: Most of the GAN-based face restoration techniques use CNN for restoration, ignoring the global information and overall uniformity of face restoration, which leads to unsatisfactory restoration results. Based on this problem, a joint Swin Transformer and UNet GAN face repair algorithm is proposed for face image repair. The method adopts the GAN generator-discriminator architecture as a whole, using Swin Transformer as the backbone network for capturing the global dependencies of the image, and the encoding-decoding structure of UNet for feature extraction and reconstruction in the local region. The experimental results show that the method can better handle the face image restoration task compared to previous methods.

Keywords: generative adversarial networks; face inpainting; Swin Transformer; UNet

文章编号:1007-1423(2024)06-0038-06

DOI: 10.3969/j.issn.1007-1423.2024.06.007

基于LSTM-Attention的空间目标分类研究

杨礼友*,余显冰,李 智

(四川大学电子信息学院,成都 610065)

摘要:针对空间目标特有属性及其运动趋势难以使用单一元素进行描述,以及现有空间目标分类技术准确度低等问题,构建了一种基于LSTM-Attention的空间目标分类模型,该模型无需开展额外的特征工程,能够联系空间目标序列数据的上下文信息和长期依赖关系,提取样本的局部特征并对其长期运动趋势进行建模。利用 Mini-Mega TORTORA(MMT)系统实测光变曲线进行验证,与传统方法相比,设计使用的模型拥有较高的数据处理效率,能够提 高空间目标的分类准确度并满足空间态势感知的部分应用需求。

关键词: 空间目标分类; 深度学习; LSTM; 注意力机制

0 引言

随着近年航天活动的快速发展,进入空间 的人造卫星、碎片等目标数量呈爆发式增长。 为了更好地开展空间活动,必须要对空间目标 类别和特性开展深入研究。空间目标分类是利 用探测数据对空间目标的属性、状态等情况开 展研判的过程,其核心是通过特征空间、特征 图的变换,结合原有样本库和特征库进行学习、 训练,得到准确的分类识别结果^[1]。空间目标 分类有助于了解空间态势,为空间活动提供重 要的感知基础。光度数据作为空间目标信息的 重要展现方式,与目标特性息息相关,可以形 象地描述空间目标的多尺度特征并用于目标的 分类识别。

空间目标不能发光,通常说的光度是其表 面反射光在给定单位面积或单位立体角内的光 通量,连续测光得到的数据被称为光度、光度 曲线或光变曲线^[2],光变曲线呈现的不同模式 反映了空间目标的部分属性或状态。例如,火 箭箭体等尺寸较大目标的光变曲线周期较长, 碎片等尺寸较小目标的光变曲线周期则较短。 通过研究光变曲线,可以了解空间目标的关键 信息,能够对其运动趋势、异常姿态等情况进 行分析。但是,空间目标的光度受其材料、形 状、姿态等多种因素影响,很难使用单一元素 描述其变化^[3]。

深度学习通过学习数据的分布和模式去发 现潜在的规律和特征,不需要开展额外繁琐的 特征工程,具有很强的鲁棒性,已成为空间目 标分类领域的热门工具。鹿瑶等[4]利用仿真和 实测光变曲线数据,基于深度学习完成了空间 碎片基本类型的分类识别。Furfaro 等^[5]使用卷 积神经网络搭建了一个数据驱动型学习框架, 并基于光变曲线完成了空间目标的分类,实验 结果表明这种框架优于多种传统机器学习方法。 Linares 等^[6]基于一维卷积神经网络(1D CNNs)搭 建了深度学习模型,使用模拟数据对四种不同 旋转状态的火箭箭体进行了分类。Huo 等^[7]构建 了一个由卷积神经网络和递归神经网络组成的 模型,用于提取地球同步轨道目标的不同尺度 特征。Liu 等^[8]通过分析空间目标的数据特征, 基于长短期记忆网络搭建了空间目标分类模型, 准确识别了不同姿态和类别的目标。Raut等^[9]

收稿日期: 2023-10-18 修稿日期: 2023-11-19

作者简介:*通信作者:杨礼友(1991—),男,贵州开阳人,硕士研究生,研究方向为目标识别、时间序列分析,E-mail: yangliyou110@stu.scu.edu.cn;余显冰(1998—),男,四川攀枝花人,硕士研究生,研究方向为目标识别、时间序列分析; 李智(1974—),男,四川成都人,博士,教授,研究方向为物联网与边缘计算、智能感知与认知、目标识别

通过引入计算机视觉SOTA模型实现了光变曲线 的图像分类。综上,基于深度学习技术和光变 曲线数据的空间目标分类识别方法已较为成熟, 能够使用不同的技术提取数据特征,并使用卷 积神经网络等算法实现空间目标的分类。然而, 上述方法对数据的利用还存在不足。一方面, 不断提高的探测和存储能力导致了空间目标探 测数据的显著扩大,上述方法的处理效率已不 能满足空间目标分类实时性的要求;另一方面, 光变曲线不仅存在局部特征,而且在序列上还 根据空间目标特性存在长期的趋势变化,若只 关注其局部特征而忽略长期运动趋势,会造成 部分信息丢失。

长短期记忆网络能够实现光度数据的关键 信息在隐含层中传递和存储,不重要的信息被 忽略。带有注意力机制的神经网络会自动过滤 掉光度数据的无用信息并结合多头方式实现并 行计算,同时通过自动学习光度数据中不同位 置间的联系去深入理解输入序列。综上,本文 基于LSTM-Attention搭建了深度学习模型研究空 间目标分类,并使用实测光变曲线数据进行了 验证。

1 数据集

1.1 数据源

空间目标的光学数据主要使用地基光学望 远镜进行获取,同耗资昂贵的相控阵雷达和天 基监测系统相比, 地基光学望远镜具有探测距 离远、测角精度高、综合成本低等优势[10]。本 研究使用的样本数据来自俄罗斯喀山联邦大学 建设和使用的多功能广域监视系统——MMT (Mini-MegaTORTORA)^[11],该系统结合了视域 宽和分辨率高的优势,为空间态势提供了丰富 的感知信息。MMT系统对空间目标进行常态化 监测,探测到的光度数据被存储到数据库中。 该数据库使用公开可用的轨道根数对空间目标 进行识别,提供了包括空间目标光变信息和自 然宇宙天体的观测数据,含俄罗斯以外国家的 卫星、碎片、火箭箭体等空间目标的开源参数, 同时提供目标的工作状态及光变曲线周期性参 数等数据。



图 1 MMT 官网(http://astroguard.ru)提供的光变曲线 示意图

1.2 数据预处理

MMT数据库中的原始文件不能直接用于模型训练,需要进行相应的数据预处理,从庞杂的数据字段中提取有效信息,缩小数据规模,以便于模型训练并减少训练时间。本文从MMT系统中提取数据并开展数据预处理的流程如下:

(1)从MMT官网下载原始数据文件,剔除 其中的部分多色测光数据文件。然后根据空间 目标属性和北美航空司令部(North American Aerospace Defense Command, NORAD)发布的信 息,将样本类别分为非周期性空间目标(一类样 本)、周期性卫星(二类样本)、周期性碎片(三 类样本)和周期性火箭末级(四类样本)等类别。

(2)受探测系统故障、信号丢失等影响,提 取的数据出现了部分缺失,使光变曲线难以被 后续模型开发利用,用于实验的分类模型会因 数据不平衡问题而使结果受到影响。为了解决 光变曲线的数据缺失问题,本文在提取的数据 集上使用MissForest^[12]插补算法处理缺失值。

(3)本文直接使用了MMT系统提供的光变 周期值,用以分析光变曲线的周期性并确定样 本序列的长度。对于结构较大、易于观测的卫 星和火箭箭体,本文选择了512个连续的光变值 作为样本序列,其中至少包含了两个周期;对 于碎片等结构较小或其他有较短样本序列的目 标,本文对其使用了插值的方法,使其长度达 到了512。

最终,本文用于实验的数据由3480条一类 样本、3487条二类样本、2946条三类样本和 3846条四类样本组成。实验对这些数据重新进 行了打乱和划分,随机抽取80%的数据作为训 练集,20%的作为测试集。

2 基于LSTM-Attention的分类模型

2.1 LSTM

长短期记忆网络(long-short term memory, LSTM)是一种基于递归神经网络(recurrent neural network, RNN)的时间序列处理架构^[13],其通 过引入门机制来控制信息堆积的速度并控制一 些旧状态的省略概率,避免了 RNN因重复计算 权重矩阵导致的梯度消失和爆炸问题。LSTM使 用三个门控循环单元来控制网络记忆的更新与 遗忘,能够保证重要信息或特征被持续记忆, 而不重要的信息则被忽略,实现重要信息在隐 含层中的传递和存储。



图 2 LSTM 结构图

LSTM的核心是内部状态的循环连接,具体可以用数学公式描述如下:

$$f_t = \sigma(W^f h_{t-1} + Z^f x_t + b_f) \tag{1}$$

$$i_{i} = \sigma(W^{i}h_{i-1} + Z^{i}x_{i} + b_{i})$$

$$(2)$$

$$\sigma_{i} = \sigma(W^{\circ}h_{i+1} + Z^{\circ}x_{i} + b)$$
(3)

$$h_{i} = o_{i} \odot \tanh(c_{i}) \tag{4}$$

$$c_{t} = f_{t} \odot c_{t-1} + i_{t} \odot m_{t}$$

$$(1)$$

$$m_{t} = \tanh\left(W^{m}h_{t-1} + Z^{m}x_{t} + b_{m}\right)$$
(6)

其中: σ 和 tanh 分别是激活函数和双曲正切函数。 W^{f} , W^{i} , W^{o} , W^{m} , Z^{f} , Z^{i} , Z^{o} , Z^{m} 表示权重参数, b_{f} , b_{i} , b_{o} , b_{m} 表示偏置,它们都是网络中的可学习参数。由上述公式可知,每个门控单元的计算方式都基本相同,仅有权重学习参数和偏置学习参数不同。

2.2 注意力机制

注意力机制(attention mechanism)与人类的 视觉选择机制类似,即通过将有限的注意力集中 到重点核心信息上,既而节省资源,实现信息处 理资源的高效分配,快速有效获取关键信息^[14]。 以自注意力机制为基本单元的Transformer^[15]模型 则将注意力机制广泛应用到自然语言处理^[16]、 计算机视觉^[17]等领域。

如图3所示,在使用注意力机制计算时,首 先使用Query和Key进行相似度计算,并得到注 意力权值,然后将权值进行归一化处理得到直 接可用的权重,最后将权重和Value进行加权求 和。注意力机制通过自动学习数据中不同位置 间的关联关系,能够很好地理解输入序列,其 核心计算公式为

Attention(
$$Q, K, V$$
) = Softmax $\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V$ (7)



图 3 注意力机制结构

2.3 模型结构

结合 LSTM 和注意力机制的特点以及空间目标特性,本文搭建的空间目标分类模型如图4所示,该模型由输入、一维特征提取模块、LSTM-Attention模块、特征融合模块和输出等组成。其中, $x \in R^m 和 y \in R^n 分别表示输入和输出向量, m表示光变曲线数量, n表示类别。使用一组实例训练网络后, x将映射到类别y。本文y是大小为4×1的向量,其中的元素分别代表输入光变曲线属于非周期性空间目标、周期性卫星、周期性碎片和周期性火箭箭体的概率。训练时,每个小批量的光变曲线数据由输入层传至输出层后,用于试验的网络模型会进行误差反向传播,并根据损失函数和试验参数更新可学习参数。所有小批量信息传输结束后,完$

成一个周期的训练并得到新的分类模型,同时 在测试集上评估模型的性能。



图 4 基于LSTM-Attention 的空间目标分类模型结构

一维特征提取模块负责提取光变曲线的一 维特征,其使用1D CNNs完成光变曲线的特征 提取,卷积层后面是批标准化层(Batch Norm)以 及 ReLU函数,还使用了最大池化技术对训练集 的输出信息进行简化。每个卷积层对光变曲线 的计算如下:

$$h(\mathbf{x}) = f(W^*x + b) \tag{8}$$

其中:*是一维卷积算子,W是卷积核,f是激 活函数ReLU。

LSTM-Attention模块负责捕捉光变曲线的长期运动趋势并提取局部重点特征,包括LSTM 层、注意力机制和Dropout。输入的光变曲线经LSTM计算后,时间步的隐藏状态 h_i 和各时间步输出 o_i 拼接作为注意力机制的Query,各时间步输出 o_i 线性变换后作为Value,网络的线性变换矩阵 w_k 作为Key,Query和Value相乘后输出结果矩阵,具体计算如下:

$$\boldsymbol{Q}_{t} = \boldsymbol{w}_{q}(\boldsymbol{o}_{t} + \boldsymbol{h}_{t}) \tag{9}$$

$$\boldsymbol{V}_t = \boldsymbol{w}_v \boldsymbol{o}_t \tag{10}$$

$$\boldsymbol{Z}_{t} = \tanh\left(\boldsymbol{Q}_{t}\boldsymbol{w}_{k}\right)\boldsymbol{V}_{t} \tag{11}$$

特征融合模块负责融合上述模块提取的特征并将其作为分类函数的输入,并最终输出分类结果。该模块使用拼接操作将1D CNNs模块输出的特征图和LSTM-Attention输出的特征图进行融合,然后使用Softmax对权重进行归一化处理,并根据输出向量y确定模型的分类输出结果。

3 实验与分析

3.1 实验设置

本文采用了Python3.6作为编程和数据分析 语言,使用了Tensorflow2.0.2作为深度学习框 架,并使用了1080 Ti显卡来完成深度学习模型 的训练。此外,本文的训练模型批处理方式大 小为128,初始学习率为0.001,训练轮数为 600,更新参数的优化算法为Adam。具体实验 环境见表1。

表 1 实验设置

名称	参数
操作系统	Ubuntu16.04, 64bit
编程和数据分析语言	Python3.6
开发环境	Tensorflow, Pycharm
CPU处理器	Intel®Xeon(R)CPU E5-1650v4@3.60GHz
GPU处理器	GeForce GTX 1080 Ti × 2
内存/硬盘容量	64 GB/1.5T

3.2 评价指标

本文主要使用了深度学习领域常用的准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F-score等指标对分类模型的效果进行评价,其计算公式如下所示:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$
(12)

$$Precision = \frac{TP}{TP + FP}$$
(13)

$$Recall = \frac{IP}{TP + FN}$$
(14)

$$F\text{-score} = \frac{1+\beta^2}{\frac{1}{Precision} + \frac{\beta^2}{Recall}}$$
(15)

其中: *TP*(True Positive)表示深度学习模型的预测结果和真实情况均是正例; *FP*(False Positive)和*FN*(False Negative)表示预测和真实情况不同, *FP*的真实情况是反例,但模型的预测结果是正例,*FN*的真实情况是正例,但模型的预测结果 是反例; *TN*(True Negative)表示真实和预测结果 均是反例。

3.3 结果与分析

图5显示了模型的精度和轮数的关系,可以 反映模型在训练和测试中的相关情况。其中, 单独使用1D CNNs搭建的模型在完成约400个训 练周期后趋于稳定,最后的训练精度达99.45%, 测试精度达85.41%;1D CNNs结合LSTM搭建的 模型在完成约450个训练周期后趋于稳定,最后 的训练精度达99.48%,测试精度达85.59%;本 文基于LSTM-Attention搭建的模型在完成约300 个训练周期后趋于稳定,最后的训练精度达 99.91%,测试精度达87.94%。结果表明LSTM-Attention模型在计算效率、分类准确率上优于传 统模型。



图 5 各对比模型的训练精度和测试精度对比曲线

表2显示了本文使用的其他评价指标在对比 模型上的具体结果。

米団	齿刑	评价指标		
尖別	侠型	精确率	召回率	F-score
	1D CNNs	0.77	0.81	0.79
非周期性 空间目标	1D CNNs &LSTM	0.72	0.87	0.78
	Ours	0.83	0.84	0.83
	1D CNNs	0.90	0.69	0.78
周期性 刀星	1D CNNs &LSTM	0.88	0.81	0.84
1.2	Ours	0.88	0.86	0.87
First Have Let	1D CNNs	0.74	0.86	0.80
周期性 碎片	1D CNNs &LSTM	0.86	0.80	0.83
	Ours	0.89	0.87	0.88
	1D CNNs	0.82	0.86	0.84
向期性 火箭末级	1D CNNs &LSTM	0.90	0.84	0.87
2 Chq / 1992	Ours	0.86	0.89	0.87

表 2 空间目标光变曲线分类模型性能对比

从最简单的分类精确率来看,本文使用的 基于LSTM-Attention的深度学习模型在空间目标 分类任务中取得了最佳结果,该模型比传统的 分类模型更有优势。这是因为LSTM-Attention模 块可以有效帮助卷积神经网络处理数据,使得 模型在训练过程中减少了信息损失并提高了全 局信息交互,准确提取了光变曲线的局部重点 特征,解决了传统技术在长期依赖关系学习上的 不足,更加高效地挖掘光变曲线的隐藏特征。总 的来说,基于LSTM-Attention的深度学习模型在 空间目标分类中表现出的性能显著,通过实测 光变曲线数据集进行验证,表明其优于传统模 型,实现了快速、准确的分类。

4 结语

本文基于LSTM-Attention 搭建了空间目标分 类模型,并使用实测光变曲线对其进行了验证。 该模型从整体上突出样本数据的局部特征并联 系序列的长期趋势,全方位丰富了识别特征, 解决了传统空间目标分类模型准确度不高的问题。同时,模型的输入可直接使用原始数据, 无需开展额外的特征工程,简化了空间应用操 作。与传统分类模型相比,基于LSTM-Attention 搭建的分类模型的准确率更高。在未来研究中, 将针对空间目标载荷和姿态进行研究,为空间 态势感知应用提供参考。

参考文献:

- [1] 耿文东.空间态势感知导论[M].北京:国防工业出版社,2015.
- [2] 向守平.天体物理概论[M].合肥:中国科学技术大 学出版社,2008.
- [3] MARIANI L, ZARCONE G, HOSSEIN S H, et al. Computer vision and stochastic approach for LEO objects attitude determination [C] //Proceedings of the 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). Ankara, Turkey: IEEE, 2021:1-6.
- [4] 鹿瑶,赵长印.基于光变曲线的空间碎片基本形状 分类[J].天文学报,2020,61(6):14.
- [5] FURFARO R, LINARES R, REDDY V. Space objects classifiation via light-curve measurements: deep convolutional neural networks and model-based transfer learning [C] // Proceedings of the 19th Advanced Maui Optical and Space Surveillance Technologies Conference, Maui, American, 2018.
- [6] LINARES R, FURFARO R, REDDY V. Space ob-

jects classification via light-curve measurements using deep convolutional neural networks [J]. The Journal of the Astronautical Sciences, 2020, 67 (3) : 1063-1091.

- [7] HUO Y R, LI Z, FANG Y Q, et al. Classification for geosynchronous satellites with deep learning and multiple kernel learning [J]. Appl Opt, 2019, 58: 5830-5838.
- [8] LIU T, SCHREIBER K U. Photometric space object classification via deep learning algorithms[J]. Acta Astronautica, 2021, 185:161-169.
- [9] RAUT R, DEORE R, BOBADE S, et al. Optimization and enhancement of space object identification algorithms and implementation using deep learning network[C]//Proceedings of the 2021 Asian Conference on Innovation in Technology (ASIANCON). PUNE, India: IEEE, 2021: 1-5.
- [10] 杜俊举. 空间碎片光度数据的获取与分析[D]. 威 海:山东大学,2022.
- [11] KARPOV S, BESKIN G, BIRYUKOV A, et al. Photometric calibration of a wide-field sky survey data from Mini-Mega TORTORA [J]. Astronomische

Nachrichten, 2018.

- [12] STEKHOVEN D J, BÜHLMANN P. MissForestnon-parametric missing value imputation for mixedtype data[J]. Bioinformatics, 2012, 28(1):112-118.
- [13] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9 (8):1735-1780.
- [14] BRAUWERS G, FRASINCAR F. A general survey on attention mechanisms in deep learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2023,35(4):3279-3298.
- [15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. arXiv: 1706. 03762,2017.
- [16] GALASSI A, LIPPI M, TORRONI P. Attention in natural language processing [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32 (10):4291-4308.
- [17] KHAN S, NASEER M, HAYAT M, et al. Transformers in vision: a survey[J]. ACM Computing Surveys, 2022, 54(10s): 1-41.

Research on space object classificationbased on LSTM-Attention

Yang Liyou^{*}, Yu Xianbing, Li Zhi

(College of Electronics Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract: Aiming at the problems of the difficulty of describing the unique attributes of space objects and their motion trends using a single element, as well as the low accuracy of the existing space object classification techniques, constructs a space object classification model based on LSTM-Attention. The model does not need to carry out additional feature engineering, and is capable of linking the contextual information and long-term dependencies of space object sequence data, extracting the local features of the object and modelling its long-term motion trends. Validated using real light curve observations taken from the Mini-Mega TORTORA(MMT) system, designed to use a model with high data processing efficiency compared to traditional methods, and is able to improve the classification accuracy of space objects and satisfy some of the application requirements of space situational awareness.

Keywords: space object classification; deep learning; LSTM; attention mechanism

文章编号:1007-1423(2024)06-0044-06

DOI: 10.3969/j.issn.1007-1423.2024.06.008

基于受扰判断的智慧网络移动终端数据链多层加密算法

梁正华,温权波*,邹立朋

(贵州省科技创新中心有限责任公司,贵阳 550000)

摘要: 传统智慧移动终端的数据链防护方法通常未考虑受扰判断情况,导致单层可见性算法很容易被干扰,无法 全面保护移动终端上的敏感信息。针对这一问题,提出智慧网络移动终端数据链多层加密算法。对数据链电磁干扰机 理进行分析后,根据不同趋势和不同紊乱频率下的错误率,来判断数据链是否受到干扰。基于这些分析结果和判断, 设计多层加密算法来增强数据链的安全性。第一层加密技术用于对数据链数据的编码,以确保数据不被篡改,并能够 在接收端恢复为原始数据。第二层加密技术用于保证从数据链单元接收到的数据的安全等级,避免恶意攻击对数据的 损害。第三层加密技术用于防止病毒入侵,保护从其他数据链节点接收的数据。通过多层加密技术实现移动终端数据 链的安全防护。实验结果表明,该算法下的丢包率低,整体加权防护时间较为稳定,且生成的密文长度短。

关键词:智慧网络;移动终端数据链;多密码联合防护;防护方法

0 引言

智慧网络移动终端已经开始采用一些数据 传输端口数据来实现多连接合作机制。智慧网 络移动终端和互联网的发展使得移动用户的使 用频率增加。考虑到智慧网络移动终端与收费 之间的密切关系, 经济效益对智慧网络移动终 端的发展起到积极作用。然而,产业链中的某 些环节可能存在违规行为,这会给智慧网络移 动终端用户带来严重的隐私问题和经济损失[1]。 因此,智慧网络移动终端数据链安全问题引起 了社会各界的高度关注,确保智慧网络移动终 端数据链的安全成为当前的迫切问题。数据链 问题是限制网络发展的关键因素。然而,目前 还没有找到解决智慧网络移动终端安全问题的 方法。与移动终端数据链相比,智慧网络移动 终端的数据链处理能力有限,存储空间不足, 影响了移动杀毒软件的处理效率。目前无法保 证智慧网络移动终端数据链的安全性。

因此,有关人员对这一问题进行了广泛的 研究。针对现有环境信息数据链方法中与低传 输效率和高封装损耗有关的问题,应用网络开 发环境信息数据链方法,可以显著提高数据链 过程中环境信息传输的有效性,并有效控制环 境信息包装的损失指标。董贵山等[2]提出以密 码为基因的网络安全防护体系研究方法, 以密 码为基因,构建了一个全面覆盖且安全可靠的 网络安全防护体系。该体系利用密码技术作为 核心手段,通过身份认证、数据加密、安全传 输等方式,实现对网络安全的全面保护。但该 方法的实施难度较高,且需要更高的计算和网 络资源,从而增加了成本和运维的压力。周文 等[3]提出一种基于国产密码与拟态防御融合的 一体化内生安全防护架构,该方法在数据链过 程中具有较高的空间专业性和局限性。为了实

收稿日期: 2023-09-19 修稿日期: 2023-11-26

基金项目:贵州省科技业务电子化签署关键技术研究与示范(黔科合支撑[2021]一般 298)

作者简介:梁正华(1992—),男,贵州遵义人,本科,初级工程师,研究方向为移动终端数据链应用、数据安全、密码相关; *通信作者:温权波(1994—),男,贵州安龙人,本科,中级工程师,研究方向为科技大数据分析研究、项目预算管理研究、 企业财务数据分析、企业R&D(科学研究与试验发展)政策研究,E-mail:1154260727@qq.com; 邹立朋(1983—),男,陕西 西安人,本科,副高,研究方向为科技业务管理与政策研究、大数据分析研究、知识图谱研究、企业ERP系统设计开发、 R&D(科学研究与试验发展)政策研究

现信息方法中的信息交换,应用表面的终端设 计了无线即时报告方法。使用公共服务器来构 建云环境,以向即时报告终端移动设备提供服 务,并且在无线 P2P 网络内建立邻近节点。该 方法具有很强的稳定性和响应速度,但方法的 安全性很低,需要进一步改进。

为了确保智能网络移动终端数据链的安全 性,本文提出了一项研究智慧网络移动终端数 据链的联合多层密码防护方法。鉴于现有方法 存在的不足,本文从三个方面提出了一种优化 多连接移动数据链路的新方法,并进行了实验 验证。实验结果表明,设计的多连接移动数据 链路优化方法能够在网络波动中保持相对稳定 的传输速度和数据的安全传输。这项研究有着 重要的应用价值,它确保了多线路智慧网络移 动终端数据链的安全性和稳定性。本文的方法 适用于开发加密设备的智慧网络移动终端数据 链,以提高智慧网络移动终端安全通信方法的 安全性水平。

智慧网络移动终端数据链多密码联合 防护方法设计

1.1 数据链电磁受扰情况下单层加密存在的 问题

智慧网络移动终端数据链是用于在终端之间 传输数据的通信链路,其中每个节点都承担着数 据的转发和处理任务。而电磁干扰是指外部的电 磁信号干扰了数据链路的正常运行,导致数据传 输的错误、丢失或泄漏。为了确保在传输过程中 不会发生信息泄漏或被未授权的第三方获取^[4-5], 需要对数据链电磁受扰情况进行判断。该过程包 含三个元素,描述为*R* = (*X*,*C*,*Y*),其中*X* 描述 所有影响因素的总和,*C*、*Y* 描述所有影响因子 的范围。数据链安全指标影响因素决策表见表1。

表 1	1 数据链	安全指标影	响因素决	策表
表 1	1	女生指怀彰	响囚系伏	: 朿:

b_1	b_2	b_3	•••	b_n	Ε
Y_{b11}	Y_{b21}	Y_{b31}		Y_{bnI}	Y_{bE1}
Y_{b1M}	Y_{b2M}	Y_{b3M}		Y_{bnM}	Y_{bEM}

在终端数据链通信过程中,通过分析正弦

载波的相位变化来确定等价于属性的数据链电 磁干扰的程度。利用时域信号进行处理,利用 属性的取值来确定所属等同类别。将数据链电 磁干扰机理的总影响因素划分为等效类,决定 同一因素 *E* 和等效类有几个影响因素。正弦载 波的相位会随着二进制基带信号变化进行分析, 如果属性 *E* 中存在确定的*t* 个差异,基于属性 *E* 的判断值来确定等同于属性 *E* 的数据链电磁干 扰的类别。时域信号为

$$Y = \begin{cases} Y_{b11} & Y_{b21} & \cdots & Y_{Et} \\ Y_{b12} & Y_{b21} & \cdots & Y_{Et} \\ \vdots & \vdots & & \vdots \\ Y_{b1f} & Y_{b1f} & \cdots & Y_{bEt} \end{cases}$$
(1)

在这里,第j个电网数据链安全的影响因素 得到了第j个属性的取值,并在Y_{EI}中确定。在n 中设置智慧网络移动终端数据链方法的目标数 量,其中存在m个导入口,并且m<n。引入匿 名数据链协议的私有节点数进行设置,以计算 智慧网络移动终端数据链方法的可信度,通过 计算可信度可以确保智能网络移动终端数据链 方法的安全性^[6]。

在进行受扰判断时,应该考虑智能网络移 动终端数据链方法中初始泄露节点的概率,以 及位于匿名信道和网络节点中初始安全单元位 置的概率。这些概率值可以用来计算系统的安 全性和秘密泄露单元的受扰情况。

假设源事件是匿名数据链信道I中第一个泄漏单元的前驱节点,*H*_k表示匿名数据链信道I 中的第一个泄漏节点,*k*表示匿名信道单元中的 位置事件。可通过相干解调的方式进行处理, 表示如下:

$$H = Y + H_1 + H_2 + \dots + H_k$$
(2)

当智慧网络移动终端安全方法具有公共网 络节点n + m + z时,安全封装源的可信度为c, 智慧网络移动终端安全体系的长度为k + 1,那 么假设:

$$n+m+z=t \tag{3}$$

$$\frac{n+m+z+c}{n+m+z+k+1} = q$$
(4)

如果数据链包的传输源的可信度是10,使 用最高可信度值来计算智慧网络移动终端安全 方法的安全性,则可以计算位于匿名信道和网 络节点中的初始泄漏节点*P*(*H_i*)的概率:)

$$P(H_i) = \frac{C}{n+m+z} \left(\frac{n+m+z+c}{n+m+z}\right) = (1-q)q^{-1} \quad (5)$$

初始隐藏泄漏节点的概率位于智慧网络移 动终端安全方法中的初始安全单元位置的概率 为

$$p(H_1) = P(H_i)(1 - q^k)$$
(6)

智慧网络移动终端数据链方法的秘密泄露 单元可能位于智慧网络移动终端安全方法中的 第二数据链单元位置的概率为

$$P(H_2) = p(H_1)q(1 - q^{k-1})$$
(7)

通过对电磁干扰机理的深入理解,可以看 出,单层加密方法存在明显漏洞,无法有效保 护网络安全,无法确保数据链路的安全和可靠 性。比如:如果智慧网络移动终端数据链接收 信号的细致噪声相对偏小,则干扰信号会被部 分信号压制并取代,超过规定限制误码率解调 输出错误码元,这说明上行通道安全防护过程 受扰的情况出现。

1.2 数据链受扰判据确定

通过对干扰机理进行深入分析,能够更好 地了解数据链受扰的原因和机制,从而能够采 取相应的对策来减少或避免干扰的影响。在实 现数据链与地面控制站之间的信息交互时,最 关键的是调整链路以确保稳定的信息传输。满足 地空双向通信需求,并假设当前要传输的数据链 数据流为5 Mbps,有三条网络链路*a、b和c*。根 据相应的阈值参数1.43 Mbps^[7,8],可以判断数据 链的稳定性。如果传输速率低于1.43 Mbps,就 可能出现数据链受到干扰的情况。

为了确保数据流的稳定流动,需要暂停链路*c*上的数据传输,并调整其他链路的配置来发送数据。通过调整链路的配置,可以使数据链的 传输速率保持在阈值以上,从而确保数据的稳定 流动。在本文中,将使用等式(8)所示的算法来 控制数据链设备,从而判断数据链是否受扰。

$$z = \frac{p\left(\frac{I}{H}\right) - ap\left(H_2\right)}{E\sqrt{2v\frac{p}{3}} + \left[t_r \left(3\sqrt{3v\frac{p}{3}} \times P\left(1 + 32P^2\right)\right)\right]}$$
(8)

其中: z是拥塞控制函数, a是数据受干扰的平

均值, E是受干扰的时间, p是平均传输丢失率, v是基于提取数据发送的数据量的最终决定, t,是恢复丢失数据传输所需的时间。经过双 工器滤波后进入到接收机,减少网络连接的拥 塞^[9-10]。当数据链受到同频或邻频电磁干扰时, 调整连接受干扰的电磁,接收工作信号时接收 到的干扰信号如等式(9)所示。

$$T_a = z \left| \frac{e}{a} \right| \left| \frac{1}{E} \right|^a \left| 1 + \frac{1}{E} \right|^{e^{-a}}$$
(9)

其中: *T_a*表示相关的干扰时间, *e*是通过电子网络接收的干扰数据量, *E*是传输信号的相应长度。

为了保证稳定的频率间输出信号,通常在 接收机内部安装自动控制增强方案。这种自动 增强控制方案根据输入信号的强度自动调整其 效率,以使输出信号容量保持在一个固定的水 平,从而影响下一阶段的读取和解调方案。然 而,由于存在电磁干扰,交换频率的实际输出 速率往往低于指定的容量水平,因此相关的自 动增强控制输出值也会发生变化。当自动增强 控制的输出值达到某个临界点时,错误率将开 始呈指数级增加,并且数据链路传输的信息可 能存在一定程度的失真。随着错误率不断增加, 数据链路将无法保持稳定连接,甚至可能导致 通信中断。因此,需要根据自动增强控制的趋 势和不同紊乱频率下的错误率来判断数据链受 到的干扰程度。

基于多密码防护实现移动终端数据链的 安全防护

对数据链受扰判断进行确定后,基于多密 码防护实现移动终端数据链的安全防护。对移 动终端数据链进行多密码防护已经成为常见的 操作。加密移动终端数据链已经成为移动终端 安全管理的标准措施之一。基于数据链电磁干 扰机理分析结果和数据链受扰判断,可以设计 移动终端数据链的安全防护流程。在这个流程 中采用多层加密技术来增强数据链的安全性。

1.3.1 数据链电磁干扰机理分析

分析干扰信号的特性、来源和路径,了解 干扰对数据链的影响。

1.3.2 错误率分析

在不同趋势和紊乱频率下,监测并分析数 据链的错误率变化,确定干扰对数据链的影响 程度。

1.3.3 数据链受扰判断

根据错误率的变化趋势和与频率紊乱的关 系,以及干扰的类型和严重程度,确定数据链 是否受到干扰,并评估受扰的程度。

1.3.4 多层加密技术应用

第一层加密技术:对数据链数据进行编码, 保证数据的完整性和一致性。其加密结果为

$$E_{\mathrm{Tx}}(l,d) = \begin{cases} T_a l E_{\mathrm{elec}} + l \varepsilon_{\mathrm{fs}} d^2, d < d_{\mathrm{crosover}} \\ T_a l E_{\mathrm{elec}} + l \varepsilon_{\mathrm{fs}} d^4, d < d_{\mathrm{crosover}} \end{cases}$$
(10)

其中: *lE*_{elec}是传输过程中对数据链数据进行编码所计算的代码, *le*_id²是数据链数据的最终解密流程, *le*_id⁴是接收最终可接受误码率的数据链数据。

第二层加密技术:确保从数据链单元接收 的数据具有安全等级,避免恶意攻击损害数据。 被用于确保数据链单元中传输的数据链数据达 到最高的安全防护级别。利用第二层加密技术, 能够有效预防恶意攻击对数据链数据的破坏。 经过加密后的数据链数据必须符合严格的安全 防护标准,以确保数据链的完整性和安全性。 具体计算公式为

$$E_{\rm Rx}(l) = E_{\rm Tx}(l,d) l E_{\rm elec} W_h \tag{11}$$

第三层加密技术:防止病毒入侵,保护从 其他数据链节点接收的数据。具体计算公式为

 $E_{\text{relay}}(l,d) = E_{\text{Rx}}(l) \left(lE_{\text{elec}} + l\varepsilon_{\text{fs}}d^2 \right) \quad (12)$

1.3.5 数据链防护流程实施

根据分析结果和判断结果,制定相应的安 全防护策略。实施多层加密技术,保护数据链 的安全性和机密性。定期对防护措施进行评估 和更新,以适应不断演变的干扰和安全威胁。

通过以上流程,设计出一个综合的移动终 端数据链安全防护流程,从分析干扰机理、确 定受扰程度到应用多层加密技术。借助多密码 防护技术,智慧网络移动终端的数据链得到了全 面的安全保护。目前移动终端数据链的安全工作 中广泛应用了多密码联合防护方案,并成功建立 了相应的数据链防护程序。这一举措为智慧网络 移动终端提供了可靠的数据链安全保障。

2 实验论证

为了验证本文建立的智慧网络移动终端数 据链多层加密算法的有效性,进行了对比实验。 具体实验准备及结果如下。

2.1 实验准备

为了验证智慧网络移动终端数据链多密码 联合防护方法的可能性,在仿真环境中建立了 用于智慧网络移动终端的数据链方法的实验平 台。测试网络环境分为三种类型:Wi-Fi网络、 5G网络和4G网络。测试设备由两个SIM智慧网 络移动终端和多个移动终端数据链方法网卡组 成。PC设备应向智慧网络移动终端发送信息, 使用适当的软件登记传输的交通数据的状态, 分析和处理从智慧网络移动终端提取的数据。 该设备的具体参数见表2。

表 2 设备参数表

项目	SIM7010GF1	SIM7010GF2
IP	10.15.118.160	10.58.110.268
ETH	eth1: 169.20.5.1	eth1: 169.20.3.2
UL	1700-1850	1700-1850
NAT	SIM1-eth1	SIM2-eth2
DL	1820-1895	1820-1895

以设备加密方法传输的数据为实验对象, 利用两个系列资源实现RS485和RS232之间的 数据传输和接收。为了满足智慧网络移动终端设 备应用的要求,应使用LAN 8720A以太网控制器 来实现以太网数据链,包括速度、数据链模式, 并满足接口和频带的要求,数据链中采用的数据 链模块ASR1802S平台,可以支持最大速度和最 大连接速度,这就是150 Mbps和50 Mbps终端数 据链方法的实际应用。在实验过程中,数据链 终端的数据链方法为一组5000个智慧网络移动 终端数据链数据随机作为实验数据,将智慧网 络移动终端数据链多密码联合防护方法与设备 加密方法连接起来。从5000个智慧网络移动终 端安全数据集中选择适当且可用的数据进行试 点测试。该测试设备的 IP 地址都是具有 UL 和 DL的标准移动链路,代表两个移动频带。两个 智慧网络移动终端设备在整个数据链路径上形 成多连接。传输对象是具有不同分辨率的三个 图像。同时,本文使用Mininet平台创建了一个 虚拟网络,用于各种网络负载测试。分别选择 文献[2]和文献[3]的方法作为对比方法进行实 验测试。

2.2 对比实验

为了验证本文方法的性能,选择防护时间 作为评价指标进行实验测试,不同方法的防护 时间对比如图1所示。



图 1 不同方法的防护时间对比

根据图1可知,本文方法数据链多密码联合 防护耗时变化不大,整体加权防护时间相对稳 定,无论数据链是如何加密或解密的,都不会影 响系统编码的时间。与对照组相比,本文设计的 方法防护较为稳定。通过实验可知,本文方法的 防护性能优于文献[2]方法和文献[3]方法。

表 3 实验结果对比

方法	发送信息/条	接收信息/条	丢包率/%
本文方法	5000	5000	0
文献[2]方法	5000	4945	1.1
文献[3]方法	5000	4891	2.2

根据表3可知,在发送5000个智慧网络移 动终端数据链数据为实验数据时,文献[2]方法 和文献[3]方法均存在丢包现象,无法完全接收 发送信息。本文方法接收信息丢包率为0。相比 另外两种方法丢包率更低,证明本文方法优于 另外两种方法。

为了更进一步验证本文方法的加密性能, 对比不同明文下不同方法的密文长度。选择20、 40、60、80、100字节明文,不同方法的密文长 度对比实验结果见表4。

表 4 不同方法的密文长度对比 单位:字节

明文长度	本文方法	文献[2]方法	文献[3]方法
20	14	26	34
40	14	38	25
60	14	50	42
80	14	52	24
100	14	63	52

经过对比实验分析不同明文长度下的密文 长度,本文方法在加密性能方面表现出更优秀 的特点。无论明文长度如何变化,本文方法始 终能够生成相对较短且稳定的密文长度,为固 定的14字节。与之相比,文献[2]方法和文献 [3]方法的密文长度随着明文长度的增加而变 化。本文方法具有高效和可靠的加密能力,能 在不同明文长度下实现高效的数据保护。这些 实验结果验证了本文建立的智慧网络移动终端 数据链多层加密算法的有效性,并进一步证明 了其在加密性能方面的显著优势。

3 结语

在智能网络的不断推广中,许多不同类型 的智慧网络移动终端与电网公司的内部网络互 联,实现了内部网络应用程序和数据方法的交 换。此外,移动应用技术的发展扩大了电网公 司在智慧网络移动终端上的生产、管理等。电 网部门与公共能源供应安全密切相关,也是信 息安全要求的公共经济的重要支撑事业。许多 智慧网络移动终端被连接,并且环境相对复杂。 对信息完整性和保密性的要求不断提高。在此 基础上,本文提出了智慧网络移动终端数据链 多密码联合防护研究。对比实验表明,它能够 满足实际使用的需要,智慧网络移动终端的安 全接入方法是安全可靠的。

参考文献:

- [1] 王爱兵.基于量子密码的网络信息安全动态防护方法[J].信息与电脑(理论版),2023,35(10): 212-214.
- [2] 董贵山,姬少培,颜亮,等.以密码为基因的网络安全防护体系研究[J].信息安全与通信保密,2023
 (4):57-65.
- [3] 周文,董贵山,张汝云,等.一种基于国产密码与拟态防御融合的一体化内生安全防护架构[J].信息 安全与通信保密,2023(1):50-59.
- [4] 张习发,曾嵘.基于国产密码技术的应用安全防护 设计[J].云南科技管理,2022,35(5):33-35.
- [5] 梅莉蓉,郭晓黎,张小琼.基于密码的数据安全防护 体系研究[J].信息安全与通信保密,2022(9): 48-56.

- [6] 牛淑芬,张美玲,周思玮,等.面向智慧网络移动终端的密文可验证属性基可搜索加密方案[J].移动终端数据链工程与科学,2022,44(11):1941-1950.
- [7] 任志勇,梅启梁,徐柯.基于量子密码技术的电子档案离线状态下安全防护实现[J].山西档案,2022
 (4):141-146.
- [8] 袁捷,张峰,于乐.基于高性能商用密码的电信领域
 防护实践[J].信息安全与通信保密,2022(8):
 19-29.
- [9] 许德斌.基于量子密码的数字化档案信息安全防护 算法设计[J].廊坊师范学院学报(自然科学版), 2022,22(2):8-12.
- [10] 胡宝胜,查正朋.基于商用密码技术和零信任理念 的安全防护体系实践[J].广播电视网络,2022,29 (4):46-49.

A multilayer encryption algorithm for smart network mobile terminal data link based on disturbance judgment

Liang Zhenghua, Wen Quanbo*, Zou Lipeng

(Guizhou Science and Technology Innovation Center Co., Ltd., Guiyang 550000, China)

Abstract: The traditional data link protection methods for smart mobile terminals usually do not consider interference judgment, resulting in single-layer visibility algorithms being easily interfered with and unable to fully protect sensitive information on mobile terminals. To address this issue, a multi-layer encryption algorithm for smart network mobile terminal data link is proposed. After analyzing the electromagnetic interference mechanism of the data link, determine whether the data link is affected by interference based on different trends and error rates under different disturbance frequencies. Based on these analysis results and judgments, design multi-layer encryption algorithms to enhance the security of the data chain. The first layer encryption technology is used to encode data chain data to ensure that the data is not tampered with and can be restored to the original data at the receiving end. The second layer encryption technology is used to ensure the security level of the data received from the data link unit, and to prevent malicious attacks from damaging the data. The third layer encryption technology is used to prevent virus intrusion and protect data received from other data link nodes. Implement security protection of mobile terminal data links through multi-layer encryption technology. The experimental results show that the packet loss rate is low, the overall weighted protection time is relatively stable, and the length of the generated ciphertext is short.

Keywords: intelligent network; mobile terminal data chain; multi-password joint protection; protection method

文章编号:1007-1423(2024)06-0050-06

DOI: 10.3969/j.issn.1007-1423.2024.06.009

基于 Bi-LSMT 结合注意力机制的钓鱼网站识别

尚培文*,李东帅

(辽宁工业大学电子与信息工程学院, 锦州 121001)

摘要:随着互联网的普及,钓鱼网站是人们接触到最常见的网络犯罪类型之一,造成了巨大的经济损失、信息泄漏等危害。鉴于此,提出了一种基于双向长短记忆神经网络(Bi-LSTM)结合注意力机制模型来识别钓鱼网站。预处理之后的数据,采用Word2Vec模型构建词向量,通过Bi-LSTM进行特征提取,并使用注意力机制计算注意力权重,生成最终的特征向量,通过Sigmoid函数对网站类别分类。实验结果表明,所提出的模型在对钓鱼网站检测效果达到98.3%,能够有效应对此类攻击威胁,有助于维护网络空间安全体系。

关键词:钓鱼网站;Bi-LSTM;Word2Vec;注意力机制

0 引言

现如今,人们的大部分日常活动都是基于 互联网的,包括聊天、商务、教育和购物等。 因此,用户在互联网中的敏感信息容易受到网 络犯罪的侵害。其中,钓鱼网站是一种利用社 会工程的网络犯罪,旨在欺骗人们并窃取他们 的金融账户凭据或其他敏感数据。

现有的钓鱼网站检测主流技术主要分为以 下类别:基于URL结构的检测技术,基于页面 内容的检测技术和基于黑、白名单检测技术。 黑名单方法通过匹配当前检测的URL是否存在 于黑名单库中,以此完成钓鱼网站检测。这种 方法检测速度快、误报率低,但是由于钓鱼网 站更新迭代迅速、黑名单库不能及时更新导致 对未知的钓鱼网站检测性能较差。基于页面内 容的检测技术主要利用网页内容特征,如源代 码、DOM树结构等,训练模型以实现对钓鱼网 站的检测。这种方法具有较高的准确率,但是 需要采集大量的网页内容,检测时间过长并占 用更大的网络带宽。

针对基于网页内容检测钓鱼网站的效率问题,通过URL结构特征检测方法具有较好的适用性和轻量的优点。通过提取URL结构特征识

别钓鱼网站,主要分析URL字符串,统计URL 长短、特殊符号个数和一些关键词的信息,采 用传统机器学习算法对钓鱼网站进行识别。这 种方法可以快速地提取和处理URL特征,不需 要对整个网站进行分析,从而提高检测效率。 但最终模型预测结果的准确率十分依赖特征工 程,并且手工设计的特征可能无法捕捉到URL 结构中的更深层次的语义和结构信息。

随着深度学习算法开始应用于钓鱼网站的 识别,如卷积神经网络、循环神经网络等,可 以自动学习钓鱼网站特征表示,不需要进行特 征工程,并且可以捕捉URL结构中更深层次的 语义和结构信息。因此,本文提出双向长短记 忆神经网络(Bi-SLTM)结合注意力机制的方法 来对钓鱼网站进行检测。

1 相关工作

基于钓鱼网站对互联网用户的危害,国内 外的学者以及安全人员对如何高效地检测钓鱼 网站开始了一系列的研究,涌现出了大量的识 别方案,主要有以下三类。

1.1 黑名单检测技术

黑白名单方法是钓鱼网站检测中常用的技

收稿日期: 2023-09-11 修稿日期: 2023-09-25

作者简介:*通信作者:尚培文(1996—),男,山西晋中人,硕士,研究方向为计算机信息安全技术与应用,E-mail: cowboyw@163.com; 李东帅(1996—),男,吉林四平人,硕士,研究方向为计算机信息安全技术与应用

术,其中最典型的代表是 Google 安全浏览 API^[1],Google 使用"Google Safe Browsing"的 服务来收集和更新黑名单数据。它通过定期扫 描互联网上的网页,来检测其中是否存在钓鱼 页面或其他不安全元素。Han等^[2]提出一种自 动个人白名单(AIWL)的方法来保护用户的网络 数字身份,并使用朴素贝叶斯维护。Prakash 等^[3]由近似匹配算法将URL分为多个组件与黑 名单中条目进行匹配。Hong等^[4]通过文献调查 收集了许多词汇特征,并将它们与列入黑名单 的域相结合来提高检测性能。

1.2 基于页面内容特征检测技术

基于页面内容检测技术通过比较网站页面 的内容、外观、结构等特征,判断该网站是否 为钓鱼网站。Benavides-Astudillo等^[5]使用自然 语言处理(NLP)和深度学习算法,基于可疑网 页的文本来检测网络钓鱼攻击。Zhang 等^[6] 采 用Word2Vec模型提取钓鱼网页的一系列语义特 征。由于能够避免人为偏差,基于自动特征学 习的方法具有较强的泛化能力,更适合钓鱼网 页生命周期短、迭代快的情况。Jain 等^[7] 通过 分析网站HTML源代码中的超链接来检测网络 钓鱼攻击。将超链接特定特征分为12个不同的 类别,并使用这些特征来训练机器学习算法。 Opara^[8]将HTML源代码中所有内容输入,将输 入字符长度设置为2000,超过截断、不足使 用 </PAD>字符串进行填充补齐,之后进行字符 嵌入和单词嵌入的构建,并通过CNN提取特征。

1.3 基于URL结构特征检测技术

统一资源定位符 (uniform resource locator, URL)作为网页的唯一标识,在所有基于钓鱼网 站的恶意活动中,用户都需要点击某个 URL进 行跳转。利用 URL结构特征识别钓鱼网站与合 法网站是常用的方法之一。Abedin 等^[9]使用基 于 URL 的特征进行训练,并尝试通过使用建议 的软件提案来防止零日攻击,该提案通过分析 网站的 URL 并使用随机森林算法区分合法网站 和钓鱼网站。Kara 等^[10]从钓鱼 URL 和域名中 提取 11 个特征采用 LR、KNN、DT、RF 等六种 模型进行训练。Roy 等^[11] 以网站 URL 信息为输 入,使用 LSTM、Bi-LSTM 检测恶意 URL 的检测 机制,和门控循环单元(GRU)。Xiao等^[12]提出了 一种具有注意力机制的卷积神经网络(CNN),使 用URL作为输入,来进行网络钓鱼网站的识别。

2 模型设计

本文所设计的检测模型由预处理、嵌入层、 循环网络层、注意力模块、输出层组成。模型 以经过特定符号分割后的网站URL作为输入, 使用Word2Vec构建词向量,由Bi-LSTM对特征 提取,通过注意力机制计算注意力权重,加权 得到最终的特征向量传送至输出层进行归一化 处理,将模型输出值转换为输出概率,以数字 标签 "0" "1"的形式作为检测结果。模型结构 如图1所示。



图 1 实验模型架构组成

2.1 预处理

统一资源定位符(uniform resource locator, URL)由协议、主机名、域名以及参数部分组 成。如图2所示。



图 2 统一资源定位符

在 URL 中除协议、域名外,还存在一些特

以式(1)中的特殊符号对输入的URL进行切割,分为多个词,如图3所示。



图 3 钓鱼网站URL处理

2.2 嵌入层

Word2Vec 是将词语表示为向量的算法^[13], 它能够从大量文本数据中学习词嵌入。这些生 成的单词向量捕捉了单词之间的语义和句法关 系,使得可以进行强大的自然语言处理应用。 Word2Vec已成为自然语言处理领域广泛使用的 算法,其成功启发了许多类似的算法,旨在学 习文本数据的高质量表示。

Word2Vec有连续词袋模型(CBOW)和跳词 模型(Skip-Gram)两种主要模型。CBOW的输入 是某个特定词的上下文相关词的词向量,输出 是该特定词的词向量。Skip-Gram相反,输入特 定词的词向量,输出为其对应的上下文词向量。 CBOW的训练效率高于Skip-Gram,且更为稳 定。主要在于CBOW使用上下文平均方式训练, 每个训练步骤会涉及更多样本。而在低频率词 汇处理上,Skip-Gram比CBOW更优,由于 Skip-Gram的训练过程中,虽然高频词比低频词 出现的次数更多,但是高频词仍然是单独地出 现的,而在CBOW中,低频词通常被裹在高频 词之间作为上下文里可见度较低的一部分,而 高频词经常会在上下文里连着一起出现,相反, Skip-Gram不会刻意忽略低频词汇。

鉴于网站URL结构中域名、路径多由不规则的低频词汇构成,Skip-Gram较之CBOW在低频词汇处理上更具优越性,本文使用Skip-Gram 模型对URL进行词向量训练。

2.3 循环网络层

循环网络层的工作主要使用 Bi-LSTM 进行

特征提取。它由前向LSTM和后向LSTM组成, Bi-LSTM可以有效地增加网络可用的信息量, 提高算法的上下文理解能力。

LSTM由Hochreiter和Schmidhuber于1997年 提出,是RNN的一种进化,能够通过门学习长 期依赖并长时间记住输入信息^[14]。它由细胞状 态、输入门、遗忘门和输出门组成^[15],其中遗 忘门用于决定上一刻细胞状态信息中哪些需要 保留、哪些需要舍弃;输入门用于决定哪些信 息需要更新;输出门用于决定哪些信息需要输 出。LSTM的详细架构如图4所示。



图 4 LSTM 架构图

Bi-LSTM结构如图5所示。通过这种结构, Bi-LSTM能够捕捉到当前输入上下文信息,先 前和未来的信息均可以被利用,从而更好地理 解输入序列的语义信息。



Bi-LSTM的计算公式如式(2)~(4)所示。

$$\vec{h}_{t} = LSTM(x_{t}, \vec{h}_{t-1})$$
(2)

$$\overleftarrow{h}_{t} = LSTM(x_{t}, \overleftarrow{h}_{t-1})$$
(3)

$$h_t = \vec{h}_t \oplus \vec{h}_t \tag{4}$$

2.4 注意力层

注意力机制是一种在机器学习和自然语言 处理领域中常用的技术。它模拟人类注意力的 行为,通过对输入的不同部分不同的权重占比 来实现对信息的重点关注。由 Vaswani 等^[16]首 次提出。

注意力机制的基本思想是,在生成输出时, 模型会根据当前时间步的输入和隐藏状态计算 注意力权重向量。这个向量表示了输入序列中 每个位置对当前时间步的重要性。通过提高贡 献度更高特征的权重占比,降低贡献度低的特 征的权重占比,减少噪声干扰,以提升模型的 检测性能。

在统一资源定位符中,不同部分可能具有 不同的意义,从而对整体表达产生不同的影响。 由于Bi-LSTM模型既难以考虑每次输出的权重, 又难以区分语义影响较大的特征词和普通词的 权重,因此本研究在Bi-LSTM模型中引入了注 意力机制。这样做主要是为了体现URL结构中 各个部分对于整体所要表达的不同重要性。 URL结构中的重要词被赋予了更大的权重。注 意力实现过程包括三个步骤:

在初始动作中,为了提取特征,应用了单 层感知器来输入*h*,其公式如式(5)所示。

$$\boldsymbol{u}_{t} = \tanh\left(\boldsymbol{W}_{w}\boldsymbol{h}_{t} + \boldsymbol{b}_{w}\right) \tag{5}$$

随后,考虑随机初始向量 u_u 和 u_i 之间的相 似性,它表示每个输入对整个序列的权重。然 后,使用 softmax 回归得到归一化权重向量 α_i , 其公式如式(6)所示。

$$\boldsymbol{\alpha}_{t} = \frac{\exp(\boldsymbol{u}_{t}^{T}\boldsymbol{u}_{w})}{\sum_{j=1}^{T} \exp(\boldsymbol{u}_{j}^{T}\boldsymbol{u}_{w})}$$
(6)

最后,将注意力权重向量 α_i 与输入 h_i 的加 权平均确定为输出 γ ,表示如式(7)所示。

$$y = \sum_{i=0}^{T} \boldsymbol{\alpha}_{i} \boldsymbol{u}_{i}$$
(7)

2.5 输出层

输出层由全连接层和输出结果组成,全连 接层接收注意力模块特征向量,使用Sigmiod函 数将模型的输出映射到[0,1]区间,进行二分 类判定。当输出大于阈值时,将网站判定为钓 鱼网站,否则为正常网站。并且引入Dropout机 制,该机制会以一定的概率随机丢弃一部分节 点的输出,避免模型过拟合问题。

3 实验分析

3.1 实验数据集

为验证上述模型的有效性,本实验的数据 从 Github 和 Kaggle 收集整理了共 67997 条网站数 据,数据集分为钓鱼网站和正常网站两类数据, 其中钓鱼网站 29769 条,正常网站 38228 条。比 例约为7:3。训练集与测试集样本数量比例设 置为8:2,见表1。

表1 数据集详情

	训练集	测试集	总计
钓鱼网站数据	23815	5954	29769
正常数据	30582	7646	38228
总计	54397	13600	67997

3.2 评估方法

为验证所设计检测模型效果,使用Accuracy (准确率)、Precision(精确率)、Recall(召回率) 和F1分值评估钓鱼网站识别模型的性能。计算 公式如下所示。

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$
(8)

$$Precision = \frac{TP}{TP + FP}$$
(9)

$$Recall = \frac{TP}{TP + FN}$$
(10)

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$
(11)

其中:*TP*为真阳性,*TN*为真阴性,*FN*为假阴性,*FP*为假阳性。

3.3 实验结果分析

为验证本文提出的 Bi-LSTM 结合注意力机 制的钓鱼网站识别方法的有效性,构建了机器 学习模型决策树(DT)以及深度学习模型 LSTM、

Bi-LSTM,并进行横向对比,实验结果见表2。

2	模型评价指标对比表(%)	
---	------------	----	--

Model	Accuracy	Precision	Recall	F1
DT	95.8	95.7	95.8	95.7
LSTM	95.6	95.6	95.5	95.6
Bi-LSTM	97.1	97.0	97.2	97.2
Bi-Atten	98.3	98.1	98.3	98.1

基于以上模型的对比实验结果,准确率、 精确率、召回率、F1值的比较结果表明:

(1)本文所设计的Bi-LSTM结合注意力机制的模型在检测钓鱼网站的准确率比DT、LSTM、Bi-LSTM模型分别高出2.5、2.7、1.2个百分点;在F1分别提高2.4、2.5、0.9个百分点。

(2) 从手工提取的机器学习模型的检测与 自动提取特征的深度学习模型的实验结果来看, Bi-LSTM 从 URL 结构中提取的特征是有效的, 减少了特征提取的工作。

(3)通过LSTM和Bi-LSTM的对比实验可以看出,Bi-LSTM的评估指标优于LSTM,可以看出Bi-LSTM提取上文和下文的信息,相较于LSTM的上文信息,特征更为丰富,更有利于分类。

(4)使用注意力机制对实验分类效果具有 一定程度的提升。它通过增强贡献度高的特征 权重,降低贡献度低的特征权重来减少噪声干 扰,提升模型性能。

4 结语

本文提出 Bi-LSTM 结合注意力机制的钓鱼 网站检测方法,首先对数据集 URL 进行文本预 处理,然后使用 Word2Vec 进行词向量训练,输 入 Bi-LSTM 模型进行特征提取,并加入注意力 机制进一步捕捉关键信息,提高模型的检测性 能。实验结果表明,本文设计的模型可准确地 完成对钓鱼网站的分类任务。在未来工作中, 考虑加入网站页面内容结构特征,进一步提升 检测性能。

参考文献:

- [1] Google Developer. Safe browsing API [EB/OL]. https://developers.google.com/safebrowsing/.
- [2] HAN W, CAO Y, BERTINO E, et al. Using auto-

mated individual white-list to protect web digital identities[J]. Expert Systems with Applications, 2012, 39(15):11861-11869.

- [3] PRAKASH P, KUMAR M, KOMPELLA R R, et al. PhishNet: predictive blacklisting to detect phishing attacks [C] //2010 Proceedings IEEE INFOCOM, San Diego, CA, USA, 2010:1-5.
- [4] HONG J, KIM T, LIU J, et al. Phishing URL detection with lexical features and blacklisted domains [J].
 Adaptive Autonomous Secure Cyber Systems, 2020: 253-267.
- [5] BENAVIDES-ASTUDILLO E, FUERTES W, SANCHEZ-GORDON S, et al. A phishingattack-detection model using natural language processing and deep learning[J]. Applied Sciences, 2023, 13(9):5275.
- ZHANG X, ZENG Y, JIN X B, et al. Boosting the phishing detection performance by semantic analysis
 [C]//Proc IEEE Int Conf Big Data(Big Data), 2017: 1063-1070.
- [7] JAIN A K, GUPTA B B. A machine learning based approach for phishing detection using hyperlinks information [J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10: 2015-2028.
- [8] OPARA C, WEI B, CHEN Y. HTMLPhish: enabling phishing web page detection by applying deep learning techniques on HTML analysis[C]//Proceedings of the 2020 International Joint Conference on Neural Networks(IJCNN). Scotland:IEEE,2020:1-8.
- [9] ABEDIN N F, BAWM R, SARWAR T, et al. Phishing attack detection using machine learning classification techniques[C]//Proceedings of the 2020 3rd International Conference on Intelligent Sustainable Systems(ICISS), Thoothukudi, India, 2020: 1125-1130.
- [10] KARA I, OK M, OZADAY A. Characteristics of understanding URLs and domain names features: the detection of phishing websites with machine learning methods[J]. IEEE Access, 2022, 10:124420-124428.
- [11] ROY S S, AWAD A I, AMARE L A, et al. Multimodel phishing URL detection using LSTM, bidirectional LSTM, and GRU models[J]. Future Internet, 2022, 14:340.
- [12] XIAO X, XIAO W, ZHANG D, et al. Phishing websites detection via CNN and multi-head self-attention on imbalanced datasets [J]. Computers & Security, 2021,108:102372.
- [13] MIKOLOV T.SUTSKEVER I.CHEN K, et al. Dis-

表

tributed representations of words and phrases and their compositionality [C] //Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2013;3111-3119.

- [14] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. Neural networks, 2005, 18(5/6):602-610.
- [15] GRAVES A, JAITLY N, MOHAMED A. Hybrid

speech recognition with deep bidirectional LSTM [C] //Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc, Czech Republic, 8-12 December, 2013: 273-278.

[16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] //Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017: 6000-6010.

Phishing website identification based on Bi-LSTM with attention mechanism

Shang Peiwen^{*}, Li Dongshuai

(School of Electronics & Information Engineering, Liaoning University of Technology, Jinzhou 121001, China)

Abstract: With the popularization of the Internet, phishing websites that people contact are one of the most common types of cybercrime, causing huge economic loss and information leakage. Based on this, a model based on Bi-LSTM (bidirectional long short memory neural network) combined with an attention mechanism is proposed to recognize phishing websites. Data were preprocessed, word vectors were constructed using the Word2Vec model, features is extracted using Bi-LSTM, attention weights were calculated using the attention mechanism to generate final feature vectors, and the classification of phishing websites was obtained by using the sigmoid function. Experimental results show that the model is 98.3% effective in detecting phishing sites, and can effectively deal with the threat of phishing attacks and maintain the security of phishing sites.

Keywords: phishing websites; Bi-LSTM; Word2Vec; attention mechanism

(上接第13页)

Gait recognition method based on dual-level temporal features

Wei Yongchao¹, Xu Weiqi^{2*}, Zhu Hongchao², Zhu Zihan³, Liu Weijie²

(1. Scientific Research Office, China Civil Aviation Flight Academy, Deyang 618307, China;
2. School of Civil Aviation Safety Engineering, China Civil Aviation Flight Academy, Deyang 618307, China;
3. School of Avionics and Electrical Engineering, China Civil Aviation Flight Academy, Deyang 618307, China)

Abstract: At present, most gait recognition methods focus on the modeling of a single time scale of gait sequences, ignoring the information interaction of different time scales. Based on this, a dual-scale temporal feature representation network is proposed. This method aggregates two time level features to obtain the motion representation of gait, and fuses the features on the two time scales to achieve information interaction. Through experimental verification, the performance of this method on the data set CASIA-B surpasses the mainstream gait recognition method, and the Rank-1 accuracy rate reaches 97.8%, 93.1% and 80.6% under NM, BG and CL conditions, respectively.

Keywords: gait recognition; time level; spatial features; multi-view recognition

文章编号:1007-1423(2024)06-0056-05

DOI: 10.3969/j.issn.1007-1423.2024.06.010

基于图像亮度均衡技术的霍夫圆检测研究

任建松*,来枫璟

(重庆交通大学机电与车辆工程学院,重庆 400074)

摘要:棋子定位是下棋机器人视觉系统中至关重要的环节。目前下棋机器人视觉系统中最常用的圆检测算法为霍 夫國变换法。该检测算法具有同时可一次检测多个圆,满足对棋局实时检测的需求等优点,但在暗光照环境下也暴露出不 稳定性,导致采集到的图像存在较大的局部阴影和局部反光现象,使得检测出现偏差。为了克服其检测的光敏性,增强检 测的鲁棒性,设计了一种图像亮度均衡算法。为了验证方案的可行性,对在暗光源环境下加入图像亮度均衡算法与不加图 像亮度均衡算法分别进行了霍夫圆检测的对照实验。实验结果表明,加入图像亮度均衡算法对暗光源环境下采集的棋盘图 像预处理后,再进行检测,能够矫正偏差,准确检测到图像中棋子的内轮廓圆,准确实现棋子精准定位。

关键词:亮度均衡;霍夫圆变换;下棋机器人;双三次插值

0 引言

一个完整的下棋机器人由视觉系统、对弈 系统、操作执行机构以及人机交互界面四部分 组成。其中,视觉系统充当了下棋机器人的 "眼",能实时地扫描当前的棋局信息,并输出 给对弈系统,得到当前回合的最佳走法。一旦 棋子检测发生错误,下棋机器人便会得到错误 的棋局信息,从而造成下棋机器人对弈系统逻 辑混乱。因此,在下棋机器人视觉系统中,保 证对棋子的检测与识别是棋子定位中至关重要 的一环。棋子定位最直观的方法,就是对相机 采集到的棋盘俯视图像进行圆检测。目前常用 的圆检测算法主要有随机圆拟合度评估法、最 小二乘法以及霍夫圆变换法。前两者检测到的 圆往往具有较高的精度,但对噪声异常敏感, 且无法一次性检测图像中的多个圆; 而霍夫圆 检测算法具有较强的鲁棒性,同时可一次检测 多个圆¹¹,满足对不断变化棋局的实时检测需 求,因此常常被作为下棋机器人视觉系统中圆 检测的首选方法。

本文最初的下棋机器人视觉系统设计,采 用了任帅等[2]提出的设计流程。即对采集的棋 盘图像进行中值滤波,去除噪声干扰后直接进 行霍夫圆检测。但在后来调试过程中,发现该 设计流程中霍夫圆检测的不稳定性,即对光照 的敏感性较强。具体体现在, 当采取特殊光照 条件时,由于相机曝光能力不足,再加上棋子 本身材质问题,会导致采集到的图像存在较大 的局部阴影和局部反光现象。这类现象的出现 往往会导致OpenCV中的霍夫圆检测函数出现漏 识别以及错误识别,从而给棋子定位精度带来 很大的影响。由于大多下棋机器人都在常规光 照条件下进行测试,目前国内针对上述问题的 相关文献较少。孔凡国等^[3]提出了将采集 RGB 颜色空间的棋盘图像转化为HSV颜色空间,来 避免不同光线条件对棋子颜色的影响,同时提 取图像的H分量,并通过形态学处理得到棋子 轮廓再进行霍夫圆检测。该方法更多地是克服 光照对棋子颜色的影响, 经测试, 无法有效解 决暗光源光照下霍夫圆检测错误的问题。为解 决上述问题,本文在原视觉系统设计流程中加

收稿日期: 2023-10-22 修稿日期: 2023-11-14

作者简介: *通信作者:任建松(1996—),男,重庆彭水人,在读硕士研究生,研究方向为机电一体化技术,E-mail: 1243823640@qq.com; 来枫璟(1999—),男,重庆南岸人,在读硕士研究生,研究方向为机电一体化技术

入了一种图像亮度均衡算法,即在霍夫圆检测前,对棋局图像中的光照不均匀区域进行补偿, 从而实现图像亮度均衡的效果。经过测试,该 算法能够有效解决霍夫圆检测的光敏性。

1 霍夫圆检测

霍夫圆检测是棋子定位的第一步,也是棋 子定位的关键一环。通过霍夫圆检测,可从采 集的图像中得到棋子的圆心点像素坐标值与棋 子半径值。提取到的棋子圆心像素坐标值,一 是可通过标定算法,转化为真实的世界物理坐 标值^[4],将此坐标信息提供给机械臂,以完成 对棋子的抓取。二是可通过基点定位算法,从 而将图像中的真实棋局信息转化为二维数组的 虚拟棋盘信息,以便当前棋局信息和历史棋局 信息在整个下棋机器人软件系统中传输。

1.1 霍夫圆变换原理

霍夫圆变换的基本思路是认为图像中每一 个非零像素点都有在一个圆上的潜在可能,通 过坐标变换,将原二维坐标点转化为三维坐标 点并统计投票,最后再通过投票值与设定阈值 作比较来定位图像中的圆^[5]。

在笛卡尔二维坐标系中,圆的坐标方程为

 $(x-a)^2 + (y-b)^2 = r^2$ (1) 式中: (a,b)为二维空间中某圆的圆心坐标,r为 该圆半径。由上式可知,二维参数空间中,已 知圆心坐标(a,b)与半径,便可确定空间中的某 个圆。由a,b,r表示的三维参数空间中的每个 点,都对应于二维参数空间中的某个圆^[6]。通 过遍历图像像素坐标,代入式(1),便可解出对 应上式的r值。每计算出一个(a,b,r)值,就在 其对应的累加器A(a,b,c)上加1。遍历结束后, 当坐标点累加器值大于设定的阈值时,可视为 检测到对应坐标点下的圆^[7]。

1.2 霍夫圆检测API函数

在 OpenCV 中, 霍夫圆检测算法是封装在 HoughCircles()函数中的。该函数还集成了 Canny边缘检测算法,即在霍夫圆检测前,会先 对预处理之后的原图像进行边缘特征提取,生 成二值轮廓图像之后再进行霍夫圆检测^[8]。在 二值轮廓图像中,属于边缘特征的像素点的像 素值为255(白色),非边缘特征的像素点的像素 值为0(黑色)。由于霍夫圆变换只检测图像中的 非零像素点,因此Canny边缘检测的运用大大加 快了霍夫圆检测速度。通过Canny边缘检测得到 的棋盘效果如图1所示。



图1 处理前后效果对比图

HoughCircles()函数中有多个形参,合理调整函数中的各参数值可获得图像中的理想圆。

1.3 霍夫圆检测的光敏性

由于 OpenCV 的 HoughCircles()函数中集成 了 Canny 边缘检测算法,该算法本身对光源较为 敏感。如图 2 所示,在暗光源且相机曝光不足时 的棋盘图像经过 Canny 算子提取边缘特征后,棋 子内轮廓圆出现了不连续现象。



)暗光源下灰度图 (b)暗光源下 Canny 检测图

图2 暗光源下检测对比图

暗光源现象的出现使得霍夫圆检测出现了 漏识别或错误识别的情况,即由于Canny算子提 取的个别棋子边缘特征严重失真,导致霍夫圆 检测棋子数量不准确。错误识别情况如图3所 示,由于Canny算子提取的个别棋子内轮廓边缘 特征丢失,使得霍夫圆检测错误地识别了棋子 外轮廓部分的圆特征,从而导致检测到的圆心 点发生偏移。



图3 暗光源霍夫圆检测结果

这两种问题都会对棋子定位产生严重影响。

从上述分析可知,由于Canny边缘检测算法的存在,使得HoughCircles()函数也具有了光敏性较强的特点

2 实验方法

为提高霍夫圆检测的抗光敏性, 解决漏识 别和错误识别两大问题,本文提出了在原视觉 设计流程的基础上引入图像亮度均衡算法。在 采集的原图像转灰度、中值滤波后, 对图像光 照不均匀区域进行补偿, 使图像中局部阴影严 重区域与反光严重区域亮度相互中和, 进而使 全图达到较为均衡的亮度水平。经过上述操作, 再将预处理图像输入 HoughCircles()函数, 进行 霍夫圆检测。

2.1 亮度均衡算法原理

对于一幅大小为*M*×*N*的棋局图像而言, 可计算整幅图像的平均亮度:

$$Lum_{av} = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} p(i,j)}{M \times N}$$
(2)

式(2)中: p(i,j)为像素坐标对应的亮度值。

可通过对图像分块,并计算子块的平均亮 度值来得到图像中不同区域的亮度差异情况。 假设用大小为m×n的子块对图像进行分块,则 子块亮度均值为

$$Lum_{av} = \frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} p(i,j)}{m \times n}$$
(3)

由式(2)、(3)可得每一子块亮度与全图像 亮度的偏差:

$$\Delta_{Lum} = Lum_{av_b} - Lum_{av} \tag{4}$$

由式(4)可知,在整幅图像的局部高亮度区域,则其Δ_{Lum}值为正;相反,在局部低亮度区

域,其Δ_{Lum}值为负^[9]。为了使图像达到亮度均衡 效果,则需要对Δ_{Lum}为正的区域做亮度值衰减处 理,对Δ_{Lum}为负的区域做亮度值增强处理。为避 免亮度调整过程中子块区域灰度值的衰减和增 强出现不连续现象而导致子块图像失真,综合 考虑实验工况、算法运算时间以及上位机显存等 因素,本实验采取的图像插值算法为线性插值中 的双三次插值法。该算法主要指通过传统的数字 图像线性处理技术来实现图像插值效果^[10]。

2.2 双三次插值算法原理

传统的线性插值算法主要有最近邻插值、 双线性插值、双三次插值三种算法。其中,双 三次插值算法避免了前两种算法在插值过程中 出现的马赛克效应,是三者中边缘纹理处理最 好、插值最平滑的算法^[11]。

双三次插值算法原理如图4所示。



图4 双三次插值算法原理图

首先需要将放大图像上某一未知像素点 (*u*,*v*)逆映射为原始图像的像素点(*x*,*y*)。此时像 素点为亚像素点,即*x*,*y*值均为小数值。然后确 定原图像中点(*x*,*y*)周围与之邻近的16个像素 点。接下来,在水平方向上分别用三次多项式 插值估算点*A*、*B*、*C*、*D*处的灰度值:

$$F(x,y) = \sum_{i=0}^{3} w_{i} x^{i}$$
 (5)

式中: x^i 为水平方向每一像素点的横坐标值, $w_i(\cdot)$ 为双三次插值算法的权重基函数。同样, 在垂直方向上再做三次多项式内插:

$$F(x, y) = \sum_{j=0}^{3} w_{j} y^{j}$$
(6)

(7)

综上,便可得到对未知像素点的灰度估计值。

双三次插值算法中所用的基函数 $w_i(\cdot)$ 为 Bicubic函数,其表达式为

$$w(x) = \begin{cases} 1 - 2|x|^{2} + |x|^{3}, |x| < 1\\ 4 - 8|x| + 5|x|^{2} - |x|^{3}, 1 \le |x| < 2\\ 0, |x| \ge 2 \end{cases}$$

式中: *x* 为水平或垂直方向上四点的中心点坐标 与四点水平或垂直方向的距离。

2.3 实验设计流程

本实验在微软公司研发的Visual Studio 2022编 译平台上搭建。其中,开发环境为OpenCV 4.5.5, 代码基于 C++编写。该实验设计流程可总结为 以下步骤:

1) 通过相机实时采集当前回合棋盘图像;

2) 将采集图像由RGB三通道转换为单通道 灰度图,方便输入HoughCircles()函数,并进行 中值滤波;

3) 由式(3)求取当前图像的全局平均亮度;

4)对图像进行分块,并由式(4)计算每个子块的平均亮度;

5) 将子块亮度矩阵每个元素值减去全局平 均亮度值,得到子块差值矩阵。其中,高亮度 区域的子块差值矩阵元素值为正数,低亮度区 域子块差值矩阵元素值为负数;

6)采用双三次线性插值算法,将子块亮度 差值矩阵扩展到原图像大小,得到全图像亮度 差值矩阵;

7)用全图像亮度矩阵减去全图像亮度差值 矩阵,实现原图像亮度均衡效果;

8)将矫正后的图像输入到HoughCircles()函数,进行霍夫圆检测。

3 实验结果

通过各大小子块分块后的亮度均衡化棋盘, 得到的图像如图5~图7所示。

结合未进行亮度均衡处理的图 2(a),由上 述效果图可以看出,在亮度均衡算法中,子块 大小越小,原图像亮度均衡化的效果越明显。 但分块过小也会导致亮度过补偿,由于原图像 存在的局部阴影会分散到每个棋子上,这也使 得棋子字符特征受到影响,进而影响后期对棋 子类别的分类识别。同时,分块过小也会增加 整个算法的运算时间。经过实验研究与分析, 最终确定本实验的亮度均衡算法按照128*128大 小的子块进行分块处理。



图5 64*64子块

图6 128*128子块



图7 256*256子块

将图 8 与图 2(b)对比可知,在暗光源下, 经过亮度均衡化后 Canny 算子提取的边缘轮廓特 征更加明显,没有图 2(b)棋子轮廓部分丢失的 情况,保持了棋子轮廓的连续性。虽然棋盘的 轮廓特征与到图 1(b)所示的正常光照下仍有差 距,但棋子轮廓的连续性并不会对后续的霍夫 圆检测造成影响。

最后,经过图像亮度均衡算法矫正过的暗 光源下霍夫圆检测结果如图9所示。对比亮度均 衡前的图2(b),能明显看到棋子的内轮廓圆已 能被准确地检测到。



下棋机器人视觉系统改进后,再经过数轮 运行测试。测试过程中,对照每回合真实棋局 信息与上位机显示的每回合虚拟棋盘信息发 现,下棋机器人视觉系统的识别率已达100%, 并满足实时性要求。同时,观察每回合机械臂 抓取棋子表现,已无因棋子定位误差导致的抓 空现象。

4 结语

本实验改进了下棋机器人视觉系统,能够 有效克服下棋机器人在低光照或暗环境条件下, 由于相机曝光不足造成的霍夫圆检测错误情况, 弥补了相关工况中棋子检测识别错误以及棋子 定位不准等缺陷,有效提高了该视觉系统的工 作效率。

参考文献:

- [1] 张霄,彭维.基于Hough变换的圆形物体的检测
 [J].传感器与微系统,2006(4):62-64.
- [2] 任帅,张云飞.一种智能博弈象棋机器人[J].科学 技术创新,2021(13):187-188.
- [3] 孔凡国,李智宗,刘庆,等.中国象棋机器人视觉系统设计[1].机械工程师,2020(5):1-3.
- [4] 田宇轩. 中国象棋机器人双目视觉定位和机械手运 动学的研究[D]. 重庆;重庆大学,2019.

- [5] 刘爱胤,熊根良,姚健康,等.基于霍夫圆检测的标志物多图像特征定位方法[J]. 计算机仿真,2021, 38(3):341-345.
- ZHAO S, CHEN C. Algorithm of location of chess-robot system based on computer vision [C]
 //Proceedings of the 2008 Chinese Control and Decision Conference, Yantai, Shandong, 2008: 5215-5218.
- [7] LUQMAN H M., ZAFFAR M. Chess brain and autonomous chess playing robotic system [C]//Proceedings of the International Conference on Autonomous Robot Systems and Competitions, 2016:211-216.
- [8] SU K L, LI B Y. Implementation of the chess game artificial intelligent using mobile robots [C]//Proceedings of the 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS), Kitakyushu, Japan, 2014: 169-174.
- [9] 彭兴邦,蒋建国.一种基于亮度均衡的图像阈值分 割技术[J]. 计算机技术与发展,2006(11):10-12.
- [10] 席佳祺,陈晓冬,汪毅,等.细节保持的非均匀光照 图像亮度均衡算法[J].光电工程,2019,46(4): 47-58.
- [11] 李英民. 图像双三次插值算法的研究[D]. 兰州:兰 州大学,2020.

Hough circle detection based on image brightness equalization technology

Ren Jiansong^{*}, Lai Fengjing

(School of Electromechanical and Vehicle Engineering, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: The positioning of chess pieces is a crucial link in the vision system of chess robot. Hough circle transform is the most commonly used circle detection algorithm in the vision system of chess robot. The detection algorithm has the advantages of detecting multiple circles at the same time to meet the needs of real-time detection of chess games, but it also exposes instability in dark lighting environment, resulting in large local shadows and local reflections in the acquired images, which makes the detection deviation. In order to overcome the photosensitivity and enhance the robustness of detection, an image brightness equalization algorithm is designed. In order to verify the feasibility of the scheme, a control experiment of Hough circle detection with and without image brightness equalization algorithm is carried out respectively in the dark light source environment. The experimental results show that after adding the image brightness equalization algorithm to the chessboard image collected under the dark light environment, the error can be corrected, the inner contour circle of the chess pieces in the image can be accurately detected, and the precise positioning of the chess pieces can be accurately realized.

Keywords: brightness equalization; Hough circle transformation; chess robot; bicubic interpolation

DOI: 10.3969/j.issn.1007-1423.2024.06.011

基于改进 YOLOv5 的路面坑洞检测设计

周研逸,周月娥*,沈琳芸,沈 立,赵远东

(南京理工大学紫金学院,南京 210023)

摘要:随着道路交通的增加,对于路面检测中目标的准确识别成为了一个重要的研究课题。提出了一种基于改进 YOLOv5的目标检测算法,通过优化损失函数来提高路面检测中目标的识别精度。该设计算法对路面坑洞检测准确率 达到了89.1%,相较于原始YOLOv5算法提升了7.7个百分点,同时出现漏检现象较少,具有较好的检测精度。结果表明,算法在目标检测任务中取得了较好的效果,准确性和实时性得到较高提升。

关键词:目标检测;优化损失函数;路面检测;改进YOLOv5

0 引言

随着社会现代化和城市化进程的推进,公 共道路网络不断拓展和延伸。我国的公路行业 已经从最初的建设阶段进入了修护保养阶段。 在公路建设方面,更多地采用了沥青混凝土路 面,然而,受到高温、严寒、雨水冲刷以及车 辆挤压等一系列外界因素的影响,长期以来, 道路路面会出现不同程度的损伤,如龟裂、坑 洼等现象。这不仅仅增加了行车的风险,也大 大降低了道路的安全性和质量稳定性。因此, 在公路建设后期,对于路面的检测和实时维护 显得尤为重要。只有通过及时的检测和维护, 才能确保公路的安全性、舒适性以及持久性, 为人们提供更加便捷、安全的出行环境。

目标检测技术在计算机视觉领域得到广泛 应用,然而传统算法在处理目标检测时存在困 难,尤其在路面检测任务中。

本文旨在基于改进 YOLOv5 的目标检测算 法^[1],通过优化损失函数设计,提高精度和鲁 棒性。 我们分析传统目标检测算法,发现漏检和 误检是目标检测的主要挑战。对此,本文改进 了YOLOv5算法并重新设计了损失函数,以适 应目标检测任务。同时提出了一种新的损失函 数设计方法,综合考虑目标的大小、位置和形 状等特征信息。同时引入针对目标检测的数据 增强策略,提升鲁棒性和泛化能力。

在公开的路面检测数据集上进行大量实验 验证,算法在各项评测指标上优于主流方法, 且具有良好的实时性能。

本文提出的基于改进 YOLOv5 的目标检测 算法,提高了路面检测任务中目标的检测精度 和鲁棒性。该算法在交通监控、自动驾驶等领 域具有实际意义,为相关研究和应用提供新思 路和方法。

1 YOLOv5算法

1.1 Y0L0v5算法简介

YOLOv5是一种用于目标检测的深度学习算法^[2]。它采用了一种称为单阶段检测的方法,

· 61 ·

收稿日期: 2023-10-07 修稿日期: 2024-01-31

基金项目: 江苏省大学生创新创业训练计划项目(202313654046T)

作者简介:周研逸(2003—),男,江西南昌人,在读本科,研究方向为嵌入式;*通信作者:周月娥(1980—),女,硕士, 副教授,研究方向为自动控制及人工智能,E-mail:zyenj@163.com; 沈琳芸(2002—),女,浙江湖州人,在读本科,研究方向 为金融经济; 沈立(2002—),女,江苏苏州人,在读本科,研究方向为金融经济; 赵远东(2004—),男,广东江门人,在读 本科,研究方向为电气工程

能够通过一次前向传播来同时检测图像中的多 个目标。相较于之前的版本,YOLOv5具有更高 的精度和更快的速度。它的核心思想是使用一 个卷积神经网络来直接预测目标的类别和位置。 它将输入图像分成多个网格,并在每个网格上 预测每个目标的边界框和置信度分数。同时, YOLOv5还引入了一种称为特征金字塔网络 (FPN)的方法,能够有效地处理不同尺度的目标。

与其他目标检测算法相比,YOLOv5有几个 优点。首先,它具有更好的性能和精度,能够 检测到更小和更复杂的目标。其次,YOLOv5拥 有更快的推理速度,可以实时应用于实际场景。 此外,YOLOv5的训练过程也更加简单和高效。 现阶段已经被广泛应用于各种领域,如智能监 控、自动驾驶和机器人等。它在目标检测任务 中表现出色,并为实现智能化技术提供了一种 高效而准确的解决方案。

1.2 优化损失函数^[3]

IoU_Loss 是一种用于衡量预测框和真实框 之间重叠程度的损失函数。它通过计算两者的 交集与并集之比来表示重叠程度。然而, IoU_Loss存在一个问题,即当预测框和真实框 没有重叠部分时,IoU值为0,无法反映出两者之 间的距离关系。为了解决这个问题,GIoU_Loss 被提出,它通过引入最小外接矩形和外接矩形 与并集的差集,以增加相交尺度。然而, GIoU_Loss仍然无法完全解决问题,因为当预测 框和真实框相交时,GIoU_Loss 就退化为IoU, 不能准确区分框的相对位置关系。

为了提高模型的准确性,我们提出了两种 新的损失函数:DIoU_Loss和CIoU_Loss。这两 个函数在原先的基础上考虑了预测框和真实框 之间的重合和矩中距离,还额外考虑了它们的 长宽比。当真实框完全包裹预测框时,可以直 接比较它们之间的距离,这样可以更快地让函 数收敛。通过优化这三个因素,可以更好地衡 量两个框的相似程度。也就是说,会考虑长宽 比对于判断两个框是否相似的影响。具体的计 算方法如式(1)所示:

 $CIoU_Loss = 1 - CIoU$ = 1 - $\left(IoU - \frac{Distance^2}{Distance_C^2} - \frac{\nu^2}{(1 - IoU) + \nu}\right)$ 同时*v*被用于评估长宽比的一致性,具体定 义如下:

$$\nu = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w^p}{h^p} \right)^2$$
(2)

该算法的损失函数分为三个方面:边界框 回归损失、类别损失和置信度损失。交并比 (IoU)常用于评估边界框回归损失,它的定义如 下所示:

$$IoU = \frac{Intersection}{Union}$$
(3)

从交并比中,可以得到预测框和真实框之 间的距离,这间接反映了算法检测的效果。但 是它作为损失函数有以下缺点:假如两个框之 间没有交集,那么它的 IoU 就为0,这点并不能 充分反映两者的重合程度;同时当 IoU=0时, IoU_Loss=1-IoU=1,不存在梯度回传,这会导 致算法无法继续学习和训练,所以在 YOLOv5的 边界框回归损失函数上采用了 CIoU_Loss。

EloU的惩罚项目是以CloU的惩罚项为基础,将长宽比的影响因素分解开,使其单独地计算出目标框和锚框的长度和宽度。该损失函数的重叠损失和中心距离损失仍然和CloU中的计算方法相同,但是宽高损失会使目标盒和锚 盒宽度和高度的差异最小,从而加快了收敛。EloU惩罚项公式如下所示:

$$L_{\text{EIoU}} = L_{\text{IoU}} + L_{\text{dis}} + L_{\text{asp}}$$

= 1 - IoU + $\frac{\rho^2(\boldsymbol{b}, \boldsymbol{b}^{\text{gt}})}{c^2} + \frac{\rho^2(\boldsymbol{w}, \boldsymbol{w}^{\text{gt}})}{C_w^2} + \frac{\rho^2(h, h^{\text{gt}})}{C_b^2}$
(4)

2 训练算法及其实验结果分析

2.1 训练环境

(1)

本次训练数据集选用公开路面坑洞的数据 集。使用LableImage标注工具将1000张照片进 行特征物标记,从而生成目标。本次研究使用 的是PASCAL-VOC格式的数据集。

当前所进行的实验平台主体上由两个部分构成,其中硬件平台CPU为i7-12700H,GPU平台为NVIDIA GeForce RTX2060。其中软件平台配置具体为:Ubuntu18.04软件操作系统Anaconda3-2020.07-Linux-x86_64、Cuda10.2、Cudnn7.6.5、Python3.8、PyTorch1.7、OpenCV3.4、Torch1.6.0。

2.2 模型算法评估指标

对于算法的评估,需要对这三个指标进行 判断:mAP、Precision和Recall。计算公式如下,表1为计算公式的含义。

$$Precision = \frac{TP}{TP + FP} \times 100\%$$
(5)

$$Recall = \frac{TP}{TP + FN} \times 100\%$$
(6)

$$mAP = \frac{\sum_{L=1}^{N} P(L) \Delta R(L)}{M}$$
(7)

表 1 公式含义

公式 符号	含义
TP	模型正确预测正样本的个数
FP	模型将负样本错误检测为正样本的个数
FN	正样本被错误识别为负样本的个数
Recall	召回率,即模型正确检测出正样本的个数所占的比率
N	测试集中样本总数
М	检测任务中类别的种数
P(L)	模型在同时识别L个样本时精确率的值
R(L)	模型识别的样本个数从1-1变为1时,召回率的变化 情况

交并比(intersection over union, IoU)是目标 检测识别算法中用于评估预测检测框和实际检 测框之间重叠程度的重要指标。IoU被广泛运用 于基于深度学习与其他目标检测算法的研究中, 被视为综合计算评估算法模型性能的重要标准。 当IoU值较高时,表示预测区域与实际边界较大 的相交区域,这表明模型的预测准确度较高、 性能较好。相反,当IoU值较低时,说明模型的 性能较差。经过路面坑洞检测识别算法的脚本 测试,我们得出的结论是,该模型的IoU值大于 0.5,这意味着模型在预测路面坑洞时具有较高 的准确性和性能。

2.3 实验数据分析

图 1~图4所示,为训练后各指标变化图。 YOLOv5算法测试mAP值为81.4%,通过优化损 失函数后改进的 YOLOv5算法测试mAP可达到 89.1%。



2.4 网络模型

改进的YOLOv5算法训练数据集所运用的权 重文件为YOLOv5s.pt。YOLOv5s.pt是YOLOv5 的最小版本,它的模型体积相对较小,同时在 保持很高的检测精度的情况下,具有较快的运 算速度,适用于在算力受限的设备上运行。

对训练后的数据图片进行验证,通过图片 置信值大小评估算法优劣性。在数据集图片中 通过识别框框出特征物 "pothole",并标注出置 信值数值,如图5和图6所示。



图 5 pothole1(mAP=0.9)



图 6 pothole2(mAP=0.6)

由此可见,算法识别特征物精准度是较高的。但是根据不同情况下图片清晰度不同所反 馈的情况来看,图片清晰度不高会使识别效果 置信值降低,但未出现漏检情况。这说明改进 后的YOLOv5算法在各类情况下对特征物的检 测都较好,优化损失函数可以增强YOLOv5算 法对图片信息的处理,更好地应对特殊情况。

3 结语

本文介绍了一种改进的基于YOLOv5的路面 检测算法。相比于传统的路面检测方法,该算法 在检测路面缺损方面取得了显著的提升,通过优 化损失函数修正了传统方法的不足。尽管目前的 成果在路面缺损检测方面表现出一定的提升,但 算法仍存在一些不足之处。借助现代技术发展, 希望能够提高路面维护工作的效率,并为实现更 好的路面检测和维护贡献一份力量。

参考文献:

- [1] 林莉,姜麟,张志坚.基于AE-Tiny YOLOV3的小 目标检测模型[J].软件导刊,2022,21(3):55-61.
- [2] 马琳琳,马建新,韩佳芳,等.基于YOLOv5s目标检 测算法的研究[J].电脑知识与技术,2021,17(23): 100-103.
- [3] 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.

Design of road pothole detection based on improved YOLOv5

Zhou Yanyi, Zhou Yue'e*, Shen Linyun, Shen Li, Zhao Yuandong

(Zijin College, Nanjing University of Science and Technology, Nanjing 210023, China)

Abstract: With the increase of road traffic, the accurate identification of targets in pavement detection has become an important research topic. To enhance the recognition accuracy of targets in pavement detection, an improved YOLOv5-based target detection algorithm is proposed in this design, which optimizes the loss function. In comparison to the original YOLOv5 algorithm, compared with the original YOLOv5 algorithm, the accuracy of road potholes detection reaches 89.1% surpassing the original YOLOv5 algorithm by a significant margin of 7.7%. At the same time, there were fewer missed detection phenomena, and the design algorithm had a good detection accuracy. The proposed algorithm yields excellent outcomes in the object recognition task, as demonstrated by the results, and its precision and real-time performance are enhanced.

Keywords: object detection; optimized loss function; road surface detection; improve YOLOv5

(上接第19页)

Document level event argument recognition based on BERT and graph attention network

Wang Kai^{*}, Liao Tao

(School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China)

Abstract: Event argument recognition is one of the sub-tasks of event extraction. Its purpose is to identify event-related arguments and corresponding argument roles in text. The research shows that sentence dependency is helpful to the task of event argument recognition. However, it is easy to introduce irrelevant arguments in the construction of discourse dependency relations to generate noise, which is poorly handled by existing methods. To solve this problem, an article level event argument recognition model based on BERT and graph attention network is proposed. This model solves the noise problem from two perspectives. On the one hand, it constructs more effective discourse dependency syntactic features by obtaining sufficient semantic features of the text. On the other hand, the graph attention network is used to assign different weights to different argument nodes to eliminate invalid arguments. The experimental results on RAMS corpus show that the proposed method can effectively solve the noise problem in the context-dependent syntactic relation, and obtain a good result of context-level event argument recognition.

Keywords: document level event argument recognition; dependency syntactic relation; BERT; graph attention network.

DOI: 10.3969/j.issn.1007-1423.2024.06.012

基于Transformer的蛋白质相互作用预测研究

靳晓宏,韦文山*

(广西民族大学电子信息学院,南宁 530000)

摘要:研究蛋白质相互作用(PPIs)在生物学中具有非常重要的意义,蛋白质相互作用及其相互作用类型的研究对于 理解正常和疾病状态下的细胞生物学过程至关重要,从而有助于治疗靶点的识别和新药物的设计。针对传统蛋白质相 互作用预测模型预测精度不够高的问题,提出了一种基于Transformer的模型TPPI对蛋白质进行相互作用的预测。五倍 交叉验证实验结果表明,该模型在人类蛋白质数据集上预测的准确率为0.944,表现出了较好的预测准确性。

关键词: Transformer; 蛋白质; 相互作用; 预测

0 引言

蛋白质-蛋白质相互作用(PPIs)是生物体内 重要的生物学过程,涉及到细胞增殖、分化、 代谢以及信号转导等多种生命活动, PPIs 的识 别是PPIs 研究的重要环节。目前,常用的PPIs 识别方法包括酵母双杂交、免疫共沉淀、串联 亲和纯化、质谱等技术。其中,酵母双杂交和 免疫共沉淀技术较为常见,但难以大规模开展, 而串联亲和纯化和质谱技术则可以实现大规模、 高效率的 PPIs 筛查。氨基酸序列是蛋白质的一 级结构,最容易获得。在过去的几十年里,通 过高通量测序技术确定的蛋白质序列数量迅速 增加,这为利用蛋白质序列预测 PPIs 提供了前 提和可能性。近年来,随着机器学习技术的发 展,人们提出了一系列基于统计数据和特征工 程的计算方法。这些方法在 PPIs 预测方面比高 通量生物技术更有效,如支持向量机、随机森 林、KNN、多层感知器和集成极端学习机。一 些工作对识别 PPIs 的机器学习方法进行了系统 的评估。这些方法中的大多数密集特征工程使 得预测任务变得更加繁琐和琐碎。

深度学习技术的最新进展在许多领域提供 了重要的贡献,包括生物信息学、语音识别、 计算机视觉和自然语言处理。深度学习架构在 解决一些具有挑战性的生物信息学问题方面已 经显示出了强大的能力,如RNA结合位点预 测11、药物-靶点结合识别22和蛋白质结构预 测[3]。与机器学习相比,深度学习方法可以在 没有先验知识的情况下从大规模原始数据集中 自动提取有效特征。因此,许多基于深度学习 的方法已经被提出用于 PPIs 预测^[4]。例如, Hashemifar 等^[5]提出了一个框架 DPPI,使用堆叠 卷积神经网络(CNN)捕获蛋白质序列高维位置 特异性谱的特征; DNN-PPI^[6]模型使用 CNN 和长 短期记忆神经网络,利用从蛋白质一级序列中 自动学习到的特征来预测 PPIs; PIPR^[7]体系结 构在暹罗架构中加入深度残留循环 CNN, 在 PPIs识别方面显示出了良好的性能; Song 等^[8]提 出了融合蛋白质序列特征和结构特征的TAGPPI 模型来研究PPIs。

前人的研究工作,虽然在预测蛋白质相互 作用方面取得了相对较好的结果,但是预测准 确率有待进一步提高,因此本文采用基于Transformer 的方法对前人的TAGPPI模型进行改进, 得到TPPI模型,并在人类蛋白质数据集上进行 五倍交叉验证,得到的预测准确率为0.944,相 较于在TAGPPI模型上得到的预测准确率0.925 提高了1.9个百分点。TPPI模型的工作流程如 图1所示。

收稿日期: 2023-10-18 修稿日期: 2024-02-01

作者简介: 靳晓宏(1999—), 女, 山西阳高人, 硕士研究生, 研究方向为生物信息学; *通信作者: 韦文山(1965—), 男, 广西南宁人, 硕士生导师, 教授, 研究方向为现代教育技术, E-mail: wws1996@163.com



图 1 TPPI 模型的工作流程图

1 材料与方法

1.1 数据集

为了与现有的其它几种先进的用于蛋白质--蛋白质相互作用预测的方法进行比较,使用了 从DIP数据库中收集的人类蛋白质--蛋白质正样 本和从Negatome数据库中收集的人类蛋白质--蛋 白质负样本,共有1698种人类蛋白质、正负样 本各为1160对。

1.2 数据预处理

利用 Elnaggar 等^[9]提到的基于 Transformer 模型的蛋白质语言模型进行蛋白质的向量表示,如图2所示。



图 2 生成蛋白质向量表示的流程图

对一个蛋白质序列先进行标记化,并添加 位置编码。生成的向量通过Transformer模型, 为每个输入氨基酸创建上下文感知嵌入,得到 蛋白质的向量表示,输出大小为L×1024,L是 输入蛋白质的氨基酸个数。

1.3 基于GATv2的结构特征的计算

根据蛋白质的结构文件构建了蛋白质的接触 图文件,构建的蛋白质的接触图的大小为 $L \times L$, 然后把经由基于Transformer模型的蛋白质语言模 型(pLM)得到的蛋白质的向量表示赋值给由接触 图形成的蛋白质图(蛋白质图中的顶点代表氨基 酸,边代表两个氨基酸间有接触),此时生成的 蛋白质图中的每个顶点都有了特征,将顶点有特 征的图送入一层GATv2层,之后送入最大池化层 和包含128个神经元的线性层操作,以确保结构 特征提取模块的输出大小固定,最后,对于一对 蛋白质空间图 $G_i 和 G_j$,我们提取出了它们的结构 特征向量,分别为 $F_{g1} \in \mathbf{R}^{1\times 128} 和 F_{g2} \in \mathbf{R}^{1\times 128}$ 。

1.4 基于TextCNN的局部序列特征的计算

采用堆叠的三层卷积层和三层最大池化层 进行蛋白质局部序列特征的提取,它的输入是 基于 Transformer 模型的 pLM 编码得到的蛋白质 的向量表示 $L \times 1024$,在提取蛋白质局部序列 特征的过程中,将L的最大值定义为1200,以 确保 TextCNN 模块的输出向量大小固定,对于 长度小于 1200 的序列用 0 填充输入矩阵。对于 输入的蛋白质向量表达经过第一次卷积层的输 出大小为 1198 × 128,在经过第一次卷积层的输 出大小为 1198 × 128,在经过第一次最大池化 层之后的输出大小为 399 × 128,重复上述的操 作三次,得到蛋白质对局部序列的特征大小分 别为 $F_{s1} \in \mathbf{R}^{1 \times 128}$ 和 $F_{s2} \in \mathbf{R}^{1 \times 128[10]}$ 。

1.5 预测模块设计

在最后的预测模块将上述步骤得到的蛋白质的局部序列特征(F_{s1} 和 F_{s2})和结构特征(F_{gc1} 和 F_{g2})通过 $F_{gc1} = (1 - w)F_{g1} + wF_{s1}$ 和 $F_{gc2} = (1 - w)F_{g2} + wF_{s2}$ 进行融合,将融合后的特征进行连接并输入到MLP层后经过Sigmoid函数得到最终的预测结果,0或1。其中,1代表输入的蛋白质对有相互作用;0代表输入的蛋白质对没有相互作用。

1.6 评价方法与评价指标

本研究采用五倍交叉验证的方法来评估 TPPI模型的性能。在每次实验中,数据被平均 分成五份,取其中四份作为训练集,剩下的一 份作为测试集,这样可以确保训练集和测试集 之间没有相同的数据对。最终结果取五倍交叉 验证结果的平均值^[10]。

从七个方面进行二元预测的评估,分别是 准确率(Accuracy, ACC)、精度(Precision, Prec)、 召回率(Recall, R)、特异性(Specificity, Spec)、 F1 分数(F1 score, F1)、ROC 曲线下面积(area under curve, AUC)、PRC 曲线下面积(the area under the precision-recallcurve, AUPRC)。

2 实验结果及分析

2.1 实验工具和模型参数设置

TPPI模型搭建和实验主要应用了深度学习 框架PyTorch,使用Python编程语言。同时,实 验利用了GPU强大的计算能力,即NVIDIA Quadro RTX,24 GB内存。TPPI模型利用Adam 进行优化,在人类蛋白质数据集上训练epoch设 置为50,学习率和batchsize分别为0.001和32。 为了避免过拟合,在训练时采用了Dropout技 术。其他参数采用PyTorch提供的默认值。

2.2 实验设计

为了将TPPI模型与其它方法进行比较,在 相同的数据集(人类蛋白质数据集)上进行了蛋 白质相互作用的预测。模型都采取五倍交叉验 证,最后的结果都取五折结果的平均值。

2.3 TPPI 模型与其它方法的比较

TPPI模型在人类蛋白质数据集上的结果见 表1。根据表1可以得出,在人类蛋白质数据集 上,模型的平均准确率(ACC)为0.944、平均精 度(Prec)为0.953、平均召回率(Recall)为0.935、 平均特异性(Spec)为0.954、平均F1分数为 0.944、平均ROC曲线下面积为(AUC)为0.986 和平均PRC曲线下面积(AUPRC)为0.987。

表 1 模型在人类蛋白质数据集上五倍交叉验证的结果

Fold Set	准确率	精度	召回率	特异性	F1分数	AUC	AUPRC
1	0.960	0.981	0.938	0.982	0.959	0.986	0.986
2	0.938	0.938	0.938	0.938	0.938	0.980	0.980
3	0.942	0.971	0.911	0.973	0.940	0.988	0.989
4	0.944	0.946	0.942	0.946	0.944	0.988	0.990
5	0.938	0.930	0.946	0.929	0.938	0.986	0.988
Average	0.944	0.953	0.935	0.954	0.944	0.986	0.987

TPPI模型与前人模型比较的结果见表2。由表2可以得出,与其它四种模型PIPR^[7]、 SCNN^[7]、TransPPI^[11]和TAGPPI^[8]相比较,TPPI 模型性能显著提升。这表明该模型能更好地对 蛋白质的相互作用进行预测,具有更高的预测 准确率。

表 2 不同方法在人类蛋白质数据集上的性能比较

Model	准确率	精度	召回率	特异性	F1分数	AUC	AUPRC
PIPR	0.923	0.900	0.953	0.950	0.925	0.977	0.975
SCNN	0.937	0.927	0.948	0.947	0.937	0.983	0.981
TransPPI	0.936	0.915	0.961	0.911	0.938	0.987	0.987
TAGPPI	0.925	0.923	0.929	0.921	0.925	0.977	0.974
TPPI (ours)	0.944	0.953	0.935	0.954	0.944	0.986	0.987

注: 粗体字为该列中的最佳结果。

3 结语

本文对TAGPPI模型进行了改进,使用基于 Transformer的蛋白质语言模型对蛋白质序列进 行编码,同时使用GAT的改进版GATv2进行蛋 白质结构特征的提取,节省了对计算资源的使 用量,提高了模型预测的准确率。

参考文献:

- [1] ZHANG S, ZHOU J T, HU H L, et al. A deep learning framework for modeling structural features of RNA-binding protein targets [J]. Nucleic Acids Research, 2016, 44(4): e32.
- [2] ZHENG S J, LI Y J, CHEN S, et al. Publisher Correction: predicting drug-protein interaction using quasi-visual question answering system [J]. Nature Machine Intelligence, 2020, 2(9):551.
- [3] JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold [J]. Nature, 2021, 596(7873):583-589.
- [4] 刘桂霞,王沫沅,苏令涛,等.基于深度神经网络的 蛋白质相互作用预测框架[J].吉林大学学报(工学 版),2019,49(2):570-577.
- [5] HASHEMIFAR S, NEYSHABUR B, KHAN A A, et al. Predicting protein-protein interactions through sequence-based deep learning [J]. Bioinformatics, 2018,34(17):i802-i810.
- [6] LI H, GONG X J, YU H, et al. Deep neural network based predictions of protein interactions using pri-

mary sequences[J]. Molecules, 2018, 23(8):1923.

- [7] CHEN M H, J-T JU C, ZHOU G Y, et al. Multifaceted protein-protein interaction prediction based on Siamese residual RCNN[J]. Bioinformatics, 2019, 35 (14):i305-i314.
- [8] SONG B S, LUO X Y, LUO X L, et al. Learning spatial structures of proteins improves protein-protein interaction prediction [J]. Brief Bioinform, 2022, 23 (2):bbab558.
- [9] ELNAGGAR A, HEINZINGER M, DALLAGO C,

et al. ProtTrans: toward understanding the language of life through self-supervised learning [J]. IEEE Trans Pattern Anal Mach Intell, 2022, 44 (10) : 7112-7127.

- [10] 蔡标,葛成,徐晴,等.基于Transformer 网络的抗癌 肽的预测[J].现代计算机,2022,28(18):9-15.
- [11] YANG X D, YANG S P, LIAN X Y, et al. Transfer learning via multi-scale convolutional neural layers for human-virus protein-protein interaction prediction [J]. Bioinformatics, 2021, 37(24):4771-4778.

Transformer-based protein interaction prediction research

Jin Xiaohong, Wei Wenshan*

(College of Electronic Information, Guangxi Minzu University, Nanning 530000, China)

Abstract: The study of protein interactions (PPIs) is of great significance in biology, and the study of protein interactions and their interaction types is essential for understanding cell biological processes in normal and disease states, thereby aiding the identification of therapeutic targets and the design of new drugs. In this paper, aiming at the problem that the prediction accuracy of traditional protein interaction prediction models is not high enough, a Transformer-based model TPPI is proposed to predict protein interactions. The results of the five-fold cross-validation experiment show that the prediction accuracy of the model on the human protein dataset is 0.944, showing good prediction accuracy.

Keywords: Transformer; protein; interaction; forecast

(上接第25页)

Lightweight human body detection algorithm suitable for edge devices

Zhou Ning, Tao Qingchuan^{*}, Peng Boxing

(College of Electronics Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract: A lightweight human detection algorithm based on YOLOv7 is proposed to address the issue of complex human detection networks that perform poorly when deployed on edge devices. The algorithm first replaces the original network ELAN module with the improved ShuffleNev2 basic module; Next, add SE attention and SPPF pooling at the end of the backbone network; Then, in the Neck section, the improved GSConv is used to replace the standard convolution, and the GSConv based VoVGSCSP is introduced to replace the ELAN-W module. The validation results on GPU and Sophon SE5 show that this lightweight human detection algorithm loses 2.6% accuracy compared to YOLOv7, but significantly reduces computational complexity. The inference speed on Sophon SE5 reaches 54 FPS, which is 39 FPS higher than YOLOv7.

Keywords: YOLOv7; target detection; ShuffleNet; edge computing; Sophon SE5
DOI: 10.3969/j.issn.1007-1423.2024.06.013

基于改进的LSD线特征提取算法研究

谭帅奇1,孙 皓2*,李健铭1

(1.重庆工商职业学院电子信息工程学院,重庆 401520; 2.重庆理工大学管理学院,重庆 400054)

摘要:基于LSD 直线段检测算法可能导致直线提取碎片化的问题,提出了改进的LSD 线特征检测算法,根据角度和距离对直线进行筛选,解决了直线提取碎片化问题,采用LBD(line band discriptor)描述相邻帧图像之间的线特征相似度,提高了视觉 SLAM 位姿优化的计算效率,在公开数据集上对LSD 算法与改进后的算法进行对比实验。结果表明,改进后的 LSD 在匹配精度上提升了 27.8%。

关键词:视觉 SLAM; LSD-SLAM; 线特征

0 引言

在人造弱纹理场景下,为了解决视觉SLAM 算法在弱纹理环境中特征点提取效果不佳的问题,引入线特征,旨在提供更丰富的环境信息, 解决此问题。

直线段检测算法(LSD)是一种基于直接法的 视觉SLAM算法,LSD-SLAM是LSD算法的第一 个版本,由Engel等^[1]于2014年提出。该算法使 用线段作为特征点,通过直接法进行匹配,实 现机器人的定位和地图构建。LSD-SLAM使用了 半密集的深度图像,通过直接法进行匹配,从 而避免了特征点提取和匹配的过程,提高了算 法的效率和鲁棒性,成为了基于直接法的SLAM 算法的代表之一,但LSD算法也会使得直线提 取碎片化。为解决传统LSD算法将一条直线分 成多段的问题,本文提出了改进的LSD线特征 检测算法。

1 线特征提取

本文提出的LSD算法改进思路为:首先对 图像进行预处理,利用线段密度滤波器对图像 内的线段区域进行筛选,若区域内线段密度大 于阈值,则剔除该区域;然后对剩下区域计算 像素值梯度,当梯度值大于阈值则进行LSD线 特征提取;最后将提取得到的线段根据线段之 间的夹角大小关系来决定合并还是剔除,最终 得到提取后的线特征图像。

LSD算法的主要思想是将图像中的像素点 分成若干个支撑域,然后在每个支撑域内寻找 直线段。如图1所示。



图 1 LSD 线段支持域^[2]

对于每个支撑域,计算其内部的梯度方向和 梯度大小,像素点 $P_i(x,y)$ 在x轴方向的像素梯度 $g_x(x,y)$ 和在y轴方向的像素梯度 $g_y(x,y)$ 分别为 $g_x(x,y)=$

$$\frac{i(x+1, y) + i(x+1, y+1) - i(x, y) - i(x, y+1)}{2}$$
(1)

基金项目:重庆开放大学重庆工商职业学院2023年度科研项目(NDYB2023-02)

收稿日期: 2024-01-12 修稿日期: 2024-02-01

作者简介:谭帅奇(1997—),男,重庆人,硕士研究生,研究方向为计算机视觉;*通信作者:孙皓(1997—),女,重庆人,硕士研究生,研究方向为数智管理,E-mail:907493839@qq.com;李健铭(1998—),男,重庆人,硕士研究生,研究方向为 图像处理

 $g_y(x,y) =$

· 70 ·

$$\frac{i(x, y+1)+i(x+1, y+1)-i(x, y)-i(x+1, y)}{2}$$
(2)

根据 $g_x(x,y)$ 与 $g_y(x,y)$ 得到 $P_i(x,y)$ 点的梯 度幅度G(x,y)和梯度方向*LLA*,分别为

$$G(x, y) = \sqrt{g_x^2(x, y) + g_y^2(x, y)}$$
(3)

$$LLA = \arctan\left(\frac{g_x(x, y)}{-g_y(x, y)}\right)$$
(4)

根据梯度幅度G(x,y)计算梯度强度 τ_{xy} :

$$\tau_{xy} = \begin{cases} 0, |G_{xy}| \leq G_s \\ 1, |G_{xy}| > G_s \end{cases}$$
(5)

根据梯度强度 τ_x 构建梯度强度阈值 ρ_x :

$$\rho_{xy} = \frac{\sum_{x=1}^{K} \sum_{y=1}^{K} \tau_{xy}}{k^* k}$$
(6)

利用ρ_{xy}来进行筛选候选点,若像素点梯度 幅值大于阈值,则设为候选点。将候选点组合 成直线段,找到一个最小外接矩形与其对应。 对于检测到的直线段,进行非极大值抑制和阈 值处理,得到最终的直线段。

LSD算法的优点是能够检测出较短的直线 段,并且对于噪声和图像变换具有较好的鲁棒 性。尽管LSD算法在实时性和线段提取方面表 现出色,但它仍存在一些不足之处,比如某些 特定的线段可能会被复杂地分割成多个部分。

2 基于角度与距离的线段合并

在线段合并中,本文采用角度、水平距离 和垂直距离对线段集合进行筛选,得到候选线 段组,然后对剩下的线段进行排序,选出次长 的线段,重复上述过程,直到所有的线段都被 分组为止。这样可以减小计算量、提高匹配效 率,同时保证匹配的准确性。

设LSD算法提取的直线 L_i 的长度为 l_i ,水平 方向的夹角为 θ_i ,构成直线两段的坐标分别为 $(x_{Ai}, y_{Ai}), (x_{Bi}, y_{Bi}),合并步骤如下:$

 LSD算法是一个局部区域算法,在局部 区域提取的线段中,长度越长的线段其特征维 度越高,因此需先对所有线段按长度的降序排
 序,记为集合P₁:

$$P_{L} = \{L_{1}, L_{2}, L_{3}, \cdots, L_{n}\}$$
(7)

2) 对线段集合 P_{L} 中每条线段 L_{i} 比较水平方向的夹角 θ_{i} ,将角度差小于角度阈值 θ_{i} 的线段筛选到集合 L_{θ} :

$$L_{\theta} = \left[L_i \in L: \left(\left| \theta_i - \theta_i \right| \right) < \theta_i \right]$$
(8)

3) 将集合 L_{θ} 中线段按照水平距离筛选,水 平距离差小于阈值 d_{μ} 的线段放入集合 L_{h} :

$$L_{h} = \left[\forall L_{i} \in L_{\theta} : \left\{ \min\left(\left| x_{Si} - x_{S1} \right|, \left| x_{Si} - x_{E1} \right|, \left| x_{Ei} - x_{S1} \right|, \left| x_{Ei} - x_{E1} \right| \right\} \right]$$
(9)

4) 将按照水平距离筛选得到的线段按照垂 直距离进行筛选,垂直距离差小于阈值*d*_a的线 段放入集合*L*_a:

$$L_{v} = \left[\forall L_{i} \in L_{h} : \left\{ \min \left(\left| y_{Si} - y_{S1} \right|, \left| y_{Si} - y_{E1} \right|, \right. \right. \right. \\ \left| y_{Ei} - y_{S1} \left|, \left| y_{Ei} - y_{E1} \right| \right) < d_{v} \right\} \right]$$
(10)

5) 将经过长度、角度和距离分组后得到的 线段集合*L*_v记为*P*_L, 选取直线*L*_j, 若*L*_j自身的长 度小于最短长度*d*_w, 则给予合并得到集合*L*_d:

$$L_{d} = \left[\forall L_{i} \in P_{L}: \left(d_{ij} < d_{i} \right) \right]$$
(11)

在集合L_a中,如果两条线段的夹角小于阈 值,则可以将这两条线段合并成为一条新的线段 M。此时,需要检查新线段M与原来的长线段L₁ 之间的角度误差是否小于角度阈值,如果小于, 则可以用新线段M来代替原来的长线段L₁:

$$L_{M} = \left[\forall L \in L_{d} : \left| \theta_{i} - \theta_{1} \right| < \theta_{s-i}^{*} \& \left| \theta_{M} - \theta_{1} \right| < \frac{\theta_{s}}{2} \right] (12)$$

经过线段合并的LSD算法提取的直线,具 有特征显著和鲁棒性高的特性,为线特征匹配 提供了丰富的纹理信息。如图2所示。

图 2(a)是 LSD 算法提取线特征后,未经过 线段合并和剔除的效果,图 2(b)是 LSD 提取后 经过线段合并的效果,可以看出,经过线段合 并后的效果明显优于未合并的效果。

表1为Hough直线检测算法、LSD算法与本 文算法在线特征提取上的性能对比,通过表中 数据可知,本文算法在提取数量上分别较前两 种方法减少了77.6%和27.8%,避免了直线检测 碎片化问题,提高了线特征匹配的速度与精度。 同时平均线段长度也介于二者之间,线段长度 过长或过短都会影响匹配效果,本文算法取得 了较好效果。



(a) LSD 算法



(b) 本文算法

图 2 LSD 算法与本文算法的线特征提取效果对比

表 1 各个算法性能比较

算法类型	Hough	LSD	本文算法
提取数量/个	174	54	39
耗时/s	0.65	0.41	0.48
平均线段长度/cm	1.94	5.49	4.27

通过表1数据可知,本文算法在特征提取上 只比LSD算法多耗时0.07 s,体现在系统运行 上,高出17%;但在提取平均线段的长度上, 则减少了22.2%。

3 实验结果与分析

在 EuRoc 数据集上对本文算法与 LSD-SLAM 进行性能对比,从定位精度和建图效果两方面 定量衡量性能差异。本文选用绝对轨迹误差 (ATE)^[3]进行评价,首先根据时间戳将各时刻的 真实位姿与估计位姿关联,然后使用奇异值分 解将两个轨迹对齐,使轨迹的出发点一致。最 后,统计所有时刻(假设为*n*)真实位姿与估计位 姿之间的差异,并采用均方根误差(RMSE)^[4]进 行计算:

$$RMSE(Tr_g, Tr_e) = \sqrt{\frac{1}{n} \sum_{i=0}^{n} trans} (T_i^g T_i^e)^2 \quad (13)$$

其中: Tr_{s} 代表整条真实轨迹, Tr_{i} 代表整条估计轨迹, $i(0 \le i \le n)$ 代表时刻, T_{i}^{s} 、 T_{i}^{e} 分别表示 i时刻真实位姿和估计位姿, 符号 trans(·)表示位姿的平移向量, 对应于相机光心在空间中的三维位置。

在 EuRoc 数据集上,本文使用绝对轨迹误 差对各算法的定位精度进行对比。表2展示了各 算法的 RMSE。

表 2	EuRoc数据集上本文方法与LS	D-SLAM的	
	RMSE比较	单位:	m

Sequence	ORB-SLAM2	LSD-SLAM	Ours
MH_01	0.241	0.107	0.043
MH_02	0.191	0.213	0.025
MH_03	0.499	0.131	0.015
MH_04	0.377	0.024	0.018
MH_05	0.371	0.192	0.014

由表2可以看出,在室内图像序列MH_01至 MH_05中,线特征较为显著的情况下(弱纹理场 景),本文算法定位精度显著高于ORB-SLAM2与 LSD-SLAM。

4 结语

本文根据像素梯度变化构建图像局部支持 域,计算支持域内的梯度方向,由梯度方向阈 值构建最小外接矩阵,从而提取图像中的线特 征。根据LSD线特征提取算法的缺陷,提出了 对LSD算法的改进策略,从线段之间的夹角、 水平距离和垂直距离出发,对LSD所提取的直 线进行筛选,再对分类后的线段集合根据端点 长度进行合并,解决直线提取碎片化问题,并 对比了LSD算法和本文改进算法分别提取图像 线特征的效果,结果表明改进后的算法显著优 于原算法。

参考文献:

[1] ENGEL J, SCHPS T, CREMERS D. LSD-SLAM: large-scale direct monocular SLAM[C]//Proceedings of the European Conference on Computer Vision. Springer, Cham: Computer Vision Press, 2014: 1357-1362.

- [2] CARUSO D, ENGEL J, CREMERS D. Large-scale direct SLAM for omnidirectional cameras [C] //Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015:141-148.
- [3] 王睿.单目视觉 SLAM 关键技术研究[D]. 徐州:中国矿业大学,2020.
- [4] 吴雪梅,刘志强,张天龙,等.双模型结合进一步降 低预测均方根误差和均方根相对误差的方法[J]. 分析化学,2015,43(5):754-758.

Research on improved LSD line feature extraction algorithm

Tan Shuaiqi¹, Sun Hao^{2*}, Li Jianming¹

School of Electronic Information Engineering, Chongqing Technology and Business Institute, Chongqing 401520, China;
 School of Management, Chongqing University of Technology, Chongqing 400054, China)

Abstract: In view of the problem that the LSD straight line segment detection algorithm may lead to the fragmentation of straight line extraction, proposes an improved LSD line feature detection algorithm, which filters straight lines based on angles and distances, solves the problem of fragmentation of straight line extraction, and uses LBD descriptors to describe The similarity of line features between adjacent frame images improves the computational efficiency of visual SLAM pose optimization. A comparative experiment was conducted between the LSD algorithm and improved algorithm on a public data set. The results show that the improved LSD improves matching accuracy by 27.8%.

Keywords: visual SLAM; LSD-SLAM; line features

(上接第31页)

Research on improved text detection algorithm for prosecutorial scenarios based on DBNet

Yu Xiao*, Lin Shiji

(School of Electrical Engineering and Automation, Tianjin University of Technology, Tianjin 300384, China)

Abstract: In the text detection of inspection scene, the existing detection algorithms still have the problems of high false detection rate and high missing detection rate. By improving the existing feature extraction network, introducing the efficient channel attention and spatial attention module CBAM, and improving the differentiable binary function, the improved network is applied to the text detection of inspection scene. The accuracy, recall and F value of the improved algorithm on ICDAR 2015 data set increased by 2.2, 5.4 and 4.2 percentage, respectively, to 89.2%, 63.6% and 74.3% compared with those before the improvement. Experimental data show that the improved DBNet text detection algorithm has a significant improvement in convergence speed and detection accuracy.

Keywords: text detection; prosecutorial scenes; sifferentiable binarization; deep learning; CBAM

实践与经验

文章编号:1007-1423(2024)06-0073-07

DOI: 10.3969/j.issn.1007-1423.2024.06.014

基于不确定性的机坪管制系统风险评估

廖 勇*,黄 杰,方 娜,赵一阳

(中国民用航空飞行学院空中交通管理学院, 广汉 618300)

摘要:对机坪管制系统进行风险评估是提高机场安全水平的一个重要措施,但如何科学客观地评价机坪管制系统, 得到贴近现实的评估结果是一个难题。基于此,建立了机坪管制系统风险评价体系,从组合权重法的观点出发,通过 乘法归一化合成评价权重,并在模糊集理论的基础上引入DS证据论对某机场机坪管制系统进行实例评估,得出人员因 素是机坪管制系统安全体系的重点这一结论,对机坪管制系统安全改进有较好的指导意义。

关键词: 机坪管制; 组合权重法; 不确定性; 隶属度函数; DS证据理论; 风险评估

0 引言

近些年来,我国民航业发展迅速,民航的 运输总周转量、旅客运输量比同期都有较大增 长。根据交通运输部公布的数据,我国民航在 2023年一季度的运输总周转量,高达239.9亿吨 公里,同比增长39.7%。近几年航班量的急剧增 加,导致机场管制情况愈加复杂,管制压力加 大^[1]。为了应对相关状况,民航局做出大型机 场开展机坪移交的决策,即塔台管制系统将机 坪管制移交给机场相关方,便于提高管制效率。 2013年8月民航局发布《关于推进航空器机坪 运行管理移交机场管理机构工作的通知》的有 关要求^[2]。为进一步推进机坪运行管理移交工 作,2018年4月民航局颁布《航空器机坪管制 移交工作总体方案》^[3]。

目前对于风险评估的研究中,针对不同的评估目标采用不同的方法,比如层次分析法^[4]、灰 色关联法^[5]等可以消除评估信息的不确定性; Zhao等^[6]在民用机场安全综合评估中运用相关方 法;在专家可信度确定方面,国内外学者采用方 法为DS证据理论^[7]。风险识别是风险评估关键准 备工作,对此Edwards^[8]教授开发了SHEL模型, Hawkins^[9]进一步完善。另外对于评估方法的研究 总体上来说是从基于认知走向基于模型,在认知 方面,王燕平^[10]基于机坪事故案例建立机坪安全 评价指标体系,提出风险管理方案;Adhikari^[11] 提出了在新环境下航空安全和安保的风险框架。 而对于模型的运用,高浩然等^[12]基于熵权法从人 机环管四个方面选取指标建立机场运行安全效益 评价模型;夏桂书^[13]建立了机坪安全系统评估模 型;陈明亮等^[14]提出了基于绩效的机坪运行评估 模型;Wong 等^[15]建立改进的事故频率模型来评 估机场运行风险。

当前对于机坪管制系统方向的研究较少, 特别是基于管制移交的现状下,机坪管制的风 险评估系统需与时俱进,适应当下新环境。另 外在风险评估中,专家评价有较强的主观性和 不确定性,证据理论能降低主观性的影响并且 处理不确定性信息。因此采用隶属度函数将专 家的评价定量化,再运用DS证据理论进行数据 融合,得到可信度较高的结果。

收稿日期: 2024-02-22 修稿日期: 2024-03-06

基金项目:四川省科技重点研发项目(2023YFG0163);中国民航飞行学院大学生创新创业训练计划项目 (S202210624129)

作者简介:*通信作者:廖勇(1983—),男,四川资阳人,教授,博士,研究方向为空中交通管理和综合交通优化, E-mail:cafucairport@163.com; 黄杰(1999—),男,江西赣州人,硕士研究生,研究方向为空中交通运输; 方娜(1998—),女, 四川绵阳人,硕士研究生,研究方向为空中交通运输; 赵一阳(1997—),男,山东烟台人,硕士研究生,研究方向为航空运行管理

1 机坪管制的评价指标体系

机坪管制通过危险源识别构建指标体系, 以此确保机坪安全,并采取必要的措施来规避 风险。需要从多个方面进行,从事故因果连锁 理论和"4M"理论出发,结合民航机场安全审 计指标体系^[16],对中国民用航空安全信息系统 (Aviation Safety Information System of CAAC)中近 5年统计的800起机坪管制系统不安全事件进行 原因分析。可以将机坪管制风险评价指标分为 两级,第一级由四个指标组成,即人员因素 (*A*)、设备因素(*B*)、环境因素(*C*)和管理因素 (*D*)。*A、B、C*和*D*可继续划分为第二级的各项 具体指标,通过这些指标来建立具有针对性的 评价指标体系,如图1所示。



图 1 机坪管制风险评估的评价指标体系

2 机坪管制风险评估建模

2.1 基础理论

为了确定各指标对于机坪管制系统的重要 程度,需要对其进行权重计算,权重的大小代 表了指标的重要性,权重越大则表明该指标对 于机坪管制系统越重要,引发危险的可能性也 越大。计算权重的方法一般分为主观赋权、客 观赋权和组合赋权三种。其中主观赋权与客观 赋权均为单一赋权,在评估过程中会受到不同 程度的影响,得到不够准确的权重值。例如单 一使用层次分析法进行指标权重的计算,往往 会偏向于专家的主观意识,不同的专家给出的 权重可能会有相悖的结果;而熵权法作为一张 客观赋权法,其忽略了指标本身的重要程度, 往往会忽视决策者主观的意图,在单一使用的 过程中可能导致权重值与预期的结果相差甚远。 在此基础上,本文采用AHP—熵权法组合赋权 计算,可以在一定程度上避免结果有过强的主 观因素,同时也规避了结果与实际相悖的问题, 尽可能得到符合实际的权重^[17]。

2.1.1 AHP确定主观权重

层次分析法(analytic hierarchy process, AHP) 是一种用于进行决策和评估的多准则分析方法, 由美国运筹学家托马斯·萨蒂(Thomas L. Saaty) 于1970年提出。核心是将问题分解为若干个层 次和准则,通过对准则之间的比较和权重分配, 最终得出决策结果。

2.1.2 熵权法确定专家权重

熵权法是一种根据指标变异性的大小来确 定客观权重的赋权方法,用于解决具有多个指 标或准则的决策问题。该方法基于信息熵的概 念,旨在确定各个指标在决策中的权重,从而 更好地考虑各指标之间的相对重要性。主要有 以下步骤:

假设有n个待评价对象,m个指标,构造原 始数据矩阵 $A = (x_{ij})_{n \times m}$,其中 x_{ij} 代表第i个评价 对象的第j个指标的数值。

第一步,数据规范化处理

使用极差变换法,对数据进行标准化处理, 将指标的绝对值转化为相对值。

$$x^{*} = \frac{x_{ij} - \min(x_{1j}, x_{2j}, \dots, x_{nj})}{\max(x_{1j}, x_{2j}, \dots, x_{nj}) - \min(x_{1j}, x_{2j}, \dots, x_{nj})}$$
(1)

第二步,数据归一化,处理公式为

$$y_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}$$
(2)

第三步,求各指标熵值

信息熵是信息论中的一个概念,用于衡量不 确定性或混乱程度。本文用信息熵来衡量各指标 的离散程度。信息熵越大,表示指标值越分散。

$$e_{j} = \frac{1}{\ln n} \sum_{i=1}^{n} y_{ij} \ln y_{ij}, j = 1, 2, \cdots, m$$
(3)

第四步,通过熵值计算各指标权重

通常情况下,指标的权重等于其对应信息 熵占总信息熵的比例。

$$\begin{cases} d_{j} = 1 - e_{j}, j = 1\\ w_{j} = \frac{d_{j}}{\sum_{i=1}^{m} d_{j}}, j = 1 \end{cases}$$
(4)

2.1.3 AHP—熵权法组合赋权

本文分别使用层次分析法与熵权法来获得 各指标的权重值*w^a_i*和*w^b_j*,再使用乘法归一化方 法,得到指标的组合权重*w_i*。

$$w_{j} = \frac{w_{j}^{a} w_{j}^{b}}{\sum_{i=1}^{m} w_{i}^{a} w_{j}^{b}}, \ j = 1, 2, \cdots, m$$
(5)

其中: w_i为第j个指标的组合权系数,w^a_i为第j 个指标的层次分析法权重系数,w^b_i为第j个指标 的熵权法权系数。

2.1.4 确定隶属度矩阵

模糊集理论是一种用于处理模糊和不确定性 问题的数学方法,它由美国数学家洛特菲·扎德 (Lotfi A. Zadeh)提出。在模糊集理论中,可以通 过设置不同的隶属函数来灵活地划分集合,表 示不确定事物中满足模糊概念的程度,从而更 好地适应不同问题的特性和要求^[18-19]。本文采 用高斯隶属度函数表示正态分布特性:

$$y = e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
 (6)

其中: μ是函数的中心, σ表示宽度, 即专家对 评估结果的不确定程度, σ越小表明专家对评 估结果的确信度越高。

通过模糊集把机坪管制风险划分为5个风险 等级,分别表示 I 级(很低)、II 级(低)、II 级 (中等)、IV级(高)、V级(很高),量化值区间 为[0,1],对不同量化风险等级的函数中心点 $\mu = \{0, 0.25, 0.5, 0.75, 1\},则隶属度函数为$

$$\begin{cases} y_{VL}(x,\sigma) = e^{-\frac{(x-0)^2}{2\sigma^2}} \\ y_L(x,\sigma) = e^{-\frac{(x-0.25)^2}{2\sigma^2}} \\ y_M(x,\sigma) = e^{-\frac{(x-0.5)^2}{2\sigma^2}} \\ y_H(x,\sigma) = e^{-\frac{(x-0.75)^2}{2\sigma^2}} \\ y_{VH}(x,\sigma) = e^{-\frac{(x-1)^2}{2\sigma^2}} \end{cases}$$
(7)

根据式(7)构建隶属度矩阵,指标A的隶属 度矩阵为

$$H_A =$$

$$\begin{bmatrix} y_{VL}(a_1,\sigma_{a_1}) \ y_L(a_1,\sigma_{a_1}) \ y_M(a_1,\sigma_{a_1}) \ y_H(a_1,\sigma_{a_1}) \ y_{VH}(a_1,\sigma_{a_1}) \end{bmatrix} \\ y_{VL}(a_2,\sigma_{a_2}) \ y_L(a_2,\sigma_{a_2}) \ y_M(a_2,\sigma_{a_2}) \ y_H(a_2,\sigma_{a_2}) \ y_{VH}(a_2,\sigma_{a_2}) \\ y_{VL}(a_3,\sigma_{a_3}) \ y_L(a_3,\sigma_{a_3}) \ y_M(a_3,\sigma_{a_3}) \ y_H(a_3,\sigma_{a_3}) \ y_{VH}(a_3,\sigma_{a_3}) \end{bmatrix}$$

$$(8)$$

指标B、C、D的隶属度矩阵同理。

在利用隶属度进行风险评价时,应通过 专家对各个风险评价指标进行等级划分,V= {VL,L,M,H,VH},其量化值定义为V=(0.1,0.3, 0.5,0.7,0.9)。评估等级的定性描述如下:VL (很低),可以忽略,不需要采取任何措施,或 者目前风险处理措施有效,不需要采取额外的防 护措施;L(低),可轻微接受,保持控制并进行 监控,保障安全性;M(中),有条件接受,需要 进一步实施风险处置措施,以提升安全性;H (高),提高警惕,严格落实控制举措,降低风 险;VH(很高),严格控制,必须采取风险处置 措施降低风险,并准备应急方案,必须高度重 视并采取措施进行风险规避或风险转移。

2.1.5 DS证据理论合成法则

证据理论(Theory of Evidence),也称为 Dempster-Shafer理论,是一种用于处理不确定性 和推理推断的数学框架。它由阿瑟·普·邓普 斯特(Arthur P. Dempster)和格伦·肯尼斯·肯贝 尔(Glenn Shafer)于20世纪60年代提出,旨在处 理不确定性问题,特别是在推理和决策中涉及 到不完备和冲突信息的情况。将这一理论运用 在机坪管制的风险评估,可以进一步融合不同 专家的主观意识带来的偏差,使得多因素复杂 系统的风险评估更加有代表性。

合成法则作为DS证据理论数据融合的核 心,在形式上较为简洁便于运算。当证据之间 存在高度冲突时,K趋近于1,由于归一化因子 的存在,在计算的过程中极可能造成融合结果 与实际经验相悖,而这种悖论问题的存在将直 接影响证据理论在决策当中的可靠性与正确 性^[20-21]。

机坪管制系统的风险评估指标体系是依靠 模糊集建立的,风险因素较多,且依靠专家经 验进行初步分级评价,不同专家之间可能存在 着高度的证据冲突。为了建立全面且可靠的机 坪管制系统风险评估模型,规避人的不同主观 因素带来的负面影响,应针对传统 Dempster 合 成法则进行一定的改进。本文采用了基于权值 分配的 DS 合成法则,用于解决证据冲突问题。

通过引入冲突概率分配函数f(A),将多个 证据间的冲突概率依据各个命题的平均支持程 度进行加权分配,则取 $f(A) = K \cdot q(A)$,基于权 值分配的DS合成法则为

 $m(A) = \begin{cases} 0, A = \emptyset \\ \sum_{A_i \cap B_j \cap C_k \cap \dots \neq A} m_1(A_i) * m_2(B_j) * m_1(C_k) * \dots + Kq(A), A \neq \emptyset \end{cases}$ (Q)

其中: $K = \sum_{A_i \cap B_j \cap C_i \cap \dots \neq \emptyset} m_1(A_i) * m_2(B_j) * m_1(C_k) * \dots,$ $q(A) = \frac{1}{n} \sum_{1 \le i \le n} m_i(A), 且 q(A) 表示为证据对于A$ 的平均支持程度。

基于权值分配的DS合成法则对数据进行融合,可以规避证据高度冲突带来的影响,对于 判断各因素之间的关系及判断机坪管制系统运 行风险程度更加可靠。

2.2 基于DS证据论的评估模型

基于上述的理论,为机坪管制风险评估建 模的过程主要分为以下几步:

第一步,专家组打分,邀请业内专家组成 专家组对机坪管制系统各评价指标的重要程度 进行打分,打分方式采用1~9比度法,结果通 过判断矩阵来呈现,从而汇总各位专家意见。

第二步,对评价指标体系的各个指标进行 赋权,通过 AHP 与熵权法两种方法,进行 MATLAB 相关运算,分别获得各指标的权重值 $w_i^{\alpha} 和 w_i^{b}$,再通过公式(5)进行组合赋权,得到各 个指标的组合权重。

第三步,邀请专家组分别对二级指标的风 险等级进行评价,用量化值进行衡量,以及给 出对此等级的不确定度。得到专家打分结果。

第四步,采用高斯隶属度函数,对专家打 分结果进行计算处理,得到隶属度矩阵。再将 矩阵归一化,即可得到二级指标的基本概率分 配表。 第五步,基于权值分配的DS合成法则对基本概率分配表里面的数据进行融合,进行MATLAB运算即可得到用于评价分析的结果。

第六步,根据组合赋权得到的一级、二级 权重对 DS 融合的结果进行矩阵运算,可以得到 评价结果,用于风险评估分析。

根据上述简介,该模型的全部流程如图2所示。



图 2 评估模型流程图

3 实例分析

以西南某一机场机坪管制系统为例,根据 上述风险评估模型进行风险评估。

3.1 指标权重的确定

业内专家对机坪管制系统风险评估的指标重 要程度进行打分,根据比度法确定判断矩阵。例 如针对一级指标采用比度法得出的判断矩阵为

[1	3	5	3
1/3	1	3	3
1/5	1/3	1	1/3
1/3	1/3	3	1

将判断矩阵换算成各指标的比值,利用 MATLAB工具,使用层次分析法计算各指标的 权重值,再根据熵权法计算出各个指标的熵和 熵权值,通过组合公式将两者组合,得到组合 权重值。

通过层次分析法,可得一级指标的权重为:

 w_a =(0.5083,0.2653,0.0752,0.1512)。该判断矩阵的随机一致性比率 *CR*=*CI/RI*=0.07<0.1,即一致性可接受。通过熵权法,可得一级指标权重为: w_b =(0.2063,0.4131,0.1268,0.2538)。通过AHP—熵权法组合赋权,可得一级指标权重为: w_j =(0.3997,0.4177,0.0363,0.1463)。同理可得二级指标的权重,结果见表1。

表 1 机坪管制系统评价指标权重

准面	一级指标		一级指标		二级指标		
层	层次 分析法	熵权 法	组合 权重	层	层次 分析法	熵权 法	组合权 重
				A_1	0.3821	0.2671	0.3374
A	0.5083	0.2063	0.3997	A_2	0.5256	0.3065	0.5325
				A_3	0.0923	0.4264	0.1301
				B_1	0.5725	0.2161	0.4478
В	0.2653	0.4131	0.4177	B_2	0.3234	0.3236	0.3788
				B_3	0.1041	0.4603	0.1734
				C_1	0.0746	0.3855	0.1226
С	0.0752	0.1268	0.0363	C_2	0.7121	0.1497	0.4546
				C_3	0.2133	0.4648	0.4228
				D_1	0.1551	0.3682	0.1796
D	0.1512	0.2538	0.1463	D_2	0.5312	0.2881	0.4813
				D_3	0.3137	0.3437	0.3391

3.2 构造隶属度矩阵

邀请三位业内专家,按要求分别评价二级 指标的风险等级,给出各影响因素所处的风险 等级以及对此等级的不确定度,结果见表2。

表 2 三位专家打分结果

北右	北右 评价		量化值			不确定度	
1百 1小	等级	专家1	专家2	专家3	专家1	专家2	专家3
A_1	中	0.7	0.5	0.5	0.15	0.1	0.1
A_2	高	0.7	0.9	0.7	0.1	0.1	0.1
A_3	高	0.7	0.7	0.7	0.1	0.1	0.1
B_1	低	0.3	0.5	0.3	0.15	0.1	0.1
B_2	中	0.5	0.5	0.7	0.1	0.1	0.15
B_3	低	0.5	0.3	0.3	0.15	0.1	0.1
C_1	很低	0.1	0.1	0.1	0.1	0.1	0.1
C_2	高	0.7	0.7	0.7	0.1	0.15	0.1
C_3	低	0.1	0.3	0.3	0.1	0.1	0.15
D_1	很低	0.3	0.1	0.1	0.2	0.1	0.1
D_2	低	0.3	0.3	0.3	0.1	0.1	0.1
D_3	中	0.5	0.5	0.5	0.1	0.1	0.1

将表2结果代入隶属度矩阵式,将得到的矩 阵进行归一化运算后分别得到三位专家基本概率 分配表,如专家1的基本概率分配结果,见表3。

表 3 专家 1 基本概率分配

指标	很低	低	中	青	很高
A_1	0	0.0074	0.2734	0.6292	0.09
A_2	0	0	0.1316	0.8576	0.0108
A_3	0	0	0.1316	0.8576	0.0108
B_1	0.09	0.6292	0.2734	0.0074	0
B_2	0	0.0404	0.9192	0.0404	0
B_3	0.0026	0.1655	0.6638	0.1655	0.0026
<i>C</i> ₁	0.6511	0.3485	0.0004	0	0
C_2	0	0	0.1316	0.8576	0.0108
C_3	0.6511	0.3485	0.0004	0	0
D_1	0.1638	0.4890	0.3060	0.0401	0.0011
D_2	0.0108	0.8576	0.1315	0.0001	0
D_3	0	0.0404	0.9192	0.0404	0

3.3 基于D-S证据论的数据融合

利用 MATLAB 的算法进行数据融合,结果 见表4。

表 4 DS证据理论数据融合结果

指标	很低	低	中	高	很高
A_1	0	0.0226	0.7716	0.1828	0.0230
A_2	0	0	0.0653	0.7679	0.1668
A_3	0	0	0.0506	0.9455	0.0039
B_1	0.0317	0.5032	0.4501	0.0150	0
B_2	0	0.0226	0.7716	0.1828	0.023
B_3	0.007	0.6651	0.2793	0.0478	0.0008
<i>C</i> ₁	0.7198	0.2798	0.0004	0	0
C_2	0	0.0013	0.1000	0.8789	0.0198
C_3	0.2039	0.6844	0.1096	0.0021	0
D_1	0.4952	0.4038	0.0891	0.0116	0.0003
D_2	0.0040	0.9454	0.0506	0	0
D_3	0	0.0091	0.9818	0.0091	0

3.4 机坪管制系统的模糊综合评价

根据组合赋权得到二级权重,对合成后的 二级风险指标数据进行再融合,得到一级评价 结果,再通过一级权重对一级评价结果进行融 合,得到二级评价结果,可表示为

$$\begin{cases} \boldsymbol{M} = \begin{cases} \boldsymbol{m}(A_i) = \sum_{j=1}^{n_i} w_{ij} \boldsymbol{m}(A_{ij}) \\ \boldsymbol{m}(A) = \sum_{i=1}^{4} w_i \boldsymbol{m}(A_i) \end{cases}$$
(10)
$$\boldsymbol{S} = \boldsymbol{M} \boldsymbol{V}$$

其中: S为机坪管制系统风险综合评价结果; M为二级评价矩阵; V由评价集决定,取值为各风险评价等级量化值, $V = (0.1, 0.3, 0.5, 0.7, 0.9)^{T}$ 。

得到的评价结果见表5。

评价 等级	指标	很低	低	中	盲	很高
	A	0	0.0076	0.3017	0.5936	0.0971
一级评	В	0.0154	0.3492	0.5423	0.0843	0.0089
价结果	С	0.1745	0.3243	0.0918	0.4004	0.0090
	D	0.0909	0.5306	0.3733	0.0052	0.0001
二级评 价结果	U	0.0261	0.2383	0.4051	0.2877	0.0428

表 5 评价结果

则机坪管制系统风险综合评价结果S = MV = 0.5167。

根据以上研究结果,可见该机坪管制系统 风险接近于中高风险,说明一些风险因素已然 存在,应当依据评价结果进行系统化的监督管 理,加强防范。

此外,通过对各二级指标的数据整理与分 析,可以看出人员因素当中工作人员违规操作 (A₃)、设备因素中机坪保障设备完好性(B₂)、环 境因素中的航空气象条件(C₂)和管理因素中的 与各单位协调沟通配合度(D₃)的影响最为显著。 为此可从本质安全的角度出发,制定相应的风 险管控制度,规避风险的发生;同时应建立完 善的规章制度,确保机坪管制员安全上岗,操 作专业。由于机坪管制系统的复杂性,决定了 风险因素在机坪布局中难以消磨,安全事故的 发生往往伴随着多个风险因素共同作用,针对 现场具体的机坪管制系统运行流程,应依靠完 备的风险评估模型进行风险分析,进一步确定 适应于实际的风险管控措施。

4 结语

(1) 通过机坪风险的识别建立了机坪管制系

统的风险评价指标体系,采用AHP—熵权法对 机坪管制系统进行赋权,并基于模糊集理论构 建了机坪管制系统风险判断模型。

(2) 将 DS 证据理论应用于机坪管制系统风 险评价,对传统 DS 证据论合成法则进行了改 进,规避了高度冲突带来的负面影响,融合了 不同专家主观意识带来的偏差,提高了多因素 复杂系统的风险评估可靠性。具体评价了某机 场的机坪管制系统风险因素,结果表明该机坪 管制系统风险为中高等级,说明该机坪管制系 统中一些风险因素已然存在。

(3) 与其他风险评价相比,本评估模型综合 了组合赋权法与模糊集理论,并将基于权值分 配的 DS 证据论引入风险指标的数据融合过程, 最大限度地考虑了各风险指标的模糊性与不确 定性。并通过组合赋权在一定程度上避免结果 有过强的主观因素,同时也规避了结果与实际 相悖的问题,提高了评价结果的准确性与可靠 性。基于本文提出的机坪管制风险评估模型, 能够适应机坪管制系统的风险特征,因此可运 用于评价机坪管制系统的风险状况,为制定机 坪管制系统的管理措施提供理论依据。

(4)本模型对于影响因素较多的系统依旧存 在着不足之处,未来应针对 DS证据理论的合成 法则进行进一步改进,优化冲突概率的分配并 减少计算量,对于风险的评价尽可能减少人为 主观因素带来的偏差。

参考文献:

- [1] 程万臻,大型机场机坪管制安全绩效评估研究
 [D].德阳:中国民用航空飞行学院,2023.
- [2] 薛勇. 机坪管制移交的思考与启发[J]. 中国航务周 刊,2022(34):54-56.
- [3] 民航局综合司.航空器机坪管制移交工作总体方案[A].2018.
- [4] 何秋钊,汪家保,冯晓磊,等.基于安全等级的机场 安全运行综合评价指标体系构建[J].中国战略新 兴产业,2017(28):103-105.
- [5] 潘卫军,邱文彬.机场机坪安全管理研究[J].科技和 产业,2014,14(1):133-135,170.
- [6] ZHAO J, SHI L, ZHANG L. Application of improved unascertained mathematical model in security evaluation of civil airport [J]. International Journal of Systems Assurance Engineering and Management,

2017,8(3):1989-2000.

- [7] 田文杰,徐吉辉,祝娜,等.基于Z-number和改进 DS证据理论的风险评估方法[J].火力与指挥控制, 2023,48(1):43-49.
- [8] EDWARDS E Y. The determination of the shear characteristics of soil by the direct shear test[J]. Geotechnique, 1965, 15(4):249-293.
- [9] HAWKINS F H. Human factors in flight[M]. Alder Shot: Ashgate Publishing Limited, 1987: 345-368.
- [10] 王燕平. 停机坪安全风险评价与管理[D]. 南京:南 京航空航天大学,2017.
- [11] ADHIKARI S, MIRCHANDANI S. Integrating risk assessment modeling with aviation cybersecurity framework[C]//AIAA AVIATION 2020 FORUM, 2020.
- [12] 高浩然,黎泽君,郑涛.基于熵权法的机场运行安全 效益评价研究[J]. 航空计算技术,2023,53(5): 24-28.
- [13] 夏桂书.机场停机坪安全系统风险综合物元分析模型[J].中国科技论文,2013,8(11):1184-1188.
- [14] 陈明亮,张元,陈艳秋.基于安全绩效的机坪运行风 险评估研究[J]. 民航学报,2018,2(6):90-94.

- [15] WONG D K Y, PITFIELD D E, CAVES R E, et al. The development of a more risk-sensitive and flexible airport safety area strategy: part I: the development of an improved accident frequency model[J]. Safety Science, 2009, 47(7):913-924.
- [16] 李茜. 中国机场安全审计的作用、成效及发展[J]. 中国民用航空,2011(9):51-52.
- [17] 周长春,肖荃益,苏月.基于层次分析-熵权法的 GBAS全寿命周期风险评估[J]. 航空计算技术, 2022,52(1):21-25.
- [18] 张清华,王进,王国胤. 粗糙模糊集的近似表示[J]. 计算机学报,2015,38(7):1484-1496.
- [19] 杜玉琴,侯福均,于倩,等.不确定区间隶属度语言 变量及其应用[J].运筹与管理,2017,26(4):84-88.
- [20] LEFEVRE E, COLOT O, VANNOOREN-BERGHE P. Belief function combination and conflict management [J]. Information Fusion, 2002, 3 (2):149-162.
- [21] 高帆.基于DS证据理论的数据融合关键技术研究 [D]. 西安:西安理工大学,2023.

Risk assessment of apron control system based on uncertainty

Liao Yong^{*}, Huang Jie, Fang Na, Zhao Yiyang

(College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan 618300, China)

Abstract: Risk assessment of ramp control system is an important measure to improve airport safety level, but how to evaluate ramp control system scientifically and objectively and get the assessment results close to reality is a difficult problem. Based on this, a risk assessment system for apron control system was established. From the perspective of combined weight method, the evaluation weights were combined by multiplication and normalization, and DS evidence theory was introduced to evaluate the apron control system of an airport on the basis of fuzzy set theory, and a conclusion was drawn that personnel factors are the focus of the apron control system safety system. It has a good guiding significance for the safety improvement of apron control system.

Keywords: apron control; combined weight method; uncertainty; membership function; DS evidence theory; risk assessment

文章编号: 1007-1423(2024)06-0080-07 DOI: 10.396

DOI: 10.3969/j.issn.1007-1423.2024.06.015

基于CBM的共用信息系统装备维修保障体系分析与架构设计

庞军伟¹,张 鹏^{2*},侯兴明¹

(1. 航天工程大学研究生院,北京100084; 2. 北京空间信息中继传输技术研究所,北京100084)

摘要:随着信息化和数字化技术的不断发展,企业和组织之间信息共享和协作的迫切需求驱使着共用信息系统的 产生。企业和组织可以通过共用信息系统装备实现资源的共享和互通,避免信息孤岛和资源浪费,提高信息协同和决 策效率。然而环境、人为等因素会导致共用信息系统装备出现故障,使系统无法正常工作。虽然已有相应的装备维修 体系,但针对共用信息系统的装备维修方法落后,架构老旧且维修可扩展性低。为解决该问题,以基于状态的维修理 论(CBM)为基石,结合模糊理论、灰色模型以及现有共用信息系统维修体系,分析基于CBM的共用信息系统维修保障 体系所需功能和流程,构建基于CBM的全新共用信息系统装备维修架构。所提出架构优化了传统装备维修保障模式, 为维修管理工作提供了科学的参考依据。

关键词: CBM理论; 共用信息系统; 模糊理论; 灰色模型; 装备维修保障

0 引言

大数据驱动的共用信息系统能够有效满足 企业与组织之间信息共享和协作的需求,通过 共用信息系统,企业和组织之间可以实现资源 的共享与互通,避免信息孤立和资源浪费,提 高信息协同和决策效率。例如,多个企业可以 通过共用同一套信息系统来管理供应链和物流, 实现信息共享和协作,提高运营效率和质量; 多个政府部门可以共用同一套信息系统来管理 公共服务和资源,提高政府服务的效率和质量。 然而,环境及人为因素会影响共用信息系统设 备的正常使用,进而导致系统出现问题,给企 业或组织等带来巨大经济损失。目前已有相应 的装备维修体系,但现有共用信息系统的维修 体系方法落后且架构老旧,无法满足共用信息 系统装备维修需求。因此,需要基于新的技术, 对共用信息系统维修保障体系所需功能和流程 进行分析,并基于此为共用信息系统设计新的、 合理的装备维修保障体系架构。

基于状态维修^[14](condition based maintenance, CBM)是一种全新的装备维修理论,具有减少装 备故障、缩短维修时间,提高系统的可用度和 主要部件的使用寿命等优势,被广泛应用于多 个领域,包括航空机务、车辆保障、电力电子 等。此外,基于模糊理论^[5]和灰色模型^[6]的故障 诊断与预测方法也被广泛应用于多领域的装备 维修保障方案中。因此,本文以共用信息系统 为研究对象,以CBM维修理论为基石,结合模 糊理论、灰色模型以及现有维修保障体系,分 析共用信息系统维修保障业务需求,探索构建 信息系统维修保障系统 CBM 体系结构及框架, 优化维修保障模式,为维修管理工作的科学决 策提供参考依据。

1 共用信息系统装备维修需求和挑战

共用信息系统装备^[7]主要担负为各类型应 用系统的话音、视频、数据等业务的落地转接 提供地面信道链路、信号格式转换、设备电源 优质稳定供电等任务。其可分为三类:①传输

收稿日期: 2023-06-09 修稿日期: 2023-12-11

作者简介: 虎军伟(1986—),男,安徽临泉人,在读研究生,研究方向为装备保障; *通信作者:张鹏(1986—),男, 辽宁本溪人,本科,工程师,研究方向为计算机网络、信息工程,E-mail:554277650@qq.com; 侯兴明(1970—),男,山西介休人, 硕士,教授,研究方向为装备保障 设备,主要包括在网运行的各类型光端机、网管 系统及附属配套设备;②程控交换设备,主要包 括各类型程控交换机及附属配套设备;③通信电 源设备,主要包括交流供电系统、直流供电系统、 电源监控系统及附属配套设施、防雷接地系统。

1.1 共用信息系统装备维修需求

共有信息系统物理层设备维修需满足以下 要求:①维修影响域要小:共用信息系统底层 物理设备作为通信的最基础、最底层传输手段, 一方面从保障对象上讲,无论是光传输设备还 是供电系统都不是只为单一应用系统的业务提 供保障服务, 而是同时承载多种应用系统的通 信服务,另一方面从保障地域上讲,作为一栋 楼、一个营区甚至一片区域的基础通信保障节 点,特别像光纤通信系统作为目前有线通信的 基础骨干网络,是通信中的主要传输方式,在 对上述装备进行维修时要尽量减小其维修时所 造成的影响范围;②维修时效性强:光传输、 程控交换、通信电源装备从列装之日起一直处 于常态化运行工作状态,一旦上述系统发生故 障,并且一段时间内无法抢修将会导致整个单 位业务无法通信、指挥通联手段丧失、系统断 电瘫痪。因此能及时准确压缩故障点、快速准 确进行快速维修,保证光传输、程控交换、通 信电源装备维修时间,在主设备保护机制的前 提下,减少全局性中断的影响时间,对装备维 修保障工作提出了很高要求。

1.2 共用信息系统装备维修挑战

根据单位装备管理相关规定,装备维修保 障主要以计划维修为主,即责任单位应根据所 属装备和执行任务实际情况制定装备维护保养 计划并报装备主管机关备案。执行重大任务前, 责任单位应安排时间对参加任务的装备进行彻 底的检查、维护和调整,排除故障,提高系统 的稳定性和可靠性,确保参加任务的装备战技 指标满足任务要求。责任单位对有寿命期限的 设备部件、元器件要进行使用时间登记、跟踪 和寿命到期预警,及时更换到寿命器件。一般 情况下,设备部件、元器件在超过90%寿命期 或发现异常后要进行强制更换。

传统的维修虽然可以避免设备损坏带来的

问题,但依然存在着一些问题。首先,维修方 法落后,装备维修工作依据主要有依靠设备使 用时间和设备故障,不管是坏了以后再维修还 是制定维修计划都以维修保障人员的经验来判 定。对于基层单位来说,虽然每天与设备打交 道,但是一般停留于操作使用层面,在维修保 养上可开展、能开展的工作相对较少, 往往是 查缺补漏式的维修模式。其次,维修保障信息 化水平不足,目前装备都配有管理软件,可实 现对设备状态、告警信息的实时监视,但对于 装备维修保障层面讲,又存在各自独立、装备 全寿命管理信息化程度不高的情况。不能从装 备维修的角度来实现装备管理的信息化。最后, 维修停机时间长,一般组织大、中修或者设备 出现故障都是影响整个系统的正常运行,从而 带来上层用户被迫影响导致业务中断,即使具 备主备切换的能力,也存在主用装备维修时单 系统运行时间长的风险。

因此,需要结合共用信息系统装备维修要 求,分析设备维修保障功能和流程,基于新的 理论方法设计更为合理的设备维修架构,以解 决传统设备维修的缺陷,提供高效可靠的系统 设备维修保障。

2 基于CBM的共用信息系统维修保障 体系分析与架构设计

本节首先分析基于CBM的共用信息系统维 修保障体系所需功能和流程,然后据此设计基 于CBM的共用信息系统维修保障架构。

2.1 CBM模型

CBM利用传感器或者加装式检测设备监测 装备或设备的各类状态,将状态信息收集、分 类、存储并处理加工,而后根据状态信息的变 化利用各类算法实时评估装备或设备的技术状 态,预测装备或设备故障发生时机,结合维修 决策需求给出科学合理的维修建议^[8]。在确定 维修需求时,考虑装备的运行状态信息是 CBM 最突出的特点。维修决策时间的更新由新的状 态信息来驱动,根据更新的维修决策时间,通 过采取维修措施能够尽可能避免故障发生。这 样可以带来一系列效益,例如减少维修费用和 维修任务。然而,状态监控只是 CBM 的前提与 基础,因此不能简单地将CBM理解为状态监控 维修。CBM的核心在于根据获取的状态信息进 行数据处理和寿命预测等活动,其最终目标是 确定维修需求并做出维修决策。图1展示了典型 CBM维修流程图。



图 1 典型CBM 维修流程图

(1)设备状态采集:通过部署的传感器或内 部状态信息将装备的参数实时进行上报,这里 的数据采集信息要进行筛选,要结合装备使用 手册、维护保障手册以及日常装备维修保障的 工作经验,将那些会威胁装备运行状态、危及 装备寿命安全的参数进行采集,有针对性地提 高设备状态采集的有效性,同时要判定采集的 状态信息是否是在装备寿命期间内合理发生的信 息,如果不合理则属于中断类的信息,例如是外 部问题导致的设备状态信息变化,这类信息不作 为维修保障状态采集的信息,不用上报,按照通 信中断的事件进行处理即可,处理完毕设备运行 状态恢复正常;如果合理则属于性能劣化的信 息,例如装备本身硬件故障导致的设备状态信息 变化,此类状态信息为CBM采集的信息。

(2)基础数据分析:将设备状态采集的信息 在此阶段进行分析,主要目的一是符合状态信 息的准确性,通过分析多种状态信息来判定采 集的信息是否准确,可采取人工符合或多信息 验证的方式,避免出现误判的情况,提高检测 故障征兆的一致性;二是匹配维修准则,在已 知维修方式的知识库中去匹配所采用的维修手 段或维修方式,比如更换、检修、保养、清洁 等维护方法,特殊情况下出现疑难复杂的故障, 匹配不出已知维修手段,在后续维修完毕后更 新进知识库,从而不断迭代更新。

(3)状态维修决策:按照维修方式、维修类型、维修时机等进行设备维修工作,并对维修结果进行评估分析,同时对优化后的维修方案

进行更新、存档。

2.2 功能分析

共用信息系统维修保障系统对信息系统进 行在线或离线的状态监控、信息采集,从而对系 统整体健康状态进行实时评估,并判断系统的故 障机理、预测最佳状态,进而根据信息系统的维 修保障需求和健康状态,在适当的时机实施维修 保养,降低信息系统的平均故障时间。因此,须 从如图2所示的四个主要功能模块进行需求分析 以设计更合理的共用信息系统维修保障体系。



图 2 基于CBM的共用信息系统维修保障功能需求

(1)功能分析1(状态信息采集):该功能 模块主要有两种实现方式,一是依托现有共用 信息系统各类型设备的网管系统或者第三方开 发的监控系统,通过网络连接相关网管监控系 统,筛选收集与设备寿命状态相关的信息流; 二是结合实际,对没有监控手段的设备增加传 感器等设备状态采集点,将信息引接至状态采 集功能模块。系统状态信息是多层次、多方面 和动态的,特别是信息系统,具备各信息特有 的状态参数,既需要实时的状态信息,也需要 定期进行检查得到的检测信息。通过完备的检 测方法手段,实现信息系统状态的信息采集功 能,为实施CBM提供可靠的数据,是维修保障 系统构建的基础和前提条件。

(2)功能分析2(数据分析):由于采集的状态信息对于在基层装备维修保障人员直接提供的决策意义不大,单一信息来源不可能全面反

映设备真实状态,因此数据分析功能模块主要 是对采集状态信息进行分类,通过故障诊断、 状态评估、寿命预测等环节对设备真实运行状 态进行判定。由于共用信息系统具有军民融合 的特点,很多系统直接采用商用产品,因此数 据分析模块还要搜集民用相关品牌型号以及设 备使用维护手册,通过多种分析手段(如第3.2 节所述的模糊诊断)有效地对设备状态进科学分 析,查出故障点或劣化点。利用状态信息采集 得到系统的状态数据,可以定期或不定期地对 系统状态做出评估,分析系统的性能衰退趋势, 当系统出现劣化征兆时及时报警。借助各种预 测技术(如第3.2节所述的灰色预测)来预测关重 件的剩余寿命,在严重停机故障发生之前及时 实施维修。利用有效的预测功能制订并实施维 修计划,降低平均故障时间。

(3)功能分析3(決策支持):该功能模块辅助维修保障人员作出维修保障决策,可依据数据分析的结果匹配可行的维修策略,协助维修保障人员规划与解决问题,制定维修方案。决策支持还要与现有的装备管理制度、管理系统、维修器材、仪器仪表等手段进行协同,便于科学判定维修方案的可行性。维修决策的对象主要包括维修类型、维修方式、维修时机。例如,根据信息系统的状态和使用情况确定其维修类型^[9-10](大修、中修、小修)、维修范围(总体、部件)、维修时机(何时维修),并需要对维修方案进行优化。

(4)功能分析4(人机交互):该功能模块需要提供人机交互界面,实现维修保障人员对 CBM维修的可视化。

2.3 维修保障流程分析

当前共用信息系统维修保障方式主要有修复 性维修、定期维修两种。基于CBM构造新型共用 信息系统装备维修保障的主要流程如图3所示。

(1)故障申报阶段:当共用信息系统装备发 生故障或者有故障先兆时,一般归为三种维修 需求。事后维修主要是出现突发的设备损坏故 障,可能是自然环境、防雷接地、人为操作等 情况引起的装备损坏,一般突发性、偶然性较 强,必须采取紧急维修的方式恢复装备能力,全 面性较强。CBM维修作为前两者的补充,基于对 日常设备运行状态情况进行提前维修,一般灵活性、针对性较强。三种维修方式相辅相成、协同工作,每次维修完对维修效果的评估输入到CBM系统中,提升CBM决策评估的准确性。



图 3 共用信息系统维修保障流程

(2)任务响应&下达阶段:根据不同维修需 求生成对应的维修计划,并组织单位维修管理 部门和技术骨干进行评审,审核通过后下达维 修任务。

(3)维修准备阶段:目前主要有三种维修方式,现场维修由装备使用单位或者维修作业单位到装备使用现场或指定地点实施;后方维修由装备使用单位将故障装备交付维修作业单位实施;巡检巡修由维修作业单位按方案计划到装备单位实施。三种维修方式的选择主要根据装备故障影响的程度以及装备使用单位维修能力而确定,无论哪种方式都要做好人员、设备、器材、技术资料、脱密处理等准备,以及维修环境、后勤保障等条件的保障。

(4) 维修实施阶段:装备维修的过程中注意 及时跟进维修进度,保障任务、规避风险。

(5) 验收与信息完善阶段:维修结束后及时 组织修理质量的验收确认和质量归零工作,同 时将维修装备的故障现象、维修方法、维修效 果等信息反馈给CBM系统数据库。

2.4 基于CBM的共用信息系统维修保障架构

从一般化的CBM系统以及相关思路入手,结 合中心现有装备实际和技战术指标,以光传输、 程控交换、通信电源为数据采集对象,维修保障 体系可采用如图4所示结构,主要包含数据采集 层、数据处理层、功能模块层、业务应用层。



图 4 基于CBM 的共用信息系统维修保障系统架构

(1)数据采集层:数据采集层主要依托信息 系统采集终端,利用原有的传输协议,加装监 视器对信息系统基本状态信息、交互信息等进 行状态监测和采集。通过采集终端将设备板卡 信息、开关使用次数、光功率值、电池内阻、 电池电压、电池电流、服务器内存使用率、 CPU温度等不同类型数据进行分类和储存。

(2)数据处理层:主要借助采集终端的分析 处理能力,在数据传输到数据中心之前,对采集 的数据进行预处理并分析,实时显示信息系统状态,能够提供实时数据显示和历史数据查询。

(3)功能模块层:数据处理与应用层主要将 数据中心的各类信息系统数据深入加工处理并 提供应用。按照数据种类和时间进行排列并储 存,结合被监测系统的状态,进行剩余寿命预 测评估。根据各参数阈值及时预警各类预测评 估结果,为科学制定维修保障计划提供真实有 效的依据。

(4) 业务应用层:将装备状态判定、设备管

理、智能决策数据变成可视化页面,为各类保 障人员提供可视化参考,并生成图文报表,帮 助信息系统用户和保障人员实时掌握系统状态。

3 故障诊断及预测关键技术

故障诊断及预测是设备维修保障体系的基本功能,其关键技术包含基于CBM理论、基于 模糊理论的故障诊断技术和基于灰色理论的故障预测技术。

3.1 基于CBM的故障诊断技术

系统与设备的大部分故障是一个渐变的过程,并具有一定规律。这个变化过程可绘制*P-F*曲线来表征状态规律^[11-12],如图5所示。依据 *P-F*曲线,其中*O*点是系统和设备状态劣化的实际起点,称为故障萌发点;*P*点开始存在异常的状态,称为潜在故障点(potential failure),这一点可通过相应技术手段准确地检测到并预警;*F*点为装备或设备最终失效的时间点,称为功能故障点(functional failure)。从*P*点到*F*点之间的时间长度称为*P-F*间隔。



图 5 P-F间隔曲线

分析可知,为了降低系统或设备停机率、 预防功能性故障的发生,曲线的F点以前是合 适的维修时机。为了最大限度地利用装备及其 部件的有效寿命,应该选择在P点之后进行维 修。因此,合适的维修操作时机应该在P点和F 点之间确定。

3.2 基于模糊理论的故障诊断技术

在使用传统的数学工具对设备故障进行诊断时,由于故障原因繁多且相互交织和影响,因此设备故障诊断具有一定的模糊性,存在定量诊断的困难。在模糊理论中,故障原因和故障症状之间的模糊关系可以通过模糊逻辑来描

述。基于隶属函数和故障原因与故障症状之间 的模糊关系,故障原因和故障症状的识别问题 能够得到有效解决,从而解决使用传统的数学 工具进行定量诊断困难的问题。

设备故障通常具有多种不同的故障症状, 使用*n*种状态变量对这些故障症状进行描述,记 作 X_1, \dots, X_n 。根据故障症状严重程度, X_i 按表 取不同的值,形成故障症状的模糊向量X =(X_1, X_2, \dots, X_n)。故障症状空间由所有故障症状 模糊向量组成。设 Y_1, Y_2, \dots, Y_m 表示*m*种故障原 因,称 $Y = (Y_1, Y_2, \dots, Y_m)$ 为故障原因向量。故 障症状和故障原因之间的模糊诊断矩阵*R*可以 基于X和Y构造:

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{r}_{11} & \cdots & \boldsymbol{r}_{1m} \\ \cdots & \cdots & \cdots \\ \boldsymbol{r}_{n1} & \cdots & \boldsymbol{r}_{nm} \end{bmatrix}$$
(1)

其中: $r_{ij}(i \in \{1, \dots, n\}, j \in \{1, \dots, m\})$ 表示故障 症状 X_j 对于故障原因 Y_j 的模糊隶属度。对于任意 给定的故障症状模糊向量 $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbf{X}$, 通 过矩阵 \mathbf{R} 计算对应的故障原因向量 $\mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathbf{Y}$ 。其中,

$$(y_1, y_2, \dots, y_n) = (x_1, x_2, \dots, x_n) \odot \begin{bmatrix} r_{11} & \cdots & r_{1m} \\ \cdots & \cdots & \cdots \\ r_{n1} & \cdots & r_{nm} \end{bmatrix}$$
(2)

而^①为模糊算子, $M(\cdot, +)$, $M(\wedge, \vee)$ 是常用 的模糊算子模型。 $M(\cdot, +)$ 是一种加权平均型诊 断模型, 兼顾了主次因素,并且考虑了所有因 素的影响。在求得y以后,故障检查的先后顺序 可以通过比较各分量的相对大小来确定。而 $M(\wedge, \vee)$ 仅考虑主要因素,并不考虑次要因素, 是一种主因素决定型诊断模型,适用于将单一 主要症状视为综合特征的情况。

3.3 基于灰色理论的故障预测技术

灰色理论的研究对象是那些在信息上部分 已知、部分未知的小样本或者贫信息的不确定 性系统。它的主要目标是在对已知信息的生成和 开发的基础上实现有价值的信息提取,从而正确 描述和有效监控系统的运行行为和演化规律。

由于受到环境的干扰,离乱的情况经常出 现在表示系统行为特征的原始数据中。灰色数 列或灰色过程被用来表示这种离乱的数列。在 灰色理论中,对少量或不确定的原始数据序列 进行变换后,能够建立微分方程,进而建立称 为灰色模型(Grey Model)的模型,简称 GM 模 型。基于灰色模型进行故障预测如算法1所示。

算法1 基于灰色模型的故障预测
输入:
$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$$

1. $x^{(1)}(k) = \sum_{i=1}^{k} x^{0}(i);$
2. $x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\};$
/* 计算邻值生成数 $z^{(1)}(k) */$
3. $z^{(1)}(k) = 0.5x^{(1)}(k) - 0.5x^{(1)}(k-1), k \in \{2, 3, \dots, n\};$
/* 构建GM灰色微分方程 */
4. $x^{(0)}(k) + az^{(1)}(k) = b;$
/* 构成数据矩阵 B 与数据列 $Y; */$
5. $Y = \begin{bmatrix} x^{(0)}(2) \\ \vdots \\ x^{(0)}(2) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ \cdots & \cdots \\ -z^{(1)}(n) & 1 \end{bmatrix};$
6. $\hat{a} = [a, b]^{T} = (B^{T}B)^{-1}B^{T}Y;$
/* 构建白化方程 */
7. $\hat{x}^{(1)}(k+1) = \begin{bmatrix} x^{(1)}(0) - \frac{b}{a} \end{bmatrix} e^{-at} + \frac{b}{a}, k \in \{1, 2, \dots, n\};$
8. 离散模型式根据k值输入计算预测累加值 $\hat{x}^{(1)}(k);$
9. $\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k);$
10. $\epsilon^{(0)} = x^{(0)} - \hat{x}^{(0)};$
11. $E^{(0)} = \epsilon^{(0)}/x^{(0)}$

在给定故障诊断原因序列y的基础上,通过 算法1,构建故障原因预测灰色模型,对可能出 现的情况进行一定程度预测,以便于设备故障 维修和预防。

4 结语

随着新技术、新材料和新工艺的广泛应用, 装备维修保障面临着更大的挑战。共用信息系 统装备维修也不例外,其采用了计算机、大规 模集成电路和软件系统,具有高度的信息化和 智能化水平,其维修保障变得更加困难。鉴于 装备维修保障体系的内涵发生了质的变化,因 此需要设计新的、更合理的共用信息系统装备 维修保障架构。针对现有共用信息系统维修方 法老旧、框架落后及可扩展性低等问题,本文 梳理了故障诊断及预测关键技术,分析了基于 CBM的共用信息系统装备维修功能需求和流程, 设计了新的共用信息系统装备维修保障架构。 所提架构可用以优化传统共用信息系统装备维 修保障模式,其对提高共用信息系统维修保障 进程有重要实践意义。

参考文献:

- [1] 黎漫斯,陈春良,尚永爽,等.基于CBM的航空装备
 保障費用动态分析[J].火力与指挥控制,2014,39
 (10):61-65.
- [2] 张杰.基于CBM的信息设备状态检修研究与尝试[J].现代计算机(专业版),2016(4):70-74.
- [3] 马飒飒,贾希胜,夏良华.装备维修工程 CBM 综述[J].装备指挥技术学院学报,2008(2):111-116.
- [4] 张仕新, 昝翔, 李浩, 等. 状态维修理论及剩余寿命
 预测的研究现状与展望[J]. 兵工自动化, 2014, 33
 (9):15-20.
- [5] 张英芝,申桂香,吴甦,等.基于模糊理论的数控车 床故障分析[J].中国机械工程,2009,20(19): 2354-2357.

- [6] 张英芝,侯胜冬,王志琼,等.基于灰色理论的加工 中心故障分析[J].吉林大学学报(工学版),2022, 52(2):433-438.
- [7] 吴松涛, 蔺希杰, 樊俊宏, 等. 装备维修保障能力评估优化研究[J]. 军事运筹与评估, 2023, 38(1): 34-37.
- [8] 韩新平,张国海,胡国栋,等. CBM 在车辆维修中关 键技术研究[J]. 物流科技,2012,35(5):86-88.
- [9] 柳卫东.汽车制动系统 FTA 法的故障诊断研究 [D]. 西安:西北工业大学,2007.
- [10] 赵媛媛,陈妹.国外车辆故障预测与健康管理系统的最新发展[J].国外坦克,2017(11):38-41.
- [11] 李建伟.基于运行状态的水泥设备维修维护管理系 统设计[D].哈尔滨:哈尔滨工业大学,2014.
- [12] 张耀辉,王少华,韩小孩,等. 状态维修决策的研究
 现 状与展望[J]. 装甲兵工程学院学报,2013,27
 (2):6-13.

A maintenance and support mode for shared information system equipment framework based on CBM

Pang Junwei¹, Zhang Peng^{2*}, Hou Xingming¹

Graduate School of Aerospace Engineering University, Beijing 100084, China;
 Beijing Institute of Space Information Relay Transmission Technology, Beijing 100084, China)

Abstract: With the continuous development of information and digital technology, the urgent need for information sharing and collaboration between enterprises and organizations has driven the emergence of shared information systems. Enterprises and organizations can achieve resource sharing and interconnectivity through shared information system equipment, avoid information silos and resource waste, and improve information collaboration and decision-making efficiency. However, shared information systems still face some challenges. For example, the existing shared information system maintenance methods and frameworks are outdated and have low scalability. To solve these problems, a new shared information system equipment maintenance and support system framework based on CBM was constructed, which integrated fuzzy theory, grey model and the existing shared information system maintenance and support modes and provides a scientific reference for maintenance management work.

Keywords: CBM theory; shared information system; fuzzy theory; grey model; equipment maintenance and support

文章编号:1007-1423(2024)06-0087-07

DOI: 10.3969/j.issn.1007-1423.2024.06.016

知识增强大模型在信息系统故障分析中的应用研究

张海龙*,黄文锋,路 翔,张 磊

(中国电子科技集团公司第十五研究所,北京 100083)

摘要:典型的信息系统故障分析与处理任务高度依赖经验数据。将大语言模型应用于这一专业任务可以显著提高故障 分析处理的效率与准确性。但通用大模型在专业知识问答任务中存在准确性低、知识不足且陈旧的问题。为此,设计了基 于微调大模型与知识增强框架的故障分析系统。系统通过LoRA方法基于故障分析知识数据微调ChatGLM2-6B模型,通 过LangChain框架融合微调大模型与故障分析知识库。对比实验表明,该系统在故障分析主观问答任务中具有更好的性 能表现,提高了信息系统故障分析的准确性与自动化水平。

关键词: 大模型; 故障分析; 知识库; 智能运维

0 引言

典型的智能运维故障分析任务通过分析多 KPI指标及服务调用关系进行故障判定与根因搜 索,属于主动检测,是数据挖掘专业的一个分 支,具有很强的专业性。然而,众多面向用户 的信息系统都会遭遇使用过程中出现故障的情 况,需要进行被动的故障分析与处理,因此相 关集团公司都建立了产品的运维知识库,一些 公司甚至建立了专业的客户技术支持中心,出 售故障诊断与处理服务。在进行此类故障分析 与处理的过程中,基于人工经验建设的知识库 通常会发挥巨大的作用。但使用此类知识数据 需要专业的技术人员使用搜索引擎进行检索与 分析,费时费力,而且无法实现自动关联分析。 当前,随着大语言模型技术的蓬勃发展,基于 大模型进行运维知识数据的自动学习与关联分 析已经成为可能[1-2]。

通用大模型基于庞大的语料数据库训练生 成,拥有出色的语义表示和逻辑推理能力,可 以分析理解知识问答任务中复杂的内容表示与 关联关系信息,并输出接近人类水平的答案。 在将大模型部署到专业应用场景中时,由于其 具有很强的通识能力,相比专业小模型能够生 成更准确和自然的输出结果。然而,受限于训 练数据的样本格式和知识范围,大模型的性能 表现依然具有局限性,存在垂直场景落地困难、 知识更新缓慢、产生幻觉事实(Hallucinatory Fact)等问题^[3],难以满足专业领域对分析结果 专业程度的要求。比如在智能运维故障分析领 域,由于预训练时使用的样本数据与故障分析 知识问答任务需求存在显著差异,直接将通用 大模型部署应用时,分析结果往往缺乏准确性 和专业性。

为了增强大模型在专业领域知识问答任务 中的能力,各开源大模型都提供了监督微调的功 能。大模型监督微调任务对硬件资源的需求较 低,且能学习到专业领域数据集的知识与模式, 能够提升大模型在专业领域的性能表现,但会出 现"灾难性遗忘"(Catastrophic Forgetting)现 象^[4]。同时,由于训练成本高且训练时间长,

收稿日期: 2023-12-17 修稿日期: 2024-01-18

作者简介:*通信作者:张海龙(1988—),男,山东临沂人,博士,高级工程师,研究方向为数据挖掘、智能算法设计, E-mail:zhlcuc@163.com; 黄文锋(1982—),男,湖北丹江口人,博士,工程师,研究方向为大语言模型、多智能体协同; 路翔(1989—),男,山西长治人,硕士,工程师,研究方向为数据挖掘、体系对抗仿真;张磊(1982—),男,天津人,博士, 工程师,研究方向为多模态内容理解

大模型存在通识知识更新速度缓慢的问题。为 了改善这些问题,开源框架LangChain^[5]提出通 过文本向量化存储与向量语义检索,将交互输 入、领域知识库、通用大模型及外部工具连接 起来,改善大模型知识更新滞后的缺陷,同时 增强大模型对旧知识的记忆能力。

基于以上分析,为了提高信息系统故障分 析与处理的效率,改善大模型知识不足与陈旧 的问题,本文提出基于微调大模型与知识增强 框架的故障分析系统。首先通过故障分析知识 数据及指令模板进行大模型指令微调,构建故 障分析专业大模型;然后通过优化的LangChain 框架连接输入故障信息、微调大模型与故障分 析知识库,构建面向故障分析主观问答任务的 智能故障分析系统。系统支持运维人员通过输 入故障描述信息获得故障原因分析及故障解决 方案。本文使用OpsEval¹¹中提出的主观题评价 指标对模型性能进行评价。实验结果表明,智 能故障分析系统在故障分析任务中具有更好的 性能表现,能够有效解决大模型在故障分析与 诊断领域的适配难题。

1 故障分析微调大模型

ChatGPT^[6]大模型在多种自然语言处理任务 上表现优异,并出现智力涌现现象,但其是闭 源模型,难以面向专业领域进行适配。同时,国 内外涌现出一批开源大模型,比如LLama^[7]、 ChatGLM^[8]、MOSS^[9]等。其中清华大学的Chat-GLM模型专门针对中文问答和对话进行了优化, 可以在消费级的显卡上进行本地部署与微调。当 前推出的ChatGLM2-6B^[10]模型在MMLU、CEval、 GSM8K、BBH等数据集上取得了显著的性能提 升,在同尺寸开源模型中具有较强的竞争力。

实际测试发现,在进行主观故障分析问答 时,比如询问出现某类软硬件故障如何解决的 问题,通用大模型的回答结果往往在内容上不 准确,在格式上不符合要求。因此,为了提高 大模型分析处理专业故障问题的性能表现,本 文选用 ChatGLM2-6B 作为基座大模型,通过 LoRA^[11]方式在故障分析知识数据上进行微调训 练,构建面向故障分析主观问答任务的专业大 模型。

1.1 LoRA 微调大模型

随着自然语言处理技术向着大模型的方向发 展,尤其是ChatGPT出现后,模型预训练参数已 经达到千亿级别。由于训练成本高昂,传统的全 参数微调方法已经不适用于大多数垂直应用场 景。为此,研究人员提出了多种低成本微调方 法。目前主流的方法包括 Adapter-Tuning^[12]、 Prefix-Tuning^[13] Prompt-Tuning (P-Tuning)^[14] Low-Rank Adaptation(LoRA)、P-Tuning v2^[15]等。 其中LoRA 方法通过冻结预训练的模型参数,将 可训练的秩分解矩阵注入到Transform架构的每 一层,极大地减少了微调时可训练参数的数量, 是一种高效的参数微调框架。LoRA微调过程如 图1所示,其中左边部分是预训练大模型的权重 参数,输入输出维度都是d,在微调训练期间被 冻结;右边部分是LoRA 增加的可训练权重参 数,分为A、B两个低秩分解矩阵,秩为r。其 中A矩阵使用随机高斯分布进行初始化,维度 为 $r \times k$,负责将输入数据维度降至r维; B矩阵 使用0进行初始化, 维度为d×r, 负责将数据 维度升至d维。设图中预训练权重参数矩阵为 $W_0 \in \mathbb{R}^{d \times d}$,待更新的参数矩阵为 $\Delta W = BA$,其 中 $B \in \mathbb{R}^{d \times r}$, $A \in \mathbb{R}^{r \times d}$ 。则微调模型最后的输出 数据为两部分输出数据之和:



图 1 LoRA 微调示意图

LoRA 方法通过控制分解矩阵的秩r,可以 灵活调整可训练参数的数量,且并行求和的模 型结构设计即保留了原始预训练参数,也没有引 入额外的推理延迟,可以显著改善微调大模型 "灾难性遗忘"的问题,是一种高效的参数微调 方法。因此,本文使用LoRA方法基于故障分析 知识数据对ChatGLM2-6B大模型进行参数微调。

1.2 基于LORA的故障分析微调大模型

首先对故障分析知识数据进行预处理,按 照"指令-问题-答案"的指令微调样本模板构 建故障分析微调数据集。其中微调数据集的指 令部分统一设置为"请提供以下故障的解决方 案",通过指明具体任务目标提高微调大模型在 故障分析主观问答任务上的性能;问题部分是 对故障现象的具体描述;答案分为两部分,前 面是对故障原因的分析,后面是根据故障原因 提供的对应解决方案。

本文使用从互联网收集的信息系统故障处 理知识数据构建了包含610条样本的故障分析微 调数据集。在对 ChatGLM2-6B 模型进行微调训 练时,使用的 GPU为 NVIDIA GeForce RTX 3090 (24 GB),秩r设置为8,训练轮数设置为10。 微调后,LoRA 方法的可训练参数与原始 Chat-GLM2-6B 预训练参数合并生成微调大模型参数。

2 智能故障分析系统

指令微调后的 ChatGLM2-6B 模型能够输出 在内容及格式上适配故障分析主观问答任务需 求的信息,但存在"灾难性遗忘"及知识更新 缓慢的问题。为了改善这些问题,本文基于微 调大模型与 LangChain框架设计了一种外挂故障 分析知识库的智能故障分析系统。

2.1 LangChain框架

当前的通用大模型存在几个痛点^[16],一是 训练数据滞后,导致大模型不能回答最新知识 信息的相关问题;二是不能实时访问网络或者 知识数据库,无法获取专业的领域知识信息;三 是输入Token数量受限,无法接受大量的语料输 入。因此,为了帮助开发人员基于大模型快速构 建端到端的应用程序,LangChain框架被提出。

LangChain 是一个开发框架,能够将大模型 与外部知识数据进行连接,辅助大模型与外部 环境进行交互以及通过代理使用外部工具,包 含 Model I/O、 Retrieval、 Chains、 Memory、 Agents、Callbacks 六个部分。其中的 Model I/O 包括 Language models、Prompts、Output parsers 三部分,提供大模型调用接口及格式化问答模 板。Retrieval包括 Document loaders、Document transformers、Vector stores、Retrievers 四部分, 提供文档加载与切分,文本向量化存储与检索 功能。Chains是 LangChain 中为连接多个与 Language Model的交互过程而抽象的重要组件,它 可以将多个组件组合在一起以创建一个单一的、 连贯的任务。Memory可以帮助大语言模型补充 历史信息的上下文,包括提供历史对话记录或 向量数据库数据等。Agents为根据用户输入灵 活地调用其它工具提供了相关支持。

2.2 基于LangChain的智能故障分析系统

本文构建的智能故障分析系统架构图如图2 所示。



图 2 智能故障分析系统架构图

系统首先在故障分析知识数据集上使用 LoRA方法微调通用ChatGLM2-6B模型。并使用 RecursiveCharacterTextSplitter及text2vec工具将 故障分析知识库文档切分成更小的文本块并向量 化编码后,存储到FAISS向量数据库中。然后, 针对用户提出的问题,系统将其向量化后在向量 存储数据库中进行相似文本块检索。最后,系统 将检索到的文本块拼接后作为提示信息,与用户 的原问题一同输入微调ChatGLM2-6B模型,生成 最终的大模型故障分析结果。

3 实验对比与分析

在本文的故障分析主观问答场景中,微调 样本数据的输入问题信息由指令、故障描述两 部分构成,输出答案信息由故障分析、解决方 案两部分构成。其中指令部分固定为"请提供 以下故障的解决方案";故障描述部分简要描述 信息系统出现的故障现象;故障分析部分对故 障现象的可能原因进行分析描述;解决方案部 分根据故障分析的结果提供相应的解决办法。 除了指令部分,其它部分全部从由运维人员编 写的运维知识库文档中抽取而来。

在 OpsEval^[1]中,专家发现使用 BLEU 和 ROUGE 等传统指标对主观问答类任务进行评测 的准确性较低,因此专门面向智能运维主观问 答场景设计了基于人工打分的专家评估(Expert Evaluation)方法。专家评估指标要求专业人员基 于三个标准在(0,3)的范围内对输出信息进行打 分,分数越高,性能越好。打分的第一个标准是 流畅性(Fluency),用于评估输出信息的语言流畅 性以及回答格式是否符合主观问题的要求;第二 个标准是准确性(Accuracy),用于评估模型输出 的精确性和正确性,包括是否充分覆盖了真实答 案的关键点;第三个标准是证据量(Evidence), 用于检查模型的输出是否包含足够的论证和证 据支持,以确保答案的可信度和可靠性。本文 使用 BLEU、ROUGE 以及专家评估三个指标对 原始大模型、微调大模型以及故障分析系统进 行对比评价,分数越高,模型性能越好。

首先从本文构建的故障分析微调数据集中 随机选择100条样本数据对模型进行打分评价, 结果见表1。

表1中原始ChatGLM2-6B表示清华大学发 布的原始版本,微调ChatGLM2-6B表示使用故 障分析知识库微调后的版本,智能故障分析系 统表示本文基于微调ChatGLM2-6B及LangChain 框架构建的故障分析系统。从表1可以看出,在 BLEU、ROUGE及专家评估三个指标上,智能 故障分析系统的得分最高,微调ChatGLM2-6B 的得分次之,原始ChatGLM2-6B的得分最低。 这表明,经过LoRA微调后,微调ChatGLM2-6B 模型能够输出更符合标准答案的分析结果。而 智能故障分析系统通过集成微调大模型与专业 知识库,能够生成最符合标准答案的分析结果。

本文在表2中展示了三个典型的故障分析样 本数据在不同模型上的故障分析结果。

表2表明,相比于原始ChatGLM2-6B,微调 ChatGLM2-6B模型首先能够按照微调数据集的 模板格式,将分析结果分成"故障分析"与

模型	DIEU	DOLLCE	专家评估			
	DLEU	RUUGE	流畅性	准确性	证据量	总分
原始ChatGLM2-6B	2.37	3.17	1.06	0.74	0.68	2.48
微调 ChatGLM2-6B	3.85	5.41	1.74	1.58	1.39	4.71
智能故障分析系统	4.71	7.83	2.31	2.44	1.92	6.67

表 1 测试集数据在 BLEU、ROUGE 以及专家评估三个指标上的得分

表 2 典型故障分析样本的分析结果

主观问题1 公司内部邮件系统的条个外接系统加在线编辑工具 公文演转系统 个人办公真面等电视无法访问的问题
工机的检1 公司将前两门不死的岁于开放不死知住我编辑工经、公关机径不死、十八万公未回守田况无法仍当的问题
 第決方案: 1.重新启动相关系统:尝试重新启动在线编辑工具、公文流转系统、个人办公桌面等系统,这可能会网络连接的问题; 2.确认网络连接:确认公司内部网络连接是否正常,可以尝试打开一个网页来检查网络连接是否畅通 3.检查网络连接:如果网络连接正常,请尝试打开相关的系统程序,如果仍然无法访问,可以尝试即 技术支持来解决问题; 4.重新安装软件:对于一些软件问题,可以尝试重新安装软件并重新启动系统; 5.修改系统设置:对于一些系统设置问题,可以尝试修改系统设置并重新启动系统; 6.升级系统:如果以上措施都没有解决问题,可以尝试升级相关系统或更换设备,以解决网络连接问
请注意,这些解决方案可能不适用于所有情况,具体问题需要具体分析。如果以上措施无法解决问题 议联系技术支持或设备制造商具或进一步的帮助

续表2	
微调ChatGLM2-6B	故障分析: 1. 网络问题; 2. 外接系统问题; 3. 邮件系统配置问题; 4. 权限问题。 解决方案: 1. 检查外接系统与邮件服务器之间的网络连通性,确保网络设备及防火墙的配置正确; 2. 检查外接系统的服务器、数据库、API接口等是否正常运行; 3. 检查邮件系统的配置参数是否变更,与外部系统的接口集成是否正常; 4. 检查用户在邮件系统和外接系统中的权限设置是否正确。如有需要,调整用户权限。
智能故障分析系统	故障分析: 系统访问权限问题; 2. 网络问题; 3. 邮件系统接口集成问题; 4. 外接系统服务问题。 解决方案: 外接系统升级时使用了非标准访问接口,导致鉴权服务组件运行异常,外接系统接口访问失败,需要重启外接系统鉴权服务组件; 检查网络配置是否正常,确保防火墙配置正确,网络连接稳定; 确保邮件系统服务调用接口状态正常,网络地址、端口号等参数设置正确; 检查外接系统服务调用接口状态是否正常,外接系统服务运行是否正常。
主观问题2	内部办公电脑挂载的存储服务器无法访问,检查发现存储服务器死机,Web管理控制台无法访问,硬盘指示灯变成黄灯并闪烁。
原始 ChatGLM2-6B	 解决方案: 1. 紧急处理:尝试重启存储服务器; 2. 检查存储服务器连接:检查存储服务器是否连接到内部办公电脑。如果未连接,请将其连接并尝试重启存储服务器; 3. 检查网络连接:确保内部办公电脑与存储服务器之间存在正常的网络连接。尝试重新连接网络并重启存储服务器; 4. 进入安全模式:在无法正常启动存储服务器的情况下,尝试进入安全模式; 5. 使用系统还原:如果无法进入安全模式,请尝试使用系统还原; 6. 重新挂载存储服务器:如果上述措施都不能解决问题,请尝试重新挂载存储服务器并将其连接到内部办公电脑。 如果以上步骤仍然无法解决问题,请考虑联系IT部门寻求帮助。
微调 ChatGLM2-6B	故障分析: - 网络异常; 2.存储服务器异常; 3.硬盘故障; 4.Web管理控制台异常。 解决方案: - 通过测试工具检查内部办公电脑是否可以通过网络正常连接到存储服务器; - 通过串口或USB连接,检查存储服务器是否可以正常连接并登录; - 如果硬盘指示灯变成黄灯并闪烁,请尝试重新安装硬盘并重启存储服务器,或者更换其他存储设备; - 如果网络正常,Web管理控制台无法访问,请尝试重启管理工具或者联系管理员处理。
智能故障分析系统	故障分析: 存储服务器服务异常; 2. 网络连接异常; 3. Web管理控制台服务异常; 4. 硬盘出现故障; 5. 服务器未断电的情况下拔出硬盘造成硬盘故障,存储服务器死机,硬盘黄灯闪烁。 解决方案: 通过串口或USB连接的方式检查存储服务器能否本地访问; 通过工具检测办公电脑到存储服务器的网络是否正常; 如果网络连接正常,Web管理控制台无法访问,请尝试重启Web管理控制台服务; 格式化硬盘或者更换硬盘; 服务器未断电的情况下拔出硬盘造成硬盘故障,进入BIOS中,查看硬盘状态,修复硬盘恢复数据,在raid卡配置界面选项 "PD Mgmt"界面,检查磁盘状态,此时04、05、06盘位的磁盘故障了。其他两块故障硬盘的状态显示为Rebuild状态,等待半小时后,重启系统,进入操作系统后,重新登录管理控制台,查看相应服务是否正常运行,检查硬盘故障灯是否恢复正常,此时黄灯不再闪烁,正常亮绿灯。
主观问题3	多台工作站电脑均接入到华为交换机CE6851下,其中部分工作站工作正常,部分工作站在开机后一段时间 后网络断开,光网卡接口指示灯熄灭,交换机侧接口指示灯熄灭。在交换机侧更换接口以后,也是同样现象。
原始ChatGLM2-6B	 解决方案: 1.检查网络连接:检查故障工作站与交换机之间的网络连接,包括IP地址、子网掩码、网关等是否配置正确; 2.检查工作站的网络配置:检查工作站网络配置是否正确,包括IP地址、子网掩码、网关、DNS服务器等; 3.检查交换机接口配置:检查交换机接口配置是否正确,包括VLAN、端口数量、端口模式等; 4.检查网络设备状态:检查网络设备(如路由器、防火墙等)是否处于正常工作状态,并检查其与交换机的接口配置是否正确; 5.联系网络管理员进行故障排查:若以上步骤无法解决问题,建议联系网络管理员进行故障排查,包括检查网络拓扑、设备参数配置等。

はもつ

块衣2	
微调ChatGLM2-6B	故障分析: 1. 网络配置问题; 2. 工作站问题; 3. 交换机接口配置错误; 4. 接口启用链路震荡保护功能。 解决方案: 1. 检查故障工作站与交换机之间的网络连接配置是否正确; 2. 检查工作站电脑是否存在异常网络连接,是否感染病毒、恶意软件; 3. 检查交换机接口配置是否正确,检查VLAN、端口模式配置参数; 4. 检查交换机是否启用链路震荡保护功能。
智能故障分析系统	故障分析: 1. 交换机接口启用了链路震荡保护功能; 2. 工作站问题; 3. 交换机配置问题; 4. 网络问题。 解决方案: 1. 交换机接口默认启用链路震荡保护功能,在10秒以内发生5次以上up/down后接口将会被自动shutdown。 当交换机接口状态显示为ERROR DOWN(link-flap)时,建议按照以下思路处理:①检查历史告警。 如果过去长时间内,接口存在较多的Up/Down告警,表明接口工作状态不稳定,建议排查链路两端的光 模块和光纤是否正常;②恢复接口状态。默认情况下接口ERROR DOWN后不会自动恢复,需要执行 shutdown和undo shutdown命令手动恢复。用户也可以在接口未Error-Down时,通过执行error-down auto-recovery cause link-flap interval interval-value命令配置接口ERROR DOWN (link-flap)后自动恢复; 2. 检查工作站电脑是否感染病毒、恶意软件,是否存在异常网络连接; 3. 检查交换机接口参数是否配置正确; 4. 检查工作站的网络参数配置是否正确。

"解决方案"两部分;其次,微调ChatGLM2-6B 模型能够将故障问题的可能原因进行分类,并 提供对应的解决方案,且解决方案更加具体, 覆盖一定的标准答案关键点。这说明微调Chat-GLM2-6B模型学习到了微调知识库数据集的问 答格式与内容,证明基于专业领域数据集对大 模型进行微调是有效和必要的。进一步分析, 相比于其它两个模型,智能故障分析系统生成 的分析结果在格式上更加符合应用场景需要, 在内容上更加具体且有针对性,覆盖了最多的 真实答案关键点。这些实验结果说明,本文基 于LangChain框架设计的故障分析系统可以改善 大模型在故障分析主观问答任务中知识匮乏的 缺陷,从而提高智能运维故障分析的准确性与 智能化水平。

4 结语

通用大模型在智能运维故障分析任务中存 在准确性低、知识不足且陈旧的问题。为了改 善这些问题,本文基于LoRA 微调方法及Lang-Chain集成框架构建了外挂知识库的智能故障分 析系统。实验结果表明,智能故障分析系统在 故障分析主观问答任务中具有较高的性能,有 效解决大模型面向智能运维故障分析领域的适 配难题。其中的LoRA 微调方法能够帮助大模型 学习到微调数据集的数据格式与知识内容, LangChain集成框架能够进一步改善大模型在故障分析主观问答任务中知识匮乏的缺陷。同时,本文对于通用大模型面向其它专业领域的应用落地也具有一定的指导意义。

参考文献:

- [1] LIU Y H, PEI C H, XU L L, et al. OpsEval: a comprehensive task-oriented AIOps benchmark for large language models[EB/OL]. arXiv:2310.07637,2023.
- [2] MIAO Y K, BAI Y, CHEN L, et al. An empirical study of netops capability of pre-trained large language models[EB/OL]. arXiv:2309.05557,2023.
- [3] MAYNEZ J, NARAYAN S, BOHNET B, et al. On faithfulness and factuality in abstractive summarization
 [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 2020:1906-1919.
- [4] TONEVA M, SORDONI A, COMBES R T, et al. An empirical study of example forgetting during deep neural network learning[C]//Proceedings of the 7th International Conference on Learning Representations, New Orleans, United States, 2019.
- [5] GUU K, LEE K, TUNG Z, et al. REALM: retrieval-augmented language model pre-training [C]//Proceedings of the 37th International Conference on Machine Learning, Online, 2020; 3929–3938.
- [6] OPENAI. GPT-4 technical report [EB/OL]. arXiv: 2303.08774,2023.

- [7] Meta AI. Introducing LLaMA: a foundational, 65-billion-parameter large language model [EB/OL].
 [2023-12-01]. https://ai. facebook. com/blog/largelanguage-model-llama-meta-ai.
- [8] DU Z, QIAN Y, LIU X, et al. GLM: general language model pretraining with autoregressive blank infilling[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 2022: 320-335.
- [9] Fudan University. Moss: training conversational language models from synthetic data [EB/OL]. [2023-12-01]. https://github. com/OpenLMLab/ MOSS.
- [10] ChatGLM2-6B [EB/OL]. [2023-12-01]. https://github.com/THUDM/ChatGLM2-6B.
- [11] HU E, SHEN Y L, WALLIS P, et al. LoRA: low-rank adaptation of large language models [C] //Proceedings of the 10th International Conference on Learning Representations, Online, 2022.
- [12] HOULSBY N, GIURGIU A, JASTRZEBSKI S, et al. Parameter-efficient transfer learning for NLP [C] //Proceedings of the 36th International Conference

on Machine Learning, Long Beach, United States, 2019:2790-2799.

- [13] LI X L, LIANG P. Prefix-Tuning: optimizing continuous prompts for generation [C] //Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, Online, 2021: 4582-4597.
- [14] LESTER B, AL-RFOU R, CONSTANT N. The power of scale for parameter-efficient prompt tuning [C] //Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online and Punta Cana. Dominican Republic, 2021: 3045-3059.
- [15] LIU X, JI K X, FU Y C, et al. P-Tuning v2: prompt tuning can be comparable to fine-tuning universally across scales and tasks [C] //Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 2022:61-68.
- [16] KADDOUR J, HARRIS J, MOZES M, et al. Challenges and applications of large language models [EB/ OL]. arXiv:2307.10169,2023.

Research on the application of knowledge enhanced LLM in fault analysis of information system

Zhang Hailong^{*}, Huang Wenfeng, Lu Xiang, Zhang Lei

(China Electronics Technology Group Corp 15th Research Institute, Beijing 100083, China)

Abstract: The typical fault analysis and processing tasks of information system are highly dependent on empirical data. Applying large language model(LLM) to this professional task can significantly improve the efficiency and accuracy of fault analysis and processing. However, there are issues such as low accuracy, insufficient and outdated knowledge when using LLM in the professional knowledge answering tasks. Therefore, a fault analysis system based on fine-tuning LLM and knowledge enhanced framework is designed. The system uses LoRA to fine-tune ChatGLM2-6B on the knowledge data of fault analysis, and integrates LLM with the knowledge base of fault analysis through LangChain. Comparative experiments show that the system has better performance in the subjective question-and-answer task of fault analysis, and improves the accuracy and automation level of fault analysis in information system.

Keywords: LLM; fault analysis; knowledge base; AIOps

文章编号:1007-1423(2024)06-0094-05

DOI: 10.3969/j.issn.1007-1423.2024.06.017

改进哈里斯鹰算法用于阵列天线方向图赋形

朱智坚,华 伟*

(四川大学电子信息学院,成都 610065)

摘要:由于目前群智能优化算法在解决直线阵列天线方向图的旁瓣电平抑制和零陷控制问题时容易陷入局部最优 以及收敛速度较慢的问题,提出了一种改进哈里斯鹰优化算法。该算法首先引入单鹰探索策略来扩大算法的全局搜索 范围,然后引入自适应控制理论,用以提高算法的搜索精度并加快算法的收敛。仿真结果表明,与粒子群算法、布谷 鸟搜索算法、蚁群优化算法以及标准哈里斯鹰优化算法相比,所提算法在压低直线形阵列天线方向图的旁瓣电平和控 制零陷方面更具优越性。

关键词: 阵列天线; 方向图; 零陷控制; 旁瓣抑制; 哈里斯鹰优化算法

0 引言

相比于传统的单天线系统,由于阵列天线 具有高主瓣增益、低旁瓣电平等优势,被广泛 应用于雷达系统和卫星通信等领域中^[1-3],其中 直线阵列的方向图赋形一直是人们所重点关注 的内容。通过对直线阵列单元的物理位置、馈 电幅值以及馈电相位进行控制,可以有效地在 满足主瓣宽度要求的情况下压低旁瓣电平并获 得可控零陷。因此,寻找出一种高效、稳定的 控制算法来完成对阵列天线的方向图赋形一直 是天线设计领域所工作的重心^[46]。

由于群智能优化算法具有超越传统数值 方法和解析方法的灵活性,已有不少学者将 群智能优化算法用于阵列天线的方向图优化 中。遗传算法^[7](genetic algorithm, GA)被运 用于直线阵列的余割平方波束赋形,但算法 存在收敛速度较慢的问题;Greda等^[8]将粒子 群算法(particle swarm algorithm, PSA)用于直线 形阵列天线的方向图优化,目的在于抑制最大 旁瓣电平并获得零陷控制,但算法易陷入局部 最优;Khodier^[9]将布谷鸟搜索算法(Cuckoo search algorithm, CSA)用于直线阵列天线的综 合赋形,但算法容易受到种群内参数的影响, 且稳定性较差。

哈里斯鹰优化¹⁰(Harris Hawks Optimization, HHO)算法是 2019 年算法研究者通过仔细观察 哈里斯鹰群对猎物的捕食过程后抽象出的一种 新型优化算法。该算法面对通用数学问题的求 解能力较强,但其在面对阵列天线的方向图优 化问题时容易陷入局部最优,且收敛效果并 不好,无法被应用于实际的天线设计中。针 对此现象,本文提出了一种改进哈里斯鹰算 法(Improved Harris Hawks Algorithm, IHHA), 该算法引入了单鹰探索策略以及自适应权重控 制策略以提高其全局检索能力以及寻优精度, 并应用于直线形阵列天线的方向图优化。

1 阵列天线模型

对于由*N*(*N*为偶数)个单元组成的直线阵列 天线,当天线单元沿*x*轴对称放置时,其阵因子 表达式如下:

$$S(\phi) = \sum_{n=0}^{N-1} I_n e^{j(kx_n \cos \phi + \theta_n)}$$
(1)

其中: I_n 为第n个天线单元的馈电幅值,k为波数, x_n 为第n个天线单元的坐标位置, ϕ 为扫描

收稿日期: 2024-01-06 修稿日期: 2024-02-27

作者简介:朱智坚(1999—),男,四川资阳人,硕士研究生,在读硕士研究生,研究方向为电磁场与电磁波;*通信作者: 华伟(1966—),男,河南郑州人,博士,副教授,从事领域为移动边缘计算及电磁计算,E-mail:huaw@scu.edu.cn

(7)

角度, θ_n 为第n个天线单元的馈电相位。

为了抑制直线阵列天线最大旁瓣电平,可 将适应度函数表示如下:

$$f_1(\phi) = \operatorname{Max}\left(20 \lg \left| \frac{S(\phi)}{S(\phi_M)} \right| \right)$$
(2)

其中: φ_M为主瓣方向。

为了在限制最大旁瓣电平的条件下引入零 陷控制,可建立如下适应度函数:

$$f_2(\phi) = C_1 \cdot f_1(\phi) + C_2 \cdot \sum_i 20 \lg \left| \frac{S(\phi_i)}{S(\phi_M)} \right| \quad (3)$$

等式右边第一项用于限制最大旁瓣电平, 第二项用于控制零陷点。其中, C₁和C₂分别为 权重因子, φ_i为零陷方向。

2 改进哈里斯鹰算法

2.1 单鹰探索策略

在哈里斯鹰优化算法的全局搜索阶段,某 一只鹰的位置更新策略与其他鹰的位置呈现出 较强的关联性,这就使得哈里斯鹰优化算法在 鹰数目较少的情况下有较差的寻优效果。针对 此不足之处,本文将单鹰探索策略引入哈里斯 鹰优化算法的全局搜索阶段,目的在于进一步 加强随机扰动并弱化个体之间的位置关联,使 得全局搜索阶段的寻优范围更广。在该策略中, 单一个体的搜索位置不仅仅会根据其他个体的 位置做出更新,还会由概率根据自身位置朝着 边界进行探索,更新后的寻优方法如下:

X(t + 1) =

$$\begin{aligned} \left| X_{\text{rand}}(t) - r_1 \right| X_{\text{rand}}(t) - 2r_2 X(t) \Big|, q \ge \frac{2}{3} \\ X(t) + M(t+1) \Big(B_2 - X(t) \Big), \frac{1}{2} < q < \frac{2}{3} \\ X(t) + M(t+1) \Big(X(t) - B_1 \Big), \frac{1}{3} < q \le \frac{1}{2} \\ X_{\text{best}}(t) - X_m(t) - r_3 \Big(B_1 + r_4 \big(B_2 - B_1 \big) \Big), q \le \frac{1}{3} \end{aligned}$$

$$\tag{4}$$

$$M(t+1) = \begin{cases} \frac{M(t)}{a}, 0 \le M(t) \le a\\ \frac{1-M(t)}{1-a}, a < M(t) \le 1 \end{cases}$$
(5)

其中: *t* 为当前迭代次数, *X* 为鹰的位置, *X*_{rand} 为随机选择的鹰的位置, *X*_{best} 为当前种群的最优

位置, X_m 为种群的平均位置, B_1 和 B_2 分别为搜 索空间的下限和上限, r_1 、 r_2 、 r_3 、 r_4 以及q为取 值处于(0,1)的独立随机数, M为混沌扰动因 子^[11], M(0)为取值区间为(0,1)的随机数, a为 取值为1/2的映射中点。

2.2 自适应权重控制策略

为了提升算法的寻优能力以及收敛速度, 本文将权重因子引入哈里斯鹰优化算法的四种 围捕机制中。当迭代次数较小时,鹰群将以较 大的权重因子进行位置更新,使单一个体能更 广泛地进行空间探索,有利于避免陷入局部最 优。当迭代次数较大时,鹰群将以较小的权重 因子进行位置更新,使算法的寻优变得精细。 所引入的权重因子及位置更新策略表示如下:

$$\omega = 1 + \left(\frac{t}{T_{\text{max}}}\right)^2 - \frac{2t}{T_{\text{max}}}$$
(6)

 $X'_{\text{best}} = \omega \times X_{\text{best}}$

其中:T_{max}为最大迭代次数。

2.3 IHHA 算法步骤

改进哈里斯鹰算法的主要步骤如下:

Step1 初始化哈里斯鹰群的位置以及扰动因子;

Step2 计算出适应度函数的值,找到当前种 群内最好的位置 X_{best} ;

Step3 计算出目标的逃逸能量;

Step4 在全局搜索阶段,根据公式(4)对鹰 群位置进行更新;

Step5 在局部开发阶段,根据公式(7)对鹰群位置进行更新;

Step6 如果算法满足迭代停止条件,则结束;否则返回步骤Step3继续运行。

3 仿真结果与分析

为了验证所提出的改进哈里斯鹰算法在阵 列天线旁瓣电平抑制以及零陷控制上的有效性, 本文针对不同阵元数目和不同优化目标进行了 多组实验,并将仿真结果与传统哈里斯鹰优化 算法、布谷鸟算法、粒子群算法以及蚁群算法 的结果进行比较。为了保证实验条件的一致性, 上述算法的种群规模均被设置为50,最大迭代 次数为100。完成实验所用到的环境为Inter Core i9-12900H、32 GB内存、Windows10操作 系统、MATLAB R2018b。

3.1 最大旁瓣电平抑制

对48阵元直线阵列天线的旁瓣电平采用适 应度函数式(2)进行优化,目标在于得到一组馈 电电流幅值使得最大旁瓣电平尽可能的小,实验 中阵元的馈电相位以及阵元间距均等同于均匀直 线阵列天线,馈电方式采用对称馈电(解空间的 维度设置为24)。所有算法独立运行50次,各算 法的最优解所对应的阵因子方向图如图1所示。



图 1 48 阵元副瓣电平抑制

其中,改进哈里斯鹰算法所获得的最大旁瓣 电平为-30.82 dB,哈里斯鹰算法(图中以HHA表 示)为-29.39 dB,布谷鸟算法为-28.34 dB,粒子 群算法为-23.41 dB,蚁群算法(图中以ACA表 示)为-22.02 dB。由仿真结果可以得出,改进哈 里斯鹰算法的效果最好,相比于均匀直线阵列 (图1中以UN表示)的最大旁瓣电平-13.16 dB 改善了17.66 dB,改进哈里斯鹰算法所对应的阵 因子方向图如图2所示。



图 2 48 阵元 IHHA 寻优结果

图3描述的是上述算法在迭代过程中最大旁 瓣电平的变化曲线,可以看出本文所提出的改 进哈里斯鹰算法的寻优能力以及收敛能力都具 有明显优势。



图 3 迭代曲线

各算法在50次蒙特卡罗实验中的统计结果如 图4所示,数值结果见表1。由表1可知,改进哈 里斯鹰算法得出的最大旁瓣电平的最优值为 -30.82 dB、均值为-29.77 dB,方差为0.4788 dB, 三者都是各项数据中最小的,说明在直线阵列天 线的最大旁瓣抑制问题上,相比于其他各对比算 法,改进哈里斯鹰算法的寻优效果更具优越性。



图 4 蒙特卡罗实验统计结果

表 1 50次蒙特卡罗实验统计结果 单位:dB

最大旁瓣电平	IHHA	HHA	CSA	PSA	ACA
最优值	-30.82	-29.39	-28.34	-23.41	-22.02
最差值	-28.52	-27.14	-24.49	-19.97	-18.96
平均值	-29.77	-28.38	-26.47	-21.61	-20.38
方差	0.4788	0.6216	0.9584	0.6184	1.1987

3.2 旁瓣电平抑制及零陷控制

为了在抑制旁瓣电平的同时获得指定方向 角的零陷,本节共设置了两个实验,利用改进 哈里斯鹰优化算法、哈里斯鹰优化算法、布谷 鸟搜索算法、粒子群算法以及蚁群优化算法对 直线阵列天线的方向图采用适应度函数式(3)进 行优化,其中*C*₁=1,*C*₂=4。各算法的优化对象 为阵列天线的馈电相位,实验中阵元的馈电电 流幅值以及阵元间距均等同于均匀直线阵列天 线,馈电方式均采用对称馈电。

实验一是对18阵元直线阵列天线的旁瓣电 平进行抑制并获得64°方向上的零陷,仿真结果 如图5所示,并将数据列入表2中。



图5 18阵元副瓣电平抑制与零陷控制

表 2 18 阵元零陷控制数据统计 单位:dB

指标	IHHA	HHA	PSA	CSA	ACA
最大旁瓣电平	-15.01	-14.28	-13.51	-13.73	-13.39
64°零陷深度	-140.61	-123.76	-101.96	-119.84	-98.47

其中,改进哈里斯鹰优化算法得出的最大 旁瓣电平为-15.01 dB,相比于哈里斯鹰优化算 法、粒子群算法、布谷鸟搜索算法和蚁群优化 算法分别改善了 0.73 dB、1.50 dB、1.28 dB、 1.62 dB,并在 64°方向上获得了-140.61 dB 的零 陷深度。实验二是对 36 阵元直线阵列天线的旁 瓣电平进行抑制并获得 26°方向上的零陷,仿真 结果如图 6 所示,数值结果见表 3。由数据可 知,改进哈里斯鹰算法得出的结果最优,将最 大旁瓣电平限制到-16.48 dB 的同时在 26°方向 上获得了-155.58 dB 的零陷深度。综合上述实 验的仿真结果可以看出,改进哈里斯鹰优化算 法在直线形阵列天线的最大电平抑制及零陷的 联合控制问题上,其寻优效果具有明显优势。



图 6 36 阵元副瓣电平抑制与零陷控制

表3 36 阵元零陷控制数据统计 单位:dB

指标	IHHA	HHA	PSA	CSA	ACA
最大旁瓣电平	-16.48	-15.63	-14.56	-15.29	-14.32
26°零陷深度	-155.58	-138.94	-106.84	-127.93	-104.48

4 结语

针对目前群智能优化算法在面对阵列天线 方向图优化问题时易陷入局部最优及收敛较慢 的困境,提出了一种改进的哈里斯鹰优化算法。 该算法首先将单鹰探索策略引入原哈里斯鹰算 法的全局搜索阶段,用于提高种群的多样性; 其次将自适应权重因子引入原哈里斯鹰算法的 局部开发阶段,用于动态地变换寻优步长。根 据仿真结果可知,在直线形阵列天线方向图的 最大旁瓣电平抑制以及零陷控制方面,改进哈 里斯鹰优化算法与布谷鸟搜索算法、粒子群算 法、哈里斯鹰优化算法以及蚁群优化算法相比, 无论是在寻优能力还是稳定性上都更具优势。 未来将会把改进哈里斯鹰优化算法用于优化更 复杂的阵列天线模型,例如对稀疏阵列天线的 方向图进行具备应用前景的综合优化。

参考文献:

[1] 陈安榕.小型化可重构无源波束形成网络研究
 [D].成都:电子科技大学,2022.

- [2] 梁峰,程友峰,廖成.一种基于方向图重构机制的宽
 带宽视角相控阵天线[J].电子测量技术,2022,45
 (18):161-166.
- [3] 刘书含.宽带可重构阵列天线系统研究[D].成都:
 电子科技大学,2022.
- [4] 孙斌.改进的多目标飞蛾扑火算法及其在阵列天线 综合中的应用[D]. 杭州:杭州电子科技大学,2022.
- [5] 李铭琦. 遗传算法在阵列天线方向图综合中的应用[J]. 中国新通信,2018,20(9):116-117.
- [6] 国强,李佳莹,王亚妮.改进郊狼算法用于直线阵列
 零 陷综合[J].西安电子科技大学学报,2021,48
 (6):172-178.
- [7] 丁扬,王郑杰,伍捍东,等.基于遗传算法的余割平 方波束赋形天线设计[J]. 微波学报,2020,36(S1):
 119-122.

- [8] GREDA L, WINTERSTEIN A, LEMES D L, et al. Beamsteering and beamshaping using a linear antenna array based on particle swarm optimization [J]. IEEE Access, 2019, 7:141562-141573.
- [9] KHODIER M . Comprehensive study of linear antenna array optimisation using the cuckoo search algorithm [J]. IET Microwaves, Antennas Propagation, 2019,13(9):1325-1333.
- [10] HEIDARI A A, MIRJALILI S, FARIS H, et al. Harris hawks optimization: algorithm and applications[J].
 Future Generation Computer Systems, 2019, 97: 849-872.
- [11] 单梁,强浩,李军,等.基于Tent映射的混沌优化算 法[J].控制与决策,2005,(2):179-182.

Improved Harris hawks algorithm for pattern shaping of linear antenna arrays

Zhu Zhijian, Hua Wei*

(College of Electronics Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract: In order to better solve the problem about side lobe level suppression and nulls control of linear antenna arrays, an improved Harris hawks algorithm (IHHA) is proposed. The algorithm first uses a single eagle search strategy to expand the global search scope of the algorithm, and then introduces adaptive control theory to improve the search accuracy of the algorithm and accelerate the convergence of the algorithm. The simulation results show that compared with particle swarm optimization, cuckoo search algorithm, and colony optimization algorithm and standard Harris hawks optimization algorithm, the proposed algorithm has more advantages in reducing the side lobe level and controlling nulls of the linear antenna arrays.

Keywords: antenna array; beam patterns; nulls control; sidelobe suppression; Harris hawks algorithm;

DOI: 10.3969/j.issn.1007-1423.2024.06.018

智慧城市下基于属性加密的隐私数据保护

时豪,田野*

(太原师范学院计算机科学与技术学院,晋中 030619)

摘要: 智慧城市中涉及的数据涵盖了人们的个人信息、位置数据、交通数据等敏感信息。为了满足智慧城市中的 数据共享和隐私保护需求,结合属性基可追踪加密技术,为智慧城市的隐私数据保护提供一种新的解决方案并详细介 绍了方案的构成。根据属性来细粒度地控制不同用户对数据的访问权限,有效提高隐私数据的安全性。

关键词: 隐私保护; 密文策略属性基加密; 智慧城市

0 引言

2012 年开始,我国工信部、科技部以及住 房和城乡建设部等国家部委陆续开展智慧城市 建设试点,截至 2017 年底我国已有超过 500 个 城市明确提出或正在建设智慧城市^[1]。伴随着该 理念的发展,英美等西方国家加大了智慧城市建 设的部署及建设,智慧城市已经成为现代城市发 展的重要方向,通过信息技术的应用,实现了城 市各个领域的数字化、智能化和互联互通。

随着智慧城市的快速发展,大规模的数据 收集和共享成为支撑城市运行、提供各种智能 服务的基础。大量个人隐私和敏感数据(例如个 人身份信息、位置数据、健康记录、交通数据, 监控视频等)的泄露或滥用可能导致个人隐私暴 露、身份盗用、不当监控、金融欺诈、个人名 誉损害以及其他安全风险,因此,隐私数据保 护成为一个紧迫而重要的问题。

目前,隐私保护在智慧城市领域已经引起 了广泛关注,但是大多集中在分析智慧政府应 用所面临的法律、技术以及制度困境及相对应 的政策建议上。Gürses等^[2]强调城市中的智能家 居、智慧家居、智慧交通以及智慧医疗等已构 成智慧城市环境中的一部分,它们依靠用户家 里的智能传感器和微型通信器连入互联网和物 联网。在数据隐私的安全和管理方面,智慧城 市的复杂性正在增长,因此需要直面智慧城市 建设带来的数据隐私挑战,不仅是为了建立安 全连接的需要,也为了数据管理和访问规则。 吴迪^[3]研究了边缘计算赋能智慧城市的机遇, 对相应的隐私保护困境进行了前瞻性分析,并 针对这一挑战提出政策建议和对策。张琴^[4]揭 示了智慧城市治理中个人信息安全所面临的风 险及其成因,并从立法保障、制度建设以及机 制完善等维度提出因应策略。盖宏伟等^[5]着眼 于电子政府向智慧政府的转型,分析了智慧政 府应用所面临的法律、技术以及制度困境,并 提出相对应的政策建议和消解对策。

在具体的数据隐私保护方法上,尚未有定 论。牛淑芬等^[6]提出一种基于身份的隐私保护 性广播加密算法,在传统的基于身份的广播算 法上使用拉格朗日插值实现用户身份的匿名。 冷超等^[7]基于QKD应用的基础原理,与经典数 据传输网络融合,解决了量子保密通信和经典 通信协同应用问题,实现了智慧城市更安全可 靠的数据加密传输。郭俊^[8]提出了一种基于区 块链属性加密的智慧城市通信数据安全共享方 法,将所有加密数据存储在区块共享链上,运 用多节点参与联合计算的方式,实现了对数据 的高效访问、通信数据的安全共享。属性基加

收稿日期: 2023-11-03 修稿日期: 2023-12-04

作者简介:时豪(1997—),女,山东泰安人,硕士研究生,主要研究方向为属性基加密;*通信作者:田野(1980—), 女,山西阳泉人,博士,副教授,主要研究方向为信息安全,E-mail:ty_sunshine@163.com

密技术作为一种新兴的加密方案,已经在隐私 保护领域得到广泛关注。它可以通过属性来控 制数据用户的访问权限,保护数据拥有者的隐 私。然而,该技术在智慧城市领域的应用仍处 于起步阶段,有待深入研究和探索。随着智慧 城市的发展,对属性基加密技术的需求和研究 也在不断增长。

本文结合属性基加密技术,能够对智慧城 市下的隐私数据提供细粒度的访问控制和数据 的可追溯。一方面通过不同部门的不同属性实 现不同级别访问权限,形成不同的属性集以访 问该属性集能够满足访问策略的数据,例如城 市规划部门人员在授权后可以访问城市的交通、 环境、车流轨迹等数据,公共安全部门人员在 授权后可以访问视频监控、安全监控等信息, 增强智慧城市数据的安全性和隐私性。另一方 面利用可追踪技术,在发现隐私泄漏时,可以 追踪到泄露人员,增加了数据的可追溯性和责 任可追究性。

密文策略属性基加密技术 1

密文策略属性基加密是一种基于双线性映 射的加密方案, 它将访问策略直接应用到密文 上, 根据用户的属性是否能够满足访问结构来 对用户的解密权限进行控制。每个用户都有一 组属性来描述了其身份和权限。访问策略指定 了用户必须具备的属性集合,以便能够解密和 访问相应的密文。只有具备满足数据的访问策 略要求的用户,即其属性集合与访问策略匹配, 才能成功解密密文并获取数据。利用属性基加 密能够实现隐私数据的安全性和机密性保护, 同时利用属性基加密的可追踪技术,追踪到泄 露数据人员, 增强数据安全性和责任追究能力, 为智慧城市的持续发展提供一个可行的方案。

智慧城市下基于密文策略属性加密的 2 隐私数据保护方案设计

基于密文策略属性加密系统一般由四个实 体组成:属性中心、云服务器、数据所有者和 数据用户。流程图如图1所示。每个实体在该系 统中具有不同的分工,具体如下。

(1) 属性中心。属性中心是完全可信的一个

可信认证中心,负责初始化系统公私钥、分发 确定用户的属性,为用户生成用户私钥。

(2) 云服务器。云服务器负责存储智慧城市 中产生的加密状态下的隐私数据,以及根据智 慧城市平台中生成的搜索陷门提供搜索服务, 并将结果返回给用户。

(3) 数据所有者。在智慧城市中,数据所有 者一般为边缘设备。边缘设备产生的视频监控 数据等个人隐私数据由数据所有者负责加密, 并将加密后的数据上传到云服务器中存储。

(4) 数据用户。数据用户向属性中心提交属 性集合,属性中心将生成的用户私钥发送给合 法用户。数据用户可以生成搜索陷门提交给云 服务器,接收云服务器传来的半解密的密文并 对其进行解密,获取搜索结果。



图 1 密文策略属性基加密系统流程图

2.1 系统初始化

在系统的最开始,需要进行系统的初始化, 由属性中心执行。初始化算法根据选取的一个 安全参数,生成系统的公钥和主私钥。

属性中心选取素数 p 阶的乘法循环群(G, G_r),随机选取群G的一个生成元g,选取一个 双线性映射 $e: G \times G \to G_r$, 选取哈希函数 H_1 : $\{0,1\}^* \to \{0,1\}^*, \ H_2:\{0,1\}^* \to Z_{n^\circ}$

算法随机选取 g, u, h, w, v \in G 和 α , a $\in Z_n$, 选取一个随机数k,将k保存在主私钥中。

公钥为 $pp = (G, G_{\tau}, e, g, u, h, w, v, g^{a}, e(g, g)^{\alpha})$ (1)私钥为: ms

$$sk = (\alpha, a, k) \tag{2}$$

2.2 用户私钥生成算法

私钥生成算法由属性中心和用户交互执行, 根据用户的每一个属性与一个随机值运算,形 成每个用户不同的私钥。

用户私钥生成分为以下四部分:

a) TA使用AES加密方法以主密钥中的*k*对
 *ID*加密,生成*K'* = *Enc_k*(*ID*),计算*c* = *H*₂(*K'*),
 发送给用户USER。

b) USER 随机选取一个 $r_u \in Z_q$, 对 c 进行签 名, $\xi = g^{\frac{1}{c+r_a}}$, 用户的签名验证公钥设置为 $PK_{sign} = g^{r_a}$, 将 $c = H_2(K')$ 、签名、验证公钥, 发送给 TA。

c) 属性中心收到 $c = H_2(K')$ 、签名、验证公 钥,验证签名。

如果验证成功,将H₂(enc(ID))、签名、验 证公钥进行AES加密,二次加密保证签名信息 不被篡改。

d) KeyGen (pp, msk, $H_2(enc(ID)), \phi \subseteq Z_p$)

 $\rightarrow sk_{ID,S}$

算法由属性中心运行,输入公共参数、系统主密钥和用户的ID,属性集合,生成与用户 属性相对应的用户私钥。

算法随机选取 $r, r_1, r_2, \dots, r_{|\phi|} \in Z_p$, 计算: $K = g^{\alpha t (a + H_2(K'))} \cdot w^r$, K' = Enc(ID), $K_2 = w^r$, $L = g^r$, $L' = g^{\alpha r}$, $Sign = Enc(H_2(K') ||\xi|| PK_{sign})$, 对于每个 $\tau \in \phi$, $K_{\tau,2} = g^{r_\tau}$, $K_{\tau,3} = (u^{v_\tau} h)^{r_\tau} v^{-(a + H_2(K')) \cdot r}$, 解密密钥 $sk_{ID,S}$ 为

$$\left\langle K = g^{\alpha (a + H_2(K'))} \cdot w^r, K' = Enc(ID), \right.$$

$$\left. K_2 = w^r, L = g^r, L' = g^{ar}, \left\{ K_{\tau,2} = g^{r_\tau}, \right.$$

$$\left. K_{\tau,3} = (u^{v_\tau} h)^{r_\tau} v^{-(a + H_2(K')) \cdot r} \right\}_{\tau \in \phi}, Sign$$

$$(3)$$

2.3 数据拥有者上传加密数据

数据拥有者选择文件上传时,按照要求将文件选择并上传,同时填写文件的关键字等相关信息和加密文件要求的访问策略。加密算法根据填写的 信息和用户要求的访问策略加密文件并上传。

加密算法分为两部分:第一步,使用对称 加密算法对上传的数据加密;第二步,对加密 后的数据的存储地址、关键字等信息、数据拥 有者规定的访问策略使用属性基加密算法进行 加密,生成加密索引。 具体算法如下:

a) 共享数据密文的生成

该算法选择上传文件F,首先用对称密钥ck加密共享数据文件F,得到共享数据密文 C_F = Enc(F,ck)。采用 CP-ABE 加密机制对对称密钥ck进行加密,生成密文如下:

$$C_{key} = ck \cdot e(g,g)^{\alpha s} \tag{4}$$

b) 关键字密文的生成

 $Encrypt(pk, w, (M_{l \times n}, \rho)) \to ct \qquad (5)$

算法输入一个系统公开参数pp、待加密关 键字w和一个访问结构 $(M_{i \times n}, \rho)$ 。算法随机选取 $\vec{y} = (s, y_2, \dots, y_n)^{\mathsf{T}} \in Z_p^{n \times 1}$,其中s是用于分享的 随机秘密。

算法计算内积 $\lambda_i = M_i \vec{y}$ 并得到分享向量 $\vec{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n)^{\mathrm{T}}$,其中 M_i 是矩阵M的第i行。 算法随后随机选取 $t_1, t_2, \dots, t_k \in Z_p$ 。

索引CT为

$$\left\langle \left(M_{l \times n}, \rho \right), C_{0} = g^{s}, C_{0}' = g^{as}, C_{1} = g^{H_{2}(w)} \cdot g^{s}, C_{2} = v^{s}, \\ \left\{ C_{i,1} = w^{\lambda i} v^{ti}, C_{i,2} = (u^{\rho(i)} h)^{-t_{i}}, C_{i,3} = g^{ti} \right\}_{i \in [I]} \right\rangle$$

$$(6)$$

最后,算法输出索引 $\langle C_{key}, CT \rangle$ 以加密状态 传输到智慧城市平台中。

2.4 智慧城市平台中需要数据的用户搜索 陷门生成

搜索用户输入检索关键字,后台的陷门生 成算法利用用户的私钥生成随机搜索陷门,陷 门包含关键字,发送给云服务器。由于云服务 器是半可信的,因此必须将搜索陷门加密以保 证陷门关键字的安全性:在用户生成搜索陷门 时,用户将生成一个随机值参与陷门的生成, 保证陷门在每次搜索时都是不同的,同时将随 机值保存在搜索用户中。当合法用户申请访问 云上存储的密文时,只有数据用户私钥中的属 性与密文中的访问策略相互匹配,云服务器才 能为合法用户返回半解密的密文。

该算法由DU执行。

该算法选择查询关键字q,随机选择 $\theta \in Z_p$,保证陷门的不可连接。计算

$$\begin{cases} T_{K} = K^{\theta} = g^{\alpha\theta/(a + H_{2}(K'))}W^{\theta \cdot r} \\ T_{L} = L^{\theta} = g^{r\theta} \\ T_{L'} = (L')^{\theta} = g^{ar\theta} \end{cases}$$
(7)

対于每个
$$\tau \in \phi$$
,
 $\left\{ T_{K_{r,2}} = (K_{\tau,2})^{\theta} = g^{r_{\tau} \cdot \theta}, T_{K_{r,3}} = (K_{\tau,3})^{\theta} = (u^{V_{\tau}}h)^{r_{\tau} \cdot \theta}V^{-(a+H_{2}(K')) \cdot r \cdot \theta} \right\}_{\tau \in [|\phi|]}$
(8)
最后, 陷门为
 $T_{d} = \left\langle T_{1} = g^{-H_{2}(q)} \cdot g^{\theta}, T_{2} = v^{\theta}, T_{K}, T_{L}, T_{L'}, K', \left\{ T_{K_{r,2}}, T_{K_{r,3}} \right\}_{\tau \in [|\phi|]} \right\rangle$
(9)

2.5 搜索算法

云服务器接受到搜索陷门后执行搜索,首 先执行关键字搜索,遍历所有文件,执行简单 的双线性配对即可得到符合关键字的索引的集 合。在关键字符合的情况下,继续判断该用户 的属性是否满足文件要求的访问策略,最后将 满足条件的文件进行部分解密。将检索结果传 给用户并由用户后台将名称展示到检索结果上。 用户选择一个文件进行下载。由于判断用户的 属性是否能够解密要求的文件比较费时,将搜 索步骤分为两步可以有效减少搜索所需要的时 间。云服务器进行检索并将密文发送给DU。

该算法由CS执行。CS收到DU发来的搜索 陷门Td后,

a)首先对关键字进行筛选,逐个判断所有
 索引能否使等式 e(C₁ · T₁, v) = e(g, C₂ · T₂)相
 等,如果等式成立,继续计算;

b) 如果 DU 的属性名集合满足访问策略, CS 计算得到 $I = \left\{ i \middle| \rho(i) \in \left\{ c_j \right\}_{j=1}^{|\phi|} \right\} \subset \{1, 2, \dots, l\},$ 其中: $\omega_i \in Z_p$ 满足 $\sum_{i \in I} \omega_i M_i = (1, 0, \dots, 0)_{\circ}$ 然后,执行如下搜索计算: $T = e(T_K, C_0^{K'} \cdot C_0') = e(g, g)^{c_K \cdot \theta} e(w, g)^{(a+c)sr\theta}$ (10) $D = \prod_{i \in I} \left(e(T_L^{K'} T_{L'}, C_{i,1}) e(T_{K_{r,2}}, C_{i,2}) e(T_{K_{r,3}}, C_{i,3}) \right)^{\omega_i}$ $= \prod_{i \in I} \left(e(g, w)^{(a+c)r\theta \cdot \lambda i} \right)^{\omega_i} = e(g, w)^{(a+c)rs\theta}$ (11)

将 $Z = T/D = e(g, g)^{\alpha s \theta}$ 和密文CT发送给用户。

2.6 DU 解密:DU 收到 Z后,求解出对称密钥 对文件进行解密

由数据用户 DU 执行,输入密文 CT,部分 解密密文 Z 和用户私钥 SK。DU 计算得到相应的 对称密钥 ck:

$$ck = \frac{C_{key}}{\left(Z\right)^{1/\theta}} \tag{12}$$

从而使用对称密钥ck解密得到共享数据文F。

2.7 追踪算法:Trace(pp,msk,sk_{ID,s}) →ID或者⊥

算法从 $sk_{D,s}$ 中取出K' = enc(ID),通过主私 钥msk中的k解密获得ID。

3 结语

本文通过结合属性基加密技术提出了智慧 城市环境下隐私数据保护的新方法。提出可追 踪的访问控制方案,通过属性中心分配的属性 分享数据决定访问权限,增强数据的安全性和 隐私性。引入属性基加密的可追踪技术,使得 通过泄露的私钥可以追踪到数据拥有者,增加 了数据的可追溯性和责任可追究性。并具体给 出了方案的构造,为智慧城市的隐私数据保护 提供了一种新的解决方案。

参考文献:

- [1] 黄沣爵,杨滔,张晔珵.国内外智慧城市研究热点及 趋势(2010—2019年):基于 CiteSpace 的图谱量化 分析[J].城市规划学刊,2020(2):56-63.
- [2] GÜRSES S, TRONCOSO C, DIAZ C. Engineering privacy by design [C] //2011 Paper Presented at the Fourth Conference on Computers, Privacy and Data Protection, 25-7 January 2011, Brussels, 2011.
- [3] 吴迪.边缘计算赋能智慧城市:机遇与挑战[J]. 人民论坛·学术前沿,2020(9):18-25.
- [4] 张琴.智慧城市治理中个人信息的权益解析和权利 保护[J].南京社会科学,2020(11):93-98,107.
- [5] 盖宏伟,牛朝文.智慧政务的信息安全风险及防控 策略[J].辽宁行政学院学报,2021(4):50-54.
- [6] 牛淑芬,方丽芝,宋蜜,等.智慧城市中隐私保护性 广播加密算法[J].计算机工程与科学,2022,44
 (6):1003-1012.
- [7] 冷超,杜忠岩,王题,等.量子保密通信技术及其在 智慧城市中的应用研究[J].邮电设计技术,2023
 (4):33-37.
- [8] 郭俊.基于区块链属性加密的智慧城市通信数据安全共享方法[C]//2023年第三届创新人才培养与可持续发展国际学术会议论文集.香港:香港新世纪文化出版社,2023:260-264.

文章编号:1007-1423(2024)06-0103-04

DOI: 10.3969/j.issn.1007-1423.2024.06.019

基于神经网络的OCR技术在自动阅卷系统中的应用研究

王 睿*,林 凯

(安徽商贸职业技术学院信息与人工智能学院, 芜湖 241002)

摘要:为提高学校阅卷的准确率和效率,采用OCR技术扫描试卷,识别考生的作答文字,针对传统OCR技术缺点,使用人工智能中的神经网络技术改进OCR技术,针对OCR的文本检测和文本识别两大步骤,分别使用CTPN模型进行文本检测,使用CNN、Seq2Seq、Attention模型进行文本识别,大大提高识别的准确率和效率,为下一步的自动阅卷提供支持。

关键词:神经网络; OCR技术; 自动阅卷

0 引言

在传统的考试阅卷中, 阅卷老师需要对主 观题进行手动批阅, 不仅耗费了老师大量的时 间, 还容易出现批错的现象, 或由于主观的因 素, 造成评分不标准。随着计算机技术的发展, 考试自动阅卷成为了主流趋势, 自动阅卷不仅 仅节约时间, 降低阅卷成本, 公平公正, 更能 对试卷进行统一管理和分析等统计工作。要想 自动阅卷, 首先要将纸质的试卷扫描成电子版, 然后将答案与标准答案进行对比, 判断学生答 案的正确性, 最后给出分数。其中最重要的一 环是纸质试卷学生作答情况的识别, 只有识别 正确的学生作答情况才能有效地判断作答的正 确性。

1 OCR技术介绍

OCR(optical character recognition)文字识别 是指电子设备检查纸上打印的字符,然后用字 符识别方法将形状翻译成计算机文字的过程, 即对文本资料进行扫描,然后对图像文件进行 分析处理,获取文字及版面信息的过程。OCR 的应用场景包括证件识别、车牌识别、文件文 本识别、自然场景文字识别、票据文本识别等 场景。

传统 OCR 识别采用统计模式,处理流程较 长,包括图像的预处理、二值化、连通域分析、 版面分析、行切分、字切分、单字符识别和后 处理等步骤^[1]。误识率和识别速度是衡量 OCR 技术的一个重要指标。由于每个人手写字体的 风格不同,手写字的质量也不同,采用传统的 OCR 识别误识率比较高,识别的速度比较慢。 后期采用机器学习进行 OCR 技术的改进,如可 利用 Sklearn KNN 分类器来搭建图像识别模型, 提高图像识别的准确率,实验表明能够达到 96% 的准确率。本文采用深度学习的神经网络 技术进行图像模型的改进,进一步提高模型的 准确率。

2 基于神经网络的OCR模型

2.1 神经网络概念

神经网络是一种人工智能方法,属于深度 学习,神经网络可在有限的人类协助下,帮助

· 103 ·

收稿日期: 2023-10-01 修稿日期: 2023-11-13

基金项目: 安徽省高校优秀人才支持计划重点项目(gxypZD2020056); 安徽商贸职业技术学院"三平台两基地"项目(2020ZDX05); 安徽商贸职业技术学院"双高计划"项目(2020sgxm05-4)

作者简介: *通信作者:王睿(1986—), 女, 吉林松原人, 在读博士, 副教授, 研究方向为人工智能、大数据, E-mail: 251931451@qq.com; 林凯(1996—), 男, 安徽宣城人, 硕士, 助教, 研究方向为人工智能、深度学习

计算机制定明智的决策。神经网络架构的灵感 源自人脑。人脑细胞称为神经元,构成了一个 复杂、高度互联的网络,并能互相发送电信号, 以帮助人类处理信息。同样,人工神经网络由 多个人工神经元组成,它们共同合作以解决问 题。简单神经网络架构分为三层:输入层、隐 藏层、输出层。如图1所示。信息通过输入层 进入人工神经网络。输入节点对数据进行处理、 分析或分类,然后将其继续传递到下一层,隐 藏层从输入层或其他隐藏层获取其输入,输出 层提供人工神经网络对所有数据进行处理的最 终结果。深度学习神经网络,拥有多个隐藏层, 包含数百万个链接在一起的人工神经元。层与 层之间需要包括一个非线性激活函数,需要有 一个对输入和输出都隐藏的层,还需要保持高



度的连通性,由网络的突触权重决定。

图 1 神经网络架构

主要的神经网络结构有循环神经网络 (RNN)、深度神经网络(DNN)、卷积循环神经 网络(CNN)。CRNN 是一种识别文本的方法, 该模型主要用于解决基于图像的序列识别问题, 特别是场景文字识别问题^[2]。

2.2 神经网络的OCR模型基本步骤

OCR 的实现路线,首先输入要识别的内容,然后对图像进行预处理,预处理一般包括 二值化、文字定位和倾斜校正等步骤。二值化 将三原色构成的多色图像转换成单色图像,这 就是预处理的第一步,其目的是减少冗余信息 量,加快速度^[3]。预处理之后,检测到是否是 文字,接着识别检测到的文字是什么,最后将 检测结果输出。具体如图2所示。OCR 的实现 过程中最重要的是文字检测和文本识别。



图 2 OCR基本步骤

2.2.1 文字检测技术

在进行字符识别之前,需要先确定输入图 像上文字的位置和范围,这就是文本检测的作 用^[4]。文字检测是文字识别的前提条件,先找 到文字才能识别。在文字检测上采用了 CTPN模 型(如图3),该模型的核心思想是将图片按宽度 为16像素分成很多个小格,检测每一个小格中 是否包含文本,同时预测文本的高度和宽度。最 后将多个检测结果融合,形成最终的文本框。在 CTPN模型中引入 CNN 与 LSTM 深度网络,能提 高检测的有效性。CTPN模型的优点是将文本检 测任务转化为一连串小尺度文本框的检测,引入 RNN 提升文本检测效果,边界优化 (side-refinement)提升文本框边界预测精准度。

在卷积层部分,使用了VGG16用于特征提取,在VGG的最后一个卷积层Conv5,CTPN用了3×3的卷积核来对该feature map做卷积,这个



图 3 CTPN 模型
Conv5 特征图的尺寸由输入图像来决定,而卷 积时的步长却限定为16,感受野被固定为228 个像素。卷积后的特征将送入 BLSTM 继续学 习,最后接上一层全连接层 FC 输出我们要预测 的参数: 2k 个纵向坐标 y, 2k 个分数, k个 x 的 水平偏移量。

2.2.2 文字识别技术

随着注意力机制的发展, CNN+Seq2Seq+ Attention 文字识别技术得到越来越多的人认可。 Seq2Seq模型主要包括编码器和解码器两部分, 两者都由GRU构成, GRU输入输出的结构与普 通的 RNN 相似,其核心思想与 LSTM 相似。与 LSTM相比, GRU内部少了一个"门控", 参数 比LSTM 少,但是却也能够达到与LSTM 相当 的功能。考虑到硬件的计算能力和时间成本, 因而在这里选择更加"实用"的GRU。引入 Attention 机制的 Seq2Seq 模型能够避免传统编 码器-解码器结构中所依赖的固定语义向量 c。 Attention 机制将模型的输入与输出直接建立语 义关系,通过数据训练得到不同的权重,这使 得网络能够更加广泛地选取信息,进而提升答 案生成效果。序列 $[X_1, X_2, X_3, X_4]$ 通过编码器 Encoder编码成特征向量,通过 Attention 模块生成 C_1, C_2, C_3 , 再经过解码器 Decoder 解码成序列 $[Y_1, Y_2, Y_3]$,这样就完成了序列 $[X_1, X_2, X_3, X_4]$ 到 序列 $[Y_1, Y_2, Y_3]$ 的映射。



图 4 Attention 机制

3 模型构建与训练

3.1 数据集

实验选取 MNIST 手写数据集,是美国国家 标准与技术研究院收集整理的大型手写数字数 据库,这个数据集共含 60000 个用于训练的样 本和10000个用于测试的样本,由250个不同的 人手写而成,包含了0~9共10类手写数字图 片,每张图片都做了尺寸归一化,都是28×28 大小的灰度图,每个像素的值为0~255,通道 数为1,如图5所示,但是它们并不是作为图像 文件存储的,而是28×28的二维数组中。数组 中的每个元素对应数组中的每个像素。

2222222222222222222 322 83333333333333333333 2 3 333 444444 444444 4444 5 555555 555555555555 55 5 77777777 77 7 777777 777 3999999999999999999999999999

图 5 数据集示例

3.2 模型设置

训练轮数: EPOCH = 15

batch大小: BATCH = 100

N_BATCH = int (mnist. train. num_examples / BATCH)

需要输出中间信息的间隔步数: STEP = 1000

隐层节点数目: N_HIDDEN_NODE = 300 样本输出结果如图6所示。

In [5]:	print("训练集个数为:" + str(mnist.train.images.shape[0])) print("测试集个数为:" + str(mnist.test.images.shape[0])) print("图片维度为:" + str(mnist.train.images.shape[1]))	
	训练集个数为: 55000	
	测试集个数为: 10000 图片维度为: 784	

图 6 样本输出结果

模型信息设置如图7所示。

In []	<pre>import tensorflow as tf from tensorflow.examples.tutorials.mnist import input_data</pre>
In []	<pre>old_v = tf.logging.get_verbosity() tf.logging.set_verbosity(tf.logging.ERROR)</pre>
	DATA_PATH = '/mnist/input_data' mnist = input_data.read_data_sets(DATA_PATH, one_hot=True, source_url="http://yann.lecun.com/exdb/mnist/")
	EPOCH = 15
	BATCH = 100
In []	N_BATCH = int(mnist.train.num_examples / BATCH)
	STEP = 1000
In []	N_HIDDEN_NODE = 300
	MODEL_PATH = "/single_hidden_layer_NN_model/my_model"

图 7 模型信息设置

设置好后,运行模型,经过训练模型的准确率高达0.9746,训练过程中准确率和损失率如图8所示。



图 8 准确率和损失率的变化曲线

4 结语

本文对OCR技术和应用进行了分析,特别 是OCR技术在自动阅卷系统中的应用。基于神 经网络算法和模型的OCR技术,对文字识别的 准确率更高,识别的速度更快,实现学校阅卷 的自动化,智能化,提高老师的工作效率。但 是如何让识别算法具有更高的鲁棒性仍然有待 进一步深入探讨^[5]。

参考文献:

- [1] 王日花.基于深度学习的智能OCR识别关键技术 及应用研究[J].邮电设计技术,2021(8):20-24.
- [2] 张少宇.基于人工智能机器学习的文字识别方法 研究[J].电脑编程技巧与维护,2022(9):154-156, 176.
- [3] 王栋.人工智能 OCR 技术的应用研究[J]. 电子技 术与软件工程,2022(1):122-125.
- [4] 冯亚南.基于深度学习的光学字符识别技术研究[D].南京:南京邮电大学,2020.
- [5] 邵蕾.基于深度学习的智能阅卷系统的算法设计
 [D].南京:南京邮电大学,2021.

Research on the application of OCR technology based on neural network in automatic marking system

Wang Rui^{*}, Lin Kai

(Department of Information and Artificial Intelligence, Anhui Business College, Wuhu 241002, China)

Abstract: In order to improve the accuracy and efficiency of school marking, OCR technology is used to scan test papers and identify candidates' response text. In view of the shortcomings of traditional OCR technology, neural network technology in artificial intelligence is used to improve OCR technology and target OCR text detection and text recognition. In two major steps, the CTPN model is used for text detection, and the CNN, Seq2Seq, and Attention models are used for text recognition, which greatly improves the accuracy and efficiency of recognition and provides support for the next step of automatic marking.

Keywords: neural network; OCR technology; automatic marking

DOI: 10.3969/j.issn.1007-1423.2024.06.020

基于GA-XGBoost算法的河南省粮食产量预测研究

付金鹏*,王 哲

(华北水利水电大学信息工程学院,郑州 450000)

摘要:粮食问题关乎国家命运,是国民经济发展基础中的基础。粮食产量的变化直接关系到我国的粮食安全和农业结构的优化调整。为提高河南省粮食产量预测的精度和效率,对河南省粮食产量等相关数据进行归纳分析,利用皮尔逊(Pearson)相关性影响分析确定主要影响河南省粮食产量的因素。针对XGBoost模型容易过拟合、预测不精准的问题,引入遗传算法(GA)对其学习率、树的深度等进行优化,以更准确地预测河南省粮食产量。仿真结果表明:相比于传统的XGBoost模型,GA-XGBoost模型具有更高的预测精度,RMSE仅为0.034。因此,GA-XGBoost预测模型可以对粮食产量进行更为准确的预测。

关键词:粮食产量预测;XGBoost算法;遗传算法;皮尔逊(Pearson)相关性

0 引言

"民以食为天",粮食是国家发展的战略物 资,是人民生活的必需品。河南地处中原,素 有天下粮仓的美誉,对河南省粮食产量进行预 测有助于保障粮食安全,带动农业发展。为精 准预测粮食产量,近年来,国内研究学者采用 了不同的方法来预测粮食产量,并取得了一系 列成果。邬粒等[1]利用机器学习技术预测重庆 市的粮食产量,并分析其影响因素。曾笑等^[2] 采用了NSGM(1, N)模型来预测河南省的粮食 产量,但未考虑外在影响因素对粮食产量的影 响。武慧荣等^[3]将GM(1, 1)模型与MLP神经网 络相结合,用于预测大宗货物的运输需求。虽 然将模型相结合,但预测不够精准。上述方法虽 然能够预测粮食产量,但是模型单一,对影响因 素考虑不够全面。郑佳芳等^[4]将XGBoost与LSTM 模型结合起来,针对高校招生网站流量进行预 测。发现单一XGBoost在处理复杂数据结构时有 效性不如组合模型。俞国燕等[5]通过引入局部搜 索蛇形算法(LSSA)优化XGBoost参数,进一步提 升了预测模型的精确度。康鹭莲等修的研究利用 XGBoost 分析影响粮食产量的多重因素证明

XGBoost模型稳定。徐宁等^[7]采用XGBoost对埃塞 俄比亚粮食产量进行预测,效果良好,证明其 在气候变化和农业产量预测中的应用价值。通 过不断的算法改进和参数优化,XGBoost在各个 领域的应用将更加广泛和深入^[8]。

综上所述,针对模型单一、预测精度低需 用算法改进的问题,因此本文在进行河南省粮 食产量预测时,首先构建XGBoost预测模型, 再使用遗传算法对其进行优化,从而提高河南 省粮食产量预测的准确率。

1 数据来源及影响因子确立

1.1 数据来源

依据《河南省统计年鉴》《中国农村统计年鉴》、国家统计局相关网站,选取有效灌溉面积X1(千公顷)、化肥施用量X2(万吨)、农村用 电量X3(亿千瓦小时)^[9]、农业机械总动力X4 (万千瓦)、农林牧渔劳动力X5(万人)、粮食作 物播种面积X6(千公顷)、受灾面积X7(千公 顷)共七个影响因素作为分析对象。

1.2 Pearson相关性分析确定影响因子

图1详尽地展示了七个关键农业因素之间

收稿日期: 2023-11-20 修稿日期: 2023-12-13

作者简介:*通信作者:付金鹏(1998—),男,河南漯河人,硕士研究生,研究方向为农业工程与信息技术,E-mail: 1223899215@qq.com; 王哲(2000—),女,河南商丘人,硕士研究生,研究方向为农业工程与信息技术

的皮尔逊(Pearson)相关性系数。从图1可以清 晰地观察到,有效灌溉面积、化肥施用量、农 村用电量以及农业机械总动力间具有显著的正 相关性。这表明这些因素是在相似的外部驱动 下同步增长,又或者它们中的某一因素的增加 可能推动了其他因素的增长^[10]。与此相对照, 农林牧渔劳动力和受灾面积与上述因素均表现 出了负相关性。这可能意味着当这两个因素增 长时,上述的四个因素有可能减少,又或者它 们之间存在一种稳定的相互制衡关系。此外, 粮食作物播种面积与前四个要素有着强烈的正 相关,但与受灾面积却呈现中度的负相关。综 上所述,有效灌溉面积 X1(千公顷)、化肥施用 量 X2(万吨)、农村用电量 X3(亿千瓦小时)、 农业机械总动力 X4(万千瓦)、农林牧渔劳动力 X5(万人)、粮食作物播种面积 X6(千公顷)、受 灾面积 X7(千公顷) 这七个因素之间相关性较 强,可以作为河南省粮食产量(X8)主要影响因 子,达到预测的精确标准。



图 1 影响因子相关性分析

2 GA-XGBoost河南省粮食产量预测 模型建立

2.1 XGBoost预测模型建立

采用了XGBoost算法对河南省粮食产量进行 预测。XGBoost是一个强大的梯度提升框架。以 河南省粮食产量的七个影响因子作为输入变量, 在进行建模前,首先对数据进行拆分,其中 80%用于模型的训练,而余下的20%用于模型 性能的评估。经过大量实验得到,当XGBoost的 学习器为4个时,进行11次迭代训练,每轮都 优化了前一轮的预测误差,精度提升。XGBoost 预测模型详细构建过程见表1所示。

表 1 XGBoost 预测模型构建的详细过程

	步骤	详细过程
步骤1:	模型初始化	首先选择并初始化一个XGBoost模型。此时可以设定一些初始参数,如学习率、树的深度
步骤2:	特征权重计算	使用梯度提升方法,根据前一轮迭 代的残差来计算样本的权重,使得 模型在后续迭代中更加关注难以预 测的样本
步骤3:	树构建与优化	使用步骤2中的权重数据,进行节 点分裂,递归地构建决策树。通过 正则化项和权重计算,XGBoost确 保树的构建能增加模型的准确性
步骤4:	模型加权更新	根据学习率和新构建的树预测值更 新模型权重。这一步骤使得模型在 每轮迭代中逐渐接近真实目标
步骤5:	迭代优化	重复步骤2至步骤4,直到达到设定 的迭代次数或模型性能满足既定标 准。通过多轮迭代,XGBoost不断 优化模型,增强其预测准确性

2.2 GA-XGBoost 算法设计

遗传算法是一种受自然选择和遗传学原理 启发的全局优化技术。它通过模拟生物界的遗 传和进化机制,如选择、交叉和变异,来迭代 地搜索最佳或近似最佳的解。XGBoost是一种集 成学习方法,在影响因素多的情况下面临过拟 合问题。遗传算法可以有效地用于调整XGBoost 的学习率、树的深度。遗传算法结合XGBoost预 测模型可以在确保模型性能的同时,降低过拟 合的风险,使得模型更加稳定。依靠遗传算法 的全局搜索特性,可以获得XGBoost的最优学习 器和树的深度的组合,进而提高模型的泛化能 力和准确性。

第一步,设置参数并定义编码。设置种群规 模为50、最大迭代次数为15。通常使用二进制编 码。例如max_depth范围为[1,15]时,需要4位 二进制数字来表示这15种可能的值;max_depth 为9时,其二进制表示为"1001"。

第二步,初始化种群。种群中的个体是由

XGBoost模型的超参数集合组成的染色体,染色体的长度为超参数的数量。具体地,可以选择 XGBoost的学习率、树的最大深度作为染色体,种群中的每个个体都代表了一个 XGBoost模型 的超参数组合。

第三步,定义适应度函数。选用均方误差 的倒数作为适应度函数。适应度函数公式为:

$$Fitness = \frac{1}{\sum_{i=1}^{n} (y_{\text{pre}} - \hat{y}_{\text{test}})^2}$$
(1)

第四步,轮盘赌选择法就是通过模拟赌博 的方式,从种群中选择下一代的个体。使用轮 盘赌选择方法选择两个父代个体,并根据交叉 概率进行交叉操作。采用单点交叉,即随机选 择一个交叉点,将两个父代个体的染色体在交 叉点之后交换。根据变异概率对交叉后的个体 进行变异操作。

$$P_i = \frac{Fitness_i}{\sum\limits_{i=1}^{c} Fitness_i}$$
(2)

第五步,评估种群。使用适应度函数对新生 成的个体进行评估,并将新个体加入到种群中。

第六步,判断终止条件。如果达到了最大 迭代次数或找到了满足的解,即达到了预先设 定的适应度值阈值,则终止算法。否则,回到 第四步继续进行选择、交叉和变异操作。

第七步,通过遗传算法优化后,输出适应 度最高的个体,即具有最佳超参数组合的XGBoost 模型。

3 结果与分析

收集河南省1978—2021年共44年的粮食产 量数据,其对应年份的七个影响因素利用皮尔 逊相关性分析确定相关性,将其七个影响因素 作为关键输入变量。

3.1 评价指标确定

为评估GA-XGBoost预测模型的有效性和准确性,考虑均方根误差(RMSE)、均方误差(MSE)、决定系数(R²)三个指标作为模型评价指标,具体公式为:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left| y_{\text{pre}_i} - y_{\text{test}_i} \right|^2}$$
(3)

$$MSE = \frac{1}{n} \sum_{i=1}^{m} W_i (y_{\rm pre} - \hat{y}_{\rm test})^2$$
(4)

$$R^2 = 1 - \left(\frac{SSR}{SST}\right) \tag{5}$$

式中: ypre、ytest分别为预测值和真实值。

3.2 仿真实验与分析

3.2.1 模型预测结果分析

通过XGBoost和GA-XGBoost两种方法对河 南省粮食产量进行预测。预测结果对比如图2 所示,由图2可得,GA-XGBoost预测模型在预 测精度方面明显高于单一XGBoost预测模型。



图2 预测结果对比图

3.2.2 模型误差分析

由于均方误差、均方根误差和决定系数三 种评价指标可以较好地反映预测结果的准确性, 所以输出了两种预测模型在这三个评价指标上 的对比图,如图3所示。从图3可以直观地看出 优化后的预测模型降低了误差。



图 3 评价指标对比图

通过表2也可看出优化过的河南省粮食产

量预测模型的各项误差最小,具有更高的预测 精度。

表 2 预测结果误差对比

模型	\mathbf{R}^2	MSE	RMSE
XGBoost	0.90	0.0076	0.087
GA-XGBoost	0.98	0.0011	0.034

4 结语

采用皮尔逊相关性分析影响因素,确定有 效灌溉面积、化肥施用量、农村用电量、农业 机械总动力、农林牧渔劳动力、粮食作物播种 面积、受灾面积为河南省粮食产量的主要影响因 素,并作为模型输入变量,满足预测精度要求。 选用遗传算法优化XGBoost预测模型,完成对训 练模型的训练、测试以及预测误差分析;经过结 果分析得出,利用遗传算法优化后的XGBoost模 型具有更高的预测精度。

参考文献:

- [1] 邬粒,邹黎敏,周科.基于机器学习的重庆市粮食产量预测及影响因素分析[J].中国农机化学报,2023, 44(10):185-193.
- [2] 曾笑,章礼明.基于NSGM(1,N)的河南省粮食产 量预测[J].安徽农学通报,2023,29(19):15-20.

- [3] 武慧荣,陈少阳,崔淑华.基于GM(1,1)-MLP神经 网络模型的大宗货物运输需求预测[J].公路交通 科技,2023,40(10):233-240.
- [4] 郑佳芳,游贵荣.基于XGBoost和LSTM模型的高校招生网站流量预测研究[J].重庆科技学院学报 (自然科学版),2023,25(5):45-50.
- [5] 俞国燕,左仁意,严俊,等.基于LSSA-XGBOOST 改进算法的高体鰤鱼类体质量预测模型[J].渔业 研究,2023,45(5):427-437.
- [6] 康鹭莲,易小波.2008—2021年成都市粮食产量波动及影响因素分析[J].南方农业,2023,17(17): 122-125,130.
- [7] 徐宁,李发东,张秋英,等.基于机器学习和未来气候变化模式的埃塞俄比亚粮食产量预测[J/OL]. 中国生态农业学报(中英文),1-18[2023-12-10]. http://kns.cnki.net/kcms/detail/13.1432.S.20230817. 1358.001.html.
- [8] 于涧,洪欣,于泽翔,等.基于BP神经网络的江苏省 粮食产量预测[J].沈阳师范大学学报(自然科学 版),2023,41(4):316-320.
- [9] 王镜淳,穆月英.主产区粮食生产变动及其影响因素分析:基于河南省县域数据[J].农业展望,2022, 18(3):50-57.
- [10] 杨凡雨,刘黎明,袁承程.湖南省粮食产量波动的影响因素分析及预测[J].中国农学通报,2020,36
 (29):153-160.

Grain yield prediction based on GA-XGBoost algorithm in Henan Province

Fu Jinpeng^{*}, Wang Zhe

(College of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450000, China)

Abstract: The food issue is related to the fate of the country, and the food issue is the foundation of the national economic development foundation. The change of grain output is directly related to the food security and the optimization and adjustment of agricultural structure in our country. In order to improve the accuracy and efficiency of grain production forecasting in Henan Province, relevant data such as grain production in Henan Province were summarized and analyzed, and the main factors affecting grain production in Henan Province were determined by Pearson correlation impact analysis. Aiming at the problem that XGBoost model is easy to overfit and inaccurate in prediction, genetic algorithm(GA) was introduced to optimize its learning rate and tree depth, so as to predict grain yield in Henan Province more accurately. The simulation results show that compared with the traditional XGBoost model, GA-XGBoost model has higher prediction accuracy, RMSE is only 0.034. Therefore, GA-XGBoost forecasting model can make more accurate prediction of grain yield.

Keywords: grain yield forecast; XGBoost algorithm; genetic algorithm; Pearson correlation

开发案例

文章编号:1007-1423(2024)06-0111-06

DOI: 10.3969/j.issn.1007-1423.2024.06.021

基于PyQt的双目图像采集软件设计

杨昌其,翟梓希,惠国腾*,刘晓东

(中国民用航空飞行学院空中交通管理学院, 广汉 618307)

摘要:在采集实际管制员工作场景双目图像信息的过程中,使用该软件设计对获取到的不同通道图像信息进行水 平拼接编码并为视频数据,可以使双目画面在时间维度对齐,方便数据存储、管理和使用。通过实际测试,该软件设 计在运行稳定性、图像信息压缩性能、跨平台运行能力方面表现良好。该设计可以解决实际工程及课题研究中的双目 图像数据采集问题,具有较强的通用性。

关键词: PyQt; 双目图像; 图像采集; 软件设计

0 引言

双目图像处理的相关研究是计算机视觉与 多模态数据融合的重要交叉研究领域,它是指 利用不同渠道针对目标进行图像信息采集,并 根据特定的数据合并规则对图像进行融合,可 以有效地结合不同源的目标图像特点,抑制各 个图像源的噪声,进而突出目标图像的数据特 征,甚至可以减少决策过程中的不确定性因 素^[1]。例如,红外可见光图像融合可以有效改 善图像的色彩度、对比度、纹理细节等指数, 提高图像的信息丰富程度,在立体视觉、空管 安全、行人检测、车辆识别、农业生产、安防、 军事等领域具有广泛的应用。

在获取双目图像数据时,相关研究学者通 常采用两个独立单目摄像头或单个双目摄像头 作为图像采集设备,由于设备之间的时间和拍 摄角度通常会有明显误差,所以对采集后的数 据在时间序列和图像几何形态上进行对齐具有 基础性支撑作用。图像在时间序列上对齐是指 获得在同一时间节点下不同图像采集设备所采 集的图像数据, 使之能够更加准确、真实地反 映实际拍摄时的场景信息。Li 等[2] 提出了 ALBEF (ALign the image and text representations BEfore Fusing them), 使用 image-text contrastive loss、momentum distillation 等操作来解决多模态 数据的时间序列对齐,并在处理VQA、NLVR2 等数据集上,相较于 VILLA (Vision and Language Large scale Adversarial training)模型有所提 高。李强等^[3]以Blackfin533 DSP和CPLD为基础 设计了一套包含采集、存储、处理的嵌入式双 目视觉系统,并通过改进的 Census 算法对图像 信息进行实时匹配。郑洪庆等[4]通过使用 SAA7113H视频解码模块、SAA7121H视频编码 模块,实现了系统逐路采集现场图像,并取得 了良好的时效性。

PyQt是Python编程语言实现的Qt库,具有 多个版本;Qt是一个基于C++编程语言设计的 库,可以用来创建可视化的图形用户界面 (GUI),具有强大的跨平台运行能力、丰富的功

收稿日期: 2024-02-22 修稿日期: 2024-03-05

基金项目:四川省科技计划项目(2023YFSY0025)

作者简介:杨昌其(1974—),男,贵州思南人,硕士,教授,硕士生导师,主要研究方向为空中交通安全管理、空管信息处理 等;翟梓希(1992—),女,四川成都人,硕士研究生,主要研究方向为空中交通安全管理、空管信息处理等;*通信作者:惠国腾 (1993—),男,山东菏泽人,硕士,实验师,主要研究方向为民航信息技术、民航系统建模与仿真等,E-mail:huiguoteng@sina.com; 刘晓东(1966—),男,四川广汉人,硕士,教授,硕士生导师,主要研究方向为民航信息处理、大数据分析、教育信息科学与技 术等

能组件^[5]。其源代码是开源开放的,这为软件 开发从业人员提供了极大的便利,也促使Qt应 用在各种实际学习、研究、生产场景中。

本设计针对空中交通管制工作场景,以使 用红外可见光图像双目摄像头采集实际管制员 工作场景画面信息的实际需求为出发点,设计 一种能够实现跨平台对双目摄像头的图像信息 同时进行采集录制,使之能够严格对齐图像进 行时间维度的便捷软件工具。

1 基本原理

1.1 双目摄像头原理

双目摄像头通常是指一个图像采集设备中 包含了两个摄像头,且该两个摄像头之间有固 定的距离,由于两个摄像头采集场景画面的角 度、距离会有不同而产生视差,与人类的眼睛 类似,利用该视差可以获取物体的三维立体信 息^[6]。相较于单目摄像头,其采集画面的幅度 更加宽广,采集的信息也更加丰富。相较而言, 双目摄像头的复杂程度、成本、技术难度也会 更高,例如,在采集画面信息时通过单个接口 同时获取两个摄像头的画面信息需要专用工具。

本文以实际研究为背景,需要采集管制员 工作场景的可见光图像和红外图像用于管制员 疲劳检测的研究,遂选定市场通用的红外可见 光双目活体检测摄像头为图像采集设备,其外 形和基本成像原理如图1所示。



1.2 成像原理

红外图像的基本原理是指红外相机通过接 收目标辐射出红外信号进行成像的图像采集方 式,该成像原理可以在恶劣的可见光污染情况 下对目标的轮廓细节进行良好的表达,通常具 有不易受可见光、天气等因素影响、过滤伪装 信息等优势^[7]。但也存在图像对比度低、成像 信噪比低、无法感知物体色彩等不足。可见光 图像作为日常生活中的主要图像呈现方式广为 大众接受,具有色彩度强、对比度高、纹理细 节保留清晰等优势,但也存在容易受其他可见 光源、天气状态干扰和容易被伪装信息欺骗等 劣势^[8]。使用双目摄像头进行原数据采集,可 实现数据互补,二者具备良好的互补性,在管 制员疲劳检测的研究过程中发现更加丰富的图 像信息会直接提高目标关键点位信息提取的速 度和精度。

2 软件设计

2.1 软件基本工作流程

软件的基本工作流程如图2所示。



图 2 基本工作流程示意图

主要包含了初始化、图像信息采集、双线 程管理、画面展示、画面录制等主要功能部分。 初始化的主要目的是检测设备正常工作、软件 正常启动并提供可视化的图形界面;图像信息 采集部分的主要功能是通过双目摄像头获取现场 需要采集的画面信息,单目画面分辨率为480* 640,并及时对采集的画面进行水平拼接(可见光 图像在左,红外图像在右),使其形成单幅图像。 为了提高软件使用的友好性、用户的直观感受, 特对采集的图像信息进行实时展示,并开启双线 程,其中线程1用于图形界面的用户交互和图像 信息展示,线程2用于对画面图像信息形成的视 频流数据进行及时保存,双线程运行的优势在于 可以直接提升用户交互体验、提高图像信息采集 帧率(frames per second, FPS);根据实际的画面 数据采集需求,本文对需要采集的工作场景以 480*1280分辨率双目的画面依次进行采集,为了 便于画面数据的管理和存储,遂将最终的采集结 果以.avi的视频格式呈现。

2.2 图形化界面

为了便于实际使用,对用于和用户交互的图 形化界面进行了极简设计,只保留了必要的信息 按钮,其图形化界面主窗口示意图如图3所示。



图 3 图形化界面(GUI)图

通过图3的标题栏可以看出软件的图标 (logo)、名称和版本号。在标题栏下方有两个功 能区,分别是"录制视频""播放视频",其功 能如其名称。"录制视频"区域内各按钮的功能 如下:"选择摄像头",可通过下拉列表的方式 为使用者直接展现可选择使用的摄像头名称; "启动摄像头",软件启动时默认不开启摄像头, 点击此按钮后即可开启摄像头;"开始录制", 待开启摄像头后,可点击此按钮开始视频录制; "设置保存路径",在开始录制之前设置好视频 的保存路径。"播放视频"区域内各按钮的功能 如下:"选择视频",通过此按钮选择需要播放 的视频;"播放按键",选择需要播放的视频之 后点击此按钮即可播放;"暂停按键",可实现 视频播放时的暂停播放功能,再次点击即可继 续播放;"停止按键",可实现视频的停止播放 功能,点击即可实现停止播放。图像信息展示 部分占据了图形化界面的主要区域,通过画面可 以看出左侧画面为双目摄像头的可见光通道所采 集的图像信息,右侧画面为红外通道图像信息, 经前文介绍此处展示的图像为拼接之后的单幅画 面。为了便于获取图像采集的时间信息以及帧率 信息,遂将画面的时间戳(以s为单位)信息和每秒 所采集的帧数(FPS)信息在整体画面的左上方进行 展示。例如,通过图3中的信息可以看出该画面 采集的时间为2023年11月5日18时25分35秒, 当前每秒采集的画面数量为16幅。

3 实验

3.1 实验环境

本设计通过 Python 编程语言在 IDLE 环境 下,依赖 OpenCV、Numpy、PyQt5 等库作为支 撑,编写软件工具来采集双通道的视频数据。 其软硬件环境信息见表1。

表 1 软硬件环境信息

项目	配置信息
操作系统	Windows11 22H2
CPU	Core(TM) i7-8700K
内存	16 GB
GPU	GTX 1080 Ti
开发语言	Python 3.10
依赖库	OpenCV、Numpy、PyQt5等
开发环境	PyCharm Community Edition 2023.3.2
摄像头	奥尼S500JW

为了便于横向对比,本实验以实际雷达管制席位和模拟雷达管制席位作为数据采集对象, 所采集画面的单通道分辨率为480*640,对双通 道画面进行拼接之后单幅画面分辨率为480* 1280。随后通过采集不同时长的视频数据来进 行视频帧率和视频压缩率的对比研究,以验证 该软件设计的逻辑合理性和运行稳定性。

3.2 帧率测试

帧率(FPS)是衡量视频采集的重要指标,它标志该视频的流畅度、精准度、时间延迟甚至存储和传输等基本的视频参数^[9],所以对帧率的研究是对本设计的实际效果验证的直接手段。

本次关于帧率的设计共采集了四个样本数据,整体分为画面相对运动和相对静止两个大部分,每个部分数据有两条样本数据,每条样本数据共采集了60s,数据详情见图4。由于视

频写入帧率设定为15 FPS,所以在实际采集数 据时采取步长为15的帧画面能保证每秒的时长 内都有帧画面被采集到,体现了实验的合理性、 均衡性和鲁棒性。



图 4 帧率取样值图

通过对采集到的四个样本数据(长度为60)进 行基本分析,得到四个样本的最小帧率均为10; 最大帧率为34出现在样本1的数据中;画面相对 静止场景下会高于画面相对运动场景下的帧率, 相当于平均每秒多采集1.56幅480*1280的画面。 通过帧率的标准差^[10]可以分析出帧率的波动性, 其中样本4的标准差最大,表示其帧率波动性最 大;样本2的标准差最小,表示其帧率波动性最 小,并且整体呈现出画面相对运动的场景下波动 性会高于画面相对静止场景,这一结论是与我们 日常主观感受一致,详细数据见表2。

米日	画面相	对运动	画面相对静止		
关日 -	样本1	样本2	运动 画面相对静止 样本2 样本3 样本4 10 10 10 31 32 32 15.97 18.55 17.35 17.95 17.95		
最小帧率	10	10	10	10	
最大帧率	34	31	32	32	
肺室均齿	17.02	15.97	18.55	17.35	
帜伞均值 -	16.49		17.	.95	
帖泰仁宠圣	6.23	5.50	7.06	7.16	
帜华 你准左一	5.86		7.11		

表 2 帧率测试数据

3.3 稳定性测试

在前文测试的基础之上,进一步对软件运 行的稳定性进行了测试,通过对不同场景下不 同时长的视频数据进行连续拍摄采集,观察软 件运行是否出现故障、报错以及是否可以处置 特情等。

本次测试共设计了七个视频样本数据,最 短采集9s,最长连续采集7264s(约2h)。考虑 到实际使用时的各种突发特情,本测试验证了样 本"视频5"~"视频7"数据采集时分别出现了 暴力退出、突发断电、线程中断等情境下均能正 常采集视频数据。测试的详细信息见表3。

3.4 压缩测试

由于本软件设计的基本出发点是针对实际 管制员工作场景进行画面采集,并进行疲劳检 测以及相关的分析,所以直接使用了画面数据, 而非视频数据。直接采集画面数据会占用非常 大的存储空间、大幅增加数据管理的复杂度, avi格式的视频数据具有兼容性强、占用存储空 间小、管理便捷和良好的压缩性能等优点^[9],

名称	视频时长/s	特情	视频大小/KB	总帧数/FPS	解压后图像大小/KB	平均帧率	压缩倍数
视频1	9	正常采集	1229	150	18001	16.67	14.65
视频2	7264	正常采集	1259807	112828	13077766	15.53	10.38
视频3	3606	正常采集	683520	55293	6491180	15.33	9.50
视频4	2869	正常采集	716082	46221	5166303	16.11	7.21
视频5	3768	暴力退出	951148	59570	6784598	15.81	7.13
视频6	467	突发断电	124717	7339	842449	15.72	6.75
视频7	2621	线程中断	671029	40267	4719877	15.36	7.03
总计	20604		4407531	321668	37100174	15.61	8.42

表 3 稳定性测试数据表

方便进行数据处理,所以本文最终采用此视频 格式管理图像数据。经过测试,该视频格式的 压缩倍数(率)¹⁰区间为[6.75,14.65],实际总平 均压缩倍数为8.42,实际加权压缩倍数为8.95, 如图5所示。



图 5 视频平均帧率和压缩倍数图

3.5 跨平台测试

在 Windows8、 Windows10、 Windows11、 Raspberry pi、JetsonNano等平台上进行了实际测 试,均能正常稳定运行。

4 结语

本文针对空中交通管制工作场景,以使用 红外可见光图像双目摄像头采集实际管制员工 作场景画面信息的实际需求为出发点,以 PyCharm为软件开发环境,依赖OpenCV、PyQt5、 Numpy等Python库,设计出了一种双目摄像头 的图像信息采集录制工具,并对其帧率、稳定 性、压缩能力、跨平台能力等方面进行了测试, 取得了良好的效果。该软件设计具备较强的稳 定性、实用性,可以在双目摄像头视频(画面) 采集时解决双通道图像信息时间维度(模态),对 齐后以视频数据形式保存画面信息;经测试该设 计可以在目前主流操作系统下稳定运行,实现了 跨平台使用。该设计解决了实际工程及课题研究 中遇到的双目图像数据采集问题,在类似的研究 中具有一定的推广价值和示范意义。

参考文献:

- [1] 何俊,张彩庆,李小珍,等.面向深度学习的多模态
 融合技术研究综述[J].计算机工程,2020,46(5):
 1-11.
- [2] LI J N, SELVARAJU R R, GOTMARE A D, et al. Align before fuse: vision and language representation learning with momentum distillation[EB/OL]. arXiv: 2107.07651,2021.
- [3] 李强,黎福海,陈嘉玲,等.基于DSP的双目视觉系 统[J]. 计算机系统应用,2011,20(11):20-22,49.
- [4] 郑洪庆,程蔚.一种双目图像采集系统的设计[J].
 宁德师范学院学报(自然科学版),2016,28(4):
 437-443.
- [5] 陶文玲,侯冬青. PyQt5与Qt设计师在GUI开发中的应用[J]. 湖南邮电职业技术学院学报,2020,19
 (1):19-21.
- [6] 刘然.双目视觉与障碍物探测方法研究[D].西安: 西安石油大学,2018.
- [7] LI S T, KANG X D, FANG L Y, et al. Pixel-level image fusion: a survey of the state of the art[J]. Information Fusion, 2017, 33:100-112.
- [8] MA J Y, MA Y, LI C. Infrared and visible image fusion methods and applications: a survey [J]. Information Fusion, 2019, 45: 153-178.
- [9] 李源.基于树莓派的无线视频监控系统关键技术研究[D].郑州:郑州大学,2017.
- [10] 郑翠芳. 几种常用无损数据压缩算法研究[J]. 计算 机技术与发展,2011,21(9):73-76.

Design of binocular image acquisition software based on PyQt

Yang Changqi, Zhai Zixi, Hui Guoteng^{*}, Liu Xiaodong

(College of Air Traffic Management, Civil Aviation Flight University of China, Guanghan 618307, China)

Abstract: During the process of capturing binocular image information in the actual air traffic controller working scenario, using this software design to horizontally stitch and encode the acquired images from different channels into video data enables the alignment of binocular images in the time domain, facilitating data storage, management and utilization. Through practical testing, this software design demonstrates good performance in terms of stability during operation, image compression efficiency, and cross-platform compatibility. This design can address the binocular image data acquisition needs in both practical engineering and research projects, demonstrating strong versatility.

Keywords: PyQt; binocular images; image acquisition; software design

(上接第102页)

A privacy protection scheme based on attribute encryption in smart cities

Shi Hao, Tian Ye*

(College of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, China)

Abstract: The data involved in smart cities covers sensitive information such as people's personal information, location data, and traffic data. In order to meet the needs of data sharing and privacy protection in smart cities, this paper provides a new solution for privacy data protection in smart cities with attribute-based traceable encryption technology and introduces the composition of the solution in detail. According to attributes to fine-grained control of different users to the data access rights, effectively improve the security of private data.

Keywords: privacy protection; attribute based encryption of ciphertext policy; smart city

文章编号:1007-1423(2024)06-0117-04

DOI: 10.3969/j.issn.1007-1423.2024.06.022

基于ZigBee的智慧园区LED路灯无线测控系统设计

柴西林1*, 邵照勇2

(1. 兰州石化职业技术大学电子电气工程学院, 兰州 730060; 2. 中国石油兰州石化公司, 兰州 730060)

摘要:基于ZigBee无线传感器网络,结合智慧园区的建设需求,设计实现了LED路灯光强感测控制系统,主要包括硬件设计和软件设计两部分。硬件核心部分含LED感测终端节点、ZigBee中心模块(协调器)的设计,搭载两级高性能MCU在网络边缘提供边缘计算能力,具体主要涉及电源电路、串口电路、LED驱动电路、接口电路的设计等。软件部分首先是在IAR开发环境和Z-Stack协议栈的基础上完成对协调器控制程序、终端节点控制程序的设计调试;最后基于VS软件开发实现了具有交互式GUI界面的监控软件,实现了远程访问和控制功能。

关键词:无线传感网; LED 灯; 光强感测; 边缘计算

0 引言

随着物联网、移动互联网、云计算等新一轮 信息技术的快速发展和深入应用,建设"智慧园 区"已成为发展趋势。而路灯照明系统作为"智 慧园区"的主要组成部分,也正在向智能照明和 绿色照明方向快速发展。传统的路灯照明采用定 时或手动控制,既不能灵活控制路灯照明,也不 能实现节能减排的目标,局限性越来越明显。

在此背景下,本研究依据智慧园区环境路 灯照明系统设计应用,主要采用高性能 MCU (STC1235A)和 ZigBee 模块设计实现单灯控制 器,该控制器具有无线通信功能并且功耗较低。 基于 ZigBee 无线技术,具有网络自组织能力、 高扩展性、稳定性和可靠性。此外,它还能接 收和执行各种命令,并将结果传送给协调器^[1], 同时两级高性能 MCU在网络边缘提供边缘计算 能力。这种设备具有独立控制能力,能够控制 参数与管理软件同步。系统操作简单、安装方 便、成本较低、管理容易。

1 系统总体设计

系统总体设计如图1所示。



图 1 系统总体设计

LED 路灯控制端通过光强度传感器采集 LED 灯的光照强度的信息,然后经过 ZigBee 无 线网络将信息传输到边缘网关,数据经初步处 理后通过 Wi-Fi/GPRS 无线通信网(外网)传送至 网络监控中心;接收到数据后通过监控中心的 进一步处理判断,把监控中心控制命令再经边 缘网关发送到控制节点,从而实现 LED 路灯的 远程控制。在无外网部署时,也可仅由边缘网 关通过串口连接本地监控 PC 机,再经 ZigBee 无 线网络实现光强感测及 LED 灯的无线控制;外 部网络恢复后,系统可自动同步本地监控 PC 机 备份的数据至网络监控中心服务器,这就保证 了数据的有效性、正确性和完整性,提高了系

基金项目: 甘肃省高等学校创新基金项目(2021B-342);甘肃省教育科学"十四五"规划课题(GS[2023]GHB1468) 作者简介: *通信作者:柴西林(1984—),女,甘肃兰州人,硕士,副教授,主要从事嵌入式系统和物联网、模式识别与智能 系统研究,E-mail:chai.xilin@163.com; 邵照勇(1985—),男,山东菏泽人,本科,高工,主要研究方向为自动化与仪器仪表

收稿日期: 2023-10-21 修稿日期: 2023-11-27

统的可靠性。因此,本文主要探讨经本地监控 PC机-协调器+网关-LED灯感测终端节点实现 的系统技术方案。

2 系统硬件设计

2.1 硬件设计框图

硬件核心模块设计主要由LED 感测终端节 点、边缘网关和监控中心等组成,如图2所示。 设计中主要融合了计算机技术、ZigBee 无线通 信技术以及单片机技术等,LED 灯控制系统通 过 ZigBee 终端节点和边缘网关向监控中心发送 光强传感及 LED 灯实时状态信息,监控中心(上 位机)记录并管理接收到的数据,并据此下达 LED 灯控制命令。



图 2 系统硬件设计

2.2 硬件电路分模块设计

2.2.1 LED 感测终端节点模块电路设计

该模块电路由 DL-20 无线串口模块、GY-30 光强传感器、STC12C5A60S2 芯片、LED 灯驱动 电路和电源等最小系统电路组成。终端节点通 过光强度传感器把采集的 LED 灯光照强度信息, 经过 ZigBee 无线网络到边缘网关模块,再把信 息传输到监控中心。终端节点模块电路设计如 图 3 所示。

2.2.2 边缘网关模块

该模块电路是由ZigBee模块、STC12C5A60 S2处理芯片及其最小系统电路、各通信接口等 组成。通过ZigBee无线自组网,可将终端节点 采集的数据传送到上位机监控中心,上位机进 行数据分析处理,并且发送控制命令,继续通 过ZigBee无线网络传送控制命令到LED感测终 端节点,从而实现路灯的远程控制。搭载的高 性能处理器不仅可在网络边缘提供边缘计算能 力,同时作为边缘网关可支持ZigBee无线网络 协议与Wi-Fi/4G/5G等网络之间的转换^[2]。该模 块电路设计如图4所示。







图4 边缘网关模块电路设计

2.3 ZigBee 无线网络设计

以应用 ZigBee 协议栈进行数据传输为目标,在无线传感器网络中,大多数的传感器节点负责数据采集工作,如温度、湿度、压力、烟雾浓度等信息,以协调器为中枢建立 ZigBee 无线网络,终端节点通过自组网自动加入无线网络,然后终端节点按周期采集光强数据并发送给协调器,协调器接收数据后通过串口发送给主机^[3]。实现过程如图5所示。



对于协调器来说,只需通过串行接口将接收 到的数据发送到监控主机即可;对于终端节点来 说,需要定期收集测量数据,可通过读取相应传 感器的测量值获得所需的数据^[46]。使用 ZigBee 协议栈进行无线传感器网络开发时,将传感器操 作有关的函数放在 APP 目录下,如图6所示。



图 6 ZigBee协议栈无线传感器网络开发

3 系统软件设计

系统核心模块协调器网关、LED 路灯传感 节点控制程序流程图分别如图7、图8所示。



图 7 协调器控制程序流程图



图8 终端节点控制程序流程图

4 系统软硬件联合测试

系统应用程序主要运用VS软件进行编程, 设计了监控系统的控制界面,实现了LED路灯 实时状态、光强感测数据等信息的显示及控制 模式的选择等功能^[78]。将控制命令发送给处理 模块,用户根据监控界面上的有关信息进行相 应的现场LED路灯的控制,同时以交互式界面 的形式显示给用户。

无线网通过ZigBee 连接好LED灯传感节点 控制端和边缘网关,再通过串口连接本地监控 调试PC,运行程序,选择自动模式,设定光强 阈值为1000LX,当实际光线强度小于设定阈值 时,LED灯为开启状态,其GUI界面及实际控 制效果如图9、图10所示;当实际光线强度大 于设定阈值时,LED灯为关闭状态,其GUI界 面及实际控制效果如图11、图12所示。也可选 择手动控制模式。



图 9 LED灯为开启状态的GUI界面



图 10 LED 灯为开启状态实际效果



图 11 LED灯为关闭状态的GUI界面



图 12 LED 灯为关闭状态实际效果

5 结语

本设计主要融合了计算机技术、ZigBee无 线通信技术以及单片机技术等,组建了智慧园 区LED路灯监控系统,主要设计并实现了LED 感测终端节点、边缘网关和监控中心(上位机) 等模块软硬件。系统通过ZigBee终端节点和边 缘网关向监控中心发送光强传感及LED灯实时 状态信息,监控中心记录并管理接收到的数据, 并据此下达LED灯控制命令。在完成了系统硬 件和无线网络组建的基础上,设计了交互式的 监控中心软件,通过GUI界面实现了友好的人 机交流,根据需要可以选择自动控制和手动控 制两种模式。终端节点模块化设计方便园区大 规模部署,可根据园区实际需求,组网后通过 边缘网关实现单灯控制、区域控制和全局控制; 稍作改进还可以扩展应用于LED景观灯系统。

参考文献:

- [2] 边缘计算产业联盟.边缘计算参考架构3.0,边缘计 算产业联盟白皮书[R/OL].2018.http://www.ecconsortium.org/Lists/show/id/334.html.
- [3] 吴晓. 基于 ZigBee 与 Labview 技术的室内检测软件 设计与实现[D]. 苏州:苏州大学,2017.
- [4] 张秀娟.基于ZigBee的LED灯控制系统设计[J].
 科技创新与应用,2022,12(1):81-84.
- [5] 魏志荣,梁志勋,赵家祺,等.基于无线传感器网络的智能LED景观灯控制系统的设计与实现[J].光源与照明,2022(1):73-76.
- [6] 李旭亮,邓国强.基于ZigBee+GPRS智能控制的 LED 路灯绿色照明低碳经济新技术[J].照明工程 学报,2010,21(z1):78-81.
- [7] 许梦羊,罗素芹,陈国栋,等.基于ZigBee和GPRS 感知天气的太阳能LED路灯控制系统[J].照明工 程学报,2017,28(5):54-58.
- [8] 刘勃妮,肖军,刘洲洲,等.智能太阳能LED路灯控 制系统的研究[J].自动化技术与应用,2022,41
 (1):27-29.

Design of wireless measurement and control system for LED street lights in smart park based on ZigBee

Chai Xilin^{1*}, Shao Zhaoyong²

(1. School of Electronic and Electrical Engineering, Lanzhou Petrochemical University of Vocational Technology, Lanzhou 730060, China; 2. Petro China Lanzhou Petrochemical Company, Lanzhou 730060, China)

Abstract: Based on the ZigBee wireless sensor network, combined with the construction requirements of the intelligent park, the LED street light intensity sensing control system is designed and realized, which mainly includes two parts: hardware design and software design. The hardware core part includes the design of LED sensing terminal node and Zigbee central module (coordinator). It is equipped with two-level high-performance MCU to provide edge computing capability at the edge of the network, it mainly involves the design of power circuit, serial port circuit, LED driver circuit, interface circuit, etc. Its software part first completes the design and debugging of the coordinator control program and terminal node control program based on the IAR development environment and Z-Stack protocol stack. Finally, a monitoring software with an interactive GUI interface was developed and implemented based on VS software, achieving remote access and control functions.

Keywords: wireless sensor network; LED lights; light intensity sensing; edge computing